Instruments and Observing Methods
Report No. 131

WMO Solid Precipitation Intercomparison Experiment (SPICE) (2012 - 2015)

R. Nitu, Y.-A. Roulet, M. Wolff, M. Earle, A. Reverdin, C. Smith, J. Kochendorfer, S. Morin, R. Rasmussen, K. Wong, J. Alastrué, L. Arnold, B. Baker, S. Buisán, J.L. Collado, M. Colli, B. Collins, A. Gaydos, H.-R. Hannula, J. Hoover, P. Joe, A. Kontu, T. Laine, L. Lanza, E. Lanzinger, GW Lee, Y. Lejeune, L. Leppänen, E. Mekis, J.-M. Panel, A. Poikonen, S. Ryu, F. Sabatini, J. Theriault, D. Yang, C. Genthon, F. van den Heuvel, N. Hirasawa, H. Konishi, H. Motoyoshi, S. Nakai, K. Nishimura, A. Senese and K. Yamashita

WEATHER CLIMATE WATER

WORLD METEOROLOGICAL ORGANIZATION

This publication is available in pdf format, from the WMO Library website:

https://library.wmo.int/opac/

Chair, Publications Board
World Meteorological Organization (WMO)
7 bis, avenue de la Paix
P.O. Box 2300
CH-1211 Geneva 2, Switzerland

Tel.: +41 (0) 22 730 8403
Fax: +41 (0) 22 730 8040
E-mail: Publications@wmo.int

NOTE

The designations employed in WMO publications and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of WMO concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The mention of specific companies or products does not imply that they are endorsed or recommended by WMO in preference to others of a similar nature which are not mentioned or advertised.

The findings, interpretations and conclusions expressed in WMO publications with named authors are those of the authors alone and do not necessarily reflect those of WMO or its Members.

This publication has been issued without formal editing.

FOREWORD

The World Meteorological Organization (WMO) through its Members, recognizes that in this era of Big Data, and of the increased complexity and diversity of observing technologies, we must not forget that a weather datum is the result of measurements and complex data processing, and that the traceability and understanding of measurements are critical to ensuring that data is fit-for-purpose and that they support the provision of environmental intelligence for government and business sector decision makers.

The Commission for Instruments and Methods of Observation (CIMO), as one of the eight Technical Commissions of WMO, focuses its work on ensuring the provision of accurate weather and climate data by promoting and facilitating the international standardization, compatibility, and sustainability of meteorological measurement systems used by Members within the WMO Integrated Global Observing System, and the traceability and consistency of their observations and data. CIMO plays a leadership role in facilitating the evolution of measuring methods and equipment critical to Member's needs, by fostering innovation and engagements with meteorological instrument manufacturers, the scientific community, and other developers, while remaining truthful to the principles of traceability and understanding of observations.

To achieve its mission, CIMO supports and coordinates joint initiatives of Members with respect to observing systems, to meet their critical needs in support of a broad range of hydro-meteorological and climate applications. This is the only approach to ensure results that exceed what each Member would achieve, alone, for ensuring the consistency of data, across borders, at national, regional, and global level.

In the context of the increasing diversity of observing technologies, CIMO has organized several intercomparisons of existing and emerging observing technologies, to inform Members about the impact and opportunities offered by advances in technology. Their results have been critical in ensuring the sustained reliability of observing instruments and systems, addressing the need to align the technology developments with the requirements of users of data, and the challenges of operational environments, across the globe, as well as supporting the capability building in developing and least developed countries, to close the technology and knowledge gap between them and the developed countries.

Solid precipitation and snow on ground are among the most complex and challenging parameters to observe, measure, and report on. The significant evolution of measurement techniques over the last three decades, as well as the emergence of new applications (for example, climate change, nowcasting, water supply budgets, avalanche forecast and warnings, satellite ground validation) have increased the demand for precipitation and snow data with increased temporal and spatial resolutions. Aligned with its mission, CIMO assumed in 2010, the lead role in organizing an internationally coordinate intercomparison for assessing the impact of automation on the measurement of snowfall, snow depth and solid precipitation in cold climates, the WMO Solid Precipitation Intercomparison Experiment (WMO SPICE).

SPICE has been one of most inclusive, most complex, and most internationally engaging intercomparison of CIMO. Involving the participation of well over 100 experts from

sixteen countries from five continents, with the participation of more than twenty instrument manufacturers, and with intercomparisons organized on twenty sites hosted by diverse organizations (operational, research, private sector), SPICE has provided a model for international collaboration and engagement.

The results of SPICE are impressive and cover a large range of topics. They provide comprehensive insight on the impact of measurement techniques and the influence of environmental factors, on the quality of precipitation and snow data. The results documented in this report include the detailed characterization of current and emerging measurement instruments, together with recommendations on how to address noted limitations and biases, making them very relevant to the operational and research programs of Members.

In addition to this SPICE report, as published by CIMO, a number of scientific papers have been already published in reputable scientific journals, with specific results from the SPICE intercomparison, thus offering a wealth of information and insight for all users of instrumentation and solid precipitation and snow data. It is clear that more work needs to be done to translate all recommendations made in applicable guidelines, on improvements to instruments and systems, as well as on the harmonization of data. I am confident that CIMO in collaboration with other Technical Commissions of WMO, Members, and manufacturers, will continue working together to address these recommendations, and ensure the expected sustainability and quality of solid precipitation and snow data.

Perhaps, the most important legacy of SPICE is the investment in people made by all those who contributed to this project. SPICE has fostered an incredibly closely knitted global expert community, which has brought into its fold a large number of young scientists and instrument experts. This is the strongest assurance that the principles of data quality, sustainability, standardization, and traceability will continue to be upheld, as the technologies evolve and diversify and the demands for data increase.

I wish to express my sincere gratitude and that of the WMO Commission for Instruments and Methods of Observation to all SPICE contributors both within WMO CIMO experts as well as the private vendors community. In particular I want to congratulate Mrs Rodica Nitu who has proven a wonderful leadership since 2010 for what has represented by far one of the most challenging inter-comparison ever in CIMO. My gratitude goes also to M. Yves-Alain Roulet for his significant involvement in the final drafting of the current report as well as the WMO Instruments and Methods of Observation Unit staff for their invaluable continuous support in SPICE.

(Prof. B. Calpini)

President

Commission for Instruments and Methods of Observation

Instruments and Observing Methods

Report No. 131

Solid Precipitation Intercomparison Experiment

2012-2015

Project Lead: R. Nitu (Canada)

Final Report Coordinator: Y.-A. Roulet (Switzerland)

Precipitation Data Analysis Leads: M. Wolff (Norway), R. Rasmussen (USA), E. Lanzinger (Germany)

Snow on the ground Data Analysis Leads: C. Smith (Canada), S. Morin (France)

Data Analysis Team: M. Earle (Canada), A. Reverdin (Switzerland),

J. Kochendorfer (USA), K. Wong (Canada)

World Meteorological Organization

2018

Table of contents

ACKNOWLEDGEMENTS

The SPICE project was made possible by the dedication, commitment, collaboration, and hard work of a large, multi-disciplinary, multi-stakeholder, international team representing fifteen WMO Member countries. Together, the team ensured that the project principles were reflective of the current operational, services, and development needs, and that the delivery of results remained pragmatic and achievable, in spite of geographical differences, many setbacks, changing circumstances, and the numerous competing priorities faced over the five years of this project.

The collaboration established during SPICE is a very fertile ground for sustaining an international community of practice, which brings together internationally recognized experts, young scientists and engineers.

As a common practice for WMO intercomparison, an independant International Organizing Committee (IOC) was elected, nominated by WMO:

- Rodica Nitu (Canada)
- Haihe Liang (China)
- Yves-Alain Roulet (Switzerland)
- Francesco Sabatini (Italy)
- Eckhard Lanzinger (Germany)
- Jordy Hendrikx (New Zealand)
- Valery Vuglinski (Russia)
- Bruce Baker (USA)

The following is the alphabetical list, by country, of organizations, scientists, engineers, data managers, and support technical staff without whom this project would not have been possible.

AUSTRALIA (Guthega Dam)
Snowy Hydro Ltd.: Shane Bilish, Andrew Peace

CANADA (Bratt's Lake, CARE, Caribou Creek)
Environment and Climate Change Canada: Jeff Anderson, Lauren Arnold, Nelson Cabeceiras, Bruce Cole, Michael Earle, Jeffery Hoover, Charmaine Hrynkiw, Paul Joe, Eva Mekis, Hagop Mouradian, Rodica Nitu, Amber Peterson, Sorin Pinzariu, Phil Raczynski, Amal Samanter, Craig Smith, Erica Tetlock, Emilia Vaserbakh, Stephnie Watson , Emma Wattie, Kai Wong, Daqing Yang, Lillian Yao

CHILE (Tapado)
Centro de Estudios Avanzados en Zonas Áridas (CEAZA): Shelley McDonell

CHINA (no SPICE site)
Chinese Meteorological Administration: Haihe Liang

FINLAND (Sodankylä)
Finnish Meteorological Institute: Markku Ahponen, Riitta Aikio, Osmo Aulamo, Henna-Reetta Hannula, Mika Hyötylä, Anna Kontu, Timo Laine, Leena Leppänen, Ilkka Mikkola, Jyrki Mattanen, Antti Poikonen, Veikko Postila, Niina Puttonen, Anita Sassali, Sami Suopajärvi, Riika Ylitalo

FRANCE (Col de Porte)
DREAL Auvergne Rhône Alpes, SPC Alpes du Nord, Grenoble: Vincent Bontemps, Arnaud Coupin, Alain Gautheron

EDF-DTG, Grenoble: Paul Carrier, Frédéric Gottardi

Univ. Grenoble Alpes, Irstea, UR ETNA, Grenoble: Hervé Bellot, Florence Naaim, Xavier Ravanat

Univ. Grenoble Alpes, Université de Toulouse, Météo-France, CNRS, CNRM, Centre d'Etudes de la Neige, Grenoble: Marie Dumont, Erwan Le Gac, Yves Lejeune, Samuel Morin, Jean-Michel Panel, Laurent Pezard, Lucas Rihn, Jacques Roulle

Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP, Institut des Géosciences de l'Environnement, Grenoble: Romain Biron, Thomas Condom, Jean-Paul Laurent, Laurent Arnaud, Christophe Genthon, Luc Piard, Ghislain Picard

GERMANY (no SPICE site)
Deutscher Wetterdienst: Eckhard Lanzinger

ITALY (Forni Glacier)
ARPA Lombardia: Eraldo Meraldi

CNR-IBIMET, Institute of Biometeorology: Francesco Sabatini

Ev-K2-CNR: Gian Pietro Verza

University of Genoa, Department of Civil, Chemical and Environmental Engineering: Luca Lanza

University of Genova, Electrical, Electronics and Telecommunication Engineering and Naval Architecture Department: Matteo Colli

University of Milano, Department of Environmental Science and Policy: Roberto Sergio Azzoni, Chiara Compostella, Guglielmina Adele Diolaiuti, Maurizio Maugeri, Antonella Senese

Italian Meteorological Service, RESMA: Emanuele Vuerich

Regional College of Alpine Guides of Lombardy

JAPAN (Joetsu, Rikubetsu)
Japan Meteorological Agency: Kenji Akaeda, Kotaro Bessho, Kohei Honda, Yoshihisa Kimata, Yoshihiko Tahara, Akihito Umehara

National Research Institute for Earth Science and Disaster Resilience: Sento Nakai, Katsuya Yamashita, Kotaro Yokoyama

National Institute of Polar Research: Naohiko Hirasawa, Hiroyuki Konishi

Osaka Kyoiku University: Hiroyuki Konishi

KOREA, REPUBLIC OF (Gochang)
Korea Meteorological Administration: Eun-Jin Choi, Sang-Ok Han, Bok-Haeng Heo, Hee-Jin In, Deuk-Kyun Rha, Jae-Gwang Won

Kyungpook National University, Department of Astronomy and Atmospheric Sciences: GyuWon Lee, SooRok Ryu, Ik-Soo Son

National Institute of Meteorological Sciences, Observation Research Division: Jeong-Eun Lee, Eunha Lim, Hyelim Kim, Young-San Park

NEPAL (Pyramid International Laboratory Observatory, operated by Italy)
University of Milano, Department of Environmental Science and Policy: Antonella Senese

Ev-K2-CNR, Italy: Gian Pietro Verza, Elisa Vuillermoz

NEW ZEALAND (Mueller Hut)
National Institute of Water and Atmospheric Research: Jeremy Rutherford, Andrew Williams, Christian Zammit

NORWAY (Haukeliseter)
Norwegian Meteorological Institute: Ragnar Brækkan, Ketil Isaksen, Knut Ove Nygård, Ole-Jørgen Østby, Mareile Wolff

Scanmatic AS, Staubø: Daniel Hatlevoll

POLAND (Hala Gasienicowa)
Institute of Meteorology and Water Management: Maciej Karzynski

RUSSIAN FEDERATION (Valdai, Voljskaya)
Roshydromet: Arkadiy Koldaev, Anton Tomofeev, Valery Vuglinski

SPAIN (Formigal)
Spanish National Meteorological Agency, Aragon Regional Office, Zaragoza: Javier Alastrué, Beatriz Arilla, Isabel Arranz, Samuel Buisán, José Luís Collado, Jesús Ezquerra, Jesús Laguía, Pablo López, Rafael Requena, Ismael Sanambrosio, Amadeo Uriel

Spanish National Meteorological Agency, Headquarters, Madrid: Manuel Gil, María López, Fortunato Marquez, Juan Ramón Moreta, José María Romero

SWITZERLAND (Weissfluhjoch)
Federal Office of Meteorology and Climatology, MeteoSwiss: Jean-Marc Aellen, Serge Brönnimann, Jacques Grandjean, Bertrand Henchoz, Audrey Reverdin, Yves-Alain Roulet, Yann Salvi, Floor van den Heuvel

Snow and Avalanche Research Institute: Charles Fierz, Franz Herzog, Christoph Marty, Marc Ruesch

USA (Marshall)
National Center for Atmospheric Research: Andrew Gaydos, Al Jachcik, Scott Landolt, Ania Naoumova, Roy Rasmussen, Julie Theriault

National Oceanic and Atmospheric Administration: Bruce Baker, John Kochendorfer

WMO
Global Cryosphere Watch: Barry Goodison

WMO Observing Systems Division: Isabelle Rüedi, Roger Atkinson, Miroslav Ondráš

EXECUTIVE SUMMARY

The Solid Precipitation Intercomparison Experiment (SPICE) was conducted as an internationally coordinated project, initiated and guided by the Commission for Instruments and Methods of Observation (CIMO) of the World Meteorological Organization (WMO). The SPICE field experiments took place between 2013 and 2015, with a preparatory stage during the winter of 2012/13.

SPICE was carried out as a major international effort, and has been remarkable for the diversity of organizations which hosted SPICE tests, contributed with instruments, and were engaged in the data analysis and the derivation of results. In addition to National Meteorological and Hydrological Services, research organizations, academia, and the private sector played active roles and made unique contributions. Field experiments were conducted at twenty sites located in fifteen countries, on all continents except Africa and Antarctica, as outlined in Section 2 of this report. The instrument manufacturing community made a significant contribution to SPICE, as more than twenty instrument manufacturers provided instruments measuring precipitation amount, snow depth, and snow water equivalent. Each instrument model was tested on one or more sites in different climate regimes and over a large range of environmental conditions, providing a solid foundation for the results presented in this report.

As presented in Section 1 of this report, the core objectives of SPICE focused on providing guidance on the use of automatic instruments measuring and reporting solid precipitation, snow depth, and snow water equivalent in cold climates, and on the related data aspects over various time periods (i.e. minute, hour, day, season).

Due to changing circumstances at the local level, the tests on some sites were conducted only partially. In spite of this outcome, the significant engagement provided an effective mechanism for increasing the awareness of the complexity of measuring precipitation in cold regions, and has facilitated building linkages between experts in different parts of the world, leading to the creation of an informal but strong community of practice.

Reflective of the complexity of the intercomparison, this report has been structured to account for the relevant new knowledge acquired through the Intercomparison. In preparation of this final report, specific, topic-driven, results have been published, already, in peer reviewed journals. The most notable is the Inter-Journal Special issue: “WMO-SPICE and its applications”, published in Hydrology and Earth System Sciences and other relevant journals of the European Geosciences Union, https://www.hydrol-earth-syst-sci.net/ special_issue400_78.html.

The results of the intercomparison allowed the team to address most of the project objectives. These include the traceability and the configuration of field references using automatic instruments for the unattended measurement of solid precipitation and snow on ground in cold climates. A second set of results published in this report are focused on recommendations for adjustments that account for the undercatch of solid precipitation due to gauge exposure, and are presented as a function of data available at operational sites, such as wind and temperature.

A third set of results are focused on the instruments tested. For each instrument type and model, the report includes individual Instrument Performance Reports (IPRs), provided in Annex 6. These include specific assessments of individual instrument performance relative to a field reference, on the configuration and use of the instrument, on notable operational aspects, as well as recommendations for potential improvements.

Specific operational challenges, such as the use of wind shields, the heating of instruments, addressing snow capping, the detectability of light precipitation, and the challenges of measuring snow on the ground, have been assessed throughout SPICE, and are documented in this report. This documentation is reflective of the variety of climates where tests were conducted and the collective experience of participants

Overall, the SPICE final report highlights three aspects which made SPICE possible, which are important takeaways for any future initiative, and are recommended as foundational principles for operational programmes supporting multiple applications, concurrently.

First is the fact that complex questions regarding Earth System observations can be addressed only through international, multi-disciplinary, multi-stakeholder collaboration; as the environment knows no borders, it is critical to ensure broad ownership of results and understanding of limitations of observations, and to ensuring consistency and availability of information.

Secondly, the SPICE experience and outcomes demonstrate the critical role and the necessity of standardization for generating and exchanging relevant and consistent data and results; this refers, but is not limited to, instrument configuration, instrument siting and exposure, maintenance and operations, and, of equal importance, the metadata vocabulary and semantics, and the data models used.

The third aspect is related to the need for accessibility and distribution of information on methods, the data traceability, and the accessibility to algorithms processing individual measurements in data over various time scales; the value of data increases with its full understanding.

This report, however, does not provide a relative comparison of the instruments tested. As tests were conducted on multiple sites, in a variety of climate conditions, the results have shown that the environmental conditions, the instrument and site configurations, and the operational and maintenance aspects play a critical role in the quality and consistency of data. The team felt that a direct comparison of instruments under test would not be representative of the complexity of factors influencing the measurements, and instead focused on reporting on individual performances and provided instrument-based recommendations for addressing noted issues.

The findings of SPICE have allowed the team to put forward a comprehensive list of conclusions and recommendations, as presented after this summary. These are addressed to a broad audience, including manufacturers, operators of instruments, and users of data. The report outlines recommendations for future work, inviting WMO and other interested parties to consider follow-up initiatives to pilot the findings and the recommendations of this report, and ensure that they have a real impact, in practice.

The project team (listed on the front page of the report and in the acknowledgment section above) would like to thank their home organizations, which supported their engagement and the field tests, the WMO, the instrument providers, and all participants for their significant support in carrying out SPICE.

CONCLUSIONS AND RECOMMENDATIONS

This chapter summarizes the outcomes and conclusions of SPICE and provides a list of recommendations, including for future work to advance the understanding and use of solid precipitation datasets. This summary is presented by reiterating the objectives of SPICE as defined at the beginning of the project (identified by bold and italicized text below) and outlining the corresponding progress and remaining gaps. More detailed results and recommendations are disseminated throughout the report, typically at the end of each chapter. Based on experience gained during SPICE, a set of recommendations for future WMO intercomparisons is also provided at the conclusion of this chapter.

Overall, the individual national contributions to SPICE have enabled the development of significant observational and data management infrastructure, specific to the operational configurations of each contributing country. It is strongly recommended that SPICE sites, together with their data collection and management capabilities and the expert capacity developed over the course of this project, are maintained to facilitate the long term assessment of national precipitation and snow datasets. This will allow for the assessment of changes in these datasets over time and the impact of changes in the methods of observation or instrument configuration.

As SPICE was a demonstration project of the Global Cryosphere Watch (GCW), it is suggested that the existing SPICE sites are integrated into the GCW Surface Observing Network as CryoNet stations. This will facilitate the contribution of these stations to further development and testing of operational best practices for the measurement of solid precipitation, and the implementation of many of the recommendations from this report, as a joint effort of the international expert community. They would provide excellent sites for global evaluation/validation of satellite products and model outputs.

SPICE has fostered the development of a new generation of young experts in all contributing countries, and it is strongly recommended that their engagement continues in relevant projects (e.g. as part of the collaboration within the GCW framework) to support the building of expert capacity and a community of practice, internationally.

I. **Recommend appropriate automated field reference system(s) for the unattended measurement of solid precipitation. Define and validate one or more field references using automatic instruments for each parameter being investigated, over a range of temporal resolutions (e.g. from daily to minutes).**

This objective was achieved through the assessment and comparison of several field reference configurations, with emphasis on the traceability of reference data and linkages with previous work, notably the previous WMO solid precipitation measurement intercomparison (Goodison et al., 1998). The following results were obtained, as presented in Section 3.2 of this report:

- Reference data for solid precipitation were derived from configurations using automatic instruments as a composite, rather than as a single measurement dataset, incorporating precipitation amount, precipitation type, and the environmental conditions over the specified assessment interval. The duration of the assessment interval was found to be an important factor impacting the reference dataset and subsequent analyses.
- Comparison between the manual bush gauge (R0) and DFIR (R1) at the Valdai SPICE site showed that, on average, the bush gauge caught 5-6% more than the DFIR for snow and mixed precipitation. This comparison was done on a daily scale, only.

- Comparison of automatic bush gauge (R0aG) and DFAR (R2G) configurations at Caribou Creek showed that the mean catch of the DFAR exceeded that of the automatic bush gauge by 7%. These results were derived using 60 minute event datasets for two winter seasons.
- Comparison of DFAR (R2G) and DFIR (R1) measurements at the CARE SPICE site over three winter seasons showed that the reported precipitation amount from the automatic configuration (R2) was about 7% below the corresponding manual (R1, Tretyakov gauge) value. The measurement interval used for this assessment was daily, similar to that used for the comparison of R0 and R1 reference configurations.
- The impact of the type of automatic gauge used in the DFAR was assessed through the comparison of parallel DFAR (R2) reference configurations at three SPICE sites: CARE (Canada), Bratt's Lake (Canada), and Gochang (Rep. of Korea). Each of these sites had both an R2 configuration with a Geonor T-200B3 (R2G) and an R2 with an OTT Pluvio^2 (R2P), reflecting the decision that each SPICE site could use either gauge type in their field reference configuration. This assessment demonstrated similar performance for these two reference systems, indicating that the reference data are not critically influenced by the specific type of automatic weighing gauge used.
- For the measurement of snow depth, the SPICE results demonstrated the feasibility of using multiple automatic instruments to derive a composite reference dataset as an alternative to manual measurements. This method is applicable over shorter time intervals than those typically used for manual measurements (e.g. daily or twice per day). It was also shown that large spatial variability in snow depth can often make instrument intercomparisons more difficult to assess.
- For the measurement of snow water equivalent, SWE, the results of SPICE show that more work is required to define a reliable automated field reference for use in various climate conditions.

The resulting recommendations are as follows:

- Further assessments should be conducted for the field reference configurations recommended by SPICE, including in other climate regimes and over longer periods. To achieve this, it is recommended to set up "super sites", with all references recommended in this report collocated, to allow for further assessment and characterization of their relative performance for different climate regimes, reporting intervals, operating conditions, and field limitations. This would support the need for traceability of observations in the absence of a primary standard.
- For operational networks, operators are strongly encouraged to configure and maintain continuously at least one comprehensive test site with high quality measurements, e.g. a DFAR, and including precipitation detectors, where it would be possible to also operate their national gauge(s) in the configuration(s) used operationally. This would enable networks to characterize the measurement uncertainty of national gauge(s) and to provide recommendations for long term network management based on solid precipitation data quality.
- Wind speed and direction measurements at gauge height and at 10 m, and air temperature measurements at 2 m (to be used for discriminating snow, mixed and rain events), are

recommended as the minimum standard ancillary measurements at operational sites where solid precipitation is monitored. As the use of temperature for determining the precipitation type has limitations, it is recommended that operational sites use precipitation type sensors to enable the accurate identification of precipitation type, where possible. It is strongly recommended that future work is undertaken to characterize the ability of these instruments to determine precipitation type accurately, in all climate types, and that further improvements are made to address the limitations of sensors outlined in this report.

- The results in this report were based on site event datasets with a 30 minute event definition. Operationally, different processing intervals are used for different applications. Accordingly, further work is recommended to assess if the DFAR is a reasonable reference configuration for other time scales, and to assess the impact of different processing intervals on the quality and traceability of data.
- The importance of a robust data quality control scheme has been demonstrated (see Section 3.3.2 for the description of the QC procedures developed for SPICE). It is recommended that, for operational purposes, quality control protocols are developed based on the results in this report, to ensure that operational data meet the same, or higher, quality standards.
- For the measurement of snow depth, future work is recommended to further develop and refine the derivation of reference datasets using multiple automatic instruments as new technologies become available. Future work is also recommended on the use of camera images for the semi-automatic derivation of snow depth data from snow stakes. Additionally, further work is recommended to address the instrument specific improvements proposed in the individual instrument performance reports (Annex 6).
- Regarding the measurement of SWE, as no automatic measurement could be validated as a reference, it is recommended to continue to use manual measurements as the reference for instrument intercomparisons, recognizing the inherent errors and biases in the various manual SWE measurement devices and techniques.

II. Assess/characterize automatic systems (both the hardware and the associated processing) used in operational applications for the measurement of Solid Precipitation (i.e. gauges as "black boxes"):

a. Assess the ability of operational automatic systems to robustly perform over a range of operating conditions;

The sensors under test were evaluated in various climates and operational conditions. Instrument-specific results and recommendations are presented in the instrument performance reports (see Annex 6), which include considerations of operational performance and maintenance requirements based on the experiences of site managers during the SPICE field campaign. Most of the instruments tested were proven to be robust and reliable if operated within their recommended operating conditions (as defined in user manuals).

b. Derive adjustments to be applied to measurements from operational automatic systems, as a function of variables available at an operational site: e.g., wind, temp, RH;

Transfer functions were developed to adjust precipitation measurements from operational automatic systems using only wind speed measurements, or both wind and temperature

measurements. The methodology for the derivation and application of transfer functions is outlined in Section 3.7 and peer-reviewed publications referenced in this chapter. The transfer functions developed and tested using SPICE data showed promising results for most of the sites. It should be noted that sites in mountainous regions with complex terrain, like Formigal and Weissfluhjoch, may be subject to other phenomena such as turbulent flows due to increased surface roughness. This will affect gauge catch efficiency, and the assumptions made in deriving the transfer functions will not be fully applicable. Here is especially relevant, that the wind measurements very likely not represent the conditions at the gauge due to the complexity of the landscape, thus introducing an error to the transfer function caused by a less exact wind measurement.

Overall, the results presented in this report demonstrate the potential for the derivation of universal transfer functions for a given precipitation gauge configuration that can be applied to measurements in different climate regimes, while recognizing the limitations of their use for measurements conducted in complex terrain.

It is recommended to apply the methods and results from the SPICE project to national and regional precipitation data/measurements in various climatic regimes. Further work is needed to derive and use tailored transfer functions for specific environments, e.g. mountainous regions. Additionally, it is recommended that operational programs actively disseminate information on the availability of adjustment procedures that could be applied to raw (uncorrected) operational measurements of solid precipitation, and conduct targeted assessments of the impact of their application, as recommended by SPICE.

c. *Make recommendations on the required ancillary data to enable the derivation of adjustment procedures to be applied to data from operational sites on a regular basis, potentially, in real-time or near real-time;*

See II.b above. It is recommended that all solid precipitation measurement programs and projects ensure that the required ancillary measurements used for the derivation of adjustment procedures are available operationally and are disseminated with the precipitation data. As noted above, any site reporting solid precipitation needs to include, at minimum, the measurement of wind speed at gauge height or at 10 m and temperature at 1.5 to 2 m above ground, and to the extent feasible, an indication of precipitation type. Note that the specific height of wind speed measurements will impact the transfer function coefficients. Beyond additional instruments measuring ancillary data, it is recommended that operating conditions for precipitation gauges (type of measurement method, presence of a shield, presence/operation of a heating system) are made available together with the raw data, in order to allow for sound application of adjustment techniques. Importantly, data provided by operational systems should include information on whether or not they have been adjusted.

d. *Assess operational data processing and data quality management techniques;*

The SPICE results are based on a standardized data processing approach that includes data quality control, precipitation event selection, and data archival (see Sections 3.3 to 3.5). These results offer a baseline for understanding the performance of different technologies used operationally, in different countries.

It is recommended that additional activities are organized to apply these results to derive time series solid precipitation data and products, for different time intervals, and for specific applications,

testing different approaches for operational data processing and data quality management, including the application of adjustments.

e. Assess the minimum practicable temporal resolution for reporting a valid solid precipitation measurement (amount, snowfall, and snow depth on the ground);

Based on the SPICE assessment results, a 30 minute time interval is a reasonable minimum practicable temporal resolution when using a 0.25 mm threshold for the detection of precipitation.

It is recommended that additional work is organized to assess the practicability of higher temporal resolutions, and the linkages with specific applications (e.g. radar calibration). Higher temporal resolutions are technically achievable, but may be limited in application by the signal-to-noise ratio of the measuring device; this is applicable to both the measurement of solid precipitation and snow depth.

f. Evaluate the ability to detect and measure trace to light precipitation.

Light or trace precipitation (identified by minimum 0.1 mm reference gauge accumulation) has also been studied; the results are presented in Section 4.2.5. For accumulation values below 0.1 mm, it is difficult to distinguish between noise and precipitation (see Section 3.4.3). Specifically, the ability to measure under various climate conditions (e.g. as a function of wind speed), with different instruments, and in different configurations (with and without a wind shield) has been assessed. In general, the weighing gauges showed good ability to detect light precipitation. The use of ancillary data (further work is needed to identify which data and how to measure it) is, nevertheless, highly recommended to identify blowing snow events that can confound the interpretation of results. The assessment of the applicability of transfer functions has been extended to light precipitation events. In general, good agreement was found between the adjusted data and the reference data (see Section 4.2.5.1.3).

Further assessments are recommended in conjunction with the work recommended above (II.a to II.e).

III. Provide recommendations on best practices and configurations for measurement systems in operational environments:

The instruments assessed during SPICE were configured and operated as recommended by manufacturers. This included the use of specific data outputs, including those for which the internal algorithm is proprietary.

As the expertise of instrument users varies significantly, and as instruments are used for a wide range of applications, the manufacturers are strongly urged to support these diverse communities by providing access to both raw data (for advanced users) and derived data products (for general use), along with detailed descriptions of the datasets available and the corresponding algorithms. As demonstrated in SPICE, the measurement of solid precipitation and snow on the ground is complex, due to the influence of environmental conditions, local topography, and characteristics of precipitation. Further relevant improvements are possible only through collaboration between manufacturers and instrument users.

a. On the exposure and siting specific to various types of instruments;

The siting classification from the WMO Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8, 2014), called the CIMO Guide in the rest of the document, (Part I, Chapter

1, Annex 1.B) provides criteria for siting and exposure to optimize the quality of precipitation measurements. Consistent results and conclusions can be drawn from the SPICE experiment and the response obtained from the different sites. The exposure has a direct impact on the maximum wind speed at the gauge location, and, hence, on the wind-induced error that can be expected. As an example to illustrate this, Sodankylä (sheltered site in a forest clearing) showed wind speeds up to 4 m/s, whereas wind speeds over 10 m/s were common in Haukeliseter (very exposed site). The mean catch ratio of a precipitation gauge against the reference is, accordingly, significantly lower for the latter.

A sheltered location will foster accumulation of snow on gauges and surrounding infrastructure (i.e. any snow accumulated will not be removed by wind), and, hence, increase the probability of capping situations (see Section 4.2.1). More analysis is needed to understand this process (e.g. as a function of precipitation type, the threshold for precipitation rate that can cause capping, etc.) and develop ways to mitigate it in an operational context.

At present, there is no siting classification for non-catchment type instruments; however, measurements from a shielded laser disdrometer (rectangular fence with metallic slats, provided by the manufacturer) show lower accumulation than the same instrument without a wind shield (see the IPRs for the Thies LPM in Annex 6). This suggests that the influence of wind on precipitation measurements by optical disdrometers cannot be assessed and adjusted in the same way as for weighing gauges. More work is needed to understand the impact of wind and other environmental parameters on precipitation measurements using non-catchment type instruments.

Similarly, no siting classifications presently exist for snow depth and SWE sensors. Wind can affect snow depth measurements by ultrasonic and laser sensors by impacting the spatial distribution of snow (e.g. blowing and drifting snow), and hence, the spatial representativeness of the measurement. In sheltered locations, snow accumulation on the sensor and sensor infrastructure (mast, boom) may perturb the sensor's field of view and impact it's capability to make a snow depth measurement (see Section 4.2.6).

It is recommended that for all instruments, the recommendations on siting and exposure made in this report, and the influence of local conditions, are included in the CIMO Guide. This will ensure that they are readily available for use by operational programs, leading to improvements to existing siting classifications, or to new siting classification guidelines, where none are presently available.

For non-catchment type instruments, it is recommended that additional assessments should be conducted to better understand the relationships between the reported precipitation amount and the field configuration (e.g. with or without shields), the exposure, and wind speed.

b. On the optimal gauge and shield combination for each type of measurement, for different collection conditions/climates (e.g., arctic, prairie, coastal snows, windy, mixed conditions);

In general, for weighing gauges and tipping bucket gauges, double-shields are recommended over single-shields, and single-shields over unshielded configurations. The benefits of double-shields relative to single-shields in terms of improving catch efficiency are not well demonstrated by the present results, but have been reported elsewhere (Watson et al., 2008; Smith, 2009; Rasmussen et al., 2012; Kochendorfer et al., 2017b). The influence of different climate regimes, which dictate the characteristic snow type(s) and wind conditions, must be taken into account when designing measurement sites. Summary results are presented in Table 4.5 in Section 4.1.1.3.3.1.3 and provide an overview of relative performance in different shield configurations at sites in different climate

regimes. While it is demonstrated clearly that the use of a shield (single or double) will increase the catch efficiency, there are other consequences that must be taken into account, such as increased vibration of the instrument due to wind (see Chapter 4.2.4 on measurements in high wind conditions), noise of the metal slats, and enhanced risk of capping or blowing snow from accumulation on the shield. Detailed results and recommendations can be found in Section 4.1.1 for weighing gauges and Section 4.1.2 for tipping bucket gauges. Detailed results for each instrument can also be found in the respective IPRs (Annex 6).

Transfer functions can be used to determine which sites could benefit most from the use of a shield, provided the necessary ancillary measurements are available (i.e. assessing the influence of gauge undercatch at characteristic wind speed and temperature conditions).

Based on the results documented in this report, it is recommended that the operational aspects related to the use of shields are captured in best practices, distributed to operational programs, and potentially included in the CIMO Guide.

Given the limited results available on the improvements from using double shields and their optimal configuration, it is recommended to further evaluate their impact, as a function of the configurations available.

c. *On instrument specific operational aspects, specific to cold conditions: use of heating, use of antifreeze (evaluation based on its hygroscopic properties and composition to meet operational requirements);*

Heating is recommended for all types of measurements and sensors evaluated in SPICE in order to increase the likelihood of continuous measurement, provided that sufficient power is available to support the operation of heaters. Heating of weighing gauges and tipping bucket gauges will reduce the risk of capping. In the case of tipping bucket gauges, heating is required to melt solid precipitation and enable its measurement. Heating will also prevent the gauge from being completely filled with snow, and depending on the specific heating configuration, can reduce the time lag between snowfall and the measurement. However, there are potential drawbacks of heating that require consideration when designing a measurement site. Overheating can lead to warm air turbulence above the orifice of the gauge, which can reduce the catch efficiency (chimney effect, see Section 3.1.3.4.4). In the case of tipping bucket gauges, it can also evaporate precipitation before it is measured.

At low temperatures, the contents of the buckets of weighing gauges can freeze (partially), resulting in inhomogeneity of the content, which will impact the weight measurement. To prevent this process, the use of antifreeze is recommended (see Section 4.2.3). Generally, a mixture of either ethylene glycol and methanol or propylene glycol and methanol is used. The specific antifreeze solutions for different networks or countries vary depending on the range of temperatures experienced and national regulations for the use and/or disposal of materials (see Annex 7). Since the antifreeze mixture is hygroscopic, it needs to be saturated with water to prevent water absorption from the environment. For gauges without embedded algorithms able to detect and filter out evaporation episodes, an oil layer on top of the bucket content is recommended, in order to prevent evaporation (see Annex 7). The oil layer also mitigates the evaporation of the antifreeze solution.

Where applicable, improvements of gauge shape in order to prevent snow accumulation on the shoulder/ring of instruments should be investigated, in collaboration with manufacturers.

For automatic measurements of snow on the ground, heating can prevent snow accumulation on the mounting infrastructure, and on the sensor itself, in cases of high snowfall rate and low wind speed. Heating can also cause melting and subsequent formation of ice on or in the sensor, which can affect the measurement.

It is recommended that these findings are included in the recommended practices for operational programs (e.g. the CIMO Guide) and are distributed to instrument users.

d. On instruments and their power management requirements needed to provide valid measurements in harsh environments;

In remote locations or harsh conditions there may be limited or no power availability, which may limit the potential for the operation and/or heating of specific sensors. The total power requirements for the optimal operation of sensors is a critical element to be included in the specifications for purchasing instruments and for station configurations, and this may require some tradeoffs. For the automatic measurement of snow depth, for example, optical sensors usually have higher power requirements than acoustic sensors.

Instrument manufacturers are encouraged to address the recommendations made in this report while optimizing the power consumption of instruments. Additionally, it is recommended that manufacturers recommend integrated and efficient, autonomous power solutions to address their instrument power needs, thus expanding the range of operation of these instruments. This would enable their instruments to be used more reliably in remote conditions, and by users who may not have the capability to develop complex power solutions, while maintaining the quality of measurements and data provided. This is a critical need, as the instruments are expected to operate unattended in remote conditions, and at low temperatures, often with reduced daylight at higher latitudes, and at higher elevations, in both hemispheres.

As, increasingly, the users of instruments are more generalists, the technical solutions promoted and offered by manufacturers of instruments and validated through intercomparisons become critical to sustaining long term data quality when data is provided from diverse networks.

e. On the use of visibility to estimate snowfall intensity

No specific assessments of the feasibility of using visibility for estimating snowfall intensity were conducted in SPICE. Nevertheless, past contributions can be found in the literature, e.g. work using observational data from the the Marshal Field site in Boulder, Colorado, USA (Rasmussen et al., 1999). The study include simultaneous liquid equivalent snowfall rate from a weighing gauge in a DFAR, crystal types, and both automated and manual visibility measurements. Both the observations and theory showed that the relationship between liquid equivalent snowfall rate and visibility depends on the crystal type, the degree of riming, the degree of aggregation, and the degree of wetness of the crystals, leading to a large variation in the relationship between visibility and snowfall rate (varying from a factor 3 to 10, with a wide degree of scatter). The main cause for this scatter is the large variation in cross-sectional area to mass ratio and terminal velocity for natural snow particles.

Based on this study, presented in Section 4.4, it is not recommended that the liquid equivalent snowfall rate be estimated using visibility.

f. On appropriate target(s) under snow depth measuring sensors;

During SPICE, several sites used and assessed the performance of prepared surface targets beneath snow-depth sensors. In terms of properties, the targets should:

- Provide a level and stable surface upon which snow can accumulate and melt, as naturally as possible;
- Provide a reflective surface for either a sonic pulse or an optical beam for increased reliability of snow-depth measurements, particularly when the accumulated snow is minimal, intermittent, or zero.

The impact of the target properties (e.g. color, size, and material, artificial or natural) on the measurement of snow depth is presented in Section 4.2.6.2. As only one target type was tested at each site, and each site experienced different environmental conditions, comparative results are not available. Based on the results available, it appears that the sonic sensors are more susceptible to noise related to natural grass targets, and that their measurements benefit from the use of targets, while this is not as important for optical sensors.

It is recommended that the target surface is representative of the surface environment where the measurement will take place (e.g. colour/reflectivity), both during the accumulation phase and during melting. The user must take into account the potential drawbacks of using artificial targets, which can lead to errors in data. Among the most important are:

- The freeze-thaw cycle in heavier soils, which can cause frost heave and affect the relative distance between the sensor and the target, leading to zero-snow-depth drift (ZSD), as documented in Section 4.2.6.3;
- The settling of ground under the target, which can create ZSD and impact the levelling of the target;
- Differences in the absorption of solar radiation between the target and the surrounding ground surface.

Ideally, the distance between the snow depth and the target should be the same at the end of the accumulation season as it is at the beginning. A shift in this distance could be caused by a change in the mounting infrastructure of the sensor, a settling or heaving of the target area relative to the sensor, or both, and needs to be assessed at the beginning and end of each winter season. Details of ZSD assessment and related adjustments are provided in Section 4.2.6.3.

g. Consideration will be given to the needs of remote locations, in particular those with power and/or communications limitations.

Several SPICE sites were located in remote locations with power, communication, and access limitations: New Zealand, Australia, and Forni Glacier (see site reports in Annex 9); Pyramid Observatory (Nepal); and Tapado (Chile). While no formal intercomparisons were performed on any of these sites (none were S1 or S2 sites), their participation in SPICE contributed valuable experience and lessons learned regarding the operation of remote sites.

Key factors to consider in remote locations include how to address the power limitations and the management of instrument operations and maintenance, data collection and communication, and, in particular, instrument heating.

When designing solutions for remote locations, the following elements must be considered:

- The choice of sensor technology to find the right balance between the performance of measurements and unattended operation over extended periods with limited power (e.g. sonic vs. optical sensors for snow depth measurements);
- The availability of (near) real-time data communication versus the need to store data locally and retrieve it periodically;
- The availability of multiple sources of power for redundancy, with specific attention given to stations at high latitude during periods with minimum or no sun, or in regions with extended cloud coverage, for which solar power may be limited;
- While heating of instruments typically allows them to perform more effectively in cold conditions, the limited power resources in remote locations necessitate careful selection of the heating algorithm used;
- Specific instrument maintenance requirements also need to be taken into consideration. For example, in the case of weighing gauges, the decision of use needs to take into account the amount of precipitation on site (i.e. how often the bucket would fill, requiring it to be manually emptied), the requirement for using antifreeze and oil, the frequency of visits to the site for maintenance, and the environmental regulations regarding the handling of the bucket contents.

IV. Assess the achievable uncertainty of the measurement systems evaluated during SPICE and their ability to effectively and accurately report solid precipitation.

a. Assess the sensitivity, uncertainty, bias, repeatability, and response time of operational and emerging automatic systems;

The sensors under test have been assessed using methodology developed for each sensor type (see Section 3.6). The results of the assessment for each sensor type tested and related recommendations are available in the Instrument Performance Reports in Annex 6.

Overall, the results demonstrate that all available automatic instruments are able to detect and report solid precipitation reliably over short intervals of 30 minutes, as assessed relative to the DFAR reference configuration.

When applied consistently, the application of transfer functions developed in SPICE (Section 3.7) has been shown to reduce the measurement bias, but not the uncertainty (Kochendorfer et al., 2017a-c, 2018).

Measurements of solid precipitation using heated tipping bucket gauges are subject to response delays, resulting from evaporation and the time required to melt precipitation prior to its measurement. Delay times vary depending upon the specific gauge and heating configuration (heater location and power, algorithm) and environmental conditions, and can impact the timeliness of gauge reports in operational settings. An extensive characterization of response times was conducted for the gauges tested in SPICE; the results are reported in Section 4.1.2.

b. Assess and report on the sources and magnitude of errors including instrument (sensor), exposure (shielding), environment (temperature, wind, microphysics, snow particle and snowfall density), data collection and associated processing algorithms with respect to sampling, averaging, filtering, and reporting.

The results presented in this report, reflecting the performance at different sites testing the same SUT in identical configurations, show that the environmental conditions (exposure, climate, precipitation type, etc.) have a significant impact on the sensor performance – often more than the choice of the sensor itself. The impact of the environment on measurements is illustrated in Figure 4.6 to Figure 4.14 in the weighing gauges results section 4.1.1.3.3.1.1, where the catch ratio is presented for each gauge type and configuration installed at different SPICE sites.

While there are limitations of the measuring technologies, the understanding of how the physical configuration of instruments interacts with the environment is critical to limiting the sources of errors.

For weighing gauges and, to a certain degree, for the other instruments measuring solid precipitation, wind shielding is a widely used means of mitigating precipitation measurement errors. The individual IPRs provide specific information on the measurement limitations, by sensor type (see Annex 6).

The non-catchment type instruments have the potential to address some of the limitations related to measuring precipitation with catchment-type instruments; however, their use operationally would only be possible with additional work to improve the representation of particle size distributions and density for use in deriving precipitation accumulation from measurements.

The use of transfer functions is strongly recommended to address the limitations of solid precipitation measurements, where available for the measurement technology used. The summary of the development and use of universal transfer functions using SPICE datasets can be found in Section 3.7. Additional peer reviewed contributions based on this methodology have already been published (Kochendorfer et al. 2017a-c, 2018).

One major achievement of SPICE is the development of processing algorithms for QC and precipitation event selection to produce the site event datasets (see Sections 3.3 and 3.4). These provide a foundation for operational data quality control procedures. It is recommended that the methods and results presented in this report are developed further and implemented operationally, and that this methodology is adapted for time series datasets.

V. Evaluate new and emerging technology for the measurement of solid precipitation (e.g. non-catchment type), and their potential for use in operational applications.

SPICE assessed the performance of non-catchment type instruments for the measurement of solid precipitation. Instruments tested included disdrometers, present weather sensors, and evaporative plates, currently rarely implemented in operational programs. Their performance and the related operational considerations and recommendations are available in Annex 6, by instrument model. A summary of findings is presented in Section 4.1.3.

In general, it was found that the catch ratio of optical sensors over longer periods (e.g. one winter season) ranges between 0.8 and 1.3, but can vary on a much broader scale for individual precipitation events (from undercatch to overcatch by a factor of two). This indicates that their use for precipitation accumulation measurements over shorter time intervals (typically one hour) are unreliable. The reason for this is that non-catchment instruments do not measure the mass of snowflakes, but detect presence of hydrometeors and attribute mass based on the estimated size and assumptions of particle shape (spherical) and density. Additional details are available in Section 4.1.3.

The potential for the operational use of non-catchment type sensors has been recognized. They are, for instance, less sensitive to high winds, which could make them more reliable for measurements at more exposed sites. In order to reach this goal, it is recommended that additional work be conducted to understand and improve the instrument specific internal algorithms, in a joint effort between manufacturers and the user communities.

All of these instruments are systems applying proprietary algorithms, and users have no access to the assumptions and decisions implemented in these algorithms for translating detected signals into data. For this reason, it is strongly recommended that instrument manufacturers provide accurate descriptions of the assumptions and processing applied internally to derive precipitation amounts and other associated information from detected signals. Advanced understanding of the algorithms used will enable users to select instruments according to the intended application of data, and for specific operational conditions. At the same time, this will provide a sound base for feedback to manufacturers, for further improvements, in the process of evolving these instruments for broader operational use.

The operational use of these sensors requires specific considerations:

- The power requirements for maintaining the operation of instruments within their nominal parameters;
- The orientation with respect to the prevailing wind direction, which can play a role during precipitation events, depending on the physical configuration of the instrument;
- Whether shielding around the sensing volume is beneficial or not (e.g., see the Thies LPM IPR in Annex 6);
- Understanding the data made available by the instruments (linked to the disclosure of the internal processing algorithms), and how they could be used;
- Potential advantages of the non-catchment nature of these instruments for field operations (e.g. they may require field intervention less frequently than weighing gauges, which require the bucket to be emptied periodically);
- Potential risks to data continuity in the case of power interruptions (e.g. any precipitation occurring during power interruptions will not be recorded).

Recognizing the high scatter in the catch ratio results, no effort has been made to try to develop transfer functions for non-catchment type instruments.

For emerging technologies for the measurement of snow on the ground, no assessments in the field were performed, but a literature review was conducted (see Section 4.1.4.5). These instruments belong primarily to the research community; however, if such instruments are to be made available to a broader user community, including operational networks, it would be highly desirable to carry out extended testing of “packaged” versions of the instruments, involving not only research groups, but also field testing sites operated by national operational organizations. This would help to evaluate and improve system operation, allow for expeditious calibration of sensors and devices, and encourage the development of new sensors and technologies that further the scientific goal of understanding, measuring and modeling changes in the seasonal snowpack.

VI. Configure and collect a comprehensive data set for further data mining or for specific applications (e.g., radar- and/or satellite-based snowfall estimation). Enable additional studies on the homogenization of automatic/manual observations and the traceability of automated measurements to manual measurements.

The data collected during SPICE have been centrally quality-controlled (see description of the procedure in Section 3.3.2) and stored in a central database, hosted, at the time of the release of this report, by NCAR. The dataset will be made available after the release of the report through WMO.

SPICE has produced a very valuable dataset, which will remain available for further analysis and data mining by the community. This dataset from multiple sites (both Raw and QC'ed Data) constitutes an important resource to be considered as a tool for the organization of short practical training courses in the frame of the activities of the RTC's network (WMO Regional Training Centers). The dissemination of the methodologies for data quality control and data processing may be enhanced in coordination with NMHS's and the WMO Education and Training Office.

To the extent possible, all SPICE sites are encouraged to continue operating, and to make data available, together with all associated metadata (e.g. reflecting any modifications to the configuration over time, etc.). This will allow for the establishment of long-term datasets, which will open up additional avenues of research for meteorological, hydrological and climatological purposes. The satellite and radar communities could use these datasets as accurate and quality-controlled ground-truth information for verification and validation purposes.

It is recommended that additional work is undertaken to assess the impact of using corrected datasets from operational stations (including for archived data) on specific applications (e.g. climate studies) and for different time scales, and provide input to further refine the correction procedures, as required. This work should be undertaken in collaboration with precipitation data centers, to ensure a feasible uptake of recommended methodologies, and consistent application of results.

Recommendations for future intercomparisons:

Experience gained during the organization, coordination, and execution of SPICE provides the basis for several recommendations for future WMO intercomparisons:

- Recognizing the complexity of the desired results, intercomparisons should be organized with a much more targeted focus, to enable better management of outcomes and timelines.
- For intercomparisons as complex as SPICE, it is recommended that multiple goals are managed through multiple standalone projects, delivered by fewer organizations, which could mitigate more easily the changes in resources and expertise available over the durations of projects.
- Availability of resources for intercomparisons need to be clarified at the onset of the project to give all parties a good understanding of the necessary commitment to be made.
- Intercomparisons should be time bound to ensure the availability of required resources over their duration; longer term projects should be fully funded at the onset to mitigate impacts of resource limitations, which can delay delivery of key objectives or necessitate reductions in the scope of objectives to be delivered.
- Intercomparisons should be organized with a clear link to the users of data, and should include data users when defining the expected results.

- Intercomparisons should address the needs of developing countries and the transition to automatic observations.
- Intercomparisons should include objectives for end-to-end solutions, covering all aspects from sustained measurements to the delivery of data to data centres, at timescales representative of their primary applications. These solutions are of particular interest for remote areas, where resources and infrastructure may be limited.
- Instrument manufacturers should play an active role in providing solutions for end-to-end systems, and use the opportunity of intercomparisons to validate the operation of these solutions.

Recognizing the need to protect the intellectual property of instrument manufacturers, it is recommended that the processing of measurements be well described, to enable users to select the appropriate data output for their application and to integrate data from multiple instruments, operated by different networks and agencies. The latter would represent an important step toward the goal of achieving a fully global, integrated observing system.

1. MOTIVATION AND INTERCOMPARISON PRINCIPLES

1.1 Definition of the objectives

Solid precipitation is both complex and challenging to observe and measure (Rasmussen et al. 2012). Since the first intercomparison of solid precipitation measurements organized by the World Meteorological Organization (WMO) between 1989 and 1993 (WMO/TD-No. 872, 1998), significant advancements have been made in automatic instrumentation for measuring solid precipitation and snow on the ground (SoG). In addition to these advancements, and linked to the introduction of digital electronic components and signal processing, new applications have also emerged that require precipitation data with increased temporal and spatial resolutions. These applications include climate change, nowcasting, water supply budgets, avalanche forecasts and warnings, and satellite ground validation.

These are important data. Snow on the ground and snowfall data are used widely by weather and hydrologic forecasters, climate researchers, water resource managers, construction engineers, snowplow operators, airport managers, winter resort managers, farmers, and many others.

New instruments that measure precipitation without capturing it (non-catchment-type instruments) are increasingly used operationally and are based on the principles of light scattering, microwave backscatter, and mass and heat transfer, among others. In parallel, the more traditional catchment-type instruments, tipping bucket and weighing gauges, have been developed further to include new features (e.g. heating, on-board digital signal processing, temperature and wind-impact compensation, software corrections) that expand the range of operating conditions and data products available.

The measurement of precipitation globally, in different climates and with different site exposures, instruments, and configurations, is beset with spatial and temporal inhomogeneity that has serious consequences for the accuracy and consistency of local and global precipitation time series (Sevruk, 1994). The results of the 2008 CIMO survey on national summaries of methods and instruments for solid precipitation measurement at automatic weather stations (Nitu and Wong, 2010) indicates that a variety of automatic instruments are being used worldwide for measuring solid precipitation, with multiple instrument types being used even within the same country. This variety exceeds by far that of manual standard precipitation gauges used previously (Goodison et al., 1998).

The WMO's Commission for Instruments and Methods of Observation (CIMO) agreed in 2010 to organize an intercomparison for assessing the impact of automation on the measurement of snowfall, snow depth, and solid precipitation in cold climates. The organization of the WMO Solid Precipitation Intercomparison Experiment (WMO-SPICE[1]) was endorsed at the Sixteenth Congress of WMO and the work commenced in 2011. Building on results and recommendations from previous studies and intercomparisons, the SPICE objectives focus on the use of automatic instruments for measuring and reporting:

- Precipitation amount over various time periods (i.e. minute, hour, day, season) as a function of the precipitation phase, with a focus on solid precipitation
- Snow on the ground (snow depth). As snow-depth measurements are closely tied to snowfall measurements, the intercomparison investigated the linkages between them.

[1] The terms "WMO-SPICE," "SPICE," and "CIMO/WMO SPICE" refer to the same project; the term "SPICE" will be used in the report.

SPICE provides guidance on the use of modern automated systems for measuring precipitation amount and snow depth, and recommends appropriate automated field reference system(s) for the unattended measurement of solid precipitation in cold climates. Differences between various automated instruments under test and the field working reference systems (FWRSs) established for the intercomparison are documented and used to assess the performance of instruments. Where available, the results include comparisons of automatic and manual measurements of solid precipitation.

Recommendations for adjustments that account for the undercatch of solid precipitation due to gauge exposure are presented as a function of variables available and recommended for an operational site, such as wind, temperature, and precipitation type. Additionally, the sources and magnitude of errors due to instrument characteristics, field exposure, shielding, environmental conditions, and data processing methods are investigated.

The objectives of SPICE were developed through consultations with stakeholders from the National Meteorological Services, the WMO Technical Commission for Hydrology (CHy), the WMO Technical Commission for Climatology (CCL), the World Climate Research Program – Working Group on Nowcasting, the WMO Executive Council for Polar Observations, Research, and Services (EC-PORS), the Global Cryosphere Watch (GCW, for which SPICE has been used as a demonstration project), and the Global Climate Observing System (GCOS).

A detailed description of the SPICE objectives and deliverables is provided in Annex 1.

1.2 Intercomparison principles

SPICE was led by an International Organizing Committee (IOC) with representatives from Canada, China, Germany, Italy, New Zealand, Russia, Switzerland, and the USA. The SPICE experiment was conducted across multiple test sites in both hemispheres. The goals were twofold: to address the complexity of measuring solid precipitation in various climate regimes and to assess impacts of both the environment and site configuration on the measurements and analysis of results.

The SPICE tests started informally in the winter of 2011/12, building on existing capacity in Canada, Germany, Finland, Italy, Norway, Switzerland, and the USA. The formal intercomparison started in December 2012, and it concluded in 2015 at the end of the winter season in the Southern Hemisphere.

The IOC recognized the need for a flexible approach for the configuration of the field references. It was necessary for results to be linked to the previous intercomparison (Goodison et al., 1998) while providing a working field reference with increased temporal resolution, and to ensure the transferability of the results from participating sites while also recognizing the physical limitations on some of the sites.

Taking into account these expectations, the following three configurations of the SPICE Field Working Reference System (FWRS) were endorsed (see Section 3.1.3.2 for a detailed description):

- R1: Double Fence Intercomparison Reference (DFIR) + Tretyakov gauge (manual collector) + Tretyakov shield, designated in the 1989-1993 intercomparison as a secondary field reference (WMO/TD No. 872, 1998)
- R2: DFIR + automatic weighing gauge (AWG) + shield. The model and the configuration of the AWG and its shield were specified at the end of the 2011/12 pre-SPICE experiment (see Section 3.1.3.2).

- R3: A combination of automatic weighing gauge(s) and windshields with sufficient characterization and history to have a degree of confidence for the purpose of meeting specific objectives, as agreed between the host country and the IOC. The characterization of R3 systems must be done in relation to the R1 and R2 systems. This is a pragmatic approach for sites contributing to meeting the SPICE objectives, but where the installation of a DFIR is not feasible (e.g. complex terrain with heavy wet snow).

Given these proposed configurations, the possible test site designations were as follows:

- S1: Sites where references of type R1, R2, and R3 are available. The presence of R3(s) allows for its (their) characterization against the R1 and R2.
- S2: Sites where references of type R2 and R3 are available (no manual measurements being made). The presence of the R3(s) allows for its (their) characterization against the R2.
- S3: Sites where, due to the site limitations, only field references of type R3 are feasible.
- S4: S3 sites, which enabled the investigation of specific issues, such as the operability of gauges in certain environments (e.g. high and/or remote mountain environments), their data was not used for the derivation of transfer functions.

The presence of the R3 configuration on S1 and S2 sites enables the transferability of results between the participating sites by enabling the characterization of the R3 as a function of the R1 and R2 configurations.

This site and reference terminology has been introduced to allow for easy differentiation among the different configurations, but is not intended to be a classification mechanism. A more extensive description of the different reference types and their traceability can be found in Section 3.2.

Twenty sites in sixteen countries expressed interest in participating in the WMO-SPICE initiative and met the criteria for one of the reference site categories. (See Figure 1.1.) The participating sites were in Australia (Guthega Dam); Canada (Egbert (hereafter called CARE), Bratt's Lake, Caribou Creek), Chile (Tapado); France (Col de Porte); Finland (Sodankylä); Italy (Forni Glacier); Japan (Rikubetsu, Joetsu); Nepal (Pyramid Observatory); New Zealand (Mueller Hut); Norway (Haukeliseter); Poland (Hala Gasienicowa); Republic of Korea (Gochang); Russian Federation (Valdai, Voljskaya); Switzerland (Weissfluhjoch); Spain (ARAMON-Formigal); and the United States of America (Marshall).

Legend

1.	Caribou Creek, Saskatchewan, Canada	11.	Haukeliseter, Norway
2.	Bratt's Lake, Saskatchewan, Canada	12.	FMI/Sodankylä Arctic Research Centre, Finland
3.	Marshall Site, Colorado, USA	13.	Valdai, State Hydrological Institute, Russia
4.	CARE, Ontario, Canada	14.	Voljskaya Observatory, Gorodec, Russia
5.	Tapado AWS, Región de Coquimbo, Chile	15.	Pyramid Observatory, Nepal
6.	Formigal, Spain	16.	Gochang, Korea
7.	Col de Porte, France	17.	Joetsu, Japan
8.	Weissfluhjoch, Davos, Switzerland	18.	Rikubetu, Hokkaido, Japan
9.	Forni Glacier, Italy	19.	Guthega Dam, New South Wales, Australia
10.	Hala Gasienicowa Station, Poland	20.	Mueller Hut Weather Station, New Zealand

Figure 1.1. Locations of participating sites.

The configuration of each participating site is provided in the corresponding commissioning reports, available at:

http://www.wmo.int/pages/prog/www/IMOP/intercomparisons/SPICE/SPICE.html

All participating sites had one or more FWRS as detailed in Section 3.1.3.3 of this report. Also, each site had one or more sensor under test (SUT). All instruments tested in SPICE were provided either by the host organizations, or by instrument providers (i.e. manufacturer, distributor, or instrument user not operating a SPICE test site).

The principles agreed upon by the IOC to determine the allocation of the instruments were as follows (IOC-2 Final Report, Boulder, 2012):

1. In response to the second and third letters of invitation issued by the WMO for participation in SPICE, members and manufacturers indicated interest in providing instruments for inclusion in the experiment. Once included, these instruments complemented the suite of instruments proposed by the site hosts and already available for the intercomparison. In this report, a WMO member (not hosting a SPICE intercomparison site) or a manufacturer proposing instruments for inclusion is generically recognized as an instrument provider.

2. The instrument providers proposed a wide range of instruments for the measurement of precipitation amount, SoG, and snow water equivalent reflecting the operating principles currently used for operational and scientific applications.
3. The IOC decided to allocate instruments principally to S1 and S2 sites to ensure fairness for the proponents and ensure the most relevant results, as the transfer functions were to be derived from those sites. Given the available capacity of the sites and the interest to test the instruments in a variety of climatological conditions, the IOC contacted the instrument providers that proposed more than one instrument of the same model, seeking their agreement for installing all proposed instruments at the start of the intercomparison, without keeping any spares. In addition, the IOC sought their cooperation in dealing with instrument failures, when and if needed over the course of the experiment.
4. The IOC assessed the submissions made by the instrument providers and all instruments proposed were accepted for participation in SPICE. The allocation of instruments to the participating sites took into account site capacity, project objectives, site climatology, and site objectives.
5. The IOC decided that the primary focus of the Marshall Site (USA) would be the assessment of instruments measuring precipitation amount. For that reason, at least one unit of each of the proposed models of weighing gauges was installed and tested on this site. Additionally, based on the site capacity, one unit of most of the tipping-bucket-type gauges selected was also assigned to this site. The second unit of the available weighing gauges and the balance of the tipping-bucket-type gauges were distributed between the following sites: CARE and Bratt's Lake (Canada), Sodankylä (Finland), Haukeliseter (Norway), Weissfluhjoch (Switzerland), Guthega Dam (Australia), and Mueller Hut (New Zealand).
6. For the assessment of the snow-depth/SoG instruments and snow-water-equivalent (SWE) instruments, the sites of Sodankylä (Finland) and Hala Gasienicowa (Poland) were designated as primary sites, and the instruments proposed by manufacturers for these types of measurements were allocated here. Additional instruments proposed by manufacturers were deployed at Caribou Creek (Canada), Col de Porte (France), and Formigal (Spain). Additionally, tests for the assessment of snow-depth measurements and their relationship to snowfall were organized on most of the participating sites, including CARE (Canada), using instruments owned and operated by the site proponents.
7. The non-catchment-type instruments proposed by the instrument providers were distributed to several sites, taking advantage of their range of climatological conditions and complementing the availability of similar sensors proposed by the site proponents. Most of the submitted non-catchment-type instruments were located at the Marshall site to allow for their assessment alongside the represented weighing-type gauges and tipping buckets. The focus was on the measurement of precipitation amount.

Following the principles defined earlier, a summary of the instruments included in the intercomparison and their allocation to the different SPICE sites can be found in Annex 5 (by instrument model and by site).

Calibration and installation of the instruments contributed by instrument providers were done according to the manufacturers' requirements (as specified in the corresponding instrument manuals). For the reference instruments in R2 and R3 configurations, the configuration was defined by the IOC to ensure consistent set-up across the sites (heating, antifreeze and oil, data sampling, etc., see Section 3.1.3.4.2 for a detailed explanation).

Data collected during SPICE were archived and quality controlled at a central location hosted by the National Center for Atmospheric Research (NCAR), Boulder, CO, USA. A data protocol was developed to regulate the use of SPICE data during and after the completion of the project and all stakeholders were asked to adhere to it. (See Annex 3.)

A standard methodology was developed to derive a common final precipitation data set for the analysis, as detailed in Section 3.4. The basic principle was to select precipitation events with a high level of confidence. Following automatic and manual Quality Assurance/Quality Control (QA/QC), three separate data sets were produced for each site:

1) Site Event Data Set (SEDS): events with precipitation
2) Site Non-Event Data Set (SNEDS): events with no precipitation
3) Site Light Event Data Set (SLEDS): events with light precipitation (do not meet criteria for SEDS, but precipitation was observed)

The threshold and decision algorithms to classify events within these three categories are described in Section 3.4. Similarly, a standard methodology for the quality control of the SoG data was developed, although independent of the NCAR archive. This methodology is also described in Section 3.4.

The methodology and associated algorithms developed within SPICE and presented in the following sections of this report do not reflect those of specific members, but were based on best practices related to the project and its objectives. Each user (WMO members) must decide whether those algorithms should be applied to their data and, if so, to what extent, assessing how a change of algorithm may affect the data.

2. SITE DESCRIPTIONS

Author: Francesco Sabatini

The first call for expressions of interest in SPICE participation was issued to WMO members and Association of Hydro-Meteorological Equipment Industry (HMEI) members in 2011.

A second letter of invitation was issued by the WMO in 2012. The letter was accompanied by two questionnaires, developed to gather detailed information on proposed sites and on the proposed instruments (see Annex 2.)

Proposals for potential test sites and participating instruments were received at different stages. The IOC reviewed all submissions, selected participating sites and instruments, and allocated instruments to the respective sites. Sixteen countries and 20 field sites (see also Figure 1.1) were selected in the following climate zones:

- Alpine climate: Australia, Chile, France, Italy, Nepal, New Zealand, Norway, Switzerland, Spain
- Northern Boreal: Finland
- Continental climate: Canada, USA, Russian Federation
- Maritime climate: Republic of Korea, Japan

The location and climate characterization of each site is given in Table 2.1, together with the site type, corresponding to its configuration. Site position and elevation are also provided, along with the climate zone according to the Köppen climate classification (e.g. Peel et al., 2007). Specific climate zones noted in Table 2.1 include oceanic (Cfb), humid continental (Dfb), cold semi-arid (BSk), subarctic (Dfc), humid subtropical (Cfa), and subtropical highland (Cwb).

The site proposals included information on the configuration of references for measuring precipitation amount and/or snow on the ground. Additionally, the site proponents proposed instruments for the intercomparison as SUT. These instruments and their configurations reflected either current national standards for the measurement of solid precipitation or were of specific interest to the proponent. Given the interest in increasing understanding of the national methods of measurement, the IOC acceptance of a site for participation in SPICE was an implicit an acceptance of the proposed configurations and instruments for inclusion in the experiment.

Each site manager was responsible for configuration of the experiment on their site. Prior to the official start of the experiment, the site configuration was commissioned following a procedure developed and approved by the IOC. The IOC reviewed the commissioning reports and proposed any required amendments, prior to formal acceptance. For each site, an initial testing phase was required to ensure that all instruments and equipment worked correctly, and to identify possible errors or malfunctions of instruments and/or equipment.

Table 2.1. Location, type, and climate zone (using the Köppen climate classification) of each SPICE test site. Note: in the report, Sodankylä is refered to as Northern boreal since its characteristics are not representative for an arctic or sub-arctic climate.

COUNTRY	Site	Type	Lat [°]	Lon [°]	Elevation [m asl]	Climate Zone
AUSTRALIA	Guthega Dam	S3	-36.38	148.37	1586	Cfb
CANADA	Bratt's Lake	S2	50.20	-104.71	585	Dfb
CANADA	CARE	S1	44.23	-79.78	251	Dfb, subject to lake effect
CANADA	Caribou Creek	S0a	53.94	-104.65	519	Dfb
CHILE	Tapado	S4	-30.16	-69.91	4318	BSk, Glacier Plateau
FINLAND	Sodankylä	S2	67.37	26.63	179	Dfc
FRANCE	Col de Port	S3	45.30	5.77	1325	Cfb
ITALY	Forni Glacier	S4	46.40	10.59	2631	Cfb
JAPAN	Joetsu	S2	37.12	138.27	11	Cfa
JAPAN	Rikubetsu	S2	43.48	143.76	217	Dfb
KOREA, REP. OF	Gochang	S2	35.35	126.60	52	Cfa
NEPAL	Pyramid Nepal	S4	27.96	86.81	5050	Cwb
NEW ZEALAND	Mueller Hut	S4	-43.72	170.06	1818	Cfb
NORWAY	Haukeliseter	S2	59.81	7.21	991	Cfb
POLAND	Hala Gasienicowa	S4	49.24	20.00	1520	Dfb
RUSSIAN FED.	Valdai	S0	57.98	33.25	194	Dfb
RUSSIAN FED.	Volga	S1	56.68	43.42	100	Dfb
SPAIN	Formigal	S2	42.76	-0.39	1800	Cfb, with Atlantic influence
SWITZERLAND	Weissfluhjoch	S2	46.83	9.81	2537	Cfb
USA	Marshall	S1	39.95	-105.20	1742	BSk

The climatology of each site in terms of temperature, wind, and precipitation is presented in Table 2.2. Readers can link their own site with a SPICE site according to the characteristic environmental conditions.

A detailed description of each SPICE site is available in Annex 4. This includes site location, layout (with pictures), and instrument list.

Table 2.2. Climatology of SPICE sites, with T_{mean}: daily average mean air temperature [°C], T_{max}: daily average maximum air temperature [°C], T_{min}: daily average minimum air temperature [°C], Total SFL: Average total snowfall [cm], Total PRP: total precipitation [mm], WS_{avg}: daily average wind speed [m/s], WS_{max}: daily average maximum wind speed [m/s].

COUNTRY	Site	T_{mean} [°C]	T_{max} [°C]	T_{min} [°C]	Total SFL [cm]	Total PRP [mm]	WS_{avg} [m/s]	WS_{max} [m/s]	Ref. Winter period	Obs. Period
AUSTRALIA	Guthega Dam	1.6 °	6.5	-2.6	280	1024	4.3	8.2	May-Sep	2006-15
CANADA	Bratt's Lake	-2.1	3.9	-8.1	106	206	5.3	24	Sept-May	
CANADA	CARE	-0.9	3.6	-5.4	157	430	3.7	n/a	Oct-Apr	
CANADA	Caribou Creek	-4.5	1.1	-10.0	138	252	2.6	10.9	Sept-May	
CHILE	Tapado	-6.0	-1.9	-11.7	329	- -	4.3	16.5	May-Sep	
FINLAND	Sodankylä	-0.4	4.1	-5.0	181	527	2.7	4.2	Oct-Apr	1981-2010
FRANCE	Col de Porte	-0.1	3.1	-3.4	557	794	1.4	10.4	Dec-Apr	1960-61 - 2011/12
ITALY	Forni Glacier	-1.4	4.0	-5.3	77	1562	5.0	20.5	Jan-Dec	
JAPAN	Joetsu	3.9	8.0	0.3	618	1298	2.5	--	Dec-Mar	1981-2010
JAPAN	Rikubetsu	-1.4	5.8	-8.8	420	388	1.6	--	Dec-Mar	
KOREA, REP. OF	Gochang	0.8	5.4	-3.6	80	99	2.9	6.9	Dec-Feb	2010-15
NEPAL	Pyramid Nepal	-2.3	6.0	-18.1	--	306	2.18	8.94	Jan-Dec	2004-2013
NEW ZEALAND	Mueller Hut	--	--	--	--	--	--	--	May-Sep	
NORWAY	Haukeliseter	-2.0	0.9	-4.4	--	594	--	--	Nov-Mar	
POLAND	Hala Gasienicowa	--	--	--	--	--	--	--	Nov-Mar	
RUSSIAN FED.	Valdai	0.1	3.0	-4.4	125	509	3.4	18	Sept-May	
RUSSIAN FED.	Volga	--	--	--	--	--	--	--	Nov-Apr	
SPAIN	Formigal	-2.1	0.2	-6.1	374	403	--	--	Nov-Mar	
SWITZER-LAND	Weissfluhjoch	-4.4	12.1	-23.6	740	586	2.2	12.6	Oct-Apr	1999-2015
USA	Marshall	4.0	11.5	-3.5	194	229	--	--	Oct - Apr	

3. APPROACH AND METHODS

Given the complexity of this project, a standardized approach was needed that would make use of robust and broadly applicable methods for investigation and analysis. Outcomes from the previous intercomparison and existing standards (e.g. from the CIMO Guide) were used as guideposts. The definition and implementation of standardized reference configurations is an important component of this approach. The descriptions of reference configurations for both solid precipitation and SoG measurements are presented in detail in Section 0. For solid precipitation measurements, the DFIR-shield was used as part of the reference system (see Section 3.1.3). Since SPICE is assessing automatic instruments, the use of a DFIR-shield with a shielded automatic gauge inside was defined as a DFAR[2] (double fence automatic reference) to avoid confusion when relating the automatic measurement to the manual reference (see Section 3.2.1). For the SoG reference, a composite of available instruments was used, including manual measurements of snow depth (via snow stakes), SWE (using snow tubes), and a mean of available automated instruments (where available, see Section 3.1.4).

An important consideration when dealing with different reference systems is to ensure traceability, as briefly described in Section 1. Significant effort was taken to interrelate reference types. The results of this work are presented in Section 3.2.

Detailed descriptions of data acquisition and management were crucial to understand the differences and similarities in data from different sites. All data were processed to a common time resolution to facilitate intercomparison. As well, data produced during the SPICE campaign were centralized in one database, hosted by NCAR (see Section 3.3). From the raw data archived in this database, standardized data sets were produced for all analyses. This ensures consistency among the different datasets considered in each component of the analysis. The methodology used to produce these data sets is presented in Section 3.4.

Each instrument type submitted by instrument providers has been analyzed and evaluated against the reference measurement. The methodology used to assess the performance of each instrument is presented in Section 3.6. To ensure that users will have comparable information, and to reduce risks of presenting results biased towards one type of technology, great care has been taken to present the performance of instruments using different technologies and principles through common templates. These results are presented in the instrument performance reports (IPRs) in Annex 6. It is noted, however, that some aspects of the analysis are specific to particular technologies, and hence, some differences exist among the IPR content.

Another key objective of SPICE was to assess the possibility of deriving transfer functions to account for (and, ideally, to correct) wind-induced error in solid precipitation measurement. The concept of "universal" transfer functions, which can be applied to data from instruments with a specific configuration in different climate conditions, was investigated. The methodology and results relevant to this objective are presented in Section 3.7.

[2] In this report, both the terms "DFIR-fence" (referring to the wooden double fence) and "DFAR" (referring to the system composed of an automatic gauge within a DFIR-fence) are used.

3.1 Description of the reference configurations

Authors: Rodica Nitu, Paul Joe

3.1.1 The issue of field references

The *WMO Guide to Meteorological Instruments and Methods of Observation* (WMO-No.8, 2014), states that "Intercomparisons of instruments and observing systems, together with agreed quality-control procedures, are essential for the establishment of compatible data sets."

However, the guide notes that many meteorological quantities cannot be directly compared with metrological standards. As a result, there are no absolute references for such variables as visibility, cloud-base height, and precipitation. Intercomparisons are invaluable here. The guide recommends that host countries include at least one reference instrument in the intercomparison, and if no recognized standard or reference exists for the variable(s) to be measured, a method to determine a reference for the intercomparison should be identified.

Where no reference instrument exists, the guide states, instruments should be compared against a relative reference selected from the instruments under test. Of course, care must be taken to exclude unrepresentative values from the selected data subset.

The measurement of precipitation and SoG cannot be traced directly to absolute references given the difficulty in defining the "true" amount of precipitation falling or already fallen relative to the amount measured at any single point. For SPICE, pragmatic and feasible FWRSs have been defined to enable the compatibility and reproducibility of results among the participating sites. The approach is similar to that applied for the WMO solid precipitation intercomparison conducted between 1986 and 1993 (WMO/TD No. 872, 1998).

To enable a broad understanding of SPICE's approach, the foundational principles governing the measurements and reporting of results from this project complement those defined in the *International Vocabulary of Metrology (VIM) — Basic and general concepts and associated terms* (JCGM, 2008). According to VIM, a reference "can be a measurement unit, a measurement procedure, a reference material, or a combination of such" (VIM 1.1, Note 2). A reference measurement procedure is "accepted as providing measurement results fit for their intended use in assessing measurement trueness of measured quantity values obtained from other measurement procedures for quantities of the same kind" (VIM 2.7). VIM also defines the metrological comparability of measurement results as those "that are metrologically traceable to the same reference" (VIM 2.46) and reference data as the "data related to … a system of components of known composition or structure, obtained from an identified source, critically evaluated, and verified for accuracy" (VIM 5.16).

The FWRS data are considered to be the SPICE reference data, defined by VIM 5.18 as the "quantity value used as a basis for comparison with values of quantities of the same kind".

3.1.2 Criteria for WMO-SPICE field working reference systems

To achieve relevant results, SPICE requires that the field working reference systems are well understood and accepted. The FWRSs for SPICE have been selected to meet specific criteria relevant to measurement type. For the assessment of automatic instruments, references using automatic instruments are required to provide similar sampling or reporting frequency. It is not feasible to make manual measurements over shorter temporal scales (e.g. minutely) due to both the effort required and the measurement resolution of the manual methods (e.g. manual precipitation measurement by weight or volume).

The FWRS systems for SPICE need to:

- Be robust (produce data in a variety of conditions continuously)
- Be consistent and repeatable (produce the same value under the same conditions)
- Have sources of error that are understood in terms of biases, variances, and correlations; ensure confounding factors are identified, measured, and understood
- Be feasible to implement
- Collect, generally, the most solid precipitation relative to other test configurations on a given site, particularly in strong winds.

Each of these characteristics is described in greater detail below.

Robustness: The reference data need to be available at all times and under all conditions. This refers to the stability of the physical setup of the instruments (e.g. resistant to high winds, snow, and ice storms); the reference operates continuously (e.g. continues to measure in heavy wet snow); and data are collected, recorded, and transmitted without loss. This applies to all sensors needed as components of the reference system, including those for ancillary measurements.

Reliability: The key characteristic for a reference is that it produces the same value under the same conditions. This, pragmatically, requires that two or more identical co-located sensors produce (or nearly produce) the same values (precision); there is low variance in data when the conditions are steady; the bias is known; and bias-free values are produced in controlled, ideal conditions (calibration).

Understanding: The *WMO Intercomparison on Solid Precipitation, 1986-1993* (WMO/TD - No. 872, 1998) identified that the mean wind speed was the major environmental factor that impacted the catchment efficiency of instruments tested. Particle aerodynamics have been demonstrated to be a significant factor in this reduction of catch efficiency. Noise and artifacts due to external influences (e.g. diurnal temperature variations) can also impact significantly the data from automated gauges. Measurements of these confounding factors are needed to understand the errors. Lack of these measurements would result in unexplained variance in the analysis phase.

Implementable: For the intercomparison, there are many sites with instruments installed in remote locations and with limited space. A range of field reference systems are needed to enable tests at all participating sites, and to allow for the development of transfer functions between various levels of reference systems from various sites.

3.1.3 SPICE FWRS configuration for falling precipitation

3.1.3.1 Historical perspective

3.1.3.1.1 *Double Fence Intercomparison Reference*

The first WMO intercomparison on solid precipitation (WMO/TD - No. 872, 1998) recommended a field reference for the measurement of solid precipitation based on the experience and instrumentation available at that time. This field reference was referred to as the DFIR and recognized as the secondary field reference for the measurement of solid precipitation (the primary field reference being the bush gauge, as described in WMO/TD - No. 872, 1998). As described in the intercomparison report (WMO/TD - No. 872, 1998), the DFIR is an “octagonal vertical double-fence inscribed into circles of 12 m and 4 m in diameter, with the outer fence 3.5 m high and the inner fence 3.0 m high surrounding a Tretyakov precipitation gauge mounted at a height of 3.0 m. In the

outer fence there is a gap of 2.0 m and in the inner fence of 1.5 m between the ground and the bottom of the fences." (Figure 3.1).

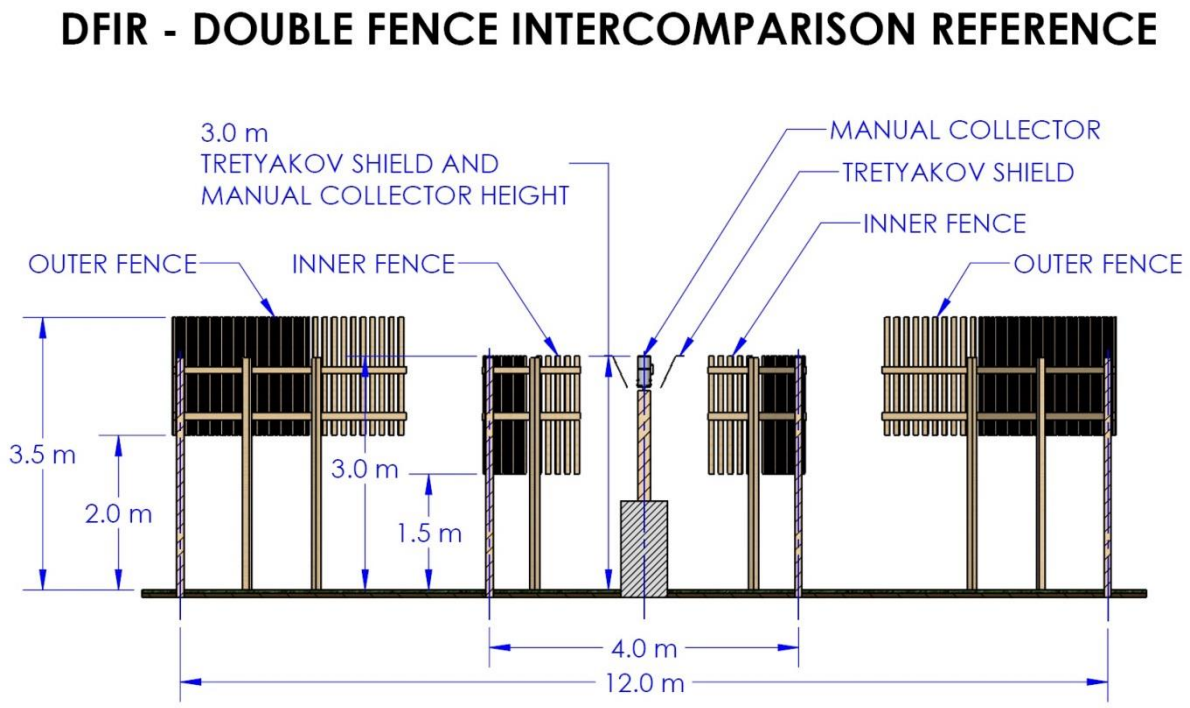


Figure 3.1. Cross-section of WMO Double Fence Intercomparison Reference (DFIR) (Drawing by J. Hoover, Environment and Climate Change Canada).

3.1.3.1.2 *Tretyakov gauge*

The Tretyakov precipitation gauge (Figure 3.2) at the center of the DFIR is a tin cylinder with an opening of approximately 200 cm^2, surrounded by a fixed-slat shield, known as a Tretyakov shield. It was introduced in the former Union of Soviet Socialist Republics (USSR) at the end of the 1950s. As the Tretyakov precipitation gauge/collector was used broadly at the time of the WMO intercomparison of 1986-1993 and had the most complete documentation of its performance for a wide range of climatic conditions, it was designated as the working network reference gauge for that intercomparison.

Figure 3.2. Tretyakov gauge and shield (CARE, Canada).

3.1.3.1.3 *History of the DFIR*

According to the report of the previous WMO intercomparison (WMO/TD-No. 872, 1998), the use of fences to protect precipitation gauges from the wind are attributed to Swiss meteorologist Heinrich Wild in the second half of the 19th century and to the Russian scientist G.I. Orlov in the early 20th century. The fence installed around a gauge collecting snow by Wild in Russia was a single square of 5 x 5 m, 2.5 m high. The gauge orifice was at 1 m above ground. Considerably more snow was reported by this gauge relative to a similar collector gauge, which was used unshielded (Wild, 1885). The use of a double-fence can be attributed to Orlov (1946), likely used for the first time in 1936 in Russia. This was an octagonal double-fence 2.5 m high, with an inner fence 4 m in diameter and an outer fence 12 m in diameter. The height of the gauge orifice was 1.7 m above ground.

As reported in Golubev (1986), three different types of double fences were tested at the Valdai experimental site (Russian Federation) between 1965 and 1972. The so called "bush gauge" was used as the reference configuration, comprising a standard snow gauge installed in a wooded area of about 100 x 100 m. The gauge was surrounded by bush cut to the level of the gauge orifice. In the experiments on this site, the catch ratios of the gauges in the double fences, defined as the ratio of the precipitation accumulation reported by each test gauge relative to that of the reference bush gauge over a set period of time, varied from 92% to 96%. Similar results were obtained during the 1986-1993 WMO intercomparison (Figure 3.3).

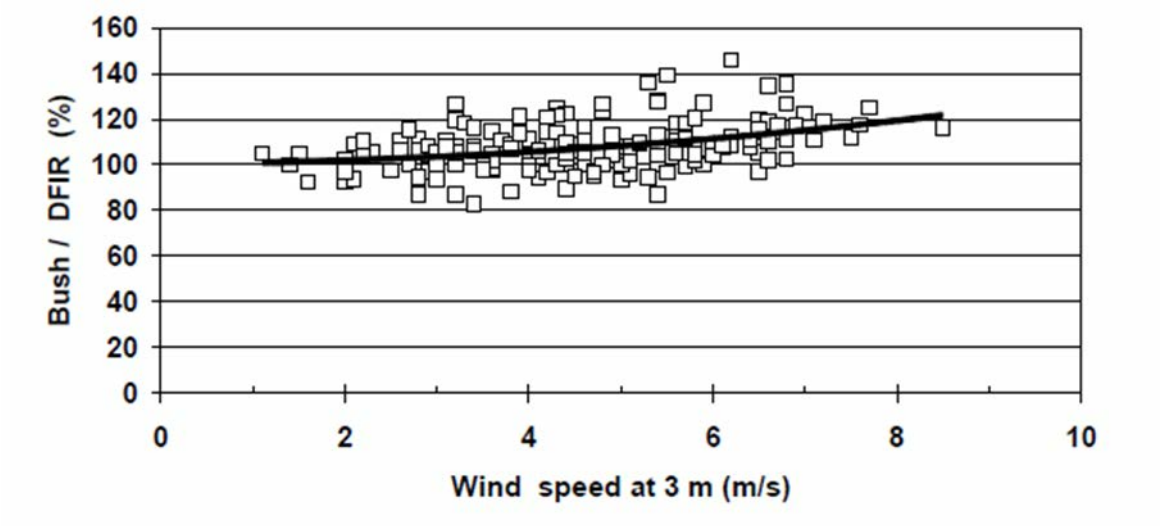


Figure 3.3. Catch ratio of the bush gauge vs. DFIR measurements as a function of wind speed and the associated transfer function (curve fitting) derived for Valdai Experimental Station, former USSR (WMO/TD - No. 872, 1998).

3.1.3.1.4 *References in 1986-1993 WMO solid precipitation intercomparison*

In preparation for the *WMO Intercomparison on Solid Precipitation, 1986-1993*, the Organizing Committee examined a range of previously recommended configurations. These were*:*

- Bush-shield: bush encircling the gauge and cut-off regularly to the level of the gauge orifice
- Double-fence: large octagonal or 12-sided, vertical or inclined lath fences encircling the gauge. The diameter of the outer fence is 6-12 m and that of the inner fence 3-4 m.
- Forest clearing: distance from trees to the gauge roughly equals the height of the trees.
- Snow board measurement: taking into account the melting and evaporation of snow during periods when no snow drifting or blowing occurs
- Dual-gauge approach: two adjacent gauges, one shielded and one unshielded

Recalling previous results, it was acknowledged that a gauge situated in a natural bush shelter would provide the best estimate of "ground true" precipitation (i.e. the highest amount) and was considered as the primary standard. The hydrological station at Valdai was and has remained the only site where DFIR measurements were assessed against measurements from gauges surrounded by bush (maintained at gauge height).

3.1.3.2 Configuration and operation of the SPICE field working reference systems

The experience of the *WMO Solid precipitation Intercomparison of 1986-1993* has played a significant role in the definition of the references for SPICE. Additionally, in preparation for SPICE, the criteria have been refined through experiments conducted during the winter of 2011/12 in Canada, Germany, Finland, Italy, Norway, Switzerland, and the USA.

For the configuration of the WMO SPICE references, the IOC adopted the octagonal double fence as defined by WMO/TD - No. 872, 1998 and used automatic gauges in place of the manual Tretyakov gauge.

To clearly differentiate the octagonal double fence from the complete FWRS, the IOC decided to use the term DFIR-fence when referring only to the octagonal double fence. Furthermore, the IOC

decided (IOC-2 Final Report, Boulder, 2012) to refer to the configuration consisting of a DFIR-fence with a shielded automatic instrument in the center as the double fence automatic reference (DFAR), as presented in Figure 3.4. Since 1985, DFIRs have been operated for research purposes, at many locations around the world. Theriault et al. 2015 has shown that the orientation of the wind to the octagonal double fence may impact the refence amount collected by the DFAR system.

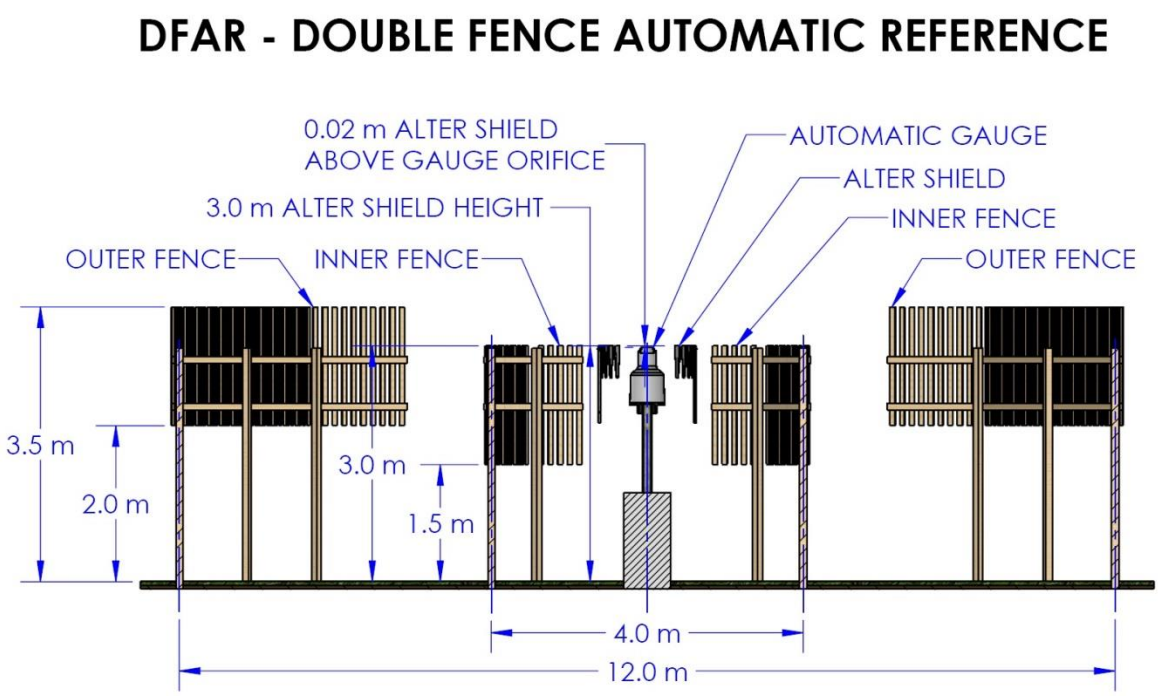


Figure 3.4. DFAR Cross-section (Drawing by J. Hoover, Environment and Climate Change Canada).

The first meeting of the IOC recommended the FWRS configurations for SPICE and the nomenclature for the intercomparison sites (IOC-1 Final Report, Geneva, 2011).

The IOC decided that the participating sites are responsible for purchasing the instruments for their own reference systems and that data from the gauges used as references would not be shared with the instrument providers during the duration of SPICE.

Additionally, the IOC recommended that the clearance below the outer fence of the DFIR-fence should be 1.5 m above the 30-year-average maximum snow height (IOC-2 Final Report, Boulder, 2012) at the site.

Several levels of reference systems have been defined to allow for the organization of tests on sites with various conditions and facing different capacity limitations. These levels are outlined in the following sections. A list of sites and corresponding reference levels can be found in Table 2.1.

3.1.3.2.1 *Field working reference system type R0*

The FWRS type R0, the bush gauge, comprises multiple manual Tretyakov gauges, each with a Tretyakov shield, surrounded by a uniform bush growth of the same height as the gauges (WMO/TD-No. 872, 1998). One of these gauges is also surrounded by a wooden fence, as shown in Figure 3.5.

The only site with this configuration continues to be the Hydrological Station at Valdai (Russian Federation), also one of the SPICE sites.

A site hosting an R0 reference is designated as an S0 type SPICE site.

Figure 3.5. Field working reference system type R0 (the bush gauge) as configured at the Valdai SPICE site, Russian Federation.

3.1.3.2.2 *Field working reference system type R0a*

To reflect the need to characterize a field reference for increased temporal resolutions, the IOC defined the R0a FWRS. This configuration is effectively an R0 system in which the manual Tretyakov gauges have been replaced by one or more single-Alter (SA) shielded automatic gauges. (See technical specifications of the single-Alter shield used in this FWRS in Figure 3.4.)

A site hosting an R0a reference is designated as an S0 type SPICE site.

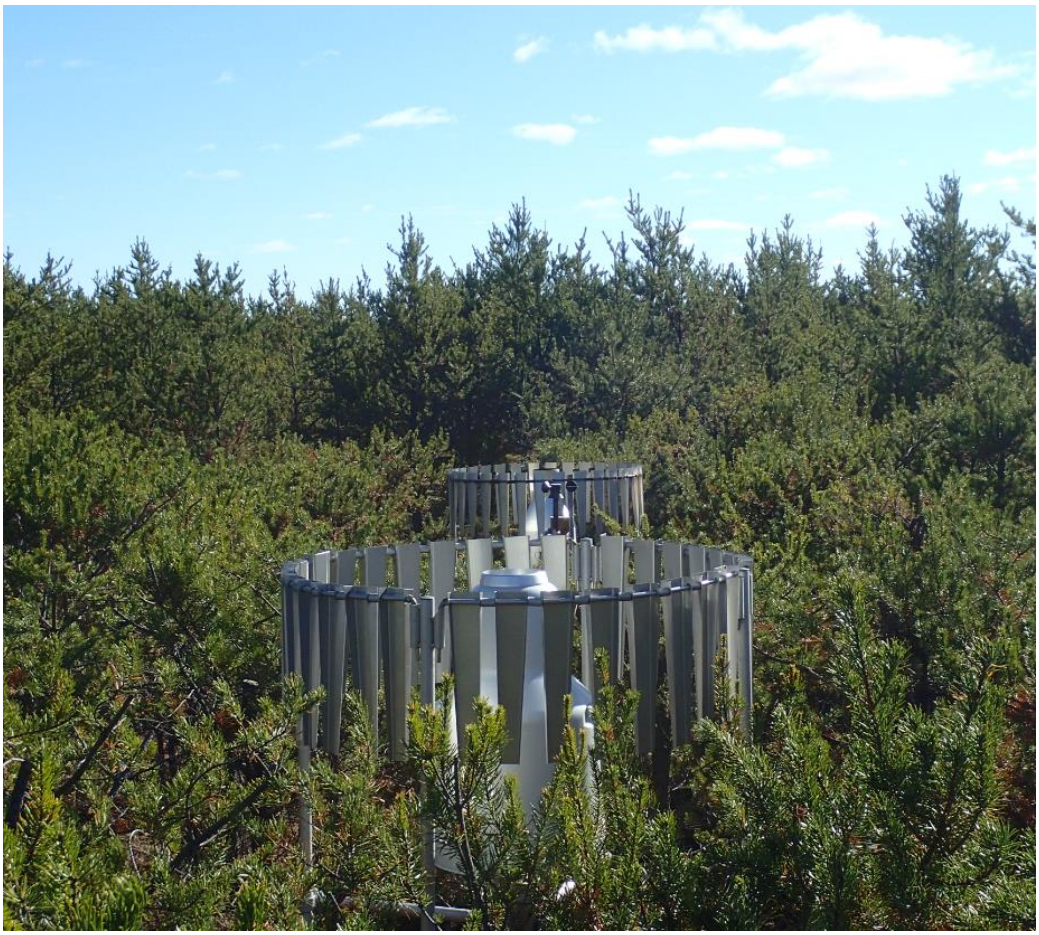


Figure 3.6. Field working reference system type R0a, Caribou Creek SPICE site, Canada. Note the two different automatic gauges in similar configurations.

3.1.3.2.3 *Field working reference system type R1*

The FWRS type R1 is the secondary field reference established by the WMO solid precipitation intercomparison of 1986-1993 (WMO/TD - No. 872, 1998). It comprises a DFIR-fence with a manual Tretyakov gauge and a Tretyakov shield. (See Figure 3.7.)

A site hosting an R1 reference is designated as an S1 type SPICE site.

Figure 3.7. Field working reference system type R1, as configured on the CARE SPICE site, Canada.

3.1.3.2.4 *Field working reference system type R2*

The FWRS type R2 consists of a SA-shielded automatic weighing gauge within a DFIR-fence, as shown in Figure 3.8. A precipitation detector is located between the inner fence of the DFIR-fence and the SA shield. (See Figure 3.13.)

A site hosting an R2 reference is designated as an S2 type SPICE site.

Figure 3.8. Field working reference system R2, as configured on the Sodankylä SPICE site, Finland.

3.1.3.2.5 *Field working reference system type R3*

The FWRS type R3 consists of a pair of identical automatic weighing gauges, one unshielded and the other single Alter (SA)-shielded, together with a precipitation detector located in close vicinity to the weighing gauges. (See Figure 3.9.)

A site operating only an R3 reference is designated as an S3 type SPICE site.

At sites where an R2 reference is also available, the type of gauges used for the R2 and R3 configurations are identical.

Figure 3.9. Field working reference system R3, as configured on the CARE SPICE site, Canada.

For the R3 FWRS, a minimum orifice height above ground of 2 m (orifice height may be higher, depending on the expected amount of snow) was adopted. It is expected that the gauge orifice will be located 1.5 m above the maximum height of the snow pack, as identified from the 30-year climate normal.

The upper rim of the single-Alter shield must be 2 cm above the gauge orifice (IOC-2 Final Report, Boulder, 2012).

The concept for the R3 reference used in SPICE was first introduced by Hamon (Hamon, 1973), who computed actual precipitation from data collected with one shielded and one unshielded gauge. Using what is known as the dual-gauge procedure, Hamon could account for precipitation losses due to the influence of wind.

Further assessments of this configuration as a field reference for the measurement of solid precipitation were conducted by Hanson (Hanson, 2004).

The configuration of shielded and unshielded gauges in the R3 FWRS is representative of the operational configurations used worldwide, as documented in the WMO-CIMO survey conducted in

2008 (Nitu and Wong, 2008). The rationale related to the representation of operational configurations was also applied during the previous WMO solid precipitation intercomparison.

3.1.3.3 Configuration of references on participating sites

The IOC decided that all sites hosting SPICE experiments and focusing on measuring precipitation amount should have at least an R3-type reference. This establishes a baseline for traceability to the R2 and R1 reference systems (present at some, but not all sites) and, ideally, the correlation of results from the participating sites.

It was recognized, however, that at some sites, it was not possible to install a reference configuration. As these sites presented value for the project in terms of enabling the investigation of specific issues, such as the operability of gauges in certain environments (e.g. high and/or remote mountain environments), these were included in the experiment with the expectation of reporting on those specific topics, and their data was not used for the derivation of transfer functions. These sites are designated as S4 type sites.

3.1.3.4 Details of field reference configurations

3.1.3.4.1 *Manual measurements perspective*

The 1986-1993 WMO intercomparison focused on manual measurements, and the data analysis was conducted for measuring intervals of 24, 12, or 6 hours (synoptic scales). Automated gauges with 15-minute sampling were also included, but they were not the primary focus of the study. The automated gauges were compared with references reporting over 24-, 12- or 6-hour intervals, which were then the standard reporting intervals.

For this intercomparison, sources of error assessed included wind effects affecting catchment efficiency, wetting losses, evaporation, rising snowpack resulting in changes in the physical setup over the season, and lack of uniform snowfall over the sample period.

With manual measurements, the sampling and the measurement are physically separated. The precipitation collector samples the precipitation at its location (e.g. within the DFIR). Then the collector is removed from its perch, taken inside a shelter, weighed or melted, and the volume measured. The measurement is, therefore, made in relatively steady conditions, in a uniform environment, and temperature, wind, or other effects on the weight scale or graduated cylinder are non-existent or minimal. Also, the measurement is stable (no noise), being made with long sampling times and the minimum measurable snowfall rate is determined by the smallest value (resolution) of the snow accumulated in the collector that can be measured with a weighing scale or by the graduated cylinder over the sampling period (generally, of several hours).

Additionally, the measuring interval and the response time of instruments used are not factors in the measurement, as the human observer would normally wait for the measuring device to stabilize before recording the reported value. The manual measurements were made at uniform intervals, and implicitly, the measurements could be interpreted as interval averages for snowfall rate.

The wind measurements corresponding to the manual precipitation measurements were averaged over the sampling interval. The wind conditions during precipitation events within sampling intervals were not identified, and correlations between the wind speed and precipitation occurrence within the sampling intervals were not considered. The winds may vary significantly within the sampling intervals from a period with precipitation to a period with clear conditions, and so the average wind speed over the entire sampling interval may not appropriately represent the wind conditions when it

is precipitating. In turn, any adjustment functions derived from the catch efficiency-wind-speed relationship may not be representative of the conditions under which the precipitation was collected (cf. description of the procedure for manual observation in Annex 7).

3.1.3.4.2 *Automatic gauges for the SPICE FWRS*

The automatic instruments for the FWRSs of SPICE were selected using principles similar to those used for the selection of reference gauges in the 1986-1993 WMO intercomparison. The instruments selected were those with the broadest operational use and with good documentation of their performance.

The 2008 WMO CIMO survey (Nitu and Wong, 2008) showed that of the instruments used operationally for the point measurement of precipitation amount, about 18% were automatic, and practically all were of catchment-type (tipping bucket-type and weighing gauges). A catchment-type gauge measures and reports only the quantity collected in its bucket(s) as detected by its transducer(s), scale, or tipping element.

Although the tipping-type gauges are more prevalent, as noted in the CIMO Guide, Part 1, Chapter 6, Measurement of Precipitation, "only the weighing-type (gauge) is satisfactory for measuring all kinds of precipitation, the use of the tipping bucket-type of precipitation gauges being for the most part limited to the measurement of rainfall." About 16% of respondents to the 2008 CIMO survey noted the use of automatic weighing-type precipitation gauges for measuring and reporting the amount of precipitation, primarily in North America and Europe.

An automatic weighing gauge weighs the precipitation collected in its bucket and calculates the precipitation amounts based on the detected mass changes of the content of the bucket or of the load. Its measurement capabilities could be characterized reasonably well under controlled conditions, relative to measurement standards (e.g. laboratory calibration of sensing elements, traceable to the gram, in the International System of Units).

The ability of weighing gauges to accurately report the amount of falling precipitation in the outdoors, however, is significantly influenced by the environment and the characteristics of precipitation.

Two gauges with wide operational use have been accepted for use in the SPICE FWRS. These are the Geonor T-200B3 gauge with three transducers and the OTT Pluvio² gauge. Given the models currently in use, the IOC agreed that Geonor gauges with 600 mm and 1000 mm capacities and the Pluvio² gauge with 200 cm^2 inlet opening were suitable for use as part of the SPICE FWRS. The decision for using either one of these gauges was based on their comparably broad use operationally (Figure 3.10) and their similar performance in the 2011/12 winter studies, which were organized in preparation of the formal launch of SPICE.

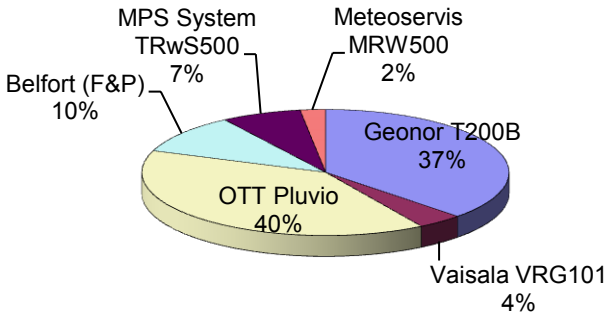


Figure 3.10. Weighing-type gauges used operationally (2008 CIMO Survey, Nitu and Wong, 2010).

Emerging technologies using heat and mass transfer (the hot plate), particle scattering (light systems), or radar (POSS, PLUDEX) were not considered as primary instruments for the SPICE reference due to lack of history, widespread experience, and full characterization. Some of these new technologies may be much more sensitive than the traditional catchment systems; however, depending on the principle of measurement, there are known limitations to their performances. For example, the laser systems are very sensitive, but are also limited in the capability to measure a large dynamic snowfall intensity range due to attenuation or saturation of the receiver.

Weather radar systems are also very sensitive, as demonstrated by the results in Figure 3.11, and allow to detect lower snowfall rates than typical weighing gauges. Nevertheless, the snowfall rates are then measured at altitude and not at the surface, and assumptions are required to estimate the snowfall density.

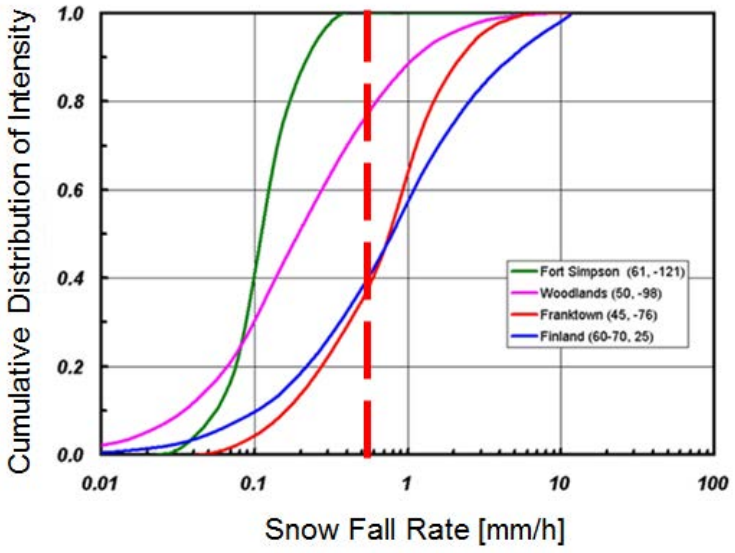


Figure 3.11. Data from a typical C band weather radar during snowfall. The vertical dashed line is set at 0.5 mm/h snowfall water equivalent, with the coloured curves various lines representing data from different locations (latitude and longitude are indicated within the brackets).

3.1.3.4.3 *Calibration of weighing gauges used in the FWRS*

The automatic weighing gauges used in the FWRS were calibrated at the beginning of the intercomparison using gauge-specific methods. For the OTT Pluvio^2, the calibration procedure from the user's manual was used. For the Geonor, a specific field verification procedure was developed. (See Annex 7) Additional calibrations were performed following each season of experiments, following procedures recommended by the manufacturer.

Over the course of the intercomparison, no adjustments in transducer-specific coefficients for Geonor gauges and no updates to the firmware version for Pluvio^2 gauges were performed. The firmware version of Pluvio^2 gauges used in the FWRS was 1.30.1, as available at the onset of the intercomparison.

3.1.3.4.4 *Heating of weighing gauges used in the FWRS*

Based on the operational practices of several participating countries (e.g. Rasmussen et al. 2001, USA) and results obtained prior to the formal start of SPICE tests, all gauges used as part of the FWRS were heated.

Heating of the gauge inlet is typically a tradeoff. Heated gauges provide a more timely response to snowfall events. The risk of capping of the gauge, in which snow accumulation on the gauge leads to partial or complete blockage of the orifice, is reduced. These benefits likely outweigh the disadvantages which include a possible "chimney effect" (in which buoyant, warm air disrupts the flow field above the gauge orifice) or evaporation/sublimation of precipitation from the inside rim before it is collected and measured, impacting the collection and measurement of precipitation. Details of heating are provided Section 4.2.1. The potential negative impacts of heating the DFAR are examined briefly in Section 3.2.2.3.4.6 and may warrant more study.

For the Geonor gauges, the heating algorithm and physical configuration of heaters is based on the method used operationally in the Climate Reference Network (CRN) of the US National Oceanographic and Atmospheric Administration (NOAA), and on NCAR's previous work for the United States Federal Aviation Administration (FAA) at the Marshall site. This algorithm attempts to maintain the temperature of the orifice rim at 2 °C while the ambient temperature is between 2 °C and -5 °C. Additionally, for temperatures below -5 °C, the heaters are activated once every 24 hours.

The heating of Pluvio^2 gauges used in the FWRS was implemented using the built-in heaters and on-board commands, and applying the same algorithm defined above for the Geonor gauges.

The Geonor inlet was heated over its entire length using two heaters, while the Pluvio^2 gauges had heat applied to the rim only, as designed by the manufacturer.

Following the 2012/13 season, there was concern from some colder and windier sites (e.g. Haukeliseter, Norway) regarding the value of the ambient temperature at which the heaters should be turned off. Based on the experience of some sites, capping had been observed at temperatures well below -5 °C. As a result, at the IOC-4 meeting (IOC-4 Final Report, Davos, 2013), it was decided that the temperature heaters was to be maintained, to ensure a rim temperature at +2 °C to +3 °C, for ambient temperatures of +2 °C and below, for both Pluvio^2 and Geonor gauges used in the reference.

3.1.3.4.5 *Use of antifreeze and oil for weighing gauges in the FWRS*

The IOC agreed that antifreeze and oil "charges" were mandatory for the gauges in the reference systems, in order to prevent the freezing of the bucket contents and to limit evaporation. Given the

wealth of knowledge of the national regulations and context, each participating team was asked to identify the compositions and quantities of antifreeze used for the experiment on each site, reflecting the national procedures and experience.

The IOC requested that an oil film be used for all reference gauges to prevent evaporation and also to mechanically minimize the hygroscopic effect of the antifreeze (powercool or propylene glycol/water mixture). Careful handling and disposing of the waste was mandatory.

Given the previous experience of several members, it was recommended that the Geonor gauges used as FWRS should never be left empty, as this could lead to a measurement error. These gauges should always be filled to at least 25% capacity.

Further considerations, including experiences and recommendations, can be found in Section 4.2.3. A site report on tests performed in a cold chamber to investigate different types of oil and antifreeze is available in Annex 7.

3.1.3.4.6 *Alter-shield configuration*

The weighing gauges in the DFIR-fence (R2 reference system) and one of the gauges in the R3 reference system were installed with Alter-type windshields. The same type of Alter shield was used in all R2 and R3 reference configurations, as well as for any other configurations included in the intercomparison where a single-Alter shield was used. (See Figure 3.12.)

The SPICE-recommended single-Alter shield configuration is provided in Annex 7.

Figure 3.12. Single-Alter shield (Bratt's Lake SPICE site, Canada).

3.1.3.5 Use of precipitation detectors in the SPICE FWRS

To help minimize false reports of falling precipitation and to increase the reliability of the reference, the binary output (Yes/No) of a precipitation detector was used as an additional information for deriving the reference data.

At the IOC-2 meeting (IOC-2 Final Report, Boulder, 2012), the use of a capacitive precipitation detector was recommended. At the same time, the IOC recognized that optical precipitation detectors are more sensitive by one order of magnitude or more, and encouraged the participating sites to add them, where possible.

During the IOC-4 meeting (IOC-4 Final Report, Davos, 2013), the committee revised its 2012 decision on the use of capacitive precipitation detectors based on findings from the 2012/13 SPICE season. It recommended that all sites use an optical precipitation detector for the FWRS, replacing the capacitive precipitation detector. The IOC strongly encouraged all sites, but in particular those operating an R2 reference, to use a laser-disdrometer-type instrument as the precipitation detector (e.g. Thies Laser Precipitation Monitor or OTT Parsivel^2) and, where possible, to collect the data on size and fall velocity distribution of the particles.

For an R2 reference, the location of the precipitation detector and/or the disdrometer-type sensor is within the inner fence of the DFIR-fence, equidistant from the Alter shield and the inner fence. The sensor was positioned 75 cm below the gauge orifice (corresponding to halfway down the inner fence) and perpendicular to the prevailing wind direction. (See Figure 3.13.) If possible, a second sensor was mounted on the experimental field to account for different wind directions.

At sites without a DFIR-fence, the precipitation detector was mounted in a wind-sheltered location or was suitably shielded.

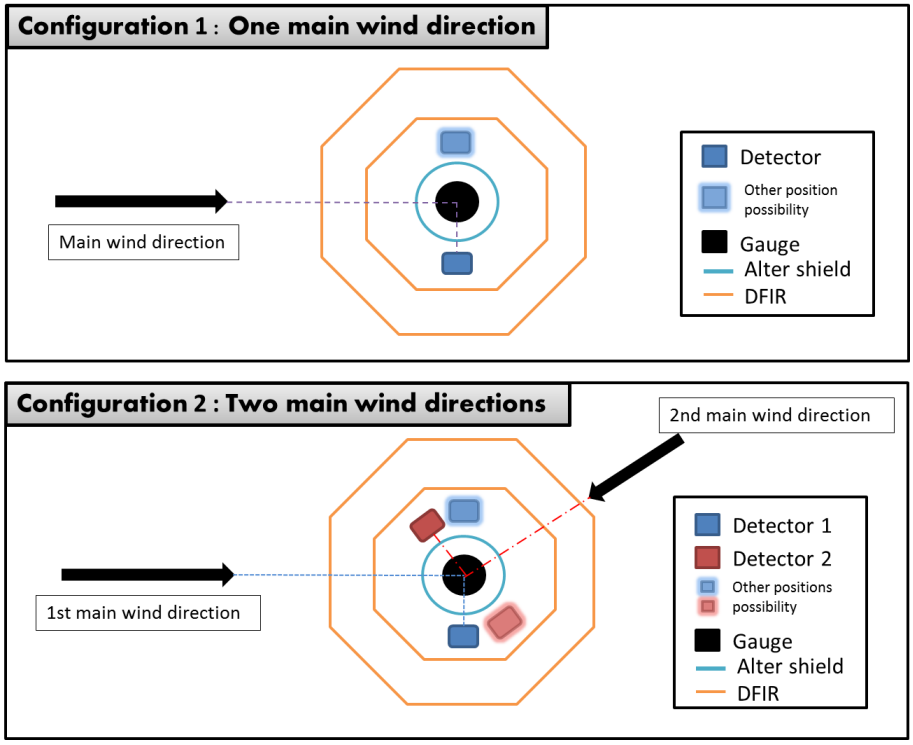


Figure 3.13. Location of the precipitation detector within the DFIR-fence in the R2 FWRS.

Figure 3.14. Thies LPM inside a DFIR-fence in R2 FWRS (CARE, Canada).

3.1.3.6 Configuration of SPICE FWRS by site

All participating sites configured their FWRS(s) as summarized in Table 3.1, below.

Table 3.1. Configuration of SPICE FWRS by site. "P" refers to reference systems using OTT Pluvio^2 gauges; "G" refers to reference systems using Geonor T-200B3 gauges.

Site	*R0*	*R0a*	*R1*	*R2*	*R3*	*Site designation*
Guthega Bay, Australia					R3G	S3
Bratt's Lake, Canada				R2G	R3G	S2
CARE, Canada			R1	R2G	R3G,P	S1
Caribou Creek, Canada		R0aG,P		R2G	R3G	S0
Tapado, Chile					R3G	S3
Sodankylä, Finland				R2P	R3P	S2
Col de Porte, France					R3G	S3
Joetsu, Japan				R2G	R3G	S2
Rikubetsu, Japan				R2G	R3G	S2
Gochang, Rep. of Korea				R2G	R3G	S2
Mueller Hut, New Zealand					R3G	S3
Haukeliseter, Norway				R2G	R3G	S2
Valday, Russian Federation	R0		R1	R2P	R3P	S0
Volga, Russian Federation			R1			
Weissfluhjoch, Switzerland				R2P	R3P	S2
Marshall, USA			R1	R2G, P	R3G,P	S1

3.1.3.7 Data-sampling strategy for SPICE FWRS

The intercomparison results are based on datasets tailored to specific objectives. One-minute datasets are the baseline for the SPICE analysis and were used to derive additional datasets. Of particular consideration is the derivation of precipitation event datasets. The data-derivation strategies are outlined in subsequent sections.

At its second meeting, the IOC recommended a data sampling and reporting interval of 6 seconds for all reference gauges, to the extent feasible (IOC-2 Final Report, Boulder, 2012). It determined that the "rawest" data and the "highest" temporal rate should be collected and be used in order to understand the signal and data processing performed by the gauge firmware. Generally, accessing the signal data is not possible, as these data are not available and/or proprietary.

Where the data collection is not feasible at 6-s frequency, one-minute intervals are acceptable. Similar sampling strategies are used for instruments under test and for ancillary data.

3.1.3.7.1 *Comparison of Geonor and Pluvio^2 data output*

The two gauges selected for use in the SPICE FWRSs, the Geonor T-200B3 and OTT Pluvio^2, operate on different principles and use different sensing elements, data collection, and processing. These differences are summarized in Table 3.2 below.

Table 3.2. Comparison of gauge characteristics and operation for Geonor T-200B3 and OTT Pluvio^2 weighing gauges.

Gauge characteristic	**Geonor T-200B3** **Reference: Geonor T-200B Precipitation Gauge User Manual, Rev: GU 20030829 and Bakkehøi et al, 1985**	**OTT Pluvio^2** **Reference: Operating instructions Precipitation Gauge Pluvio^2**
Operating principle	The bucket content is measured with a high-tension vibrating-wire (VW) transducer. Under load, the wire vibration frequency is related to the weight detected (P) based on a quadratic relation: $P = A (f – f0) + B (f – f0)^2$ Where: P = precipitation (in cm) f = frequency reading (Hz) A = Calibration constant, *given* B = Calibration constant, *given* f0 = frequency with empty bucket at calibration (Hz), *given*	The bucket content is weighed using a high-precision stainless steel load cell, hermetically sealed against environmental influences. Through internal processing (proprietary algorithm), the precipitation gauge determines the weight of the bucket and its content every 6 seconds with a resolution of 0.01 mm. The difference between this measurement and the base weight of the empty bucket gives the current bucket content.
Number of sensing elements per instrument	SPICE FWRS requires that three transducers are used per gauge.	One load cell per instrument in the FWRS.

Data sampling	The signal from a transducer is amplified into a measurable quantity read with an external data logger. The data logger samples the transducer signal by means of a user-defined strategy and logger-specific functions. The measured parameter is a frequency. The sampling of the transducer signal is, generally, not continuous. For a 600 mm gauge, the empty gauge frequency is about 1000 Hz, while a full gauge output is about 3000 Hz.	The gauge includes its own onboard processing capabilities. In the letter to IOC, dated August 23, 2012, OTT Hydromet indicated that "the Real-Time (RT) Bucket Weight (referred to as "Bucket RT" in the manual) is computed as the 6 seconds arithmetic mean of the load cell measurements sampled with a rate of 125 Hz, with static temperature correction and conversion from weight to precipitation in mm with appropriate scaling factor depending on the Pluvio^2 version. It has to be noted that Bucket RT data are not stated as precipitation output data."
Temperature dependency	The temperature of transducers is not monitored. Experimental work demonstrated that there is a correlation between the air temperature and the transducer response, and experimental temperature coefficients have been proposed, but not widely implemented (Duchon, 2004, 2008).	An integrated temperature sensor monitors the load cell temperature and an internal algorithm compensates for the temperature changes in the balance system. In the letter to the IOC, dated August 23, 2012, OTT Hydromet indicated that, "The temperature correction is gauge specific and is obtained as a result of a static laboratory calibration of each instrument over the entire temperature compensation range. The relevant temperature for this correction is measured by an internal temperature sensor and the temperature correction factors are stored in the non-volatile memory of each instrument."
Data processing	An external data logger processes the transducer output into a precipitation amount using a user defined program and transducer calibration constants provided in the calibration certificate. Data from a Geonor gauge is obtained from a system that includes the gauge, a datalogger, and user-defined processing (datalogger program).	The on-board gauge hardware processes the high-frequency data samples into several data products (see ** below). These incorporate temperature corrections (static and dynamic), evaporation corrections, wind-pumping filtering, noise filtering, and other proprietary processing.

Data output interval	The gauge signal is continuous. The data logger samples and records/reports the output as defined by the user. (See * below).	The gauge sends a preconfigured message in response to a poll command. The minimum interval at which a new message is available is 6 seconds. Every 6 seconds the Pluvio^2 calculates the bucket content using multiple raw values. Special filter algorithms are used (wind, temperature, evaporation).

*In cases where the Geonor T-200B3 was connected to a Campbell Scientific CR3000 data logger (the case for most sites using the Geonor T-200B3 as the reference gauge), the CR3000 offers the following two methods for measuring Geonor T-200B transducers:

1. PeriodAvg: This command calculates the sensor's average frequency over a specified number of cycles within a defined interval. The CR3000 permits a maximum interval of 1 second for this measurement and is user defined. The number of cycles set to be read, and how often they are read, can vary from user to user. For example, the CARE SPICE site uses a sampling strategy that consists of reading 250 cycles every 6 seconds for each Geonor transducer; an associated measurement timeout of 300 milliseconds is used. The time associated with the reading of the 250 cycles is converted to frequency by the logger.
2. PulseCount: This command calculates the sensor's frequency by counting the number of pulses over a specified time period. The CR3000 does not restrict the time period for this measurement. The number of pulses counted over the given interval is converted to frequency.

Comparison:

- PeriodAvg and PulseCount methods are comparable in terms of accuracy.
- PeriodAvg is preferable for fast measurements (less than 1 sec) and freeing the CR3000 pulse counter channel for other functions.
- PulseCount is preferable for continuous measurements over a longer time period, from several seconds to minutes (Duchon, 2004).

** OTT Pluvio^2 data products (note that RT indicates 'real-time' outputs and NRT indicates 'non-real-time' outputs, as detailed further below):

- Intensity RT (fixed-update interval: 1 minute)
- Accumulated RT/NRT (since the last measurement sample)
- Accumulated NRT (since the last measurement sample)
- Accumulated total NRT (since the last reset)
- Bucket RT
- Bucket NRT
- Temperature load cell
- Status Pluvio^2 (since the last measurement sample)

The data are available as real-time and non-real-time values. For real-time outputs, the measurement is available within 1 minute of the precipitation event occurring. For non-real-time

outputs (NRT), the Pluvio^2 outputs the measurement with a 5 minute delay. If very light precipitation is involved (< 0.1 mm/min), the output delay is up to 65 minutes.

The IOC recommended that for Geonor gauges, the preferred method of frequency measurement using CR3000 data loggers is the period-averaging method, because of the increased temporal resolution (IOC-2 Final Report, Boulder, 2012).

As the Pluvio^2 gauges use a proprietary algorithm to collect and process the gauge measurements, the IOC agreed to poll the gauges every 6 seconds, as much as feasible, to maximize the availability of gauge data. Longer intervals are acceptable, depending on site capabilities. Additionally, the IOC asked OTT Hydromet for information on the derivation of the Bucket RT data; the information provided is included in Table 3.2.

3.1.4 Configuration of field references for the measurement of snow on the ground

Authors: Craig Smith, Rodica Nitu, Samuel Morin

3.1.4.1 Overview

The measurement and reporting of snow depth and its linkage with snowfall are key deliverables of SPICE. SPICE recommends appropriate automated field reference system(s) for attended and unattended measurements of snow depth, and provides guidance on the performance of modern automated systems used operationally.

Additionally, the objectives of SPICE included the assessment of the capabilities of automated sensors to determine the SWE of accumulated or freshly fallen snow, linking these measurements to the site reference gauge precipitation measurements and snow depth measurements (where possible).The *2009 International Classification for Seasonal Snow on the Ground* prepared by the International Association of Cryospheric Sciences (IACS) Working Group on Snow Classification (Fierz et al., 2009) defines the parameters related to snow accumulated on the ground and their standard methods of measurement. These have been used as guidance by the IOC in defining the configuration of field reference systems for each of these parameters. They are:

(2.3) Height of snowpack, snow depth (HS)

Snow depth denotes the total height of the snowpack, i.e., the vertical distance in centimeters from base to snow surface. Unless otherwise specified snow depth is related to a single location at a given time. Thus, manual snow-depth measurements are often made with one or more fixed snow stakes. On the other hand, portable snow depth probes allow for measurements along snow courses and transects. Automated measurements of either snow depth or snow thickness are possible with ultrasonic and other fixed and portable snow-depth sensors.

(2.4) Height of new snow, depth of snowfall (HN)

Height of new snow is the depth in centimeters of freshly fallen snow that accumulated on a snow board during a standard observing period of 24 hours. Additional observation intervals can be used, but should be specified. Height of new snow is traditionally measured with a ruler. After the measurement, the snow is cleared from the board and the board is placed flush with the snow surface to provide an accurate measurement at the end of the next interval.

(2.5) Snow water equivalent (SWE)

Snow water equivalent is the depth of water that would result if the mass of snow melted completely. It can represent the snow cover over a given region or a confined snow sample over the

corresponding area. The snow water equivalent is the product of the snow height in meters and the vertically-integrated density in kilograms per cubic meter (Goodison et al., 1981, p. 224). It is typically expressed in millimeters of water equivalent, which is equivalent to kilograms per square meter or liters per square meter, thus referring to the unit surface area of the considered snow sample.

A key factor influencing the measurement of snow on ground and snowfall is the challenge due to drifting snow conditions and the associated challenge in obtaining a representative "mean" value of snow depth or height of new snow, using a point measurement. When the snow depth is measured by an observer, the observer's judgment and the availability of multiple measurements over a representative area would likely ensure the accuracy of measurement (Goodison et al., 1981, p. 192). This becomes a significant challenge when the point measurements are taken at sites where only automatic instruments are available. In SPICE, only a handful of sites were able to organize and sustain reference measurements taken by human observers at standard intervals. Given this limitation, as well as the fact that SPICE assesses the reporting of snow on ground and snowfall over much shorter intervals (hours, minutes) than are typical of manual measurements, other methods have been implemented to provide the reference observations for the assessment, as described in the following sub-sections.

3.1.4.2 Definition of SPICE field references for the measurement of SoG

The field references for the measurement of SoG were defined during the IOC-4 meeting (IOC-4 Final Report, Davos, 2013).

3.1.4.2.1 *Total snow depth*

Recognizing that all sites did not have the same resources for making manual snow-depth measurements, the SPICE reference measurement for total snow depth, referred to as the snow total reference (STR), is divided into four classifications:

STR0: Two or more ruler-based manual measurements at the periphery of the footprint, outside the field of view (FOV) of each automatic sensor, conducted at least once per day, at the same time, with minimum disturbances of the snow pack under and around the sensor. An observer is required. Although this is the Level 0 reference, it creates site disturbance (snow-pack modification), which needs to be considered.

STR0a: Manual observation of four graduated stakes at the corners of the automated snow-depth-sensor footprints at least once per day. An observer is required. Graduated stakes should have cm graduations and be observed as close as possible to the level snow pack to the nearest half centimeter. The stakes should be placed 40 cm outside of the sensor FOV to avoid impacting the snow characteristics within the sensor FOV.

STR0b: Hourly camera observations of four graduated stakes at the corners of the automated snow-depth-sensor footprints. Where possible, small LED lights should be installed to enable nocturnal observations. No observer is required, although extracting the snow depth valuess from the photos is typically a manual exercise.

STR1: Manual snow-depth transect of a minimum of 10 points (preferably at fixed points or using graduated stakes at fixed intervals of 3-10 meters), conducted at least once per day near the automatic snow-depth-sensor array to assess the variability of snow depth over the observing site. An observer is required, and it should be recognized that the integrity of the snow pack in the observation field needs to be preserved.

3.1.4.2.2 *Snow water equivalent*

The reference measurement for SWE is designated SWR0 and is a manual measurement conducted biweekly near each automatic snow-depth sensor, just outside the FOV, following procedures described in the *WMO Guide to Hydrological Practices, Volume 1* (WMO-No. 168). Samples should not be taken within 30 cm of a previous sample, and the core can be used to partially refill the sample hole. Known snow-sampler biases/errors need to be considered (Farnes et al., 1983). Snow-pit measurements require a larger distance between observation locations and will be farther away from the SWE sensors by necessity. A precise description of equipment and procedures used at each site is required. This manual measurement requires an observer.

The reference measurements used at each site are described in greater detail in subsequent sections and summarized in the instrument performance reports for each SoG instrument. (See Annex 6).

3.1.4.3 SoG reference methods

A consistent reference configuration for all sites was desired; however, differences in local conditions and availability of resources (e.g. human observers) led to a degree of variability among the SPICE sites examining the performance of instruments reporting SoG. Table 3.4 summarizes, by site, the snow depth and SWE reference measurements used for the intercomparison analysis.

Table 3.3. List of primary reference techniques for SPICE SoG sites.

SPICE Site	Manned (M) Unmanned (U)	Type of reference	Primary Reference Technique
CARE	M	STR0a, STR1	Daily visual snow stake observations
Caribou Creek	U	SWR0, STR0b	Bi-weekly snow surveys, hourly web-cam photos of snow stakes (when available)
Col de Porte	M	STR0b, SWR0	Weekly snow ruler and snow profiles, hourly web-cam photos of snow stakes (when available)
Forni Glacier	U	STR0b	Hourly photos of snow stakes
Gochang	M	STR0a, STR0b	Hourly web-cam photos of snow stakes
Pyramid Observatory	M	STR0a	Daily visual snow stake observations
Sodankylä	M	STR0a, SWR0	Bi-weekly snow surveys, hourly web-cam photos of snow stakes
Weissfluhjoch	M	STR0a, SWR0	Daily visual snow stake observations, daily snow board (new snow) observations, bi-weekly snow profiles

The primary (non-sensor) reference techniques outlined in Table 3.4 do not involve the use of automated sensors. An alternative means of obtaining a reference measurement is to use a mean value obtained from all of the automated sensors in the intercomparison, according to the principles outlined in Section 3.6.2.2. By averaging multiple sensors and sensor types, systematic and random biases related to independent sensors should average to zero. This method is also detailed below.

3.1.4.3.1 *CARE*

The primary (non-sensor) snow-depth reference measurement for SoG at the CARE site is the daily visual observation of 62 graduated snow stakes that are distributed throughout the instrument field. Stakes are distributed across the site (see site layout in Annex 4) for an assessment of the spatial variability of snow depth. This measurement serves as the STR1 reference. Some stakes are placed in proximity to the snow-depth instrumentation, just outside the sensor FOV, as shown in Figure 3.15. This serves as the STR0a reference for the CARE site. The entire transect serves as the STR1 reference for the site.

The daily visual observation of the snow stakes occurs between 1000 and 1500 UTC, with the snow depth measured to the nearest 0.5 cm. The start and end time of the observation circuit is recorded and takes approximately 25 minutes. Some stakes are photographed by the observer to record events of interest.

The focus of the snow-depth sensor intercomparison at CARE are the sensors installed on pedestals 12A, 20, and 11A; each pedestal is configured with 3 snow-depth targets and 4 snow-depth sensors. For the intercomparison, the reference measurement is generally either a mean of the stake observations at the four corners of the target under the SUT or the mean of all 12 stake observations at each pedestal. Because of the spatial variability in snow depth at this site, it is not advisable to use the mean observations of all 36 stakes at the 3 pedestals as the reference measurement.

The automated sensor reference for CARE is the mean measurement of all 4 SUT at each pedestal (SR50A, SHM30, SL300, and USH-8) at a resolution of 1 minute. As with the manual measurements, it is unadvisable to use the mean of sensors on all pedestals as the automated reference.

There are no SWE reference measurements made at the CARE site.

Figure 3.15. Configuration of the STR0a reference configuration at CARE, with snow stakes at the four corners of the surface target, just outside the FOV of each automated sensor.

3.1.4.3.2 *Caribou Creek*

Caribou Creek is an unstaffed site, but hosts a SWE sensor provided by a manufacturer and several snow-depth sensors provided by the site host. Reference SWE measurements are made via a five-point snow-survey transect that runs north-south across the measurement clearing, perpendicular to the prevailing wind direction. This serves as both the SWR0 and SWR1 references for this site. Of the five SWE samples, two are taken in the bush south of the measurement clearing, two are taken in the clearing, and one measurement is taken on the north edge of the clearing. (See the site layout in Annex 4). The SWE sample points are 10 m apart with additional snow depths observed at 2-m intervals between SWE sample points. Although point 3 in the snow course is close to the SWE sensor, an additional sample is taken in proximity to the SWE sensors just outside of the instrument FOV.

Snow surveys are conducted at the site approximately every two weeks during the accumulation period (i.e. the winter period dominated by snowfall events prior to the start of significant seasonal melting) with the frequency increasing to weekly during the seasonal melt period (i.e. after maximum accumulation when snow is ablating through melt processes). SWE measurements are performed using an ESC-30 snow tube that obtains a bulk-density sample of a 30-cm^2 surface area. Samples are individually bagged and weighed.

Because the Caribou Creek site is unstaffed, daily snow-depth measurements are not possible. Therefore, there is no STR0a reference at this site. However, a web camera has been set up at the site to take hourly photos of the instruments in the clearing. These photos include at least one snow-depth stake installed in the site clearing (Figure 3.16) that could function as a STR0b reference. Artificial lighting allows for photos to be taken over the entire day.

There are too few automated snow-depth and automated SWE sensors at this site to derive an automated reference measurement for either snow depth or SWE.

Figure 3.16. Example of hourly photos of the snow stakes at Caribou Creek during the day (left) and night (right).

3.1.4.3.3 *Col de Porte*

Col de Porte performs manual snow-ruler and snow-pit measurements (i.e. vertical profile of the physical properties of the snow pack including layer density, depth, temperature, and crystal structure) on a weekly basis during the winter. Snow-depth measurements are made at three snow stakes installed in the intercomparison field. Two of these stakes (labeled “North” and “South”) are closer to the automated sensor than the third stake, and the preference is to use the average of these two closer stakes as the manual reference snow-depth measurement (STR0a) for intercomparison. The stakes are approximately 17 m from the mast where the automated measurements are made in the NE corner of the site. (See Annex 4.) The weekly manual measurements are made around noon, local time. For the 2014/15 season, additional manual snow stakes were installed closer to the automated instruments to be photographed with a web camera on an hourly basis and served as the STR0b reference.

The automated sensors for the intercomparison are located in the NE corner of the site and all measure the same relatively small area under the installation. A total of five sensors (SHM30, SR50A x 2, Dimetix, and Apical) are averaged to produce an automated reference measurement at a 1-min resolution.

3.1.4.3.4 *Forni Glacier*

The Forni Glacier site is unstaffed, but hosts several snow-depth and SWE instruments. The site has a time-lapse camera that takes hourly photos of four snow stakes installed at the corners of a weighing gauge; the stake observations serve as the STR0b reference. Snow depth (and possibly precipitation type) can be extracted from the hourly images.

3.1.4.3.5 *Gochang*

The Gochang site is staffed and includes several snow-depth instruments installed along the east side of the instrumentation compound. (See Annex 4.) Snow-depth reference measurements are made

using three fixed snow stakes located between the instruments. Observations of the stakes are made hourly using web cameras. These serve as the STR0b reference.

3.1.4.3.6 *Sodankylä*

The Sodankylä site is staffed and hosts several snow-depth and SWE instruments. Snow depth at the site is spatially very consistent across the instrument field. In Sodankylä, the manual snow-depth reference comprises four wooden snow stakes installed around the site (locations 22:40, 44:66, 65:57 and 65:37). (See Annex 4 for details of locations.) An automated web camera takes photos of all stakes two or three times a day. These serve as the STR0b reference for snow depth. Because there is very little sunlight in the midwinter, no fixed measurement time was set, and the daily snow depth was interpreted from the best photos available, typically around noon. The snow stakes are outside the FOV of the snow-depth sensors, but within 8.5 m of the sensors. Figure 3.18 shows some examples of the photographed stakes, and some of the challenges for manually extracting the depth information from the photographs are discussed below.

When required, maintenance personnel were sent to clean the stakes of snow using a special tool (Figure 3.17). The stakes were cleaned in this way at least once a month, and sometimes daily.

Figure 3.17. Leveling a mound around a snow stake at Sodankylä during site maintenance.

A weekly bulk-density sample is taken to measure SWE just outside the FOV of the SWE instrument at location 40:62. This measurement is made using a Finnish Meteorological Institute (FMI) snow sampler (typically plastic, height 70 cm, and diameter 10 cm) and mechanical balance showing directly SWE (Figure 3.19). This serves as the SWR0 reference for SWE.

The automated reference at this site is taken as the mean value reported by six instruments (SR50ATH x2, USH-8 x2, SHM30, and SL300) at 1-min resolution. The snow-depth instruments are clustered on the east side of the intercomparison field.

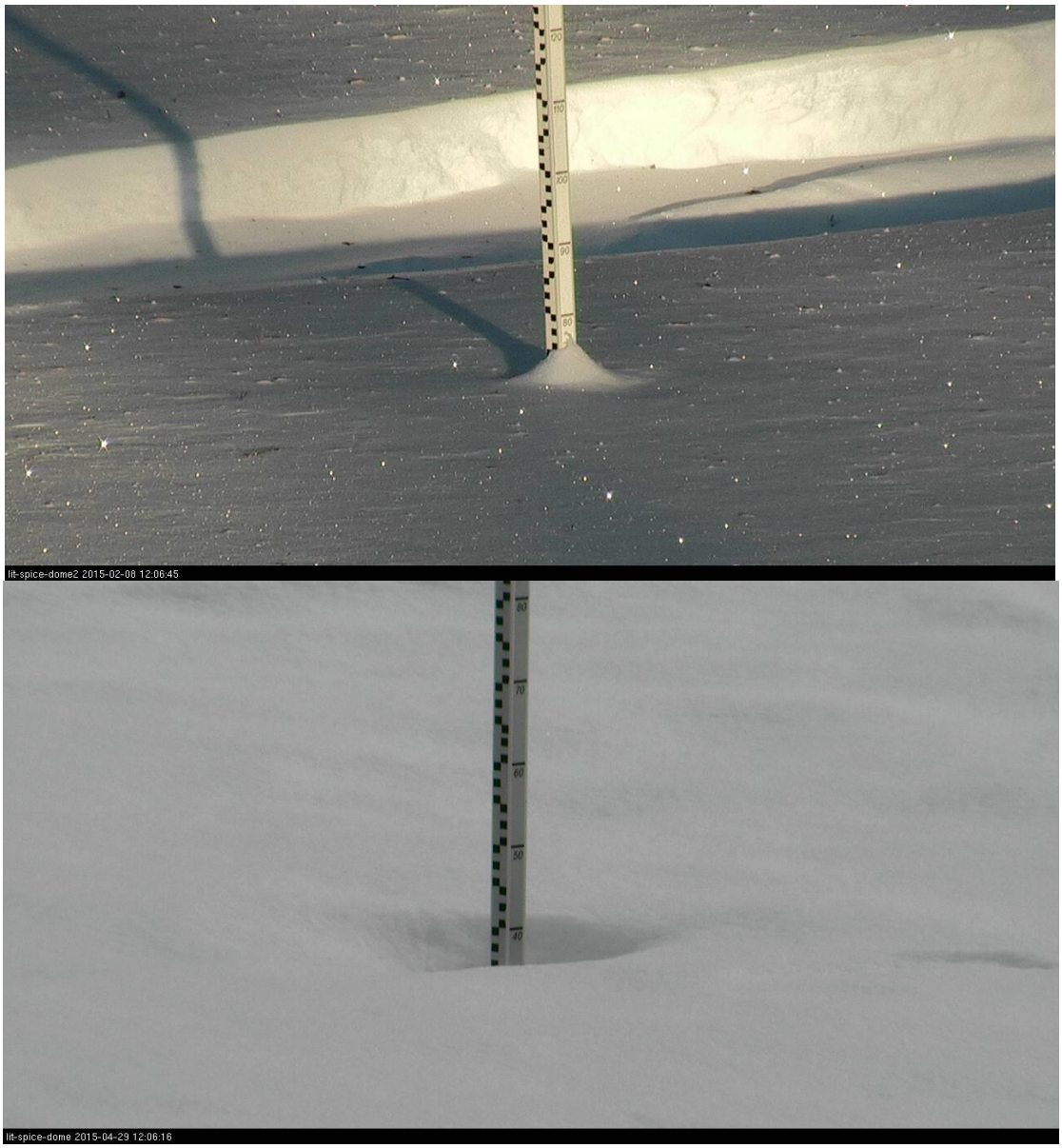


Figure 3.18. Photographs of the snow stakes in the Sodankylä intercomparison field.

Figure 3.19. Bulk-density SWE measurement using a Finnish Korhonen-Melander snow sampler.

3.1.4.3.7 *Weissfluhjoch*

Daily manual measurements are made at the Weissfluhjoch site by the Swiss Institute for Snow and Avalanche Research (SLF). The daily measurement program includes a snow-depth and SWE measurement of new snow using a weighed 1000 cm^2 sample taken from a snow board located on the southwest boundary of the instrument field. (See site layout in Annex 4.) The snow board is then cleaned and placed on top of the existing snow pack (Figure 3.20). This measurement is taken at least twice, and the results are averaged. This is not a SWR0 reference measurement, but could provide useful information for SoG analysis. Daily snow depths are also obtained via visual observation of a snow stake located approximately 17 m from the snow depth SUT, and this serves as the STR0a reference. The daily observation is made at approximately 0800 UTC. A bi-weekly snow-profile measurement is made to obtain further snow-condition information. Since there were only two automated measurements at this site, no automated reference is available for intercomparison.

Figure 3.20. SWE measurement at the Weissfluhjoch site using a snow board, sampling cylinder, and scale.

3.1.4.3.8 *Hala Gasienicowa*

The snow-depth reference at Hala Gasienicowa comprised two snow stakes installed within the measurement compound, but outside the FOV of the snow-depth sensors. (See site layout in Annex 4.) The stakes were observed daily.

3.1.4.3.9 *Challenges impacting the reference method for snow depth*

There are challenges faced when making reference measurements for snow depth at sites for the SPICE intercomparison. Specific challenges faced at the CARE and Sodankylä sites are outlined below.

At CARE, the manual snow stake measurements are made at each corner of the artificial target, the middle of which is being measured by an automatic snow depth sensor. Following sections in this report show the intercomparison between the manual reference measurements (a mean of the four snow stakes on each target) and the corresponding measurement from the automatic sensor. The intercomparison shows that the manual measurement is systematically lower than the automatic measurement. This could be caused by two mechanisms. First, due to the installation of the snow stakes at the corner of the targets (Figure 3.21), snow is trapped in the middle of the target, resulting in mounding and higher snow depths in the middle of the target as compared to the corners. Second, frost heave of the target can decrease the relative height of the target as compared to the sensor. This is a known problem at CARE, and it is illustrated in Section 4.2.6.3 by the zero-snow-depth drift analysis. Figure 3.21 shows the result of frost heave on a target that was originally installed flush with the ground surface. Because the snow stakes can float with the target, the snow-depth measurement will always be made relative to the target surface and will be equally impacted by frost heave. This increases the relative difference between the snow stake measurement and the sensor measurement.

Figure 3.21. Photo of the artificial snow-depth target at CARE and the manually observed snow stake on one corner. The snow targets were originally installed flush with the ground, but frost heave eventually pushes the target out of the ground and closer to the sensor.

At Sodankylä, the snow stake measurements are taken via web camera photos as described earlier. There are several challenges associated with the methodology. The first is obtaining photographs during the arctic winter, when daylight is quite limited. Without artificial lighting, this often limits photographic measurements to one per day. This challenge, of course, could potentially be eliminated with the lighting of the snow stakes, enabling photographs to be taken each hour.

A more significant challenge is the actual extraction of the snow depth from the stake by interpreting the depth as seen in the photo (Figure 3.18). The interpreted snow depth is the average snow depth around the stake. During snowfall, snow often mounds around the stake, making it difficult for a user to visually see the stake depth, which is level with the surface of the surrounding snow. After snowfall, the snow stake often alters the radiation budget around the stake causing snow to melt or settle, resulting in a well. The challenge for the user is to interpret what the snow depth would be in the absence of this well. In either case, an on-site observer would examine the stake from a viewpoint level with the surrounding snow and visually interpret the depth as marked on the graduated stake. This is, of course, much more difficult to do when the observation is made via a photograph, which is usually taken at an angle from well above the level of the snow. Given the

quality of the web cam photos from Sodankylä and the observed issues with mounding and welling around the stake, it is estimated that the snow depth can be extracted with an uncertainty of +/- 1 cm.

Based on the experience at Sodankylä, the following should be considered when using web cameras as an observation method for snow stakes:

- Artificial light sources are recommended for good quality photos 24 hours a day, especially at high latitudes where there is very little sunlight available in the winter.
- The photographs should be interpreted or at least checked daily (preferably), so that problems related to mounds of snow around the stakes or camera malfunctions are noticed and addressed with minimal delay.
- Maintenance personnel are required to level the snow around the snow stakes regularly.
- Positioning of the camera and the stakes should be considered carefully in order to maximize the accuracy of snow-depth estimations from the photos.

3.2 Reference traceability

As noted in Section 3.1, a DFIR-fence with an automatic gauge in the center has been selected as the SPICE R2 reference configuration. Since using the DFIR with an automatic gauge as a field reference is relatively new, one goal of SPICE has been to characterize this configuration.

Systematic errors in solid precipitation measurements have been evaluated previously, including intercomparisons organized by WMO/CIMO, such as the Solid Precipitation Measurement Intercomparison study during 1986 to 1993 (WMO/TD - No. 872, 1998). The selection of an appropriate field reference configuration is critical to assessments in which the true value of the parameter being measured, in this case precipitation amount, is unknown.

This section documents the relationships among the different levels of field references, as defined in Section 3.1, and configured on the sites participating in SPICE.

3.2.1 Assessment of DFIR vs. bush gauge (R0 vs. R1)

Authors: Daqing Yang, Craig Smith

3.2.1.1 Background

The DFIR has a long documented history, in particular from experiments conducted at the Valdai hydrologic research station in Russia, where the DFIR was tested against the bush gauge from 1970 onwards. Through the WMO-SPICE collaboration, long-term intercomparison data from 1991 to 2010 at Valdai have been made available. As outlined in Section 0, the bush gauge configuration at Valdai is a three-hectare area with shrubs surrounding two shielded Tretyakov gauges. The shrubs were maintained at the gauge height of 2 m. The bush gauge is comparable to a pit gauge, a gauge in a pit with its orifice at the ground level to reduce wind effect (WMO, 1991).

Results published by Golubev (1989) and Yang et al. (1993) using data from the Valdai station collected from 1970 to 1990 concluded that that the bush gauge measurements were systematically higher than those of the DFIR. On average, the bush gauge caught 6%, 8%, and 10% more than the DFIR for rain, mixed precipitation, and snow, respectively (Yang et al., 1993). Golubev (1989) developed an equation for adjusting the DFIR measurements using wind speed, atmospheric pressure, mean air temperature, and humidity. Further analysis showed that the effects of atmospheric pressure and humidity were negligible, and the equation could be simplified further to include only the air temperature and wind speed (Metcalfe and Goodison, 1992). Yang et al. (1993)

derived a strong linear relationship between the two gauges, except during blowing snow events. The intercomparison organized by WMO in 1989-1993 confirmed the earlier results and recommended adjustments to be applied to DIFR measurements to obtain more accurate snowfall amounts (WMO/TD - No. 872, 1998).

In the context of SPICE, and to further examine the relationship between a bush gauge configuration (SPICE R0 reference) and a DFAR (SPICE R2 reference), the Caribou Creek site organized an intercomparison between two single-Alter-shielded automatic gauges (a Geonor T-200B3 and and OTT Pluvio^2) installed inside an area of young Jack Pine trees, trimmed approximately to gauge height, and an R2 reference with a Geonor gauge installed in a clearing. The distance between the two configurations was about 125 m. Even though the site was not windy, this configuration allowed the intercomparison between the SPICE R2 reference and what is deemed to be an automated configuration of a bush gauge, denoted as R0A.

3.2.1.2 Data and methods

For the analysis of results from Valdai, the precipitation type was classified as dry snow, wet snow, mixed precipitation, or rain by site observers at the times of observations (i.e. by examining the content in the gauge bucket). Drifting or blowing-snow events were also observed and reported. The following summary of results focuses on snow and mixed precipitation data, including wet, dry and blowing snow. Specific data analyses include calculations of total snow and mixed precipitation amounts over the study period and determination of the mean catch ratios (bush gauge/DFIR and Tretyakov gauge/DFIR), mean air temperature, and wind speed for all days with snow, blowing snow, and mixed precipitation. The statistical tools used, such as the regression and correlation analyses of gauge catch ratios as a function of wind speed, were recommended and tested in the previous WMO intercomparison (WMO/TD - No. 872, 1998).

The Caribou Creek SPICE intercomparison was completed for the 2012/13 and 2013/14 winter seasons. The data from the precipitation gauges used in this analysis, the bush-shielded SA Geonor (R0G) and Pluvio^2 (R0P) and the Geonor in the DFIR-fence (R2G), were collected minutely. The data were then subjected to a manual and automated quality control process, and used to produce both a 30-minute and 60-minute Site Event Data Set (SEDS; see Section 3.4). To produce a consistent time series for intercomparison, the unfiltered bucket weight data were smoothed using a Savitzy-Golay filter followed by a noise balancing technique called the “brute force filter” (Pan et al., 2016), which results in a clean time series of accumulated precipitation for each gauge, for both winter seasons. Although the SEDS methodology is universal for much of the SPICE analysis, the second technique noted here is only used for developing comparative time series at Caribou Creek.

3.2.1.3 Results

3.2.1.3.1 *Assessment of the bush gauge vs. DFIR at Valdai (1991-2010)*

For the period from 1991 through 2010, data were collected from Tretyakov gauges installed in the bush gauge, a DFIR, and in a standard configuration (Tretyakov collector and Tretyakov shield) in an open area of the Valdai site. Over this period, 1486 observations were recorded and were classified as dry snow, wet snow, mixed precipitation, and blowing snow. Statistical analyses of the data show that the mean temperatures were -6 °C for dry snow, -3 °C for wet snow, -5 °C for blowing snow, and 1 °C for mixed precipitation. The average wind speeds at 3 m height were about 3.8-3.9 m/s for wet and dry snow, and 5.7 m/s for blowing snow. The bush gauge measurements were generally higher than those of the DFIR for all precipitation types. On average, the bush gauge caught 5%-6% more

than the DFIR for snow and mixed precipitation, and 12% more for blowing snow, respectively. The difference in mean catch ratios between snow and blowing-snow events suggests potential blowing-snow impacts on gauge observations at this site.

Figure 3.22 presents a scatter plot of the dry and wet snow catch ratio (bush gauge/DFIR) for mean 3 m wind speeds up to 8 m/s. The catch ratios are within about 90% to 120% for lower winds (below 3 m/s), and generally increase to 90% to 150% for higher wind speeds (6 to 7 m/s), suggesting that scatter in the relationship increases at higher wind speeds. The relationship shown by the regression lines in Figure 3.22are statistically significant at the 95% confidence limit, and indicate that the DFIR catch is very close to true snowfall for lower mean wind speeds, and measures, on average, about 93% of "true" snowfall for wind speeds up to 6 to 7 m/s. The results for wet snow are very similar to those for dry snow. This is an important point because it supports the use of the DFIR as a reference, given its consistency relative to the bush gauge. For blowing snow (not shown), which occurred only when the 12-hour average wind speeds were greater than 3 m/s, the catch ratios were generally quite similar to other snow types, but showed more outliers, with catch ratios as high as 180% due to snow blowing into the bush gauge. As recommended by Yang et al. (1993) and WMO/TD - No. 872 (1998), it is not practical to correct the DFIR data for blowing-snow cases because the bush gauge is not reliable for the measurement of true snowfall in these conditions.

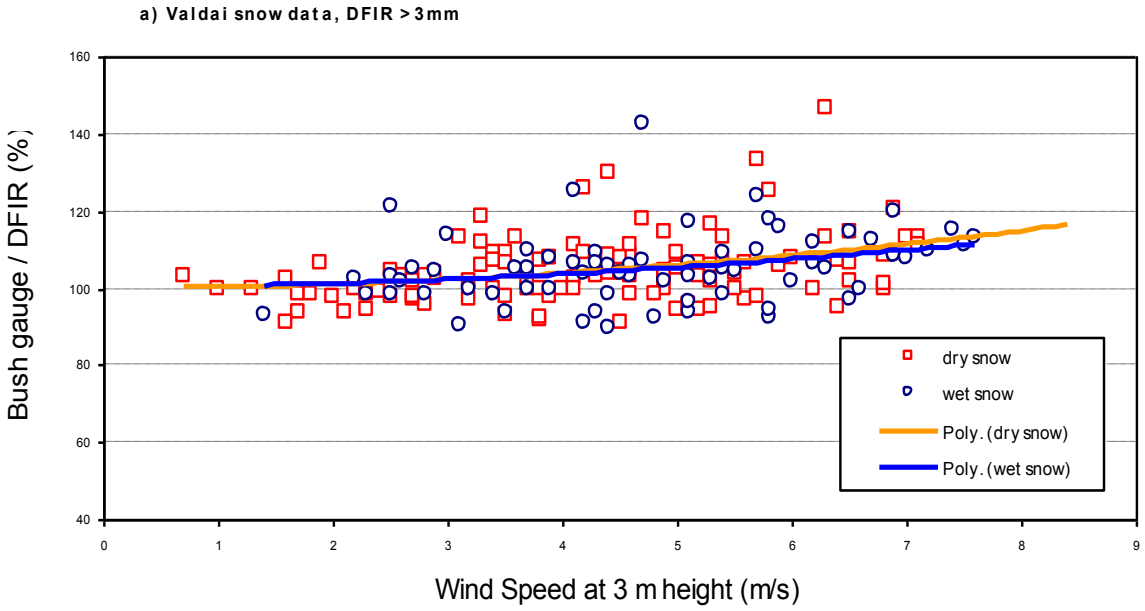


Figure 3.22. Scatter plot of the catch ratio for dry snow (red markers) and wet snow (blue markers) as a function of mean wind-speed measured at 3 m above the ground, Valdai site.

The regression equations derived from the analysis based on the 1971-1990 data are shown in Yang et al. (1993). More data were collected at the Valdai site from 1991-2010, and a similar analysis was performed by Yang (2014). Figure 3.23 compares the curves of bush gauge/DFIR catch ratios as a function of mean wind speed for snow events at Valdai for both the 1971-1990 and the 1991-2010

(shown as "this study") datasets. There are differences and similarities between the two studies. For dry snow, the bush gauge/DFIR ratios from 2014 are systematically lower than those from 1993; the differences are 2 to 4% for wind ranges of 2 to 8 m/s (Figure 3.23a). For wet snow, the ratio differences vary from 3 to 8% for wind speeds of 4 to 8 m/s (Figure 3.23b).

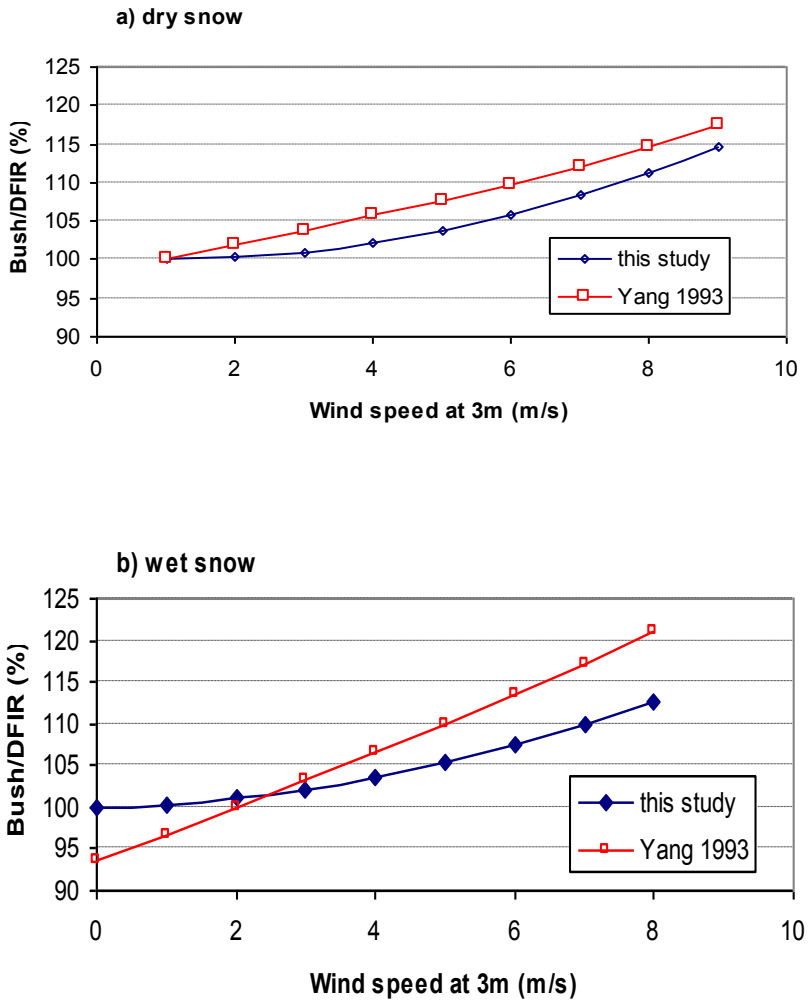


Figure 3.23. Comparison of the regression equations of bush gauge/DFIR catch ratios as a function of mean wind-speed measured at 3 m above ground, for (a) dry snow and (b) wet snow using Valdai site data from 1971-1990 (red) and from 1990-2010 (blue). The differences between the two datasets are 2 to 4% for wind speeds of 2 to 8 m/s for dry snow, and 3 to 8% for wind speeds of 4 to 8 m/s for wet snow.

3.2.1.3.2 *Automatic bush gauge vs. DFAR at Caribou Creek*

Figure 3.24 shows the cumulative precipitation of various automatic gauges tested at Caribou Creek for the 2013/14 and 2014/15 winter seasons. Total snow accumulation was about 120-150 mm and 50-70 mm, respectively, for the two winters. Comparing the seasonal totals for all configurations, the catch for both the Geonor T-200B3 and OTT Pluvio^2 gauges in the bush are very similar (less than 1% difference), but both bush gauges caught less precipitation than the Geonor in the DFIR-fence

(DFAR), even though wind speeds during precipitation events were generally higher in the clearing near the DFIR fence (Figure 3.25). The DFAR exceeded the average catch of the bush gauges by 6.8% and 7.8% for the two respective seasons, in contrast with the results from Valdai. Note that the DFAR configuration is denoted as 'DFIR' in all plots in this section, in reference to the DFIR-fence configuration.

Table 3.4. Seasonal accumulated total precipitation for automatic gauge configurations at the Caribou Creek SPICE site, 2013/14 and 2014/15.

Season	Gauge Configuration	Accumulated Total Precipitation [mm]
2013/14	Geonor-Bush	156.9
	Pluvio^2-Bush	158.3
	Geonor-DFAR	168.3
2014/15	Geonor-Bush	67.3
	Pluvio^2-Bush	68.2
	Geonor-DFAR	73.1

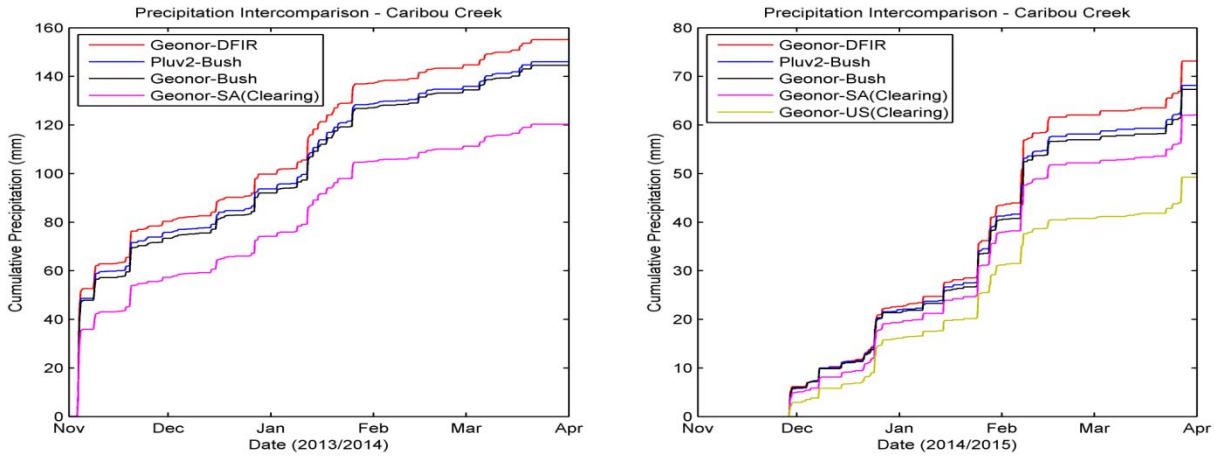


Figure 3.24. Cumulative precipitation reported by automatic gauges tested at Caribou Creek over the 2013/14 (left) and 2014/15 (right) winter seasons.

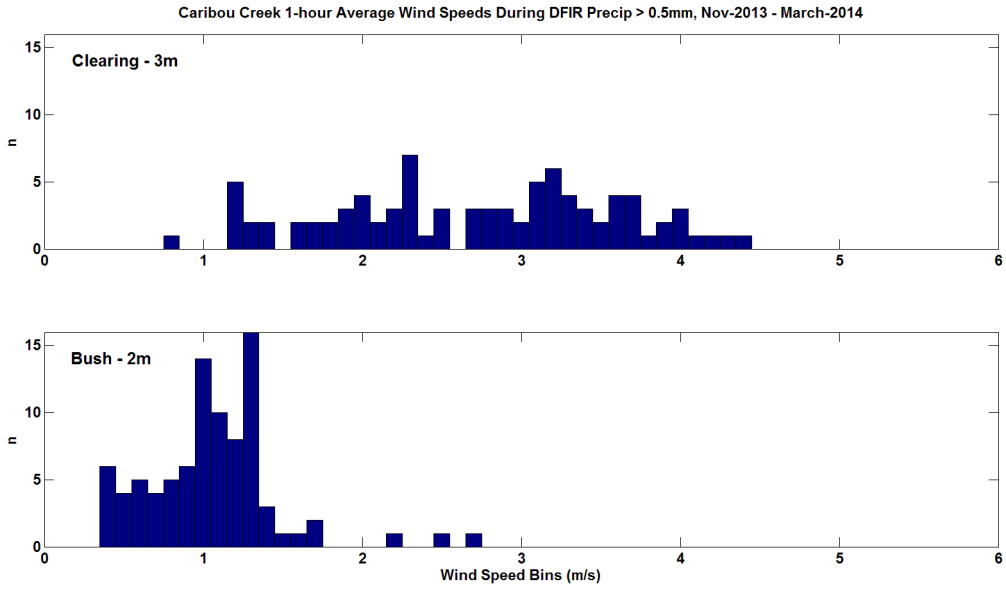


Figure 3.25. Wind-speed histogram comparing the average wind speed at 3 m in the clearing to the average wind speed at 2 m in the bush during winter 2013/14 precipitation events, Caribou Creek.

Figure 3.26 shows scatter plots of the bush gauge vs. DFAR 30-minute SEDS data for the two winter seasons. For the 2013/14 winter, the snowfall events ranged from 0.25 mm to 3 mm, and on average, the DFAR reported 11% more snowfall than the bush gauges. There is a close correlation between the two gauges with R^2 = 0.85. The results for the 2014/15 winter are similar, with the event snowfall ranging from 0.25 mm to 2 mm, the mean catch of the DFAR exceeding that of the bush gauge by 7%, and with a close correlation (R^2 = 0.73) between the two gauges.

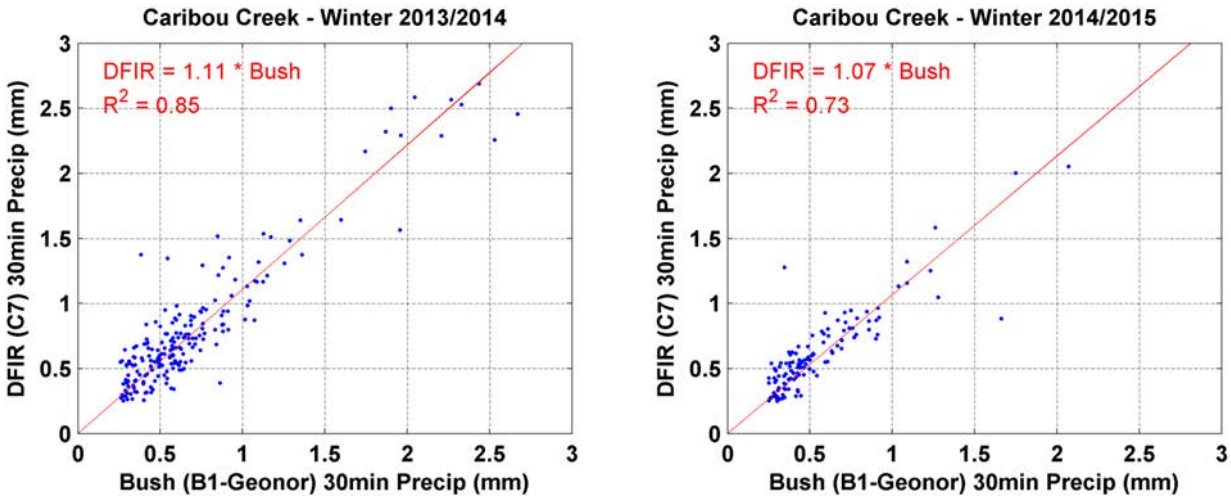


Figure 3.26. Scatter plots of the bush gauge vs. DFAR, using 30-minute SEDS data, winter 2013/14 (left) and winter 2014/15 (right), Caribou Creek.

Generally, because of the cold continental climate of this region, snowfall rates and liquid-water equivalents are lower at this site than at most of the other SPICE sites. For this reason, it is useful to compare the results for longer accumulation periods. For this purpose, a 60-minute SEDS was produced for Caribou Creek. The longer accumulation interval allows for a longer period for precipitation to accumulate and, in theory, increases the signal-to-noise ratio of the precipitation data. Figure 3.27 presents the scatter plots of the bush gauge vs. DFAR 60-minute SEDS data for the two winter seasons. The 60-minute snowfall event accumulations were larger than those for the 30-minute events, with accumulations that ranged from 0.25 mm to 5 mm for the 2013/14 winter. For most cases, the DFAR measured more snowfall than the bush gauge and, on average, reported 9% more snowfall than the bush gauge. The results for the 2014/15 winter were similar; event accumulations ranged from 0.25 mm to 3 mm, and the mean catch of the DFAR exceeded that of the bush gauge by 7%. Correlations between the two gauges were higher for the 60-minute accumulations (R^2=0.97 and R^2=0.87 for the two respective seasons) than for the 30-minute accumulations. The linear relationship, statistically significant at 90-95% confidence, may be considered as a transfer function between these gauges for this location and other regions with similar climatic conditions.

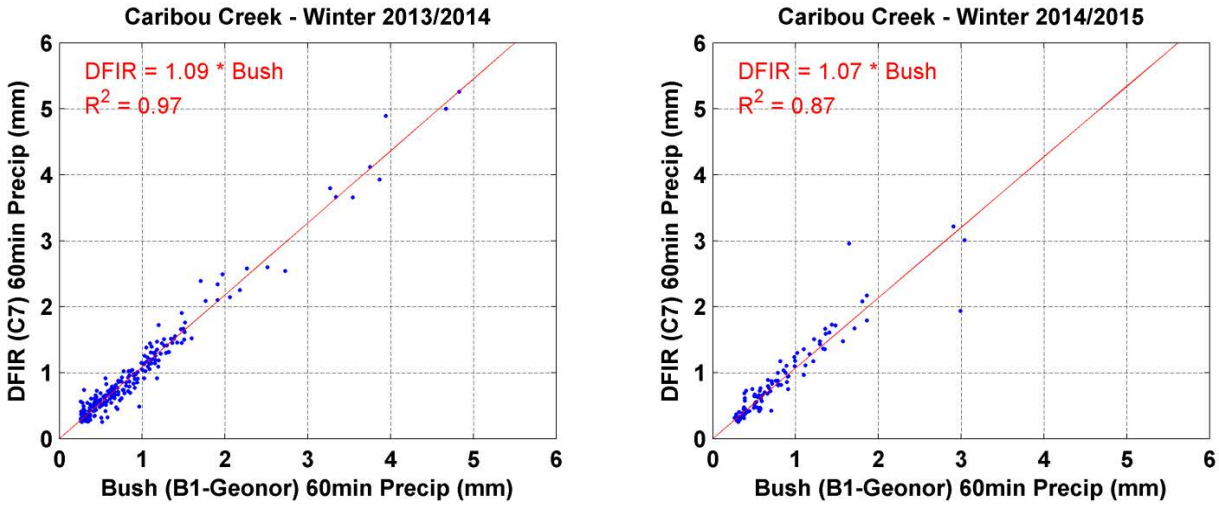


Figure 3.27. Scatter plots of the bush gauge vs. DFAR, using 60-minute SEDS data, winter 2013/14 (left) and winter 2014/15 (right), Caribou Creek.

Figure 3.28 displays the catch ratio (DFAR/bush gauge) as a function of mean wind speed at 3 m height for both the 30-minute SEDS (top, for gauge precip > 0.25 mm) and the 60-minute SEDS (bottom, for gauge precip > 0.5 mm). The catch ratios of 30-minute SEDS data vary between 0.5 to 2.5 for mean wind speeds up to 5 m/s. The scatter is higher for data collected in winter 2013/14, and it seems to increase slightly with the wind speed. The catch ratios for the 60-minute SEDS data vary from 0.7 to 1.8. As expected, this range of catch variation is much smaller relative to the 30-minute data because of reduced noise for the longer accumulation periods. The results are generally similar between the two winters, with some outliers in both. There might be a very slight tendency of increasing DFAR catch with wind speed, particularly for the 2014/15 winter, but generally the catch efficiency does not change appreciably with mean wind speed. Again, the results appear to be

contradictory to those of the Valdai intercomparison, where the bush gauge measured more than the DFIR by 3 to 8% (for wind speed between 4 and 8 m/s).

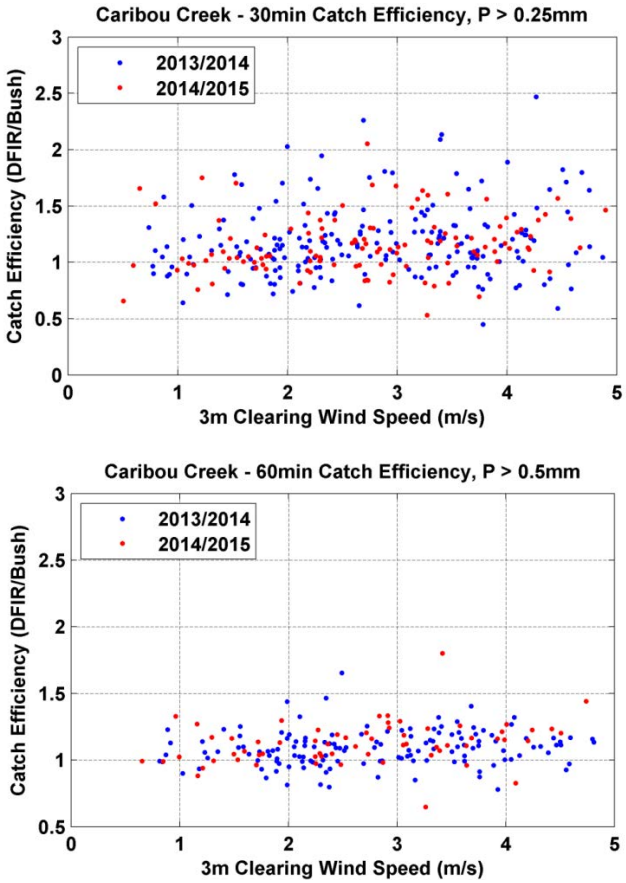


Figure 3.28. The scatter plots of catch ratio (DFAR/bush gauge) as a function of wind speed at 3 m height, for 30-minute and 60-minute SEDS data, Caribou Creek.

3.2.1.3.3 *Discussion*

Uncertainties exist in data collection and analyses for precipitation gauge intercomparison experiments. For Valdai, the observers did the classification of precipitation types at the time of the observations. Some misclassifications are likely, particularly for mixed-precipitation and blowing-snow events. Air temperature is useful to check or estimate precipitation type. At Valdai, wind speed and air temperature are the 12-hour mean values and do not represent well the weather conditions during precipitating periods. The use of 12-hour mean wind speed is one of the reasons for the higher variability in catch ratios. Data collection and analyses for shorter timescales, such as every hour or every six hours, are expected to produce better results, since wind speeds vary throughout the day. Automatic sensors (such as an optical precipitation detector or other precipitation-type sensor) could provide a better indication of the precipitation type(s) during a given assessment period.

Many blowing-snow events were recorded at Valdai at the time of observations; however, no additional information was reported in terms of blowing-snow duration and intensity, which are critical to quantify blowing-snow flux. As recommended by the previous WMO intercomparison

(WMO/TD - No. 872, 1998), the identification and separation of blowing-snow events are necessary because they represent a special circumstance when adjusting gauge data. During blowing-snow events, the bush gauge (at 2 m) caught, on average, 12% more snow than the DFIR (at 3 m height), while the average ratio of bush gauge to DFIR is only 105-106% for snow conditions. Because of the uncertainty in gauge performance in high-wind conditions, adjustments of the DFIR data when blowing snow is reported are not recommended (WMO/TD - No. 872, 1998). Since wind speeds are generally greater during blowing-snow events, the adjustment applied for undercatch during these events could be higher than warranted. This problem becomes most severe for gauges mounted close to the ground and susceptible to catching more blowing snow.

The results from the Caribou Creek intercomparison were not expected given the results from Valdai. The wind-speed histograms in Figure 3.25 show that the 3 m wind speeds in the clearing with the Geonor-DFIR are higher than wind speeds at 2 m inside the bush area with the Geonor-Bush and Pluvio^2-Bush gauges. Overall, wind speeds at the site are relatively low, below 4 m/s at gauge height during precipitation events. It is conceivable that undercatch of the bush gauges may occur at wind speeds less than 2 m/s in the bush area, while the DFIR-shield is more effective at reducing undercatch than the bush shielding at the wind-speeds experienced in the clearing.

3.2.1.3.4 *Summary*

The bush gauge at Valdai systematically catches more snow and mixed precipitation than the DFIR, which is attributed to the influence of wind speed during precipitation events. For instance, the bush gauge measures 20%-50% more snow over a 12-hour period than the DFIR for wind speeds of 6 m/s-7 m/s. Therefore, the adjustment of the DFIR for wind-induced loss is necessary to more accurately represent “true” precipitation. It is important to point out that this error changes with wind speed and precipitation type. In comparison to previous analyses (Yang et al., 1993; WMO/TD - No. 872, 1998), the more recent analysis produces similar but more reasonable results, suggesting lower snow undercatch by the DFIR relative to the bush gauge by 3%-6%. This means that the DFIR performance is better than previously documented in the past WMO intercomparison (WMO/TD - No. 872, 1998) and will influence the evaluation of national precipitation gauges against the DFIR. More effort is needed to quantify the impact through field data collection and additional data analyses at selected WMO test sites. See for example the work of Theriault et al. (2015).

Since the intercomparison data between the bush and DFIR gauges came from only the Valdai station in Russia, it is important to compare bush-shielded gauges to DFIR-shielded gauges elsewhere and using automated instrumentation. The results from the Caribou Creek site suggest that the Geonor in the DFIR-fence measured, on average, 7%-11% more snow than the bush gauge and that the catch ratio (DFAR/bush gauge) did not change appreciably with wind speed up to 5 m/s. These results are very different from those reported at Valdai. The differences in results between Valdai and Caribou Creek may be due to: a) a difference in the bush growth and structure between the sites and b) a difference in wind regimes and blowing-snow impact on gauge observations. It is important to remember that at Valdai all gauges were manual, whereas all gauges used for the Caribou Creek experiments were automatic. While there is no direct intercomparison between manual and automated bush-shielded gauges, the higher catch of the Geonor inside the DFIR-fence relative to the automated bush gauges supports the use of the DFAR as an automated reference under these conditions. However, further analysis is recommended to compare the automatic gauge data from both sites. Tests at the Caribou Creek site are continuing beyond the period used for the results

outlined in this report to better assess both reference configurations for future intercomparison studies.

3.2.2 Assessment of DFAR vs. DFIR (R2 vs. R1)

Authors: Kai Wong, Rodica Nitu, Craig Smith

3.2.2.1 Overview

This section outlines the results of the characterization of the DFAR R2 field working reference system relative to the DFIR R1 secondary field reference system, as defined in Section 0. This assessment is based on historical results from previous intercomparisons and on results derived from new data collected at the CARE (Canada) SPICE site.

3.2.2.2 Historical perspective

3.2.2.2.1 *R1 vs R2 at Jokioinen, Finland*

An early representation of a field reference system using an automatic gauge installed in a DFIR-fence (R2-type configuration) was used on the Jokioinen site in Finland during the 1986-1993 WMO Intercomparison. At the time, Jokioinen hosted a manually observed DFIR (which served as the site reference) and a second DFIR-fence with a Geonor T-200B3 gauge, as presented in Annex 3.D of Goodison et al. (1998). The data were made available after the completion of the intercomparison and not included in the final report.

Manual observations of the DFIR were made at the Jokioinen site by the Finnish Meteorological Institute from December 1988 through April 1993, twice daily at approximately 10:00 and 22:00 UTC. Present weather observations were also made at the time of observations to identify precipitation type. Air temperature and 2 m wind speed and direction were recorded and averaged over the period corresponding with each manual observation. Minimum and maximum temperatures were also reported over the same period. Figure 3.29 shows the scatter plots of the manual observations from the DFIR and the accumulated precipitation in the Geonor gauge in the DFIR-fence, for those periods when both gauges measured at least 1 mm and the maximum temperature during the period did not exceed -2 °C (threshold intended to limit the assessment to solid precipitation periods).

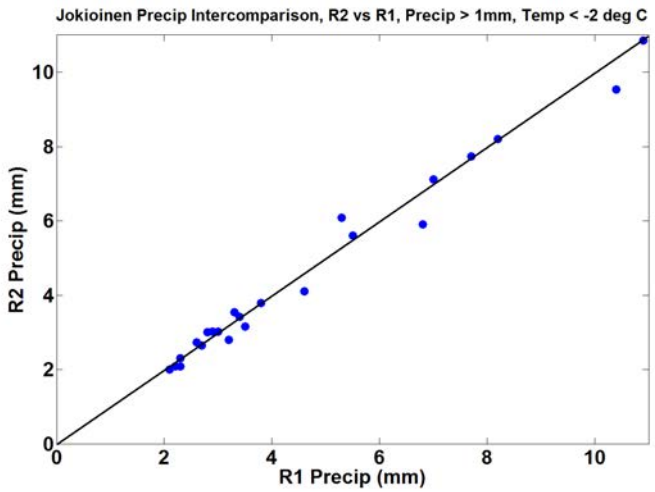


Figure 3.29. Comparison between accumulation reports from the Geonor T-200B3 gauge in DFIR-fence and the DFIR, with a 1:1 relationship overlaid. Datapoints represent daily measurements of snowfall greater than 1 mm during periods when the temperature did not exceed -2 °C.

The total reported precipitation amounts from the manual DFIR and Geonor in DFIR-fence for the 23 events shown in Figure 3.30 were 106.5 mm and 104.7 mm, respectively, with an average catch efficiency of 0.99. Overall, these results indicate that for wind speeds below 6 m/s, the specific gauge type (manual or automated (Geonor)) has minimal impact on the amount of snowfall collected.

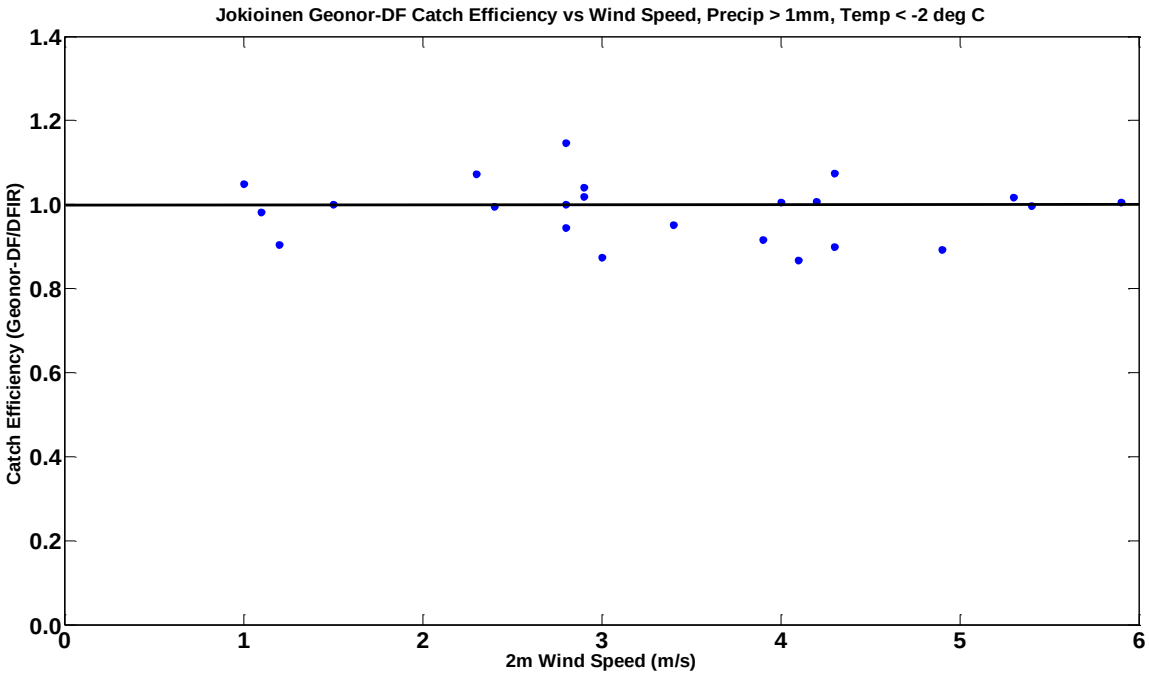


Figure 3.30. Catch efficiency – wind-speed relationship for the Jokioinen double-fence Geonor (DFAR) compared to the DFIR for daily precipitation amounts greater than 1 mm where temperatures did not exceed -2 °C. The DFIR is adjusted for wetting loss but not wind bias.

3.2.2.2.2 *R1 vs. R2 at Bratt's Lake, Canada*

In November 2003, Environment Canada began gauge/shield configuration intercomparisons at the Bratt's Lake test facility. The site hosted a twice-daily-observed manual DFIR and a second DFIR-shield to house a Geonor T-200B3 gauge (DFAR). Intercomparisons continued with these configurations until 2011. Manual observations, including present weather, were typically made at 1400 and 2300 UTC, and generally only on weekdays. As with the Jokioinen observations, the manual DFIR measurements were made volumetrically with a wetting-loss adjustment applied (Goodison et al., 1998). Automated Geonor observations were accumulated over the manual observation periods, with wind speed and temperature averaged over the same periods. Previously published results from this intercomparison (Smith, 2009, 2010) include a wind adjustment of the DFIR based on Yang et al. (1993) and show decreasing catch efficiency with increasing wind speed. As an example, Figure 3.31 presents a revision of the relationship shown by Smith (2009).

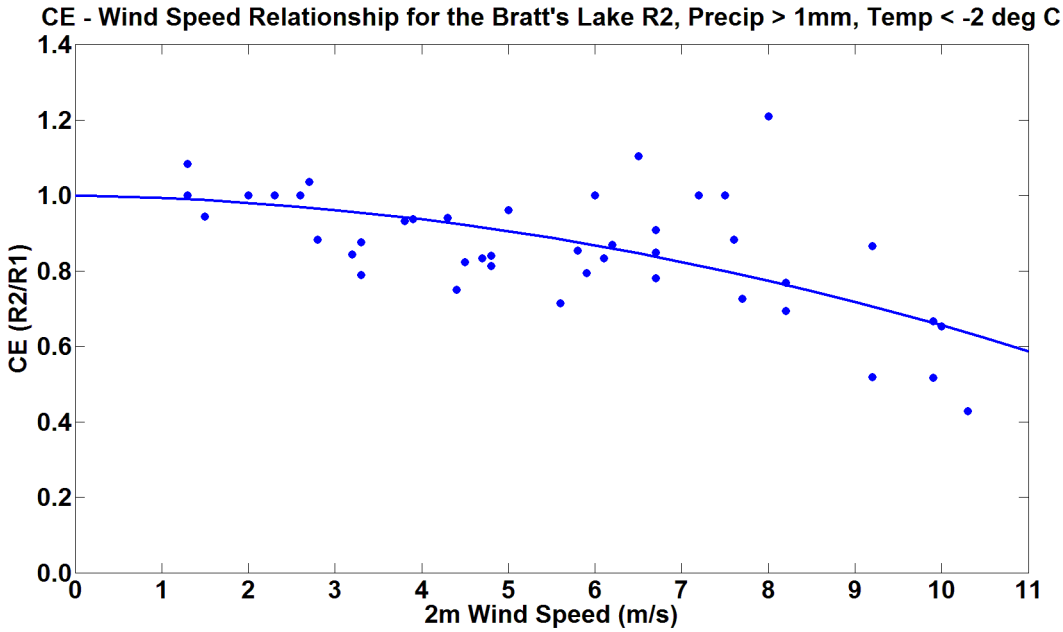


Figure 3.31. Catch efficiency – wind-speed relationship for the Bratt's Lake double-fence Geonor (R2) compared to the DFIR (R1) for daily precipitation amounts greater than 1 mm where temperatures did not exceed -2 °C.

The relationship in Figure 3.31 suggests that the catch efficiency of the Geonor gauge in the DFIR-fence begins to decrease substantially at wind speeds greater than 5-6 m/s; however, this trend is attributed primarily to the wind adjustment of the DFIR (Yang et al, 1993). To be consistent with the prior Jokioinen re-analysis and additional work conducted during SPICE, the Bratt's Lake data were re-examined without the DFIR wind adjustment. In addition, the re-examined dataset was subject to the following changes: more stringent filtering of blowing-snow events, as reported by the human observer, and an experimentally derived wetting-loss adjustment of 0.13 mm per observation, which is necessary because the DFIR observations were made using a volumetric flask.

Figure 3.32 shows the comparison between the double-fence Geonor (Geonor-DF) and the DFIR for the re-examined dataset, with the black line representing the 1:1 relationship. It appears that the catch of the Geonor in the DFIR-fence is lower than the catch of the DFIR. For the 45 events shown, the total catch of the DFIR and Geonor-DF are 117.8 mm and 106.1 mm, respectively, with an average Geonor-DF catch efficiency of 0.93. The catch efficiency-wind-speed relationship for the updated dataset is illustrated in Figure 3.33.

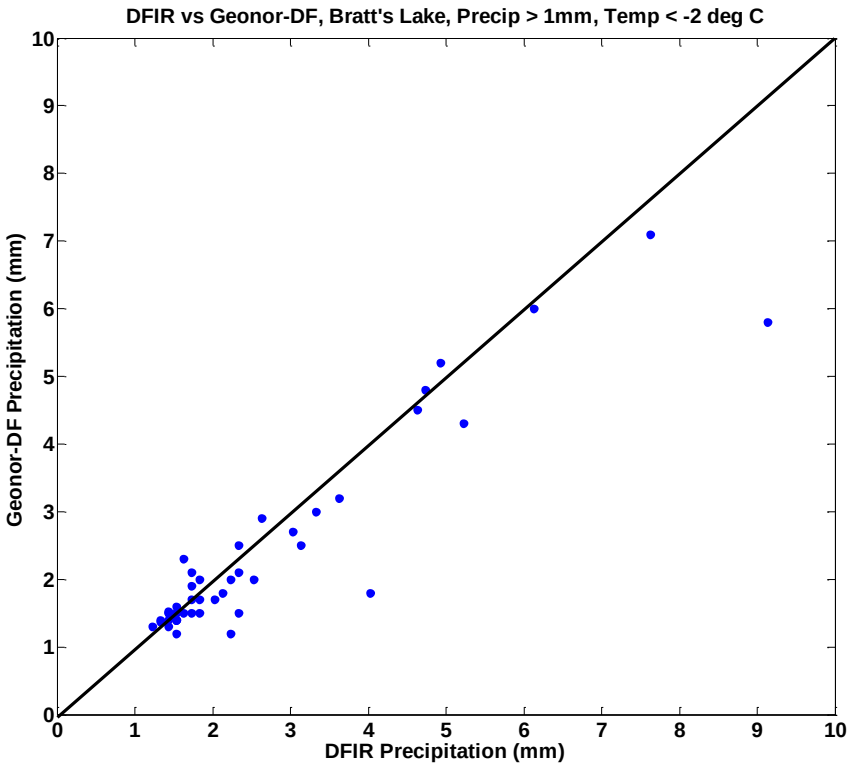


Figure 3.32. Comparison between the Geonor-DF and the DFIR, with a 1:1 relationship overlaid. Data is for daily measurements of snowfall greater than 1 mm during periods when the temperature did not exceed -2 °C.

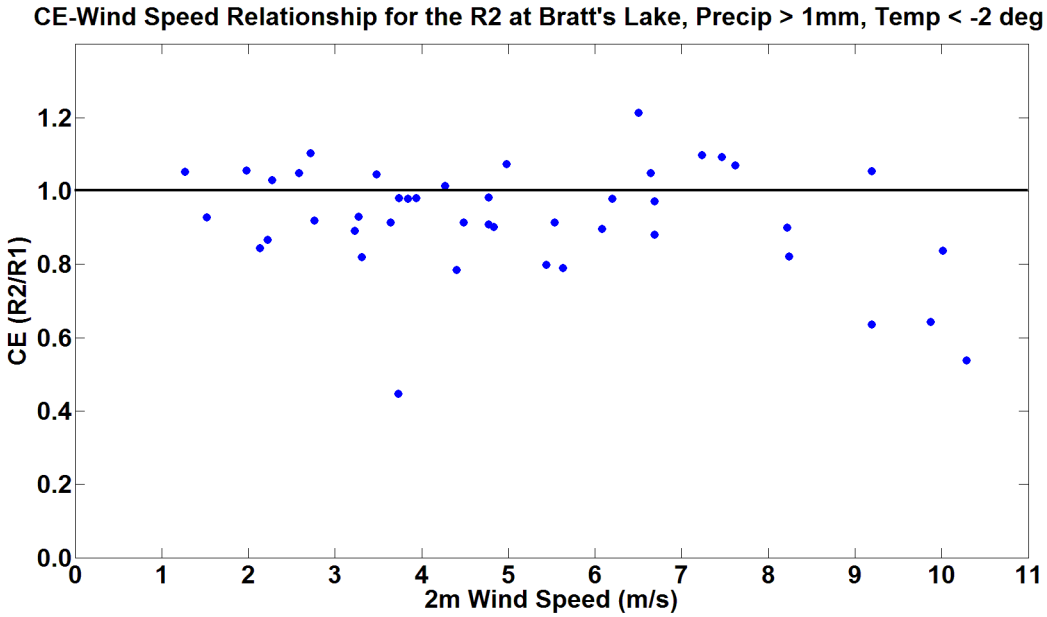


Figure 3.33. Catch efficiency – wind-speed relationship for the Bratt's Lake double-fence Geonor compared to the DFIR for daily precipitation amounts greater than 1 mm where temperatures did not exceed -2 °C. DFIR is adjusted for wetting loss but not wind bias.

3.2.2.3 WMO-SPICE: R1 vs. R2 Intercomparison at CARE

This section focuses on the results from the intercomparison of the R2 and R1 field reference systems at the CARE SPICE site. These results are relevant for providing an assessment on the performance of the R2 reference configurations installed at SPICE sites.

3.2.2.3.1 *Equipment*

The R1 reference system at CARE is described in Section 0, and consists of an octagonal vertical double-fence shield with a manual Tretyakov collector and shield at the centre. The collecting area of the Tretyakov gauge/collector is 200 cm^2. The R2 reference configuration has a similar DFIR-fence and a single-Alter-shielded Geonor T-200B3 automatic gauge at the centre, as shown in Figure 3.34. The Geonor gauge is equipped with three transducers. The gauge is heated following the approach of the US Climate Research Network, as adjusted for SPICE (IOC-2 Final Report, Boulder, 2012 and IOC-4 Final Report, Davos, 2013; It should be noted that the Geonor gauges in the Jokioinen and Bratt's Lake studies, summarized previously, were not heated. The gauge has a collecting area of 200 cm^2 and a capacity of 600 mm.

Figure 3.34. Geonor T-200B3 gauge in R2 reference configuration at the CARE site.

Wind speed was measured at 2 m and 3 m above ground using two Vaisala NWS425 ultrasonic wind sensors. The air temperature was monitored on site using a Yellow Springs International Model 44212 thermistor in a Stevenson Screen. Three present weather sensors, comprising an OTT Parsivel[2] disdrometer, a Vaisala PWD22 present weather detector, and a POSS present weather sensor, were used for the identification of precipitation type.

The CARE site layout is shown in the site commissioning report and in the site description. (See Annex 4). The distance between the R1 and R2 reference configurations is about 36 m.

3.2.2.3.2 *Data*

The data used for analysis in this section were quality controlled using filters to remove any outliers. For the R2 precipitation amounts, a Gaussian filter was used to filter out some of the high-frequency noise. Details of the SPICE QC procedure are presented in Section 3.4.

The observations of the R1 precipitation amount were made in accordance with the manual observation procedures (Earle, 2013). It should be noted that the procedures differ from those used for Jokioinen and Bratt's Lake results reported above, as the weight, instead of the volume of the collected precipitation, was measured using a calibrated scale. The scale was an Ohaus Explorer analytical and precision balance, with calibration error below 0.1 g and readability, repeatability, and linearity of measurements within 0.1 g.

The snow water equivalent in millimetres was computed from the weight of the precipitation collected using the density of water and the dimensions of the collector, as follows:

$$SWE = \frac{M}{\rho A} = \frac{M}{1\ g/cm^3 \times 200\ cm^2} = (0.005\ cm/g) \times M = (0.05\, mm/g) \times M$$

where ρ is the density of water and is assumed to be constant and equal to 1 g/cm^3, A is the collecting area and is assumed to be 200 cm^2, and M is the weight of the precipitation collected in grams.

At the CARE site, the calculated accumulated precipitation amount is reported to the hundredth of one millimetre. Assuming the uncertainty of the scale to be 0.1 g, as one standard deviation, and the collecting area to be exactly 200 cm^2 (ignoring any uncertainty), it can be derived that the laboratory uncertainty of SWE is 0.01 mm.

The R1 observations analyzed in this section of the report were made between December 8, 2012, and April 9, 2013, between December 4, 2013, and April 1, 2014, and between December 2, 2014, and March 11, 2015. The observation periods have a minimum of about six hours and a maximum of 246 hours (10 days). The average period is about 28 hours and the median is 24 hours. Only the R1 observations with periods less than 48 hours are selected for analysis.

3.2.2.3.3 *Analysis technique*

The R2 Geonor gauge amount is taken as the average of the three transducer precipitation amounts, quality controlled using the SPICE QC procedure (Section 3.3.2).The R2 increments were calculated over the R1 observation periods as the difference of the R2 gauge-reported amounts between the end and start of the R1 observation period. There was no additional filtering of R1 precipitation amounts.

The catch efficiency (CE), the ratio of R2 accumulation to R1 accumulation over the same observation period, was then computed and analyzed for different accumulation thresholds. The dependence of CE on wind speed was investigated. Following the approach used by Smith (2008 and 2009) the maximum temperature over the observation period was used to select solid precipitation events, i.e. events for which the maximum temperature over the observation period was < -2 °C were selected as solid precipitation events.

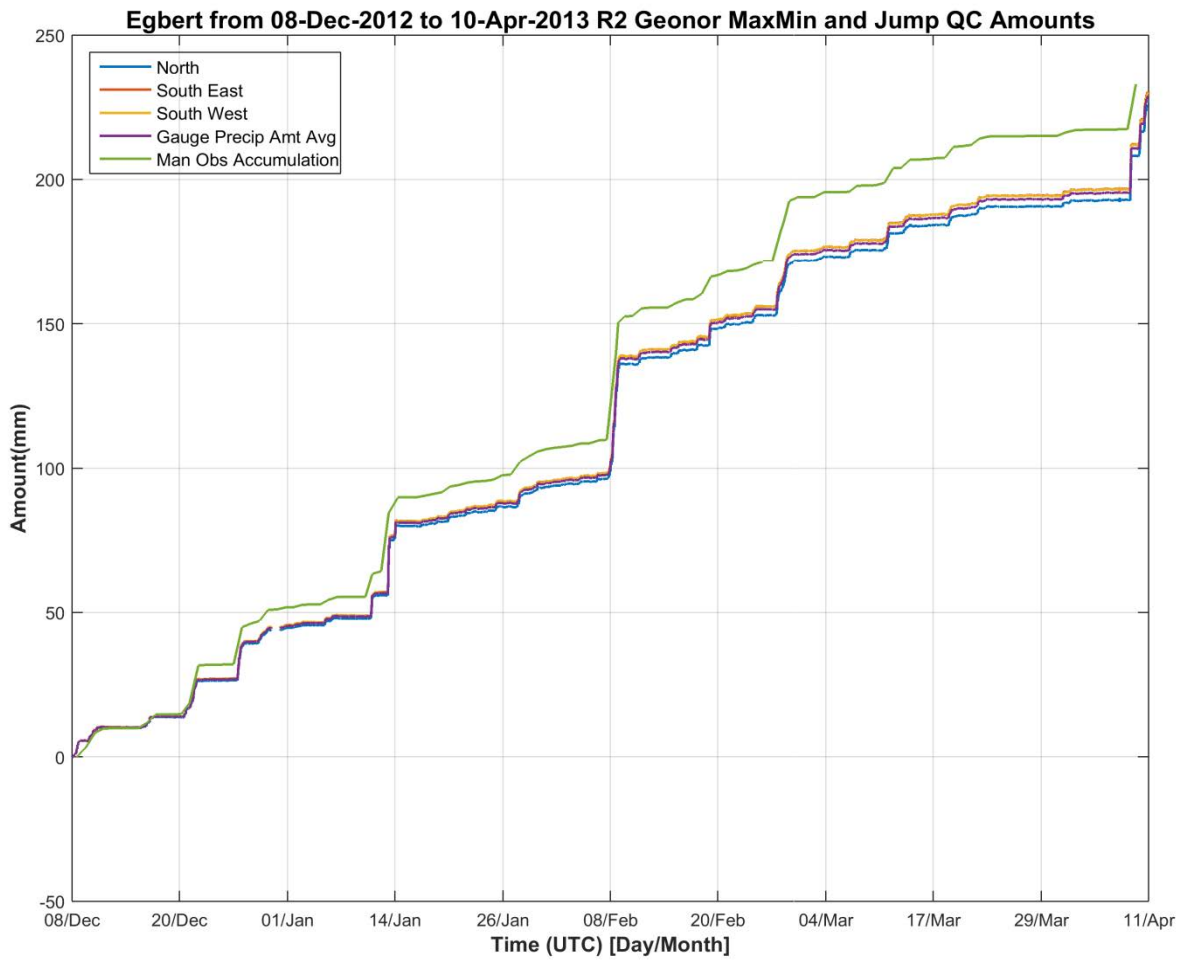


Figure 3.35. 2012/13 winter R2 transducer precipitation amounts filtered with the max-min and jump filter of the SPICE QC procedure, their average, and manual observations. Note: In the plots in this section, Egbert is used interchangeably with CARE, as Egbert (Ontario, Canada) is the closest town to the CARE site.

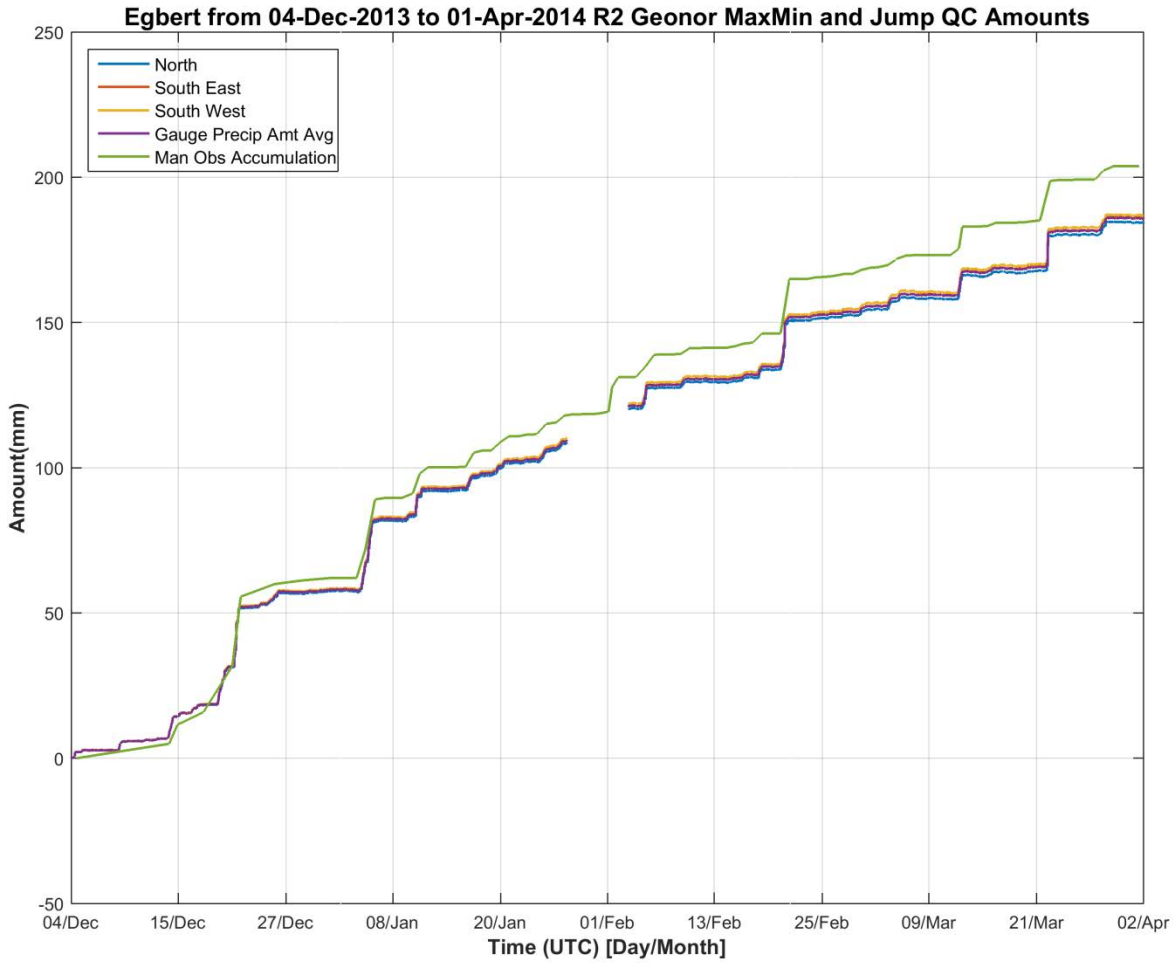


Figure 3.36. 2013/14 winter R2 transducer precipitation amounts filtered with the max-min and jump filter of the SPICE QC procedure, their average, and manual observations.

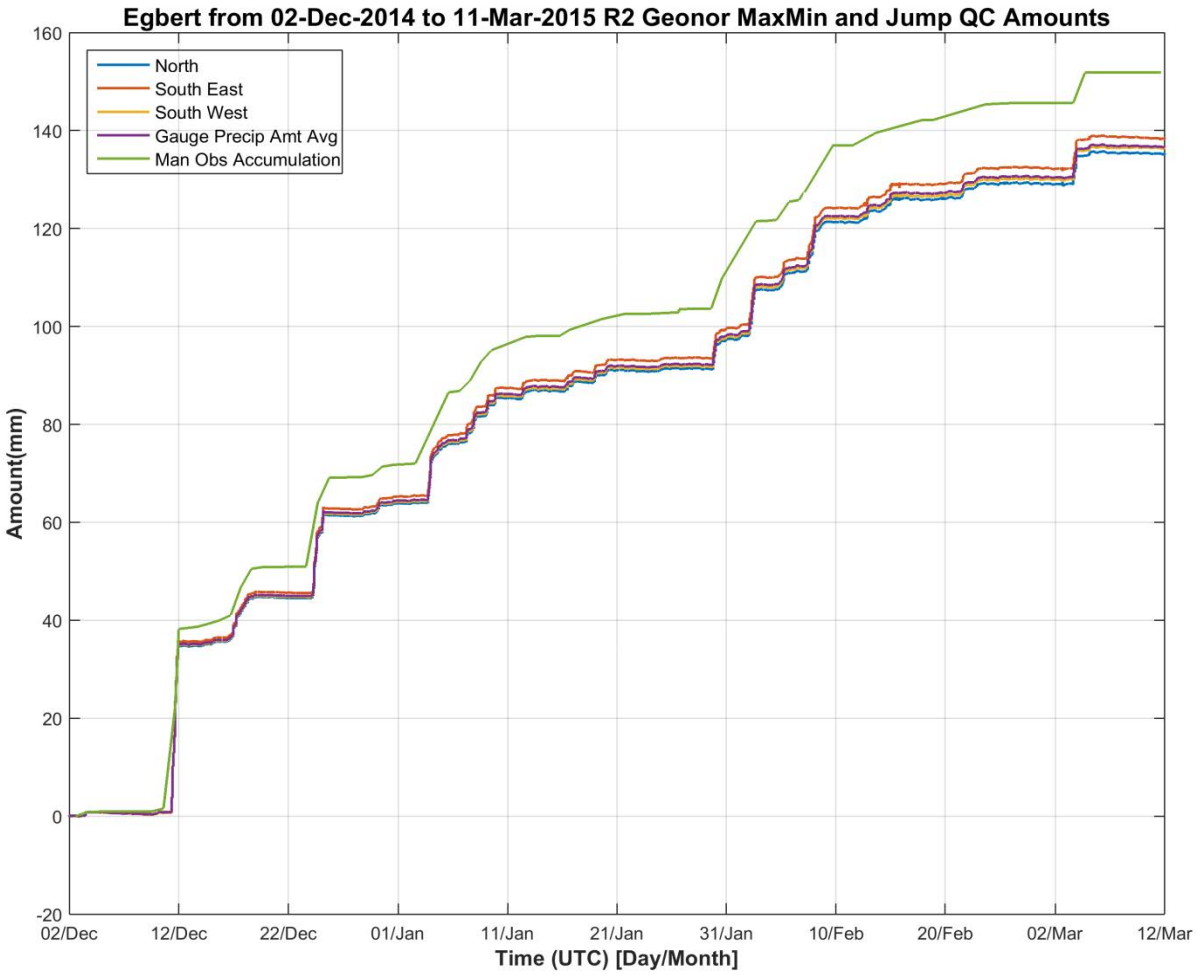


Figure 3.37. 2014/15 winter R2 transducer precipitation amounts filtered with the max-min and jump filter of the SPICE QC procedure, their average, and manual observations.

3.2.2.3.4 *Results*

Figure 3.35,Figure 3.36, andFigure 3.37 show the filtered precipitation amounts from each of the three transducers of the Geonor gauge in the R2 reference configuration, the three-wire average, and the manual observations for the 2012/13, 2013/14, and 2014/2015 winter seasons, respectively. For each season, it is apparent that the R1 configuration reports more accumulated precipitation than the R2 configuration.

The influence of wind speed on the CE was analyzed; Figure 3.38 shows a plot of CE vs. mean wind speed when the accumulation over the observation period exceeded 1 mm for both the R1 and R2 configurations. A 1 mm accumulation threshold was also used in Smith (2009). Figure 3.39 shows a scatter plot of R2 accumulation vs. R1 accumulation for each observation period, or event. There were a total of 33 events for which the accumulation reported by each configuration exceeded the 1 mm threshold: 10 events in 2012/13, 15 events in 2013/14, and 8 events in 2014/15.

Based on the average catch efficiency in Figure 3.38, the reported precipitation amount from the R2 reference configuration is about 7% below the corresponding R1 value. The potential causes of this difference are examined in the following subsections.

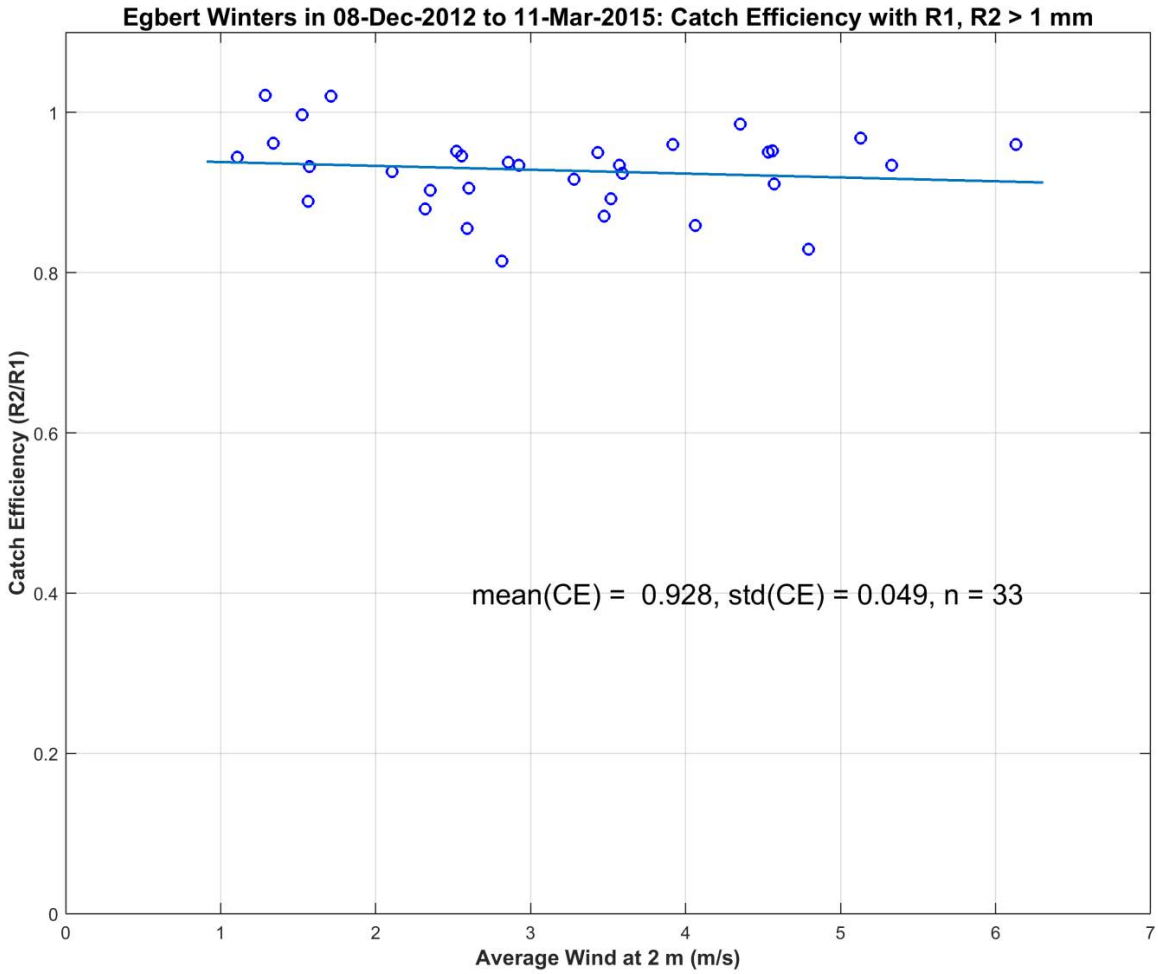


Figure 3.38. Catch efficiency vs. mean wind speed for events with accumulations of 1 mm or greater. The mean and standard deviation of the CE are also included. A linear regression test on the slope, at the 95% confidence level, showed that the slope is significantly different from zero.

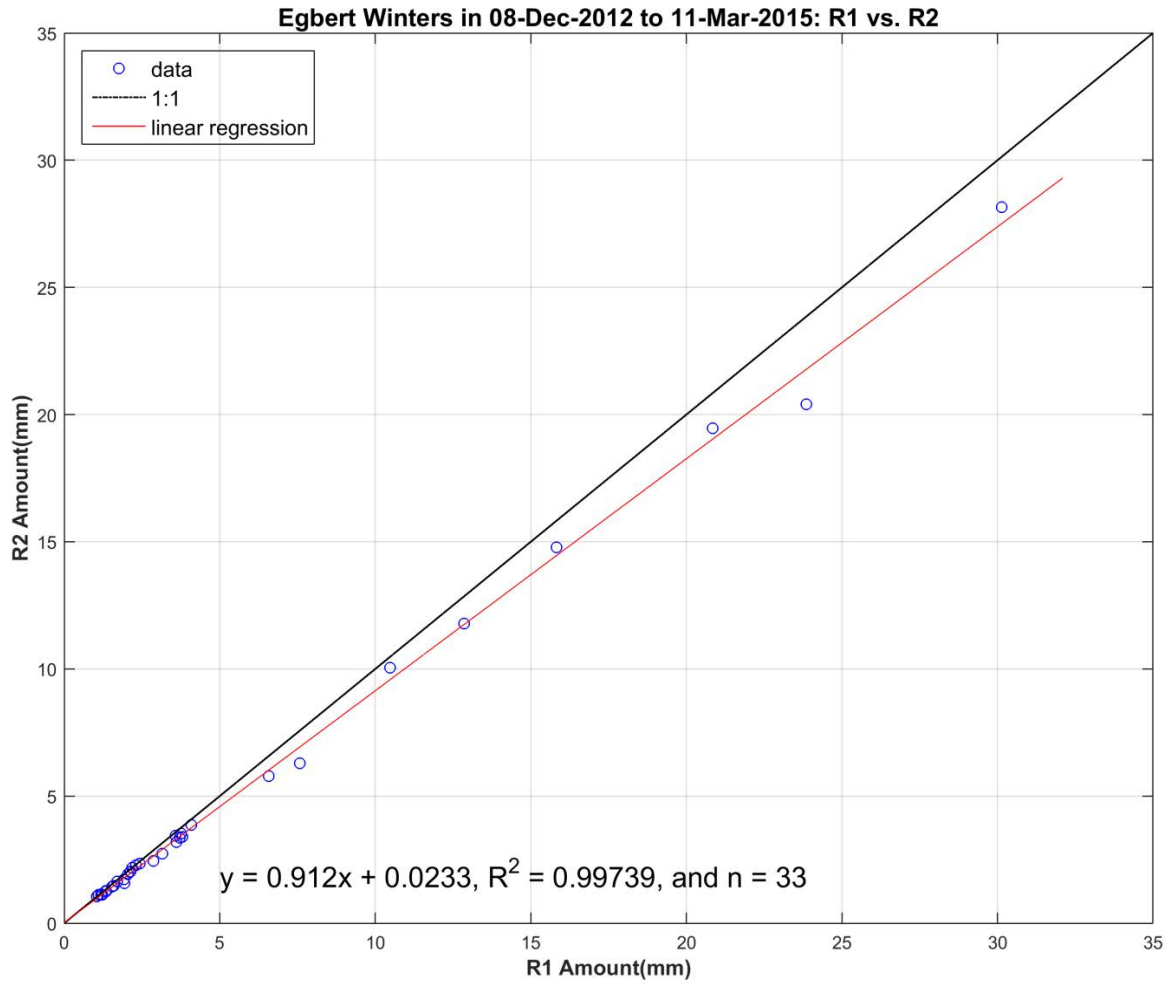


Figure 3.39. Scatter plot of R2 vs. R1 precipitation accumulations when both increments are greater than 1 mm. A linear regression test indicates that the slope of 0.912 is significantly different from 1 at the 5% significance level.

3.2.2.3.4.1 Impact of the Tretyakov gauge in R1

In a long-term intercomparison of six Tretyakov gauges conducted by Yang et al. (2013), the measurement uncertainty was assessed by determining average catch ratios for each of five gauges relative to the sixth. For snow, the average catch ratios were found to vary from 94% to 106%, suggesting a measurement uncertainty of ± 6% for Tretyakov gauges. This uncertainty provides one potential explanation for the difference in results between the R1 and R2 reference configurations observed in Section 3.2.2.

The laboratory uncertainty of the Tretyakov collector and calibration error of the scale used to weigh the collector were both within 0.1 g, or about 0.01 mm snow water equivalent, assuming a density of 1000 kg/m^3 and a collector area of 200 cm^2. These errors are not sufficient to explain the approximately 7% difference between R1 and R2 accumulation reports, however. The uncertainty of the collector is another avenue to consider. An error of about 3.7% in the radius of the collectors (r_0) would produce an offset of 7% in the computed precipitation amount; however, an error of this magnitude is unlikely, as the radius of one of the collectors was within 0.3% of r_0.

3.2.2.3.4.2 Assessment of the accumulation of frost on R1 collector

Frost can form on the inside as well as the outside wall of the R1 collector and could introduce additional errors in the measurements. As a standard practice, these instances are noted by the observer and used in the interpretation of results. According to the measurement procedure, the outside frost and any adhered precipitation is removed before the gauge is weighed (Earle, 2013).

On February 10, 2013, the observer noted that the collector was covered with frost on the outside and, to a lesser extent, on the inside. After cleaning the outside, the collector weighed 1641.6 g. With the inside also cleaned, and including the partial fin, the collector weighed 1638.6 g. The empty collector weighs 1638.3 g. Without the removal of the frost accumulation, the measurement would have been 0.05 mm/g × (1641.6 g – 1638.3 g) = 0.16 mm, likely not the result of a precipitation event. A picture of the R1 manual gauge with frost is shown in Figure 3.40.

Figure 3.40. Frost on the R1 manual gauge at CARE. (Photo credit: P. Raczynski)

In this particular case, the observer noted that it was clear the day before, and there was a morning fog on February 10. This observation was confirmed by the two weather sensors at the site. It was strongly suspected that the increase in the collector weight was due to frost resulting from the fog.

For the 33 precipitation events during the study period (events with accumulations greater than 1 mm, maximum temperatures less than -2 °C, and durations less than 48 hours), frost was reported for only two events; hence, the contribution of frost to the differences between R1 and R2 accumulation reports is not expected to be significant. An additional 18 frost events were reported for which the accumulation was below the set threshold of 1 mm per event.

3.2.2.3.4.3 Assessment of accumulation threshold and temperature limits for solid precipitation events

An assessment of the impact of the value of the precipitation accumulation threshold selected for an event was conducted. The threshold accumulation for both R1 and R2 configurations was varied between 0.1 mm and 1.5 mm, while the maximum temperature limit for a solid-precipitation event was maintained at -2 °C. The results are presented in Table 3.5, and show that the mean CE varies between 92% and 96% over the range of accumulation thresholds tested, while the standard deviation of CE increases as the threshold is decreased.

Table 3.5. Effects of different accumulation thresholds on the average catch efficiency and other parameters for R1 and R2 reference configurations at CARE.

R1, R2 threshold [mm]	Number of Events	Mean CE	St. Dev. of CE	Slope of Linear Regression of CE vs. Wind Speed	R^2
1.5	25	0.919	0.048	0.00213	0.0033
1	33	0.928	0.049	-0.0048	0.017
0.8	38	0.936	0.057	-0.00536	0.014
0.5	44	0.93	0.077	0.0056	0.0079
0.3	58	0.947	0.099	0.0155	0.035
0.1	68	0.964	0.2	0.0159	0.008

Next, keeping the R1 and R2 event accumulation threshold at 1 mm, the temperature limit for a solid-precipitation event was varied. The effects of this variation of the temperature limit are given in Table 3.6. The results show an increase in the number of events and a decrease in the CE, with an increase in the standard deviation of the CE as the maximum temperature was increased above -2 °C. Opposing trends were observed when the temperature limit was reduced from -2 °C to -5 °C.

Table 3.6. Effects of different temperature limits on the average catch efficiency and other parameters for events with an accumulation > 1 mm for R1 and R2 configurations at CARE.

Temp Limit [°C]	Number of Events	Mean CE	St. Dev. (CE)	Slope of Linear Regression of CE vs. Wind Speed	R^2
-5	23	0.94	0.046	0.00112	0.001
-2	33	0.928	0.049	-0.0048	0.017
0	43	0.931	0.06	-0.00298	0.0034
2	56	0.913	0.084	-0.00838	0.015
5	71	0.909	0.093	-0.0106	0.018

3.2.2.3.4.4 Impact of the Geonor gauge in R2

The R2 Geonor gauge was assessed by a field verification in July 2013 (Mohamed et al., 2013). The errors between the gauge amount (i.e. the average of the three transducer amounts) and the reference amount generated by water weighed to 1 kg (i.e. 50 mm in depth) have a maximum of 0.15 mm, a minimum of -0.14 mm, an average of 0.03 mm, and a median of 0.01 mm. The percentage errors between the gauge amount and the reference amount have a maximum of 0.29%, a minimum of -0.28%, an average of 0.06%, and a median of 0.02%. Thus, the calibration of the Geonor gauge in the R2 reference configuration is not likely the cause of the difference in CE observed between the R1 and R2 configurations.

3.2.2.3.4.5 Impact of heating of the R2 gauge

The impact of heating the R2 Geonor gauge was examined within the context of the difference in CE between the R1 and R2 accumulation reports. The inlet orifice of a Geonor gauge is a long tube hanging above the weighing bucket. The R2 Geonor inlet orifice is equipped with upper and lower heaters following the approach of the USCRN (C-SPICE – Precipitation Gauge Heating Summary, 2013). (Heaters (lower and upper rim) are turned on if the ambient temperature is between -5° and +5 °C inclusive and the corresponding rim temperature is below 2 °C.)

During its passage through the inlet, snow can adhere to the inner wall and will not reach the bucket to be weighed and recorded. Heating the inlet helps to mitigate the influence of adhesion, but could also lead to evaporation of some of the precipitation near or on the inlet wall. The Tretyakov collector in the R1 configuration is not subject to these adhesion and evaporation losses, as the collector is unheated and any adhered precipitation is weighed as part of the measurement.

Tests conducted at the NCAR Foothills snow machine laboratory using a Geonor T-200B3 gauge (Colli et al., 2013) showed an underestimation of accumulated precipitation when the snowfall rates are very low and the heater is active. On the other hand, no evidences of heated plumes or subsequent upward airflows were detected by tracking the falling snow flakes trajectories with a video approach. Therefore, the only source of uncertainty directly ascribable to the usage of the heater is the evaporation of the melted snowflakes along the internal heated surfaces. This effect could impact appreciably on the GEONOR T200B measurements since the size of its internal heated surface is equal to 1655 cm^2, more than four times the corresponding diameter of the OTT Pluvio^2 gauge (382

cm2). Colli et al., (2013) also observed that the wet surface of the inner collector wall could be responsible for the retention and evaporation of snowflakes when the snowfall rate is low.

3.2.2.3.4.6 Assessment of the impact of Geonor rim heating

To investigate the impact of the Geonor rim heating, the behaviour of three Geonor T-200B3 gauges in single-Alter-shields installed at the CARE site were considered, one with heaters (H6) and two without (H1 and H8), for two periods in 2013/14 and 2014/15. The heating configurations of Geonor H6 and Geonor R2 are similar, and the three Geonor gauges have the same physical and data configuration (i.e. all have three transducers and are installed in single-Alter shields). For this assessment, hourly increments were calculated for the three gauges. Precipitation events were assessed when all three gauges reported amounts equal to or greater than 0.5 mm. These hourly amounts were separated into two groups:

1. H6 Heater off: upper and lower heater status off over the hour
2. H6 Heater on: either of the heaters were on for at least 30% of the time in a given hour

Figure 3.41 shows a plot of the relative catch efficiencies for each gauge pair, H6/H1, H6/H8 and H8/H1, for the period from December 1, 2013, to May 31, 2014. The plot shows that there is on average about 7% difference in catch efficiency between the Geonor heater when it is on and when it is off. The average catch efficiency of the two Geonor gauges without heater, H1 and H8, are close to 1, as shown in the third plot.

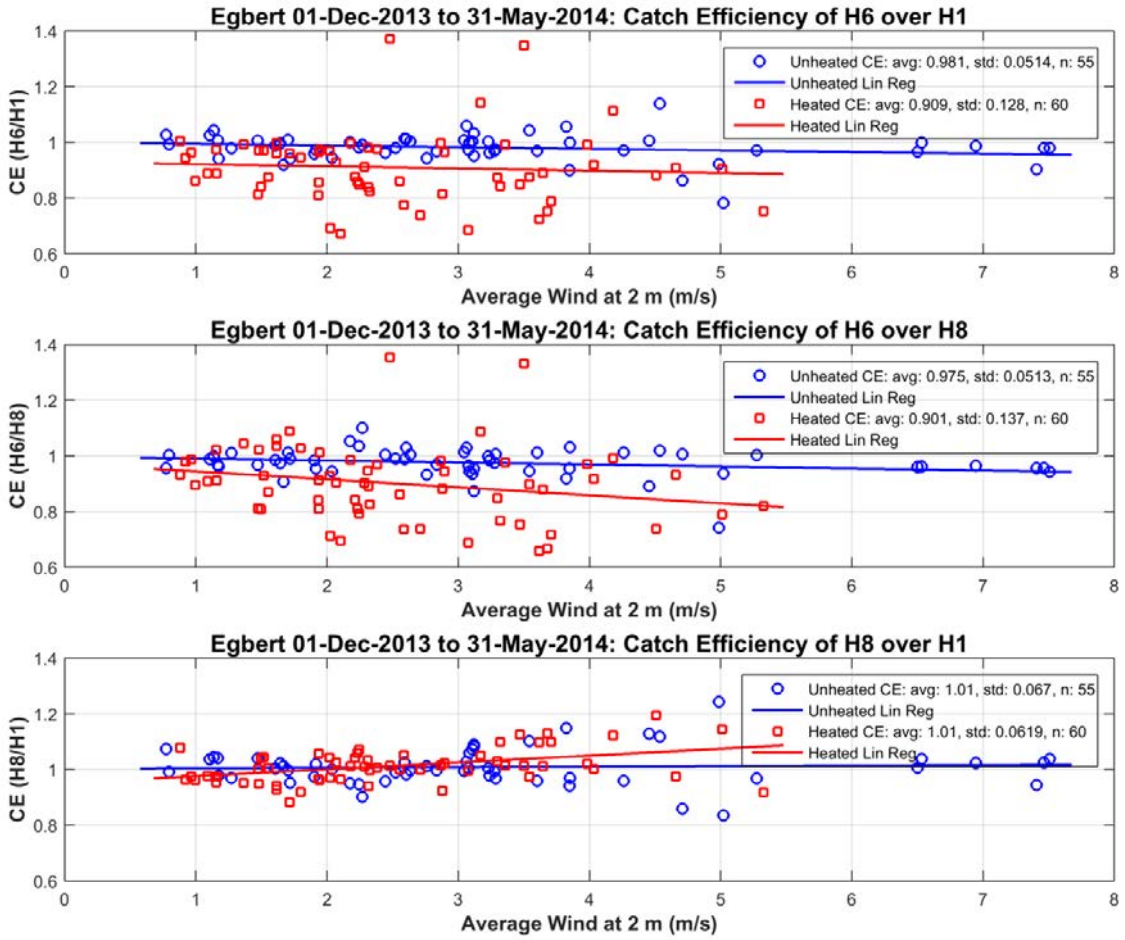


Figure 3.41. Catch efficiency of H6/H1, H6/H8, and H8/H1 for the period from December 1, 2013, to May 31, 2014.

The catch efficiency of H6/H1, H6/H8, and H8/H1 for the observation period in 2014/15 is shown in Figure 3.42. The catch efficiency differs markedly from that of 2013/14. In particular, there are groups of datapoints when the H6 heater is off for which the catch efficiency either falls below 80% or exceeds 120%.

Investigation of the specific events represented by the points in the yellow boxes (not shown) indicated the potential for wind shadowing of H8 when the wind was from the north. This was attributed to the presence of a new DFIR-fence that was installed north of H8 in November 2014 and, hence, did not impact the 2013/14 data in Figure 3.41. The dataset was filtered to remove events when the wind direction was from the north (from 20° to 320°); the filtered results are plotted in Figure 3.43 and show that the difference in mean catch efficiency is again about 6% to 7% between when the H6 heater is on and off.

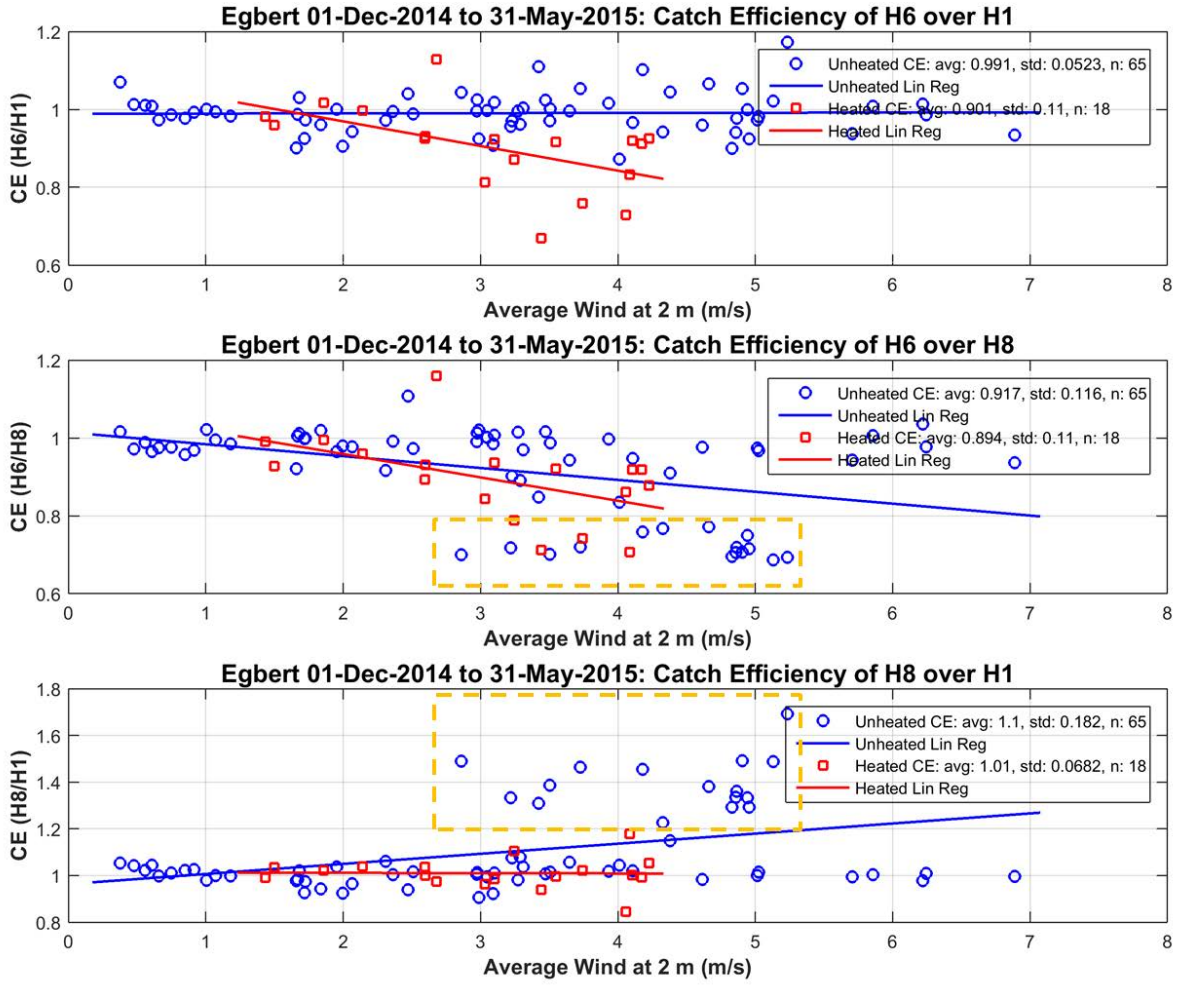


Figure 3.42. Catch efficiency of H6/H1, H6/H8, and H8/H1 for the period from December 1, 2014, to May 31, 2015. The yellow boxes indicate events with catch efficiencies ≤ 80% or ≥ 120%.

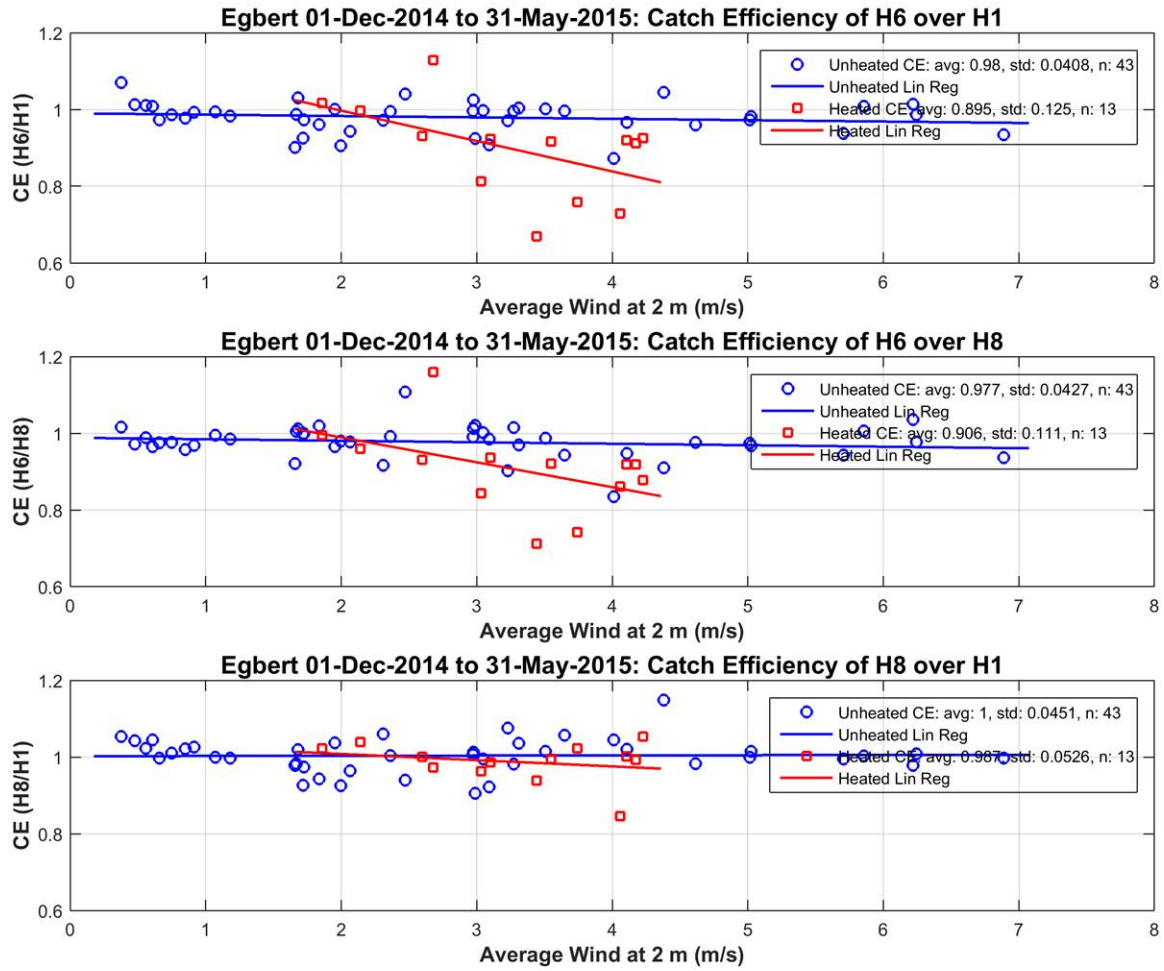


Figure 3.43. Catch efficiency of H6/H1, H6/H8, and H8/H1 for the period from December 1, 2014, to May 31, 2015, following the filtering of events with North wind (from 20° to 320°).

The consolidated results for the 2013/14 and 2014/15 seasons, filtered to remove events affected by wind shadowing, are shown in Figure 3.44. The difference in mean catch efficiency is again about 6% to 7% between when the H6 heater is on and off. The yellow boxes in Figure 3.44 identify two specific events with catch efficiencies > 120%. Case study analyses (not shown) indicated that these higher-than-expected catch efficiencies likely resulted from precipitation freezing on the rim of gauges H1 and H8 and not being collected in the respective buckets and reported. These cases serve as an important reminder of the intent of rim heating, which is to prevent precipitation from accumulating on the rim. A gauge with rim heating should catch more than one without rim heating when precipitation does, in fact, accumulate on the rim. However, rim heating does have an apparent side effect: it can result in lower-than-expected catch efficiencies on the order of 6% to 7%, as shown in Figure 3.44.

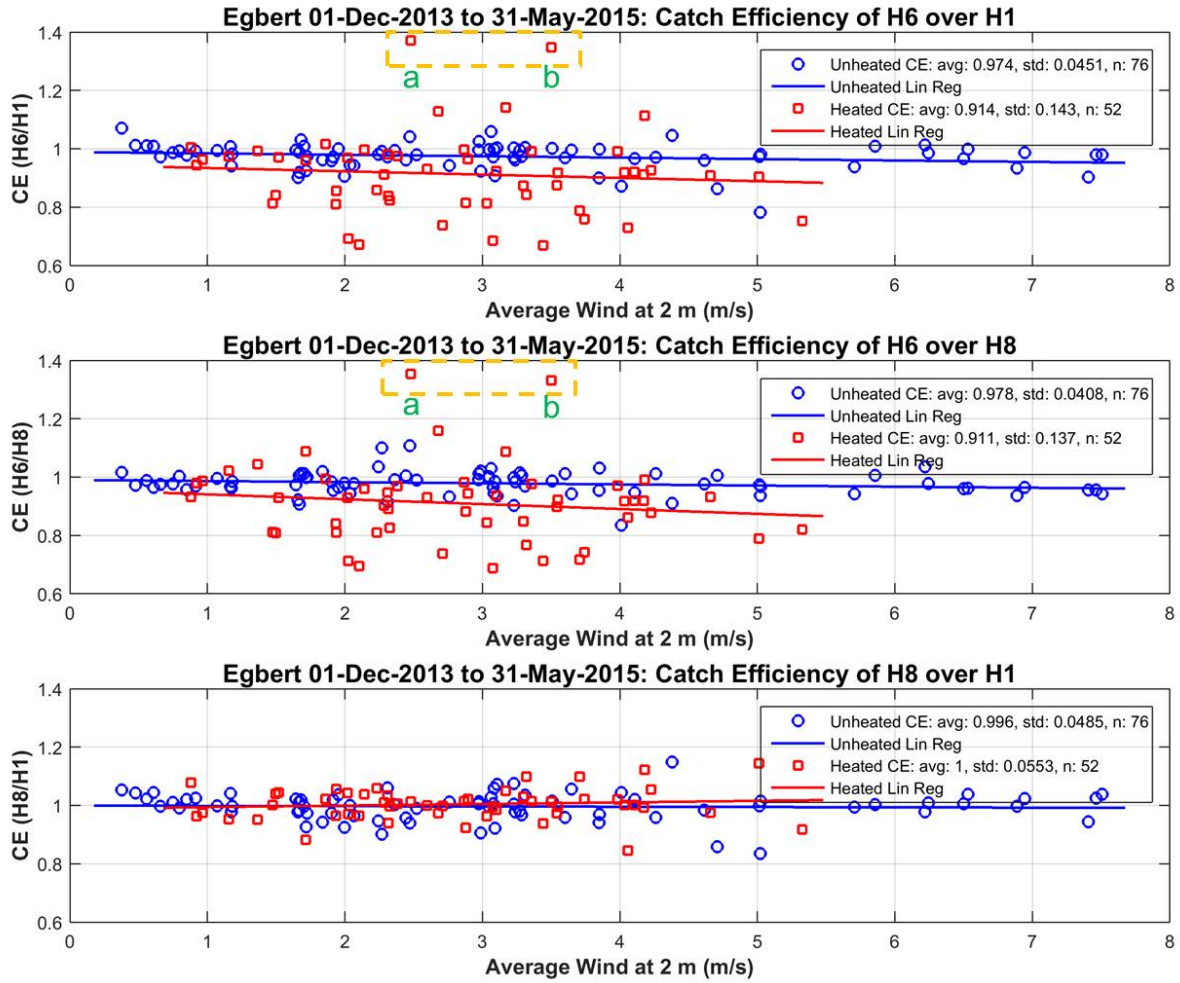


Figure 3.44. Catch efficiency of H6/H1, H6/H8, and H8/H1 for the period from December 1, 2013, to May 31, 2015. The data have been filtered by wind direction to remove events impacted by wind shadowing from the north. The yellow boxes indicate points with catch efficiencies > 120%.

Finally, the data in Figure 3.44 were replotted as a function of the average temperature over the hour, rather than as a function of the average wind speed over the hour. (See Figure 3.45) Enhanced dispersion of events is evident over the temperature range from -5 °C to +2 °C, when the H6 heating is applied according to the USCRN approach.

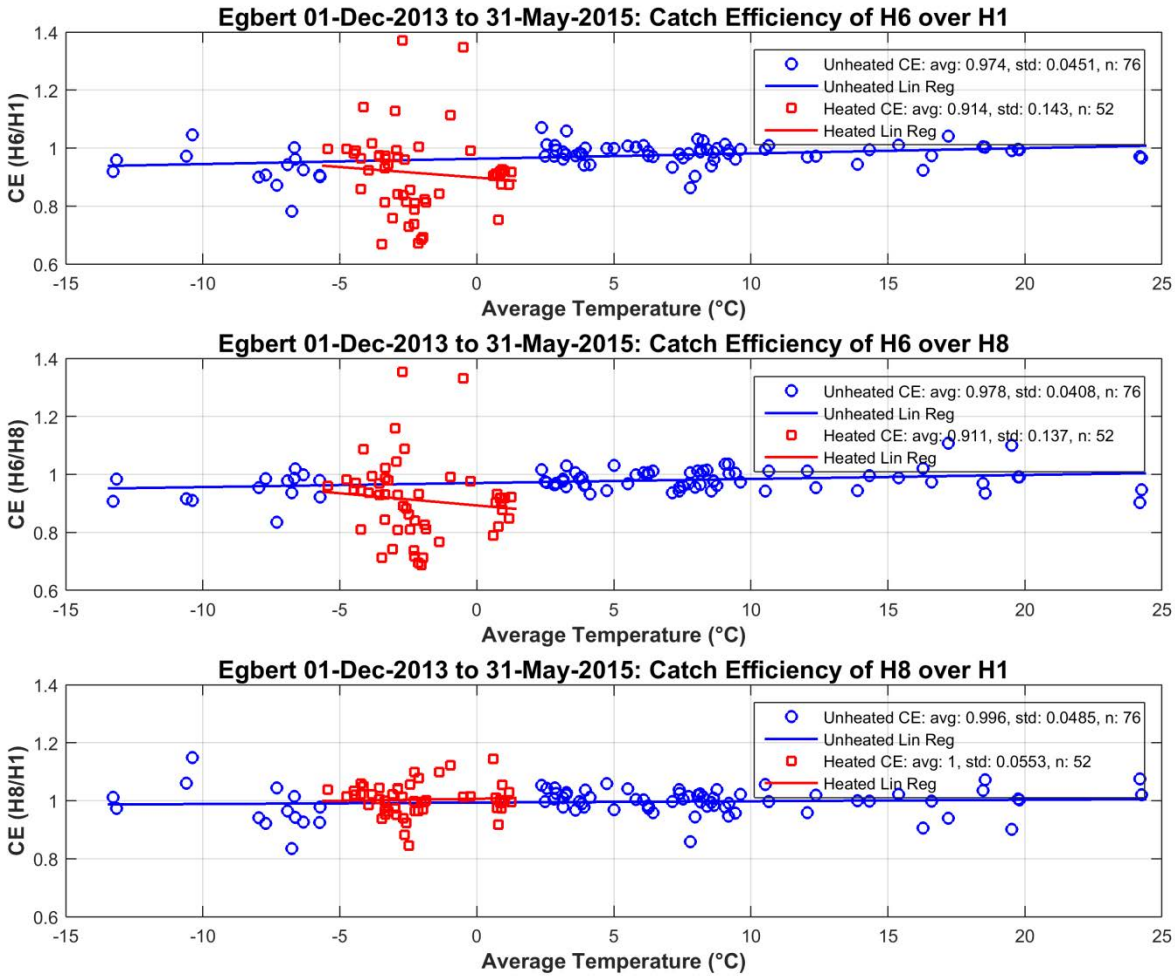


Figure 3.45. Catch efficiency of H6/H1, H6/H8, and H8/H1 for the period from December 1, 2013, to May 31, 2015, as a function of mean event temperature.

3.2.2.4 Summary

Data from a manual Tretyakov gauge and a Geonor T-200B3 automatic gauge, both inside DFIR double fences at the CARE site, were analyzed for the winter seasons of 2012/13, 2013/14, and 2014/15. The analysis shows that the dependence of the catch efficiency on the mean wind speed is not strong for the range of 1 to 6 m/s and that the automatic gauge catches about 7% less precipitation than the manual gauge. The calibrations of the scale for weighing the precipitation from the manual gauge and the transducers of the Geonor gauge were verified and associated uncertainties presented. Other effects were also considered, such as the uncertainty and area of the R1 collector, the influence of frost on the R1 observations, and the impact of the specific thresholds (accumulated precipitation, temperature) used to identify the events considered in the analysis.

An investigation of data from three Geonor gauges in single-Alter shields at CARE, one heated and two unheated, was conducted to assess the influence of heating on reported accumulation. It was found that heating does increase gauge catch efficiency in some instances (when the potential exists for precipitation freezing/accumulating on the rim), but generally results in lower catch efficiency, perhaps due to evaporation on the rim caused by the heating. The difference is about 6% to 7%, similar to the difference between the mean catch efficiencies of the R2 and R1 configurations. These

results suggest that the heating of Geonor R2 may be the main reason for the 7% difference in catch efficiency.

3.2.2.5 Comparison of previous intercomparison results

The results presented in Section 3.2.2 reflect three different experiments, organized on three sites in different climates. An overall catch efficiency of about 93% between the DFAR and DFIR has been reported for the CARE and Bratt's Lake sites, while the catch efficiency reported for Jokioinen is 99%. All three experiments used a Geonor gauge in the DFAR and a Tretyakov gauge in the DFIR. While the experiments from Bratt's Lake and Jokioinen used unheated Geonor gauges, the Egbert (CARE) DFAR used a heated Geonor in the manner defined for the field working reference for SPICE.

This report examined several factors and their potential contribution to an assessed catch efficiency of 93%. An investigation of the impacts of gauge heating performed at CARE indicated that a 6% to 7% undercatch could be attributed to the application of heat. This does not, however, explain the undercatch observed for the unheated Geonor gauge in the R2 configuration at Bratt's Lake. Additional work is required, using data from other environments, to better understand the different factors and their contributions, and to further assess the representativeness of the present catch efficiency results within the broader context of the traceability of measurements from different reference configurations.

3.2.3 Gauge-based comparison of R2 references: R2G (Geonor) vs. R2P (Pluvio^2)

Authors: Soorok Ryu, GyuWon Lee, Rodica Nitu, Craig Smith

3.2.3.1 Introduction

For assessing automatic instruments and to address the need for high temporal sampling and reporting frequency down to minutely scales, field references using automatic instruments were used in SPICE. The automatic field working reference type R2 configured for SPICE is described in Section 0. When available, data from a precipitation detector were integrated with the data from the weighing gauge for the derivation of the R2 reference dataset.

Automatic weighing gauges were selected for the DFAR based on their breadth of operational use and existing documentation of their performance, a principle similar to that used for the 1989-1993 WMO Solid Precipitation Measurement Intercomparison (Goodison et al., 1998). Two weighing-type gauges were used as part of the SPICE R2 reference system; these were the Geonor T-200B3 gauge (with 3 transducers) and the OTT Pluvio² (see Figure 3.46), both with heated rims. Each participating site configured its R2 field reference with the weighing gauge most widely used in their respective organization or program. Since both reference systems were used for SPICE intercomparisons, it is important to quantify and understand any differences in their behaviour.

In this study, an R2 system employing a Geonor gauge is referred to as R2G, while an R2 system using a Pluvio² gauge is referred to as R2P.

Figure 3.46. Weighing gauges used in R2 reference configurations: (a) Geonor T-200B3, (b) OTT Pluvio^2 (photos from CARE SPICE site).

3.2.3.2 R2 reference dataset

The SPICE reference dataset is, ideally, an unbiased, low variance, noise filtered, artifact-free data set with great sensitivity, independent of the type of gauge used. The ability to relate either the R2G or R2P reference datasets is critical for the interpretation of results for instruments tested at different locations. This section summarizes the comparative results of the two reference systems using data from three SPICE sites: CARE (Canada), Bratt's Lake (Canada), and Gochang (Rep of Korea). Each site was equipped with two DFARs, one R2G and one R2P.

The analysis was conducted using data from two winter seasons. At CARE, the periods were from December 1, 2014 to April 30, 2015, and December 1, 2015 to March 31, 2016. At Bratt's Lake, the periods were from January 17, 2015 to May 20, 2015, and December 1, 2015 to March 28, 2016. At Gochang, data for this analysis were from December 12, 2014 to February 28, 2015, and December 1, 2015 to January 26, 2016.

On the CARE site, a disdrometer-type Thies Laser Precipitation Monitor (LPM) was installed within the R2G in 2013, and was used as a precipitation detector for the derivation of the reference dataset. The gauges on the CARE and Bratt's Lake sites were heated according to the USCRN approach. On the Gochang site, the Pluvio^2 gauges used the rim heating algorithm provided by manufacturer, similar to CARE and Bratt's Lake, but all Geonor gauges were of 1000 mm capacity and were not equipped with rim heating. Details on the configuration of each of these three sites are available in Section 2 and Annex 4.

Data characteristics and sampling strategies for both gauges are described above in Section 3.1.3.7.1 and in Table 3.2. Several data products are included in the message output of the Pluvio^2. The Bucket RT is considered the lowest level data available from the gauge (considered for the purpose of this analysis as "raw" data) and was used for this analysis. This helps to ensure the direct comparability of the data from the two gauges, for the purpose of this analysis.

3.2.3.3 Data and method

3.2.3.3.1 *Derivation of precipitation data*

This analysis is based on the one-minute accumulation data from the Pluvio^2 (R2P) and Geonor (R2G) gauges, quality-controlled (QCed) by applying the SPICE QC methodology (see Section 3.3.2). For a Geonor gauge, the QC process is applied to each of the three vibrating-wire transducers, and the final Geonor minutely data is obtained by computing the average of the values from the three transducers. When the data was collected with a 6 second temporal resolution (e.g. CARE), a 1-minute value is obtained by aggregating (or mean averaging) the 6 second values for each 1 minute interval. For a Pluvio^2 gauge, minutely data were calculated based on the Bucket RT output, either as an average of the 6 second Bucket RT values, where available, or as the minutely data output collected.

Table 3.7 summarizes the information on the heating of the R2 gauges, their sampling interval on each site, and the respective ancillary measurements.

Table 3.7. Summary of ancillary measurements and temporal resolution of data, by site.

	CARE	**Bratt's Lake**	**Gochang**
Time resolutions of R2G and R2P	6s	1min	6s
Rim heating	All heated	All heated	R2G : not heated
Temperature (height), T	Vaisala HMP155 (1.5 m)	Campbell HMP45C (1.5m)	WS-T100G1 (1.6 m)
Relative Humidity (height), RH			EE180 (1.6 m)
Wind speed (height), U	Vaisala NWS425 (2 m)	RM Young 5103 (2.2 m)	JY-WS161B (1 m)

For the comparison of data from the two R2 references, R2G and R2P, a 30 minute sampling interval was used to obtain the precipitation intensity. All comparisons were made when at least one value, R2P or R2G, was positive, or all values were greater than or equal to a predefined threshold, P0. Three values for the threshold P0 were considered; these were 0.10 mm/h, 0.25 mm/h, and 0.50 mm/h. The horizontal wind speed U, air temperature T, relative temperature RH, and dew point temperature T_d have been averaged over the same precipitation sampling interval of 30 min. The dew point temperature is computed from RH and T using the formulation in Wagner et al. (2002).

In this work, the precipitation types were separated into snow, mixed precipitation, and rain, using the following temperature conditions:

$$\text{precipitation type} = \begin{cases} \text{snow} & T < -2.0\ ^\circ\text{C} \\ \text{mixed precipitation} & -2.0\ ^\circ\text{C} \leq T \leq 2.0\ ^\circ\text{C} \\ \text{rain} & T > 2.0\ ^\circ\text{C} \end{cases}$$

The approach of using the air temperature as condition to separate precipitation type for bias corrections is similar to that used in Yang et al. (1995, 1998).

Figure 3.47 shows the time series of one-minute accumulation data from R2G and R2P for each measurement season, at all three sites. The blue line is the minutely Geonor reported accumulation data and the red line is the minutely Pluvio^2 reported accumulation data.

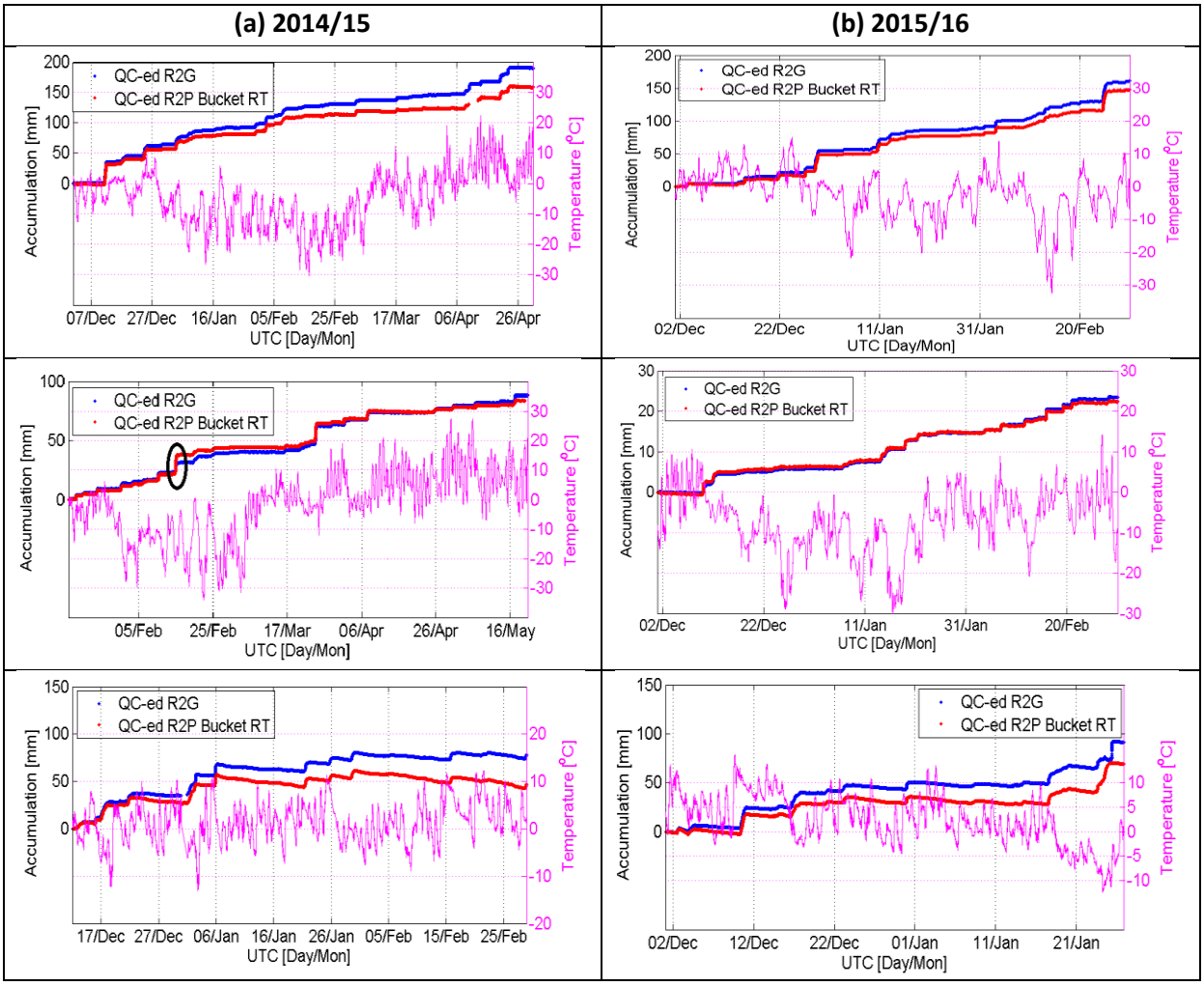


Figure 3.47. QCed accumulation data of R2G and R2P, and 1 min temperature at CARE (first row), Bratt's Lake (second row), and Gochang (third row) in (a) 2014/15 winter and (b) 2015/16 winter.

Figure 3.48 represents the time series of horizontal wind speed and relative humidity on all three sites, for each measurement season. The notable point is that the maximum values of wind speed at Gochang and Bratt's Lake were marginally higher than those from CARE.

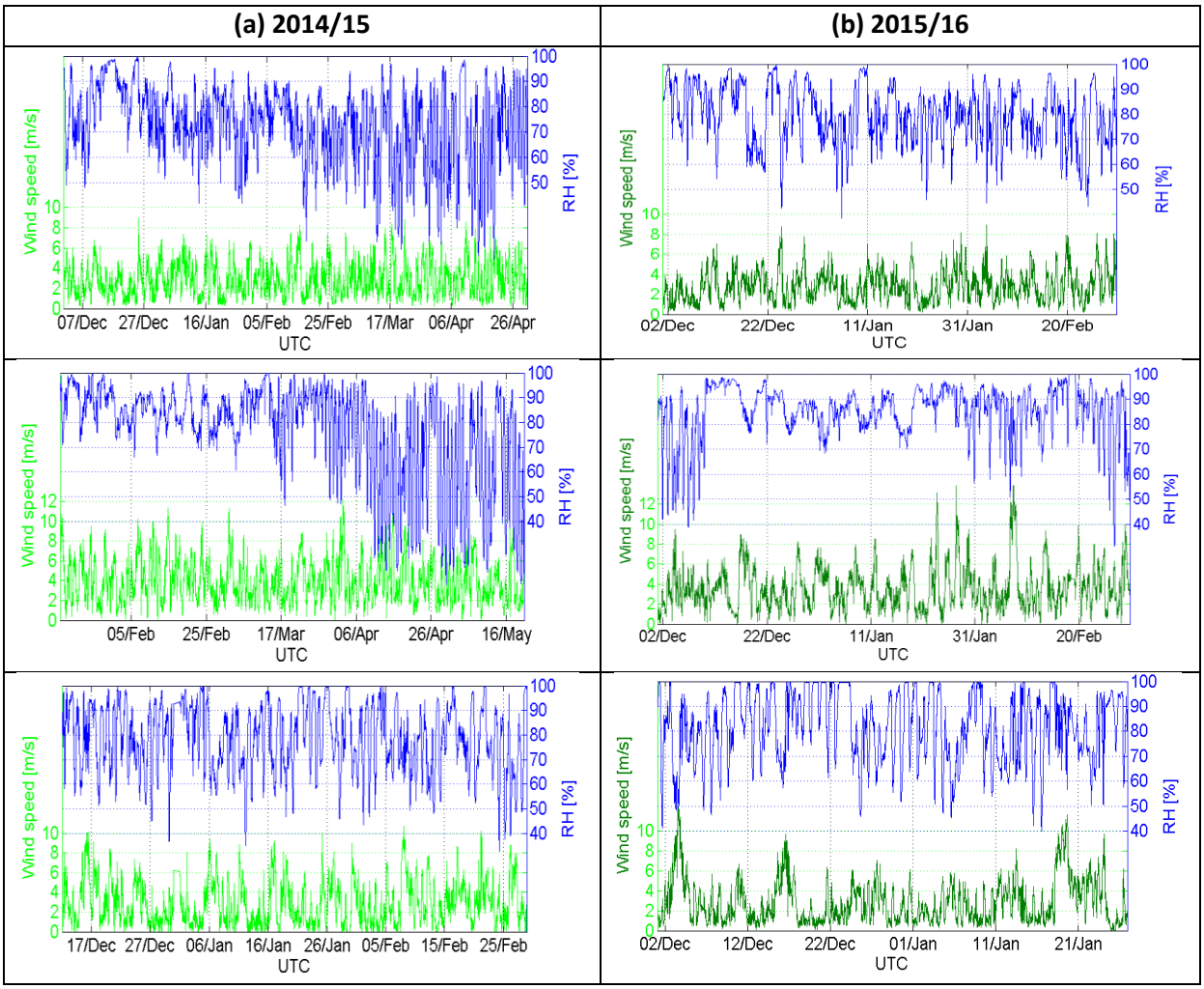


Figure 3.48. Time series of QCed wind speed and relative humidity at CARE (first row), Bratt's Lake (second row), and Gochang (third row) in (a) 2014/15 winter. and (b) 2015/16 winter.

3.2.3.3.2 *Method*

The comparison analysis between the two reference systems, R2G and R2P, was performed by examining the distribution of their differences and the ratios of the data from the two systems.

The ratio is calculated using concurrent values reported by the two systems, and the measurement of R2G was used as the denominator. Quartile statistics and transfer functions were used to describe the distributions of these ratios.

3.2.3.4 Results

3.2.3.4.1 *Environmental conditions*

Figure 3.49 shows the probability distribution function (PDF) of the number of precipitation events, for an intensity threshold of P0 = 0.25mm/h, stratified by precipitation rate, wind speed, air temperature, and relative humidity. For all events, the peaks of frequency appear for wind speeds of 5-6 m/s, 2-4 m/s, and 1-3 m/s, for Bratt's Lake, CARE, and Gochang site, respectively.

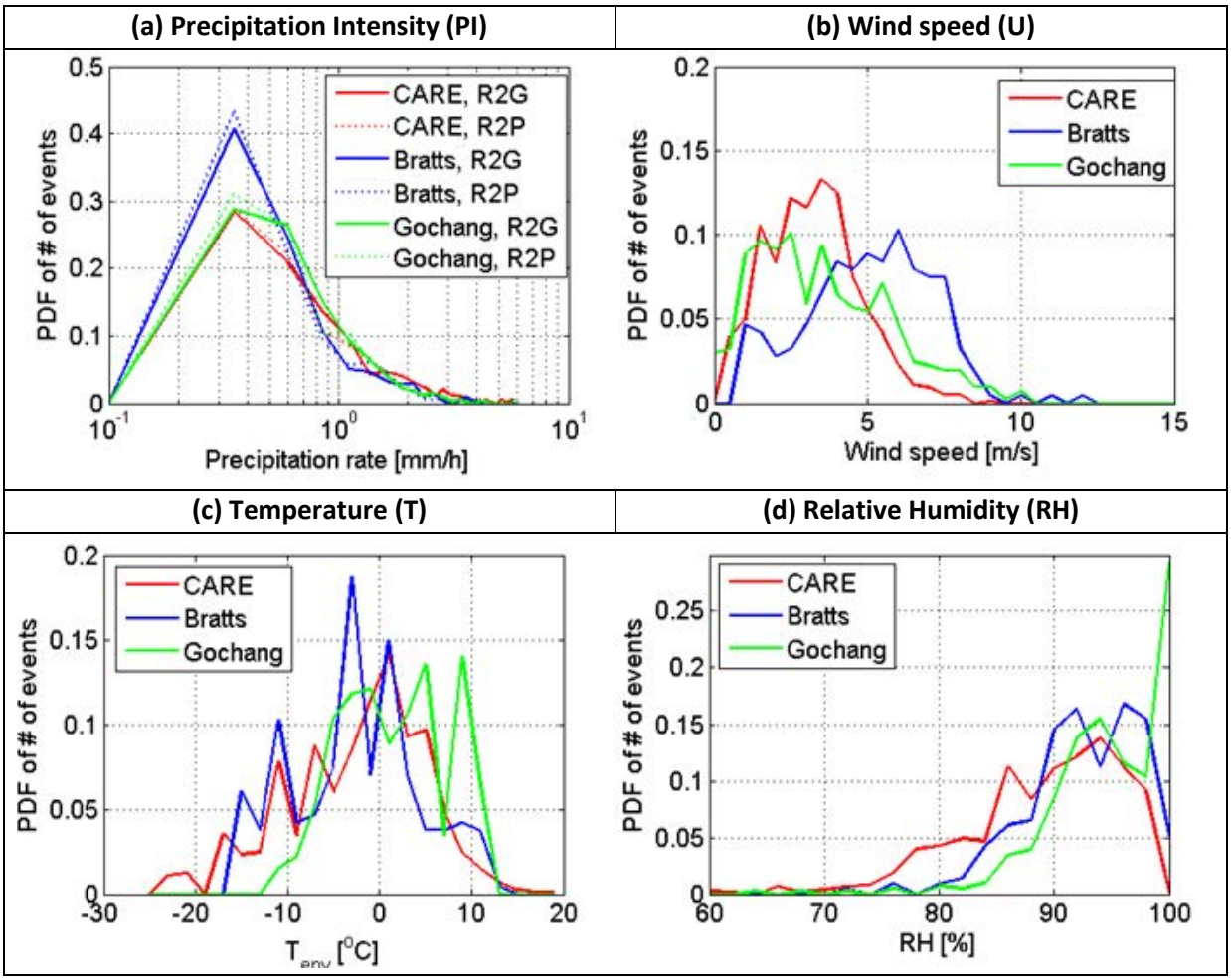


Figure 3.49. Probability density functions (PDFs) of precipitation intensity, wind speed, environmental temperature, and relative humidity for 30 minute intervals and P0 =0.25 mm/h.

3.2.3.4.2 *Comparison of sensitivity for different thresholds*

The impact of the precipitation intensity threshold on the detection and measurement of precipitation, and on the relative comparison of the R2G and R2P data, is assessed. When computing precipitation rates from accumulated precipitation values reported by automatic weighing gauges, it is important to find an optimal threshold to limit the noise, while avoiding the loss of "true" data.

Figure 3.50 shows the distributions of concurrent 30 min P2G and R2P values as a function of the dew point temperature, Td, (x-axis), the air temperature, T (y-axis), as well as the threshold P0. The blue circles represent the events when the R2G value is larger than the corresponding R2P value, while the red dots represent the events when R2P>R2G≥0. From this figure, it is observed that the agreement between two gauges increases and the shape of their distribution narrows, as the threshold P0 increases. A higher R2P value is noted in the area below the diagonal 1:1 line, and for lower P0 values. This is interpreted as the R2P values being higher than the corresponding R2G values in low relative humidity condition, and for lower P0 values.

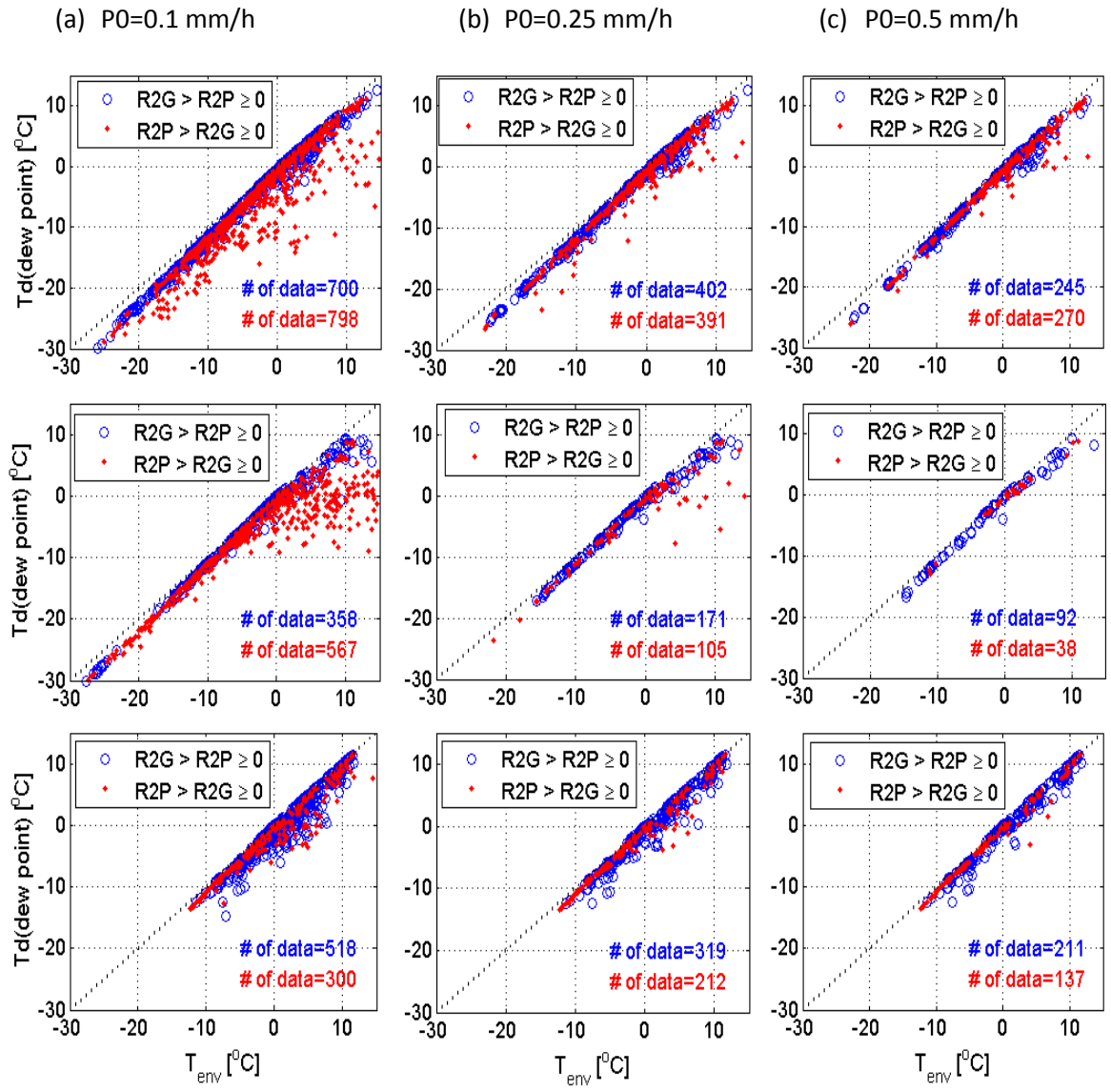


Figure 3.50. Scatter plots of temperature (x-axis) and dew-point temperature (y-axis) when R2P and R2G report different values, obtained by varying the threshold P0=0.1, 0.25, and 0.5 mm/h. The red dots and blue circles represent the events when R2P>R2G and R2P < R2G, respectively.

3.2.3.4.3 *Detectability of precipitation*

To test the detectability of precipitation by the two gauges, a targeted assessment was conducted using the data from the Thies LPM at the CARE site, for the period from December 1 2015 to February 29 2016. The 30 min intensity data from the LPM were calculated as the 30 min average of the one minute precipitation intensities reported by the sensor. Figure 3.51 shows the scattered points when one gauge recorded a precipitation amount, whereas the other did not, for different dewpoint temperatures (y-axis) and temperatures (x-axis), and for different threshold values, P0.

Similar to Figure 3.50, the red dots in Figure 3.51 represent the events when the R2P intensity is larger than that reported by R2G, while the blue dots represent the events when values reported by R2G are larger than those reported by R2P. Additionally, the cyan circles represent the events when both the R2G and the LPM report positive intensities, but R2P reports no precipitation (i.e. below the

imposed threshold), while the magenta circles represent the events when the R2P and the LPM intensities are positive, and the R2G intensity is below the threshold. When P0=0.1 mm/h, the number of datapoints represented is 175, which decreases significantly to 37 when the threshold is increased to P0 = 0.25 mm/h.

A remarkable result is that the use of the LPM as a second criterion for the identification of precipitation is important when P0 ≤ 0.1mm/h (Figure 3.51a), but its contribution is not noticeable with a higher threshold (Figure 3.51b-c). Some red dots are observed below the diagonal line when P0=0.1 mm/h, which indicates that in lower relative humidity conditions, signal is detected by the R2P but not the R2G, which in fact could include some noise. Figure 3.51d presents the cases when the R2P response appears noisy (non-zero values were determined), although the LPM reports no accumulation of precipitation.

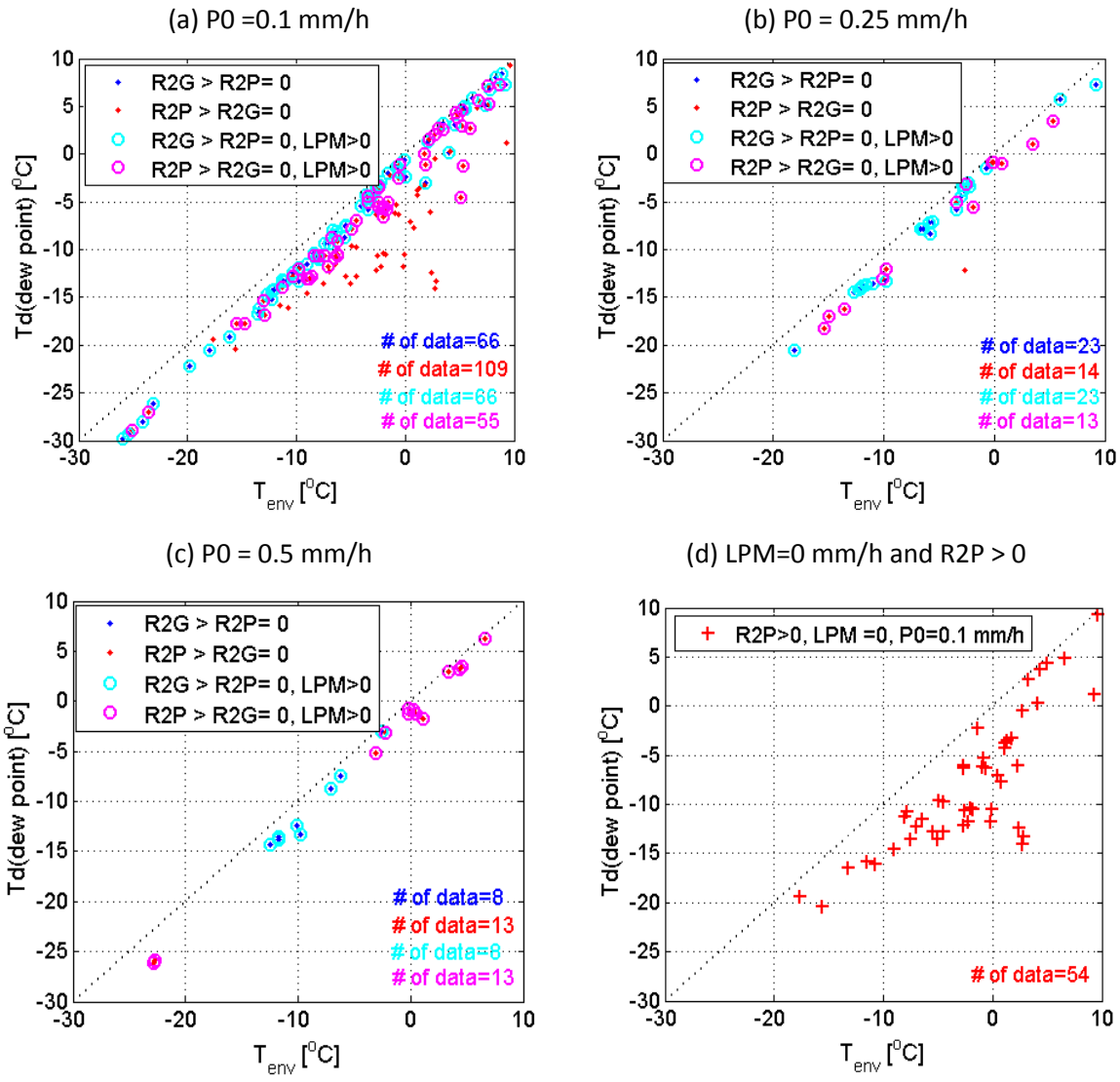


Figure 3.51. Scatter plots of temperature (x-axis) and dew-point temperature (y-axis) when one of R2P and P2G has detected precipitation but the other has not. Cyan dots and magenta dots represent cases when the LPM also reports precipitation. The LPM is located within the DFIR-fence, and the test period is from December 1, 2015 to March 23, 2016.

3.2.3.4.4 *Assessment of the R2P/R2G ratio*

Figure 3.52 shows the distribution of the R2P/R2G catch ratio (CR), by site, by precipitation type, and for the three P0 threshold levels considered. The results show that as P0 is increased, the two values, R2P and R2G, approach each other for all three sites, and for all precipitation types.

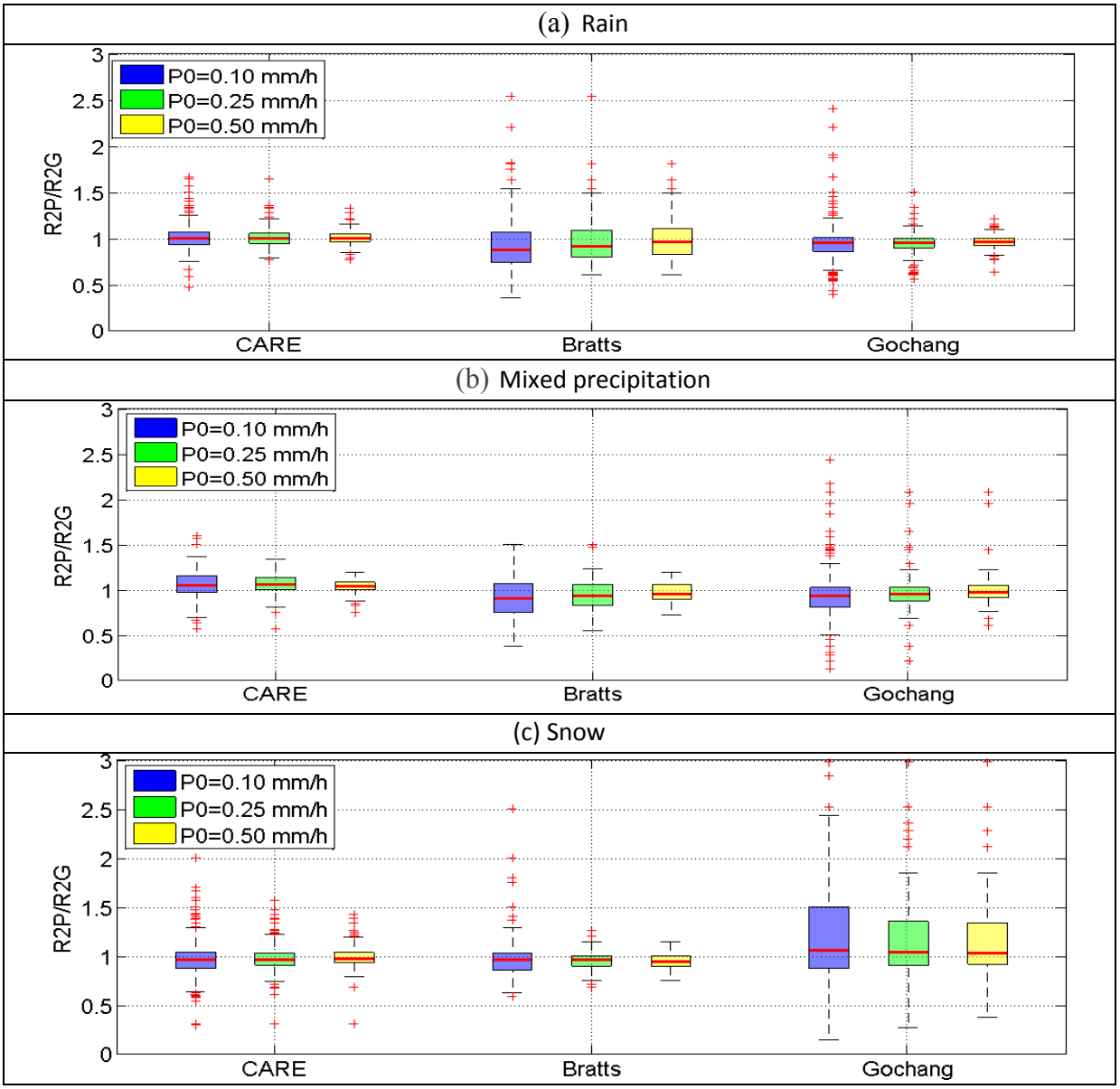


Figure 3.52. Box plots of ratio distributions using 30 min intensity data for different precipitation types (a) rain, (b) mixed, (c) snow, and for different thresholds: P0= 0.1, 0.25, and 0.5 mm/h.

An assessment of results by precipitation type is summarized below, based on the results from Figure 3.52.

3.2.3.4.4.1 Rain events

For the CARE results, the median CR is equal to 1.00 for all thresholds, indicating that no biases were detected. For Gochang, while relatively constant, the CR values hovered around 0.95, indicating slightly larger amounts reported by R2G. For Bratt's Lake, the median CR varies between 0.88 and 0.97, depending on threshold, meaning that the R2P is slightly below the R2G value. For all P0 values used, the interquartile range (IQR) is at or below 14% for CARE and Gochang, and at or below 33% for Bratt's Lake. When P0=0.5 mm/h, the IQR at Bratt's Lake is about 23%, which is larger than the 7% and 8% values from the other two sites, indicating a greater spread in the data at Bratt's Lake.

3.2.3.4.4.2 Mixed precipitation events

The results from CARE and Gochang appear to be quite similar. For CARE, R2P is slightly higher than R2G, with median values of 104-105%, for all thresholds, while the IQR of about 8% is the smallest among the three sites. At Gochang, R2P is slightly higher than R2G and the IQR decreases as the P0 is increased. The Bratt's Lake results show medians of 91% for P0=0.1 mm/h, increasing to 0.96 for P0=0.5 mm/h; at the same time, the IQR varies from 32%, for P0=0.1 mm/h to 17% for P0=0.5 mm/h, indicating marginally higher values from the R2G relative to the R2P.

3.2.3.4.4.3 Snow events

As noted previously, the R2G gauge at the Gochang site was not equipped with rim heating, which led to large differences being noted between concurrent R2P and R2G values. This can be seen from Figure 3.52c, as well, with a larger IQR than for the other sites. For CARE and Bratt's Lake, the median (Q2) varies between 0.95 to 0.98 for all P0 at or above 0.1 mm/h. The IQR values for CARE and Bratt's Lake are about 11% and 10%, respectively, for fixed P0=0.5 mm/h.

3.2.3.5 Development of R2G-R2P transfer function

The results assessed show that the R2G/R2P catch ratio varies with the threshold P0, as well as the air temperature T, and the relative humidity RH. In this section, for snow events, as well as for all types of precipitation, linear transfer functions are developed using linear regression analysis. All transfer functions are computed using the fixed threshold P0=0.50 mm/h, corresponding to the threshold used for the SPICE analysis (see Event Selection details outlined in Section 3.4), and as a function of temperature, relative humidity or wind speed.

3.2.3.5.1 *Transfer functions for snow type*

It is assumed that the catch ratio (*CR*) function is a linear function of first order of *T*, *RH*, and *U*, such that:

$$CR(T, RH, U) = a_0 + a_1 T + a_2 RH + a_3 U$$

The coefficients of this linear equation were determined using regression analysis. In the above equation, the coefficient for *U* is relatively small compared to those for *T* and *RH*. Leaving out the wind speed *U* term, the R2G/R2P transfer function becomes:

$$CR = 1.5476 + 0.0107T - 0.0054RH \ (N = 260, R^2 = 0.9726)$$

Similarly, transfer functions were determined for all precipitation types by assuming the regression equation is the following second order form:

$$CR(T, RH) = a_0 + a_1 T + a_2 T^2 + a_3 T \times RH + a_4 RH + a_5 RH^2$$

Using the data of CARE and Bratt's Lake sites, the second order transform equation is:

$$CR = 1.4318 - 0.0176T - 0.0004T^2 + 0.0002T \times RH - 0.0065RH - 0.0000RH^2$$

$$(N = 595,\ R^2 = 0.9825)$$

When removing the *T x RH* and RH^2 terms the equation for the same two sites is computed as:

$$CR = 1.2082 + 0.0010T - 0.0003T^2 - 0.0021RH\ (N = 595, R^2 = 0.9829)$$

3.2.3.6 Conclusions

We observed that each reference dataset, R2G and R2P, as well as their difference are quite sensitive to thresholds and, likely, the heating of the rim. The difference is reduced by increasing the threshold for the reference data derived from the two references, as long as both gauges are heated.

The analysis of R2G/R2P catch ratios has shown that with an increase in the precipitation intensity threshold (P0), the scatter of the differences between concurrent values of the two reference systems using either of the two gauges, decreases significantly. This is most likely due to different levels of noise in each of the systems and at each of the sites which become less significant as P0 was increased.

For snow and for all precipitation types, transfer functions dependent on *T*, *RH* were developed using linear regression analysis. These can be used to further adjust the results of similar instruments tested on different sites against field references type R2, using different weighing gauges. Note that these transfer functions were not used for the SPICE SUT intercomparisons.

Although there are noted differences between the catch of the R2P and R2G systems, partly due to different gauge conception (see Table 3.2 for a comparison of each gauge characteristics), the overall differences are generally small for all precipitation types, including snow. This justifies the use of either configuration as the DFAR for SPICE intercomparisons.

3.2.4 Methodology for using an unshielded and single-Alter shielded weighing gauge (R3) to estimate true snowfall amounts

Authors: Roy Rasmussen, Bruce Baker, John Kochendorfer, Bill Collins, Matteo Colli, Luca Lanza, Julie Theriault

3.2.4.1 Background

A methodology is presented to estimate the true snowfall at a measurement site in which only single-Alter-shielded and unshielded gauges are present. This would allow sites unable to install a DFIR-shielded gauge measurement system with a means to estimate the true snowfall amount at their site.

3.2.4.2 Assumptions and approach

The key assumption behind the two gauge-shield configuration references is that the transfer function of an unshielded gauge is sufficiently different than that of an Alter-shielded gauge. This difference can be used to determine the appropriate transfer function to a DFAR for each site. Initially, it was thought that the difference of the slope of the transfer function from an Alter-shielded gauge compared to that of an unshielded gauge (each relative to a DFAR) was sufficient to characterize the various sites, but extensive analysis (not included here) showed this not to be the case. It was shown that assuming a linear transfer function for both the shielded and unshielded gauge produces a difference function that still depends on the wind speed, making it difficult to calibrate the system of equations.

Another approach is to assume that both transfer functions are exponential following the approach of Hamon (1973). In this case, a difference equation can be derived that does not depend on wind

speed, providing an elegant solution to the problem. However, analysis of single-Alter-shielded and unshielded gauge data reveals that while the unshielded gauge transfer function can be estimated by an exponential function, the Alter-shielded gauge function usually requires a polynomial or other type of equation.

For instance, Figure 3.53a provides an example of a transfer function for an unshielded Geonor T-200B3 gauge at the NCAR Marshall site, while Figure 3.53b shows a plot of the catch efficiency data for a single-Alter-shielded Geonor T-200B3 gauge from the same site during the same 4 year time period. The accumulation period for each datapoint is 30 minutes, data are accumulated over 0.5 m/s wind-speed bins, and only snow conditions are considered. Note that the unshielded gauge catch efficiency (CE) as a function of wind speed is well approximated by an exponential function (concave down curve, CE=exp(-d*wspd)), while the single-Alter CE curve has a concave upwards shape, which is well approximated by either a two term polynomial or an exponential growing curve (CE = a+b*exp(c*wspd)); in these equations, wspd denotes the wind speed and a, b, c, and d are coefficients. The different shapes of the unshielded gauge and the single-Alter gauge curvesare supported by Lagrangian particle modeling calculations of the catch efficiency of single-Alter-shielded and unshielded gauges using flow fields generated by Computation Fluid Dynamics (CFD) solutions (Colli 2014 and Theriault et al., 2012). Both the exponential shape of the unshielded gauge curve and concave down shape of the shielded gauge catch efficiency curve are re-produced (Figure 3.54). The ratio of the unshielded gauge to the single Alter (UN/SA) shows a parabolic shape with a clear minimum at a mid-range of velocity (Figure 3.54).

The Hamon (1973) approach to analytically determine the DFIR reference precipitation using information from only single-Alter-shielded and unshielded gauges was attempted, but the resulting equations were complex and the methodology to determine the needed coefficients difficult. This is due to the different form of the CE equation for the unshielded gauge (exponential) as compared to the single-Alter-shielded gauge (polynomial or exponential growing). The Hamon (1973) approach assumed that the CE equations for the unshielded and single-Alter-shielded gauges were both exponential. The current results from the Marshall site and the CFD modeling studies show that the CE curves are different, making it difficult to apply this approach.

Instead, it is recommended to adopt a more practical approach in which users calculate the ratio of unshielded to single-Alter-shielded precipitation accumulation data as a function of wind speed and then compare the curve shape and magnitude to similar curves calculated at R2 SPICE sites with a DFAR. It is recommended that the transfer function from the site with the best match to the UN/SA ratio curve be used for the R3 site. This takes advantage of the wide climatological differences at the various R2 sites in terms of temperature, wind speed, and snow conditions. This procedure is described in the following section.

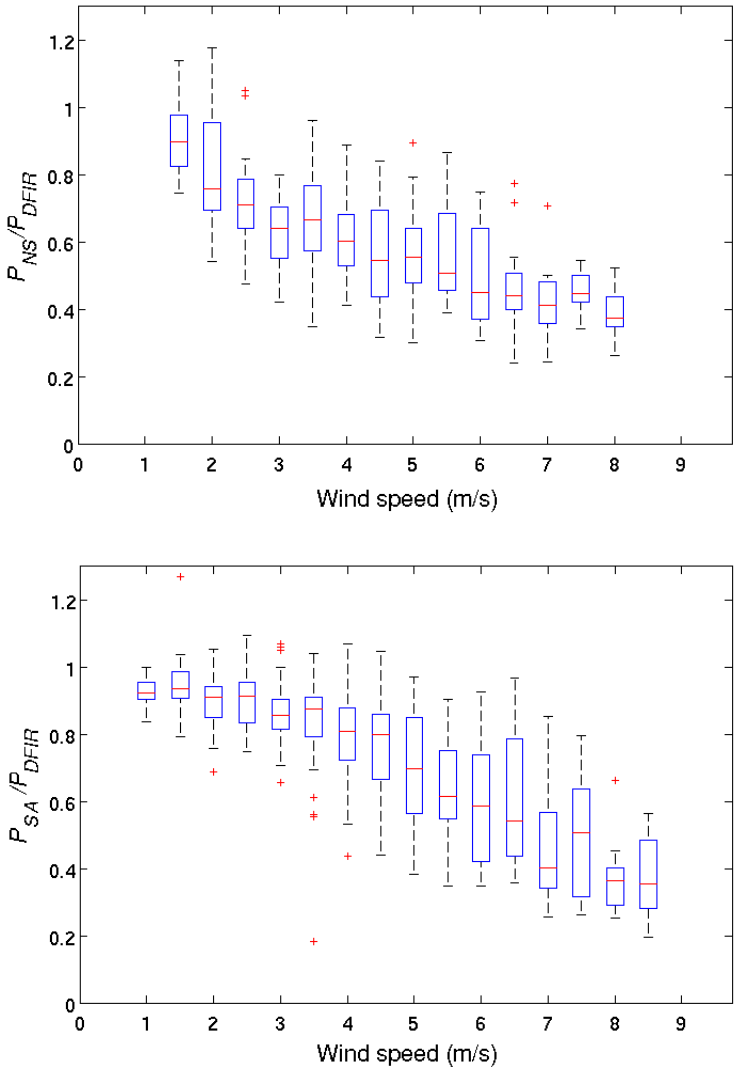


Figure 3.53. Mean wind speed dependence of catch efficiency for snow precipitation measurements made at the NCAR Marshall site by an unshielded Geonor gauge (PNS), relative to the DFAR (top) and a single-Alter shielded Geonor gauge (PSA) relative to the DFIR (bottom). The data are sampled with a 30-min period and averaged over 0.5 m/s wind-speed bins.

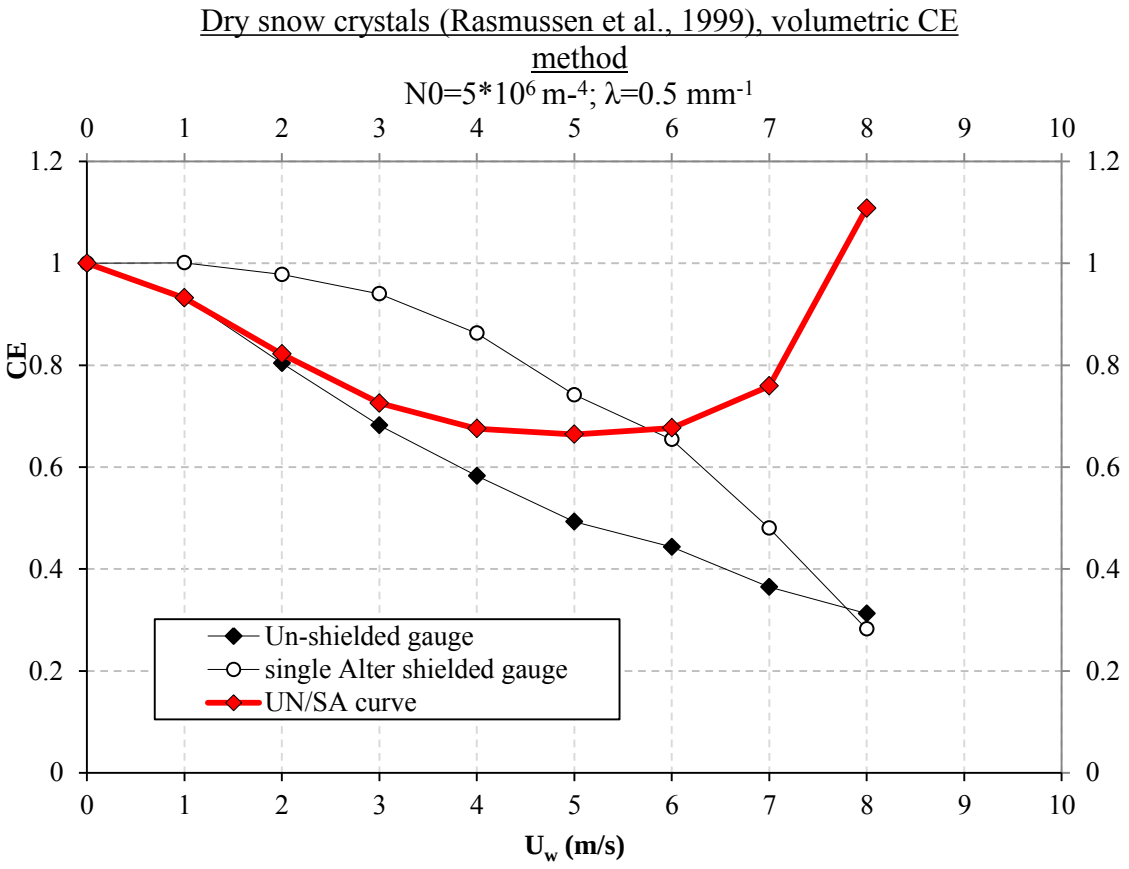


Figure 3.54. Model simulated catch efficiency for an unshielded Geonor (black diamond curve), single-Alter-shielded Geonor (open circle labeled curve) and the ratio of the unshielded CE to the Single-Alter-shielded gauge CE (red diamond curve). From Colli (2014).

3.2.4.3 Comparison to transfer functions at R2 SPICE sites

As part of their reference data collection, SPICE R2 sites operate at least one unshielded gauge and one single-Alter-shielded gauge of the same type, and at least one DFIR-shielded automated gauge of the same type (Geonor or Pluvio^2). As a result, data from these sites can be used to calculate CE curves for the unshielded gauge and single-Alter-shielded gauge (each relative to the DFIR-shielded gauge), as well as the ratio of the unshielded gauge accumulation to the single-Alter-shielded gauge accumulation as a function of wind speed.

An important variable that needs to be considered is environmental temperature, as it impacts the fall speed of the snowflakes. We therefore recommend that three separate plots be made, one for each shield configuration, for the following four temperature ranges:

- > +2 °C (rain)
- -2 to +2 °C (mixed precipitation including wet snow)
- -2 to – 4 °C (semi-wet snow)
- < -4 °C (snow)

Figure 3.55 provide examples of the CE curves for an unshielded Geonor and single-Alter-shielded Geonor (each relative to the DFIR-shielded gauge) for each of the above temperature ranges, as well as the catch efficiency curve determined from the ratio of the unshielded Geonor gauge

accumulation to the single-Alter-shielded Geonor gauge accumulation, using data from the Marshall site. The CE curves are close to 1.0 for both the unshielded and single-Alter-shielded gauges for temperatures above +2 °C (Figure 3.55), but significantly different for the colder than -2 °C curves (Figure 3.57 andFigure 3.58). These differences are attributed to the enhanced wind effects for lower-density snowflakes (relative to rain) at temperatures below -2 °C. The -2 to +2 °C curves (Figure 3.56) show CE curves in between the > +2 °C and < -2 °C curves. The CE curves for the unshielded and single-Alter-shielded gauges show increasing downward concaveness as the conditions are more conducive to snow (< -2 °C).

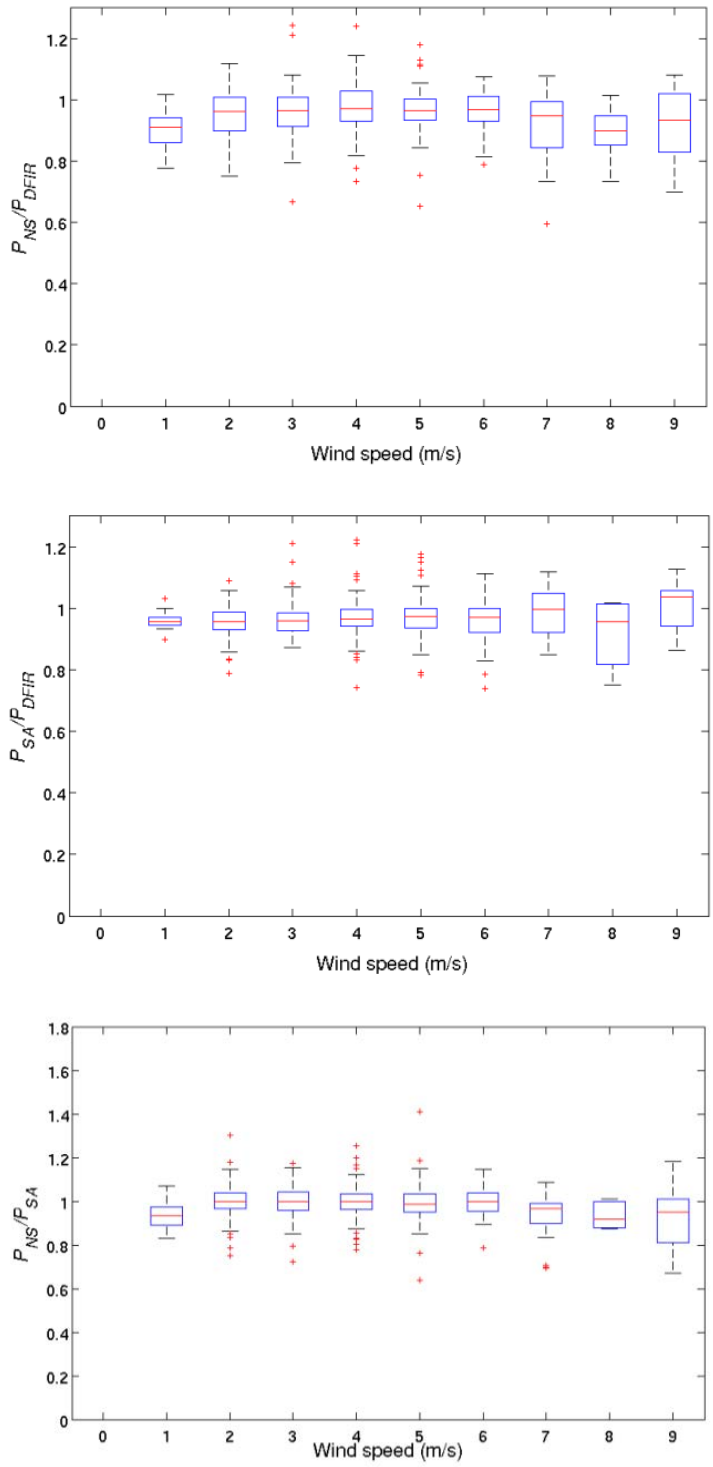


Figure 3.55. Catch efficiency for precipitation measurements made at the NCAR Marshall site in the temperature range > +2 °C for: an unshielded Geonor gauge relative to DFAR (top), a single-Alter-shielded Geonor relative to DFAR (middle), and an unshielded Geonor relative to a single-Alter-shielded Geonor (bottom). The data are sampled with a 30-min period and averaged over 1.0 m/s wind-speed bins.

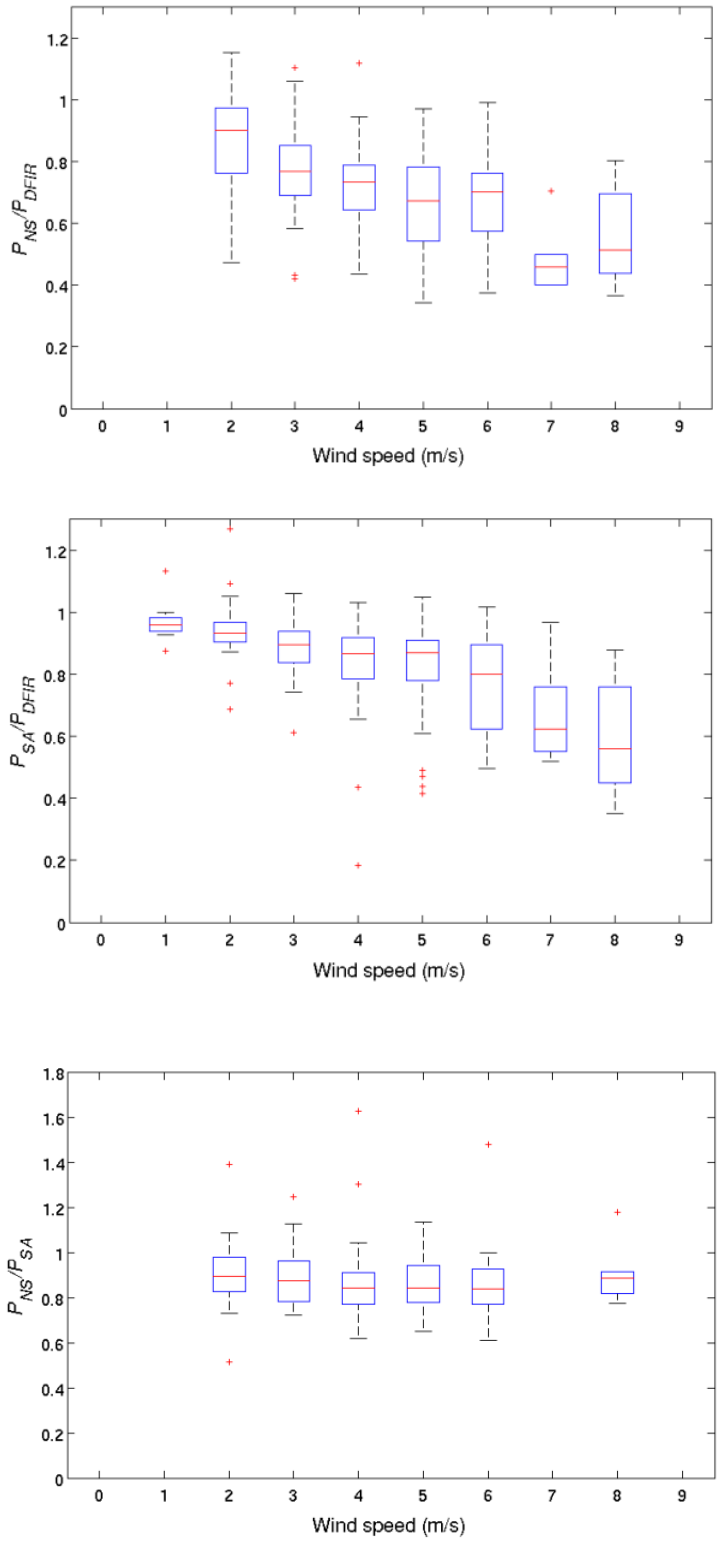


Figure 3.56. Catch efficiency for precipitation measurements made at the NCAR Marshall site in the temperature range -2 to +2 °C for: an unshielded Geonor gauge relative to DFAR (top), a single-Alter-shielded Geonor gauge relative to DFAR (middle), and an unshielded Geonor relative to a single-Alter-shielded Geonor (bottom). The data are sampled with a 30-min period and averaged over 1.0 m/s wind-speed bins.

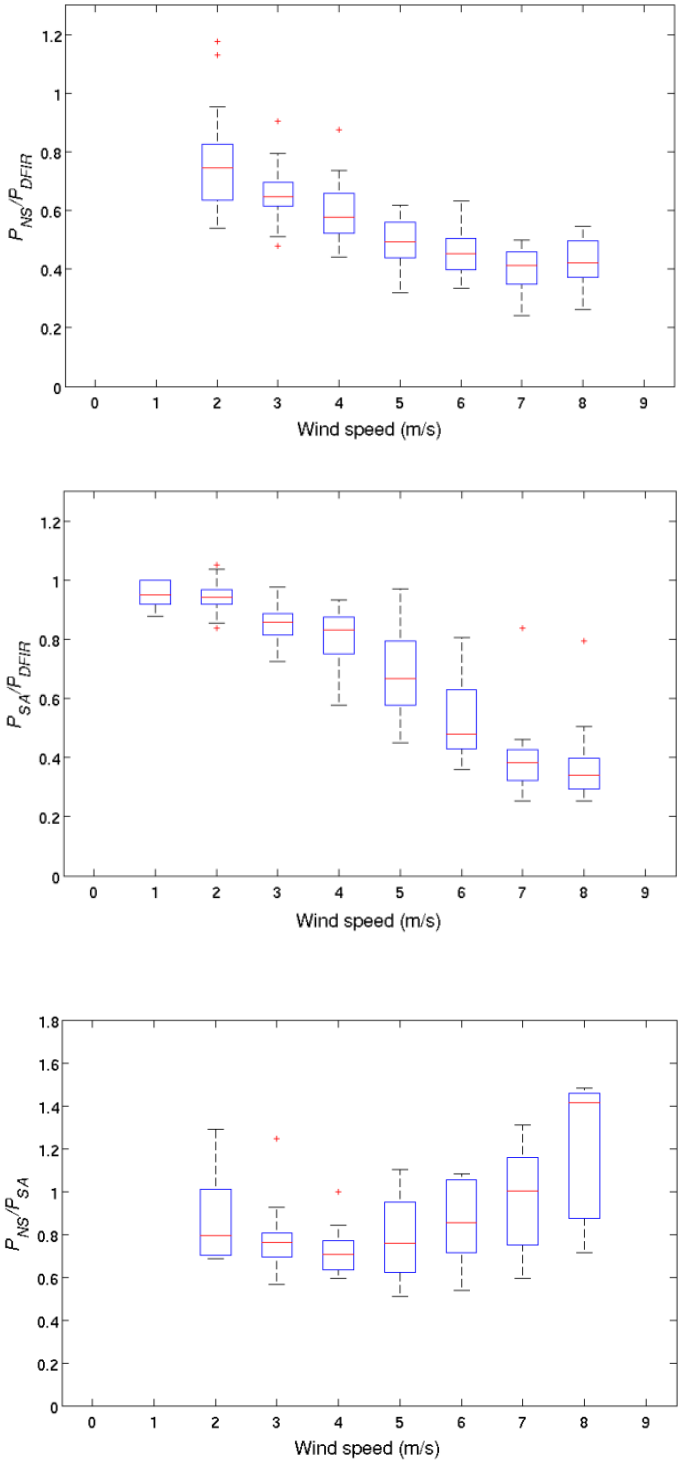


Figure 3.57. Catch efficiency for precipitation measurements made at the NCAR Marshall site in the temperature range -2 to -4 °C for: unshielded Geonor gauge relative to DFAR (top), single-Alter-shielded Geonor gauge relative to DFAR (middle), and unshielded Geonor relative to single-Alter-shielded Geonor (bottom). The data are sampled with a 30-min period and averaged over 1.0 m/s wind-speed bins.

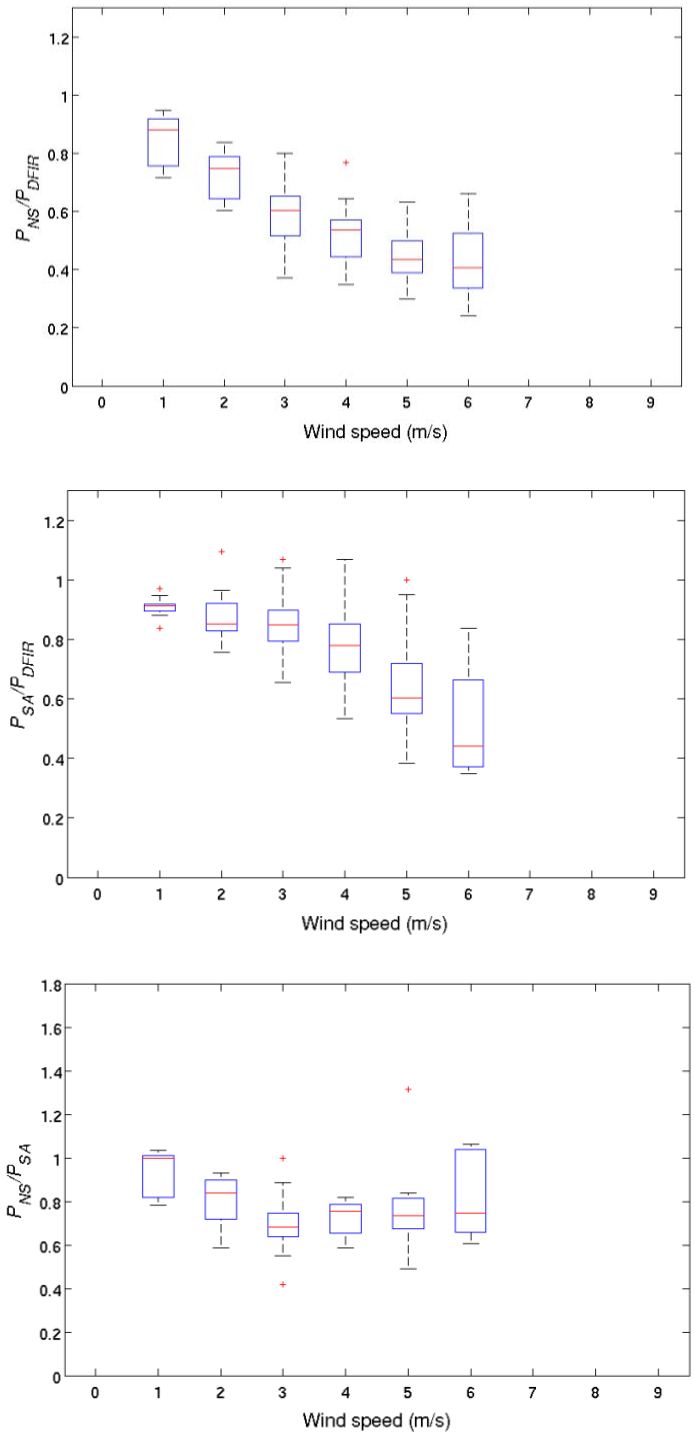


Figure 3.58. Catch efficiency for precipitation measurements made at the NCAR Marshall site in the temperature range < -4 °C for: unshielded Geonor gauge relative to DFAR (top), single-Alter-shielded Geonor relative to DFAR (middle), and unshielded Geonor relative to single-Alter-shielded Geonor (bottom). The data are sampled with a 30-min period and averaged over 1.0 m/s wind-speed bins.

3.2.4.4 Summary

A methodology has been described to estimate the true snowfall at an R3 measurement site in which only single-Alter-shielded and unshielded gauges are present. The procedure consists of comparing the ratio of the unshielded gauge accumulation to the single-Alter-shielded gauge accumulation at R2 sites as a function of wind speed and temperature to a similar quantity at R3 sites. The transfer functions at the R3 sites with the best comparison to the R2 site would be applied to the R3 site. This method was tested with SPICE data from the Marshall site only. Aditional work and testing with data from other sites is recommended.

3.2.5 Linking automated snow-depth measurements with manual snow stake observations: a methodology for uncertainty assessment

Authors: GyuWon Lee , Craig Smith

When used as a measurement reference, manual snow-depth measurements have significant limitations such as consistency, continuity, spatial and temporal resolution, and time and manpower consumption (Ryan and Doesken, 2007) that result in measurement uncertainty. Automated snow-depth sensors can be used to overcome some of these limitations, but also have their own limitations (Ryan et al, 2008; Fischer, 2011; de Haij, 2011) that can result in uncertainty. Using snow-depth data collected at the CARE site during SPICE, a methodology is demonstrated for analyzing the uncertainty of snow-depth measurements from automatic snow-depth sensors. The SPICE quality control (QC) procedures for snow-depth measurements (Section 3.3.3) are applied to the raw data sets before analysis.

The uncertainty analysis is performed using two approaches: (1) a collection of statistical measures; and (2) the propagation of error. The standard quantities for measuring the accuracy are defined under statistical measures. In the propagation of error, the uncertainty of individual instruments is calculated from the difference between two measurements of the same type. These approaches are outlined below and explained in further detail in the discussion paper by Lee et al. (2015).

3.2.5.1 Statistical measures

Standard statistical measures are used to quantify the uncertainty of snow-depth measurements. The Bias Error (BE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Bias Removed Root Mean Square Error (BRRMSE) are defined as follows:

$$BE = \frac{1}{N}\sum(y - x) \qquad (1)$$

$$MAE = \frac{1}{N}\sum|y - x| \qquad (2)$$

$$RMSE = \left[\frac{1}{N}\sum(y - x)^2\right]^{0.5} \qquad (3)$$

$$BRRMSE = \left[\frac{1}{N}\sum(y - x - BE)^2\right]^{0.5} \qquad (4)$$

where *x* and *y* are snow depths from pairs of manual measurements, manual and automatic measurements sharing the same snow target, or two instruments of the same type at different targets, and *N* is the number of datapoints for a given pair. The *NBE*, *NMAE*, *NRMSE*, and *NBRRMSE* are the normalized forms in which *BE*, *MAE*, *RMSE*, and *BRRMSE* are divided by the average of *x*.

In comparisons between manual observations, the *BE* is calculated to investigate the spatial distribution of snow depth relative to the average snow depth measured by snow stakes at each

target. The average snow depth from stakes at the same snow-depth target is considered as *x* in eq. (1) for the calculation of *BE* in comparisons between manual observations and automatic snow-depth sensors, which indicates the systematic bias of measurements from individual snow-depth sensors relative to the reference. *BRRMSE* (*NBRRMSE*) indicates the bias removed random error in snow-depth measurements.

3.2.5.2 Error propagation

The error propagation equation is used to quantify the uncertainty of manual snow-depth measurements and automatic snow-depth sensors. When *z = x1 – x2* is the difference between *x1* and *x2*, the variance of *z*, *var(z)* is expressed as follows:

$$var(z) = var(x1 - S) + var(x2 - S) - 2cov(x1 - S, x2 - S) \qquad (5)$$

where *x1* and *x2* are the snow depths from pairs of two manual measurements or two instruments of the same type, and *S* is true value of snow depth. The terms *var(xi - S)* represents the variance of the 'measurement – true' difference or square of uncertainty for *xi*, and the term *cov(x1 – S,x2 – S)* represents the covariance between x1 – S and x2 – S. Simply, we denote the uncertainty *var(xi – S)* by σ_{xi}^2. The random errors for two instruments of the same type, which have the same sampling volume and resolution, are nearly identical. Those for two manual measurements performed using the same procedure are also identical. Thus, the terms σ_{x1}^2 and σ_{x2}^2 are assumed to be identical when two manual measurements are compared and when two instruments of the same type are used. The covariance is set to be zero (*cov(x1 – S,x2 – S) = 0*) by assuming the random errors from the two measurements are not correlated. Thus, σ_{x1}^2 or σ_{x2}^2 can be calculated by:

$$\sigma_{x1}^2 = \sigma_{x2}^2 = var(z)/2 \qquad (6)$$

Even though two manual measurements are performed by the same procedure, and the two instruments are the same type, bias error can still exist in each case. Therefore, the variance of *z* in (6) can be also written as follows:

$$var(z) = \frac{1}{n}\sum_n z^2 - BE^2 \qquad (7)$$

By combining (6) and (7), the uncertainties of the σ_{x1} and σ_{x2} terms can be expressed as follows:

$$\sigma_{x1} = \sigma_{x2} = \sqrt{\frac{\frac{1}{n}\sum_n z^2 - BE^2}{2}} \qquad (8)$$

In general, the total uncertainty of *n* sensors of same types can be computed as

$$\sigma_{depth} = \left[\frac{2}{n(n-1)}\sum_{\{i=1\}}^{\{n-1\}}\sum_{\{j>1\}}^{n} var(xi - xj)\right]^{0.5} \quad (n \geq 2) \qquad (9)$$

Also, the individual uncertainty σ_{xi} of i^{th} sensor can be approximated by solving an *n(n-1)/2* by *n* overdetermined system. That is, for *n=4*, each uncertainty of i^{th} sensor can be obtained by solving the following linear equation:

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{x1}^2 \\ \sigma_{x2}^2 \\ \sigma_{x3}^2 \\ \sigma_{x4}^2 \end{bmatrix} = \begin{bmatrix} var(x1 - x2) \\ var(x1 - x3) \\ var(x1 - x4) \\ var(x2 - x3) \\ var(x2 - x4) \\ var(x3 - x4) \end{bmatrix} \qquad (10)$$

The uncertainties in manual observations are calculated from pairs of snow stakes using the equation (9). The average uncertainties for individual snow stakes, each base, and each snow-depth target are compared. The comparison among snow-depth sensors of the same type is performed to quantify the instrumental uncertainty of each sensor using equation (10).

3.2.5.3 Snow-depth measurements

The data used in this analysis were obtained at the CARE SPICE site from 14 December 2013 to 7 April 2014 and from 1 December 2014 to 11 March 2015. The CARE site layout can be found in Annex 4. The prevailing wind direction at the site is west to east with open exposure. The site has a slight downwards slope from east to west.

The manual reference measurements for snow depth at this site are explained in greater detail in Section 3.1.4.3.1 of this report. The site has three instrument pedestals for measuring snow depth (12A, 20 and 11A). Each pedestal is surrounded by three snow-depth targets and each target has four graduated (0.5 cm) snow stakes at each corner that are observed daily. Each pedestal hosts three sonic snow-depth sensors and one optical (laser) snow-depth sensor. Each ultrasonic sensor has a separate target, and the laser sensor shares a target with a sonic sensor. The sensors collect snow-depth measurements every 30 seconds. The sonic sensors are the Campbell Scientific SR50A (hereafter, SR50A), the Felix SL300 (hereafter, FEL), and the Sommer USH-8 (hereafter, SOM). The optical sensor is the Jenoptik SHM30 (hereafter, JEN).

Figure 3.59 shows the time series of snow depth at each snow stake from the manual observations at each base (a-c), and the average snow depths from the four snow stakes at each target (d). The maximum snow depths recorded at bases 12A, 20, and 11A during the observation period were 40.0 cm, 42.5 cm, and 44.0 cm, respectively. The average snow depth from the four snow stakes at each target was calculated to investigate the spatial distribution of snow depth and compare with the automatic sensors at the same target. The variation in manual snow-depth measurements between the four corners of a target is attributed to the uneven deposition of snow on the surface of a target. The manual snow-depth data are also used to analyze the uncertainty of manual snow-depth measurements.

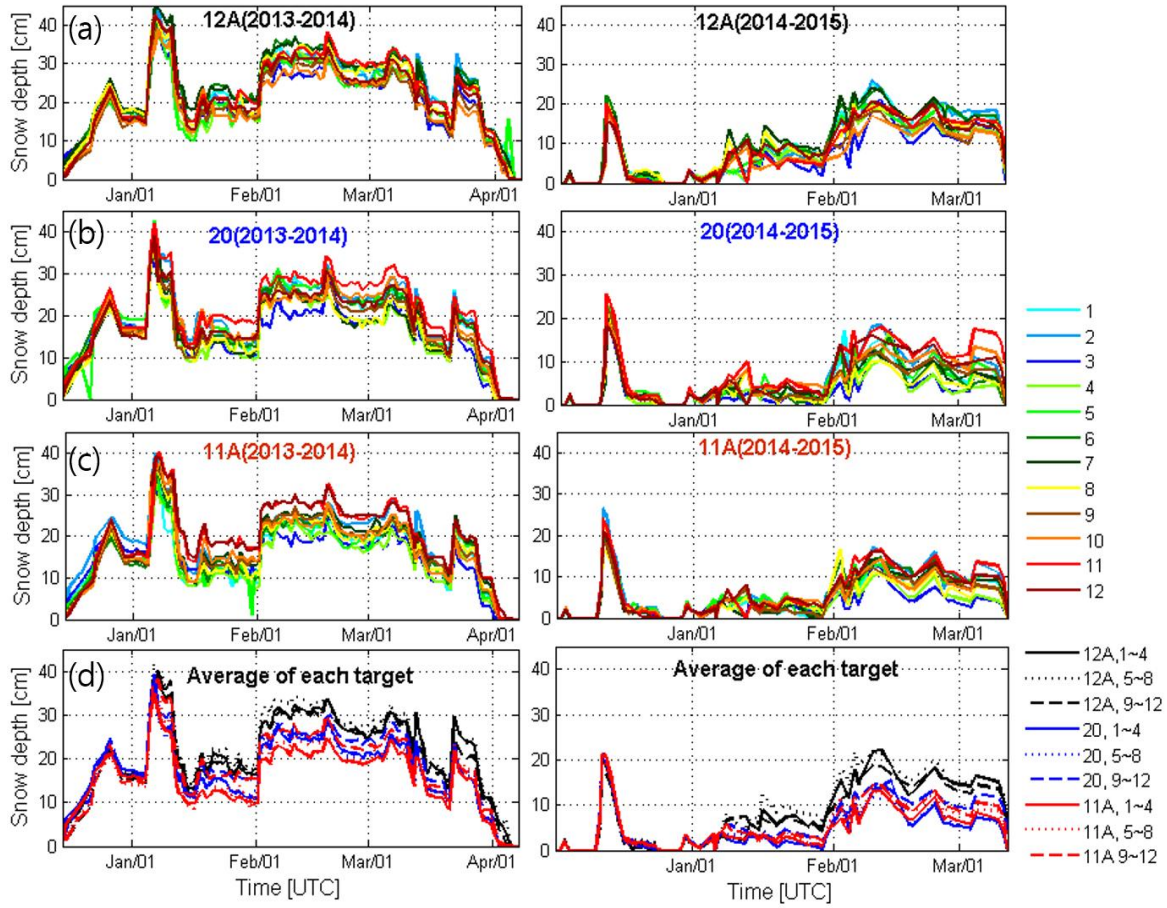


Figure 3.59. Time series of snow depth from snow stakes on base (a) 12A, (b) 20, and (c) 11A from manual observations over the period from 14 December 2013 to 7 April 2014 (left), and 1 December 2014 to 11 March 2015 (right). The line colors indicate individual snow stakes. (d) Average snow depth of four snow stakes on same snow-depth target. The line color indicates each base. The solid, dotted, and dashed lines represent average snow depth of stake numbers 1 ~ 4, 5 ~ 8 and 9 ~ 12 on the same snow-depth target.

3.2.5.4 Uncertainty of manual snow stake measurements

The BEs and uncertainties of manual snow stake measurements are calculated to analyze the spatial distribution of snow depth and uncertainty of manual snow-depth measurements (Figure 3.60). For the calculation of BEs, the average snow depth of snow stakes 1 to 4 on base 12A is considered to be the reference for the purpose of this analysis. Figure 3.60a shows that the BEs of base 12A (0.00 cm, 1.32 cm, and -0.36 cm) are the smallest, which is to be expected, given the selection of the reference for this analysis. Relative to the reference selected, the BEs of base 11A (-4.61 cm, -4.07 cm, and -2.58 cm) are the largest. From these results, it was concluded that the snow depth on base 12A is higher (east side of the experiment area) than on base 11A (west side on the experiment area). These results characterize the spatial distribution of snow depth across the experiment area, as reported by the human observer. These results also emphasize the necessity of several manual observations within the experimental site.

The uncertainties (σ_{depth}) of all snow stake pairs are shown in Figure 3.60b. The total uncertainty for all 630 pairs of stakes is 2.18 cm, for this particular configuration and measurement resolution. The

uncertainty for base 11A (1.72 cm) is the largest, while that for bases 12A (1.67 cm) and 20 (1.67 cm) are the same. The average uncertainties for each base (1.69 cm) are greater than that of each snow-depth target (1.52 cm). The uncertainty gradually increases from target (1.52 cm) to base (1.69 cm) to all pairs of stakes on the site (2.18 cm). This is due to the temporal variation of snow depth during manual observations, which was not taken into account by the long-term BE removal. Thus, the lower bound of uncertainty for manual snow-depth measurements is estimated to be in the range of 1.52 cm to 2.18 cm.

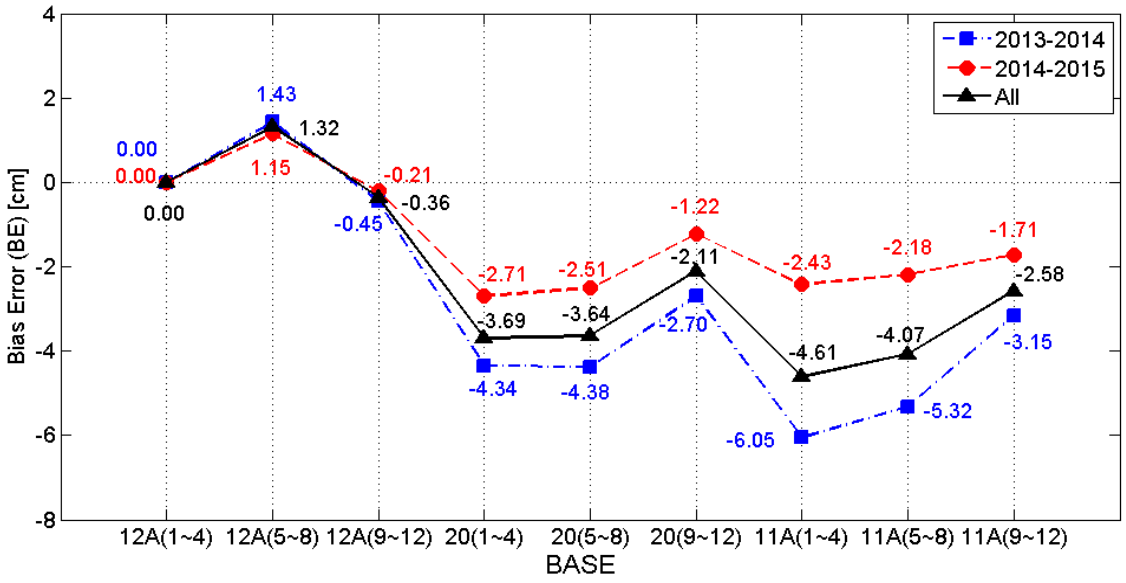
3.2.5.5 Uncertainty of snow-depth sensors

The snow depth measured by each automatic sensor was compared with the average of the manual observations at the same target, which is considered to be the reference. Figure 3.61 shows the BEs and BRRMSEs of each snow-depth sensor. The positive BEs indicate that all automatic snow-depth sensors, on all bases, measure snow depths that exceed the manual observations by 1.61 to 2.74 cm (Figure 3.61a). The BRRMSEs of snow-depth sensors on base 12A (2.58 cm, 2.74 cm, 1.93 cm, and 1.61 cm) are the largest and the those of snow-depth sensors on base 11A (1.71 cm, 2.29 cm, 1.70 cm, and 1.92 cm) are the smallest, based on the comparison among bases in Figure 3.64b. The average BEs (NBEs) of snow-depth sensors of the same type are FEL = 2.41 cm (17.6%), JEN = 1.74 cm (13.7%), SR50A = 2.14 cm (14.3%), SOM = 1.84 cm (13.9%). The average BRRMSEs (NBRRMSEs) of snow-depth sensors of the same type are shown in Figure 3.64c, and have the following values: FEL = 1.70 cm (12.6%), JEN = 1.50 cm (11.3%), SR50A = 1.64 cm (11.17%), SOM = 1.57 cm (12.6%). Given the spatial variability in snow depth implied by the base-to-base variability in bias and random errors outlined above, the differences in random errors among the different sensor types are not considered to be significant.

In general, the BE ranges from 10.5% (SOM on 12A) to 20.1% (FEL on 11A) and the random error ranges from 8.5% (JEM on 12A) to 16.7% (SOM on 11A). Again, the BEs of automatic sensors all have positive values, indicating that the automatic sensors overestimate snow depth relative to the manual measurement values. The BRRMSE values are within 3 cm; however, the patterns or NBRRMSE are not exactly same as the ones of BRRMSE, because the average snow depth measured by snow stakes at each target is different.

The snow depths measured by two sensors of the same type on different bases were compared to quantify the instrumental uncertainty of individual snow-depth sensors. It is important to note that the data quality during snow events could be poor for ultrasonic sensors, since it is a known limitation of these sensors that the sound waves are returned by the falling snow before reaching the target. This may have an impact on the calculated uncertainty. A significant bias is shown in the comparison, and should be eliminated to quantify instrumental uncertainty.

(a)



(b)

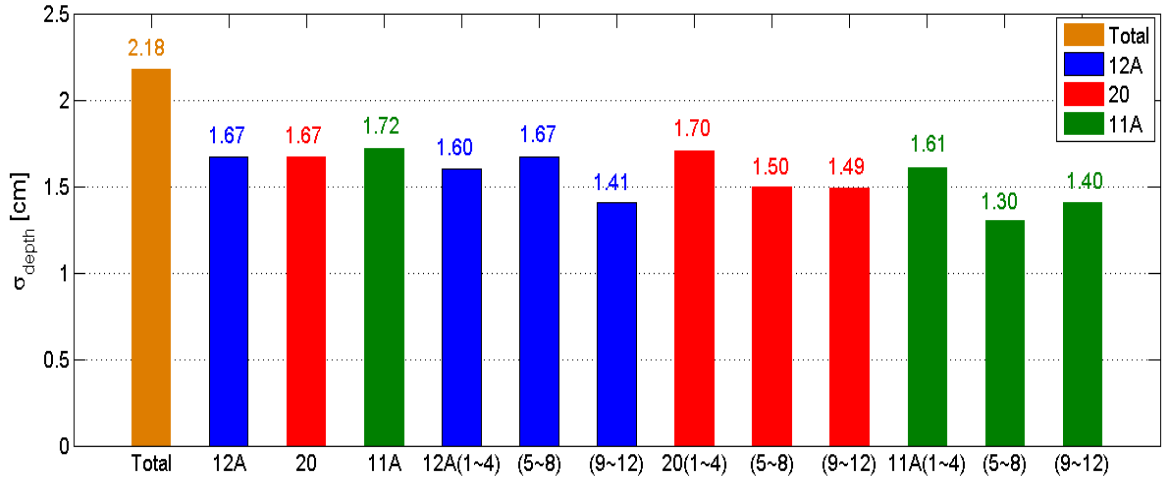


Figure 3.60. (a) BEs and (b) uncertainties of manual snow-depth measurements. The BEs are calculated for each snow-depth target. The orange bar represents the σ_{depth} for all pairs. The 2nd ~ 4th (5th ~ 13th) columns indicate the σ_{depth} for each base (snow-depth target). The color of bars indicates the same base. The blue, red, and green bars represent σ_{depth} for base 12A, 20, and 11A.

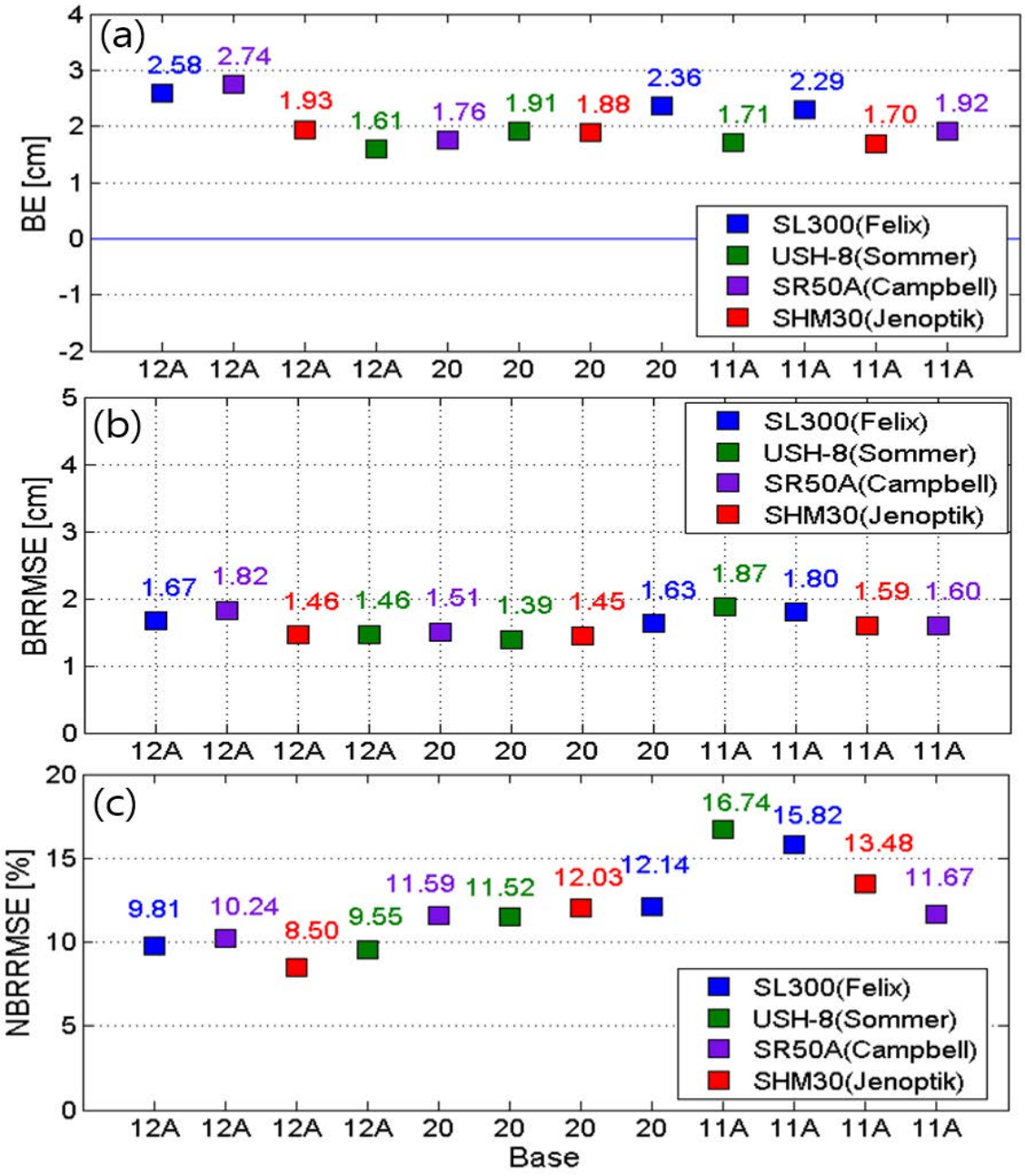


Figure 3.61. (a) BEs, (b) BRRMSEs, and (c) NBRRMSE for each snow-depth sensor.

The BEs and instrumental uncertainties of all snow-depth sensors are shown in Figure 3.62. The snow-depth sensors on base 12A are considered to be the reference for the calculation of BE (squares in Figure 3.62a), similar to the approach used for the assessment of manual observations. For the uncertainty of each sensor, the 3 by 3 system is solved, using 3 pairs of same type sensors using equation (10) for n=3. The circles in Figure 3.62a represent the spatial distribution of snow depth measured by the automatic sensors. To calculate these values, the BEs in Figure 3.60a and Figure 3.61a are added and the snow depths from sensors on base 12A are used as the reference. The BEs of snow-depth sensors on base 20 and 11A are negative. This could result from the spatial distribution of snow depth, and/or the systematic bias of snow-depth sensors. Also, the snow depths measured at bases 20 and 11A are lower than that of base 12A, based on the results from the

comparison of manual observations (Figure 3.60a). In general, all snow-depth sensors overestimate snow depth relative to the manual observations (Figure 3.61a). Thus, the larger measurements of snow-depth sensors on base 12A than those of bases 20 and 11A are parallel results with the measurement of manual observations.

When comparing each base, the instrumental uncertainties of each snow-depth sensor on base 12A (2.55 to 3.54 cm) are the largest (Figure 3.62b). The instrumental uncertainty of SR50A on 11A (0.96 cm) is the smallest in the comparison among each snow-depth sensor type. The average instrumental uncertainties of snow-depth sensors of the same type are calculated as follows: SOM = 2.29 cm, SR50A = 2.11 cm, JEN = 2.05 cm, FEL = 1.76 cm.

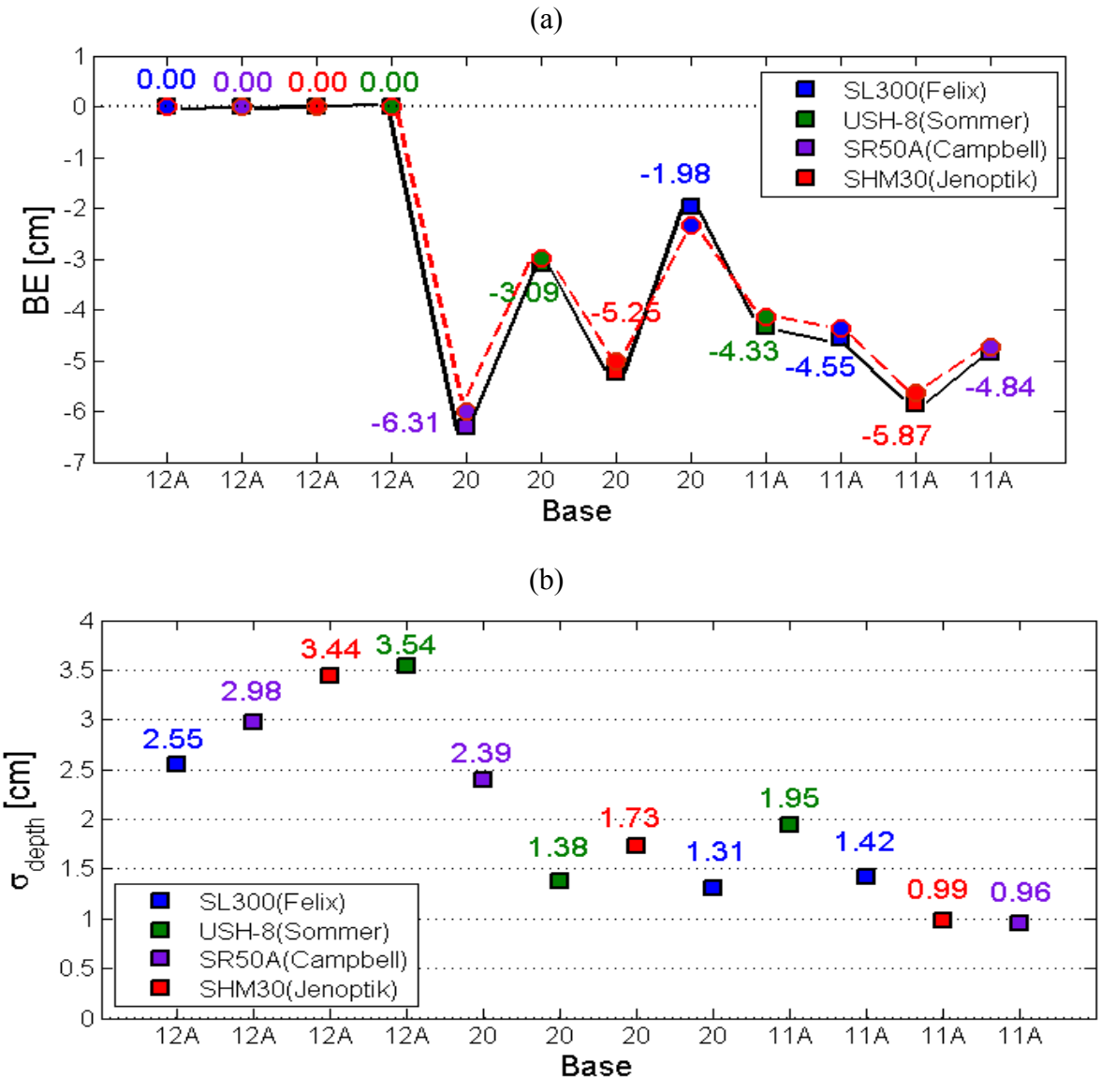


Figure 3.62. (a) BEs of each snow-depth sensor. The squares (circles) are calculated by considering the snow-depth sensors on base 12A as reference (the BEs are added and snow depth from snow-depth sensor on base 12A are then used as reference). (b) σ_{depth} of each snow-depth sensor. The blue, purple, red, and green diamonds indicate FEL, SR50A, JEN, and SOM.

3.2.5.6 Conclusions

This assessment demonstrates a method for assessing the uncertainty in manual and automated measurements of snow depth using statistical techniques and error propagation analysis. The BEs of manual snow-depth measurements provide information on the spatial distribution of snow over an area, and the comparison between manual and automatic snow-depth measurements provide information about the systematic bias of each snow-depth sensor at the CARE site. The uncertainties of manual observations for all pairs, on each base, and for each snow target were 2.18 cm, 1.69 cm, and 1.52 cm, for targets 12A, 20, and 11A respectively. The BEs of snow-depth sensors on base 12A, 20, and 11A ranged from 1.61 to 2.74 cm (10.5 to 20.1%) in comparison with manual observations at the same snow target. The average instrumental uncertainty was SOM = 2.29 cm, SR50A =2.11 cm, JEN = 2.05 cm, and FEL = 1.76 cm.

The uncertainty of measurements can vary among similar instruments collocated on the same site. The variability of results obtained through this study may indicate that other additional factors could influence the uncertainty of measurement of any sensor. The identification and treatment of such factors could improve further the uncertainty of measurement, and warrants further investigation. Two categories of factors are recognized to influence the uncertainty of measurements that would require further investigation. The first is related to the site and sensor configuration, while the second is specific to a sensors ability to detect and measure snow on the ground. The differences in the uncertainty of measurements for similar sensors would include the differences in the accumulation due to topography, wind influence, etc. Additionally, the accuracy of the measurement of the initial distance between the sensing element and the ground is critical, as is the ability to maintain this distance throughout operations (e.g. by taking steps to mitigate changes in the target area due to frost heave, or changes in the sample area due to small variations in the sensor mounting or orientation).

3.3 Derivation of SPICE datasets

Authors: Audrey Reverdin, Michael Earle, Mareile Wolff, Eckhard Lanzinger, Craig Smith

The derivation of analysis-ready precipitation and ancillary datasets entailed consecutive quality control (QC) and precipitation-event selection (ES) procedures. The process for precipitation was separate from those developed for, and applied to, snow-on-the-ground data (Section 0). The intention was to apply a consistent approach across all site precipitation and ancillary datasets to ensure the comparability of results required to address SPICE analysis objectives. Different levels of data quality control and processing correspond to different levels of SPICE data products, as outlined in Section 3.3.1. The data quality control procedures for precipitation and SoG both involve a series of automated checks and filters, and are described in Sections 3.3.2 and 3.3.3, respectively. All data were subsequently aggregated to 1-min temporal resolution and subjected to a final manual assessment. The resulting precipitation data were then input into an event-selection procedure (Section 3.4) to identify precipitating periods with a high degree of confidence and generate integrated precipitation and ancillary datasets for each site (the Site Event DataSets, or SEDS).

3.3.1 Data levels for SPICE

A system of data levels has been established to distinguish among datasets at different stages of processing and quality control. This system is built upon the existing framework used for satellite observations by the WMO and other organizations (World Meteorological Organization, 2017; National Aeronautics and Space Administration, 2010).

Level 0: The rawest output from an instrument or instrument transducer in native units (e.g. voltage).

Level 1: The time-stamped output from each individual instrument or instrument and data logger that has been converted into geophysical measurements (e.g. weight, mass, intensity). These data are generally recorded at the highest temporal resolution feasible for a particular instrument configuration at a particular site and before any significant data quality control has been applied. These data were recorded and stored at each measurement site and transferred to the SPICE data archive at the National Centre for Atmospheric Research (NCAR), Boulder, CO, USA.

Level 2: Quality-controlled datasets for one instrument on one site.

> **Level 2a:** Level 1 data that have undergone both formatting and integrity checks to ensure the correct number of records per day (e.g. 1,440 records/day for data with 1-minute sampling intervals) and the validity of field formats within a given record (e.g. number, text string). These checks are performed automatically when data are ingested into the SPICE archive.
>
> **Level 2b:** Level 2a data after a quality control procedure has been applied. The details of the procedure may vary by sensor and, in some cases, by site (due to differences in sampling, configuration, site conditions, etc.), and have been developed through consultation with site managers. Basic data-quality flags are added. For data with sampling intervals less than 1 minute, the output data and flags are aggregated to produce 1-minute values. Level 2b data are generated and made available for download at the SPICE archive. Details of the quality-control procedures and flag criteria are provided in Sections 3.3.2.3 and 3.3.2.4, respectively.

Level 3: Usage-relevant datasets derived from Level 2 data for single sensors and parameters. Processing is application dependent and may include aggregation to different temporal scales. For example, weighing gauge and ancillary data that have been aggregated to precipitation-event timescales (e.g. 30 minutes, 1 hour, or longer) are Level 3 products.

Level 4: Integrated datasets derived using lower-level datasets for multiple sensors and parameters. Weighing gauge data (Level 2) are used in concert with data from a precipitation detector or disdrometer (Level 2) to identify and characterize precipitating periods over which ancillary and other precipitation-sensor data are also extracted. The resulting site event datasets are a Level 4 product and comprise 30-minute (or longer) precipitation and ancillary data from all instruments at a given site. The SEDS are described in detail in Section 3.4.

3.3.2 Data quality control of the precipitation data

Observations collected from the field working reference systems, from the sensors under test and from instruments providing ancillary measurements have been processed in a well-characterized and consistent manner to establish a common basis for assessment and enable the intercomparison analysis. This section outlines the quality-control processing approach employed for the derivation of the SPICE precipitation datasets up to Level 2b, which is depicted in Figure 3.63. . This approach includes: (1) a series of automated checks and filters (Section 3.3.2.1); (2) the aggregation of sensor data to a temporal resolution of 1 minute (Section 3.3.2.2); and (3) manual processing to adjust for any data issues not addressed by the automatic procedure, such as the emptying and recharging of weighing gauges (Section 3.3.2.3). A system of data flags has been developed to convey details regarding the processing applied to each 1-minute datapoint for consideration in subsequent

analyses. These flags are indicated in Figure 3.63. and relevant sections below, and are summarized in Section 3.3.2.4.

3.3.2.1 Automated checks and filters

As outlined in Figure 3.63. , site sensor data are subject to the following steps: (1) a file formatting and integrity check to ensure the uniformity of file formats for processing; (2) a range check or maximum/minimum value filter; (3) a jump and baseline shift filter; and (4) a noise filter. Descriptions of each step are provided in Sections 3.3.2.1.1 to 3.3.2.1.4. Considerations regarding temperature compensation of precipitation observations from weighing gauges are discussed in Section 3.3.2.1.5.

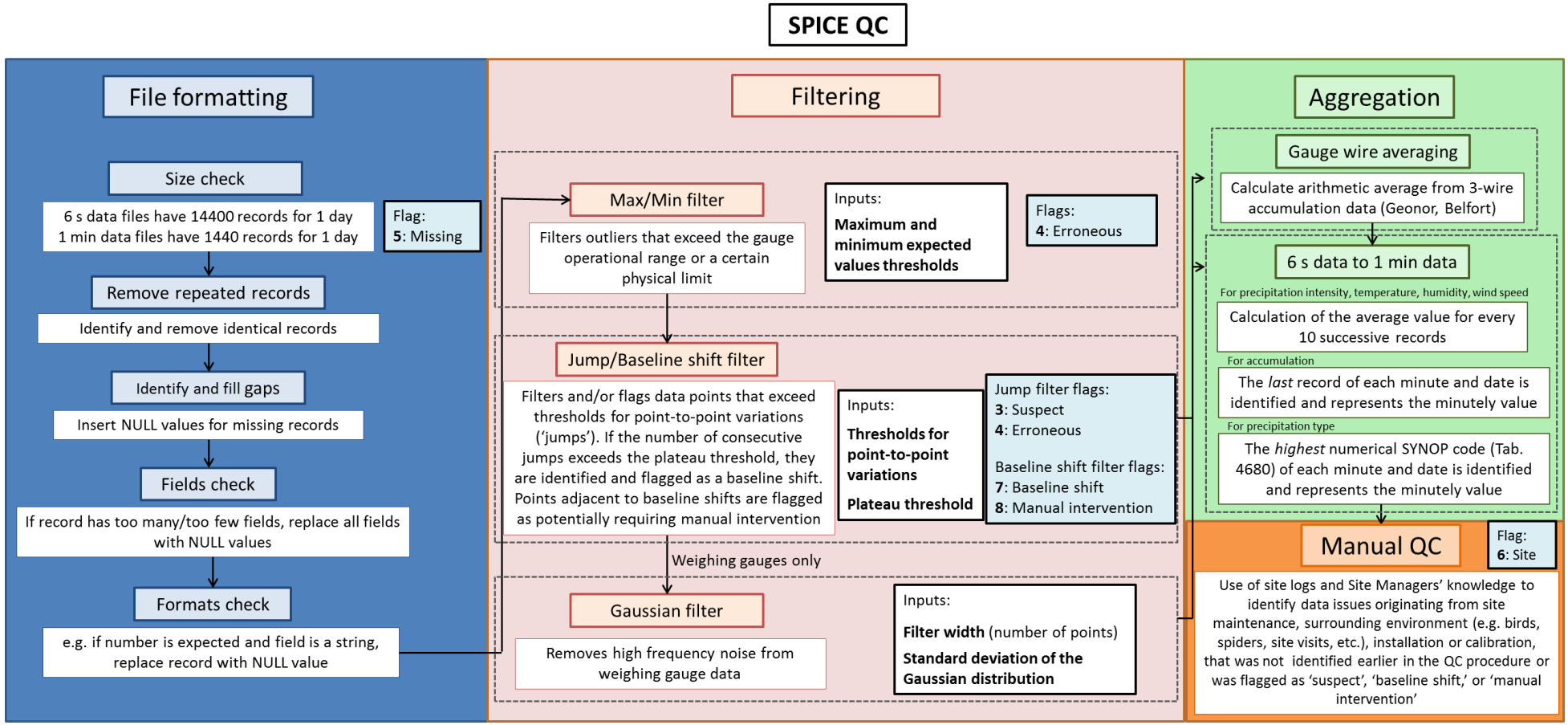


Figure 3.63. Flowchart of the SPICE quality-control procedure.

3.3.2.1.1 *File-formatting and integrity-check process*

Level 1 data files are checked to ensure that the correct number of records (time stamps) are present within a given time period and that each record contains the correct number of fields. Daily files for data with a 6-second sampling interval should have 14,400 records (10 records/minute x 60 minutes/hour x 24 hours), while those for data with a 1-minute sampling interval should have 1,440 records (1 record/minute x 60 minutes/hour x 24 hours). Any duplicate records (repeated time stamps with the same data) are removed. Any missing records are identified, the appropriate time stamps are inserted, and the corresponding fields are filled with null values (e.g. -999, NULL, or NaN). The missing records are tracked with the data presence flag (flag = 5; see Table 3.8), discussed in more detail in Section 3.3.2.4. In the event that a given record contains more or fewer fields/parameters than expected, the entire line is replaced with null values and flagged with the same data presence flag. If a given field does not match its expected format (e.g. text when a number is expected), all fields in the record are replaced with null values and again flagged. The resulting Level 2a datasets are used as inputs for the remaining steps in the QC procedure.

Table 3.8. SPICE QC data-flagging system.

Flag value	Data Classification	Data Characterization
1	Good	No issues detected
3	Suspect	Suspect performance threshold exceeded and data checked manually
4	Erroneous	Value(s) outside of gauge operational range as defined by max/min erroneous threshold and max erroneous variation from point to point. Data are removed and replaced with null values.
5	Missing	Missing datapoint
6	Site	Adapted from site logs; datapoints manually flagged to reflect maintenance, malfunction, power outage, etc. Data are removed and replaced with null values.
7	Baseline shift	Baseline shift present and data should be checked manually
8	Manual intervention	Data within specified proximity of baseline shift that should be checked manually (may be impacted by snow capping)

3.3.2.1.2 *Maximum/minimum value filter (range check)*

For each instrument parameter of interest for subsequent data analysis, a minimum and maximum expected value are defined according to the physical or mechanical constraints of the sensor or plausible/expected values of the parameter in the environment. For example, the minimum expected value for accumulated precipitation from a weighing gauge is 0 mm, and the maximum expected value corresponds to the bucket capacity. For ancillary measurements of temperature and wind direction, the minimum and maximum expected values are -50 °C and +50 °C, and 0° and 360°, respectively. If a given value lies above the maximum expected value or below the minimum

expected value, it is replaced with a null value and flagged as “erroneous” (flag = 4). (See Table 3.8) An example of maximum/minimum value filter application to weighing gauge precipitation data is provided in Figure 3.64.

3.3.2.1.3 *Jump and baseline shift filter*

Within a given dataset, “jumps” may be observed. These are intermittent deviations from the main data trend, or baseline, that fall within the maximum and minimum expected values and, hence, are not filtered out by the range check. A jump filter is employed to identify points that differ from the preceding baseline values by more than a set threshold and to flag them accordingly. In SPICE, separate thresholds are used to identify suspect and erroneous points, with the latter representing more significant point-to-point variations. Suspect jumps are flagged (flag = 3; see Table 3.8), while erroneous jumps are flagged (flag = 4) and replaced with null values. The suspect flags indicate that the corresponding data values are not necessarily erroneous, but warrant further attention during analysis. The specific values selected for the suspect and erroneous jump thresholds are meant to exceed the maximum expected increase of a given parameter per 6-second or 1-minute interval (as defined by instrument operational limits and/or site climatology). These thresholds have been defined for each instrument and parameter based on the instrument technical specifications from manufacturers, on thresholds already defined and used by national meteorological and hydrological services, and on input from SPICE site managers.

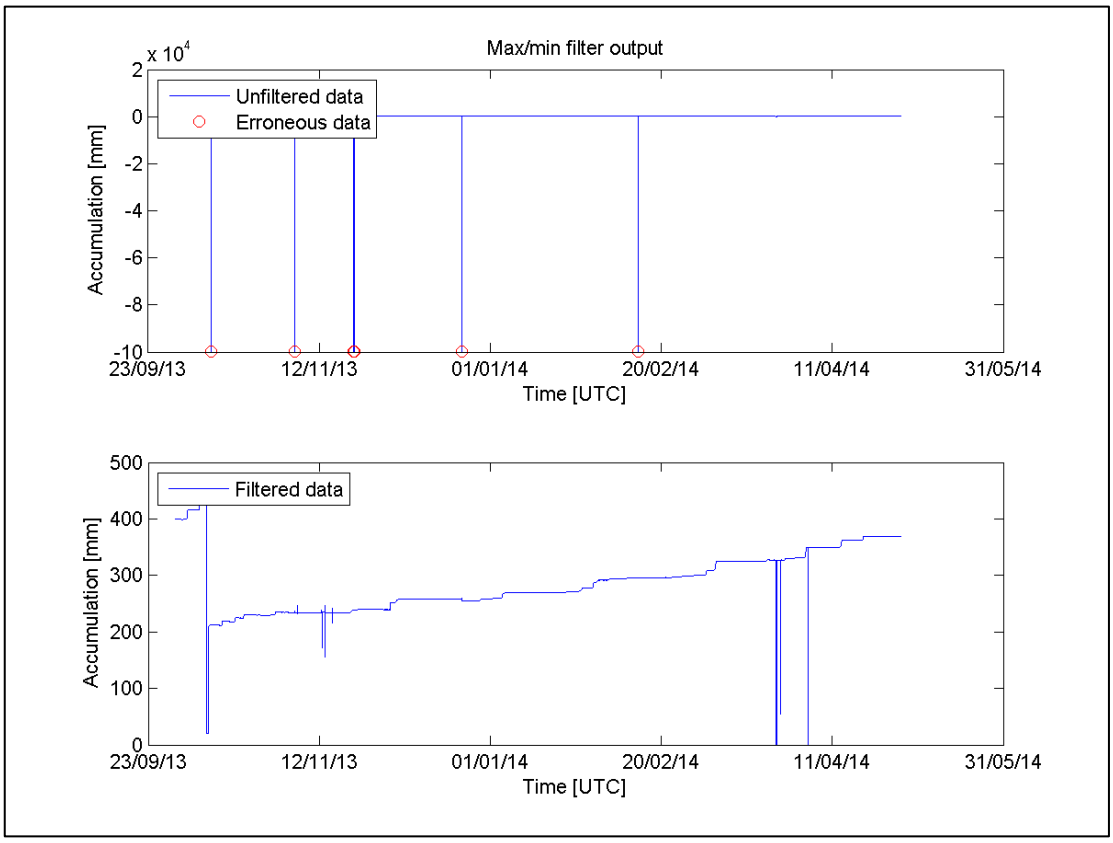


Figure 3.64. Example of the max/min value filter applied to accumulated precipitation data from a weighing gauge with minimum and maximum thresholds of 0 and 650 mm, respectively. The top panel shows the unfiltered data with erroneous data flags, while the bottom panel shows the resulting filtered data with all data flagged as erroneous by this filter being removed.

To better describe the jump and baseline shift filter process, Figure 3.65 illustrates an example of the jump filter applied to an artificial dataset. The red circles are identified as jumps relative to the last point falling within the jump threshold, the initial baseline (blue circle). When a datapoint falls back within the jump threshold relative to the initial baseline, it becomes the new baseline (green circle). As the red dots exceed the erroneous threshold in this case, the intervening points are replaced with null values in the resulting filtered dataset and flagged with number 4. In this particular example, the data values return to the initial baseline after a series of jumps. In other cases, however, jumps are not intermittent, but correspond to a change in the baseline. For weighing gauges, for example, increases in the baseline may be associated with "dumps," in which solid precipitation accumulated on the rim (a phenomenon referred to as "capping") falls into the bucket resulting in an abrupt and sometimes significant increase in accumulation (see Section 4.2.1). Decreases in the baseline may also occur and could be associated with the emptying of buckets as part of regular gauge maintenance.

To identify baseline shifts, the number of consecutive jumps is tracked. A new baseline, or plateau, is identified when the number of consecutive jumps exceeds a plateau threshold set to correspond to a specified time period. When a new plateau is identified, the associated data are not replaced with null values; rather the first point in the new plateau is flagged (flag = 7; see Table 3.8) to indicate that manual assessment is required before the period can be considered for subsequent analysis. Gauge capping and related baseline shifts may impact data before or after the shift is observed. For example, a gauge may have been capped for an extended period before observing a dump, or a gauge may remain partially capped following a dump. Also, in the case of gauge emptying, the baseline shift may be preceded or followed by a period of time during which the technicians were still working on the gauge as part of ongoing maintenance. Accordingly, the time periods preceding and following baseline shifts are also flagged (flag = 8) for manual assessment and possibly intervention. Figure 3.66a shows an example of the jump/baseline shift filter applied to an artificial accumulation dataset simulating a dump of snow falling into the bucket, while Figure 3.66b shows an example where the filter is applied to real data for a case in which a gauge bucket is emptied at the beginning of the time series. In both cases, the resulting filtered dataset still requires manual intervention to address the flagged periods, as more information (from site managers, site logs, webcam pictures, etc.) is needed to decide whether to keep the remaining flagged points.

Figure 3.64. and Figure 3.66b correspond to the same weighing gauge dataset and constitute the consecutive application of the two filters (filtered dataset from max/min filter being the input of the jump/baseline shift filter, as outlined in the QC flowchart in Figure 3.63.). As a consequence, the filtered dataset has been significantly enhanced by the two filtering steps.

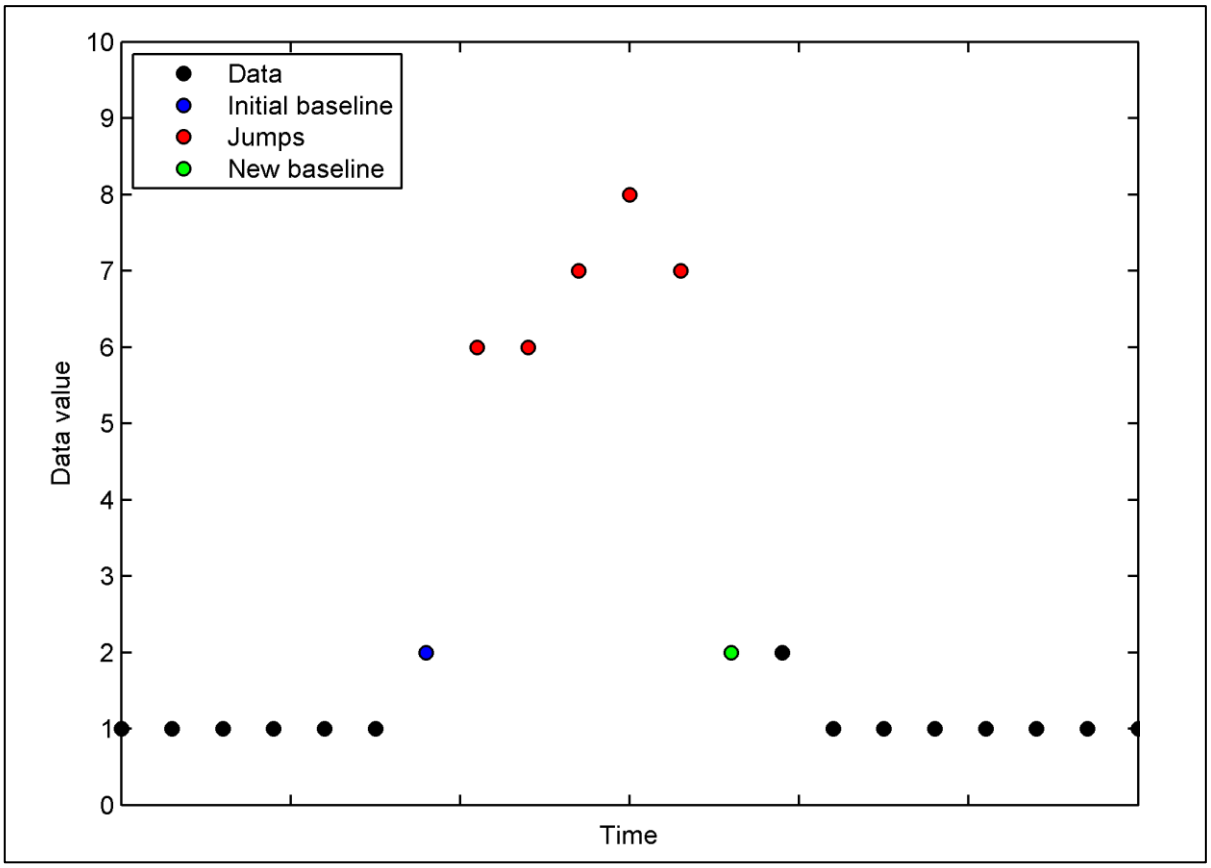


Figure 3.65. Example of jump filter application to a hypothetical dataset with suspect and erroneous jump thresholds of 2 mm/min and 3 mm/min, respectively. The flagged datapoints (red circles) are here determined to be erroneous and will therefore be removed from the dataset.

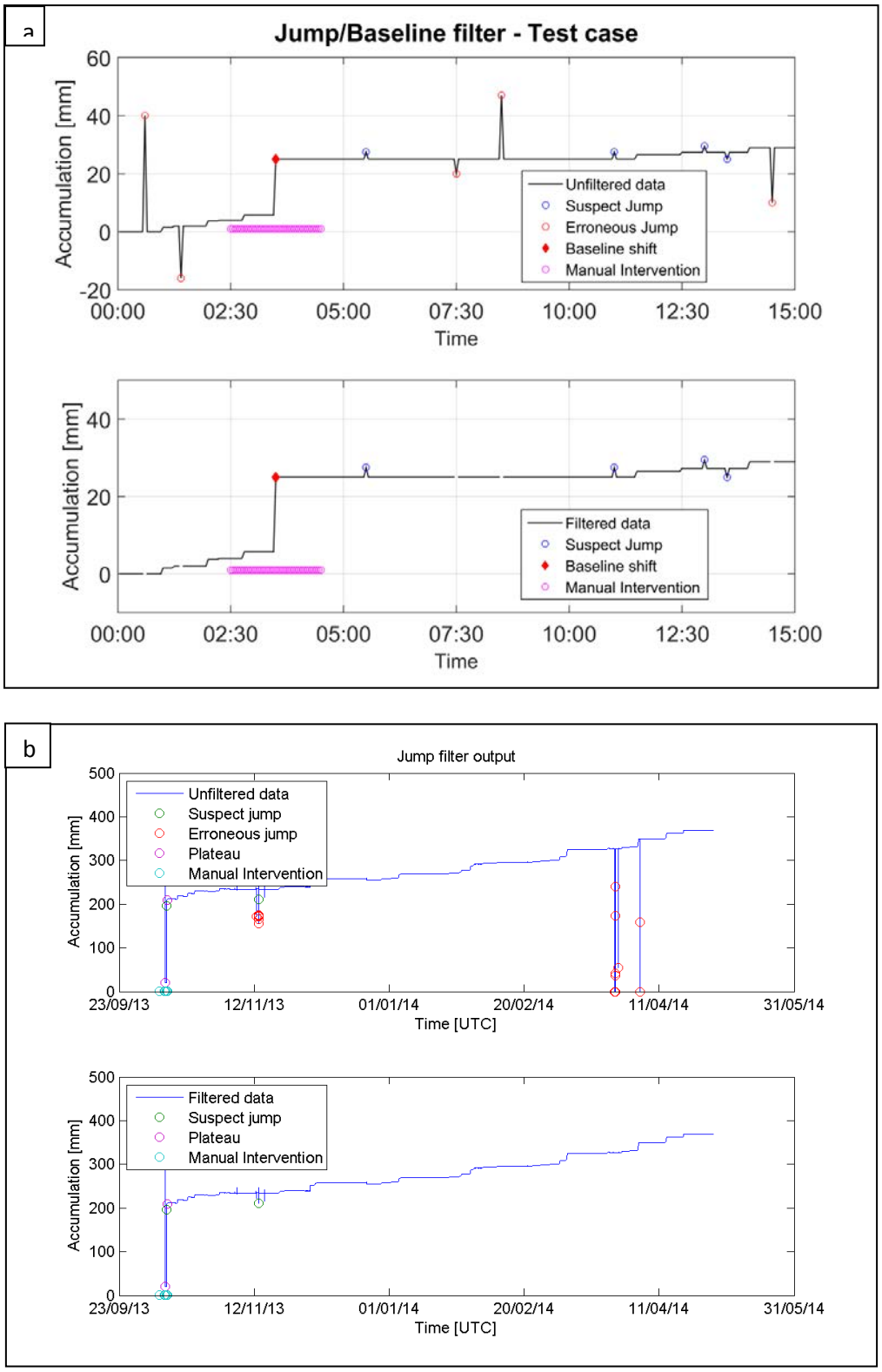


Figure 3.66. Example of the jump/baseline shift filter applied to (a) an artificial dataset simulating a dump of snow following capping and (b) accumulation data from the emptying of a weighing gauge at a SPICE site. The suspect and erroneous jump thresholds are (a) 2 mm/min and 3 mm/min and (b) 20 mm/min and 30 mm/min, respectively. Both (a) and (b) have a plateau threshold set to 60 min. The top panel in each plot shows unfiltered data with all flags detected, while bottom panel shows the resulting filtered data.

3.3.2.1.4 *Noise filter*

Precipitation accumulation measurements from weighing gauges are subject to noise, the magnitude of which increases with increasing wind speed and increasing accumulation in the bucket (Duchon, 2008). To mitigate the influence of noise, various types of filters can be employed. Several filter methods were tested using precipitation accumulation datasets from Geonor T-200B3 and OTT Pluvio^2 gauges from SPICE measurement sites (e.g. Colli, 2013). The focus was primarily on datasets with 6-second sampling intervals. In all cases, noise filters were applied after the maximum/minimum value and jump and baseline shift filters outlined in the previous section.

The majority of methods tested applied a filter of specified width (specified number of datapoints filtered in each step) to a moving window along the time series; these are collectively referred to as time-domain filters. Several time-domain filters were tested, each with a different functional form (e.g. mean, polynomial, Gaussian). Sample results are shown in Figure 3.67 and Figure 3.68. A frequency-domain approach was also tested, applying a fast Fourier-transform to convert the time-series data to the frequency-domain, to which a Gaussian filter was applied. A sample precipitation dataset from a single transducer of a Geonor T-200B3 gauge and filtered datasets using a frequency-domain Gaussian approach with different filter widths are shown in Figure 3.69.

Qualitative assessment of filtered results using these methods indicated that the moving average and Gaussian approaches were most effective (i.e. appeared to remove the most noise) for the datasets and filter parameters tested. For the Gaussian approach, the time-domain filters were more straightforward to implement and required less computation time than the frequency-domain filters. Based on this assessment, moving average and Gaussian time-domain filters were selected for additional testing and optimization as outlined in the following section.

3.3.2.1.4.1 Quantitative assessment of noise filters using artificial datasets

The quantitative assessment of data-filtering methods for precipitation datasets is complicated by the combined contributions from precipitation, noise, and any artifacts (e.g. variations due to temperature) present in measurement data. Insight can be gained from periods without precipitation, during which the “pure” precipitation signal is zero, but even in these cases, the superimposition of noise and artifacts can complicate the assessment. Further, the methods under consideration for data quality control should be tested using datasets covering the full range of expected precipitation conditions.

To address these issues, artificial datasets were generated to enable the quantitative assessment of filter performance. The datasets were generated for intensities ranging from very light (0.6 mm/hr) to very heavy (30 mm/hr) with a sampling interval of 6 seconds. Two scenarios were considered for each intensity: (1) a continuous, linear increase in accumulation at the specified intensity (linear scenario); and (2) step, or interval, increases in accumulation, with 10 x 1 minute linear increases in accumulation separated by 1-minute periods without precipitation (step scenario). The step scenario provides an avenue for testing the dynamic response of the filtering method and was used as the primary means of filter assessment. Sample datasets generated for both scenarios are shown in Figure 3.70a and Figure 3.71a.

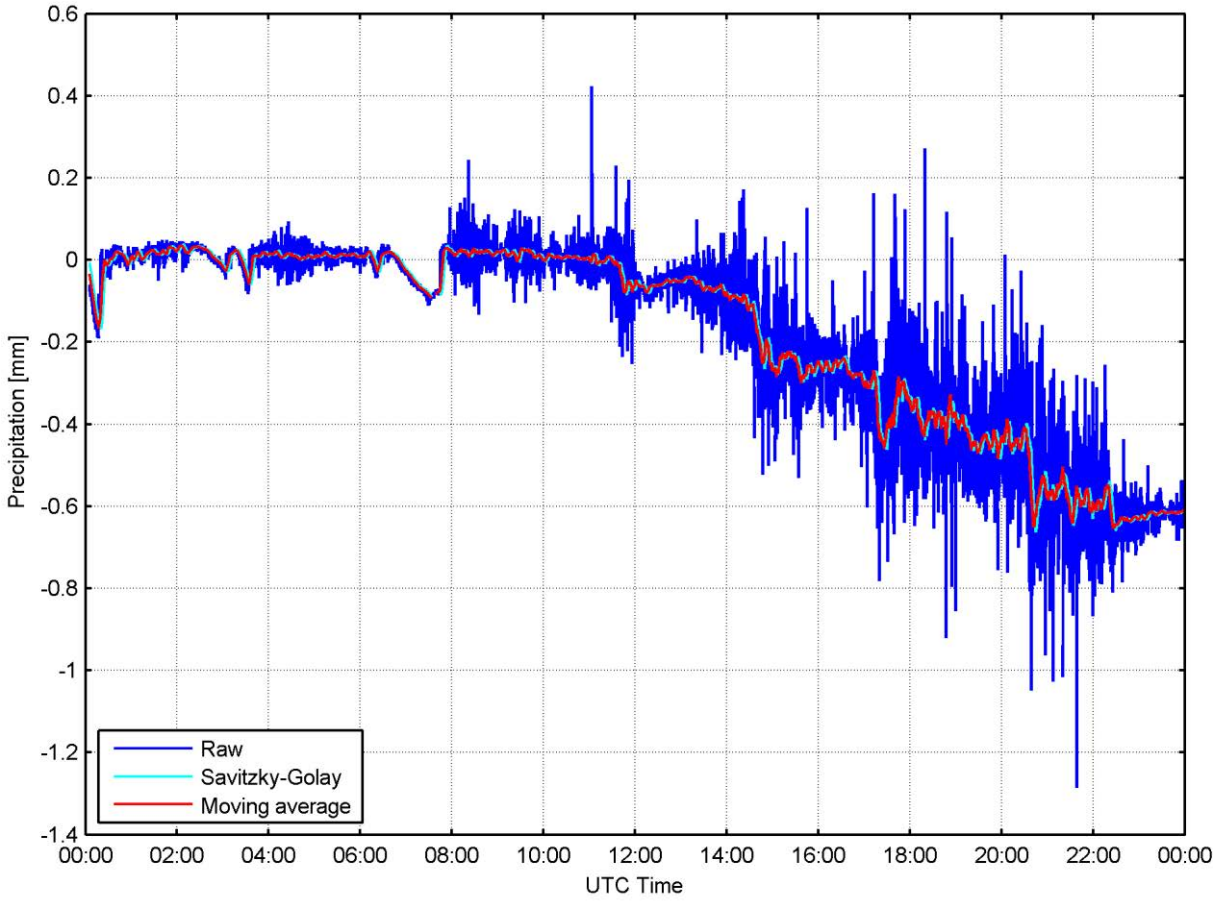


Figure 3.67. Example of filter testing on site datasets from Geonor T-200B3 gauges. Savitzky-Golay (3rd order polynomial) and moving average filters, both with widths of 5 minutes, were applied to data from a Geonor gauge in a DFIR-shield at the Marshall site on Dec. 27, 2012. Data and filtered outputs are shown for a single transducer and zeroed for visual comparison.

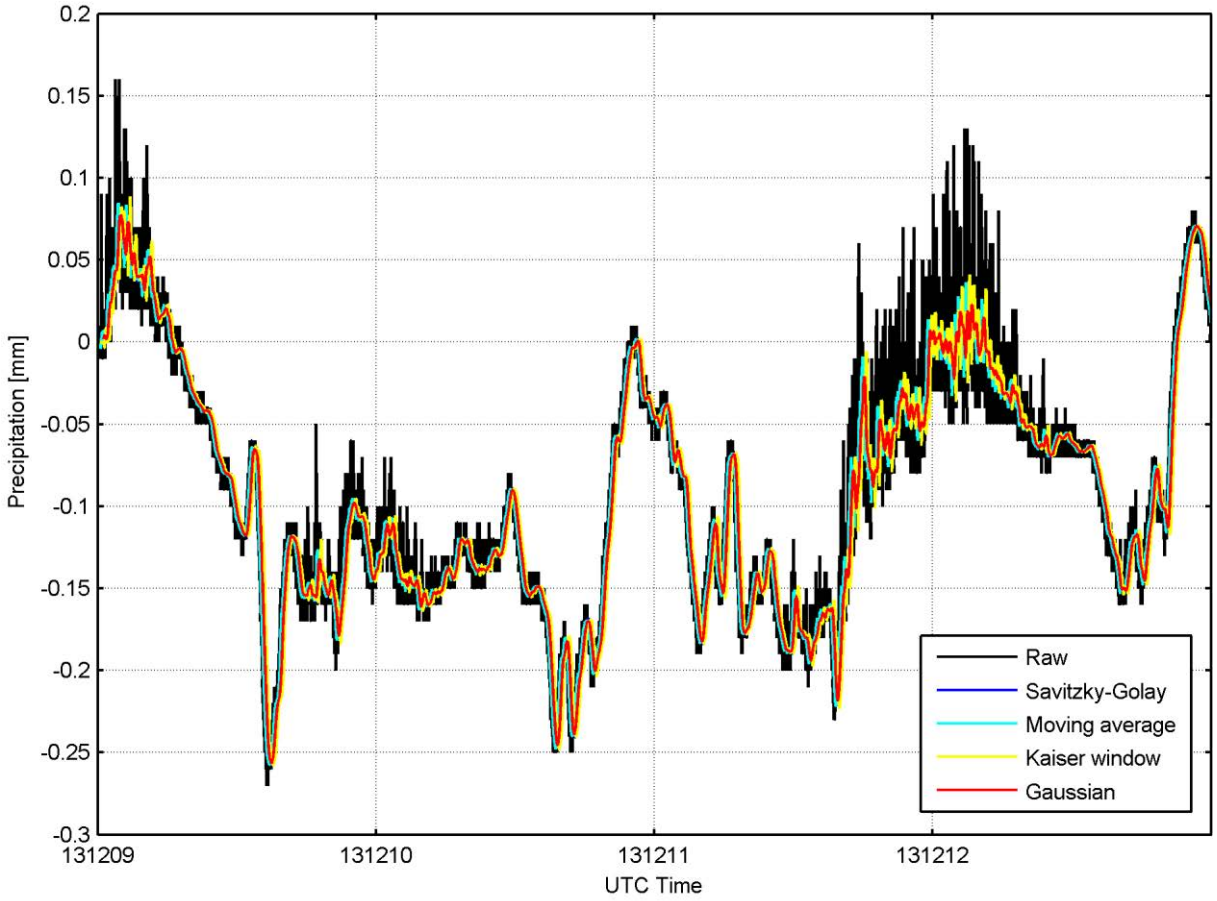


Figure 3.68. Example of filter testing on site datasets from OTT Pluvio^2 gauges. A Savitzky-Golay filter (3rd order polynomial, N = 181 points), moving average filter (N = 101 points), Kaiser window filter (N = 500 points, α = 3.4, fc = 0.0075398), and Gaussian filter (N = 301 points, σ = 50 points) were applied to 6-second Bucket RT data from a Pluvio^2 gauge in a double-Alter-shield at CARE from Dec. 9-12, 2013.

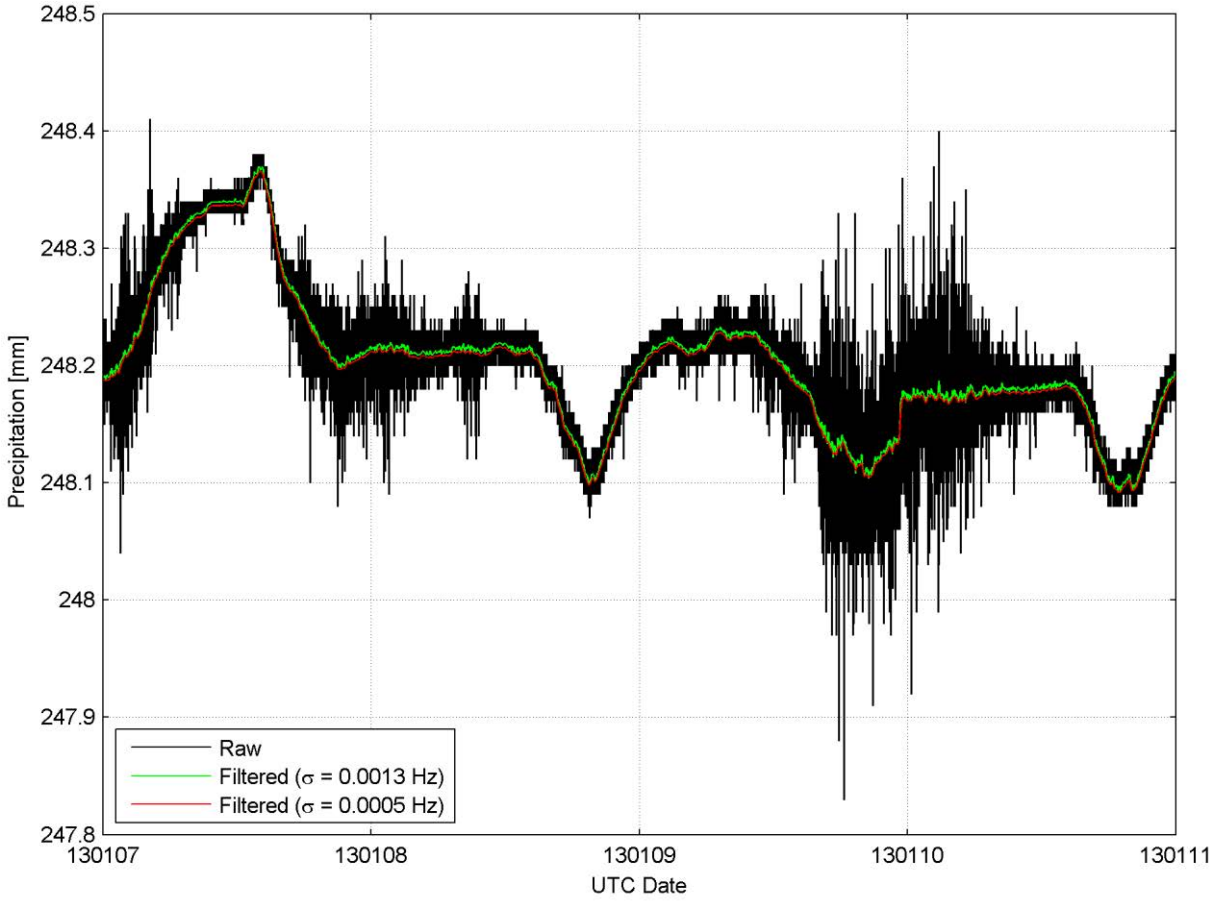


Figure 3.69. Raw and filtered datasets from a single transducer of a Geonor T-200B3 gauge at CARE when a frequency-domain Gaussian filter was employed with standard deviations of 0.0005 Hz (red curve) and 0.0013 Hz (green curve), corresponding to 5 minutes and 2 minutes in the time-domain, respectively. In both filtered datasets, the cutoff frequency corresponds to five times the standard deviation of the Gaussian distribution employed.

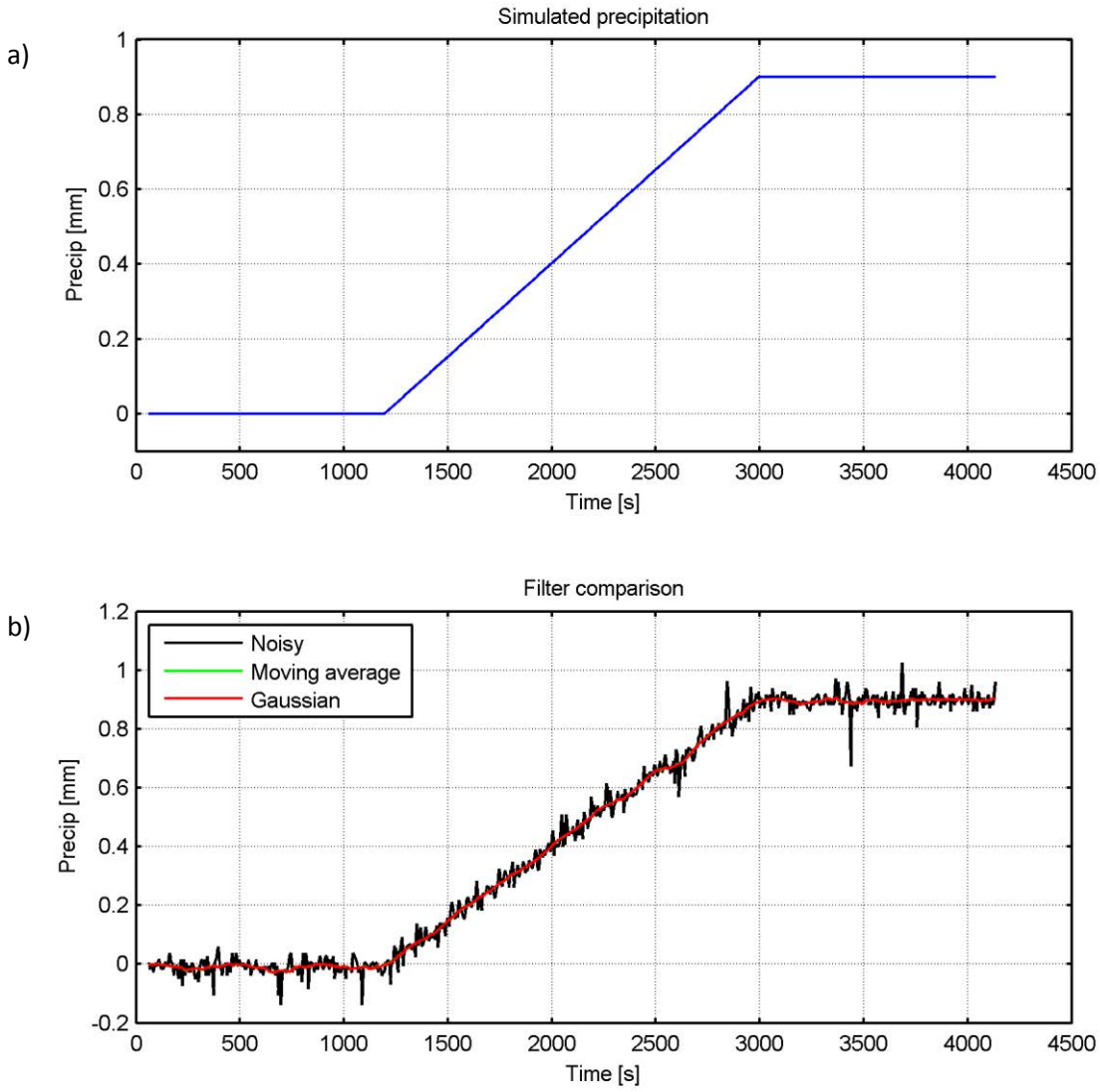


Figure 3.70. (a) Clean dataset in linear scenario for a precipitation rate of 1.8 mm/hr and (b) corresponding dataset with noise applied from Geonor noise PDF (Figure 3.73) and filtered outputs.

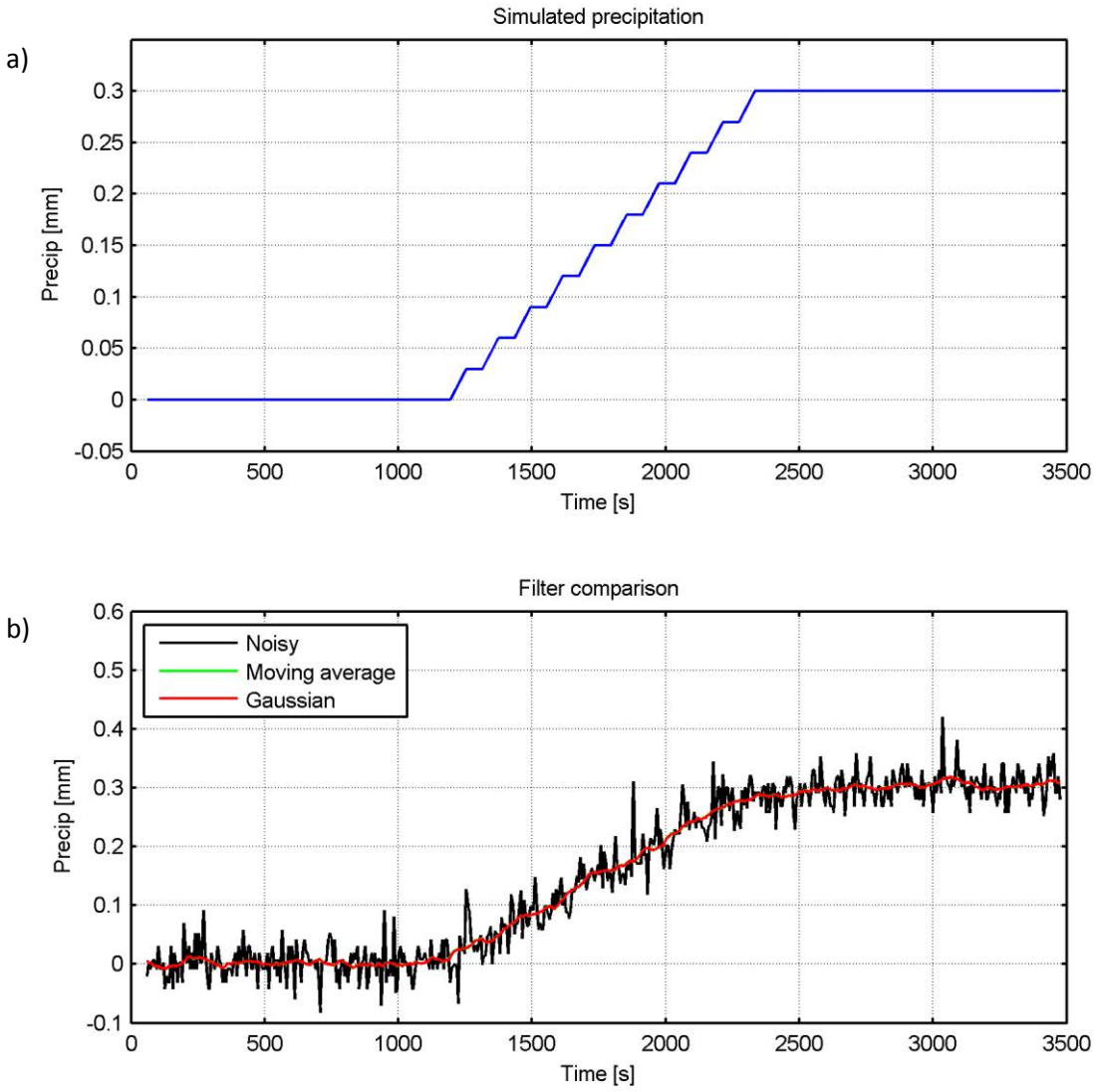


Figure 3.71. (a) Clean dataset in interval scenario for a precipitation rate of 1.8 mm/hr and (b) corresponding dataset with noise applied from Pluvio^2 noise PDF (Figure 3.74) and filtered outputs.

Filter performance was tested using noise derived from non-precipitating datasets. The magnitude of noise is typically dampened during precipitation events, so noise values observed in non-precipitating conditions are considered to represent upper-limit values for testing. Approximately 19,500 datapoints, sampled at 6-second temporal resolution (corresponding to roughly 32.5 hours of data), were selected from a Geonor T-200B3 gauge and OTT Pluvio^2 gauge at the Canadian CARE site during four non-precipitating events. Noise values were determined by taking the absolute value of the difference between each 6-second datapoint and the median value for each of the four events. Temperatures varied within 2 °C for each of the selected events, so any temperature effects on the datasets are expected to be small. An example of one of the events used for the determination of noise values is provided in Figure 3.72.

A histogram was generated from the compiled noise values from all four events for each gauge type, representing noise probability density functions (PDFs) as shown in Figure 3.73 and Figure 3.74.

Weighted random sampling with replacement was used to apply noise values from the PDFs to the "clean" datasets. Examples of the resulting "noisy" datasets for Geonor and Pluvio gauges are provided in Figure 3.70b and Figure 3.71b, respectively, for the corresponding clean linear and step datasets shown in Figure 3.70a and Figure 3.71a, respectively.

Moving average and Gaussian (time-domain) filters with widths from 1 to 10 minutes were used to filter the noisy artificial precipitation datasets for each gauge type, with intensities from 0.6 to 30 mm/hr in both linear and step scenarios. Filter performance was similar for the moving-average and Gaussian approaches. For both gauge types and filter methods, the lowest root mean square error values (indicating the best filter performance) were obtained for longer filter widths at low intensities and for shorter filter widths at higher intensities. These results were observed for both the linear and step scenarios. For the Gaussian filter, the lowest RMSE values were obtained using distributions with standard deviations equivalent to the filter width or one-half of the filter width. Filter widths between 2 and 5 minutes performed well for both gauge types.

3.3.2.1.4.2 Noise filter implemented in SPICE

Based on the assessment results, the noise filter implemented in the SPICE automatic quality-control approach for 6-second data was a Gaussian filter with a width of 2 minutes and a standard deviation of 1 minute (equivalent to one-half of the filter width). For 1-minute data, a Gaussian filter with a width of 4 minutes and a standard deviation of 2 minutes was employed. The Gaussian filter was selected over the moving average filter, despite similar performance for the datasets assessed, because the Gaussian filter gives more weight to values near the mean and less weight to values further from the mean, while the moving average applies equal weight to all values. This approach was adopted at the fourth meeting of the SPICE-IOC, as indicated in the corresponding report (IOC-4 Final Report, Davos, 2013). Examples of accumulated precipitation datasets from Geonor T-200B3 and OTT Pluvio^2 gauges that have been filtered using these parameters (6- second data) are provided in Figure 3.75.

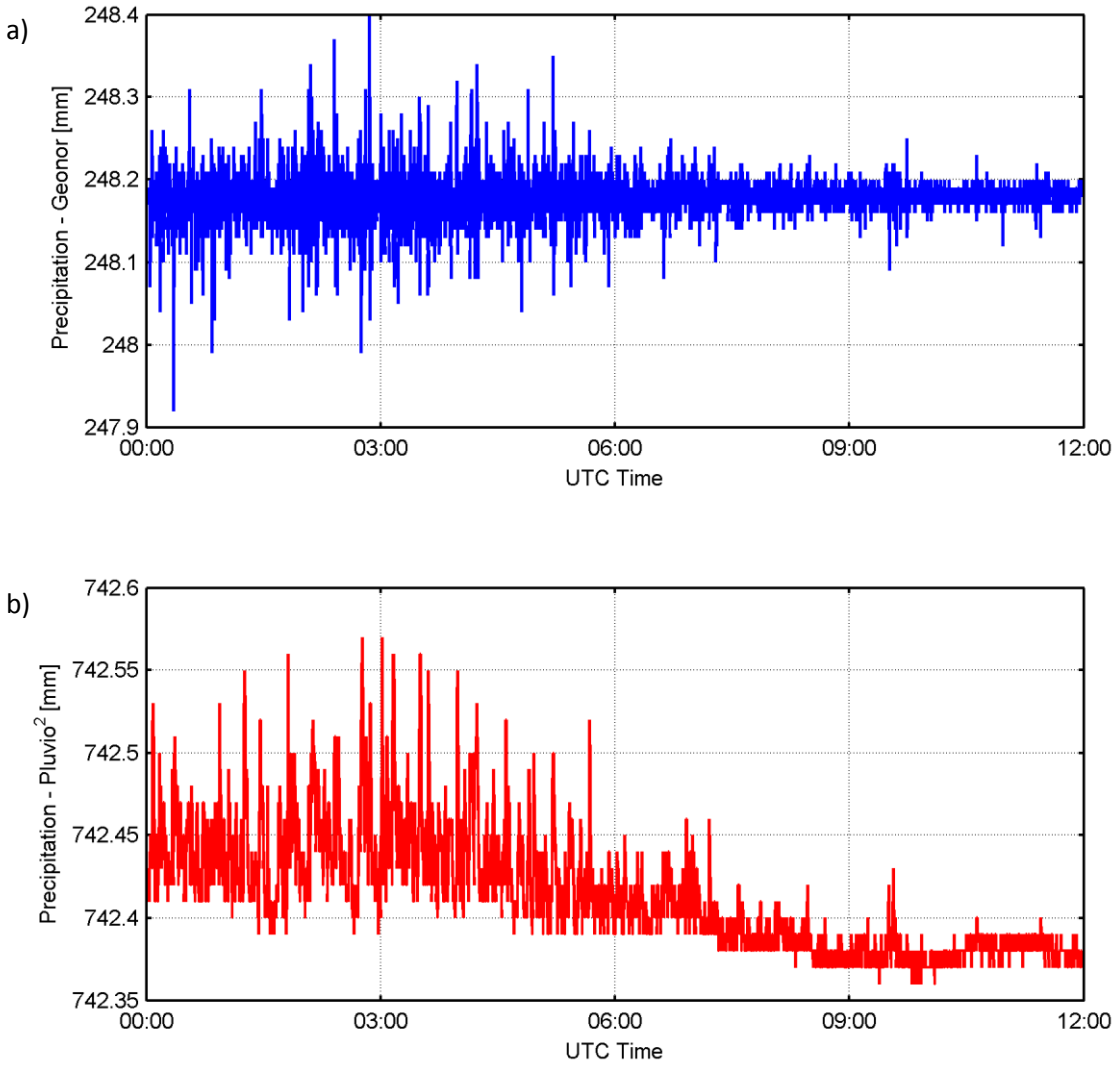


Figure 3.72. Examples of non-precipitating datasets used for the derivation of noise values for (a) a transducer from a Geonor T-200B3 gauge and (b) Bucket RT data from an OTT Pluvio² gauge at CARE on Jan. 10, 2013.

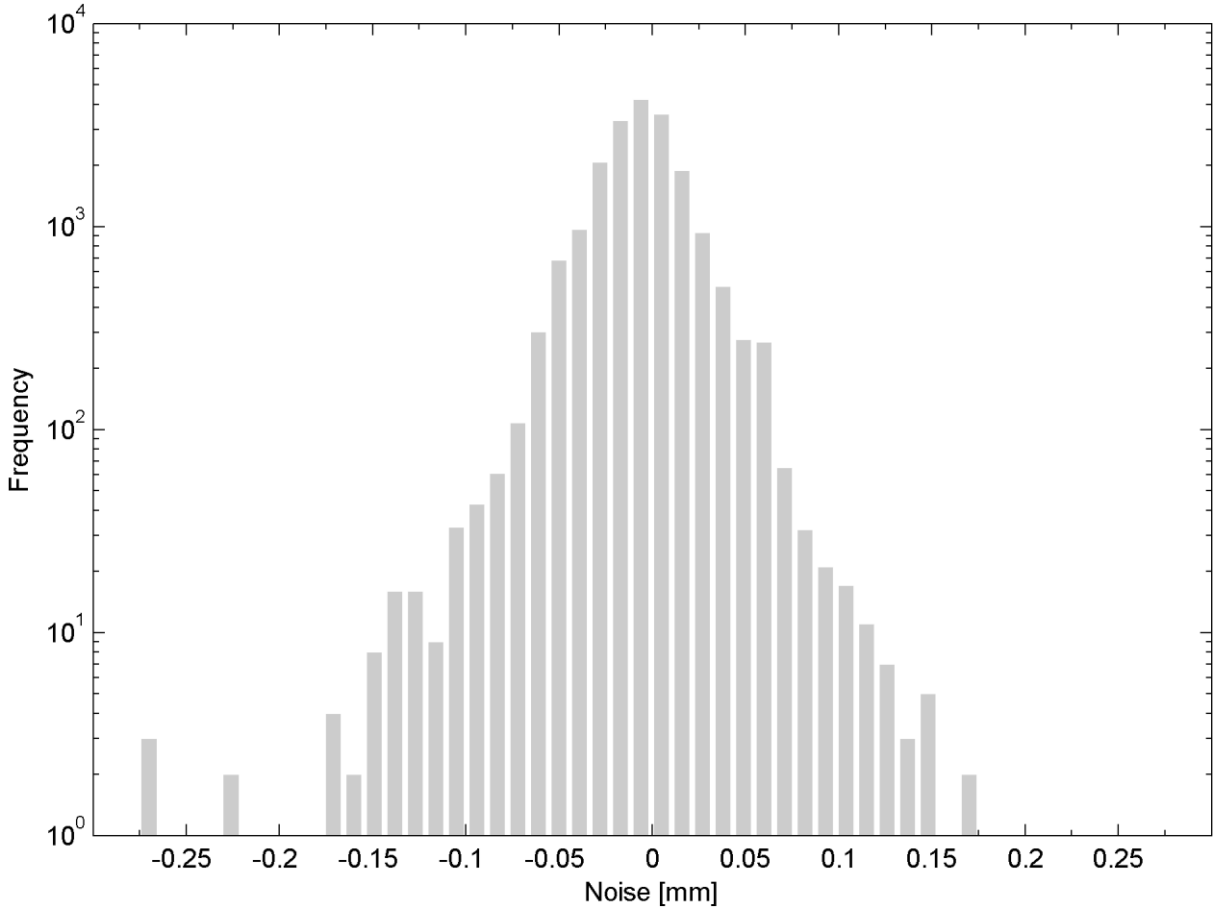


Figure 3.73. Histogram (50 bins) of noise values determined from Geonor T-200B3 precipitation accumulation measurements over selected non-precipitating events at CARE. This histogram represents the Geonor noise PDF used to generate noisy artificial datasets.

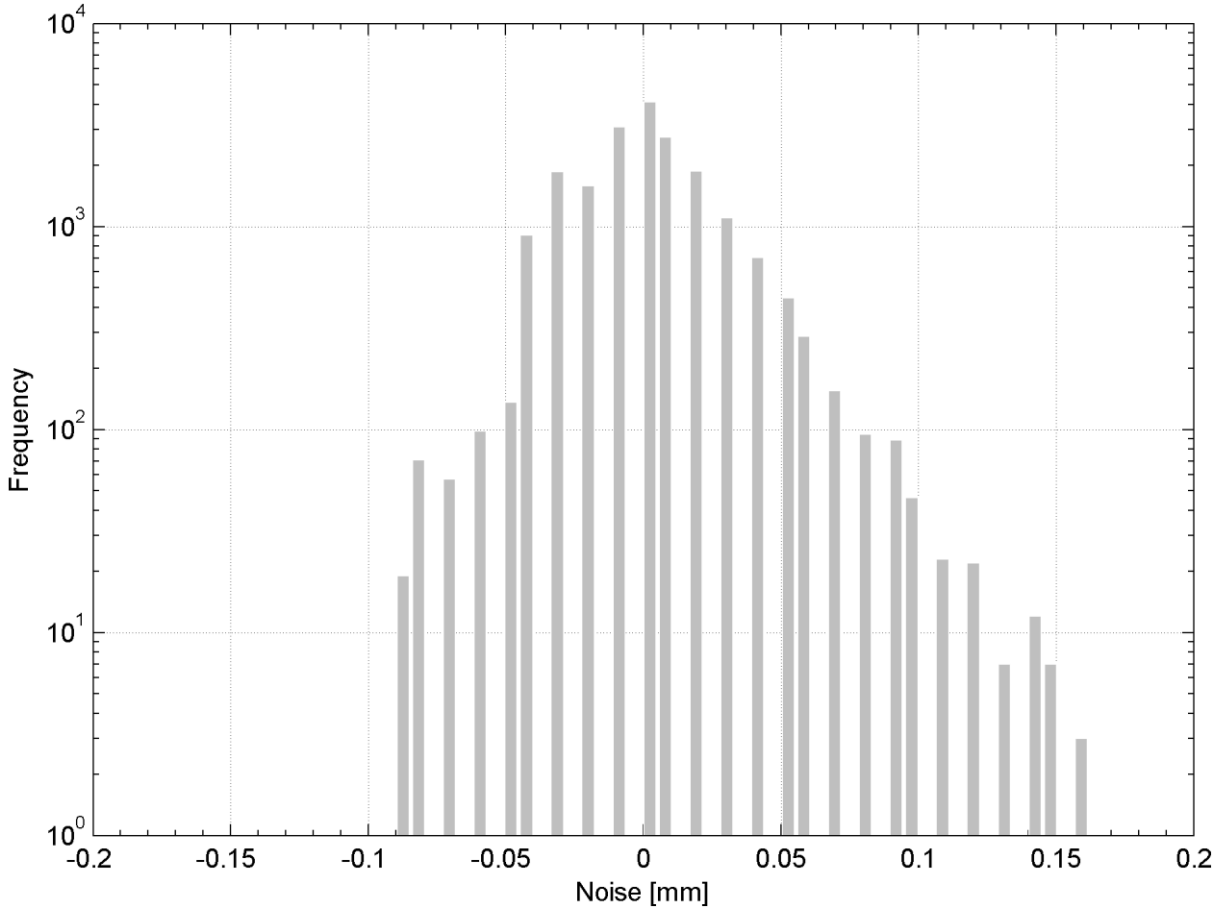


Figure 3.74. Histogram (50 bins) of noise values determined from OTT Pluvio^2 Bucket RT measurements over selected non-precipitating events at CARE. This histogram represents the Pluvio^2 noise PDF used to generate noisy artificial datasets.

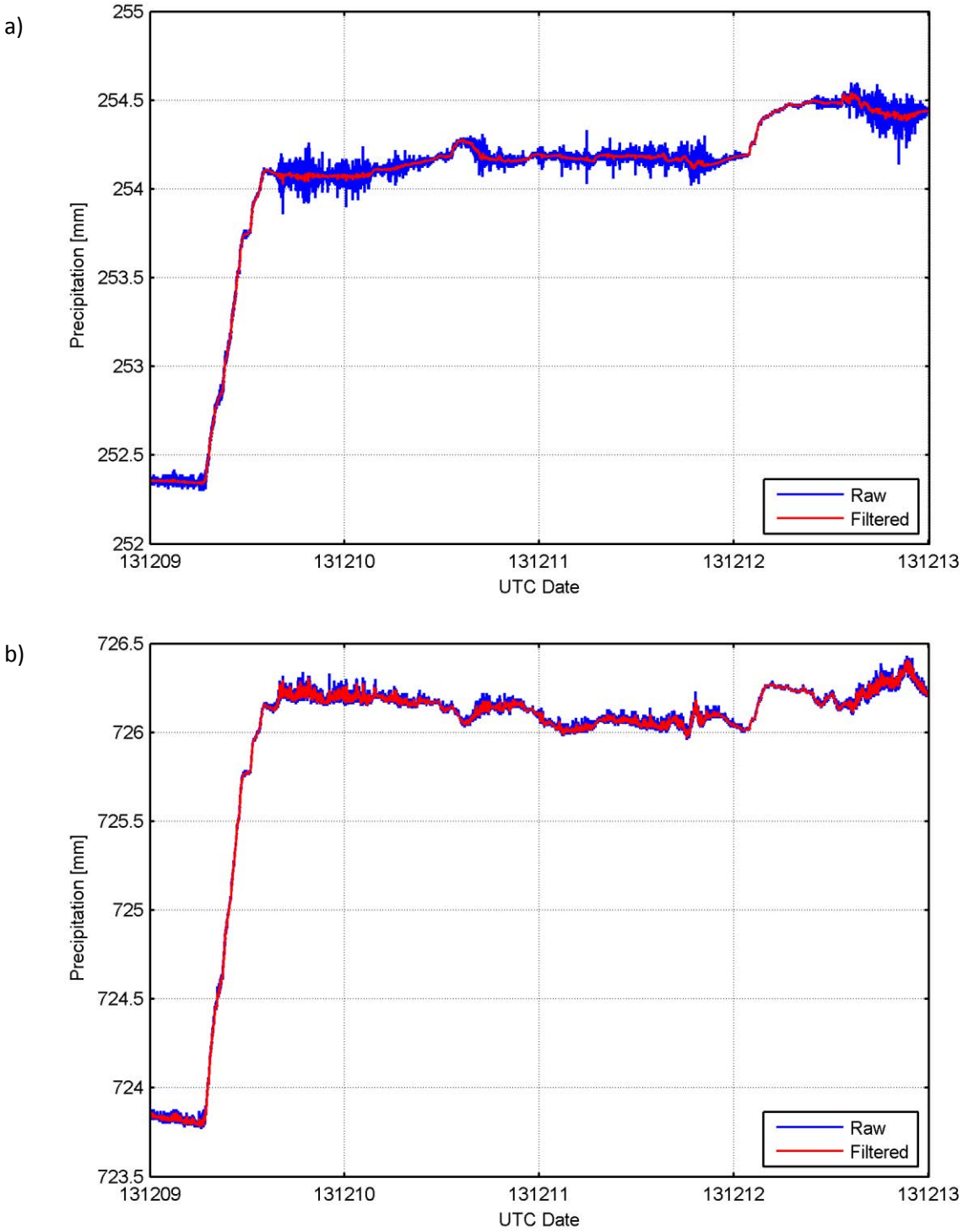


Figure 3.75. Examples of raw datasets and corresponding filtered datasets using a Gaussian noise filter with a width of 2 min and standard deviation of 1 min for 6-sec precipitation accumulation data from (a) a single transducer of a Geonor T-200B3 gauge in a single-Alter shield and (b) an OTT Pluvio^2 gauge (Bucket RT output) in a single-Alter shield at CARE from Dec. 9 to Dec. 12, 2013.

3.3.2.1.5 *Temperature compensation*

Prominent in precipitation datasets from Geonor T-200B and OTT Pluvio^2 gauges are low-frequency variations related to solar heating. These variations are typically diurnal in nature, with transducer output varying with heating/cooling in response to the daily solar cycle. These variations can artificially increase the accumulated precipitation amounts reported by weighing gauges, in particular when similar in magnitude to increases resulting from real precipitation.

For Geonor gauges, Duchon (2004) proposed the use of temperature coefficients derived from the daily accumulation-ambient temperature relationship to mitigate the influence of solar radiation on measurements from vibrating-wire transducers. Application of this approach on a broader scale would require temperature coefficients to be computed for each transducer, for each gauge under consideration. The magnitudes of coefficients vary with bucket weight (Duchon, 2004) requiring new coefficients to be computed periodically. In addition, consideration must be given to gauge orientation, solar elevation, azimuth angle, and cloud cover, which all influence the amount of incident solar radiation (Duchon, 2004). Further, the use of ambient temperature in the determination of coefficients does not capture the root cause of the observed variability – the temperatures of the vibrating-wire transducers, which are not measured by Geonor gauges – and could vary significantly from the ambient temperature measured outside the gauge.

Pluvio^2 gauges, on the other hand, measure the temperature of the load cell directly. Each load cell undergoes static temperature calibration performed by the manufacturer, used in the gauge firmware to compensate for temperature variations (see Section 3.1.3.4). The details of the compensation procedure are proprietary, and this makes it difficult to assess independently. Further, the Bucket RT data, representing the rawest possible output from the Pluvio^2, have already been compensated for temperature.

Given the caveats associated with the temperature compensation of precipitation accumulation data from Geonor T-200B3 and OTT Pluvio^2 gauges, no additional temperature compensation was applied to these data in SPICE. Instead, the focus was placed on characterizing the net accumulation and temperature variation over the specific precipitation events selected for subsequent analysis. (See Section 3.4 for details of the SPICE precipitation event-selection algorithm.) When available, other instruments (e.g. precipitation detectors or disdrometers) were employed in the event selection algorithm to facilitate the separation of real precipitation from increases in accumulation resulting from temperature variations.

3.3.2.2 Data aggregation

Quality-controlled datasets with sampling intervals of less than one minute are aggregated to generate one-minute datasets for subsequent analysis. The aggregation approach differs according to parameter type. For precipitation accumulation datasets, the aggregation step involves the selection of the last filtered datapoint from each minute. For other parameters (e.g. temperature, precipitation intensity, wind speed), aggregation generally entails a simple block average, with some exceptions for specific parameters (e.g. vector average for wind direction). For precipitation-type data reported by present weather sensors or disdrometers, the aggregation involves keeping the highest numerical SYNOP code (Tab. 4680) of each minute to represent the minutely value. The aggregation approach is more complicated for sensors with multiple transducers; the approaches considered for the three-wire Geonor T-200B3 and Belfort precipitation gauges are discussed in the following section.

3.3.2.2.1 *Aggregating data from gauges with multiple transducers*

The SPICE IOC determined that Geonor T-200B gauges used as part of the FWRS should have three active transducers, working independently (IOC-2 Final Report, Boulder, 2012). The load of the bucket, including any accumulated precipitation, is shared among the three wires. While the precipitation amount could be determined from any of the individual transducers, an aggregate of the transducer outputs is typically used for two reasons: (1) averaging transducer outputs reduces the magnitude of noise due to wind effects relative to individual transducers (Duchon, 2008) and (2) there are often differences among the transducer outputs resulting from differences in orientation (with respect to the sun, with respect to vertical) and temperature, or aspects of the configuration (unbalanced load, vibration). Accordingly, all precipitation data from gauges with three transducers used in SPICE were derived by aggregating the outputs of the three transducers, as monitored separately with data loggers.

Aggregation serves as an additional data processing step. If all transducers perform ideally, an arithmetic average is a viable aggregation approach (Baker et al., 2005). This is not typically the case, however, and different approaches have been used to account for differences among the transducer outputs. Given the requirement for uniformity across all participating sites, automated approaches are preferable. Options for automated aggregation approaches are outlined in the following paragraphs, along with the approach implemented in SPICE.

3.3.2.2.1.1 Comparison of transducer pairs

The U.S. Climate Research Network (USCRN) has used the redundancy offered by the three-wire configuration to assess which transducer data to include in the estimation of precipitation amounts (Baker et al., 2005; Leeper et al., 2015). Differences in accumulation between pairs of transducers over a set time period are compared and considered relative to a threshold value. If any of the differences exceed the threshold, the contributing wire or wires are identified and are not used to compute the aggregate value. The use of redundant information helps to safeguard against system failures (Leeper et al., 2015); however, the resulting precipitation amounts can be influenced by transducer noise and gauge evaporation (Leeper et al., 2014; 2015).

3.3.2.2.1.2 Weighted averaging

To reduce the influence of variability, or noise, in the individual transducer outputs on aggregate precipitation amounts, weighted averaging can be used, with coefficients determined by the relative magnitudes of noise among the transducers. This concept was tested on SPICE datasets using inverse variance-based weighting. Here, variance values are computed within a moving window (typically 30 minutes) for each wire, summed over a given dataset, and inverted. The weighting coefficient for a wire *i* is then as follows:

$$WC_i = \frac{1}{\sum_i \sigma_i}$$

where *WC* is the weighting coefficient and σ is the variance. These coefficients are normalized by dividing by the sum of the weighting coefficients for all three wires.

$$WC_{i,norm} = \frac{WC_i}{WC_1 + WC_2 + WC_3}$$

Applying these weighting coefficients when aggregating precipitation accumulation data (Acc) from all three wires gives less weight to noisy wires and more weight to cleaner wires, effectively reducing the magnitude of noise in the aggregate dataset, Acc_{agg}.

$$Acc_{agg} = WC_{1,norm}\, Acc_1 + WC_{2,norm}\, Acc_2 + WC_{3,norm}\, Acc_3$$

The latest aggregation approach implemented by USCRN also uses weighted averaging, and has been shown to reduce the influence of transducer noise and gauge evaporation on precipitation datasets relative to those determined from the comparison of transducer pairs (Leeper et al., 2014; 2015).

3.3.2.2.1.3 Aggregation approach implemented in SPICE

With the load being shared among the three transducers, reducing or removing the contribution from differently performing wires may bias the aggregate value too high or too low. This interdependence of transducers complicates the automated aggregation of data. For simplicity, the aggregated Level 2b data was determined as an arithmetic average. The Level 2b data for individual transducers were kept and archived. For cases in which the output from individual transducers differed significantly from the others or a transducer was malfunctioning, the arithmetic-averaged data may be compromised, necessitating the manual removal of a given wire from the average calculations.

3.3.2.3 Manual quality control procedure

The automatic QC procedure identified and addressed routine data issues in a uniform and standardized manner. Data identified and flagged as “suspect”, “baseline shift”, and “manual intervention” by the automatic procedure required manual assessment prior to analysis. Further, data issues related to site maintenance, the surrounding environment, and sensor installation and configuration were not always identified or addressed, necessitating manual assessment and intervention. Manual assessment was conducted in collaboration with site managers and made use of site logs, photos, and other metadata.

Data identified as being erroneous or compromised by events or conditions at the site were manually removed from the datasets (replaced by null values) and flagged (flag = 6; Table 3.8). Baseline shifts in weighing gauge data confirmed as being due to maintenance (emptying, calibration) were removed and flagged (flag = 6), and baselines were adjusted to provide a continuous accumulation time series for comparison with the reference (time series check). An example of such a manual correction is shown in Figure 3.76. Baseline shifts in data due to gauge capping were addressed on a case-by-case basis; further details of snow capping in precipitation datasets are provided in Section 4.2.1. In all cases, manual adjustments to the data were tracked for further reference.

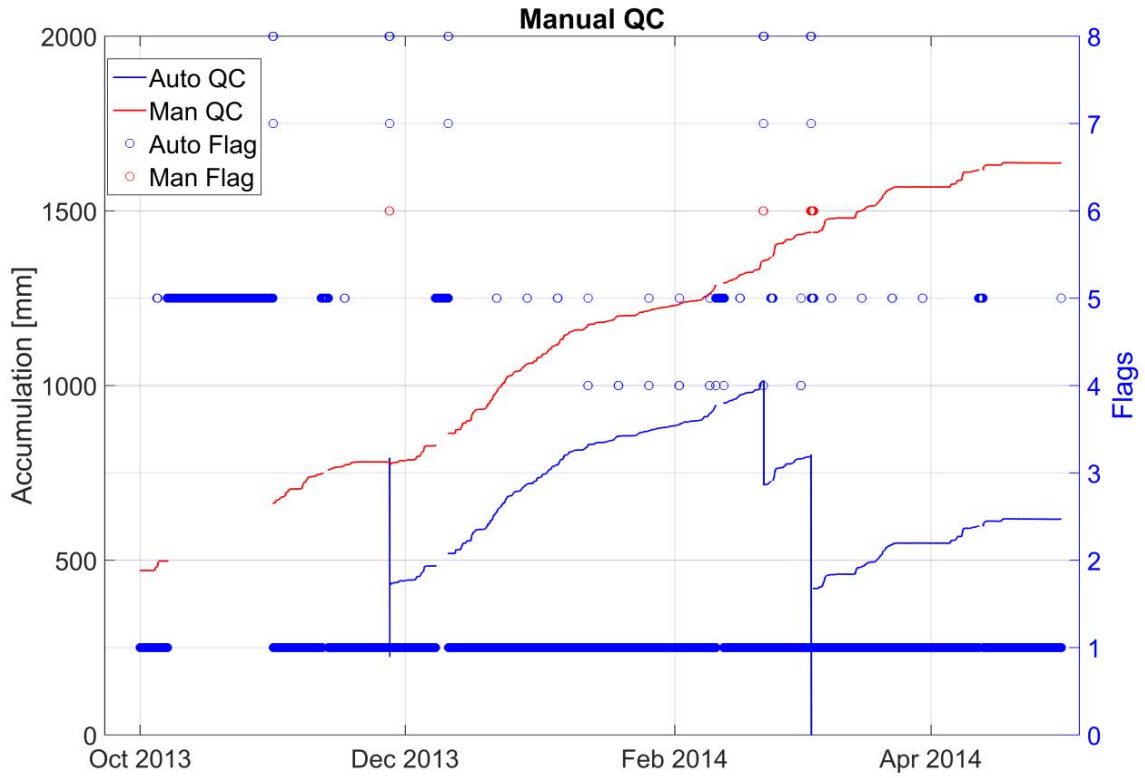


Figure 3.76. Manual QC applied to a SPICE accumulation dataset for which the gauge bucket was emptied during the measurement season.

3.3.2.4 Quality control flags

The quality control and aggregation of precipitation datasets from reference gauges, systems under test, and ancillary gauges were accompanied by a flagging procedure to provide additional insight into gauge performance and data integrity, and to support the data analysis. A system of flags was developed that follows closely the approach implemented in the *WMO Field Intercomparison of Rainfall Intensity* (Vuerich et al., 2009). This proposal is outlined in Table 3.8.

Flags were generated for each parameter of interest, for each gauge. Flags identified in datasets with sampling intervals of less than one minute were carried forward when aggregating to one-minute datasets. A threshold of 66% for flag carryover was set for the missing and erroneous flags (i.e. If 66% of points within a given minute are flagged, that flag is carried over to the one-minute aggregate value). For the remaining flags, any instances of the flag being called within this time resulted in that flag being carried forward to the one-minute aggregate value.

A given datapoint can have only a single flag value within the current system. For instances in which multiple flags were observed for a given datapoint, the following order of priority was applied:

Missing (5)> Site (6)> Baseline shift (7)> Manual intervention (8)> Erroneous (4)>Suspect (3)>Good (1)

Note that flag values of “2” are not included. This value was designated for “inconsistent” values, for example, wind direction values not equal to zero when the wind speed is zero – but was not ultimately implemented. These types of issues were identified in the manual quality-control step and flagged with the site value (flag = 6).

As stated in Section 3.3.2.1, erroneous values were replaced with null values by the maximum/minimum value and jump filters, while suspect values were only flagged. For datasets with sampling intervals less than one minute, the remaining suspect values may potentially impact the aggregate values in cases where the 66% criterion is not met. To address this concern, if fewer than 66% of points within a given minute are flagged as suspect (3), and there are no higher priority flags (flags 4 or 5 meeting the 66% criterion, or any instances of flag 7 or 8 within that minute), the resulting aggregate value was flagged as suspect (3).

For instruments with multiple transducers, flags were generated separately for the aggregate one-minute data from each transducer following the above criteria. The flags for the composite (averaged) one-minute instrument data (aggregating the contributions from each transducer) were aggregated such that the highest priority flag from a constituent transducer in a given minute was taken as the composite value.

Additional criteria were proposed for the carryover of flags to identified precipitation events. This is discussed within the context of the event selection algorithm in Section 3.4.

3.3.3 Quality control of snow-on-ground data

The quality control of the SoG automated sensor data was completed independently of the precipitation and ancillary data collected at the sites. The data were transferred offline to a central location via the NCAR database or directly from the site manager. The data then went through several phases of quality control before being archived in the NCAR database and redistributed to the SPICE analysis teams. As with the precipitation data, quality control consisted of: data aggregating and reformatting; automated range and jump filtering; manual removal of remaining outliers; and when necessary, adjustment of the zero-depth offset (used to derive snow depth from the distance-to-target measurement).

Sensor data archived by the sites (or by NCAR), often referred to as “raw” site data but technically classified as Level 1 data, were first reformatted and filed into time-consistent monthly space-delimited files that included two header lines (one containing the variable names and one containing the units). Data aggregation, if necessary, occurred at this stage (which is earlier in the QC process than for precipitation data). Any sensors with data frequencies higher than one minute were aggregated to one-minute resolution by determining the median (if measurement frequency is more than two per minute) or the mean (if the measurement frequency is two per minute). No other QC modifications were made to the data at this stage, termed the QC0 stage in the SoG quality control process. Missing data were given a universal identifier of -999 (equivalent to the null designator used by the precipitation QC process) and flagged as missing (flag=5). The data at this point were considered to be Level 2a data. The data quality flags for SoG data are outlined in Table 3.9.

Monthly QC0 files were then subjected to an automated filtering process that included site-specific range and jump filtering. Filter threshold ranges were set based on physical realities at each site (e.g. using an approximate maximum snow depth plus a 20% buffer). Data that did not pass these criteria were removed and flagged (flag=4). The flag column is added to each SoG data column, and the monthly data files were reproduced with a qc1 extension. Following the automated QC process and the production of the QC1 files, the data were visually inspected and any remaining outliers removed manually using an interactive selection process. To ensure consistency, the data were plotted over the entire season, and outliers not removed by the automated process were selected for removal and flagging. Data removed at the manual stage were flagged (flag=6) and the monthly data files

reproduced with a qc1C extension. As with the precipitation data, the SoG data were now considered to be Level 2b data.

Some of the SoG sensor data underwent a QC2 quality control stage, if necessary. The QC2 stage adjusted for incorrect zero-snow-depth offsets, but only if the season start and end offsets were the same. (Zero-snow-depth drift, where the offset changes from the start to the end of the season, is discussed in Section 4.2.6.3.) An additional flag column was added at the QC2 stage (flag=10) and the monthly data files reproduced with a qc2 extension. At each stage, the QC processes and metrics were logged and included in the metadata.

Table 3.9. Descriptions of quality flags used for SoG data. The flag numbers are consistent with the procedure developed for snowfall measurements.

Flag #	Description
1	Data OK
4	Bad data, data out of range and replaced with missing data flag
5	Missing data
6	Outlier removed manually
10	Corrected for zero offset issue

The manual reference SoG data, similar to the SoG sensor data, were quality controlled independently of the precipitation data. Manual measurements had initial data quality assurance and quality control via the observers and site managers at the time of collection and data entry. The data received from the sites was determined to be high quality, but the time series were graphically reexamined for outliers and inaccuracies. These were documented and manually corrected or removed prior to analysis.

3.4 Event selection

Authors: Audrey Reverdin, Michael Earle, Mareile Wolff, John Kochendorfer

Weather conditions during precipitation events can vary extensively over the course of a season. The different SPICE sites, characterized by different climatological conditions, increase further the diversity of precipitation events that need to be taken into account for analysis. To achieve comparable site datasets, a uniform method was required for defining and quantifying precipitation events, which could be applied to all SPICE data.

In the context of SPICE, a precipitation event was defined as a period of time when precipitation occurred with a high degree of confidence, as detected by the field working reference system. A baseline event duration of 30 minutes was established. This duration provides a balance between events being sufficiently long to be representative of snowfall events in a variety of climate and environmental conditions, while also being short enough to provide a sufficient number of events for analysis.

For each SPICE site, site event datasets (SEDS) were created for each winter season. The SEDS contain data from all precipitation instruments operating on the site, as well as from selected ancillary instruments, for all 30-minute intervals over which the FWRS reported a precipitation event.

The SEDS were derived from the one-minute quality-controlled datasets (level 2b data) and constitute level 4 data products, as defined in Section 3.4.1. The consistency of the approach described in the following sections made it possible to have comparable SEDS among all sites. The derivation of precipitation event datasets was a key component of the assessment and development of transfer functions for each sensor under test relative to the corresponding FWRS (see Section 3.7).

3.4.1 Description of the event-selection algorithm

The event-selection algorithm identifies precipitation events based on the quality-controlled data from two instruments: (1) the reference automatic weighing gauge, measuring accumulation, and (2) the precipitation detector, reporting on the presence or absence of precipitation. For S2 sites, the weighing gauge was the R2 reference; for S3 sites, the single-Alter-shielded gauge of the R3 reference was used. The precipitation detector had to be an optical precipitation detector or disdrometer, as defined in the IOC-4 Final rRport, Davos 2013 (Annex IV, p.4). For S2 sites, it was typically located near the R2 reference weighing gauge within the DFIR-fence, between the Alter shield and the inner wooden fence. At sites without a DFIR-fence, it was installed in a location sheltered from the wind.

The one-minute datasets from these two reference instruments are separated into consecutive blocks of 30-minute intervals (i.e. 00h01 - 00h30, 00h31 - 01h00, 01h01 - 01h30) over which the selection criteria are applied. The flowchart in Figure 3.77 illustrates the two steps of the event-selection algorithm applied over these intervals: event identification (step 1) and event parameter processing (step 2). In the first step, two algorithm options are considered: (1) when the precipitation detector was available on site and reported a valid output (column 1) and (2) when the precipitation detector was missing or reported invalid data (column 2). The third column in Figure 3.77 indicates when to proceed to the next stage in the algorithm. The two steps are described in detail in the following sections.

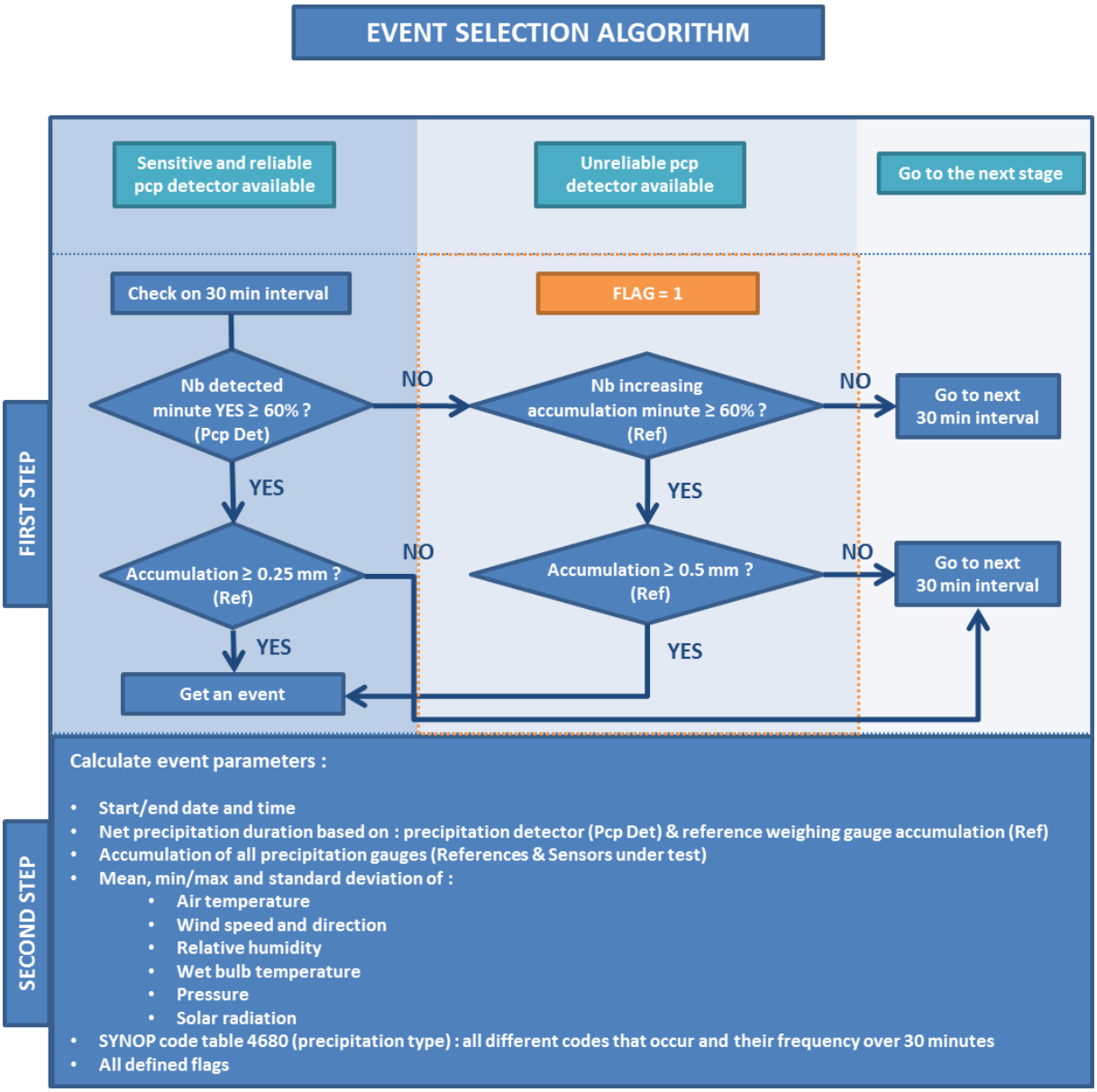


Figure 3.77. Flow chart outlining SPICE event-selection algorithm and calculated parameters.

3.4.1.1 Event identification

As a first step, data from the weighing gauge of the FWRS (including the precipitation detector if available) are examined over the 30-minute intervals. To be selected as an event, the data in this window must fulfill the following two conditions:

1) Net precipitation duration sufficiently long

The number of minutes during which precipitation is detected must be at leasr 60% of the window time (i.e. 18 minutes or longer), but does not need to be continuous. The precipitation duration is calculated based on precipitation-detector data (first column in Figure 3.77) by looking at the number of "Yes" cases that occurred during the 30 minutes. If data are missing or unreliable (second column in Figure 3.77), the algorithm examines the data from the weighing gauge in the FWRS and identifies the number of minutes during which there was increasing accumulation over the time

period. If this number exceeds 60% of the event duration, the net precipitation-duration condition is considered to be met.

2) Sufficient accumulation of reference gauge

The total accumulation in the reference weighing gauge during the 30 minute period must be equal to or greater than a defined threshold. Based on previous experience, this threshold amount was set to 0.25 mm when a reliable precipitation detector is available (first column in Figure 3.77) and to 0.5 mm over 30 minutes when such a detector is not available (second column in Figure 3.77).

A lower threshold was selected when event selection is based on a combination of data from a precipitation detector and the weighing gauge. There is a higher degree of confidence with two independently operating instruments. When only a weighing gauge was used for event selection, the threshold was more conservative.

Any 30-minute window during which these two conditions are fulfilled is considered to be a 30-minute precipitation event and is added to the SEDS. If these conditions are not fulfilled, the algorithm moves on to the next interval.

To track which procedure was applied for the identification of each event, a flag was designated to indicate if a precipitation detector was used in the process or not (i.e. Flag = 0 or 1, respectively). This flag is reported in the SEDS and appended to the aggregated quality-control flag (see Section 3.4.2 for more details).

An illustration of the application of this first step is found in Figure 3.78. In this example, the precipitation detector was not working on April 11. The selection of two events during this period was then based only on weighing gauge data using the higher accumulation threshold of 0.5 mm/30 min. The corresponding flag was also added during the process.

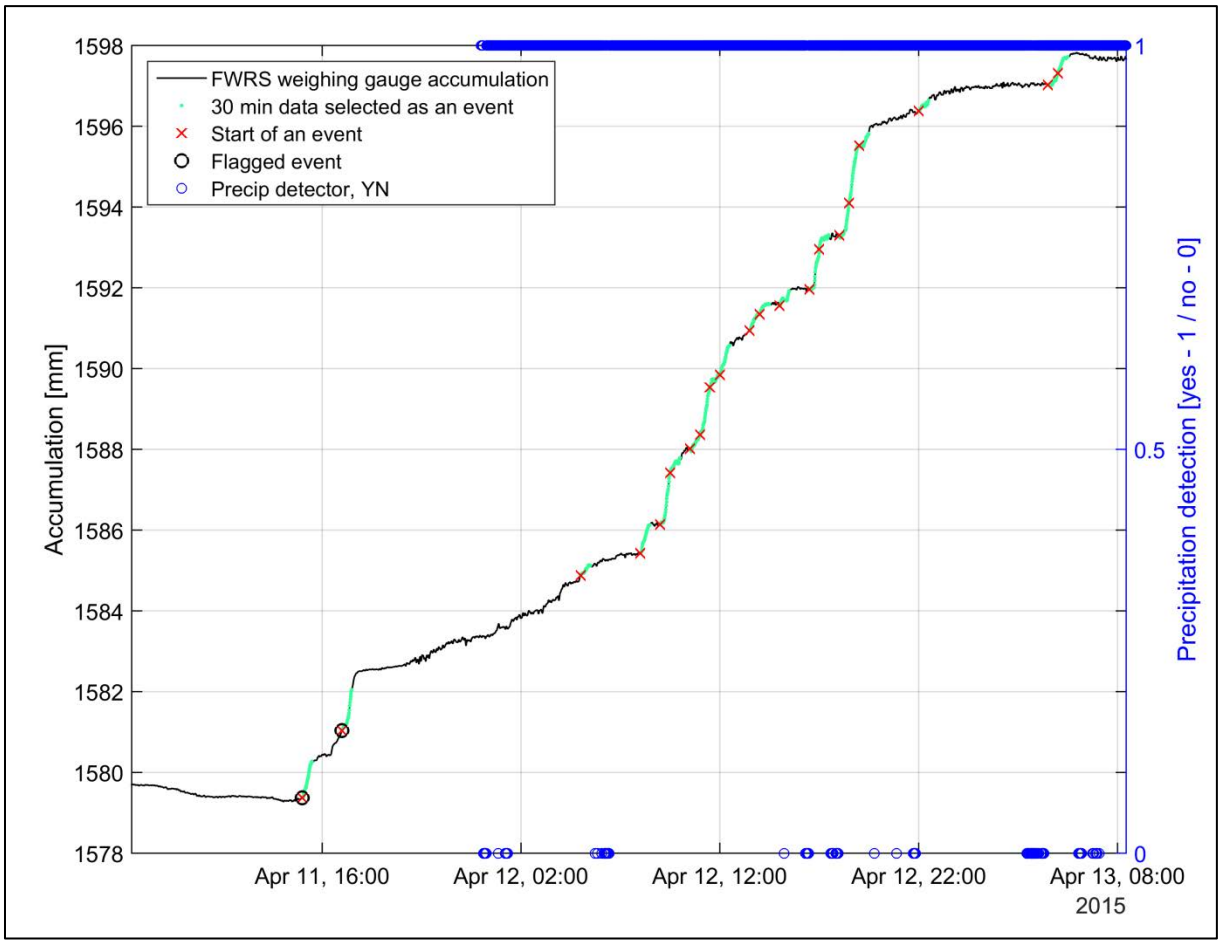


Figure 3.78. Example of event selection procedure applied to data from an R2 reference configuration. The precipitation detector in the reference configuration was not working on April 11; events selected on this date were flagged as being identified using only the accumulation data from the weighing gauge.

3.4.1.2 Event parameter calculation

For each 30-minute event identified by the first step in the procedure, the algorithm outputs several parameters to characterize the event in detail for further analysis. The list of parameters gathered in the SEDS was meant to be as consistent as possible for all sites to facilitate comparative analysis; however, since no two sites have identical equipment or sensor configurations, some adaptation was required. A general list of event parameters is outlined in Figure 3.77. The following approaches were used in the calculation of event parameters:

- The net precipitation duration of the event was calculated twice; once using the precipitation-detector data (when available) by summing the number of “minute-yes” reports from the sensor, and once using the reference-weighing-gauge accumulation data by summing the number of minutes for which increases in accumulation were observed. This provided an avenue to assess the consistency of the two instruments.
- The event accumulation was calculated by taking the difference between the last accumulation value and the first accumulation value over the 30-minute interval. For Geonor and Belfort weighing gauges, the accumulation of each individual transducer as well as the

accumulation of the average of the three transducers were computed and reported in the SEDS. For Pluvio^2 weighing gauges, the "Bucket RT" as well as the "Accumulated NRT" accumulation were computed and reported in the SEDS.

- For sensors under test outputting one-minute accumulation or intensity only, the sum, mean, minimum, maximum, and standard deviation of the data over the 30-minute event were computed and reported in the SEDS.
- For all ancillary measurements, the mean, minimum, maximum, and standard deviation of the data over the 30 minutes were also computed and reported in the SEDS. In general, the mean was computed as an arithmetic average of the data, except for wind direction where a vectorial average was used.
- The precipitation type, derived from the SYNOP code (Tab. 4680) of disdrometers or equivalent available devices, was reported in the SEDS by giving the minimum and maximum SYNOP value during the 30-minute event, as well as each individual code and their frequency (in minutes) during the event.

Altogether, the list of these computed event parameters with their corresponding statistics and flags constituted the comprehensive SEDS provided for each SPICE site and for each of the SPICE winter seasons (i.e. 2013/14 and 2014/15).

3.4.1.3 Thresholds and time intervals for SEDS

The thresholds and time intervals chosen for the event-selection algorithm are based on discussions during the IOC-4 Final Report, Davos, 2013. The objective when defining these parameters was to select precipitation events with a high degree of confidence, reducing the uncertainty related to light and/or sporadic precipitating conditions, which could lead to the selection of false or less-reliable events.

The following points give an overview of the rationale behind these choices:

a) 30-minute window

A fixed period of time is needed to report snowfall, accounting for the fact that snowfall intensities could be very low and well below the sensitivity of a weighing gauge used in the FWRS. A consistent approach is needed to compare events from different sites in various climate regimes. A period of 30 minutes is short enough to allow for stable conditions (temperature, wind speed, etc.) during the event and long enough to be reliably representative of snowfall events in a variety of climate and environmental conditions. Furthermore, 30-minute periods offer a good balance between the significance of events and the number of events detected.

When conducting the intercomparison analysis, other intervals were evaluated (e.g. 1, 3 or 6 hours) recognizing the needs of various sensors and applications such as tipping bucket gauges, snow on the ground, and light precipitation studies.

b) Net precipitation duration ≥ 60% of time

It was decided that precipitation does not need to be continuous during events, provided all other conditions are met. The detection of precipitation for 60% of the time was declared as a reasonable threshold to account for sufficient precipitation occurrence.

c) Reference accumulation ≥ 0.25/0.5 mm

The accumulation threshold needs to be sufficiently large to ensure that only genuine events are reported. The choice of a 0.25 mm accumulation threshold for 30-minute event intervals is based on previous work, illustrated in Figure 3.79. Using 30 minute periods identified by a present weather detector as snow, the effects of varying the minimum 30 minute precipitation threshold on the number of events identified (Figure 3.79a), and on the standard deviation (Figure 3.79b) and standard error (Figure 3.79c) of accumulation reports relative to the reference configuration are demonstrated. The standard error reached a minimum around 0.25 mm. The number of events with this accumulation was still very high, while the standard deviation had decreased significantly relative to events using smaller threshold values.

The threshold for 30-minute events that are selected based only on the weighing gauge accumulation was chosen to be more conservative (0.5 mm) to minimize the potential for the detection of false precipitation events.

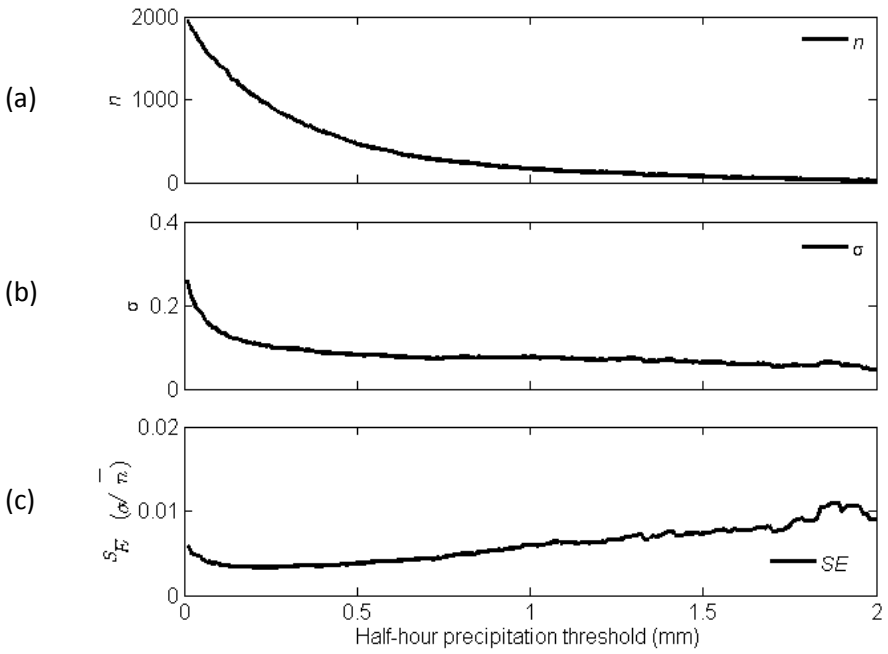


Figure 3.79. Accumulation threshold analysis for 30-minute precipitation events.

3.4.2 Event flags

In addition to data parameters, the SEDS files include an event flag for each parameter to inform of events that may be less reliable due to poor data quality or due to the way the event was selected. The event flag approach is outlined in Table 3.10. These flags are composed of two appended flags, one being an aggregation of the one minute quality control flags (as defined in Section 3.3.2.4), and the other being the flag produced by the event selection algorithm (as defined in Section 3.4.2). The aggregation of the one minute QC flags could be equal to 1, 2 or 3 depending on the percentage of "good" QC flags reported during the 30-minute event. The flag coming from the event selection

procedure, indicating which option was used to select the event (i.e. with or without a precipitation detector), is appended to the flag described above whenever it is equal to 1. That is, when the event was selected using only the reference weighing gauge data because of an unreliable or malfunctioning precipitation detector, the resulting flag values are 11, 21 or 31.

Each parameter in the SEDS file is accompanied by its corresponding 30-minute event flag.

Table 3.10. SPICE data quality flagging system for precipitation event files.

Flag value	Data Classification	Data Characterization
1	'Good'	Number of 1 minute datapoints with QC flag = 1 > 80%
11	'Good/no precip detector'	Same as 1, but event selected without precipitation detector
2	'Suspect'	60% < number of 1 minute datapoints with QC flag = 1 < 80%
21	'Suspect/no precip detector'	Same as 2, but event selected without precipitation detector
3	'Doubtful'	Number of 1 minute datapoints with QC flag = 1 < 60%
31	'Doubtful/no precip detector'	Same as 3, but event selected without precipitation detector

3.4.3 SLEDS and SNEDS

The SEDS contain 30-minute intervals for which there is a high level of confidence that precipitation occurred. These datasets enable the analysis of sensor performance in precipitating conditions. As a result of the specific event selection criteria used (see Section 3.4.1.3), 30-minute intervals characterized by light precipitation, or during which no precipitation occurred, are not identified. Additional investigations in SPICE focus on sensor performance during these light precipitation and non-precipitating periods. To accommodate these needs, two additional event files were produced for each site and each season: the Site Non-Event DataSet (SNEDS) accounting for 30-minute intervals over which no precipitation occurred, and the Site Light-Event DataSet (SLEDS) comprising all the remaining 30-minute intervals not identified by the SEDS and SNEDS. The format of these two files was exactly the same as the SEDS, since they are also based on 30-minute intervals. The criteria used to create SEDS, SLEDS and SNEDS are summarized in Table 3.11. The three files were computed for each S1 and S2 SPICE site and winter season.

Table 3.11. Criteria used for selecting events for the Site Event Datasets, Site Light Event Datasets, and Site Non-Event Datasets. "Ref Acc" refers to the accumulation of the FWRS weighing gauge, "PrecipDet_Y" to the number of minutes with precipitation detected by the precipitation detector, and "Nb_Ref_Acc_Y_min" to the number of minutes of increasing accumulation from the FWRS weighing gauge. The "flagged"' conditions refer to the selection of events performed without a reliable or existent precipitation detector.

	SEDS **Site Event DataSet**	**SLEDS** **Site Light Event DataSet**	**SNEDS** **Site Non-Event DataSet**
30 min Event Criteria	Not flagged : Ref Acc ≥ 0.25 mm PrecipDet_Y ≥ 18 min Flagged : Ref Acc ≥ 0.5 mm Nb_Ref_Acc_Y_min ≥ 18 min	Not flagged : 0 < Ref Acc < 0.25 mm PrecipDet_Y ≥ 1 min Flagged : 0 < Ref Acc < 0.25 mm Nb_Ref_Acc_Y_min ≥ 1 min	Not flagged : Ref Acc ≤ 0.05 mm PrecipDet_Y = 0 min Flagged : Ref Acc ≤ 0.05 mm Nb_Ref_Acc_Y_min = 0 min

3.5 Data Archives

Authors: Audrey Reverdin, Michael Earle, Andy Gaydos, Craig Smith

A comprehensive data archive was established, providing a first level of data quality control and monitoring the data for inconsistencies, missing data, upload times, file sizes, etc.

The data archive was proposed, hosted and operated by the National Center for Atmospheric Research (NCAR) in Boulder, CO, USA. Prior to NCAR archive upload, each site manager was required to maintain a local archive of all data collected, including the first level of data, metadata, site logs, maintenance logs, and other information relevant to SPICE. From this local archive, the data were formatted following predefined SPICE data format requirements, consisting of the development of daily ASCII files with data of six seconds or one-minute frequency (i.e. 14,400 or 1,440 records per file/day, respectively) with a file naming convention to distinguish between the different sites and instruments. The data were then transferred automatically or manually (via flash drive, removable hard drive, DVD, etc.) to a central location linked to the Internet, allowing data to be delivered by FTP to a server and, finally, propagated to the NCAR central archive. Each site had a unique login and password to access the FTP site to ensure that only the site managers could upload data for their site. The data archival system monitored the FTP site for new data and parsed the data into a MySQL database. Once the data were in the MySQL database, the raw data files were transferred to a permanent archive. At this stage, SPICE data (except SoG data) were quality-controlled to level 2b according to the described in Section 3.3.2.1. SoG data were quality controlled separately from the NCAR system, as described in Section 3.3.3.

Once the data were parsed and quality-controlled, they could be accessed through a user-friendly webpage interface. The following tasks can be done directly from the web interface: plotting data, viewing/adding site logs and maintenance notices, viewing/uploading photos, and downloading data (Figure 3.80.). Raw and quality-controlled data can be accessed and downloaded from the MySQL database (Figure 3.81.). The “Available Data” button allows checking and viewing of periods where data is available in the database.

The selected data can be downloaded as ASCII files and opened in any text editor, Microsoft Excel, or any plotting/coding program (Figure 3.82) for analysis.

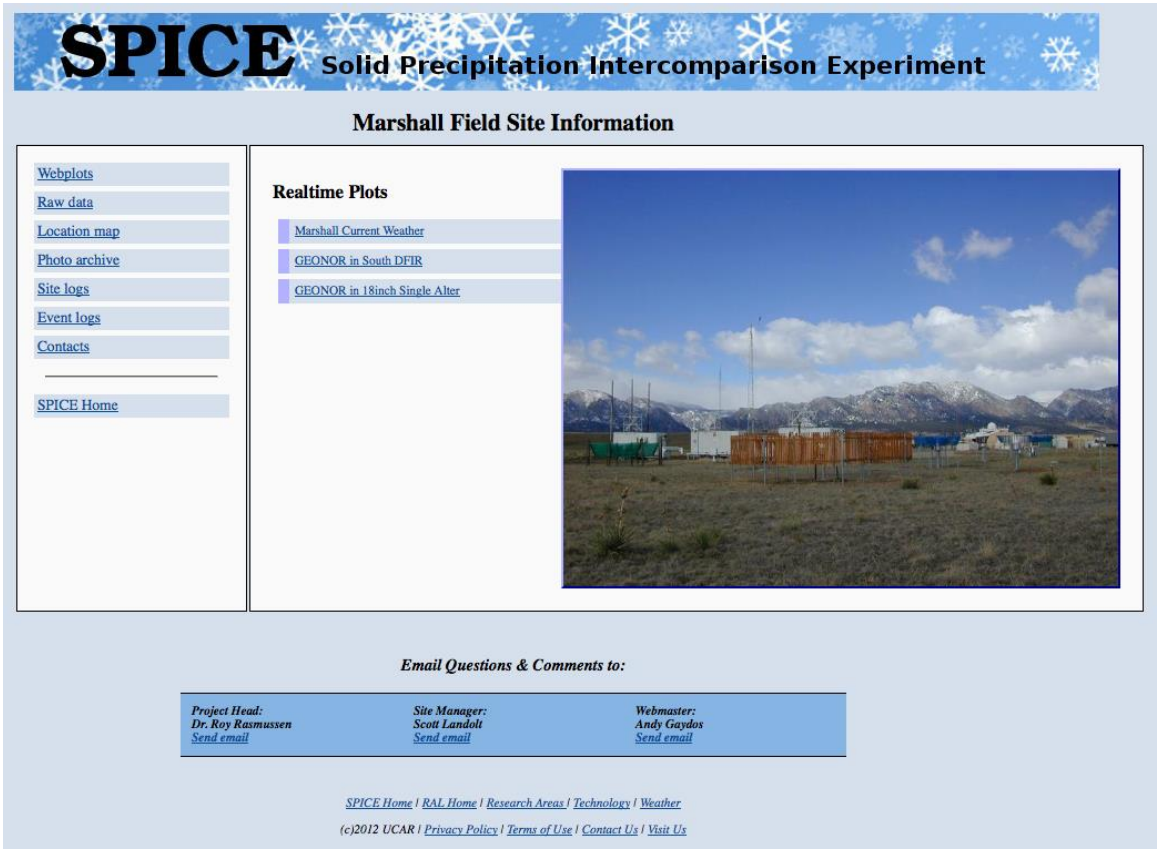


Figure 3.80. The NCAR website page to access and download SPICE data.

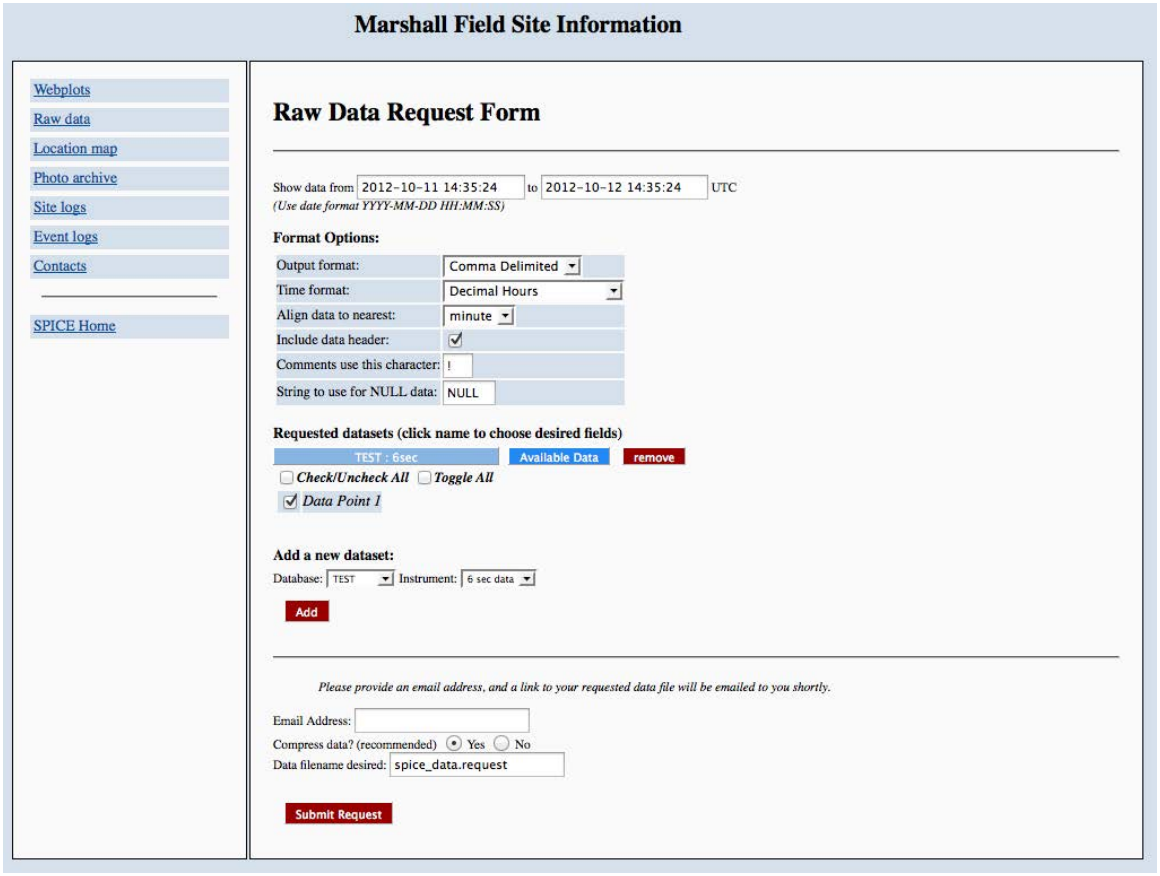


Figure 3.81. Raw data request form on the NCAR webpage.

```
!-------------------------------------------------------------------------------------------------------
!-- Raw Data output
!--   Date range: 2012-10-11 14:45:00 through 2012-10-12 14:45:48 UTC
!--
!-- (c) 2012 NCAR
!-------------------------------------------------------------------------------------------------------
!    DATABASE  DATASET            FIELD DESCRIPTION                                          UNIT
!=======================================================================================================
!1)           Year                                                                          YYYY
!2)           Month                                                                         MM
!3)           Day                                                                           DD
!4)           Dechrs                                                                        HH.HH
!5)  marshall  CURR_WEATHER        2m CS500 Temperature                                       C
!6)  marshall  CURR_WEATHER        2m CS500 Relative Humidity                                 %
!-------------------------------------------------------------------------------------------------------
!Year, Month, Day, Dechrs, AveTemp, AvgRH
2012, 10, 11, 14.750, -106.860, -73.699
2012, 10, 11, 14.767, -106.840, -73.695
2012, 10, 11, 14.783, -106.830, -73.685
2012, 10, 11, 14.800, -106.830, -73.677
2012, 10, 11, 14.817, -106.820, -73.674
2012, 10, 11, 14.833, -106.830, -73.679
2012, 10, 11, 14.850, -106.840, -73.687
2012, 10, 11, 14.867, -106.820, -73.687
2012, 10, 11, 14.883, -106.810, -73.687
2012, 10, 11, 14.900, -106.820, -73.687
2012, 10, 11, 14.917, -106.790, -73.676
2012, 10, 11, 14.933, -106.800, -73.681
```

Figure 3.82. Example of a dataset downloaded from the NCAR webpage.

3.5.1 Precipitation data

Precipitation and ancillary data from all SPICE sites were transmitted by the site managers to the NCAR archive. The Data Analysis Team (DAT) worked together with the site managers to ensure that the data were correctly transferred. Some issues related to periodic data logger desynchronization or data resolution were identified and resolved to ensure the quality of the SPICE dataset.

The availability check was essential to keep track of what happened at each site, and ensured the consistency of the SPICE dataset. It was a recursive process during the project, as new instruments were installed or removed along the way, or the format of the data files sent was changed (e.g. due to a new site manager, changes on the test field due to other projects, etc.). Therefore, some project managing tools were used to check the SPICE data availability and ensure the correct ingestion of data files at NCAR, and for reporting this information to the whole SPICE team. A comprehensive spreadsheet file containing all relevant metadata was created and periodically updated during the project in order to track the progress of data transfer.

Once the data were correctly uploaded, two seasonal time series plots for each precipitation gauge were produced and shared with the site manager: the raw versus quality-controlled data series together with the corresponding automatic QC flags (Figure 3.83a) and the quality-controlled data relative to the field working reference (Figure 3.83b), both with wind speed and temperature time series. These plots were meant to provide a sense of what happened on site during each winter season and to see how the sensor under test behaved as compared to the reference. Note that, at this point, manual quality control as described under Section 3.3.2.3 had not yet been applied.

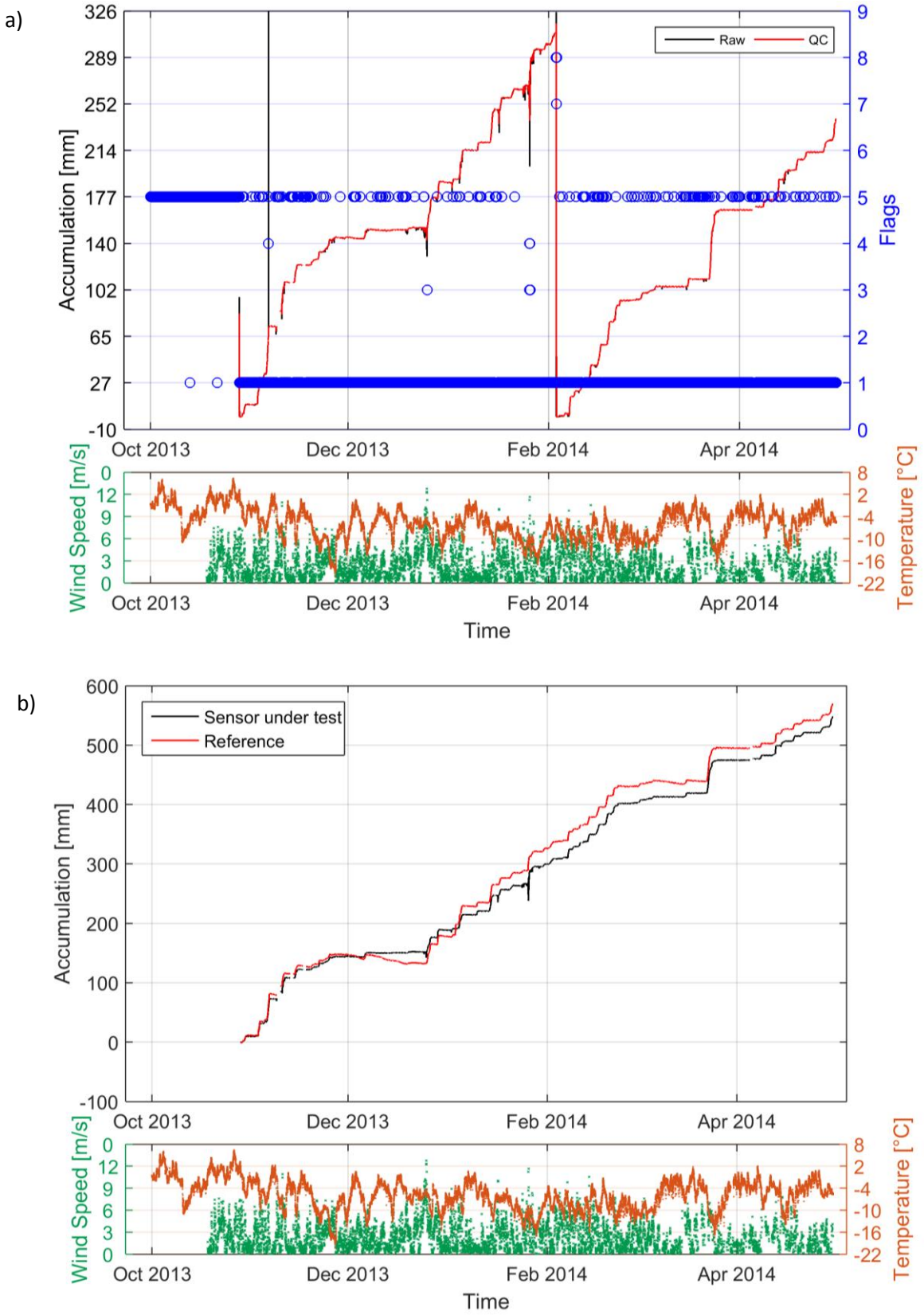


Figure 3.83. Time series plots of data from a weighing gauge under test illustrated as (a) raw versus quality-controlled datasets and associated QC flags and (b) its quality-controlled dataset relative to the field working reference gauge. Manual QC of data had not been applied at this stage. Both bottom panels represent ancillary wind speed and temperature measurements on the site.

Raw and automatically quality-controlled precipitation and ancillary data and associated flags (SPICE level 1 and 2b datasets, respectively; see Section 3.3.1) are available for download from the SPICE website. The data analysis team retrieved these data, conducted manual quality control as described in Section 3.3.2.3, archived them offline, and shared the ready-to-use final one-minute quality-controlled datasets with the SPICE team. The files are provided in MATLAB format with the following structure: time; automatic QCed data; automatic flags; manually QCed data; manual flags; and dates and indices where manual changes were done. The files are named according to the following template:

"SiteName_PcpInstrumentName_Configuration_Parameter_*Output*_Season.mat"

Where:

- **SiteName**: Name of the site, e.g. Formigal, Haukeliseter, Marshall,…
- **PcpInstrumentName**: Name of the instrument provider and model of the sensor, e.g. Geonor_1000, Campbell_PWS100, Pluvio^2,…
- **Configuration**: Configuration of the instrument on site (DFIR, Single-Alter - SA, unshielded - UN) and if it was a reference (R2, R3), a sensor under test provided by the manufacturer (SUT) or a SUT provided by the site host (UTsite), e.g. R2_DFIR, R3_SA, SA_UTsite,…
- **Parameter**: Precipitation parameter involved (accumulation – Acc, cumulative accumulation – CumAcc, BucketRT, AccNRT) and the wire or average of wires indication for Geonor and Belfort gauges, e.g. Accum_wire1, AccNRT_CumSum, TotAcc, …
- **Season**: SPICE winter season involved, e.g. 2013/14 or 2014/15

Ex: *Haukeliseter_Campbell_PWS100_UT_CumAcc_Output_2013-2014.mat*
CARE_CAE_PMB25R_UT_Int_CumSum_Output_2014-2015.mat
CARE_Pluvio2_BelfortDA_HN_UTsite_AccNRT_CumSum_Output_2014-2015.mat
CARE_R2_Geonor600_DFIR_Accum_WireSW_Output_2014-2015.mat
Weissfluhjoch_R2_PluvioDFIR_BucketRT_Output_2013-2014.mat
CaribouCreek_Geonor-600_R3-SA-C4_Accum-Avg_Output_2014-2015.mat

From these files, the SEDS, SLEDS and SNEDS were generated for each site and winter season. These event datasets were then used in subsequent analyses presented in this report. The event files were provided in both ASCII and Matlab formats to meet different user needs. The file naming convention is as follows:

"SiteName_*SEDS*_Season_TimeInterval.txt".

Where:

- **SiteName** and **Season** are the same as above;
- **TimeInterval**: Time interval on which the events were selected (only mentioned when different than the SPICE standard 30minute time interval).

Ex : *CARE_SEDS_2013-2014.txt*
Bratts Lake_SLEDS_2014-2015_360min.txt
Sodankyla_SEDS_2014-2015_60min.txt

The corresponding MATLAB files are named in a similar manner, with "_MATLAB" and ".mat" extensions.

3.5.2 Snow-on-the-ground data

The SoG data collected at SPICE sites were stored locally and transmitted to NCAR for archival. As described above, quality control of the SoG data was done offline, with data from all stages of the QC process preserved. The manual measurements of snow depth and SWE were also quality controlled and archived offline. The final versions of the SoG data, after the completion of all QC, will be returned to a central and accessible archive for perpetuity alongside the rest of the SPICE data archive.

Archived data includes:

- Raw data obtained from the site managers or the NCAR archive (.dat)
- Time-consistent raw data reformatted into monthly files in space delimited ASCII format (.qc0), aggregated to one-minute resolution where required
- Phase 1 quality-controlled data with automated removal of outliers with flags (.qc1)
- Phase 1 quality-controlled data manually inspected, with outliers removed and flagged during the manual process (.qc1C)
- Phase 2 quality-controlled data, where a zero-snow-depth offset correction is required (.qc2)
- Raw manual observations as provided by the site, usually as a text or Excel file, and a corresponding MATLAB workspace
- MATLAB scripts to read in the SoG quality-controlled data format
- Quality control documents describing data formats, QC processes, and instrument metadata

3.6 Methodology for the evaluation of the sensors under test

3.6.1 Instruments for precipitation measurements

Authors: Michael Earle, Mareile Wolff, Yves-Alain Roulet, Rodica Nitu

3.6.1.1 Instrument-specific SPICE objectives

The primary objective of the SPICE intercomparison is to assess and report on the performance of the currently available automatic systems used in operational applications for the measurement of solid precipitation (i.e. gauges as “black boxes”), covering:

- The ability of operational automatic systems to robustly perform over a range of operating conditions;
- The operational data processing and data quality-management techniques;
- The minimum practicable temporal resolution for reporting a valid solid precipitation measurement;
- The ability to detect and measure precipitation, including trace to light precipitation.

Additionally, recommendations on best practices and configurations for these measurement systems in operational environments, and on the achievable uncertainty of measurement for each of the systems reporting solid precipitation, are expected.

To meet these objectives, the sensors and systems under test (SUT) submitted by either (1) the SPICE host organizations, or (2) instrument manufacturers or distributors, were assessed for their ability to detect and to measure precipitation relative to the field reference. The assessment was based on results from tests in a field environment over two winter seasons, 2013/14 and 2014/15, conducted relative to the DFAR field reference system (reference R2) as configured on each SPICE site. Where available, systems were assessed in multiple configurations, e.g. with and without wind shields. The results were synthesized by instrument model and configuration for all sites where tested, and

provide an overall representation of the performance over the range of environmental conditions experienced. These results are presented in the Instrument Performance Reports (IPRs) in Annex 6. The methodology for the assessment presented in the IPRs is presented in the following sections.

3.6.1.2 Data derivation methodology

For this analysis, the following data derivation approach was used.

- The data used in the assessment had a 30-minute temporal resolution. This was adopted as the minimum practical interval over which a valid solid precipitation measurement can be made, and the analysis conducted verified this assumption.
- The reference data from the R2 (DFAR) system was derived using the SPICE precipitation event selection methodology, combining the accumulation reported by the weighing gauge in the DFAR with the indication on the presence or absence of precipitation from a precipitation detector or a disdrometer. A threshold of 0.5 mm/hour (or 0.25 mm/30 minutes) was applied to the R2 weighing gauge output. A second criterion was set on the precipitation detector located in the DFIR, with a minimum threshold of 18 minutes of recorded precipitation per 30-minute interval (see Section 3.4 for the description of the event selection approach). Additionally, for non-catchment-type instruments, the assessment of the SUT performance in “no-precipitation” conditions was conducted using a maximum threshold of 0.2 mm/hour (or 0.1 mm/30 minutes) for the R2 output, together with the precipitation detector showing 0 minutes of precipitation in the 30minute period. This condition was to avoid cases in which signal noise in the reference gauge data (typically under 0.1 mm/30 minutes) could be identified as precipitation.
- Table 3.12 summarizes the thresholds applied for each instrument type in categorizing 30-minute intervals as "precipitation events" or "no-precipitation events."
- The reference data set is considered the best estimation of the true precipitation amount during events, based on its traceability to manual measurements, as documented in this report.
- Precipitation phase thresholds were set as follows, using T_{min} and T_{max} during the 30-minute event to be characterized:
    - Liquid precipitation: $T_{min} \geq 2$ °C
    - Solid precipitation: $T_{max} \leq -2$ °C
    - Mixed precipitation: all remaining events not classified as liquid or solid
- This approach identified liquid and solid precipitation events with high confidence. The mixed classification had lower confidence and included events that were primarily liquid, events that were primarily solid, and/or transitions between the phases, as dictated by the temperature changes during the events. More variability in the assessment results was therefore expected for mixed precipitation events relative to liquid and solid events, but it must be noted that the classifications are not absolute and exceptions may occur depending on specific site conditions.
- The SUT data for the analysis were based exclusively on the SUT data output, quality controlled both automatically and manually using the procedures described in this report.
- For all sensors under test, the SUT 30-minute evaluation data were derived as the change in the equivalent amount reported by the sensor over the respective interval. No additional processing was applied. As the gauges tested use different operating principles (e.g. frequency of vibrating wire, bucket tips), some making available only their raw data (e.g. Geonor T200-B3) while others include more advanced, processed data products in the

output message (e.g. Pluvio^2, TRW405, CAE tipping bucket, disdrometers), the approach taken was meant to provide a simple and consistent treatment for all data. Different data processing techniques would yield different results, and advanced algorithms tailored to a sensor would address artifacts specific to each sensor.

- For the evaluation of all weighing gauges tested, a threshold was applied for the 30-minute precipitation amount, which was selected to be similar to that used for the weighing gauge in the reference system, R2, of 0.5 mm/hour or 0.25 mm/30 minutes (see Table 3.12). The use of the same threshold for all weighing gauges, similar to that used for the derivation of the reference dataset, ensured a consistent method of treatment for the data included in the study.
- For the evaluation of tipping buckets, no specific threshold was applied as the size of the bucket is in itself a threshold for the derived measurement (see Table 3.12).
- For the evaluation of the non-catchment systems, no threshold was applied, as their principles of operation lead to much higher sensitivity for these instruments, which was explored as part of this intercomparison (see Table 3.12).
- The performance report prepared for each instrument model (Annex 6) includes a summary of the combined environmental conditions (air temperature, relative humidity, wind speed, wind direction, reference precipitation rate and precipitation occurrence) for all sites where the respective instrument was tested, to provide an integrated view of the testing conducted.

Table 3.12. Thresholds used to differentiate precipitation events from no-precipitation events over 30-minute intervals, applied to the R2 reference and SUT data for the three categories of instrument types.

	Precipitation Events	No-Precipitation Events
	Reference	
Weighing gauges	R2 reference gauge Acc ≥ 0.25 mm Precip Detector recording ≥ 18 min of precip	All other cases
Tipping Bucket gauges		
Non-Catchment Type Instruments (including Hotplate)		R2 reference gauge Acc < 0.1 mm Precip Detector recording 0 min of precip
	SUT	
Weighing gauges	SUT Acc ≥ 0.25 mm	All other cases
Tipping Bucket gauges	SUT Acc ≥ SUT reporting resolution [mm]	
Non-Catchment Type Instruments (including Hotplate)	SUT Acc > 0 mm	

3.6.1.3 Evaluation of the ability to perform over a range of operating conditions

The ability of an instrument to perform over the range of operating conditions was evaluated based on the comparison with the reference data and reported using several skill scores as outlined below. This was a qualitative assessment that was interpreted within the context of the methodology outlined in this section, for each instrument type.

As the wind speed influences the amount of precipitation collected by a catchment-based sensor (weighing gauges or tipping buckets), the use and value of an accumulation threshold could

segregate events as precipitation or not, depending on whether or not the threshold was reached over the corresponding interval. This resulted in cases in which, for the same event, the reference reached its threshold and reported an event, while the SUT did not. The analysis conducted for each instrument helped each user to make appropriate decisions on the data treatment and use.

3.6.1.3.1 *Skill scores*

In the comparison of two signals with two distinct values, for instance using presence and absence of precipitation, a contingency table, as shown in Table 3.13 below, captured all possible outcomes: **x** represented the number of instances in which the reference and the SUT agreed that precipitation was present according to the methodology used (hits); **y** was the number of instances in which the reference indicated that precipitation was present while the SUT indicated its absence (misses); **z** was the number of events when the SUT reported a precipitation event while the reference did not (false alarms); and, **w** was the number of events when the reference and the SUT agreed that precipitation was absent (correct negatives). The score methodology used for this analysis was similar to that used in forecast verification, and was meant to represent qualitatively the performance of the SUT as tested in various climate conditions, based on the specific event selection criteria imposed (see Table 3.11). The same method was used for the assessment of a range of precipitation detectors and laser-based instruments reporting precipitation type (Sheppard, B and Joe P., 2000; Griesel et al, 2012).

These scores did not quantitatively assess the amount of precipitation reported by the SUT relative to the reference; rather, they represented the likelihood that the SUT would emulate the field reference in detecting precipitation in the given conditions, within a prescribed time interval. For tipping bucket gauges, which are subject to response delays relative to the reference (particularly for solid precipitation, which must be melted in the gauge funnel prior to measurement), skill scores provided insight into the timeliness of gauge reports in operational settings, with increased likelihood of misses and false alarms on account of the different principles of operation between the reference and SUT.

This information provides important insight on the performance of the SUT when operating without the concurrent presence of a field reference, i.e. how reliable the given sensor was in detecting precipitation.

The quantitative assessment of SUT relative to the field reference is assessed separately (Section 3.6.1.4).

Table 3.13. Contingency table for precipitation detection.

	Reference Precipitation	**Reference No-Precipitation**	**Total**
SUT Precipitation	x (hits)	z (false alarms)	x + z
SUT No-Precipitation	y (misses)	w (correct negatives)	y + w
Total	x + y	z + w	N = x + y + z + w

Based on the contingency table developed for each SUT on each site, the following skill scores were considered:

- Probability of detection (POD)
- False Alarm Rate (FAR)
- Bias (B)
- Heike Skill Score (HSS)

These scores are described in detail in the following sections.

POD is defined as:

$$POD[\%] = \frac{hits\ (x)}{hits(x) + misses(y)} \times 100$$

The POD gives the percentage of events, out of all the precipitation events as indicated by the reference, which will also be reported as precipitation events by the SUT. In other words, it gives the probability of the SUT agreeing on the occurrence of precipitation when the reference detected precipitation. Ideally, POD would have a value of 100%.

FAR is defined as:

$$FAR[\%] = \frac{false\ alarms\ (z)}{hits\ (x) + false\ alarms\ (z)} \times 100$$

The FAR is the percentage of precipitation events, as reported by the SUT, for which the reference data did not meet the precipitation event criteria. The FAR gives an indication of how likely it is that the sensor is not reliable when it reports the occurrence of precipitation. A larger percentage for the FAR implies a higher probability that the SUT does not recognize precipitation events in a similar manner as the reference. Ideally, the FAR would be zero.

Another measure is the Bias (B):

$$B[\%] = \frac{hits\ (x) + false\ alarms\ (z)}{hits\ (x) + misses\ (y)} \times 100$$

The bias is the ratio of the number of precipitation events, as reported by the SUT, to the number of precipitation events as reported by the reference. If B = 100%, the sensor is unbiased, meaning that the SUT detects the same number of precipitation events as the reference. However, this measure gives no information on whether the reports by the reference and sensor correspond in terms of reported precipitation amount. If B < 100% (or B >100%), the sensor “under-detects” (or “over-detects”) precipitation events relative to the reference.

HSS is defined as (Sheppard and Joe, 2000):

$$HSS[\%] = \left[\frac{(C - E)}{(N - E)}\right] \times 100$$

Or

$$HSS = \frac{2(xw - yz)}{[y^2 + z^2 + 2xw + (y + z)(x + w)]}$$

where C is the total number of correct reports of precipitation and of no precipitation from the SUT (C = x + w); E is the expectation value for the number of correct reports that would be achieved by

random guesses (with the constraint that the marginal distributions in the resultant contingency table agree with that of the actual dataset); and N is the total number of reports (x + y + z + w).

An HSS value of 0 corresponds to a sensor that has no skill, while a sensor that is always correct has an HSS of 100% (or "1"). Negative HSS values indicate that the sensor shows less skill than a random draw.

3.6.1.3.2 *SUT noise level assessment*

Ideally, when no precipitation occurs, the response of a sensor should be zero. In practice, the sensor output (either raw data or processed data using internal algorithms) is not zero and is influenced by the sensor's inherent characteristics, temperature variations, wind speed, solar irradiance, etc. (Duchon, 2004; Nemeth, 2008).

By evaluating the sensor response in the absence of precipitation, the instrument performance reports provide recommendations on how to interpret and process the response to limit the likelihood of false reports. The results also provide insight into the uncertainty of measurements and the influence of the specific accumulation threshold used in precipitation event selection.

For this evaluation, a subset of the correct negative cases was used. The cases selected corresponded to 30-minute intervals when the precipitation detector used as part of the reference system did not detect any precipitation. This condition ensured that only those events having a high degree of confidence that there was no precipitation were used. The response signal of the SUT is assessed as "sensor noise."

The results were expressed in terms of the average, standard deviation, and extreme values (min, max) of the SUT output. For weighing gauges, additional insight is provided by examining the variability of this signal as a function of the observed variation in air temperature over the interval and the corresponding wind speed.

For each SUT, a recommendation is made on the minimum threshold that could be used for operational applications. It is equal to three times the standard deviation of the average level of noise determined during the test, and calculated based on the integration of the results from all sites where the gauge is tested. This value is consistent with the goal of ensuring that over 99% of the sensor data meets the precipitation event criteria, and that the likelihood of false precipitation is below 1%.

3.6.1.4 Evaluation of the ability to measure precipitation

3.6.1.4.1 *Evaluation of the SUT ability to measure and report precipitation*

The ability of each sensor under test to measure and report precipitation was examined based on the cases when both the reference and the SUT reported a precipitation event, independently, i.e. "hits" as defined above (yes/yes, or YY cases). The results were expressed graphically and analytically.

Scatter plots and box plots of the catch ratio (SUT accumulation divided by reference accumulation, over a given 30-minute interval) were derived as a function of horizontal wind speed and air temperature for each configuration of the SUT, for each site. Additionally, accumulation-accumulation plots were generated for each instrument configuration and site, with the results stratified by precipitation type. Precipitation type was assessed using the temperature measurements during the interval, as outlined in Section 3.6.1.2.

For the evaluation of non-catchment-type instruments, additional graphs were generated, including time series (representing the cumulative sum of 30-minute YY events) and histograms and box plots of the catch efficiency. These were produced to highlight and characterize the behavior of such instruments for different types of precipitation (rain, mixed or snow). As the wind direction is likely to impact the measurement of non-catchment-type instruments due to their anisotropic shape, the potential influence of the orientation of the instrument relative to prevailing wind directions during precipitation events was also assessed using wind roses.

Analytically, the ability of a sensor under test to measure precipitation relative to the field reference was expressed using the root mean square error (RMSE), also known as operational comparability as defined in ASTM 4430 (2015), Standard Practice for Determining the Operational Comparability of Meteorological Measurements. The RMSE of the difference between simultaneous readings from two systems measuring the same quantity in the same environment is defined as:

$$RMSE = \pm \sqrt{{}^{1}/_{n} \sum_{i=1}^{n} (Xai - Xbi)^2}$$

Where:

- *Xai* = ith measurement reported by the reference system;
- *Xbi* = ith measurement reported by the system under test;
- *n* = number of data samples used for the evaluation.

The RMSE was calculated using YY cases over the entire dataset for each precipitation type, gauge type, configuration, and site tested.

The RMSE is a measure of the uncertainty of measurement for the instrument tested relative to the reference used. It provides a quantified measure of the quality of the precipitation amount data as reported by an operational sensor, and specific to the shield configuration used and site conditions when no analytical adjustments (e.g. transfer functions) were applied to measurements.

Complementary to the derivation of the RMSE, an estimate of the overall catch ratio is provided for each instrument for the two seasons of the intercomparison (or for all data available within the two measurement seasons). Catch ratios quantify the overall agreement between the SUT and reference for intervals during which they both report precipitation. These values are specific to how the datasets were derived, and are provided as guidance for the level of error expected for the seasonal precipitation values, including all types of precipitation experienced in a specific climate. A separate section of this report presents recommendations for the derivation and use of transfer functions to adjust the instrument measurements for the estimation of “true” precipitation amounts (see Section 3.7).

To characterize the influence of threshold selection on weighing gauge performance (see Section 4.2), the overall catch ratio was estimated for two threshold levels applied to the SUT dataset: the standard threshold of 0.25 mm/30 minutes and a lower threshold of 0.1 mm/30 minutes.

For tipping bucket gauges, the overall catch ratio based solely on YY cases is less representative, as it does not account for response delays. For example, the reference may record precipitation during a given interval before the tipping bucket responds (YN cases), or the tipping bucket may respond during a later interval when the reference no longer observes precipitation (NY cases). For this reason, overall catch ratios determined for all YY, YN, and NY cases are computed for tipping bucket

gauges, and are considered to be more representative than the catch ratios determined for YY cases only.

3.6.1.4.2 *Detection of light precipitation events by weighing gauges: threshold selection*

For a specific instrument, the impact of the selected threshold was assessed by using a subset of the dataset for that instrument (e.g. one site, all data) to generate contingency tables (

) for four different threshold cases: case 1, threshold equal to 0.25 mm/30 minutes; case 2, threshold of 0.1 mm/30 minutes; case 3, no threshold applied to the SUT data; case 4, threshold of 0 mm applied to the SUT data. The changes in the POD and FAR were used to gauge the appropriateness of a given threshold. As noted above, a POD close to 100% and a FAR close to 0% are ideal.

Recognizing the need for a reasonable balance between the detection of light and very light events and minimizing the risk of false reports, the goal is to find a “sweet spot” where a given gauge in different conditions would correctly detect precipitation (especially when light) and would have the minimum likelihood of falsely reporting precipitation. This information is important, as most of the operational stations use only one precipitation gauge, and the ability to verify one measurement against an independent reference is not available.

3.6.1.4.3 *Assessment of events when the reference and the SUT do not agree on the occurrence of precipitation*

The cases identified as Misses and False Alarms in the contingency tables (see

) correspond to those cases when the reference dataset and the SUT data did not agree on the presence or absence of precipitation during the same interval. For these cases, the SUT data were further examined in an attempt to understand whether the data derivation method or other factors contributed to the noted differences.

It is acknowledged that the reference system is able to catch more precipitation than a SUT, given the use of the DFIR-fence as part of the reference system. For smaller diameter shields, or no shield at all, the ability to capture precipitation is diminished, thus the same gauge in different shield configurations would detect precipitation differently, primarily as a function of wind speed. Therefore YN and NY cases for weighing gauges were characterized using histograms of the reference and SUT amounts, and average wind speed. Additional parameters relevant to the assessment and interpretation of sensor performance were considered for non-catchment-type instruments and tipping bucket gauges.

3.6.1.4.4 *Characterization of response delays for tipping bucket gauges*

Response delays for the measurement of precipitation by heated tipping bucket gauges were assessed by determining the time elapsed between the onset of precipitation as determined by the reference configuration, and the first tip recorded by the tipping bucket. Since these delays can extend beyond the 30-minute periods considered in this analysis, longer periods were considered for this assessment. These longer "tipping bucket comparison events" (TBCEs) consist of one or more consecutive 30-minute precipitation periods as identified by the reference configuration (≥ 0.25 mm precipitation, ≥ 18 minutes of precipitation identified by precipitation detector), followed by at least 180 minutes without precipitation. This extended period without precipitation is intended to allow additional time for the melting and recording of precipitation by heated tipping bucket gauges.

An example of a response assessment event illustrating a tipping bucket response delay is shown in Figure 3.84 below. Note that the event duration is truncated to more clearly depict the response delay.

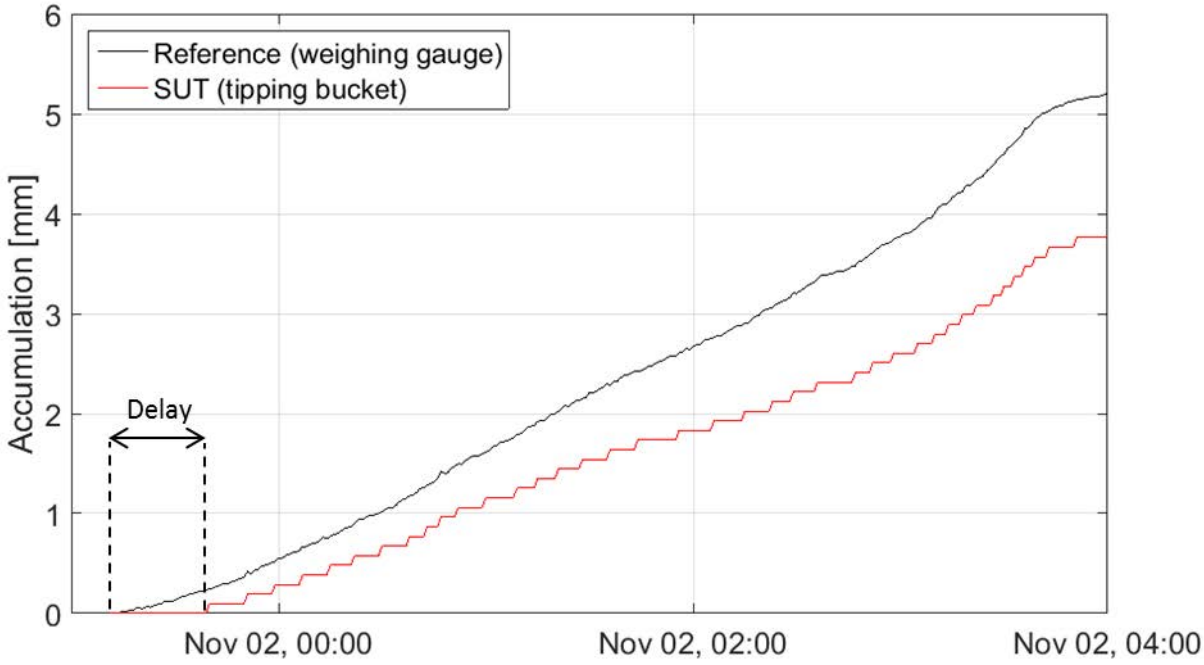


Figure 3.84. Example of tipping bucket comparison event and response delay.

Probability distribution functions were generated from the compiled delay times from all TBCEs and used to characterize the response delays for each tipping bucket gauge, at each site.

3.6.1.5 Interpretation of results

The integration of results for a specific SUT from different sites, where available, enables the evaluation of the sensor in multiple conditions (climatology, site configuration, local conditions).

The results provided information on the variability of results between sites relative to the parameters noted above. Operational issues that require additional configuration considerations were also noted.

3.6.2 Instruments for snow-on-the-ground measurements

Authors: Craig Smith, Samuel Morin

3.6.2.1 Instrument-specific SPICE objectives

The objective of the Instrument Performance Reports for the SoG sensors is to assess and document the capability of these sensors to measure snow depth and SWE as compared to a defined reference. The measurement reference(s) for SoG measurements are outlined in Section 3.1.4.2.

Following a graphical and statistical intercomparison of the SUT with the reference(s), the data quality metrics are summarized, and comments and recommendations on best practices and lessons learned from the SPICE community are provided. The objective is to capture the collective experience

with sensor installation, configuration, and operation in a variety of measurement environments that may be of interest to instrument and data users.

The performance of sensors under test, as submitted by host organizations and/or manufacturers, was assessed based on tests conducted over two winters at the participating intercomparison sites. The results of these tests were synthesized by instrument for all sites tested, and provide an overall representation of the performance over the range of environmental conditions experienced (see IPRs in Annex 6).

3.6.2.2 Data derivation methodology

For this analysis, the following data approach has been used:

- Generally, the data were collected at one-minute resolution. The snow-depth data at CARE were collected every 20 seconds and aggregated (via the median) to produce a one-minute data output. The SWE data obtained by the CS725 is output at six-hour intervals, while the SWE data obtained by the SSG1000 snow scale had a frequency of one measurement per minute.
- The data were quality controlled via a multi-phase process that involved the automated removal of non-reasonable outliers and the further removal of outliers via manual intervention. This process is described in more detail in Section 3.3.3. Only data determined to be "good" by the quality control process were included in the intercomparison.
- The manual reference dataset was derived from manual or photographic observation of graduated snow stakes in the case of snow depth, or bi-weekly bulk-density sampling for the measurement of SWE. To compare the lower-frequency manual measurements with the high-frequency automated measurements, the observation times were matched as closely as possible. For CARE, only the start time and end time of the daily snow survey was recorded, so the end time was used as the time stamp for these observations.
- The automated reference for snow depth, where applicable, was a mean of all available automated measurements at the site. The exception is CARE, where the automated reference is the mean of all of the automated measurements at each pedestal. The automated reference measurements, as with all of the automated measurements, were reported at one-minute resolution.
- Each instrument performance report includes the combined environmental conditions (air temperature, relative humidity, wind speed, wind direction, reference precipitation rate and precipitation presence) for all sites hosting a specific SUT.

3.6.2.3 Evaluation of the ability to perform over a range of operating conditions

The ability of a sensor to perform over a range of operating conditions was evaluated based on the comparison of simultaneous automated reference data and near-simultaneous manual reference data with SUT data, derived as defined above. The overall level of agreement was assessed with regression analysis, RMSE calculations, and visual intercomparison of seasonal time series. RMSE is calculated as follows:

$$RMSE = \pm \sqrt{1/n \sum_{i=1}^{n} (Xai - Xbi)^2}$$

where X_{ai} is the i^{th} measurement reported by the reference system, X_{bi} is the i^{th} measurement reported by the system under test, and n is the number of data samples used for the evaluation.

Interpretation of the intercomparison must take into account known issues with both the manual and automated reference techniques and consider the impact of spatial variability in both snow depth and SWE.

3.6.2.4 Interpretation of results

The integration of results for a specific SUT from different sites, where available, enables the evaluation of the sensor in multiple conditions (climatology, site management, local conditions).

Results in the performance reports provide information on the following parameters:

- Data processing and quality control which could be interpreted (with caution) as a measure of instrument reliability, factoring in circumstances causing data loss not related to the instrument.
- Variability of results between sites, seasons, and potential causes.
- Use of heating and mounting infrastructure and potential impacts on measurements of snow on the ground.

3.7 Transfer function development

Authors: John Kochendorfer, Michael Earle, Mareile Wolff, Audrey Reverdin

The weighing gauge transfer function development and testing using the SPICE measurements were presented in two separate manuscripts published in the WMO-SPICE special issue of *Hydrology and Earth System Sciences*, and in Wolff et al. (2015).

In Kochendorfer et al. (2017a), transfer functions were developed from the host-provided reference unshielded and single-Alter-shielded gauges available at eight WMO-SPICE sites with R2 reference configurations. This was done by combining measurements from all eight sites to create different types of multi-site, "universal" transfer functions that described catch efficiency as a function of wind speed (and air temperature). The use of multiple sites allowed for the creation of defensible transfer functions for use at all sites. Using these multi-site "universal" transfer functions, site biases and other error statistics were calculated for all eight individual sites.

In a separate manuscript (Kochendorfer et al. 2017b), transfer functions were developed for all manufacturer-provided weighing gauges. The 1500 mm Geonor transfer function was evaluated at Marshall, Bratt's Lake, Weissfluhjoch, and Caribou Creek; unshielded and single-Alter-shielded Sutron gauges were evaluated at Marshall; the unshielded and shielded MRW500 were both evaluated at Marshall and Bratt's Lake; and the unshielded TRWS 405 was evaluated at Marshall and Haukeliseter.

Gauge-specific corrections for the unshielded Sutron, MRW500, and TRWS 405 gauges did not perform significantly better than a multi-gauge, "universal" transfer function developed using the host-provided unshielded reference gauges at all eight sites (Kochendorfer et al. 2017a). Likewise, the manufacturer-provided single-Alter-shielded 1500 mm Geonor and Sutron gauges were correctable using the "universal'" single-Alter transfer function developed from the host-provided single-Alter-shielded reference gauges from eight sites (Kochendorfer et al. 2017a). The results indicated that transfer functions can be determined by the type of wind shielding, or the lack of wind shielding, rather than the type of gauge. The only manufacturer-provided gauge that required its own custom transfer function was the shielded MRW500, which was provided with a shield that was smaller than the single-Alter shield. It was found to be under-corrected by the "universal" single-Alter function and over-corrected by the "universal" unshielded correction.

In addition, transfer functions for the double-Alter, the Belfort double-Alter, and the small DFIR (SDFIR) windshields were developed in Kochendorfer et al. (2017b) using mainly host-provided gauges tested for national interests rather than specifically for WMO-SPICE. The double-Alter shield and the Belfort double-Alter shield were both evaluated at CARE and Marshall, and the SDFIR shield was evaluated only at Marshall. A comparison of all transfer functions developed is available in Figure 3.85. Both the uncorrected and the corrected measurements recorded within the more effective shields, such as the SDFIR and Belfort double-Alter, were less prone to errors, as determined by comparison to the DFAR. This was most notably observed with the Belfort double-Alter and the SDFIR shields, which required very little correction. The errors in the corrected measurements from these more effective shields were much smaller than the errors in the corrected unshielded and single-Alter-shielded gauges. These results are described in detail in Kochendorfer et al. (2017b), but they indicate that although adjustments can be used to effectively reduce biases in unshielded and other less well-shielded precipitation measurements, a DFIR-shield or another almost equally effective windshield is necessary for the most accurate measurement of solid and mixed precipitation.

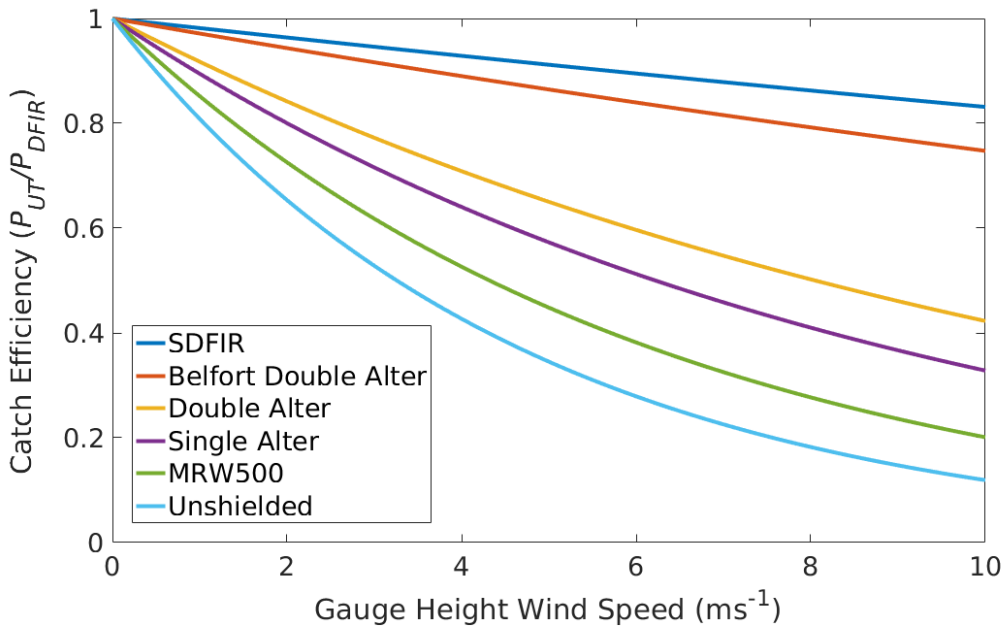


Figure 3.85. A comparison of the WMO-SPICE weighing gauge transfer functions, from Kochendorfer et al. (2017b).

4. RESULTS

This section presents a comprehensive evaluation of instruments tested, based on the IPRs (see Annex 6) and the methodology described previously (see Section 3.6). The instruments are classified into the following groups:

- Weighing gauges (Section 4.1.1)
- Tipping bucket gauges (Section 4.1.2)
- Emerging technologies (Section 4.1.3)
    - Non-catchment-type instruments (optical instruments)
    - Evaporative plates
- Snow on the ground (Section 4.1.4)

General considerations and recommendations are made by instrument type. The full details of the analysis for each SUT can be found in the IPRs.

A number of challenges and issues impacting the measurement of snow are also addressed. Besides the well-known effect of wind-induced errors, there are other external factors that can influence the quality of snow measurements. These factors can be technical or environmental, and may significantly affect the measurement of solid precipitation. The considerations discussed in Section 4.2 reflect experiences from the SPICE measurement campaign; the intent is to help operational services address external factors that may negatively impact their measurement of snow (snowfall and SoG). Among others, capping of gauges, heating configuration, use of antifreeze and oil in weighing gauges, vibrations of gauges under high-wind conditions, detection of light events, and technical issues for SoG measurements (e.g. mounting of sensors, target type) are discussed using examples from SPICE sites.

Besides the methodology developed within SPICE and the resulting SPICE analysis, there were a number of additional contributions that were site specific, applied alternate methods, or examined different instrument parameters (e.g. precipitation type). This additional work, which supported the SPICE analysis, has been collected and referenced in Section 4.2.6.

4.1 Evaluation of instruments

4.1.1 Weighing gauges

Authors: Michael Earle, Kai Wong, John Kochendorfer, Rodica Nitu, Audrey Reverdin

4.1.1.1 Introduction

The previous WMO solid precipitation intercomparison (Goodison et al., 1998) relied primarily on manual observations of snowfall. Since then, automated weighing gauges capable of measuring snowfall have become both widely available and widely used (Nitu and Wong, 2010).

Weighing gauges capture precipitation in fundamentally the same way as manual gauges – by collecting, or "catching," hydrometeors in a vessel. These are collectively referred to as "catchment-type" gauges. Weighing gauges differ from manual gauges in that that the amount of precipitation collected is automatically monitored and does not require a human observer.

Given the emergence of automated weighing gauges in operational networks, it is important to characterize their performance in different climate regimes to best inform their deployment. In WMO-SPICE, weighing gauges were tested globally at eight different sites (and climates) as depicted in Figure 4.1: Bratt's Lake (continental, high wind), CARE (humid continental), Caribou Creek

(southern boreal), Formigal (alpine with maritime influence), Haukeliseter (alpine, high wind), Marshall (dry continental), Sodankylä (northern boreal), and Weissfluhjoch (alpine with complex topography). Seven gauge models were tested, each in one or more wind-shield configurations (Table 4.1). The majority of test configurations were heated; however, this was not possible for all test gauges at all sites.

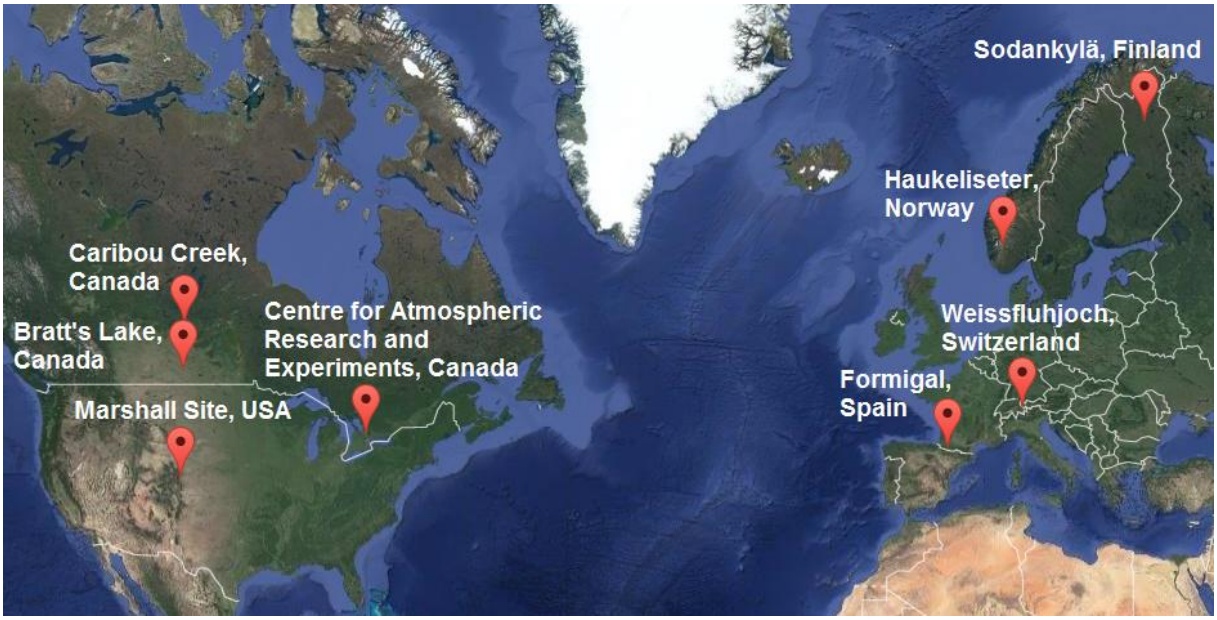


Figure 4.1. Sites hosting automated weighing gauges under test in WMO-SPICE.

Table 4.1. Weighing gauges under test in WMO-SPICE

Sensor (capacity)	Site(s)	Configurations tested*
Belfort AEPG 600 (600 mm)	Weissfluhjoch	Belfort double-Alter
Geonor T-200B3 (600 mm)	Bratt's Lake CARE Caribou Creek Marshall	Single-Alter, unshielded Single-Alter, unshielded Single-Alter, unshielded (unheated) Single-Alter, unshielded
Geonor T-200B3MD (1500 mm)	Bratt's Lake CARE Caribou Creek Marshall Weissfluhjoch	Single-Alter (unheated) Belfort double-Alter Single-Alter (unheated) Single-Alter Single-Alter
Meteoservis MRW500 (900 mm, as tested)	Bratt's Lake Marshall	Single-Alter/Tretyakov, unshielded Single-Alter/Tretyakov, unshielded
MPS TRwS405 (750 mm)	Haukeliseter Marshall	Unshielded Unshielded
OTT Pluvio^2 (1500 mm)	CARE Formigal Haukeliseter Marshall Sodankylä Weissfluhjoch	Single-Alter, unshielded Single-Alter, unshielded Single-Alter Single-Alter, Tretyakov, unshielded Single-Alter, unshielded Single-Alter, unshielded
Sutron TPG (914 mm)	Marshall	Single-Alter, unshielded

*All configurations are heated unless otherwise specified.

The performance of each test gauge, under the specific range of conditions experienced at each applicable test site, is detailed in the corresponding Instrument Performance Reports (Annex 6). The following section outlines key elements of the assessment presented in these reports, consolidating results for comparison across different gauge types, configurations, and sites. An overview of material relevant to the interpretation of the results, including background on the operation and performance of weighing gauges, as well as the methods employed in the WMO-SPICE assessment, is provided. The interpretation of results and related discussion are used to make recommendations for the selection, field deployment, and operation of automated weighing gauges.

4.1.1.1.1 *Fundamentals of precipitation measurement using weighing gauges*

The automated weighing gauges under test in WMO-SPICE (Table 4.1) employed two different methods to monitor the weight of accumulated precipitation. The Geonor T-200B3, Geonor T-200B3MD, and Belfort AEPG 600 gauges use vibrating-wire transducers that change frequency as the weight in the bucket changes. Each of these gauges employs three vibrating-wire transducers for redundancy; however, a single transducer is capable of monitoring precipitation independently. The Meteoservis MRW500, MPS TRwS405, OTT Pluvio^2, and Sutron TPG test gauges all employ a single-load cell transducer for weight measurements, similar to the strain gauges used in bathroom scales.

Weighing gauges measure the total amount of water accumulated in the gauge, irrespective of transducer type or precipitation phase. Precipitation amounts are calculated from the change in weight over a given time period. Fundamentally, this differs from manual and tipping bucket gauges that report only new or recent precipitation. This approach is advantageous: weighing gauges collect precipitation even when not operational, so seasonal totals are impacted less by site/gauge maintenance and power/data outages. Further, solid precipitation does not need to melt before being measured (as with tipping bucket gauges) or the collector removed for measurement elsewhere on the site (as with manual measurements). This results in more timely responses to incident precipitation. There are, however, two notable drawbacks to this approach: maintenance is required to empty the gauges periodically and an antifreeze/oil mixture is often required to prevent contents from freezing (see Section 4.2.3).

Heating is often applied at the gauge orifice and/or inlet to prevent snow capping and precipitation freezing on or within the inlet, both of which can delay or prevent incident precipitation from being collected and reported as discussed in Sections 4.2.2 and 4.2.3. The use of heating requires careful consideration of power requirements and the potential for heat to evaporate precipitation or disrupt the flow field above the gauge orifice (the “chimney effect”). Capping, evaporation, and chimney effects can impact seasonal precipitation totals reported by weighing gauges in spite of the advantage provided by measuring total accumulation.

Wind shields are used often to mitigate the effects of wind on the ability of weighing gauges to collect and report precipitation. Various configurations have been used in this regard (e.g. Goodison et al., 1998; Rasmussen et al., 2012). These generally comprise a single shield or two concentric shields (double shields). Within these broad categories, shield configurations may differ in the following respects: overall dimensions; mounting to or separate from the gauge post; slat dimensions; and the degree slats are able to rotate. Wind-induced undercatch is primarily a problem for solid precipitation, which has a lower density than liquid precipitation, and is, therefore, more likely to be deflected by wind away from the gauge orifice.

The specific data outputs and processing approaches employed by different gauges must also be considered. Some weighing gauges provide only the raw, unprocessed output of bucket weight; others produce a smooth and processed bucket weight and/or precipitation rate that may also help to mitigate the effects of gauge noise, temperature changes, and/or evaporation. Both approaches have advantages and disadvantages. On the one hand, smoothing and processing may be accompanied by delays in the reporting of precipitation and the potential to report false accumulation. On the other hand, unprocessed weighing gauge measurements may require a sophisticated algorithm to differentiate between noise and precipitation, remove the effects of evaporation, and/or accurately determine precipitation rates (e.g. Leeper and Kochendorfer, 2015; Leeper et al., 2015). Auxiliary sensors, such as a wetness sensor or an optical precipitation detector, can help to distinguish precipitating periods from gauge noise.

The evaluation of weighing gauge performance in different shield configurations and for different data outputs (if available) are important components of the SPICE assessment approach outlined in the following section. Each weighing gauge, in all applicable configurations and at all applicable test sites, is assessed in terms of its ability to detect and report precipitation relative to the reference configuration and in terms of how the reported accumulation values compare with those reported by the reference configuration. The relative effects of wind speed are investigated for each gauge type at all sites and for all gauge types tested at each site.

4.1.1.2 Assessment approach

The WMO-SPICE assessment approach considers the performance of a given gauge/shield combination relative to the R2 reference configuration (DFAR and precipitation detector) at the same site (see Section 3.6.1.2). The comparison is based primarily on 30-minute precipitation events during which the weighing gauge in the DFAR reports ≥ 0.25 mm of precipitation and the precipitation detector reports at least 18 minutes of precipitation occurrence. Results are consolidated for each weighing gauge type and for each test site in Section 4.1.1.3. Skill scores, root mean square error, and catch efficiency are used to quantify test-gauge performance over the entire assessment period. A detailed assessment of wind-speed effects, focusing only on solid precipitation, is also conducted for each gauge/shield combination on a per-event basis.

4.1.1.3 Results

4.1.1.3.1 *Characterization of precipitation events*

Mean characteristics are provided for 30-minute precipitation events in all precipitation types (phases) and for snow events only in Table 4.2 and Table 4.3, respectively. The values presented are intended to illustrate how the precipitation events varied from site to site, but are subject to significant variability, as indicated by the accompanying standard deviation values. The distribution of precipitation events by phase at each test site is provided in Table 4.4. It is important to emphasize that differences in the environmental conditions experienced at each site do not explain all observed differences in results. Differences in gauge siting and configuration can impact gauge performance significantly, and it is difficult to separate their contribution to errors/uncertainty from those of the environment and the operation of and internal processing specific to different gauge types.

Table 4.2. Precipitation event characteristics at each test site over the duration of WMO-SPICE. Mean event parameters are provided, with uncertainty represented in terms of the standard deviation. Results presented are for events identified by the R2 reference configuration.

Site	Climate zone	Mean event parameters		
		Accumulation [mm]	Temperature [°C]	Wind speed [m s^{-1}]
Bratt's Lake	Continental (high wind)	0.55 ± 0.35	-3.9 ± 8.2	5.7 ± 2.5
CARE	Continental (humid)	0.64 ± 0.46	-2.1 ± 6.8	3.1 ± 1.6
Caribou Creek*	Southern boreal	0.50 ± 0.27	-9.6 ± 9.2	2.6 ± 0.9
Formigal*	Alpine (maritime influence)	1.03 ± 0.80	-0.3 ± 3.4	2.6 ± 1.4
Haukeliseter	Alpine (high wind)	0.63 ± 0.44	-2.3 ± 3.0	6.6 ± 3.8
Marshall	Continental (dry)	0.89 ± 0.67	-1.1 ± 4.7	2.5 ± 1.4
Sodankylä	Northern boreal (low wind)	0.45 ± 0.23	-2.0 ± 3.9	1.5 ± 0.6
Weissfluhjoch	Alpine (complex terrain)	0.92 ± 0.61	-5.2 ± 4.2	2.7 ± 2.1

* Data and results available only for winter 2014/15.

Table 4.3. Snow-event characteristics at each test site over the duration of WMO-SPICE. Mean event parameters are provided, with uncertainty represented in terms of the standard deviation. Results presented are for events identified by the R2 reference configuration.

Site	Climate zone	Mean event parameters		
		Accumulation [mm]	Temperature [°C]	Wind speed [m s^{-1}]
Bratt's Lake	Continental (high wind)	0.53 ± 0.33	-10.4 ± 5.6	6.4 ± 2.7
CARE	Continental (humid)	0.57 ± 0.38	-7.0 ± 4.5	3.0 ± 1.2
Caribou Creek*	Southern boreal	0.41 ± 0.14	-13.9 ± 8.1	2.4 ± 1.0
Formigal*	Alpine (maritime influence)	0.95 ± 0.74	-4.4 ± 1.3	3.0 ± 1.4
Haukeliseter	Alpine (high wind)	0.60 ± 0.41	-4.3 ± 2.6	5.9 ± 3.5
Marshall	Continental (dry)	0.67 ± 0.33	-5.8 ± 2.8	2.7 ± 1.5
Sodankylä	Northern boreal (low wind)	0.43 ± 0.22	-5.3 ± 3.6	1.5 ± 0.6
Weissfluhjoch	Alpine (complex terrain)	0.89 ± 0.57	-6.9 ± 3.1	2.9 ± 2.2

* Data and results available only for winter 2014/15.

Table 4.4. Distribution of precipitation events by phase at each test site over the duration of WMO-SPICE. Results presented are for events identified by the R2 reference configuration.

Site	Climate zone	Percentage of events of each phase by number (by precipitation in mm)		
		Liquid	Mixed	Solid
Bratt's Lake	Continental (high wind)	22.7 (28.2)	26.3 (23.1)	60.0 (48.8)
CARE	Continental (humid)	26.0 (31.0)	20.5 (21.6)	53.4 (47.3)
Caribou Creek*	Southern boreal	2.6 (1.9)	30.2 (42.6)	67.2 (55.4)
Formigal*	Alpine (maritime influence)	26.4 (25.9)	44.6 (47.4)	29.0 (26.6)
Haukeliseter	Alpine (high wind)	4.3 (5.4)	55.1 (55.8)	40.5 (38.8)
Marshall	Continental (dry)	15.4 (18.6)	50.3 (55.4)	34.3 (26.0)
Sodankylä	Northern boreal (low wind)	7.5 (8.3)	50.4 (51.0)	42.1 (40.7)
Weissfluhjoch	Alpine (complex terrain)	5.7 (5.8)	17.5 (20.1)	76.8 (74.1)

* Data and results available only for winter 2014/15.

In general, the mean event accumulations for sites in boreal climates (Caribou Creek, Sodankylä) are the smallest in Table 4.2 and Table 4.3, while those for sites in alpine climates (Formigal, Weissfluhjoch) are the largest. The sites in continental climates (Bratt's Lake, CARE, Marshall) are characterized by intermediate mean event accumulations relative to those in boreal and alpine climates. The site in Haukeliseter, which has an alpine climate, has mean event accumulations comparable to those in continental climates. These values may reflect the influence of wind speed on gauge-catch efficiency to a greater extent than the other alpine sites, which are characterized by lower mean wind speeds.

Other notable observations from Table 4.2, Table 4.3, and Table 4.4 include Bratt's Lake showing the highest mean wind speeds and smallest mean accumulations among the continental sites, while Marshall shows the largest mean accumulations, warmest mean temperatures, and a larger proportion of mixed precipitation relative to solid precipitation compared to the other continental sites. For the boreal sites, Caribou Creek is characterized by lower mean temperatures and a larger proportion of solid-precipitation events than Sodankylä. Weissfluhjoch shows the largest proportion of solid-precipitation events among the alpine sites, Weissfluhjoch and Haukeliseter both show few liquid events, and roughly half of all precipitation events (by both number and total accumulation) at Formigal and Haukeliseter are characterized as mixed precipitation.

4.1.1.3.2 *Detection of precipitation relative to reference*

The ability of each weighing gauge/shield configuration under test to detect precipitation relative to the R2 reference configuration was assessed using the following skill scores: probability of detection, false alarm rate, bias, and Heidke Skill Score. For the purposes of the assessment, the gauge configurations under test are considered to report precipitation if they report ≥ 0.25 mm of precipitation in a given 30-minute assessment interval, while the reference is considered to detect precipitation when its weighing gauge reports ≥ 0.25 mm accumulated precipitation and its precipitation detector reports ≥ 18 minutes of precipitation occurrence during an assessment interval. There are four potential detection scenarios: the reference and test gauge both detect precipitation (YY cases); the reference detects precipitation, but the test gauge does not (YN cases); the reference does not detect precipitation, but the test gauge does (NY cases); and neither the reference nor the test gauge detects precipitation (NN cases). The number of events in each detection scenario for each test gauge over the duration of experiments is used to calculate the skill scores, as detailed in Section 3.6.1.

The detection criteria were selected to identify precipitation events with a high degree of confidence; however, light- and/or sporadic-precipitation events are not well represented. A detailed assessment of the performance of each test gauge during non-precipitating periods is presented in the instrument performance reports (Annex 6) and provides insight into the sensitivity of the test gauges under light-precipitation conditions. Additional tests, using different precipitation-detection thresholds, are also presented to illustrate how the results are impacted by the specific thresholds selected. A separate assessment of weighing gauge detection performance in light precipitation is presented in Section 0.

Skill scores for all test gauges, at all sites, and in all applicable shield configurations are compiled in Figure 4.2 toFigure 4.5. These scores represent the ability of each weighing gauge/shield combination to detect precipitation relative to the reference configuration over the full range of environmental conditions and precipitation types experienced at a given site, within 30 minute periods, and using the specified detection thresholds. The different shield configurations are designated as follows: Belfort double-Alter (BDA); single-Alter (SA); Tretyakov (Tret); single-Alter/Tretyakov (SA/Tret); and unshielded (UN).

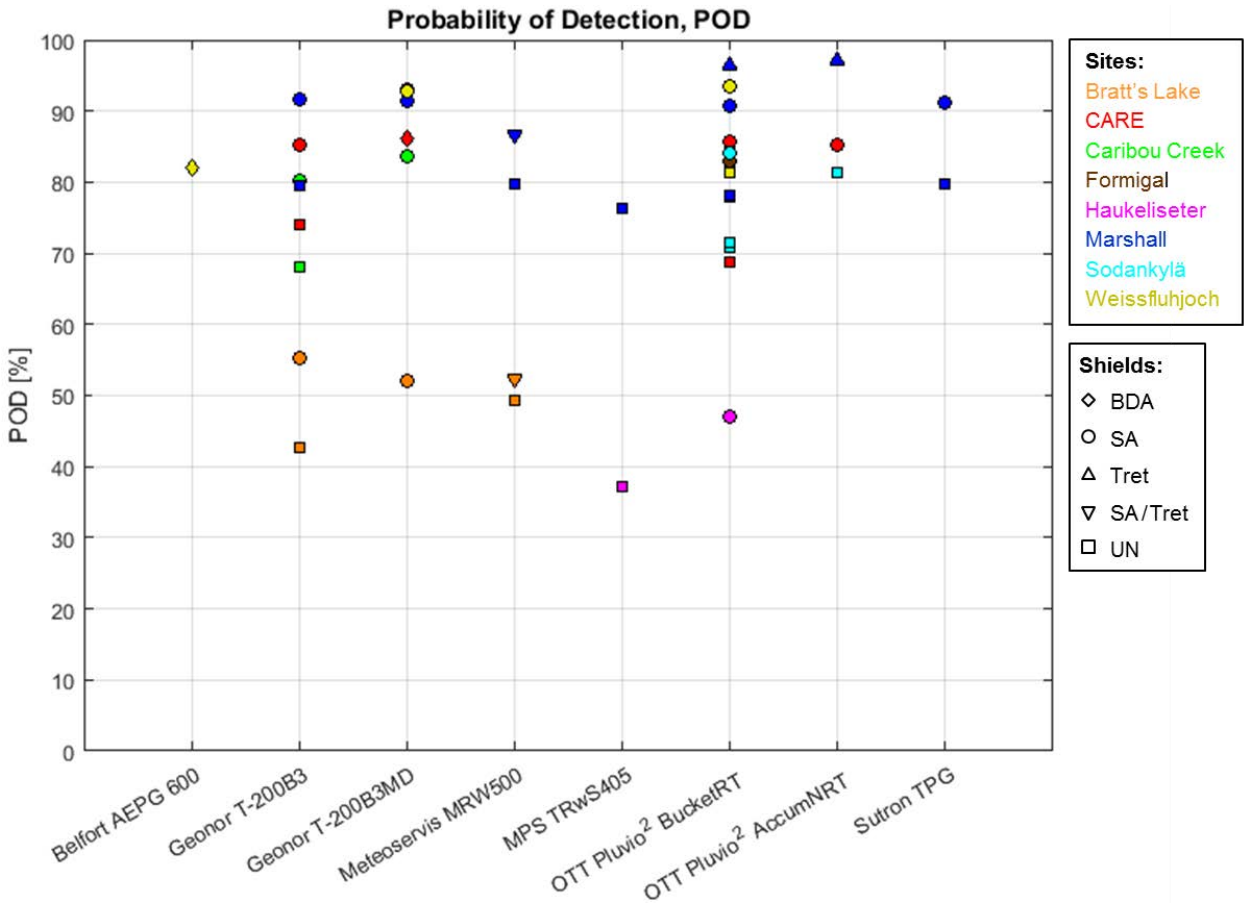


Figure 4.2. Probability of detection results for all weighing gauges under test in WMO-SPICE, including unshielded (UN) and shielded configurations Shields tested include the Belfort double-Alter (BDA), single-Alter (SA), Tretyakov (Tret), and single-Alter/Tretyakov (SA/Tret).

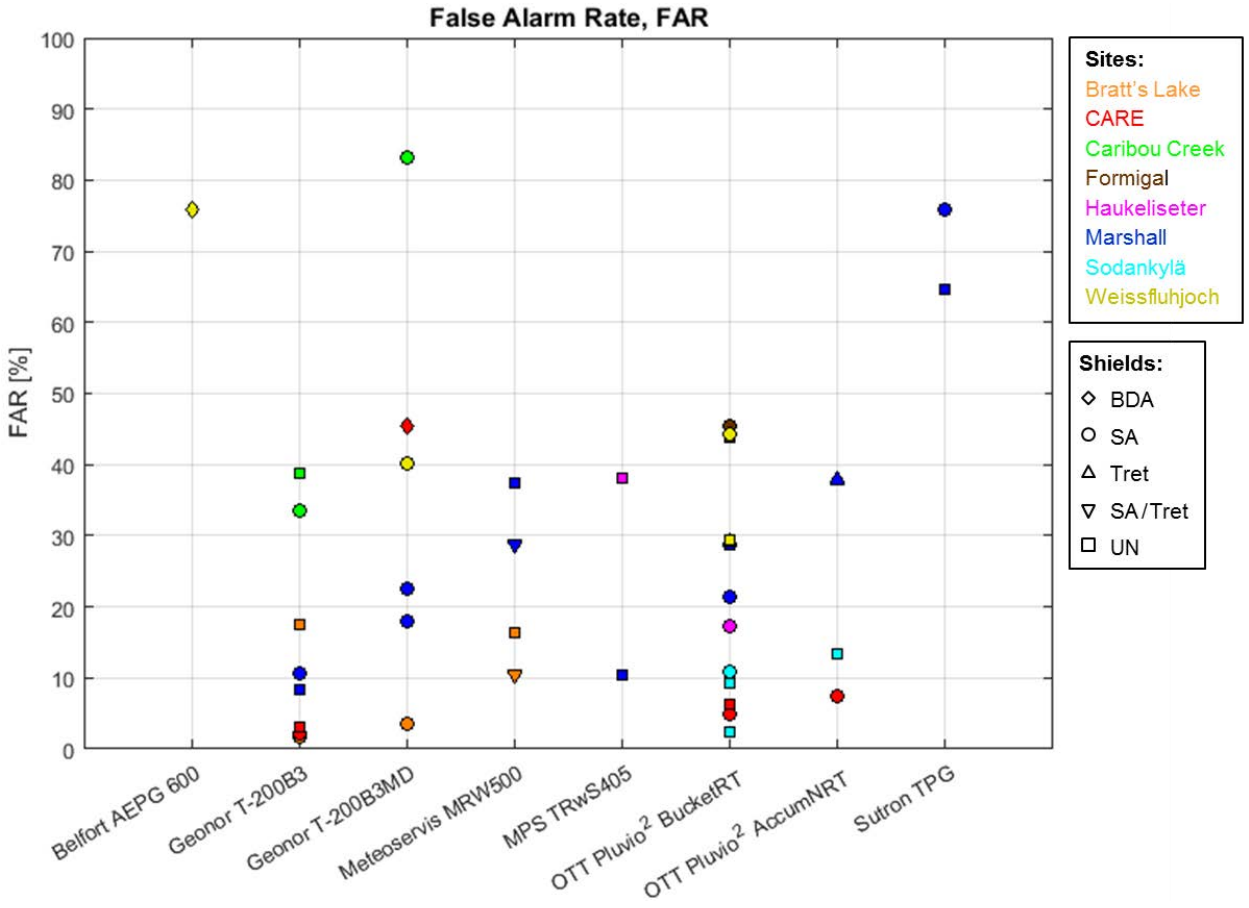


Figure 4.3. False alarm rate results for all weighing gauges under test in WMO-SPICE, including unshielded (UN) and shielded configurations. Shields tested include the Belfort double-Alter (BDA), single-Alter (SA), Tretyakov (Tret), and single-Alter/Tretyakov (SA/Tret).

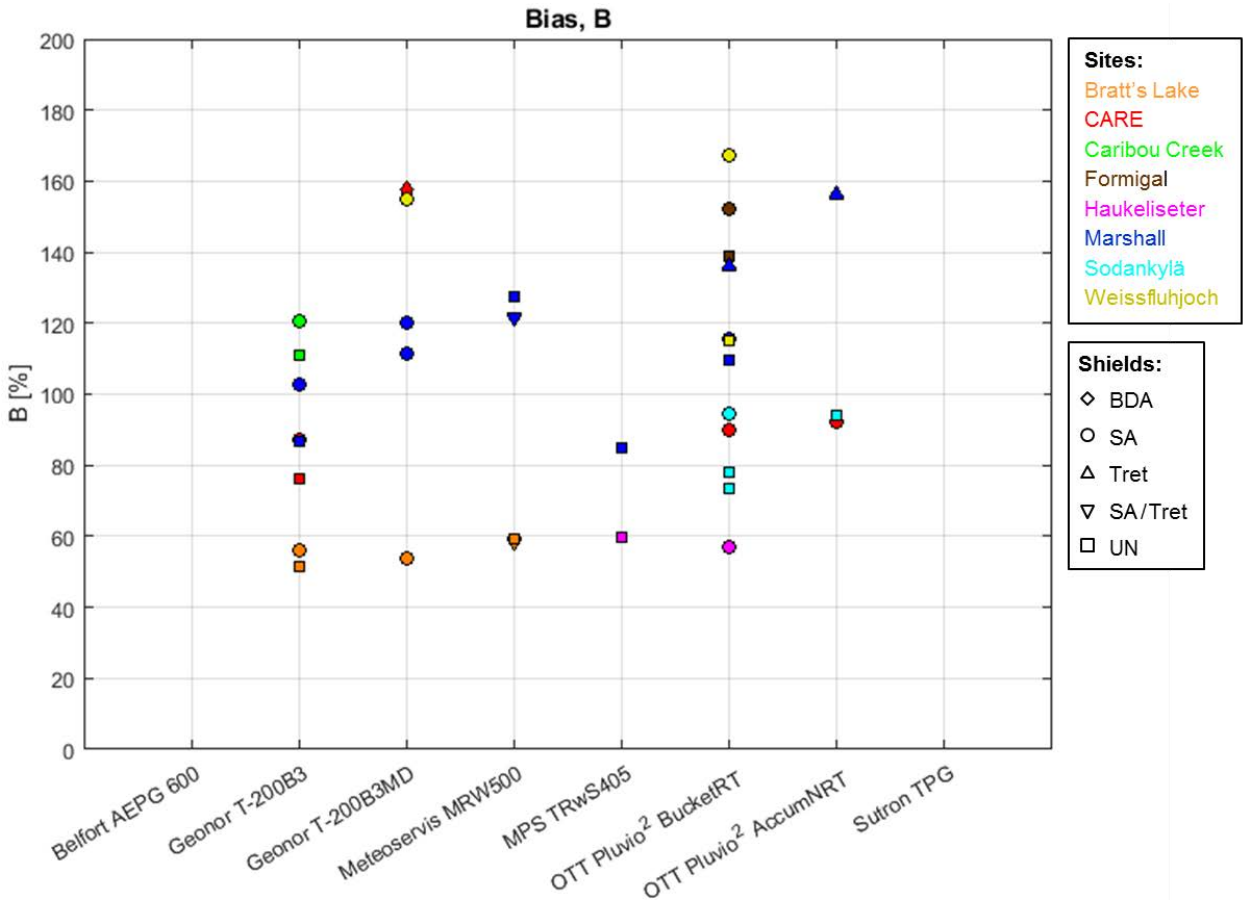


Figure 4.4. Bias results for all weighing gauges under test in WMO-SPICE, including unshielded (UN) and shielded configurations. Shields tested include the Belfort double-Alter (BDA), single-Alter (SA), Tretyakov (Tret), and single-Alter/Tretyakov (SA/Tret). The y-axis is limited to values ≤ 200% for clarity; datapoints not visible in the plot are discussed in the accompanying text.

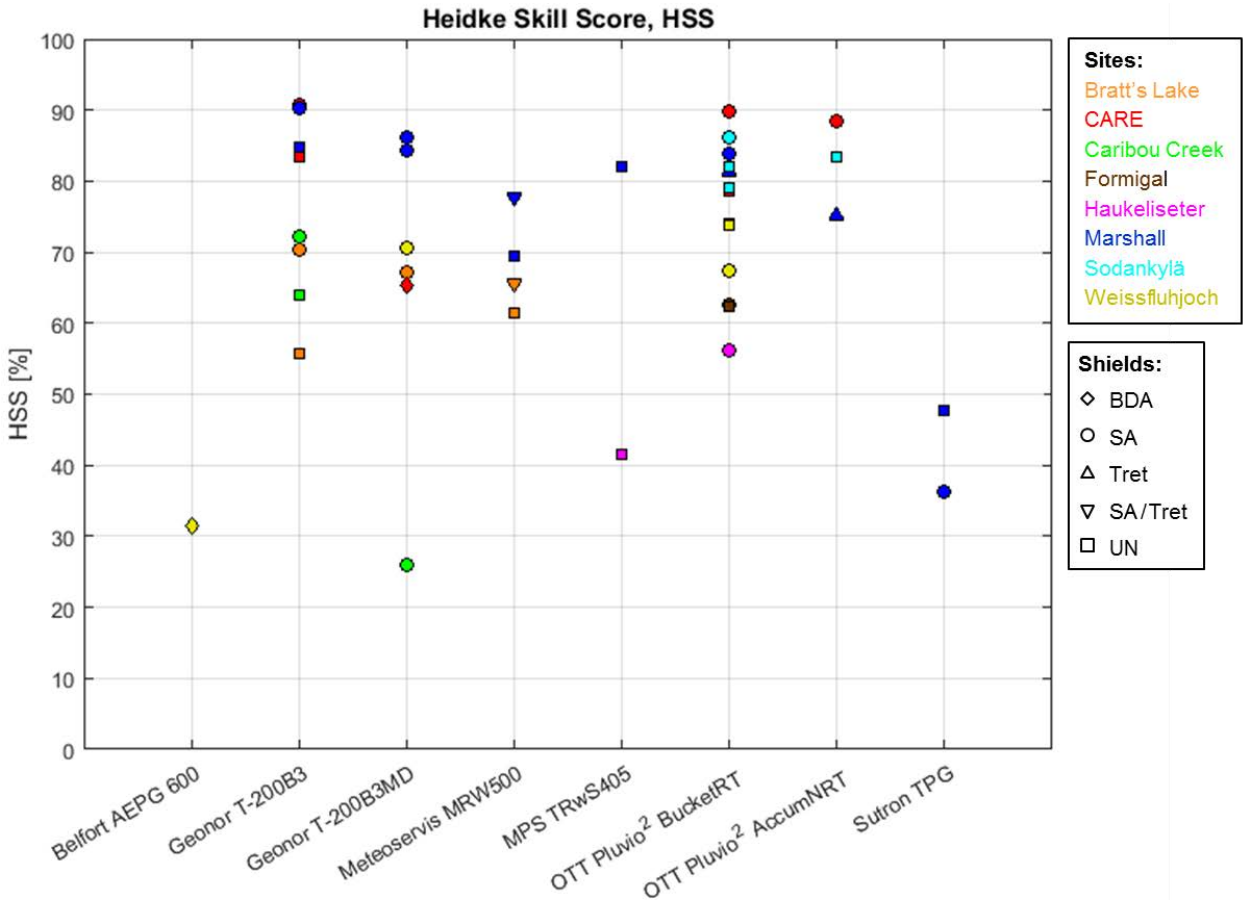


Figure 4.5. Heidke Skill Score results for all gauges under test in WMO-SPICE, including unshielded (UN) and shielded configurations. Shields tested include the Belfort double-Alter (BDA), single-Alter (SA), Tretyakov (Tret), and single-Alter/Tretyakov (SA/Tret).

The probability of detection results for all test gauge configurations are shown in Figure 4.2. Lower POD values are evident for the test gauges at Bratt's Lake (POD between ~ 40% and 55%) and Haukeliseter (POD between ~ 37% and 47%) corresponding to the two sites characterized by the highest mean wind speeds (Table 4.2 and Table 4.3). For the test gauges at all other sites, the POD is greater than about 80% for all shielded gauges and within about 70% to 80% for all unshielded gauges, irrespective of the specific site/climate regime. For the Geonor T-200B3, OTT Pluvio^2, and Sutron TPG test gauges, the POD values for the SA-shielded gauges are all approximately 10% to 15% higher than the unshielded gauges at the same site. For the Meteoservis MRW500 test gauges at Bratt's Lake and Marshall, the POD is about 5% to 7% higher for the SA/Tret shielded configuration relative to the unshielded configuration at each site.

Both Bucket RT and Accumulated NRT outputs from the OTT Pluvio^2 gauges under test at CARE, Marshall, and Sodankylä (submitted by either the site host or gauge manufacturer) are considered in the assessment. For the test gauges at CARE and Marshall (SA and Tret shields, respectively), the POD is similar for both outputs. For the unshielded test gauge at Sodankylä, the POD is approximately 10% higher for the Accumulated NRT output, which can likely be attributed to trace precipitation identified by the processing algorithm pushing the event accumulation above the detection

threshold. Note that the NRT data are offset by five minutes to account for the fixed-output delays related to processing, such that the RT and NRT data both cover the same 30-minute periods.

The false alarm rates (Figure 4.3) vary by test configuration and site, and are generally within about 40% to 50%. There is no clear trend for relative FAR between shielded and unshielded test gauges of the same type at the same site; however, trends are evident for processed and unprocessed data outputs from test gauges of the same type and configuration at the same site. Specifically, the FAR values are higher (by up to 10%) for processed Accumulated NRT outputs relative to unprocessed Bucket RT outputs for the OTT Pluvio^2 gauges at CARE (single-Alter shield), Marshall (Tretyakov shield), and Sodankylä (unshielded). The higher FAR values in these cases may result from false reports, in which the processing artificially increases the accumulation values reported.

The FAR values differ by about 5% to 7% for identical gauge/shield combinations at the same site (SA-shielded Geonor T-200B3MD gauges at Marshall, unshielded OTT Pluvio^2 gauges at Sodankylä) suggesting that gauge siting and the inhomogeneity of conditions play a role. These factors, combined with the roles of specific gauge configuration and data processing, may give rise to the notably higher FAR values for the Belfort AEPG at Weissfluhjoch and the Sutron TPG gauges at Marshall, but the details are unclear. The FAR for the SA-shielded Geonor T-200B3MD test gauge at Caribou Creek is also notably high and exceeds that for the other test configurations at the site by ~ 40%. A similar trend is observed for the BDA-shielded Geonor T-200B3MD test gauge at CARE. The assessment of gauge responses in non-precipitating conditions presented in the relevant instrument performance reports (Annex 6) indicate that each of these Geonor T-200B3MD gauges was subject to enhanced noise relative to the reference configuration, resulting from issues with the specific gauge and/or site configuration. Enhanced noise in the gauge output could result in more NY cases, thereby increasing the FAR. Another consideration is that the Geonor T-200B3MD has a longer, larger capacity bucket (1500 mm) relative to the other test gauges, resulting in differences in the center of mass and overall gauge stability that may lead to more wind-induced vibration and noise.

Bias results for the weighing gauges under test are plotted in Figure 4.4. A bias value of 100% indicates that the reference and test gauge detect the same number of events; this is the ideal scenario. Biases above/below 100% indicate that the test gauges detect more/fewer events than the reference. It follows that the test gauges with the lowest POD values (Figure 4.2), specifically those at Bratt's Lake and Haukeliseter (high-wind sites), are biased low relative to the reference. Similarly, test gauges with the highest FAR (Figure 4.3) are biased high relative to the reference; this is not immediately apparent in Figure 4.4, as the y-axis has been limited to values within 200% for clarity. The BDA-shielded Belfort AEPG gauge at Weissfluhjoch has a bias of ~ 339%, the SA-shielded Geonor T-200B3MD gauge at Caribou Creek has a bias of ~ 495%, and the unshielded and SA-shielded Sutron TPG gauges at Marshall have bias values of ~ 225% and 377%, respectively.

The bias values vary by gauge/shield configuration and by site. The test gauges at CARE and Sodankylä are generally biased low relative to the reference in terms of precipitation detection, while those at Caribou Creek, Formigal, Marshall, and Weissfluhjoch tend to be biased high relative to the reference. For test gauges with shielded and unshielded configurations at the same site, the shielded configuration typically has the larger bias value, indicating that they detect more precipitation events relative to the unshielded configuration (both relative to the reference configuration). For OTT Pluvio^2 gauges in the same configuration at the same site, larger bias values are observed for the processed gauge outputs (Accumulated NRT), which may be attributed to false reports resulting from the specific processing algorithm employed.

The Heidke Skill Score values in Figure 4.5 reflect the overall ability of test gauges to detect precipitation relative to the reference configuration, accounting for expected performance due to chance alone. The test gauge configurations with markedly higher false alarm rates have the lowest HSS values; the BDA-shielded Belfort AEPG at Weissfluhjoch, SA-shielded Geonor T-200B3MD at Caribou Creek, and unshielded and SA-shielded Sutron TPG gauges at Marshall all have HSS values below 50%. Among the other test gauges, those at high-wind sites (Bratt's Lake, Haukeliseter) show lower HSS values − between about 40% and 70% − with the lowest values for unshielded configurations. For the remaining sites and test gauges, the HSS values range from approximately 60% to 90% and are generally higher for shielded configurations relative to unshielded configurations for a given gauge and site.

It is important to emphasize that the skill score results reflect only the ability of the test gauges to detect precipitation relative to the reference configuration at a given site over 30-minute intervals. These results do not reflect the overall performance of the gauge in terms of reporting accumulated precipitation over the same or longer time intervals. An assessment of test-gauge performance in terms of accumulation reports over both 30-minute assessment intervals and the duration of experiments is presented in Section 4.1.1.3.3.

4.1.1.3.3 *Reporting accumulated precipitation relative to reference*

The catch efficiency, or catch ratio, is the ratio of accumulated precipitation reported by a test gauge relative to that reported by the reference configuration over a specified time interval. In the current assessment, the catch efficiency is calculated for 30-minute assessment intervals during which the reference and test gauge both detect and report precipitation (YY cases). The catch efficiency of solid precipitation is of particular interest, as the lower densities and slower fall velocities of particles make them more susceptible to wind-induced undercatch. The catch efficiency of all test-gauge configurations in solid-precipitation conditions is assessed as a function of mean wind speed in Section 4.1.1.3.3.1. The overall catch efficiency, reflecting the total accumulation reported by a test gauge relative to that reported by the reference configuration over the duration of formal tests is discussed in Section 4.1.1.3.3.2. The total accumulation values used in computing the overall catch efficiency are determined by summing the accumulated precipitation of the test gauge and reference over all YY cases considered in the assessment.

The catch efficiency is a useful indicator of gauge performance relative to the reference, but does not provide information about the magnitude of accumulated precipitation reported by each gauge. The root mean square error, however, considers the absolute difference in reported accumulation between the test gauge and the reference for each 30-minute interval. An assessment of RMSE results is provided for all test gauges, at all sites in Section 4.1.1.3.3.3.

4.1.1.3.3.1 Wind effects on catch efficiency

The catch efficiency of each test gauge and at each site is assessed as a function of the mean wind speed for 30-minute assessment intervals in Section 4.1.1.3.3.1.1. This assessment is limited to snow events during which the maximum temperature does not exceed -2 °C over a given 30-minute assessment interval. A similar assessment is presented for all precipitation types in the instrument performance reports (Annex 6).

To assess the influence of wind speed, box and whisker plots of catch efficiency as a function of mean wind speed are plotted for each test gauge, with results presented in 1 m/s bins. The boxes in each plot represent the range of values between the 25th percentile (lower quartile; bottom of box) and

75th percentile (upper quartile; top of box), referred to as the interquartile range. The median value is indicated by the horizontal line across the box. The whiskers below (above) the box indicate the lowest (highest) values. Outlying points are indicated by markers above or below the whiskers. To facilitate comparison, all plots have the same scale, covering mean wind speeds up to 10 m/s and catch efficiencies ≤ 2.

4.1.1.3.3.1.1 Results by gauge type

The influence of wind speed on catch efficiency for each test gauge type is illustrated in Figure 4.6 toFigure 4.14. In each figure, different colored plots are overlaid for each test configuration and/or site. For the Geonor T-200B3 and OTT Pluvio^2 test gauges, which have the largest numbers of test configurations, separate figures are provided for shielded and unshielded test configurations for clarity.

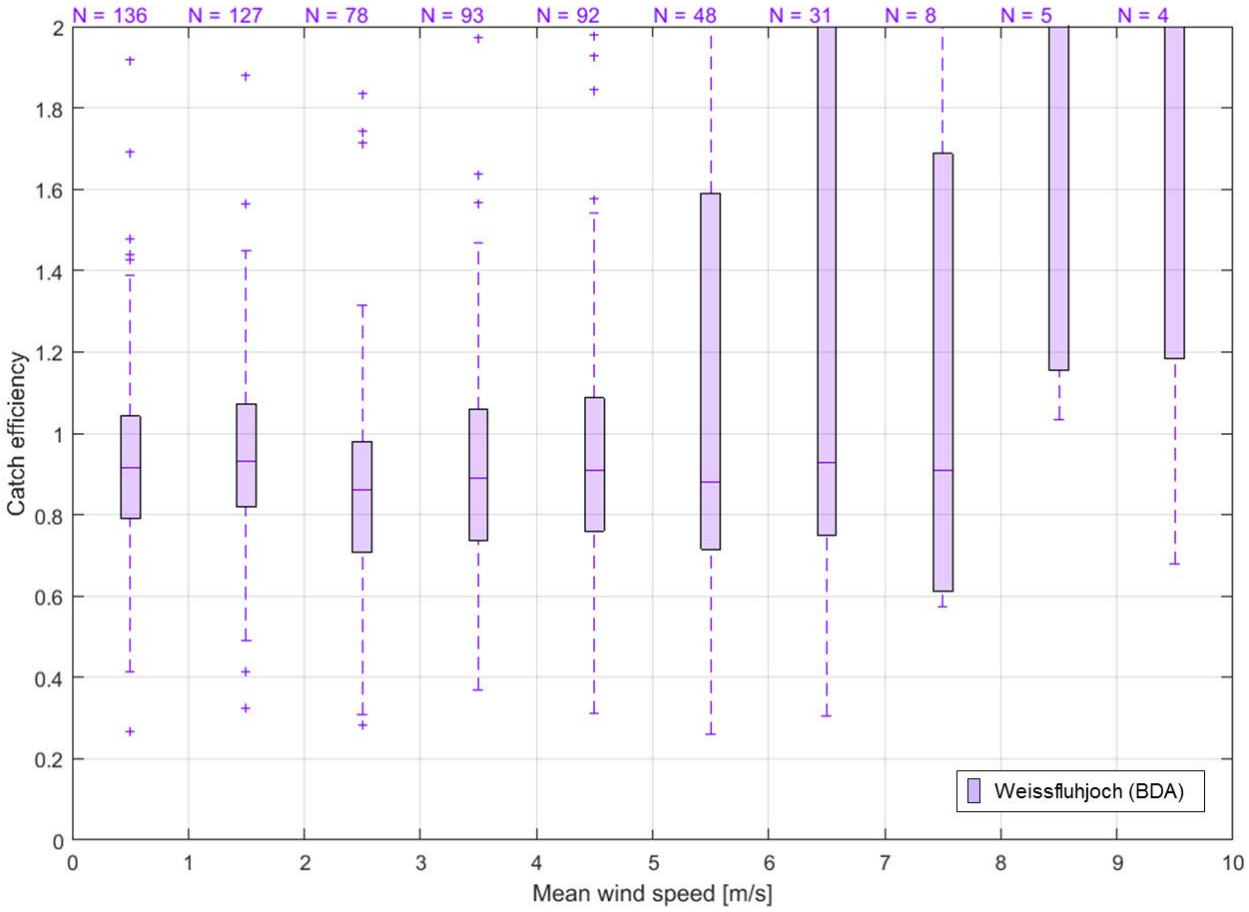


Figure 4.6. Box and whisker plot of catch efficiency as a function of mean wind speed for the Belfort AEPG 600 test gauge in Belfort double-Alter shield. The number of precipitation events in each wind-speed bin is indicated above the plot.

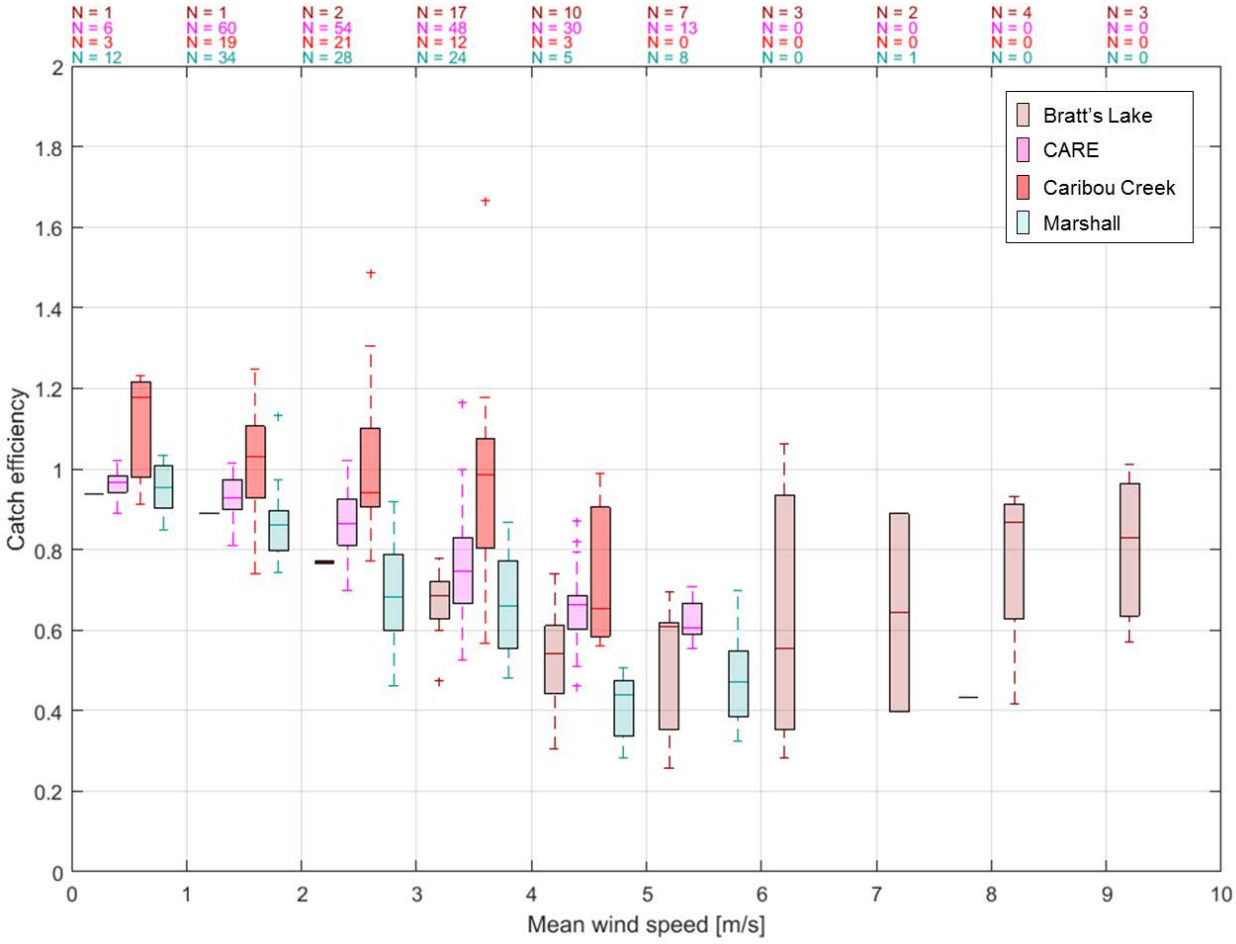


Figure 4.7. Box and whisker plot of catch efficiency as a function of mean wind speed for single-Alter-shielded Geonor T-200B3 gauges. The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

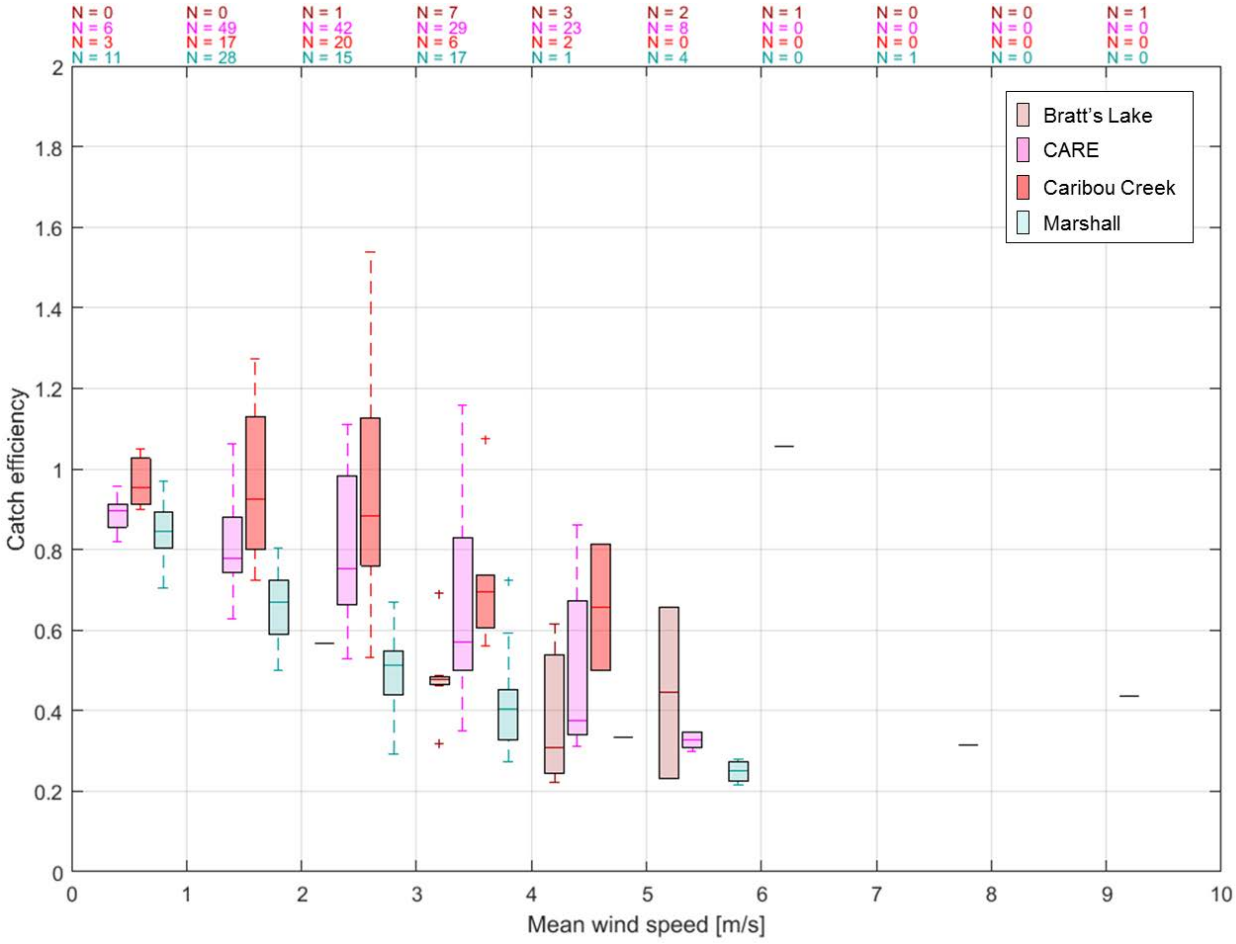


Figure 4.8. Box and whisker plot of catch efficiency as a function of mean wind speed for unshielded Geonor T-200B3 gauges. The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

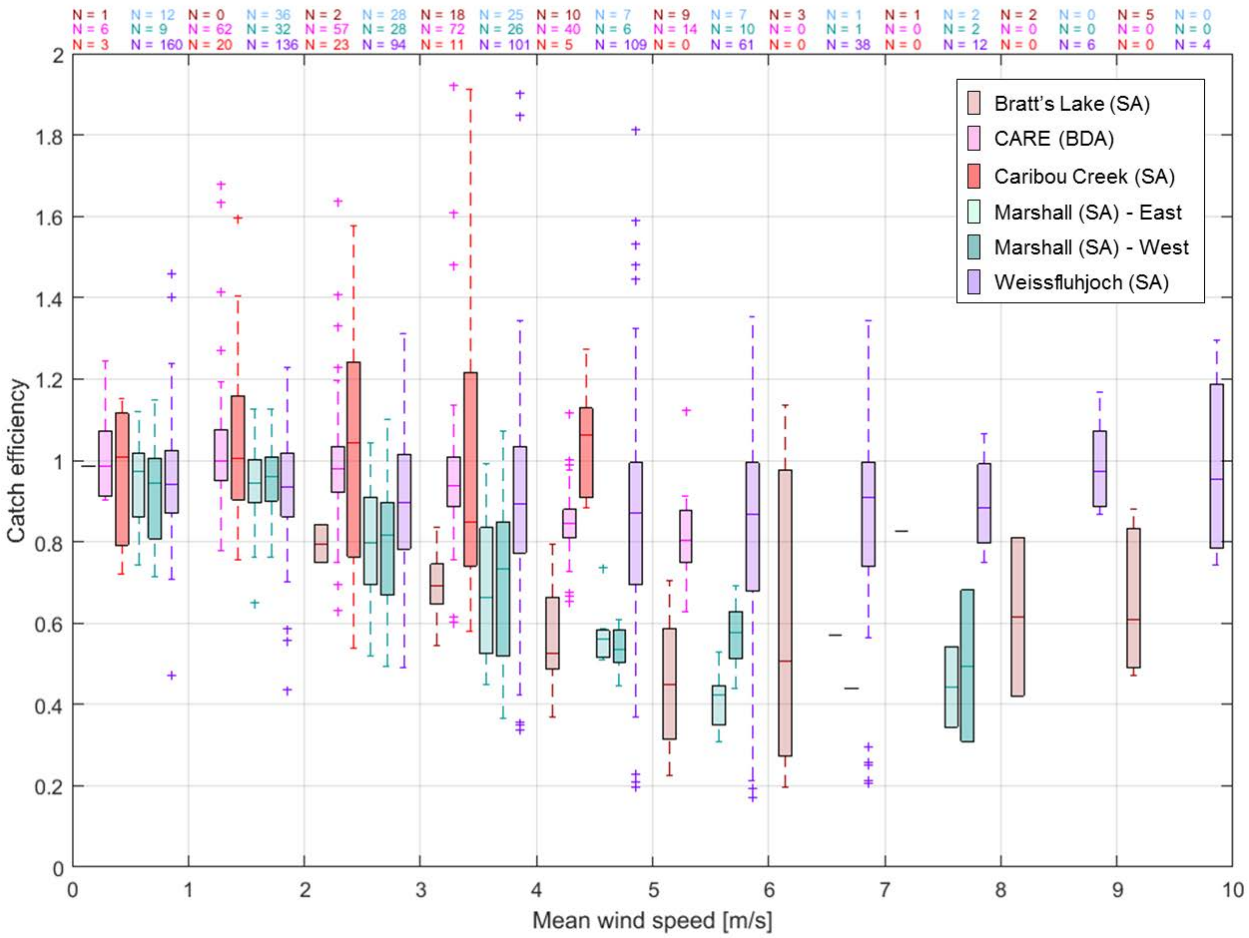


Figure 4.9. Box and whisker plot of catch efficiency as a function of mean wind speed for single-Alter-shielded and Belfort double-Alter-shielded Geonor T-200B3MD gauges. The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

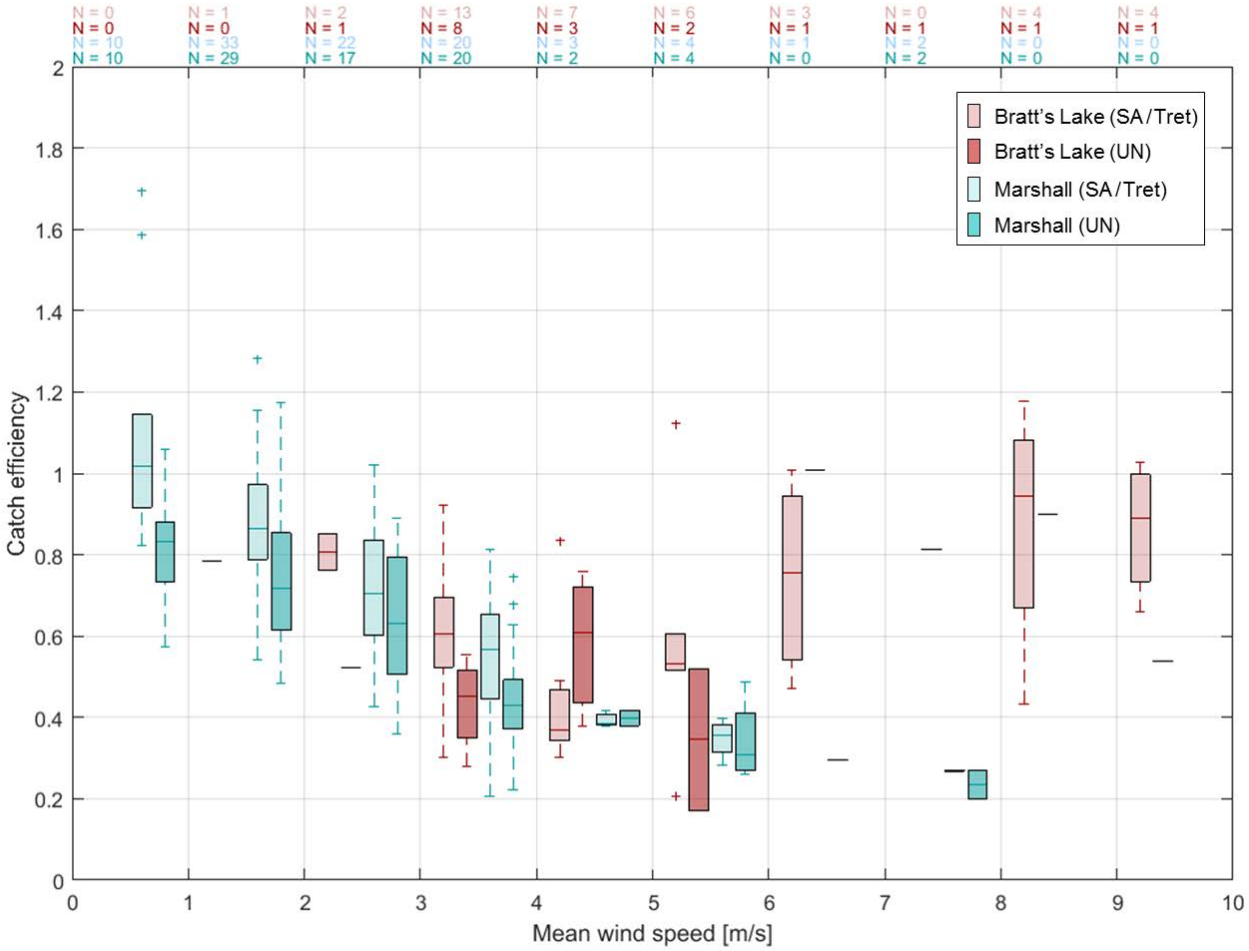


Figure 4.10. Box and whisker plot of catch efficiency as a function of mean wind speed for single-Alter/Tretyakov-shielded and unshielded Meteoservis MRW500 gauges. The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

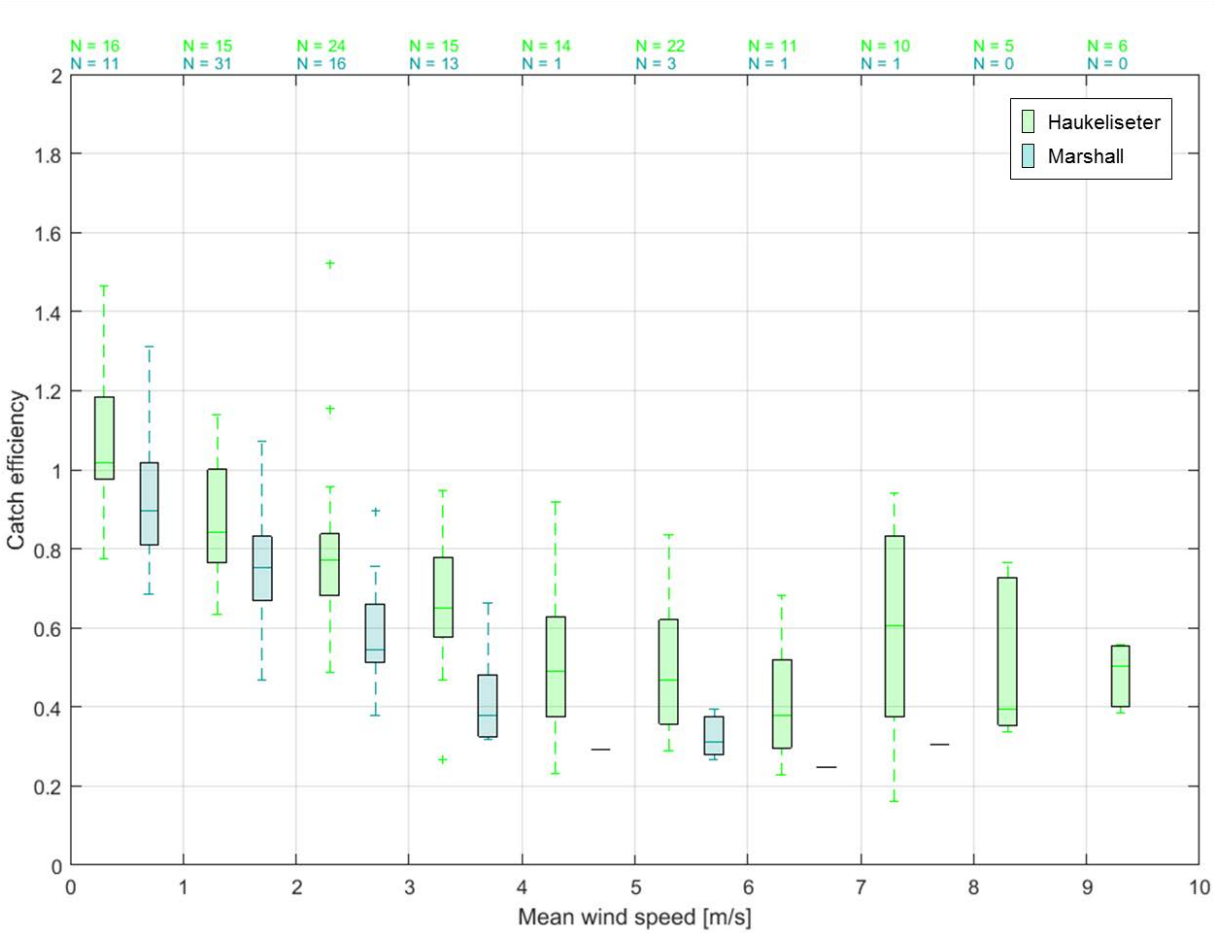


Figure 4.11. Box and whisker plot of catch efficiency as a function of mean wind speed for unshielded MPS TRwS405 gauges. The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

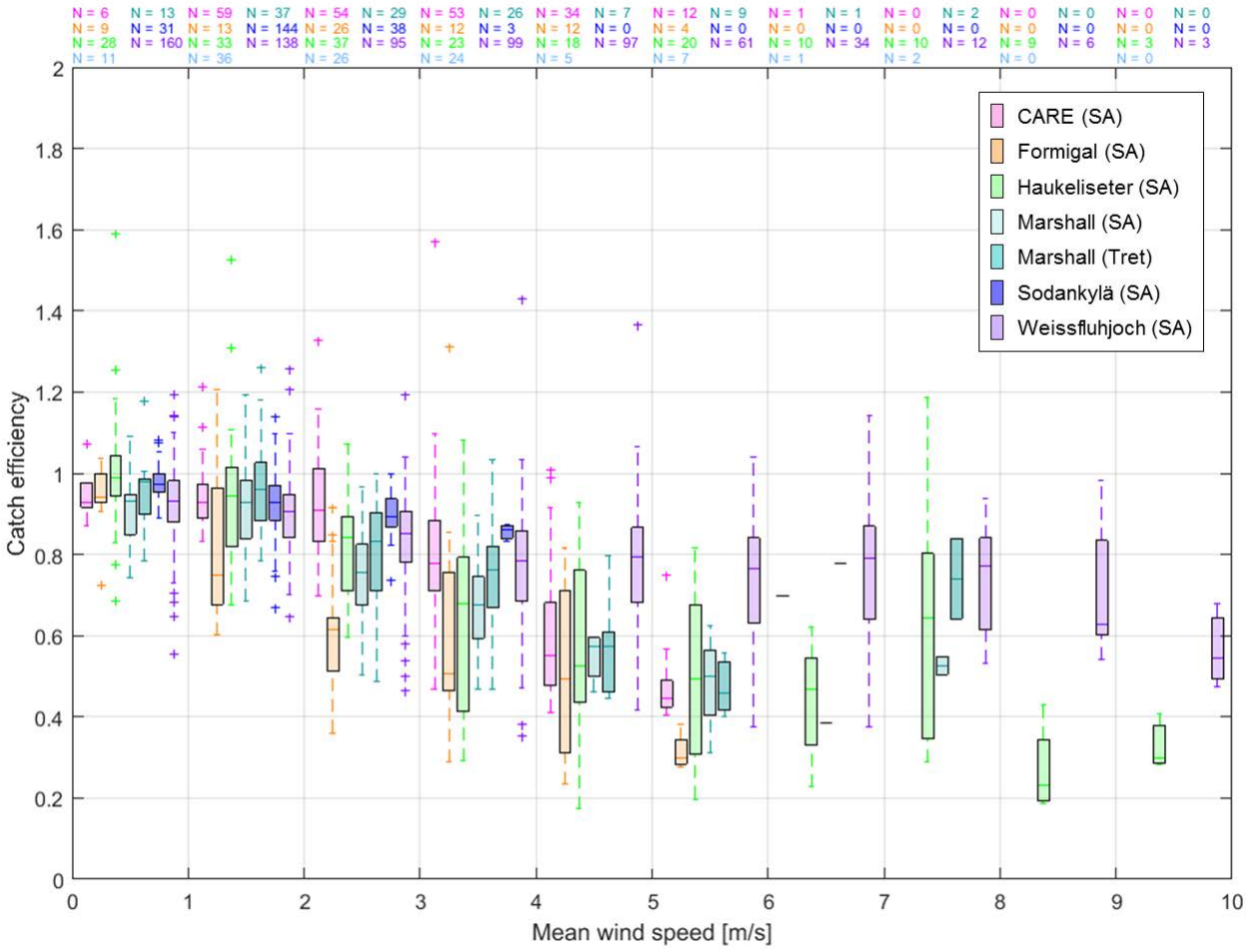


Figure 4.12. Box and whisker plot of catch efficiency as a function of mean wind speed for single-Alter- and Tretyakov-shielded OTT Pluvio^2 gauges (Bucket RT output). The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

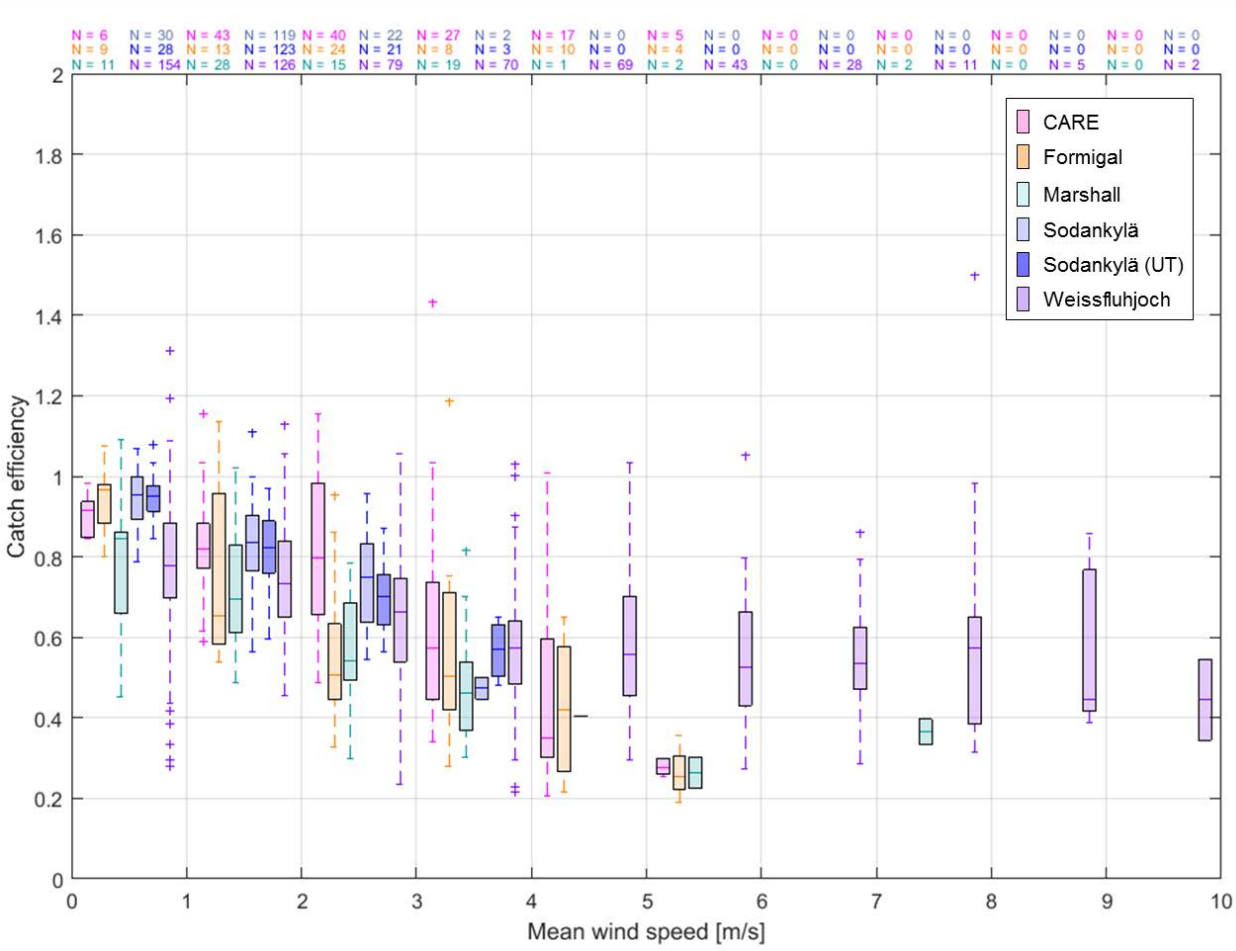


Figure 4.13. Box and whisker plot of catch efficiency as a function of mean wind speed for unshielded OTT Pluvio^2 gauges.The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot. The UT designation in the legend indicates that the gauge was submitted by the manufacturer for evaluation and is used to distinguish between gauges in the same configuration at the same site.

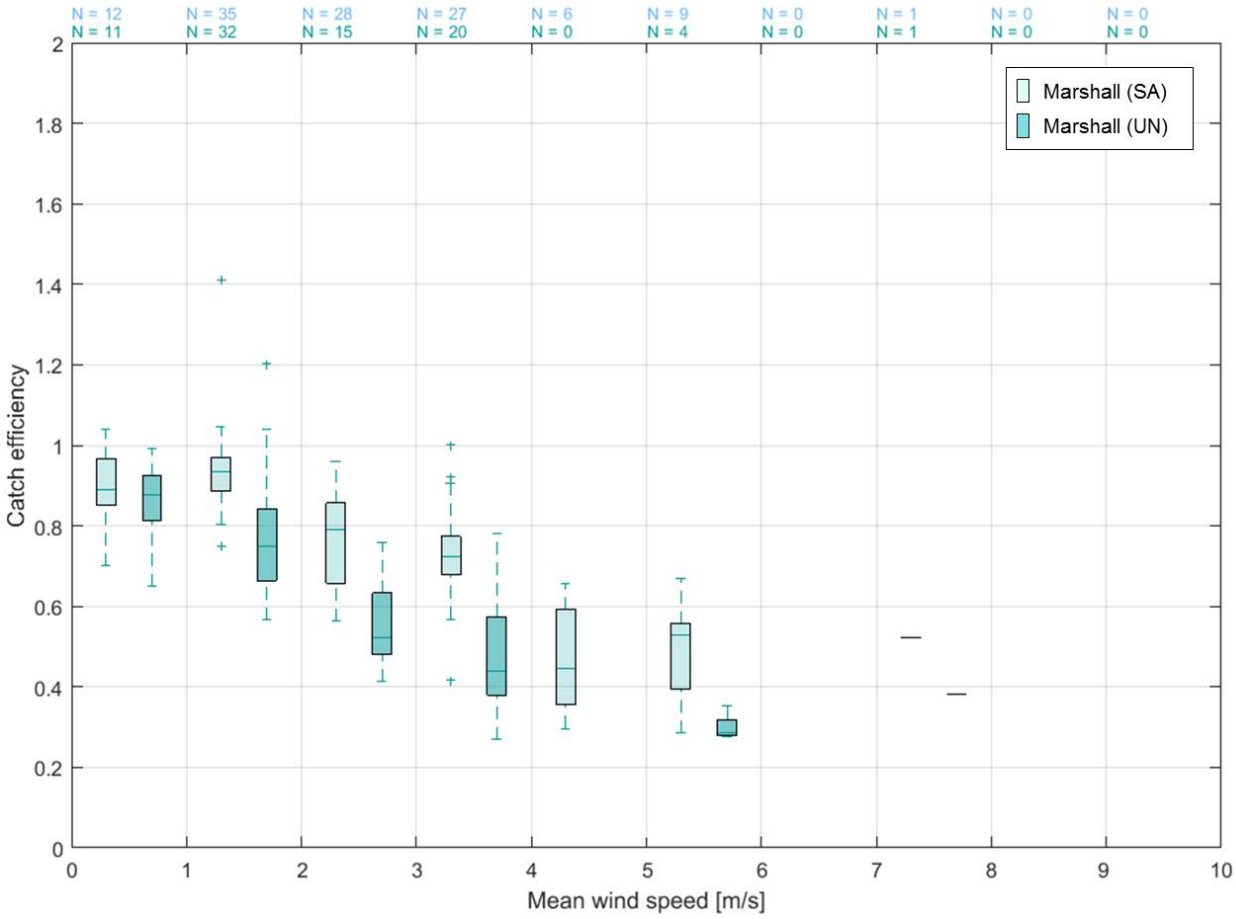


Figure 4.14. Box and whisker plot of catch efficiency as a function of mean wind speed for single-Alter-shielded and unshielded Sutron TPG gauges.The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

It is difficult to make attributions for observed differences in the CE vs. wind speed relationship for the same gauge at multiple sites. These differences are likely the combined result of differences in environmental conditions (mean event accumulation, temperature, predominant ice crystal type (e.g. Theriault et al., 2012) for solid precipitation) and differences in gauge configuration and siting. Accordingly, the identification and diagnosis of differences in results between or among sites is not a focal point of this work; rather, the focus is on identifying general trends in the results to help inform gauge selection and deployment in different climate regimes.

4.1.1.3.3.1.1.1 Belfort AEPG 600

The Belfort AEPG 600 gauge (Figure 4.6) was tested only at Weissfluhjoch and in a single configuration (Belfort double-Alter shield). Significant scatter is observed in the results, which may indicate that its specific configuration, as installed, is subject to performance limitations under the conditions tested. There is no discernible trend in catch efficiency with increasing wind speed; the median CE falls between 0.8 and 1 for mean wind speeds up to 8 m/s and exceeds the range of the plot at higher wind speeds (noting that there are relatively few events with mean wind speeds > 7 m/s). Further testing of Belfort AEPG 600 gauges is required, preferably at multiple sites in different climate zones, to better characterize how their performance is affected by wind speed.

4.1.1.3.3.1.1.2 Geonor T-200B3

The catch efficiency vs. mean wind speed relationships for single-Alter-shielded Geonor T-200B3 test gauges are shown in Figure 4.7. Three of the test sites – Bratt's Lake, CARE, and Marshall – are in continental climates, while Caribou Creek is in a southern boreal climate. The difference in climate region provides one possible explanation for the higher catch-efficiency values observed for the test gauge at Caribou Creek relative to the other gauges, but this is more likely related to the higher noise observed for the gauge at Caribou Creek. (See the instrument performance report in Annex 6). This may artificially enhance the reported precipitation amounts, giving rise to the median catch efficiencies > 1 observed at mean wind speeds < 2 m/s.

Considering the lowest median catch efficiencies in each wind-speed bin as lower limits, the median catch efficiencies for single-Alter-shielded test gauges in continental climates decrease within approximately 0.2 for each 2 m/s increase in mean wind speed, falling to approximately 0.4 for wind speeds within 6 m/s. At higher mean wind speeds, there are relatively few events detected. The increase in catch efficiency observed for the Bratt's Lake data above 6 m/s is attributed to the influence of blowing snow, which was corroborated using observer reports from a nearby airport.

For the unshielded Geonor T-200B3 gauges under test (Figure 4.8), the median catch efficiency decreases more quickly with increasing mean wind speed relative to the shielded gauges, falling to approximately 0.2 for wind speeds within 6 m/s. Compared with the shielded test gauges at the same sites (Figure 4.7), the interquartile ranges within a given wind-speed bin are greater for the unshielded gauges, indicating greater variability in the reported accumulation amounts. At mean wind speeds > 6 m/s, the unshielded gauge at Bratt's Lake reports fewer events than the shielded gauge; in this particular case, the enhanced undercatch in the absence of a wind shield mitigates the reporting of false precipitation due to blowing snow.

4.1.1.3.3.1.1.3 Geonor T-200B3MD

Geonor T-200B3MD gauges were tested at the same continental and southern boreal sites as the Geonor T-200B3 gauges, as well as at the alpine Weissfluhjoch site. All test gauges were in single-Alter-shielded configurations, with the exception of the test gauge at CARE, which was in a Belfort double-Alter shield. The catch efficiency vs. wind speed plots for all Geonor T-200B3MD test configurations are compiled in Figure 4.9. The single-Alter-shielded test gauges at the continental climate sites (Bratt's Lake, Marshall) showed similar trends, with median catch efficiencies decreasing to about 0.4 to 0.6 for mean wind speeds within 6 m/s. Given the small numbers of events detected at speeds > 6 m/s, it is difficult to assign much weight to the trends observed, but it is believed that the increase in catch efficiency observed in the Bratt's Lake data may result from blowing snow (as considered above for the Geonor T-200B3 gauges in Section 4.1.1.3.3.1.1.2). The Belfort double-Alter-shielded gauge at CARE was evidently impacted less by increasing wind speed, with median catch efficiencies of approximately 0.8 for mean wind speeds up to 6 m/s. While this can likely be attributed to the influence of the second concentric wind shield, one cannot rule out the possibility that the accumulation reports from this gauge were artificially increased by noise related to the configuration (as noted in the IPR, Annex 6), which would increase the computed catch-efficiency values.

The catch efficiencies for the single-Alter-shielded test gauges at Caribou Creek (southern boreal climate) and Weissfluhjoch show no discernible trends with increasing mean wind speed. A high degree of variability is apparent in the Caribou Creek results (as identified from the vertical extent of interquartile ranges – the boxes – and the ranges of values covered by the whiskers), which may be

attributed to noise inherent to the site configuration (see IPR in Annex 6) and to the specific configuration of the 1500 mm capacity Geonor T-200B3MD (discussed in Section 4.1.1.3.2). The cause of the consistent catch efficiency of the test gauge (values between 0.8 and 1) at Weissfluhjoch over the full range of mean wind speeds (within 10 m/s) is unclear. It is interesting to note, however, the similarity of this trend to that observed for the Belfort AEPG 600 test gauge at Weissfluhjoch, which also operates using three vibrating-wire transducers.

4.1.1.3.3.1.1.4 Meteoservis MRW500

Meteoservis MRW500 gauges were tested in single-Alter/Tretyakov and unshielded configurations at both the Bratt's Lake and Marshall test sites (continental climates). The catch-efficiency trends with increasing wind speed are comparable for the test gauges at both sites (Figure 4.10). The median catch efficiencies of the single-Alter/Tretyakov-shielded gauges exceed those of the unshielded gauges by 0.1 to 0.2 for mean wind speeds up to 4 m/s. The median catch efficiencies of the shielded and unshielded configurations fall to 0.5 to 0.6 and 0.4 to 0.5, respectively, for mean wind speeds within 4 m/s. At higher mean wind speeds, the numbers of events detected are limited, lending less credence to the trends observed; however, it appears that the median catch efficiencies for both configurations are reduced to the 0.2 to 0.3 range for wind speeds up to 8 m/s at Marshall, while the Bratt's Lake results again likely show the influence of blowing snow at mean wind speeds > 6 m/s.

4.1.1.3.3.1.1.5 MPS TRwS405

The unshielded MPS TRwS405 test gauges at Haukeliseter (alpine climate) and Marshall (continental climate) show different catch-efficiency trends with increasing mean wind speed (Figure 4.11). The median catch efficiency for the test gauge at Marshall drops off more rapidly with increasing wind speed, reaching approximately 0.3 for mean wind speeds up to 6 m/s, while that for the test gauge at Haukeliseter falls to approximately 0.5 over the same wind speed range. In fact, the median catch efficiency for the test gauge at Haukeliseter remains at approximately 0.4 or above for mean wind speeds up to 10 m/s. These differing trends are interesting: even though the climate regimes are different, the mean snow-event accumulations and temperatures in Table 4.3 are similar. These results are most likely a reflection of differences in the individual events, gauge siting, and configuration (e.g. the test gauges at Haukeliseter are installed at 4.5 m, while those at Marshall are installed at 1.85 m).

4.1.1.3.3.1.1.6 OTT Pluvio^2

The OTT Pluvio^2 gauges were tested over the broadest range of climate conditions and sites, including alpine (Formigal, Haukeliseter, Weissfluhjoch), continental (CARE, Marshall), and northern boreal (Sodankylä). The gauges were tested in single-shield (single-Alter and Tretyakov) and unshielded configurations. The catch efficiency vs. wind speed relationships for shielded test configurations are plotted in Figure 4.12. The results for SA-shielded gauges illustrate how the same gauge in identical configurations can exhibit markedly different performance at sites in different climate regimes, and even at sites within the same climate regime. The median catch efficiencies for the test gauges at alpine sites show distinctly different decreases with increasing wind speed. For mean wind speeds up to 6 m/s, the median catch efficiencies remain above 0.7 for the test gauge at Weissfluhjoch, while falling to approximately 0.5 for the test gauge at Haukeliseter and to 0.3 for the test gauge at Formigal. The test gauges at CARE and Marshall show similar trends to one another, with median catch efficiencies of about 0.4 to 0.5 at mean wind speeds up to 6 m/s and within about 0.1 to 0.15 of each other in each intermediate wind-speed bin. The median catch efficiency for the test gauge at Sodankylä, characterized by light-snow events and low wind speeds (Table 4.2) remains above 0.8 for mean wind speeds ≤ 4 m/s.

The results for the unshielded configurations are plotted in Figure 4.13. Here, the decreases in catch efficiency for the test gauges at Formigal and Weissfluhjoch (alpine sites) are similar for mean wind speeds up to 4 m/s (CE between 0.5 and 0.6), after which the values level off for the gauge at Weissfluhjoch (CE between 0.4 and 0.6 for mean wind speeds up to 10 m/s) and continue to decrease for the gauge at Formigal (CE of approximately 0.25 for mean wind speeds up to 6 m/s). The test gauges at CARE and Marshall (continental sites) show similar decreases as the test gauge at Formigal (CE of ~ 0.25 for mean wind speeds up to 6 m/s), as do the test gauges at Sodankylä (northern boreal site with CE of ~ 0.5 to 0.6 for mean wind speeds up to 4 m/s). Similar to the results observed for Geonor T-200B3 gauges (Figure 4.7 and Figure 4.8), the results for unshielded OTT Pluvio^2 gauges show greater variability relative to those for shielded gauges at the same site, as evidenced by the greater vertical extent of the boxes and whiskers in Figure 4.13 relative to those in Figure 4.12.

4.1.1.3.3.1.1.7 Sutron TPG

Single-Alter-shielded and unshielded Sutron TPG gauges were tested at the Marshall site (continental climate); the corresponding catch efficiency vs. mean wind speed results are plotted in Figure 4.14. The median catch efficiencies for the different configurations are similar for mean wind speeds ≤ 1 m/s, then decrease more rapidly with increasing wind speed for the unshielded gauge relative to the shielded gauge. For mean wind speeds between 5 m/s and 6 m/s, the median catch efficiency for the shielded Sutron TPG gauge is approximately 0.55, while that for the unshielded gauge is approximately 0.3. As these results are limited to a single test site, it is difficult to extrapolate these results to other climate regimes; however, it is expected that shielded configurations will mitigate the influence of wind-induced undercatch relative to unshielded configurations in other climate conditions.

4.1.1.3.3.1.2 Results by site (climate zone)

Wind speed impacts on catch efficiency for all test-gauge configurations at each site are presented in Figure 4.15 toFigure 4.23. In each figure, different-colored plots are overlaid for each test configuration and/or data output parameter. The Marshall site had the largest number of test configurations (see Table 4.1); in order to limit the number of test gauges plotted in a single figure, separate figures are presented for shielded and unshielded test configurations.

By presenting the results by site, rather than by gauge type (as in Section 4.1.1.3.3.1.1), the confounding effects of site-to-site differences in the environmental conditions, configuration, and data sampling on the interpretation of results are mitigated. Differences in individual gauge siting and installation/configuration may still play a role, as may the inhomogeneity/spatial variability of precipitation, but their impacts are expected to be relatively minor compared to site-to-site differences. Therefore, the results by site provide a more appropriate basis for the comparison of different gauge types and configurations for the specific subset of conditions experienced within each climate zone.

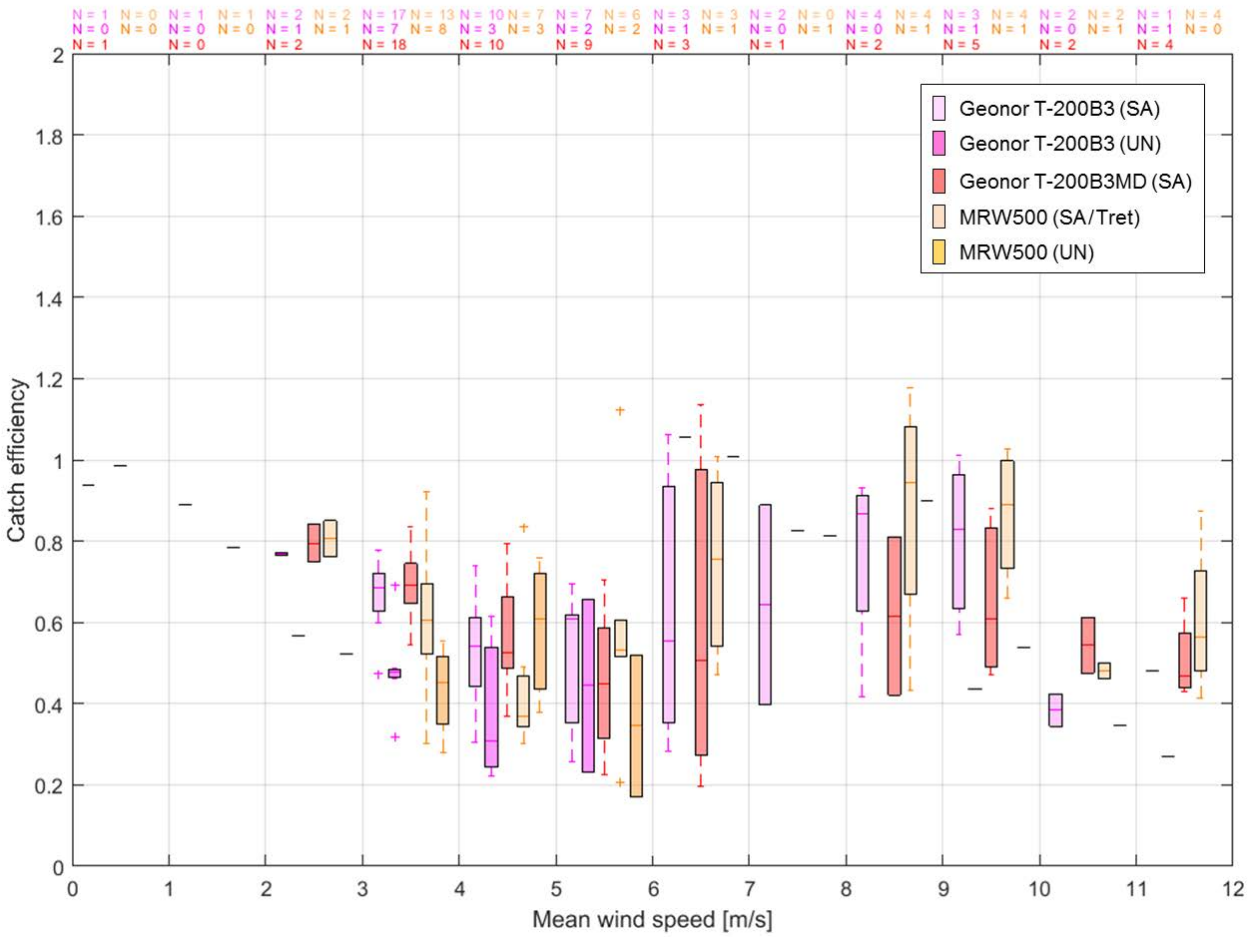


Figure 4.15. Box and whisker plot of catch efficiency as a function of mean wind speed for weighing gauges under test at Bratt's Lake. The following shield configurations were tested: single-Alter (SA), single-Alter/Tretyakov (SA/Tret), and unshielded (UN). The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

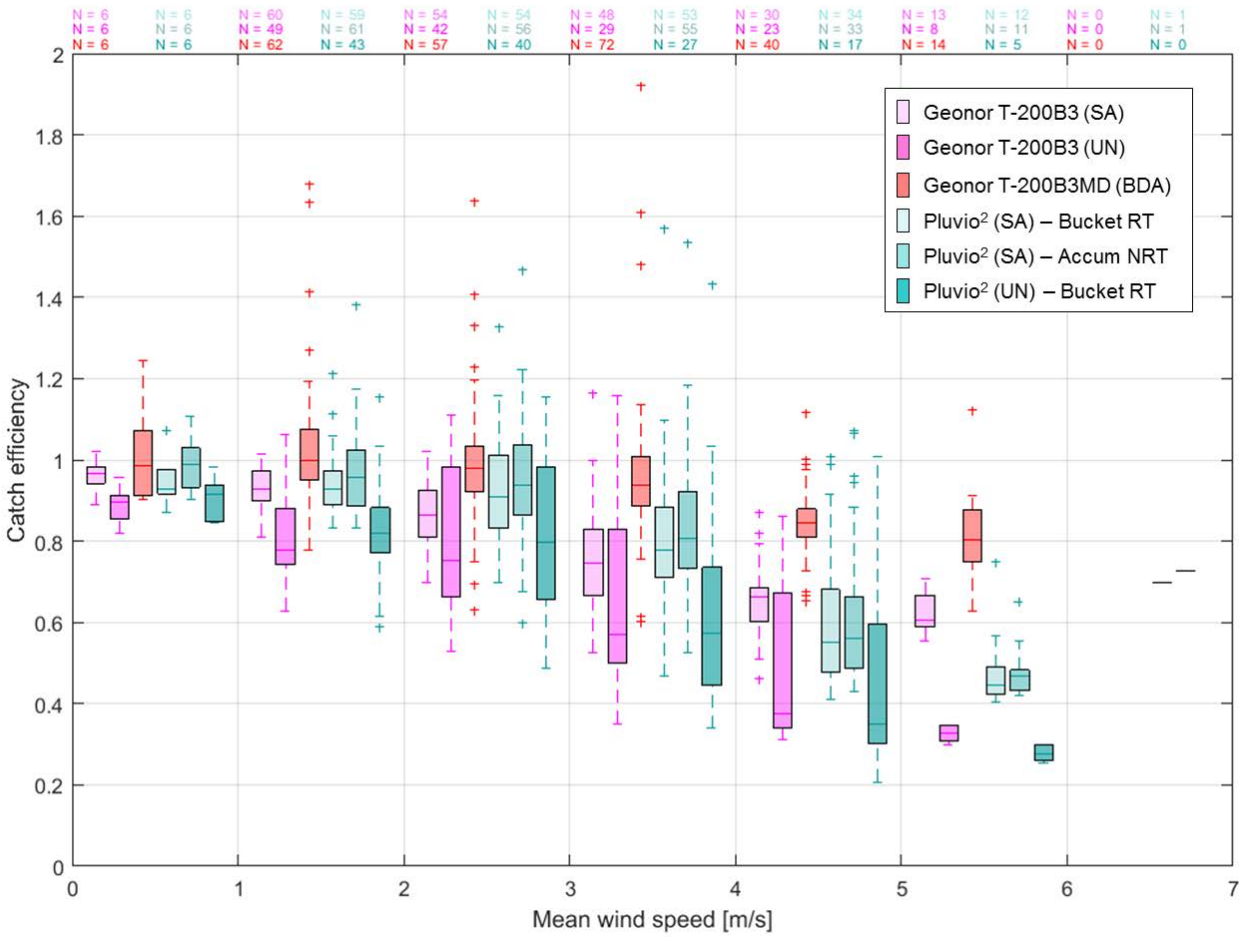


Figure 4.16. Box and whisker plot of catch efficiency as a function of mean wind speed for weighing gauges under test at CARE. The following shield configurations were tested: Belfort double-Alter (BDA), single-Alter (SA), and unshielded (UN). Results are presented for both the Bucket RT and Accumulated NRT data outputs from the single-Alter-shielded OTT Pluvio^2. The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

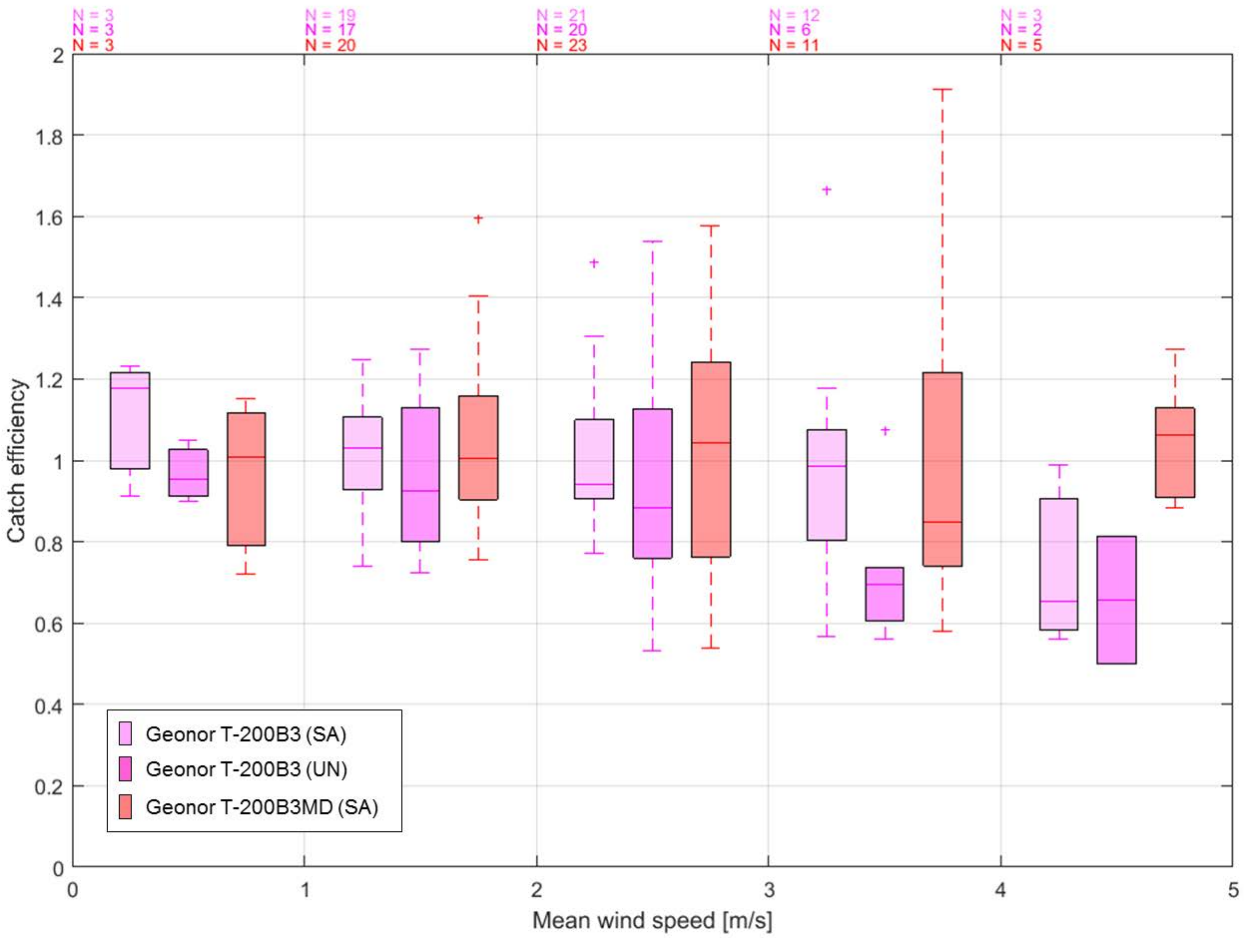


Figure 4.17. Box and whisker plot of catch efficiency as a function of mean wind speed for weighing gauges under test at Caribou Creek. Single-Alter (SA) shielded and unshielded (UN) configurations were tested. The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

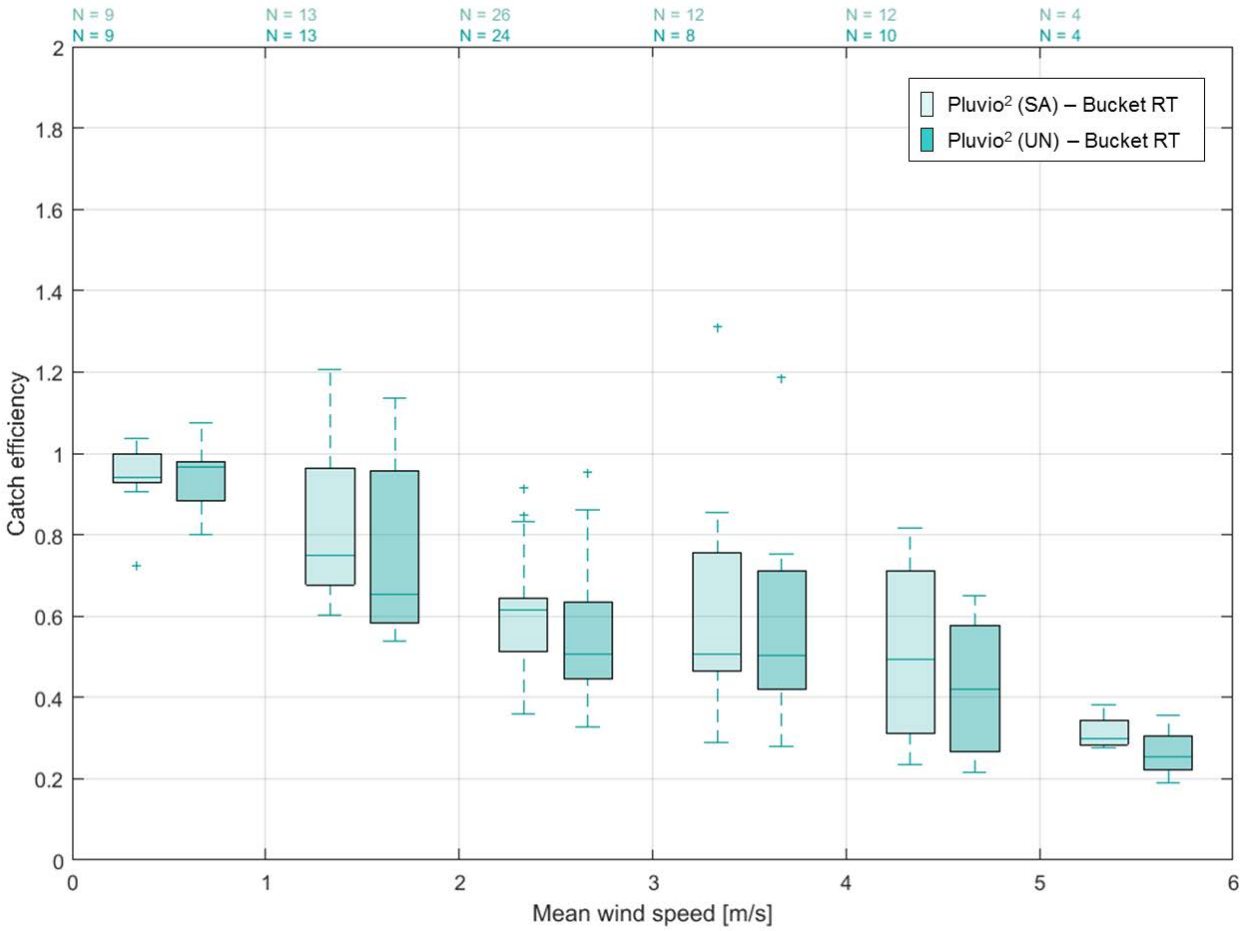


Figure 4.18. Box and whisker plot of catch efficiency as a function of mean wind speed for weighing gauges under test at Formigal. Single-Alter (SA) shielded and unshielded (UN) configurations were tested. The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

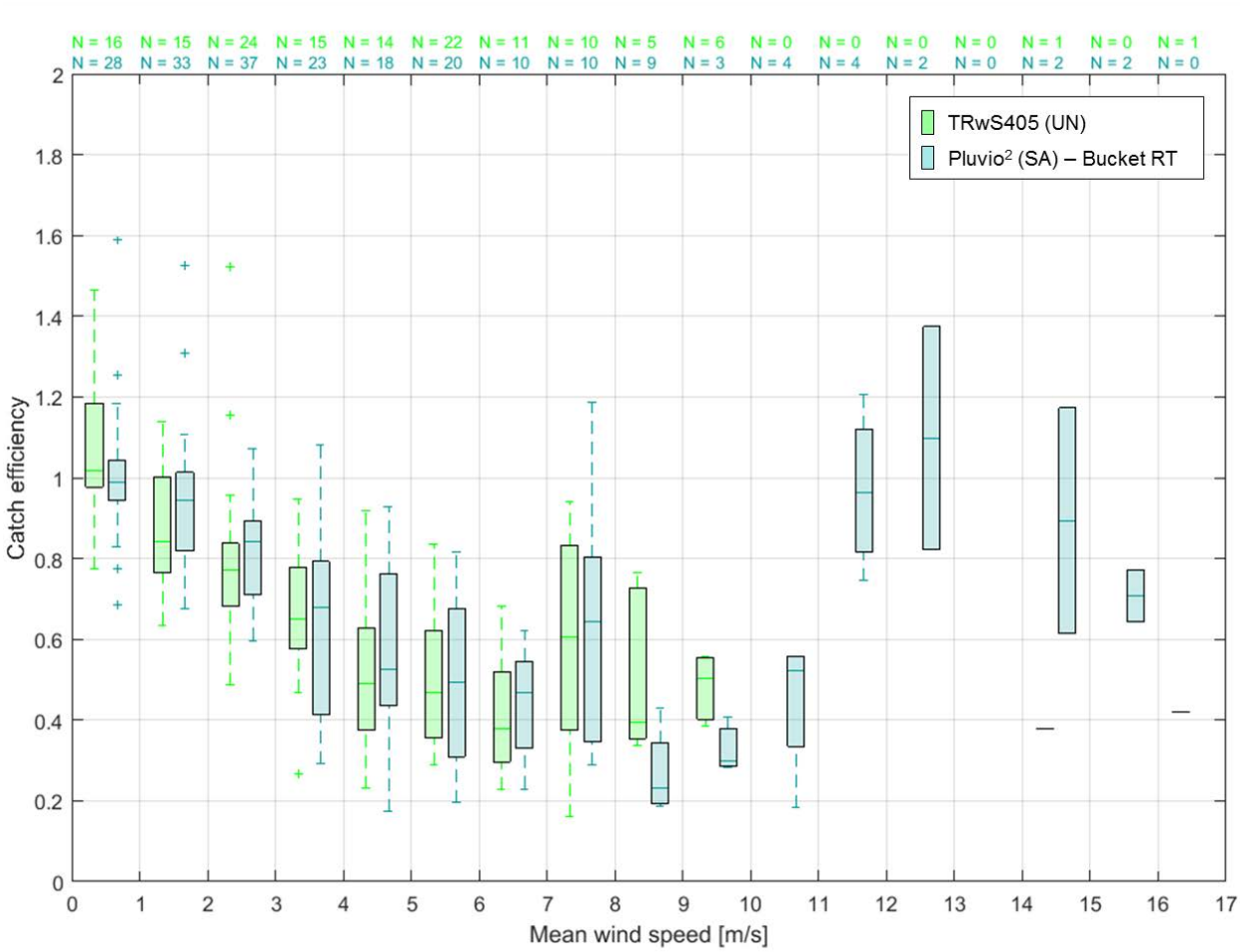


Figure 4.19. Box and whisker plot of catch efficiency as a function of mean wind speed for unshielded (UN) and single-Alter-shielded (SA) weighing gauges under test at Haukeliseter. The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

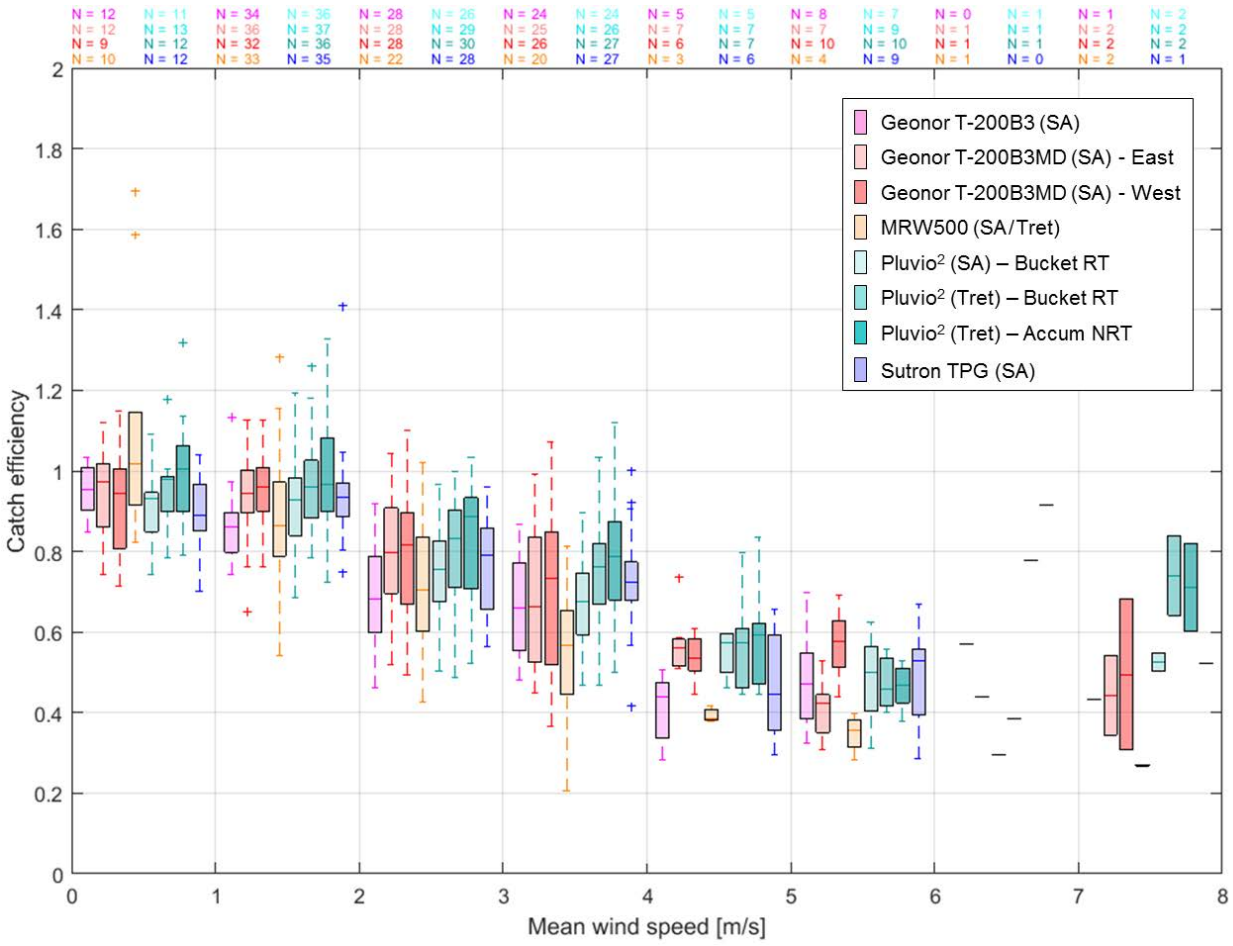


Figure 4.20. Box and whisker plot of catch efficiency as a function of mean wind speed for shielded weighing gauges under test at Marshall. Single-Alter (SA), Tretyakov (Tret), and single-Alter/Tretyakov (SA/Tret) shield configurations gauges were tested. Results are presented for both the Bucket RT and Accumulated NRT data outputs from the Tretyakov-shielded OTT Pluvio² gauge under test. The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

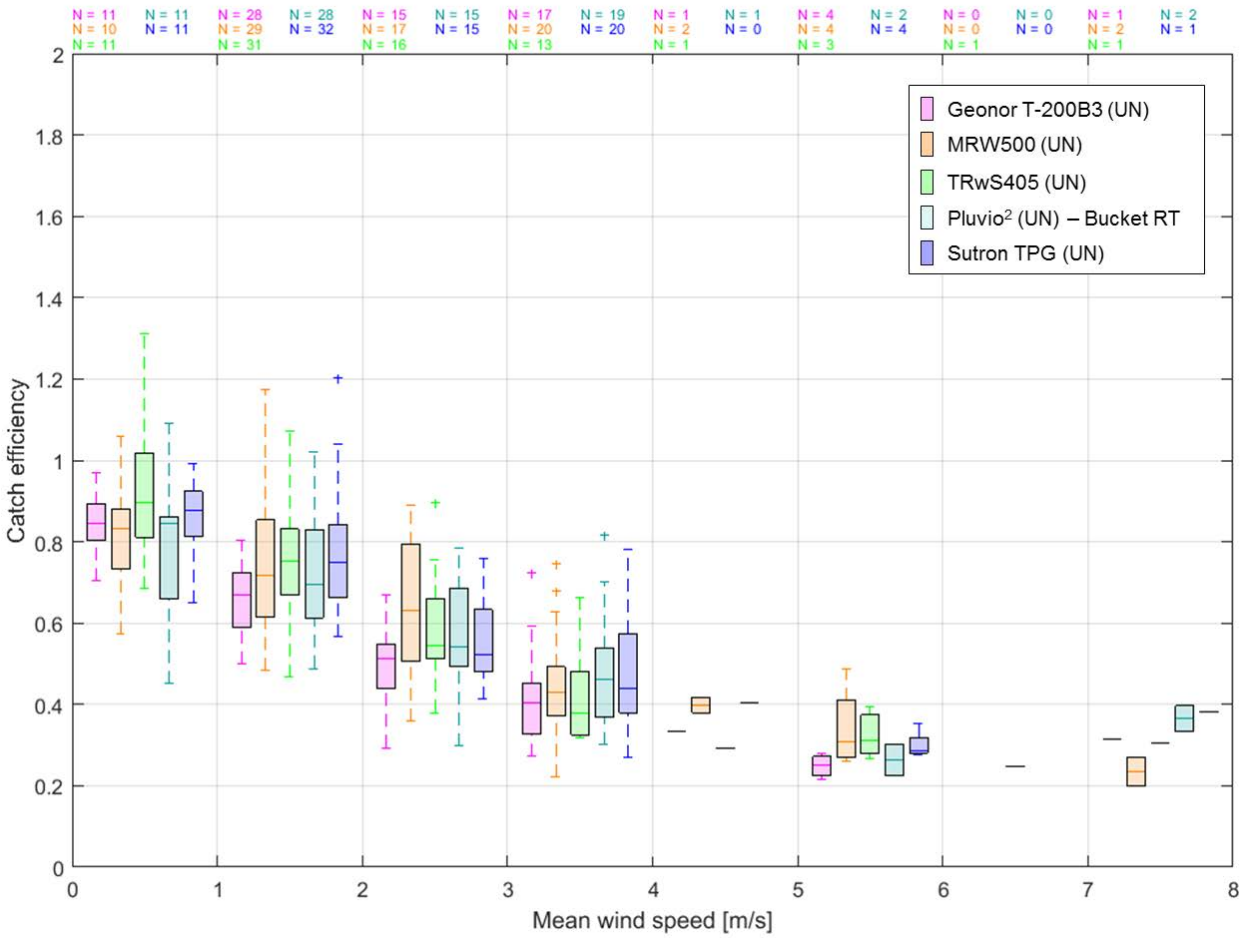


Figure 4.21. Box and whisker plot of catch efficiency as a function of mean wind speed for unshielded weighing gauges under test at Marshall. The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

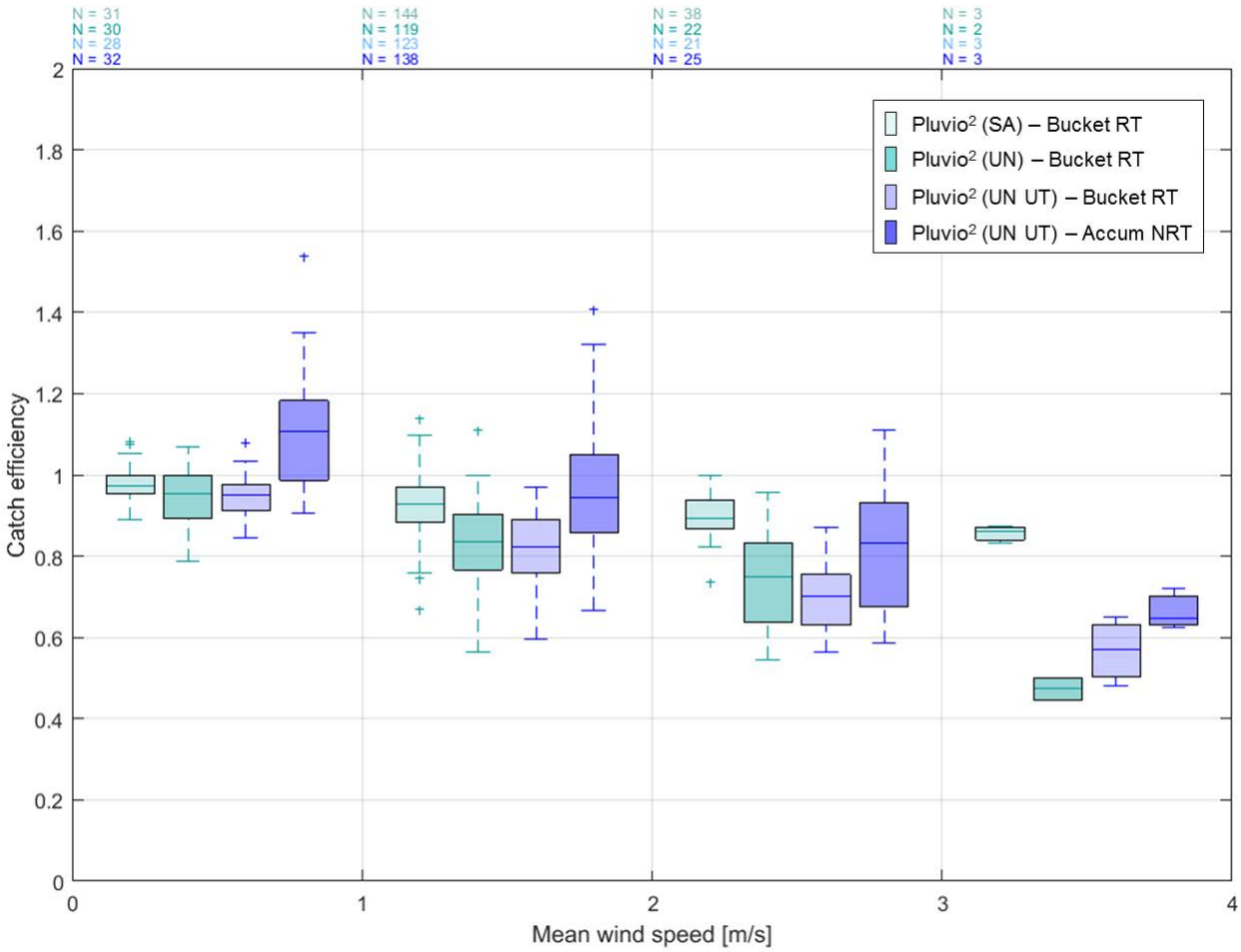


Figure 4.22. Box and whisker plot of catch efficiency as a function of mean wind speed for single-Alter-shielded (SA) and unshielded (UN) weighing gauges under test at Sodankylä. The UT designation in the legend indicates that one unshielded configuration was "under test" (submitted by the manufacturer for evaluation). For the UT configuration, results are presented for both Bucket RT and Accumulated NRT outputs. For all other configurations, results are presented for the Bucket RT output only. The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

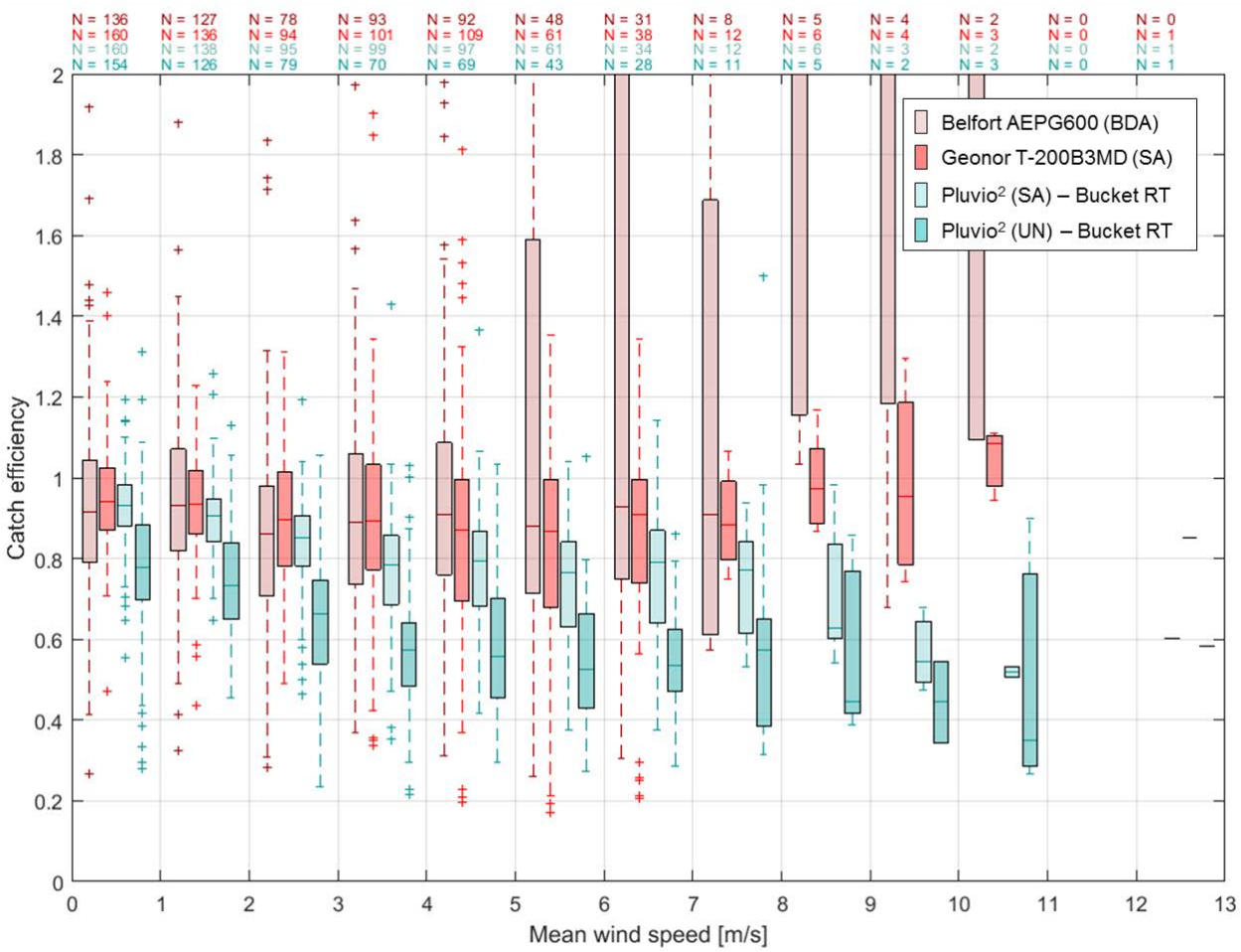


Figure 4.23. Box and whisker plot of catch efficiency as a function of mean wind speed for weighing gauges under test at Weissfluhjoch. The following shield configurations were tested: Belfort double-Alter (BDA), single-Alter (SA), and unshielded (UN). The number of precipitation events in each wind-speed bin for each test gauge is indicated above the plot.

4.1.1.3.3.1.2.1 Bratt's Lake (continental, high wind)

The assessment of trends in the catch efficiency vs. wind speed results for the test gauges at Bratt's Lake (Figure 4.15) is limited by the small numbers of events detected at mean wind speeds < 3 m/s and the apparent influence of blowing snow at mean wind speeds > 6 m/s. Over the intermediate range of mean wind speeds between 3 m/s and 6 m/s, the median catch efficiencies for the shielded gauges are all within about 0.1 to 0.2 of one another, falling from approximately 0.6 - 0.7 between 3 m/s and 4 m/s to approximately 0.4 - 0.6 between 5 m/s and 6 m/s. For the Geonor T-200B3 and Meteoservis MRW500 test gauges, the shielded configurations detect more precipitation events than the unshielded configurations. This makes it difficult to quantify trends for the unshielded configurations with confidence, but with that noted, the median catch efficiencies are still lower than those for the corresponding shielded configurations by at least about 0.2 in all mean wind-speed bins, with the exception of the MRW500 test gauges between 4 m/s and 5 m/s.

4.1.1.3.3.1.2.2 CARE (continental, humid)

The test gauge results for CARE (Figure 4.16) provide an example of how wind speed impacts on catch efficiency compare for different gauge types in the same shield configuration at the same site.

The median catch efficiencies for the single-Alter-shielded Geonor T-200B3 and OTT Pluvio^2 gauges follow similar trends with increasing mean wind speed; values are within ~ 0.1 to 0.15 in each wind-speed bin and decrease to values within ~ 0.44 to 0.6 at mean wind speeds between 5 m/s and 6 m/s. The median catch efficiencies for the unshielded Geonor T-200B3 and OTT Pluvio^2 test gauges also show similar agreement in each wind-speed bin, falling to lower values of ~ 0.28 to 0.34 at mean wind speeds between 5 m/s and 6 m/s. The difference in median catch efficiency between the shielded and unshielded configuration of each gauge type rises with increasing mean wind speed, from within about 0.1 below 3 m/s to within about 0.2 at higher speeds up to 6 m/s.

The median catch efficiency of the Belfort double-Alter-shielded Geonor T-200B3MD gauge remains above 0.8 for mean wind speeds up to 6 m/s. This increase in catch efficiency relative to the single-Alter-shielded gauges is attributed primarily to the double shield, based on previous results comparing weighing gauge performance in different shield configurations (Watson et al., 2008; Smith, 2009; Rasmussen et al., 2012; Kochendorfer et al., 2018); however, these findings may be confounded by noise in the data related to the specific gauge configuration (see IPR in Annex 6). Higher mean catch efficiencies are also observed in each mean wind-speed bin for the Accumulated NRT output from the single-Alter-shielded OTT Pluvio^2 relative to the Bucket RT output from the same gauge. The Bucket RT is close to the raw bucket weight of the gauge, while the Accumulated NRT is a processed output designed to mitigate the influence of evaporation and improve the reporting of light precipitation. As the NRT processing effectively increases the amount of precipitation reported, it is not unexpected that the observed catch efficiencies for the NRT output exceed those for the Bucket RT output. Caution must be exercised, however, when using the NRT output, given the potential for false or artificially increased accumulation reports resulting from the processing algorithm.

4.1.1.3.3.1.2.3 Caribou Creek (southern boreal)

The catch efficiency vs. mean wind speed relationships for the test gauges at Caribou Creek are shown in Figure 4.17. All test configurations – the single-Alter-shielded and unshielded Geonor T-200B3 gauges and single-Alter-shielded Geonor T-200B3MD gauge – report catch efficiencies > 1 at mean wind speeds < 3 m/s. Both shielded configurations continue to report CE values > 1 at mean wind speeds between 3 and 4 m/s, and the Geonor T-200B3MD gauge continues to report such values at mean wind speeds between 4 m/s and 5 m/s. This “overcatch” is attributed to the high magnitude of noise in measurements from Caribou Creek relative to those from other sites, which is believed to result from a ground noise issue (see corresponding IPRs in Annex 6). For the Geonor T-200B3MD, this may be exacerbated by noise related to the configuration, as the larger, taller bucket relative to the Geonor T-200B3 may be more susceptible to wind-induced vibration. While the relative magnitudes of noise may differ from gauge to gauge, it appears that the median catch efficiencies of the shielded gauges exceed those of the unshielded gauge by about 0.1 to 0.3 in each mean wind-speed bin up to 4 m/s.

4.1.1.3.3.1.2.4 Formigal (alpine, maritime influence)

The median catch efficiency of the single-Alter-shielded OTT Pluvio^2 test gauge at Formigal exceeds that of the unshielded gauge of the same type for mean wind speeds up to 6 m/s (Figure 4.18). The median catch efficiencies for the shielded and unshielded configurations are within 0.1 in each mean wind-speed bin, and fall to values within 0.2 to 0.4 for mean wind speeds up to 6 m/s. The number of solid-precipitation events may be underestimated by the SPICE classification, in which the maximum temperature over the 30-minute interval does not exceed -2 °C. Work by Buisán et al. (2017) has

demonstrated that many events classified as snow by an optical sensor at Formigal corresponded with temperatures > -2 °C. As such, the results presented in Figure 4.18 should be considered representative of performance in solid precipitation as defined within SPICE, and not necessarily representative of performance in all solid precipitation conditions experienced at Formigal.

4.1.1.3.3.1.2.5 Haukeliseter (alpine, high wind)

The unshielded MPS TRwS405 and single-Alter-shielded OTT Pluvio^2 test gauges at Haukeliseter show similar decreases in the median catch efficiency with increasing mean wind speed (Figure 4.19), with values falling from approximately 1 to between 0.4 and 0.5 over the mean wind speed range from 1 m/s to 7 m/s. The median catch efficiencies for both test configurations are within 0.1 in each 1 m/s mean wind-speed bin over this range, with the value for the shielded gauge exceeding that for the unshielded gauge in each bin. Higher catch efficiencies for the shielded gauge can be attributed to the presence of the shield, which mitigates the influence of wind speed on gauge catch. The close proximity of median catch-efficiency values over the previously noted wind speed range, despite the difference in shielding, may be attributed to the influence of TRwS405 data processing, which the manufacturer states is able to compensate for the effects of evaporation, temperature, and wind.

For mean wind speeds between 7 m/s and 8 m/s, the median catch-efficiency values for both test gauges increase to between 0.6 and 0.7, which may be related to blowing-snow events. For mean wind speeds between 8 m/s and 10 m/s, the median catch efficiencies for the shielded OTT Pluvio^2 gauge fall to between 0.2 and 0.3, while those for the unshielded MPS TRwS405 gauge drop to between 0.4 and 0.5. This reversal in trend for median catch efficiencies relative to lower mean wind speeds is likely related to differences in the processing of measurements, with the Bucket RT output from the Pluvio^2 being close to a raw output, while the output from the TRwS405 is processed to mitigate the contributions of evaporation, temperature, and wind, as noted previously. At mean wind speeds greater than 10 m/s, only the shielded gauge reports precipitation, with apparent contributions from blowing snow.

4.1.1.3.3.1.2.6 Marshall (continental, dry)

The mean wind speed dependence of catch efficiencies for shielded and unshielded test gauges at the Marshall site are plotted in Figure 4.20 and Figure 4.21, respectively. The plots were separated for clarity, but provide further perspective on the relative performance of gauges in the same shield configuration at the same site. All of the test gauges in Figure 4.20 are in single-shield configurations. For mean wind speeds up to 6 m/s (above which the number of events detected is limited), the median catch efficiencies for all test gauges, including five different gauge types, three single-shield types, and two Pluvio^2 data outputs, all agree within 0.1 to 0.2 in each wind-speed bin. The median catch efficiencies decrease with increasing mean wind speed, from above 0.8 up to 2 m/s, to above approximately 0.6 between 2 m/s and 4 m/s, to above about 0.4 between 4 m/s and 6 m/s. The median catch efficiencies for the OTT Pluvio^2 test configurations and outputs show the following trend in each mean wind-speed bin up to 5 m/s: single-Alter (Bucket RT) < Tretyakov (Bucket RT) < Tretyakov (Accumulated NRT). The highest median catch efficiencies are observed for the Accumulated NRT data, which are processed to compensate for temperature, wind, and evaporation, and include contributions from light precipitation.

The unshielded test gauges at Marshall also show very similar performance in terms of median catch efficiency dependence on mean wind speed (Figure 4.21). The median catch efficiencies of all unshielded test gauges agree within 0.1 in each mean wind-speed bin up to 4 m/s, over which range their values fall from approximately 0.8 – 0.9 to 0.4 – 0.5. The numbers of events detected drop off

at higher mean wind speeds, but the available results suggest that the median catch efficiencies are between about 0.2 and 0.3 for mean wind speeds between 5 m/s and 6 m/s.

4.1.1.3.3.1.2.7 Sodankylä (northern boreal, low wind)

For the sheltered, northern boreal site of Sodankylä, the assessment of wind-speed impacts on weighing gauge catch efficiency is limited to mean wind speeds below 3 m/s, as there are few events detected at higher mean wind speeds (Figure 4.22). The median catch efficiencies of the two unshielded OTT Pluvio^2 test gauges (Bucket RT outputs) agree within 0.05 in each wind-speed bin and remain at or above 0.7 for mean wind speeds within 3 m/s. The median catch efficiency of the single-Alter-shielded Pluvio^2 gauge exceeds those of the unshielded test gauges (Bucket RT outputs) by up to 0.2 at mean wind speeds between 2 m/s and 3 m/s. The processed Accumulated NRT output from one of the unshielded test gauges shows median catch efficiencies exceeding those determined using the Bucket RT output from the same test gauge, remaining at or above 0.9 in all mean wind-speed bins up to 3 m/s. The median catch efficiencies determined from the Accumulated NRT output exceed even those from the single-Alter-shielded test gauge at mean wind speeds below 2 m/s; however, the vertical extents of the boxes and whiskers for the Accumulated NRT output indicate catch efficiencies > 1. This suggests that the processing may overestimate precipitation amounts reported by OTT Pluvio^2 gauges.

4.1.1.3.3.1.2.8 Weissfluhjoch (Alpine, complex topography)

Four different gauge configurations were tested at Weissfluhjoch; the corresponding catch efficiencies relative to the reference configuration are plotted as a function of the mean wind speed in Figure 4.23. The catch efficiencies for solid precipitation events detected by the Belfort AEPG 600 gauge in Belfort double-Alter shield show significant variability at mean wind speeds > 5 m/s; however, the median values agree with those observed for the single-Alter-shielded Geonor T-200B3MD gauge within about 0.03 for mean wind speeds up to 8 m/s. Over this range of wind speeds, the median catch efficiencies reported by these test gauges remain within about 0.8 to 0.9. The single-Alter-shielded OTT Pluvio^2 test gauge shows similar performance; while the reported median catch efficiencies are evidently lower than those reported by the Geonor gauge in the same configuration (median values between about 0.7 to 0.9 over the same range of mean wind speeds), the median values agree within 0.1 in each wind-speed bin. The influence of shielding is apparent by comparison of the results for the shielded and unshielded OTT Pluvio^2 test gauges. The median catch efficiencies reported by the unshielded Pluvio^2 gauge are lower by up to about 0.25 for mean wind speeds up to 8 m/s with median values falling to as low as 0.5 over this range.

4.1.1.3.3.1.3 Summary of catch efficiency results for all test gauges, sites

To encapsulate the catch efficiency results presented in Figure 4.6 to Figure 4.23 and discussed in Sections 4.1.1.3.3.1.1 and 4.1.1.3.3.1.2, median values are consolidated for each site and shield configuration in Table 4.5. Results are presented for mean wind-speed ranges that increase in 2 m/s increments from 2 m/s to 10 m/s. Given the variability in results when the number of precipitation events per 1 m/s mean wind-speed bin is small, the results presented for each mean wind-speed range consider only the constituent 1 m/s bins with at least 10 events.

As median catch efficiencies for gauges in the same shield configuration at a given site show similar wind-speed dependency, the results in Table 4.5 represent the lowest median catch-efficiency value among test gauges in the same general shield configuration (double-shield, single-shield, and/or unshielded) at each site. The tabulated values can, therefore, be considered lower limits for median catch efficiency in each applicable configuration, at each site, over each range of mean wind speeds.

These values are intended to be used as a general reference and should not be taken as absolute. The box and whisker plots for individual gauges/sites (Figure 4.6 to Figure 4.23) provide a better indication of the range of potential catch efficiencies for each gauge, configuration, and site.

The requirement of 10 precipitation events per 1 m/s bin limits the breadth of results presented inTable 4.5. For example, this requirement is met for only one of the test configurations at the Bratt's Lake site, and for a single mean wind-speed bin. Blank entries for a given site, configuration, and mean wind-speed range indicate that an insufficient number of events was available from which to determine a value using the selected assessment approach. Please refer to the figures and discussion in Sections 4.1.1.3.3.1.1 and 4.1.1.3.3.1.2 and the instrument performance reports (Annex 6) for additional details regarding a particular test gauge configuration and site.

Table 4.5. Summary of catch efficiency results for weighing gauges in each shield configuration at each site in solid precipitation (30-minute events with maximum temperatures below -2 °C). Median catch efficiency values are rounded to the nearest 0.1 and represent the lowest median CE value among gauges in the same shield configuration at a given site and within a given mean wind-speed range. Results presented for each mean wind-speed range consider only the constituent 1 m/s bins with at least 10 precipitation events.

Site (climate)	Shield configuration	Median catch efficiency				
		U < 2 m/s	U < 4 m/s	U < 6 m/s	U < 8 m/s	U < 10 m/s
Formigal (alpine, maritime influence)	Single-shield	0.8	0.5	0.5		
	Unshielded	0.7	0.5			
Haukeliseter (alpine, high winds)	Single-shield	1	0.7	0.5	0.5	
	Unshielded	0.8	0.6	0.5	0.4	
Weissfluhjoch (alpine, complex topography)	Double-shield	0.9	0.9	0.9	0.9	
	Single-shield	0.9	0.8	0.8	0.8	
	Unshielded	0.7	0.6	0.5	0.5	
Bratt's Lake (continental, high wind)	Single-shield		0.5			
	Unshielded					
CARE (continental, humid)	Double-shield	1	0.9	0.8		
	Single-shield	0.9	0.8	0.4		
	Unshielded	0.8	0.6	0.4		
Marshall (continental, dry)	Single-shield	0.9	0.6	0.5		
	Unshielded	0.7	0.4			
Caribou Creek (southern boreal)	Single-shield	1	0.8			
	Unshielded	0.9	0.9			
Sodankylä (northern boreal)	Single-shield	1	0.9			
	Unshielded	0.8	0.7			

4.1.1.3.3.2 Overall catch efficiency

The catch efficiency results in Section 4.1.1.3.3.1 are based on 30-minute assessment intervals over which the mean wind speed and precipitation type (as determined from the temperature) are considered to be representative. This type of assessment is useful for characterizing gauge performance under specific environmental conditions, but does not reflect the overall ability of a test gauge to report accumulated precipitation over longer time scales (e.g. seasonal). For this purpose, the overall catch efficiency is calculated by summing the test and reference accumulations over all 30-minute intervals during which both report precipitation (YY cases).

The overall catch efficiency of each test gauge computed from YY cases in all precipitation types (liquid, mixed, solid) is presented in Figure 4.24. These values are representative of performance over the full range of precipitation conditions experienced at each test site over the duration of formal tests. (Mean event conditions for each site are provided in Table 4.1) The CE values will be impacted significantly by the relative proportion of events of each precipitation type in Table 4.3 (i.e. overall catch efficiencies are likely higher at sites with a larger proportion of liquid events).

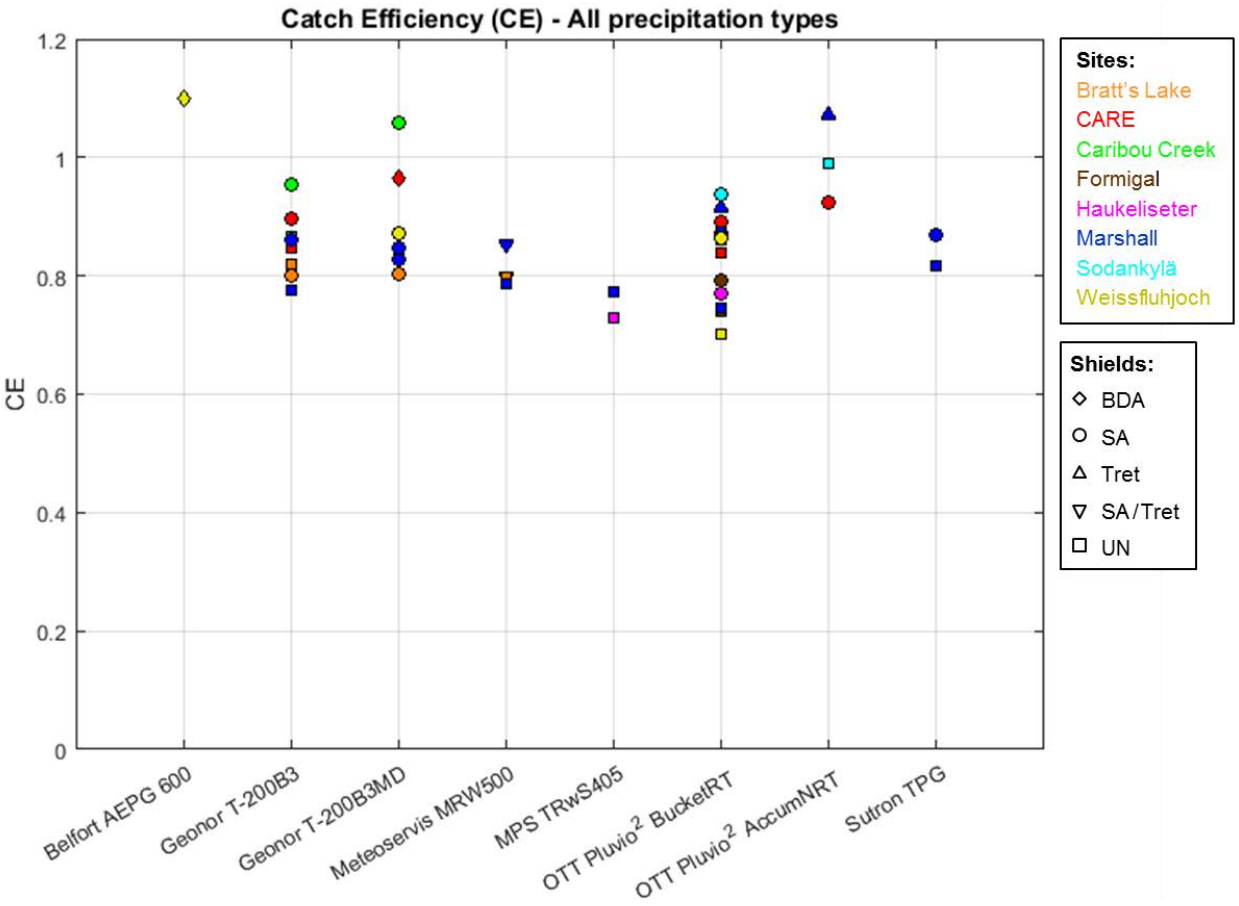


Figure 4.24. Overall catch efficiency of weighing gauges, calculated from 30-minute YY events in all precipitation types. Gauges were tested in the following shield configurations: Belfort double-Alter (BDA); single-Alter (SA); Tretyakov (Tret); single-Alter/Tretyakov (SA/Tret); and unshielded (UN).

The overall catch efficiencies for weighing gauges under test in WMO-SPICE all exceed 0.7, with the lowest values for unshielded gauges. All shielded test configurations (with the exception of the SA-shielded OTT Pluvio^2 under test at Haukeliseter, which is influenced by high winds) have overall catch efficiencies of 0.8 or higher. Specific test gauges and sites are correlated with lower overall catch efficiencies, specifically, the gauges at Bratt's Lake and Haukeliseter, due to the high winds experienced at these sites. Other test gauges/outputs and sites are correlated with higher overall catch efficiencies: the gauges at Caribou Creek and the Belfort AEPG 600 at Weissfluhjoch, due to noise/variability in the gauge outputs; the gauges at Sodankylä, due to low winds and characteristically light events; and the Accumulated NRT output from OTT Pluvio^2 gauges at Marshall and Sodankylä, which may result from the data-processing approach (potential for false precipitation reports). For the Tretyakov-shielded OTT Pluvio^2 gauge at Marshall, the overall CE may be impacted further by snow accumulating on the fixed shield and blowing into the gauge orifice.

Figure 4.25 shows the overall catch efficiency of each test gauge for solid-precipitation events, computed from only the YY cases in which the maximum temperature did not exceed -2 °C. The mean snow-event characteristics differ by site, as outlined in Table 4.3. Considering only snow events, the overall catch-efficiency values for each test gauge follow the same general trends as the values for all precipitation types (Figure 4.24); however, the influence of configuration, site, and/or data outputs are more pronounced, resulting in a broader spread of CE values. For unshielded gauges at the sites characterized by the highest mean wind speeds (Bratt's Lake, Haukeliseter), the overall catch efficiencies are as low as 0.44. For all test gauges in both SA-shielded and unshielded configurations at the same site, the overall catch efficiency of the shielded gauge exceeds that of the unshielded gauge by 0.1 to 0.2.

These results in Figure 4.24 and Figure 4.25 indicate that the overall catch efficiency is affected more by the site-specific environmental conditions and configuration, the presence or absence of a shield, and the extent of data processing than the specific type of weighing gauge, its heating configuration, and/or its principle of operation (transducer type). The size of the shield also likely plays a role. While there are no sites at which the same gauge is installed in both a double- and single-shield, the higher overall catch efficiency observed for the BDA-shielded Geonor T-200B3MD gauge at CARE relative to the SA-shielded Geonor T-200B3 gauge suggests that double-shielded gauges may be more effective than single-shielded gauges in solid-precipitation conditions. Indeed, higher catch efficiencies for BDA-shielded gauges relative to SA-shielded and unshielded gauges have been reported elsewhere (Watson et al., 2008; Smith, 2009; Rasmussen et al., 2012; Kochendorfer et al., 2018).

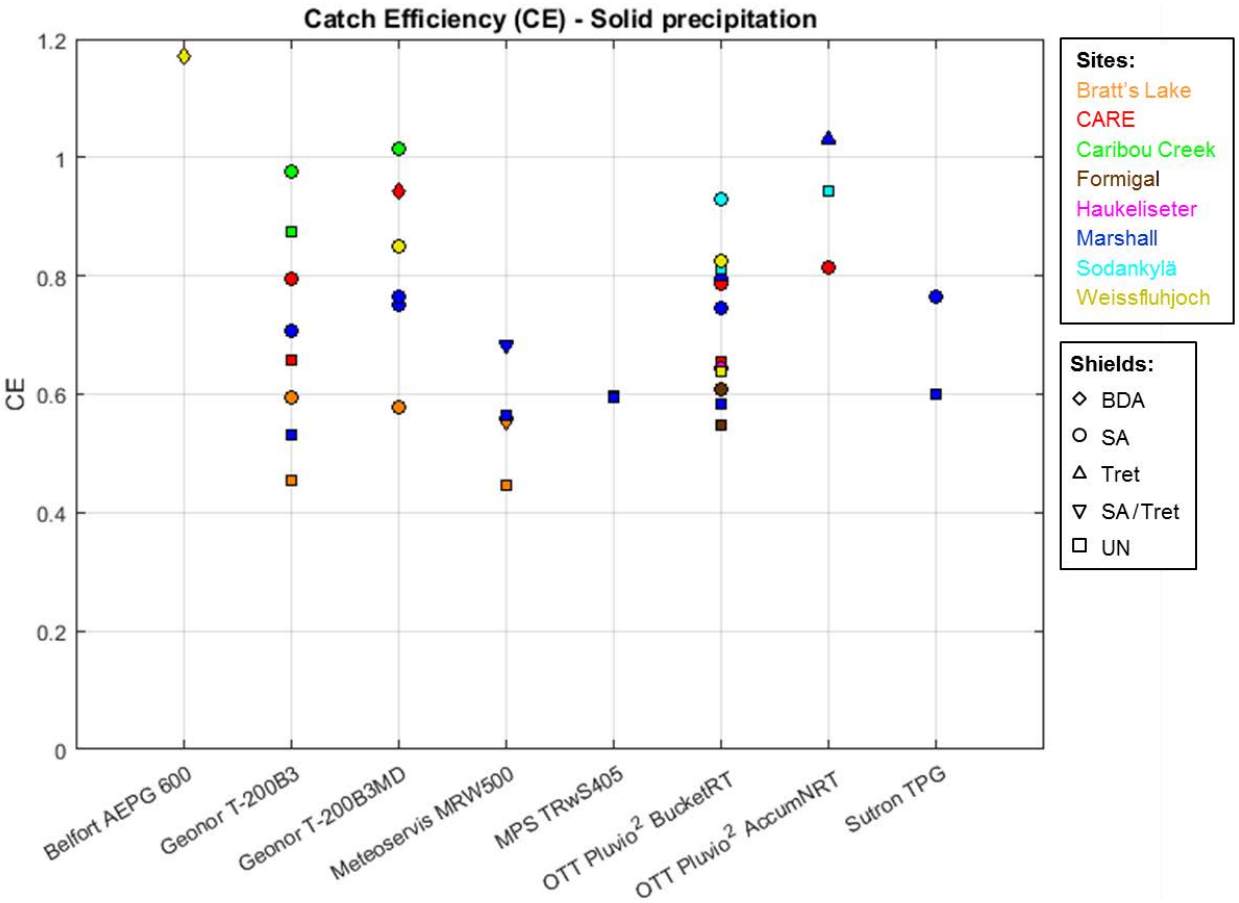


Figure 4.25. Overall catch efficiency of weighing gauges, calculated from 30-min YY events in solid precipitation only. Gauges were tested in the following shield configurations: Belfort double-Alter (BDA); single-Alter (SA); Tretyakov (Tret); single-Alter/Tretyakov (SA/Tret); and unshielded (UN).

4.1.1.3.3.3 Root mean square error

Root mean square error values are computed using YY cases for all events (results for all precipitation types) and for snow events only; the results are presented for all test gauges in Figure 4.26 and Figure 4.27, respectively. Note that these results are computed using the same events as for the overall CE results in Figure 4.24 and Figure 4.25, but consider the absolute difference in accumulated precipitation reports from the test gauge and reference. As such, the RMSE results can be considered to represent the absolute uncertainty of each test gauge relative to the reference over 30-minute assessment intervals in all precipitation conditions (all types) and in snow only.

The results in Figure 4.26 show that the RMSE varies by gauge, by site, and by configuration, with values between 0.1 mm/30 min and 0.6 mm/30 min for most gauges under test. The RMSE values for the test gauges at Sodankylä are all below 0.1 mm/30 min, which is attributed to the characteristically light precipitation events and low wind speeds. The RMSE value for the Belfort AEPG 600 gauge at Weissfluhjoch exceeds 1 mm/30 min, which is believed to result from the high data variability/noise observed for this particular gauge over the duration of the experiment. The relative magnitudes of RMSE values for a given test gauge vary depending on the characteristics of precipitation events (Table 4.2 and Table 4.4) and configurations at the sites. For a given site with shielded and unshielded gauges of the same type, the RMSE of the shielded gauge is always lower.

The RMSE results for solid precipitation in Figure 4.27 cover a broader range of values, the stratification of which can be linked to the snow-event characteristics in Table 4.3. Gauges at test sites with the lowest mean accumulations and wind speeds (Caribou Creek, Sodankylä) have the lowest RMSE values, within about 0.13 mm/30 min. Gauges at test sites with higher mean accumulations (Formigal) and/or higher mean wind speeds (Bratt's Lake, Haukeliseter) have higher RMSE values, above 0.4 mm/30 min. The Belfort AEPG gauge at Weissfluhjoch again shows results that differ from the other test gauges, with a RMSE of almost 1.2 mm/30 min (attributed to gauge variability/noise). RMSE values are lower for shielded gauges relative to unshielded gauges at a given site. Double-shielded gauges are not well-represented, but the results for test gauges at CARE suggest that double shielded gauges may reduce RMSE values relative to single-shielded and unshielded configurations. The Tretyakov-shielded OTT Pluvio^2 gauge under test at Marshall shows a higher RMSE than the other test gauges at the site. This is likely due to false precipitation resulting from the processing approach and/or snow accumulating on the fixed shield and blowing into the gauge orifice.

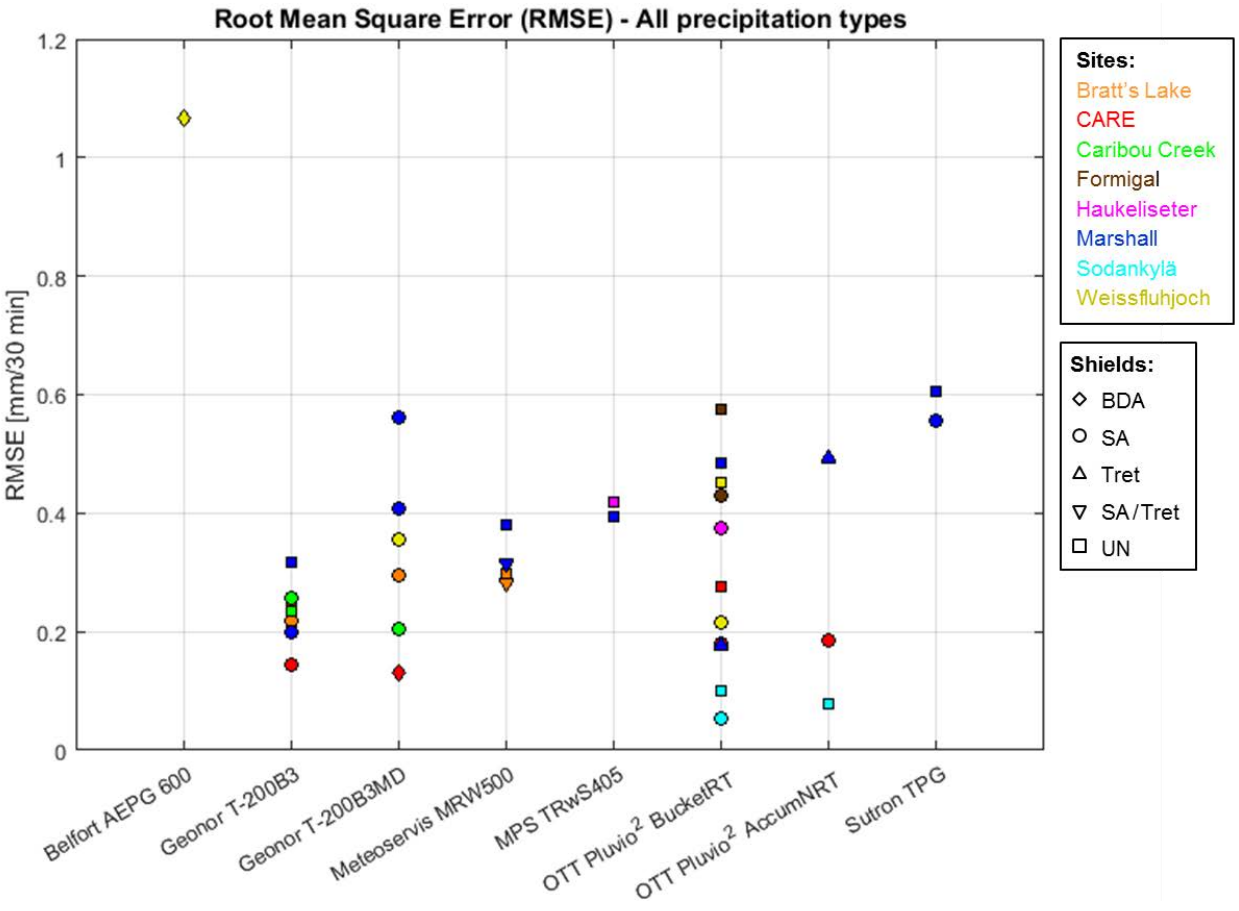


Figure 4.26. Root mean square error of accumulation reports from weighing gauges calculated from 30-minute YY events in all precipitation types. Gauges were tested in the following shield configurations: Belfort double-Alter (BDA); single-Alter (SA); Tretyakov (Tret); single-Alter/Tretyakov (SA/Tret); and unshielded (UN).

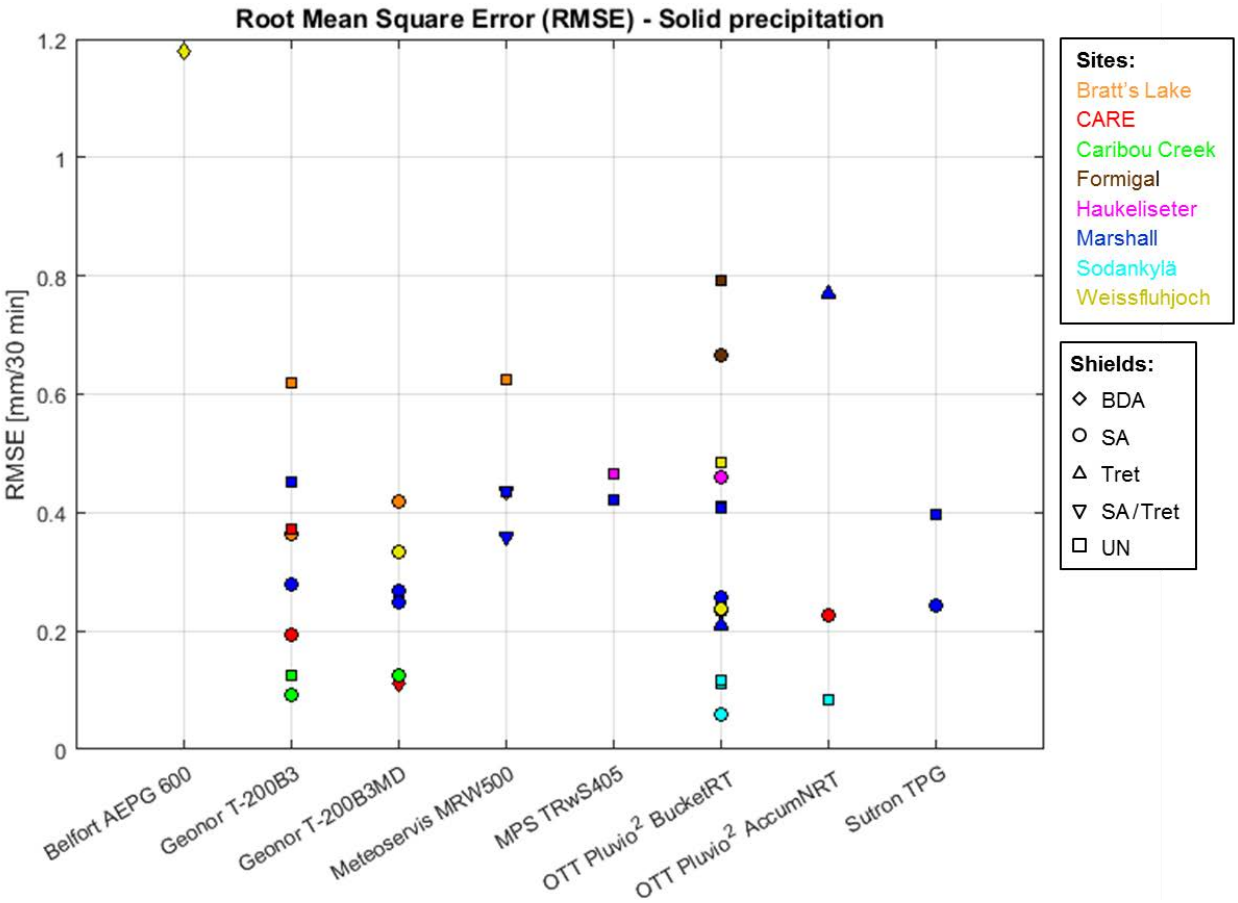


Figure 4.27. Root mean square error of accumulation reports from weighing gauges calculated from 30-minute YY events in solid precipitation only. Gauges were tested in the following shield configurations: Belfort double-Alter (BDA); single-Alter (SA); Tretyakov (Tret); single-Alter/Tretyakov (SA/Tret); and unshielded (UN).

4.1.1.3.4 *Summary of key results*

4.1.1.3.4.1 Detection of precipitation relative to reference

The ability of a test-gauge configuration to detect precipitation relative to the site-reference configuration over 30-minute assessment intervals was characterized using selected skill scores in Section 4.1.1.3.2. The scores showed general trends depending on the specific site and gauge configuration, processing of data outputs, and/or the characteristic site conditions, as outlined below.

- The probability of detection of test gauges is typically higher for shielded gauges relative to unshielded gauges at a given site and for processed data outputs that adjust for the influences of evaporation, temperature, wind, and/or trace/light precipitation. POD values are lower for test gauges at sites with characteristically higher mean wind speeds, which are subject to greater wind-induced undercatch.
- The false alarm rate of test gauges does not appear to be impacted significantly by gauge shielding, but is higher for gauge/site configurations that are more susceptible to noise or are subject to baseline noise of higher magnitude (e.g. noise from electrical interference), which can artificially increase the reported accumulation above the detection threshold. FAR values

are higher for processed outputs relative to unprocessed outputs for test gauges in the same configuration and at the same site, which may indicate artificial enhancement of the accumulation output and/or false reports by the processing algorithm.

- The bias is lower for test gauges at sites characterized by higher mean wind speeds and higher for gauge/site configurations that are more susceptible to noise or are subject to baseline noise of higher magnitude. These factors serve to decrease and increase the number of events detected relative to the reference, respectively. Gauge processing can also increase the number of events detected relative to the reference, increasing bias values relative to unprocessed outputs for test gauges of the same type and configuration at the same site.
- The Heidke Skill Score considers overall detection skill relative to chance and is typically higher for shielded test gauges relative to unshielded gauges at a given site, lower for gauges at sites with characteristically higher mean wind speeds, and lower for gauge/site configurations that are more susceptible to noise or are subject to baseline noise of higher magnitude.

4.1.1.3.4.2 Reporting of accumulated precipitation relative to reference

The ability of test gauges to report accumulated precipitation relative to the reference configuration was assessed in terms of the catch efficiency for 30-minute assessment intervals. The focus was primarily on assessing the influence of mean wind speed on the catch efficiency for solid precipitation over each assessment interval, with the results presented both by gauge type (Section 4.1.1.3.3.1.1) and by site (Section 4.1.1.3.3.1.2). Overall performance in all precipitation types and for snow events only was characterized using the overall catch efficiency and root mean square error in Sections 4.1.1.3.3.2 and 4.1.1.3.3.3, respectively. Key findings from the assessment are presented below.

- The same test gauges in identical configurations can show significantly different performance at different sites. It is difficult to assess site-to-site differences in results, given differences in environmental conditions experienced at each site and differences in how gauges are installed, configured, and sampled at each site.
- Comparison of results from different test-gauge configurations at a given site provides more insight into the relative performance of different gauges/configurations/data outputs. All test gauges are subject to the same conditions, make use of the same site infrastructure, and are generally installed, maintained, and sampled in similar ways. Differences in configuration and/or sampling among gauges may still exist, and the precipitation reported by each gauge is not necessarily the same due to differences in siting and/or the spatial variability of precipitation; however, the effects of these factors are expected to be much less significant than those resulting from site-to-site differences.
- The catch efficiency of test gauges typically decreases with increasing mean wind speed for 30-minute assessment intervals. The decrease in catch efficiency is more gradual for shielded gauges than unshielded gauges, resulting in the detection and reporting of more events at higher mean wind speeds by shielded configurations. Shielding also reduces the variability of catch-efficiency values for test gauges within a given wind-speed bin, as indicated by smaller interquartile ranges for shielded configurations.
- Gauges in the same configuration (shielded or unshielded) at the same site show similar catch efficiency trends with increasing mean wind speed, suggesting that for a given site, the

choice of shield configuration is the most important aspect of weighing gauge selection, and the specific manufacturer, model, heating configuration, and operating principle are less important.

- This finding is reinforced by the overall catch efficiency and root mean square error results; shielded gauges show systematically higher overall catch efficiencies and lower root mean square errors relative to unshielded gauges of a given type at a particular site. These trends are observed whether results are considered for all precipitation types or for snow events only.
- The overall catch efficiency values decrease with increasing mean event wind speeds (Table 4.2 and Table 4.3), increase with increasing noise magnitude, and increase for processed data outputs relative to unprocessed outputs. Overall catch efficiencies are typically > 0.7 for unshielded gauges and > 0.8 for shielded gauges in all precipitation types. Values generally decrease for snow events relative to all precipitation types at a given site, with values for unshielded gauges at windy sites as low as approximately 0.4.
- The root mean square error values increase with increasing mean event accumulation and increasing mean event wind speed (Table 4.2 and Table 4.3). The RMSE values can be considered to represent the absolute uncertainty of each test gauge configuration; values vary from about 0.05 to 0.6 in all precipitation types, and from 0.05 to 0.8 for snow events only.
- The influence of data processing on results is twofold: by taking into account the contributions of wind, temperature, evaporation, and/or trace precipitation, it is possible to improve the detection of precipitation and increase the overall catch efficiency; however, the specific algorithms employed may overestimate the adjustments in some cases, leading to false reports or accumulation reports and catch efficiencies that have been increased artificially and, therefore, erroneously.

4.1.1.4 Recommendations

Informed by the results and interpretation presented in Section 4.1.1.3 and in the Instrument Performance Reports (Annex 6), as well as the knowledge and experience of the SPICE site teams, recommendations for weighing gauge selection, configuration, and data processing are provided. These recommendations are intended to guide individuals and organizations in the implementation of weighing gauges for the measurement of precipitation.

4.1.1.4.1 *Gauge selection*

The selection of a weighing gauge for the measurement of precipitation should consider the following:

- If expected site conditions fall within the operational limits specified by the gauge manufacturer;
- If site infrastructure allows for the installation and operation of the gauge, including the power required for heating (if applicable);
- The maintenance requirements of the gauge;
- The data outputs available.

Each of these points is discussed in greater detail below.

With respect to operational limits, the key factors to consider are temperature relative to maximum and minimum expected values and bucket capacity relative to expected seasonal precipitation

accumulation. The latter ties in with maintenance considerations, as discussed later. With respect to infrastructure, sufficient power must be available for gauge operation and heating (if applicable). Efforts should be made to minimize the length of power cords, which can lead to voltage drops that may compromise gauge operation and/or heating. Longer cords can also lead to the induction of noise, which can compromise gauge data.

Maintenance is an important consideration for gauge selection, particularly for remote regions that present challenges to site visits by personnel (time and cost requirements) and/or for organizations with limited resources available for site visits by personnel. Gauges with larger bucket capacity allow for longer periods before emptying is required; however, the larger range of accumulation values may reduce the sensitivity of the gauge to small changes in accumulation, mitigating their utility in light-precipitation conditions. The amount of antifreeze required to prevent freezing over the winter (factoring in dilution) and the amount of oil to mitigate evaporation need to be considered as the addition of these fluids reduces the remaining capacity of the gauge. The complexity of gauges is another consideration, as gauges with more moving parts (e.g. mechanical pumps) provide more potential failure modes for issues requiring maintenance.

Users should select weighing gauges with data outputs suitable to their specific applications, interests, and/or comfort level with data processing. Some gauges provide raw or close-to-raw precipitation amounts derived from the bucket weight, which require further assessment and/or processing by users to account for the influences of evaporation, temperature, and wind. Other gauges provide processed outputs that account for these factors, which are more user friendly in terms of the processing ability and/or background knowledge required to interpret results. In many cases, however, gauges reporting processed outputs are effectively "black boxes" that make use of proprietary algorithms. The level of detail provided to users regarding the processing approach in these cases likely depends upon specific agreements between the user and manufacturer (e.g. non-disclosure agreements). Hence, gauges with processed outputs provide users with ready-to-use options, but may limit the ability of users to understand and/or control how the data are processed for their specific application(s). Further, depending on the specific details of the processing approach and the conditions under which measurements are taken, there is potential for the false reporting of precipitation, which can artificially increase precipitation totals. It should be noted that some gauges offer both raw or close-to-raw data outputs and processed outputs that offer users the flexibility to use either or both types of outputs, and the capability to do their own output intercomparison.

4.1.1.4.2 *Gauge configuration*

The configuration of a weighing gauge at a given site should be selected to best fit the intended application and expected conditions, while also giving consideration to the availability of site infrastructure, site resources, and the allocated budget. The results presented in Section 4.1.1.3 illustrate that the catch efficiency for shielded gauges exceeds that for unshielded gauges at a given site. In terms of improving catch efficiency, the benefits of double-shields relative to single-shields in terms of improving catch efficiency are not well demonstrated by the present results, but have been reported elsewhere (Watson et al., 2008; Smith, 2009; Rasmussen et al., 2012; Kochendorfer et al., 2018). The summary results presented in Table 4.5 provide an overview of relative performance in different shield configurations at sites in different climate regimes.

In general, double-shields are recommended over single-shields, and single-shields over unshielded configurations; however, double-shields occupy the largest space and are the most expensive to

install and maintain, and even single shields present additional expenses relative to unshielded configurations. One aspect of shielding that can impact results is the potential for wind-induced vibration of the shield increasing the magnitude of noise in the gauge data. This is a concern for shield configurations mounted to the gauge post; if the shield is mounted separately from the gauge post, vibration of the shield assembly is not expected to impact gauge reports. Another point to consider is that a double-shield may offer limited benefit relative to a single-shield at sites characterized by low wind speeds, such as Sodankylä, for which the mean wind speeds are typically under 4 m/s.

Heating of gauges reduces the potential for accumulation on or within the gauge orifice and snow capping, which can partially or fully obstruct the orifice and impact the collection and reporting of precipitation accumulation (Sections 4.2.1 and 4.2.2). At the same time, however, heating can lead to evaporation of precipitation and chimney effects, in which air warmed by the heaters distorts the flow field surrounding the orifice, both of which also impact the collection and reporting of precipitation. Heating power, the physical configuration of heaters, and associated algorithms (temperature thresholds for heaters to be turned on, set point temperatures to maintain, timed pulses of heat to remove frost/ice) vary by gauge type and/or model. These aspects should all factor into the selection of weighing gauges as discussed, giving consideration to power limitations and the expected precipitation conditions (e.g. high heating power is not likely required for a site with characteristically light-precipitation events). The configurability of the heating algorithm also warrants consideration, as the algorithm thresholds and set points can be modified for some gauges, but not others, allowing for different degrees of user control. It is recommended that prospective users consult with the gauge manufacturers to find the most appropriate heating configuration for their specific conditions. For gauge manufacturers, it is recommended that heating be configurable to optimize power consumption and/or to tailor heating for different conditions or applications, and that the related algorithm(s) are documented for the information of instrument and data users.

4.1.1.4.3 *Adjustment functions and required ancillary measurements*

The results presented in Section 4.1.1.3 illustrate that the weighing gauges tested generally under-report precipitation relative to the reference configuration (the representation of the "true" precipitation amount for this assessment), and that the extent of this under-reporting, or undercatch, increases with increasing mean wind speed in solid-precipitation conditions. With knowledge of the accumulation reports from each test gauge relative to the reference configuration at the site, it is possible to derive a relationship between catch efficiency and mean wind speed. The resulting transfer functions can be used to estimate the "true" precipitation amounts from test-gauge reports. These functions can be derived as a function of wind speed only (for precipitation events of a given type) or as a function of both wind speed and temperature (for events of all precipitation types).

Kochendorfer et al. (2017a, 2017b) derived transfer functions using weighing-gauge data from WMO-SPICE test sites. The functions were found to reduce the bias of test-gauge reports relative to reference reports, with "universal" functions derived using data from gauges at multiple sites in the same shield configurations showing similar performance as functions derived using datasets from specific gauge types. The application of transfer functions to weighing-gauge reports is recommended to mitigate the bias in test-gauge reports resulting from wind-induced undercatch. The functions derived using SPICE data may be applied to data from other gauges and test sites;

alternatively, new functions may be derived following a similar approach as that used by Kochendorfer et al. (2017a, 2017b).

The derivation of transfer functions requires the presence of a suitable site reference, such as the SPICE R2 reference configuration, which comprises both a weighing gauge in DFIR-shield and independent precipitation detector. Ancillary measurements of mean wind speed, ideally at gauge height, and temperature are required for the derivation and application of transfer functions. The duration of assessment intervals for transfer function application and development is a point of interest for potential users in different fields; for example, hourly intervals are typically used operationally, while longer intervals (e.g. 12, 24 hours) are often used in climate applications. The SPICE transfer functions were developed over 30-minute assessment intervals, but have been demonstrated to be applicable over longer intervals (Kochendorfer et al., 2017a).

4.1.2 Heated tipping bucket gauges

Authors: Michael Earle, Kai Wong, Rodica Nitu, Samuel Buisán, Audrey Reverdin

4.1.2.1 Introduction

Tipping bucket gauges are used broadly in operational networks to measure precipitation (Nitu and Wong, 2010) either as an alternative or a complement to weighing precipitation gauges (Pikounis et al., 2002). The performance of tipping bucket gauges with respect to the measurement of rainfall has been assessed and documented in previous WMO intercomparisons (Lanza et al., 2005; Vuerich et al., 2009). More recently, the addition of heaters to commercial tipping bucket gauges has extended their application to colder environments and the measurement of solid and mixed-phase precipitation.

WMO-SPICE assessed the performance of heated tipping bucket gauges over the full range of winter conditions and precipitation types experienced at sites in different climate regimes (Figure 4.28). Gauges were tested at the following sites (climates): Centre for Atmospheric Research Experiments, CARE (humid continental); Formigal (alpine with maritime influence); Marshall (dry continental); Sodankylä (northern boreal, low wind); and Weissfluhjoch (alpine with complex topography). Seven models of heated tipping bucket gauges were tested (Table 4.6). These varied in terms of the bucket/reporting resolution and heating approach (power, physical configuration, algorithm). One of the test configurations employed a wind shield to mitigate the influence of wind-induced undercatch (Figure 4.29), while all other test configurations were unshielded. Wind effects are a focal point of the assessment, as they have been demonstrated to be significant for solid precipitation, particularly for unshielded gauges (e.g. Goodison et al., 1998; Rasmussen et al., 2012).

Figure 4.28. Sites hosting heated tipping bucket gauges under test in WMO-SPICE.

Figure 4.29. EML UPG1000 heated tipping bucket gauge under test at Sodankylä with wind shield. Note: The horizontal components of the shield (top) are no longer recommended by the manufacturer due to concerns regarding the potential to accumulate snow that can then be blown by wind into the gauge orifice.

A detailed assessment of the performance of each gauge at each test site is provided in the corresponding instrument performance reports (Annex 6). The following section consolidates the key findings from these reports, both by sensor and by site, and uses these findings to make recommendations regarding the field implementation of heated tipping bucket gauges. To establish context for the interpretation of results, related discussion, and recommendations, an overview of

tipping bucket gauge operation is provided. A discussion of how the operation and heating of tipping bucket gauges impacts the assessment of performance relative to the WMO-SPICE R2 field reference configuration, which employs a weighing-precipitation gauge, is also provided. An alternative assessment methodology addressing caveats related to the comparison with weighing gauges is proposed for consideration in future investigations.

Table 4.6. Heated tipping bucket gauges under test in WMO-SPICE.

Sensor	Site(s)	Tip resolution [mm]	Reporting resolution [mm]	Maximum heating power [W]
CAE PMB25R	CARE, Marshall	0.2	0.1	300
EML UPG1000, shielded	Marshall, Sodankylä	0.1	0.1	330
HSA TBH	CARE, Marshall (x2)	0.2	0.2	70
Meteoservis MR3H-FC	CARE, Marshall, Sodankylä	0.1	0.1	555
Meteoservis MR3H-FC, ZAMG version	CARE, Weissfluhjoch	0.1	0.1	555
Thies Precipitation Transmitter, model 5.4032.35.228	Formigal	0.2	0.2	49
Thies Precipitation Transmitter, model 5.4032.45.008	Marshall	0.1	0.1	113

4.1.2.1.1 *Operating principle of tipping bucket gauges*

A schematic configuration of a tipping bucket gauge is provided in Figure 4.30. The principle of operation is detailed in instrument manuals and other publications (e.g. Pikounis et al., 2002), and outlined briefly here. A funnel transfers incident precipitation from the circular orifice to the tipping mechanism, which consists of two bucket compartments with identical capacity (corresponding to the tip resolution in Table 4.6). One of the bucket compartments is positioned below the funnel opening at a given time. Precipitation flows into the compartment until its capacity is reached, at which point the bucket tips and moves the other (empty) compartment below the opening. This triggers a magnetic reed switch that sends a signal to the data logger; meanwhile, the contents of the tipped bucket flow out of the gauge bottom through designated outlet ports. The logger counts the number of tips per sample interval, which can be used to determine the corresponding precipitation intensity and, in turn, the accumulation. For example, if one tip is recorded within a sample interval of one minute for a gauge with a 0.1 mm tip resolution, the resulting intensity is 0.1 mm/min, or 6 mm/hour.

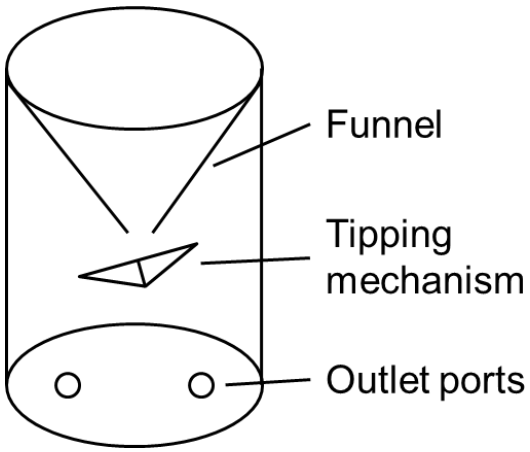


Figure 4.30. Schematic diagram of tipping bucket gauge.

The measurement of solid precipitation requires that the funnel be heated to melt the precipitation, allowing it to flow into the tipping mechanism. The mechanism is also heated to prevent re-freezing of precipitation. The outlet ports are also typically heated to prevent freezing of precipitation leaving the gauge, which can cause the gauge outlets to become blocked. Most tipping bucket gauges tested employ integrated temperature sensors and control the application of heat using set temperature thresholds. The HSA TBH gauge is designed for lower-power operation and applies heat only when the presence of snow is detected by a sensor in the funnel.

4.1.2.1.2 *Impacts of tipping bucket gauge heating*

Heating of tipping bucket gauges is required for the measurement of solid precipitation; however, heating introduces errors related to the amount of precipitation reported, as well as to the timeliness of gauge reports relative to weighing gauges. Heating of the funnel and tipping assembly can lead to evaporative and wetting losses, which reduce the amount of precipitation reported by the tipping bucket gauge (Savina et al., 2012). Losses due to the heating of tipping bucket gauges typically exceed those due to heating of weighing gauges because of the larger heated area for tipping bucket gauges (typically, only the gauge orifice or inlet tube is heated for weighing gauges) and the indirect nature of the tipping bucket measurement relative to weighing gauges, which measure precipitation directly. In fact, Zweifel and Sevruk (2002) suggested that for tipping bucket gauges, losses due to heating may be comparable to, or larger than, losses due to wind effects.

Heating and wetting losses can delay the response of a tipping bucket gauge to solid precipitation, particularly when the intensity is low (Savina et al., 2012). The time required to melt precipitation in the funnel can further delay the tipping bucket response and can be exacerbated by the specific heating mechanism. For example, the HSA TBH triggers heating only when snow in the funnel reaches the height of a snow sensor, compounding the delay required for melting once heat is applied. Melting delays give rise to another potential loss mechanism: wind blowing snow out of the funnel. Low and/or insufficient heating power may further extend melt times, providing more opportunity for losses due to wind or to the funnel reaching capacity and being unable to collect

more precipitation. At the same time, however, lower heating power could be advantageous in terms of reducing evaporative losses from the funnel relative to heating with higher power.

Savina et al. (2012) assessed losses and delays for a heated tipping bucket gauge relative to a weighing gauge for snowfall events in Switzerland. The mean delay time for the tipping bucket response relative to the weighing gauge was approximately 30 minutes. Comparisons of accumulated precipitation reports from these different gauge types (tipping bucket gauges and weighing gauges) must account for such heating losses and delays for tipping bucket gauges, as well as for delays inherent to the coarser accumulation resolution of tipping bucket gauges relative to weighing gauges (i.e. more precipitation is typically required for a tipping bucket gauge to respond). An important point raised by this study was that comparisons of snowfall observations from tipping bucket and weighing gauges should account for not only the relative precipitation amounts, but also for the relative timing of precipitation events. The assessment approach for heated tipping bucket gauges in WMO-SPICE aims to address both of these considerations.

4.1.2.2 Assessment approach

The assessment approach for WMO-SPICE (Section 3.6) is predicated on 30-minute precipitation events as identified by the R2 reference configuration, comprising a DFAR and precipitation detector (Section 0). A precipitation event is defined as a 30-minute period during which the weighing gauge in the DFAR accumulates ≥ 0.25 mm of precipitation and the precipitation detector reports the occurrence of precipitation for at least 18 minutes (Section 3.4). For consistency with the rest of the WMO-SPICE analysis, the assessment of heated tipping bucket gauges was conducted using 30-minute events. Recognizing that response delays for heated tipping bucket gauges will impact the assessment relative to the reference configuration and considering that hourly reports are commonly used operationally, the analysis was extended to 60-minute precipitation events with the same selection criteria. Results are provided for both 30- and 60-minute assessment intervals in the instrument performance reports (Annex 6); however, the consolidated results for heated tipping bucket gauges in this section are provided for 60-minute assessment intervals only.

Consolidated results are provided for each heated tipping bucket gauge type and for each test site in Section 4.1.2.3. Results include skill scores representing the ability of heated tipping bucket gauges to detect precipitation relative to the reference configuration, as well as root mean square errors and catch efficiencies, which quantify the performance relative to the reference configuration in terms of reporting accumulated precipitation. Accumulation reports and catch efficiencies are also considered on a per-event basis (relative to the reference configuration) as a function of wind speed and temperature/precipitation type.

The characterization of response delays was based on tipping bucket comparison events (TBCEs), consisting of one (or successive) 30-minute precipitation event(s) (as identified by the reference configuration), followed by 180 minutes with no precipitation (Section 3.6.1.4.4). The non-precipitating period is intended to provide sufficient time for the tipping bucket gauge to respond following the onset of precipitation and is consistent with the approach of Savina et al. (2012). Delay times are determined from the time elapsed between the start of each TBCE and the first subsequent response (tip) of the tipping bucket gauge. Probability density functions of delay times are consolidated for each tipping bucket gauge type and for each test site in Section 4.1.2.3.3. To assess the impact of snowfall intensity on delay times for each test gauge and at each site, delay times are assessed as a function of the mean snowfall intensity reported by the reference configuration during

the delay period. Snow events are identified as those for which the maximum reported temperature does not exceed -2 °C during the delay period. It should be noted that this temperature threshold does not necessarily capture all solid precipitation events at all sites. For example, work conducted by Buisán et al. (2017) using data from the Formigal test site in Spain showed that a present weather sensor classified precipitation events as snow when the mean temperatures were warmer than -2 °C.

4.1.2.3 Results

4.1.2.3.1 *Characterization of precipitation events*

To provide context for the interpretation of results from test gauges at different sites, mean precipitation-event characteristics are provided for 60-minute events in all conditions and precipitation types at each site in Table 4.7. The precipitation events at each site were separated by type/phase using the ambient temperature thresholds presented in Section 3.6.1.2; the results are presented in Table 4.8. Note that differences in the environmental conditions and relative distributions of precipitation events by phase at each site do not explain all observed differences in results. Differences in gauge siting and configuration can impact gauge performance significantly, and it is difficult to separate such contributions to errors/uncertainty from those of the environment and the operation of, and internal processing specific to, different gauge types. Further, while the mean parameter values provide perspective on how precipitation events varied from site to site, there is significant variability in the conditions experienced at each site, as reflected by the corresponding standard deviation values in Table 4.7.

The results in Table 4.7 indicate that the mean wind speeds during 60-minute precipitation events, measured at gauge height or at the standard operational height in a national network, were comparable across all sites except Sodankylä, which is a sheltered site characterized by low winds. The precipitation events at Formigal (alpine climate with maritime influence) were characterized by the largest mean accumulations and warmest mean temperatures among the test sites, while the events at Sodankylä (northern boreal climate) were characterized by the smallest mean accumulations. For the sites in continental climates, the events at Marshall were characterized by larger mean accumulations and lower mean temperatures relative to those at CARE. Due to configuration issues at CARE, data for two test gauges – the Meteoservis MR3H-FC and ZAMG MR3H-FC – were available only for the second measurement season. Table 4.7 indicates that the mean event parameters for the second season at CARE were comparable with those for events compiled over both measurement seasons, supporting the comparability of results for all test gauges at CARE.

Table 4.7. Precipitation-event characteristics at all heated tipping bucket test sites during WMO-SPICE. Mean event parameters are provided, with uncertainty represented in terms of the standard deviation.

Site	Climate zone	Mean event accumulation [mm]	Mean event temperature [°C]	Mean event wind speed [m/s]†
CARE	Continental	0.84 ± 0.72	-2.9 ± 7.0	3.1 ± 1.5
CARE (season 2)*		0.84 ± 0.71	-3.4 ± 7.9	3.0 ± 1.5
Marshall	Continental	1.07 ± 1.03	-3.6 ± 6.3	2.5 ± 1.4
Formigal	Alpine (maritime influence)	1.32 ± 1.25	-1.1 ± 3.5	2.9 ± 1.9
Weissfluhjoch	Alpine	1.05 ± 0.93	-6.8 ± 4.6	2.8 ± 2.0
Sodankylä	Northern boreal (low wind)	0.57 ± 0.38	-2.3 ± 3.8	1.5 ± 0.6

* Separate results are provided for season 2 at CARE because data for Meteoservis MR3H-FC and ZAMG MR3H-FC gauges were only available for this season.

† Mean event wind speeds were determined at gauge height for all sites except Formigal, where wind-speed measurements were taken at the standard operational height of 10 m.

The relative proportions of precipitation events by phase are presented for each site in Table 4.8. Approximately 50% of the events at continental sites were classified as solid precipitation. The proportion of mixed-phase events at Marshall was greater than that at CARE, for which the percentages of liquid and mixed-phase events were similar. The distributions for the alpine sites differed greatly; over 80% of the events at Weissfluhjoch were classified as solid (by number) with very few liquid events (3%), while at Formigal, the percentage of mixed-phase events (46%) exceeded that of solid events (38%). Mixed-phase events were predominant at Sodankylä, comprising over half of the total events (55%), with the remainder of events classified primarily as solid precipitation (40%).

Table 4.8. Percentage of precipitation events of each phase/type by number and by total accumulation at all heated tipping bucket test sites during WMO-SPICE.

Site	Climate zone	Percentage of events by number (by total accumulation)		
		Liquid	Mixed	Solid
CARE	Continental	23% (26%)	22% (23%)	55% (51%)
CARE (season 2)*		23% (24%)	24% (22%)	53% (54%)
Marshall	Continental	14% (17%)	34% (49%)	52% (34%)
Formigal	Alpine (maritime influence)	16% (15%)	46% (46%)	38% (38%)
Weissfluhjoch	Alpine	3% (3%)	16% (20%)	81% (77%)
Sodankylä	Northern boreal (low wind)	5% (4%)	55% (55%)	40% (40%)

* Separate results are provided for season 2 at CARE because data for Meteoservis MR3H-FC and ZAMG MR3H-FC gauges were only available for this season.

The mean characteristics for events classified as snow at each site are presented in Table 4.9. These results indicate that the snow events were comparable at the sites in continental climates, with similar mean event accumulations, temperatures, and gauge-height wind speeds at CARE and Marshall. The results for season 2 at CARE were similar to those for both measurement seasons, again supporting the comparability of results among CARE test gauges. Mean wind speeds during precipitation events were again comparable across all sites except Sodankylä, and snow events at Formigal had the largest mean accumulations and warmest mean temperatures. In general, the relative magnitudes of mean accumulation during 60-minute precipitation events by climate zone were as follows: alpine (maritime influence) > alpine (complex topography) > continental > northern boreal.

Table 4.9. Snow-event characteristics at all heated tipping bucket test sites during WMO-SPICE. Mean event parameters are provided, with uncertainty represented in terms of the standard deviation.

Site	Climate zone	Mean event accumulation [mm]	Mean event temperature [°C]	Mean event wind speed [m/s] †
CARE	Continental	0.77 ± 0.68	-7.8 ± 4.8	3.1 ± 1.3
CARE (season 2)*		0.85 ± 0.72	-9.5 ± 5.3	3.3 ± 1.2
Marshall	Continental	0.70 ± 0.51	-8.2 ± 4.0	2.4 ± 1.2
Formigal	Alpine (maritime influence)	1.35 ± 1.10	-4.7 ± 1.5	3.3 ± 2.3
Weissfluhjoch	Alpine	1.00 ± 0.87	-8.2 ± 3.6	2.9 ± 2.0
Sodankylä	Northern boreal (low wind)	0.58 ± 0.37	-5.7 ± 3.4	1.5 ± 0.5

* Separate results are provided for season 2 at CARE because data for Meteoservis MR3H-FC and ZAMG MR3H-FC gauges were only available for this season.

† Mean event wind speeds were determined at gauge height for all sites except Formigal, where wind-speed measurements were taken at the standard operational height of 10 m.

4.1.2.3.2 *Detection of precipitation relative to reference*

The detection of precipitation by heated tipping bucket test gauges relative to corresponding site reference configurations was assessed using the following skill scores, which are outlined in Section 3.6.1.3.1: probability of detection (POD), false alarm rate (FAR), bias (B), and Heidke Skill Score (HSS). In this assessment, the test gauges were considered to detect precipitation when at least one tip was recorded during a 60-minute assessment interval, while the reference was considered to detect precipitation when the reference weighing gauge reported ≥ 0.25 mm accumulated precipitation and the reference precipitation detector reported ≥ 18 minutes of precipitation occurrence during the same assessment interval. There were four potential detection scenarios: the reference and tipping bucket gauge both detected precipitation (YY cases); the reference detected precipitation, but the tipping bucket did not (YN cases); the reference did not detect precipitation, but the tipping bucket did (NY cases); and neither the reference, nor the tipping bucket, detected precipitation (NN cases). The number of events in each detection scenario, for each test gauge, over the duration of experiments was used to calculate the above skill scores, as detailed in Section 3.6.1.3.1.

The reference detection criteria have been selected to identify precipitation events with a high degree of confidence; however, light and/or sporadic precipitation events are not well represented. This has implications for the skill score assessment, as a proportion of the NY cases may correspond to intervals during which the tipping bucket responds to incident precipitation that does not meet the reference detection criteria. Indeed, the characterization of NY cases in the Instrument Performance Reports (Annex 6) indicates the occurrence of tipping bucket responses to light precipitation events, with reference accumulation below 0.25 mm. A detailed investigation of tipping

bucket responses to light precipitation is not considered in the present assessment, but relevant results and discussion are provided to assist with the interpretation of skill score results, as required.

Skill scores for all test gauges, at all sites, are compiled in Figure 4.31. These scores represent the ability of the tipping bucket gauges to detect precipitation relative to the reference configuration over the full range of environmental conditions and precipitation types experienced at each site, within 60-minute periods and using the specified detection thresholds. The heating power and configuration varies by gauge type (Table 4.6) with the HSA TBH being the notable exception in that the heating is triggered by a snow sensor in the funnel rather than a set temperature threshold. This difference in heating approach results in marked differences in skill score results for the TBH test gauges relative to the other gauges.

The POD represents the percentage of the total number of precipitation events identified by the reference that are also identified by the test gauge, with an ideal value of 100%. For all test gauges with heating application determined by temperature, the POD exceeded 75% in all climate zones, ranging from approximately 75 to 90% in continental climates, 80 to 85% in alpine climates, and 90 to 95% in northern boreal climates. For the TBH gauges, the POD was between about 35 and 50% on account of the difference in heating configuration.

In general, the FAR, which represents the total number of precipitation events detected by the test gauge that are not detected by the reference (ideal value of 0%), was below 30% for test gauges in all climate zones; however, there were notable exceptions. The FAR for the EML UPG1000 at Sodankylä was about 45%, which is attributed to "false tips" due to snow accumulated on the horizontal components of the wind shield (see Figure 4.29) blowing into the gauge orifice. The FAR for the TBH gauges was variable, ranging from 28% to 60%, due to the fact that the application of heat is effectively stochastic; that is, the timing of heating is dependent upon the specific conditions leading to the triggering of the snow sensor in the funnel, and does not vary in any recognizable pattern from event to event.

The Meteoservis MR3H-FC and ZAMG MR3H-FC gauges at CARE showed FAR values of 50 to 60%, notably higher than the values for the CAE PMB25R at CARE (~ 20%), which is the only other gauge at CARE with heating determined by temperature. One potential explanation is that the higher heating power of the Meteoservis and ZAMG gauges (555 W; see Table 4.6) enhances response delays relative to gauges with lower heating power, like the CAE gauge (300 W). Heating in the funnel of the Meteoservis and ZAMG gauges employs pulses to avoid over/under-heating, but constant heating applied to the region of the tipping mechanism may lead to evaporation that can delay responses into a subsequent 60-minute assessment interval (during which the reference does not detect precipitation). The numbers of NY cases for these gauges support this hypothesis; during the second measurement season at CARE, the CAE PMB25R had 90 NY cases, while the Meteoservis and ZAMG gauges had 176 and 208 NY cases, respectively. However, a comparable difference in FAR is not observed between the CAE PMB25R and Meteoservis MR3H-FC at Marshall, suggesting that the difference observed at CARE may be related to the specific gauge configurations and conditions experienced at that site.

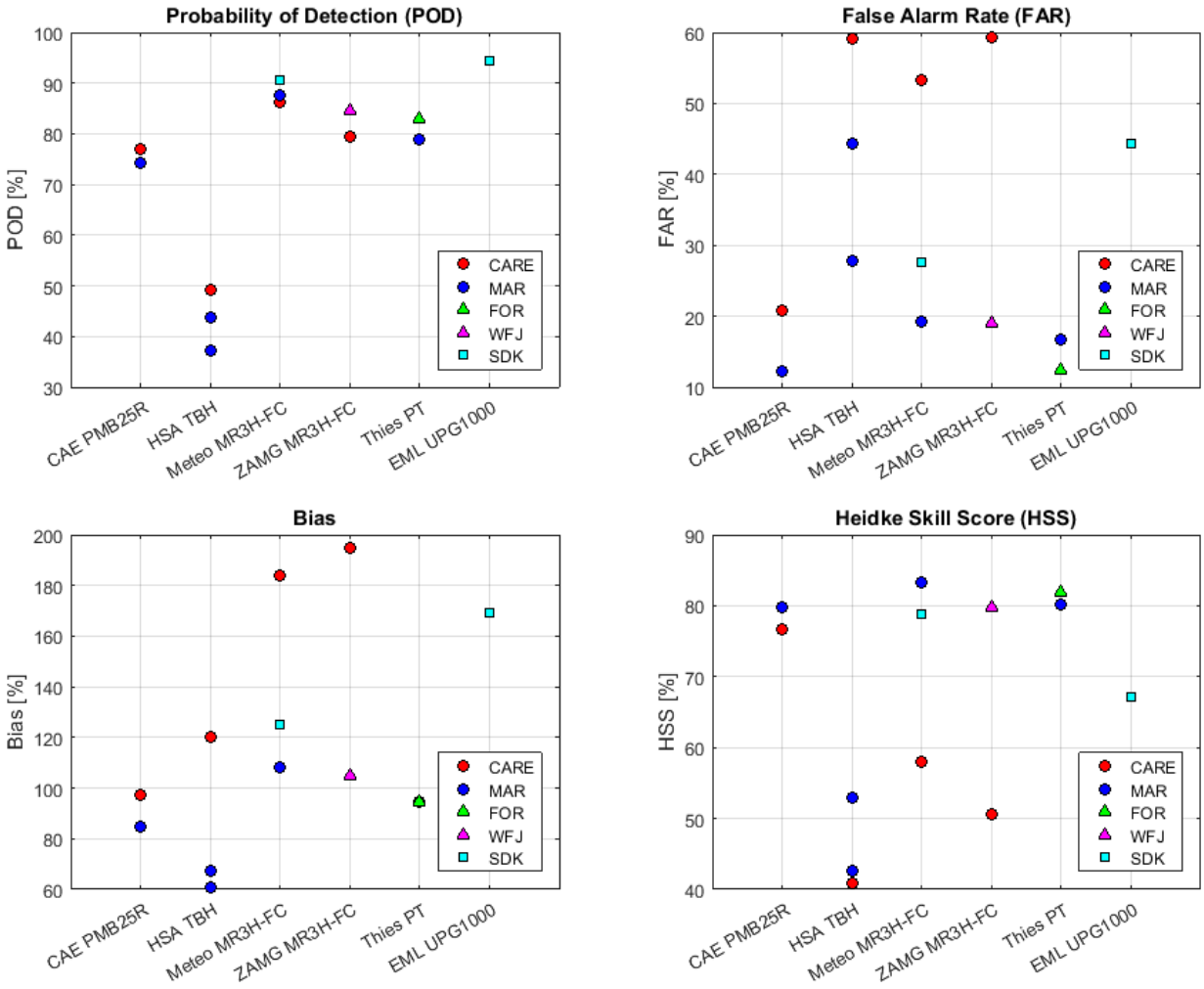


Figure 4.31. Skill scores representing the performance of each tipping bucket gauge under test relative to the reference with respect to the detection of precipitation. Scores are calculated using 60-min events in all precipitation types. Gauges in continental, alpine, and sub-arctic/northern boreal climate zones are represented by circles, triangles, and squares, respectively.

The same gauges that showed the highest FAR values – specifically, the Meteoservis MR3H-FC and ZAMG MR3H-FC gauges at CARE, and the EML UPG1000 gauge at Sodankylä – also showed the highest bias values, all above 160%. This reflects an overall tendency for these gauges to detect the occurrence of precipitation during 60-minute periods in which the reference does not detect precipitation, according to the specified criteria. For the TBH gauges with variable FAR, the bias was also variable, ranging from 60% to 120% for the test gauges at CARE and Marshall. For all other test gauges, the bias is within about ± 20% of 100%, indicating no significant positive or negative bias. In other words, there is no apparent tendency for these gauges to detect precipitation during more or fewer 60-minute intervals than the reference configuration.

The Heidke Skill Score provides a representation of the overall skill of the test gauge in terms of detecting precipitation relative to the reference, according to the specified detection criteria. The HSS considers the number of assessment intervals during which the test gauge and reference both do

or do not detect precipitation (i.e. the number of events "correctly" identified or not identified by the test gauge), while also taking into account the number of correct responses that would be expected due to chance. A sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%. The HSS was typically between 75% and 85% for test gauges in all climate zones, with lower values for the exception cases noted above. The HSS values (and potential explanations) for the exception cases are as follows: HSS within 40% to 55% for the HSA TBH gauges at CARE and Marshall (delays due to the specific heating configuration); HSS between 50% and 60% for the Meteoservis MR3H-FC and ZAMG MR3H-FC at CARE (potential influence of tipping mechanism heating on response times for the specific site configurations and conditions experienced); and HSS of approximately 45% for the EML UPG1000 at Sodankylä (false tips from snow accumulated on the shield blowing into the gauge orifice).

It is important to emphasize that the skill score results reflect only the ability of the tipping bucket test gauges to detect precipitation relative to the reference within 60-minute intervals. This is an important consideration for operational networks, which report precipitation over defined time intervals (typically hourly), but does not reflect the overall performance of the gauge in terms of reporting accumulated precipitation over the same or longer time intervals. An assessment of test gauge performance in terms of accumulation reports over both 60-minute assessment intervals and the duration of experiments is presented in Section 4.1.2.3.3.

4.1.2.3.3 *Reporting accumulated precipitation relative to reference*

The catch efficiency, or catch ratio, is the ratio of accumulated precipitation reported by a test gauge relative to that reported by the reference configuration over a specified time interval. The catch efficiency of solid precipitation is impacted significantly by wind speed, as the lower density and slower fall velocity of solid precipitation particles makes them more susceptible to deflection by wind away from the gauge orifice. The catch efficiency of solid precipitation is assessed as a function of the mean wind speed for 60-minute assessment intervals in Section 4.1.2.3.3.1.

The catch efficiency is a useful indicator of tipping bucket gauge performance relative to the reference, but does not provide information about the magnitude of accumulated precipitation reported by each gauge. The root mean square error, however, considers the absolute difference in reported accumulation between the test gauge and the reference for each 60-minute interval. An assessment of RMSE results is provided for all test gauges, at all sites, in Section 4.1.2.3.3.2.

The approaches outlined above are both based on 60-minute assessment intervals during which the reference and test gauge both detect and report precipitation (YY cases). Given the influence of tipping bucket gauge heating on response times, there are many 60-minute intervals during which the reference reports precipitation but the tipping bucket does not respond in time (YN cases). Similarly, there are many intervals during which the reference does not report any precipitation, but the tipping bucket responds to precipitation reported by the reference during an earlier assessment interval (NY cases, or "late tips"). Further, there may be YY cases during which the reference and tipping bucket are responding to precipitation that fell at different times; for example, if the reference responds to precipitation currently falling, while the tipping bucket responds to both this precipitation and to precipitation accumulated in the funnel or bucket from an earlier event. To account for these potential scenarios, which affect the catch efficiency and the RMSE values for individual assessment intervals, overall catch efficiencies were calculated using the total sums of reference and tipping bucket accumulation over all YY, YN, and NY cases over the duration of formal

tests. The overall catch efficiency values include precipitation of all types, and are presented for each test gauge, at each site, in Section 4.1.2.3.3.3.

4.1.2.3.3.1 Wind effects on catch efficiency

The catch efficiency of each test gauge (at all applicable sites) and at each site (with multiple test gauges) is assessed as a function of the mean wind speed for 60-minute assessment intervals in Sections 4.1.2.3.3.1.1 and 4.1.2.3.3.1.2, respectively. This assessment is limited to snow events during which the maximum temperature did not exceed -2 °C over a given 60-minute assessment interval. A similar assessment is presented for all precipitation types (rain, snow, mixed) in the Instrument Performance Reports (Annex 6).

4.1.2.3.3.1.1 Results by gauge type

Box and whisker plots of catch efficiency as a function of mean wind speed, either at gauge height or at the standard operational wind measurement height, are plotted for each test gauge in Figure 4.32 to Figure 4.37. The boxes in each plot represent the range of values between the 25th percentile (lower quartile; bottom of box) and 75th percentile (upper quartile; top of box) referred to as the interquartile range. The median value is indicated by the horizontal line across the box. The whiskers below/above the box indicate the lowest/highest values. Outlying points are indicated by markers above or below the whiskers.

In each plot, results are presented in 1 m/s bins, with gauges at different sites represented by different colors. In general, it is difficult to make attributions for observed differences in the catch efficiency vs. wind-speed relationship for the same gauge at multiple sites. These differences are likely the combined result of differences in environmental conditions (mean event accumulation, temperature, predominant ice crystal type for solid precipitation) and differences in gauge configuration and siting. Accordingly, the identification and diagnosis of differences in results between or among sites is not a focal point of this work; rather, the focus is on identifying general trends for gauges of a given configuration type.

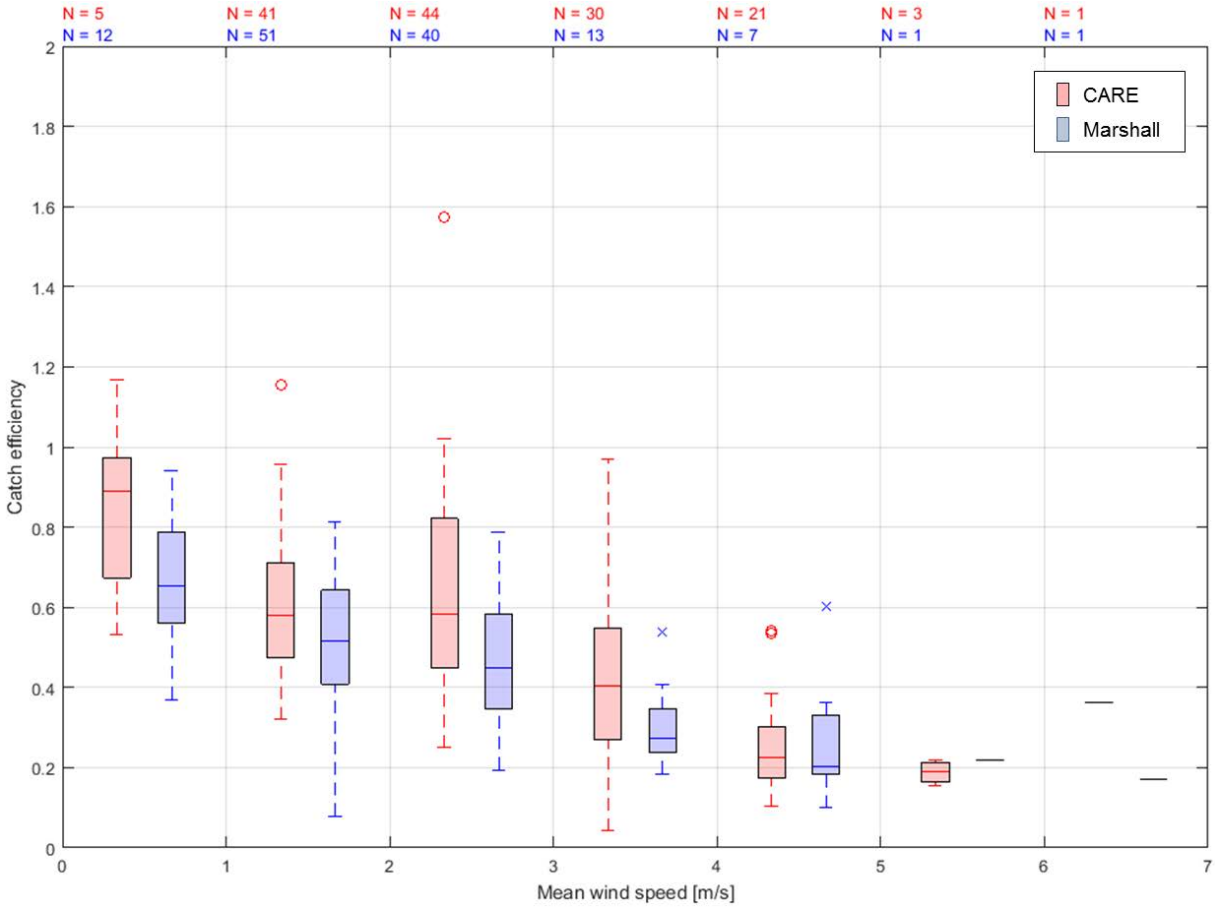


Figure 4.32. Box and whisker plot of catch efficiency as a function of mean gauge-height wind speed for CAE PMB25R gauges relative to site reference configurations for 60-minute snow events (max T ≤ -2 °C). The number of events in each wind-speed bin, for each test gauge, is indicated above the plot.

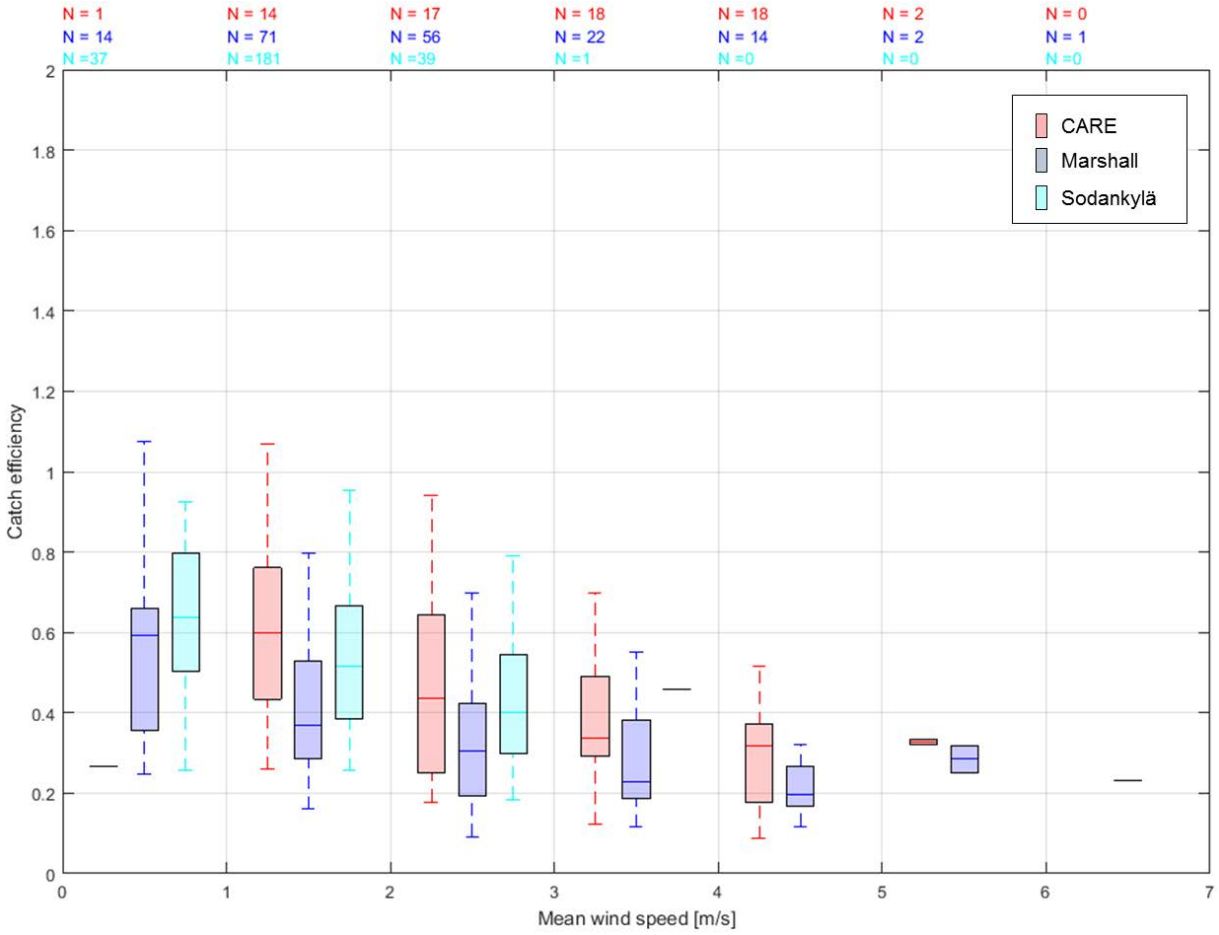


Figure 4.33. Box and whisker plot of catch efficiency as a function of mean gauge-height wind speed for Meteoservis MR3H-FC gauges relative to site reference configurations for 60-minute snow events (max T ≤ -2 °C). The number of events in each wind-speed bin, for each test gauge, is indicated above the plot.

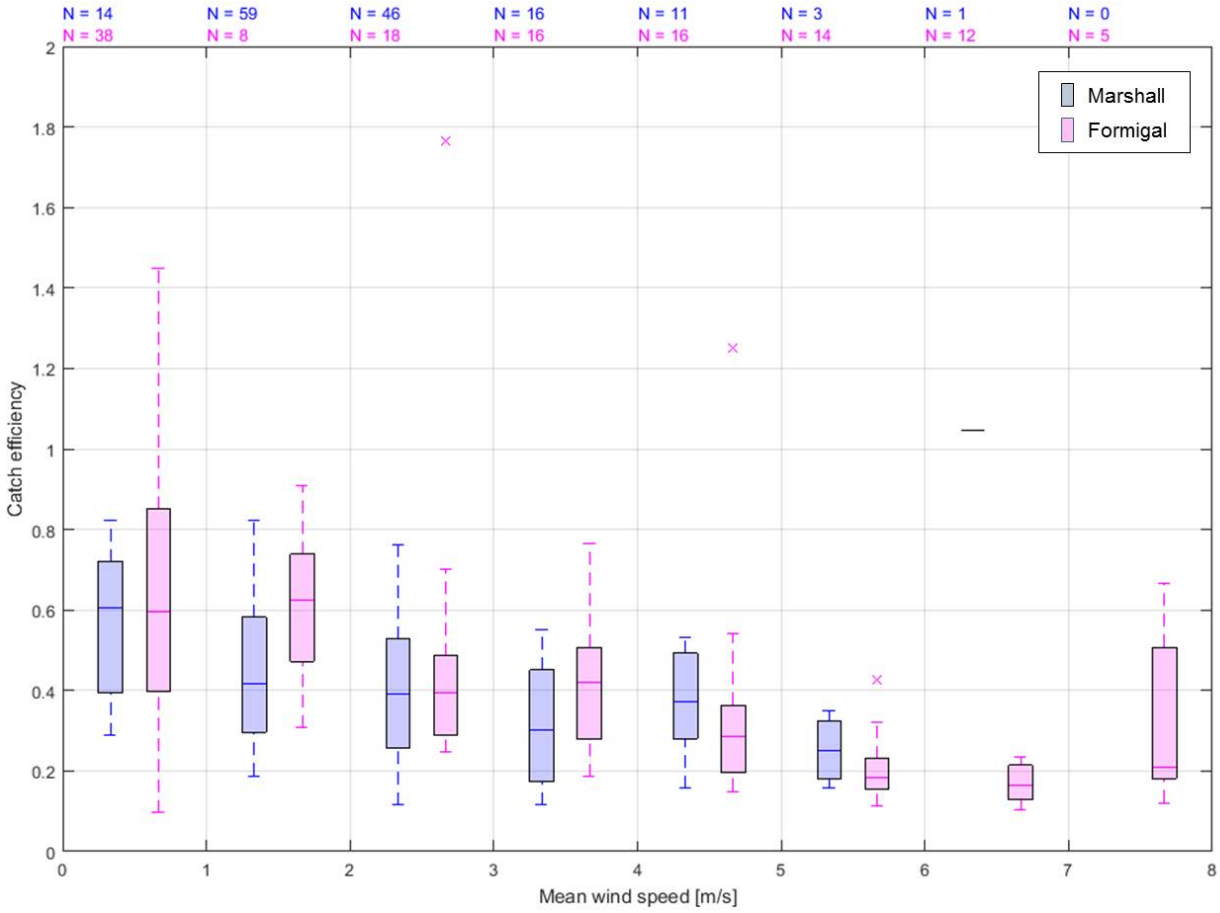


Figure 4.34. Box and whisker plot of catch efficiency as a function of mean wind speed for Thies Precipitation Transmitter gauges relative to site reference configurations for 60-minute snow events (max T ≤ -2 °C). The number of events in each wind-speed bin, for each test gauge, is indicated above the plot. Different gauge models were tested at each site (Table 1). Note that wind speed is measured close to gauge height (2 m) at Marshall, and at the operational wind measurement height (10 m) at Formigal.

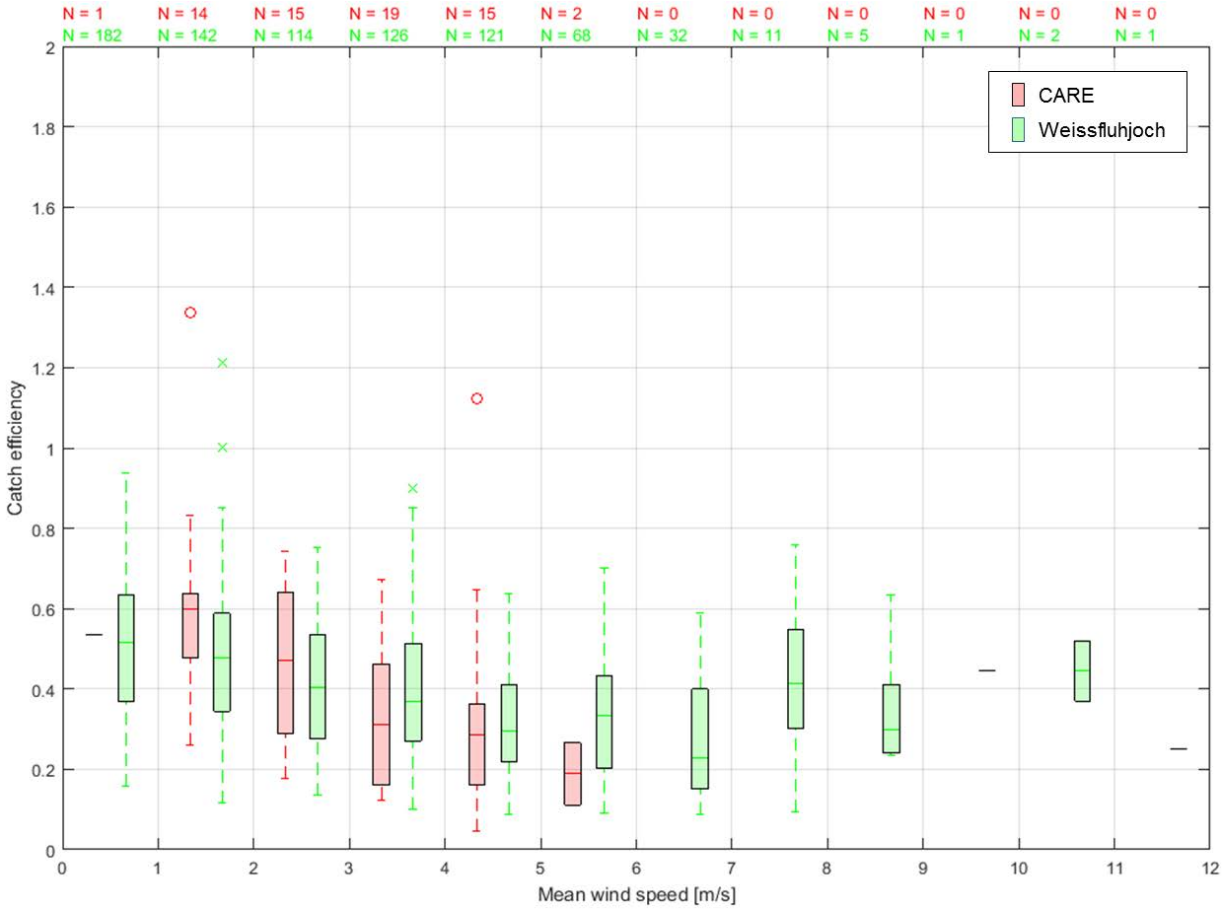


Figure 4.35. Box and whisker plot of catch efficiency as a function of mean gauge-height wind speed for ZAMG MR3H-FC gauges relative to site reference configurations for 60-minute snow events (max T ≤ -2 °C). The number of events in each wind-speed bin, for each test gauge, is indicated above the plot.

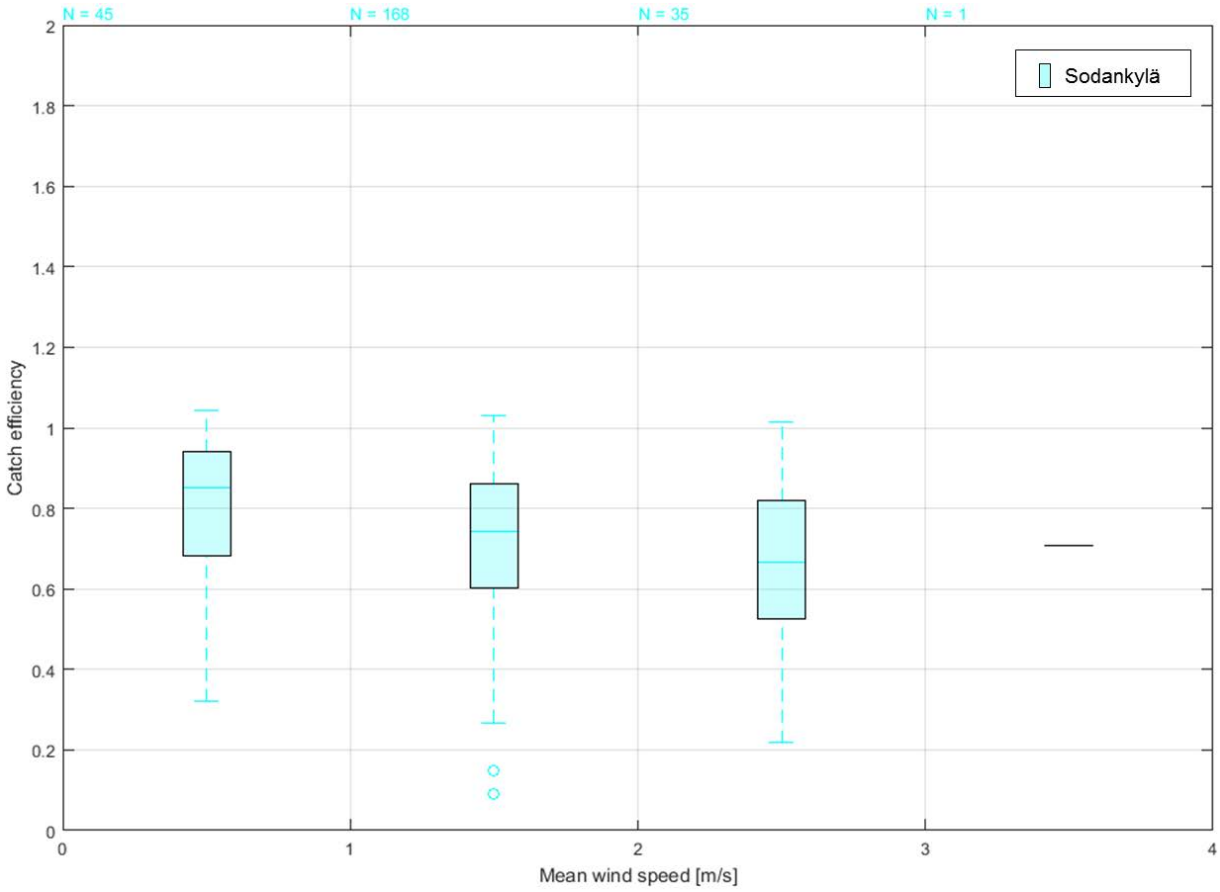


Figure 4.36. Box and whisker plot of catch efficiency as a function of mean gauge-height wind speed for EML UPG1000 gauges relative to site reference configurations for 60-minute snow events (max T ≤ -2 °C). The number of events in each wind-speed bin, for each test gauge, is indicated above the plot.

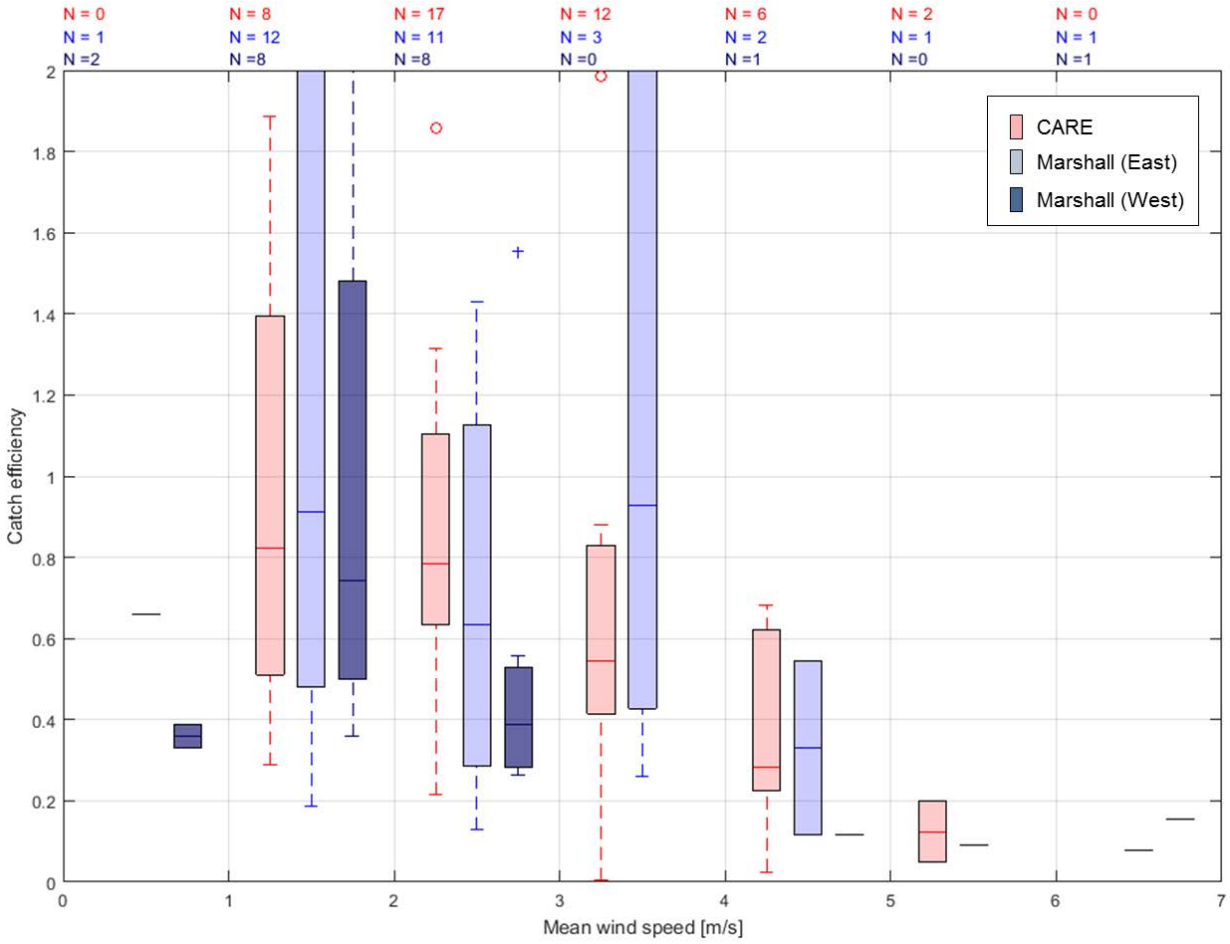


Figure 4.37. Box and whisker plot of catch efficiency as a function of mean gauge height wind speed for HSA TBH gauges relative to site reference configurations for 60-minute snow events (max T ≤ -2 °C). The number of events in each wind-speed bin, for each test gauge, is indicated above the plot.

For unshielded gauges with heating application determined by temperature – specifically, the CAE PMB25R (Figure 4.32), Meteoservis MR3H-FC (Figure 4.33), Thies Precipitation Transmitter (Figure 4.34; recall that different gauge models were tested at each site, as outlined in Table 4.6), and ZAMG MR3H-FC (Figure 4.35) – the median catch efficiency is between 0.5 to 0.9 for mean wind speeds below 1 m/s, between 0.4 and 0.6 for mean wind speeds between 1 m/s and 3 m/s, and 0.4 or lower for mean wind speeds greater than 3 m/s. For shielded gauges with heating application determined by temperature – the EML UPG1000 at Sodankylä (Figure 4.36) – the median catch efficiency is approximately 0.85 for mean wind speeds less than 1 m/s, and between 0.6 and 0.8 for mean wind speeds between 1 m/s and 3 m/s. It is difficult to generalize the shielded gauge results, due to both the limited range of wind-speed values for this sheltered, low-wind site, and to the fact that data from only one shielded gauge are available for the assessment. Further, the catch efficiency for the shielded gauge may be artificially enhanced by snow accumulated on the horizontal shield components blowing into the gauge orifice, which likely explains the higher False Alarm Rate for the shielded EML UPG1000 gauge relative to the unshielded Meteoservis MR3H-FC gauge at Sodankylä (Section 4.1.2.3.2).

For the HSA TBH gauges (Figure 4.37), significant variability in the catch efficiency vs. wind speed relationship is observed because of the heating being triggered by the presence of snow at a set height in the funnel, rather than by the ambient temperature. There are also relatively few events compared to the other gauges. Considering only the data with five or more events in a given bin, the median catch efficiency decreases with increasing wind speed: from 0.75 to 0.9 for mean gauge-height wind speeds < 2 m/s, to 0.4 to 0.8 for mean wind speeds < 3 m/s, to 0.5 to 0.6 for mean wind speeds < 4 m/s, and to 0.2 to 0.3 for mean wind speeds < 5 m/s. While this general trend is instructive, the specific catch efficiency ranges for a given wind speed should be regarded with caution given the marked variability in results, particularly that observed between the two identical TBH test gauges installed at different locations at the same site (Marshall).

4.1.2.3.3.1.2 Results by site

By comparing catch efficiency results from gauges installed at the same site, it is possible to compare the performance of different gauge types under the same conditions, recognizing that some differences in the spatial distribution of precipitation and/or the configuration of specific instruments may still exist. Results are presented only for the CARE, Marshall, and Sodankylä sites, which had multiple heated tipping bucket gauges under test. Results for Formigal and Weissfluhjoch, which both had only one heated tipping bucket gauge under test, are available in Figure 4.34 and Figure 4.35, respectively.

With the exception of the HSA TBH gauges, the solid precipitation catch efficiency for test gauges at CARE and Marshall (Figure 4.38 and Figure 4.39, respectively) decreases with increasing mean wind speed at gauge height following the general trend outlined in Section 4.1.2.3.2. The median catch efficiency varies within 0.1 and 0.2 in a given wind-speed bin, indicating similar performance for unshielded tipping bucket gauges with heating application determined by temperature. The median catch efficiency vs. mean gauge-height wind speed relationships for HSA TBH gauges at both sites follow the general trend outlined in Section 4.1.2.3.3.1.1, but should again be regarded with caution, given the apparent variability in catch efficiency results for these test gauges.

The catch efficiency results for Sodankylä are provided in Figure 4.40. For the shielded gauge (EML UPG1000), the median catch efficiency is between about 0.65 and 0.82 for mean gauge-height wind speeds between 1 m/s and 3 m/s, while for the unshielded gauge (Meteoservis MR3H-FC), the median catch efficiency is between about 0.38 and 0.65 over the same mean wind-speed range. While these results imply the benefit of shielding for heated tipping bucket gauge observations, the limited range of wind speeds experienced at Sodankylä and perceived influence of false tips from snow accumulated on the wind shield preclude any conclusive statements regarding shielding effects for this dataset.

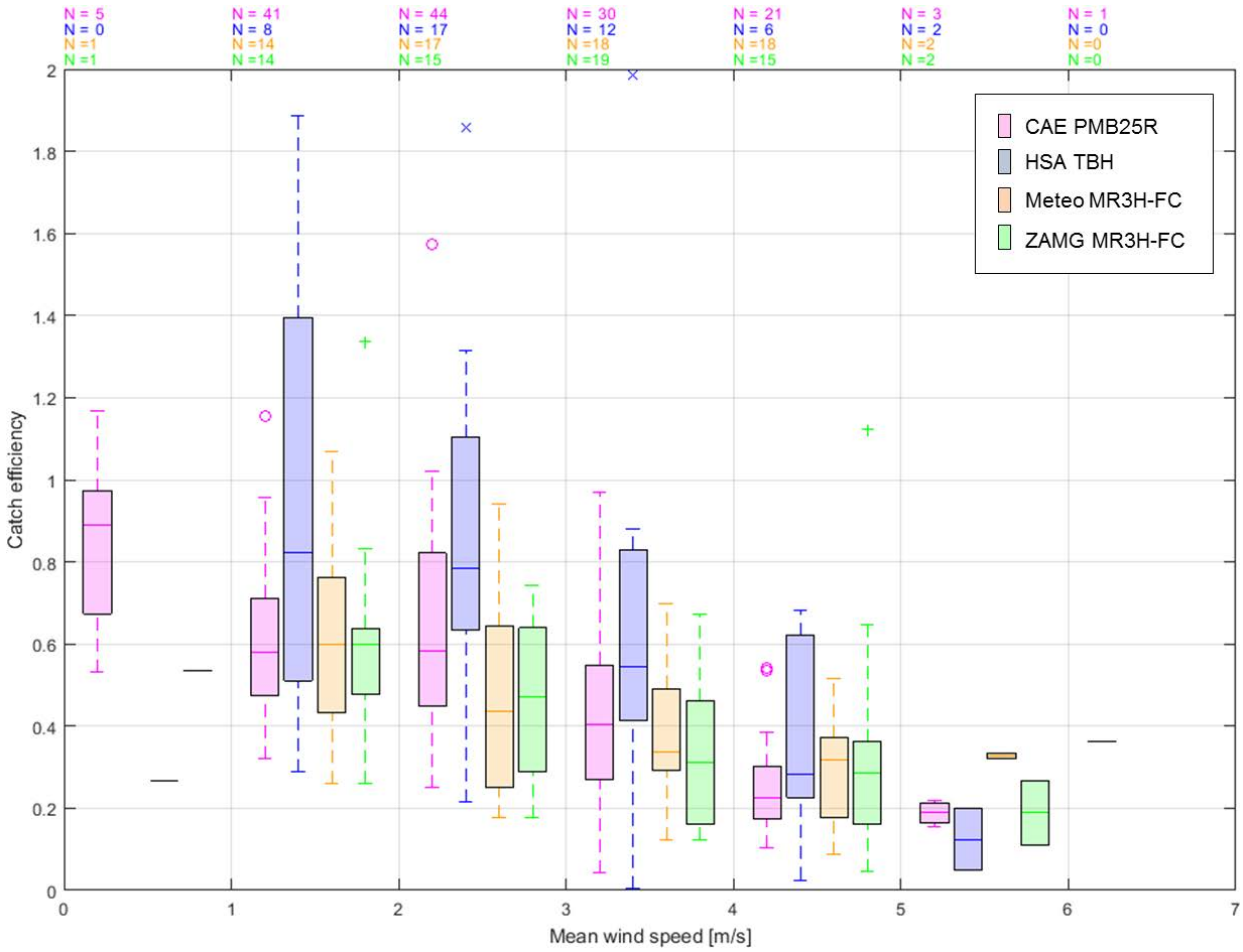


Figure 4.38. Box and whisker plot of catch efficiency as a function of mean gauge-height wind speed for gauges under test relative to site reference configuration at CARE for 60-minute snow events (max T ≤ -2 °C). The number of events in each wind-speed bin, for each test gauge, is indicated above the plot.

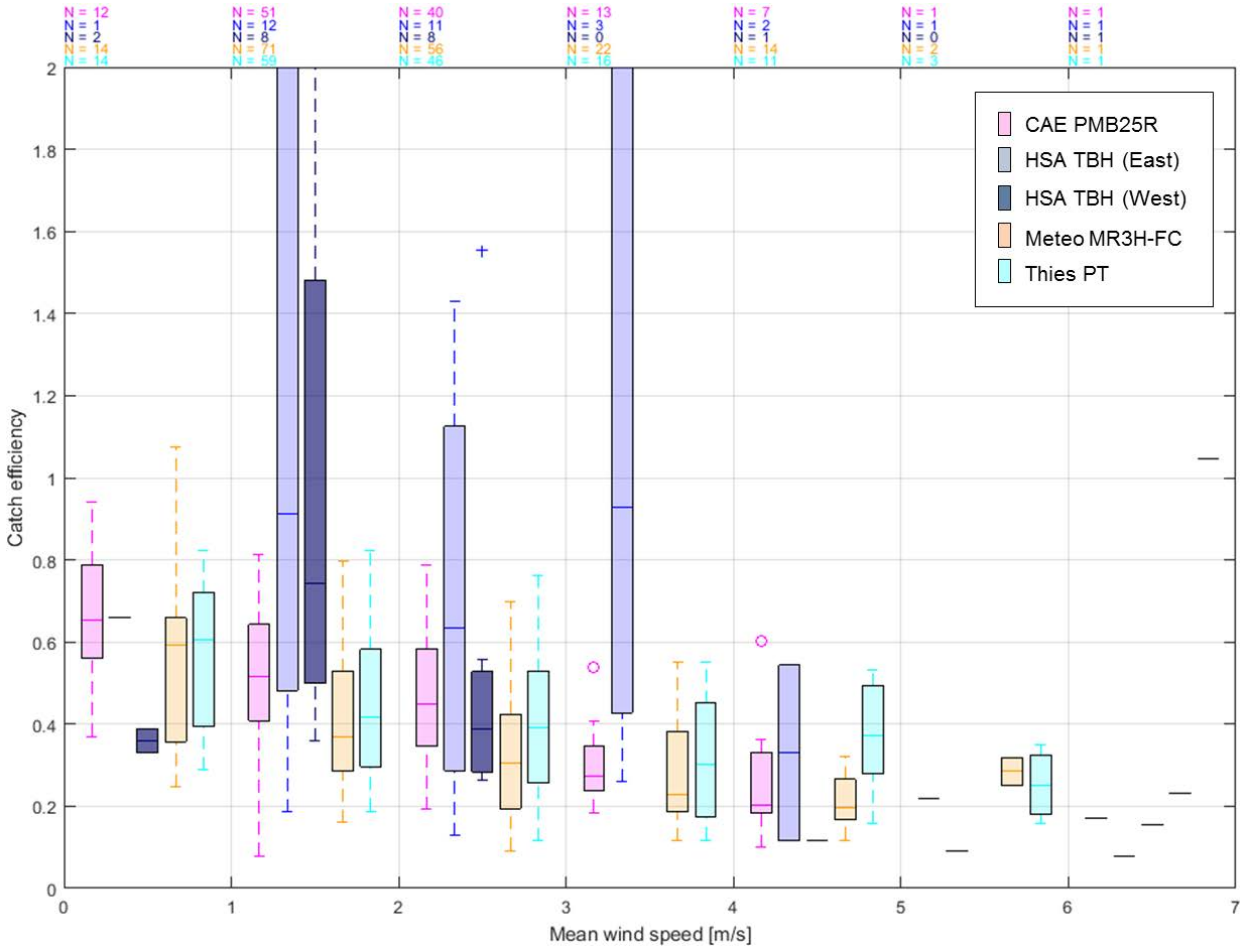


Figure 4.39. Box and whisker plot of catch efficiency as a function of mean gauge-height wind speed for gauges under test relative to site reference configuration at Marshall for 60-minute snow events (max T ≤ -2 °C). The number of events in each wind-speed bin, for each test gauge, is indicated above the plot.

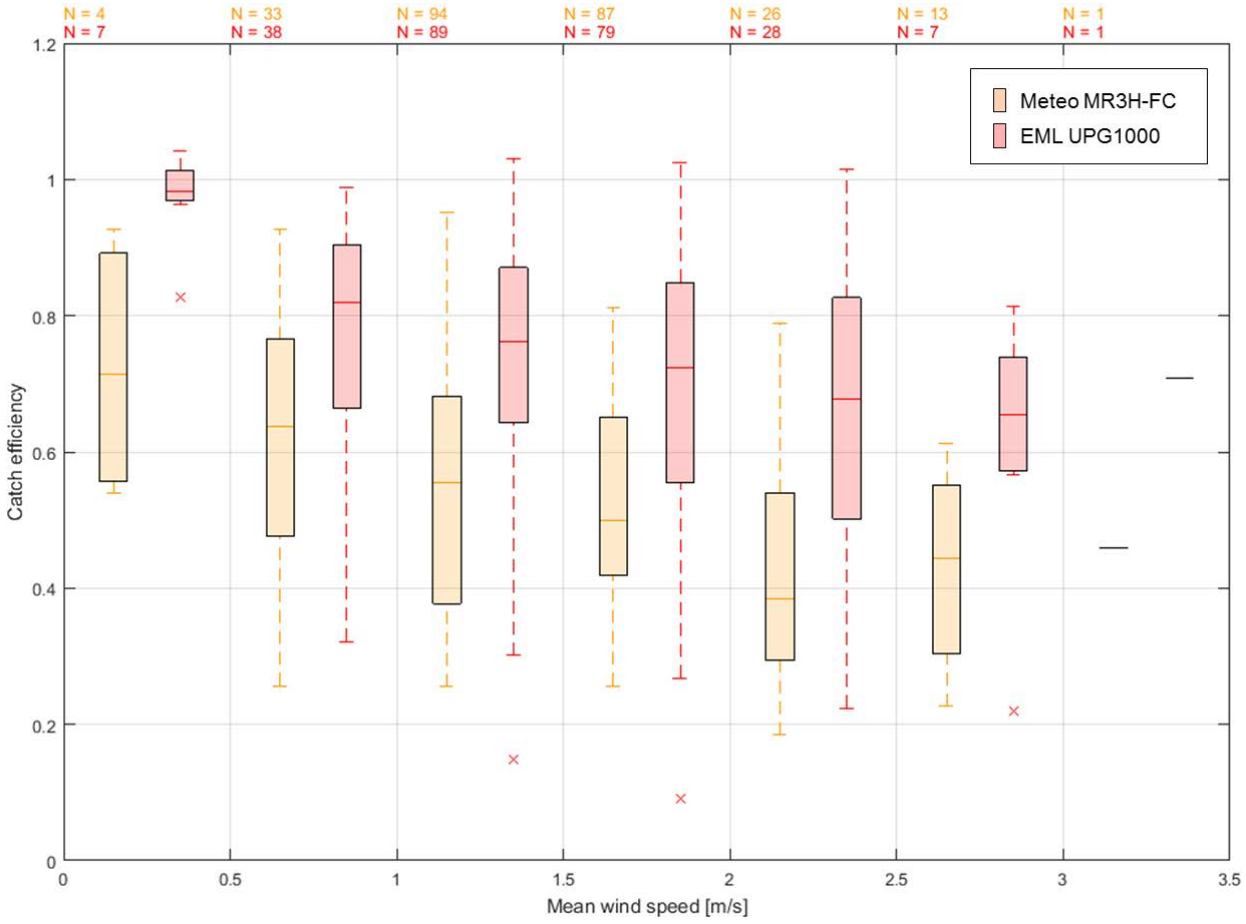


Figure 4.40. Box and whisker plot of catch efficiency as a function of mean gauge-height wind speed for gauges under test relative to site reference configuration at Sodankylä for 60-minute snow events (max T ≤ -2 °C). The number of events in each wind-speed bin, for each test gauge, is indicated above the plot.

4.1.2.3.3.2 Root mean square error

The root mean square error was calculated for 60-minute events during which the test gauge and reference both detected and reported precipitation (Section 3.6.1.4.1). The RMSE for each test gauge, at each site, for both solid precipitation events and events in all precipitation types, is shown in Figure 4.41. For solid precipitation (Figure 4.41a), the RMSE is generally between about 0.5 mm/60 minutes and 0.8 mm/60 minutes for tipping bucket gauges in continental climates, with values between 0.7 mm/60 minutes and 0.82 mm/60 minutes at CARE and between 0.5 mm/60 minutes and 0.8 mm/60 minutes at Marshall. The East HSA TBH gauge at Marshall is the notable exception, with a significantly higher RMSE of 1.2 mm/60 minutes. For solid precipitation in alpine climates, the RMSE for the test gauge at Weissfluhjoch is ~0.78 mm/60 minutes, which is similar to the values for continental climates, while a higher value of ~1.25 mm/60 minutes is observed for Formigal. These differences are attributed to the fact that the alpine test sites employ different test gauges, with different heating power (Table 4.6), and experience different snowfall conditions (Table 4.9). The lowest RMSE values for solid precipitation are observed for the test gauges in northern boreal climates, with values of ~0.2 mm/60 minutes to 0.3 mm/60 minutes attributed to the low mean wind speeds and small mean event accumulations observed at Sodankylä.

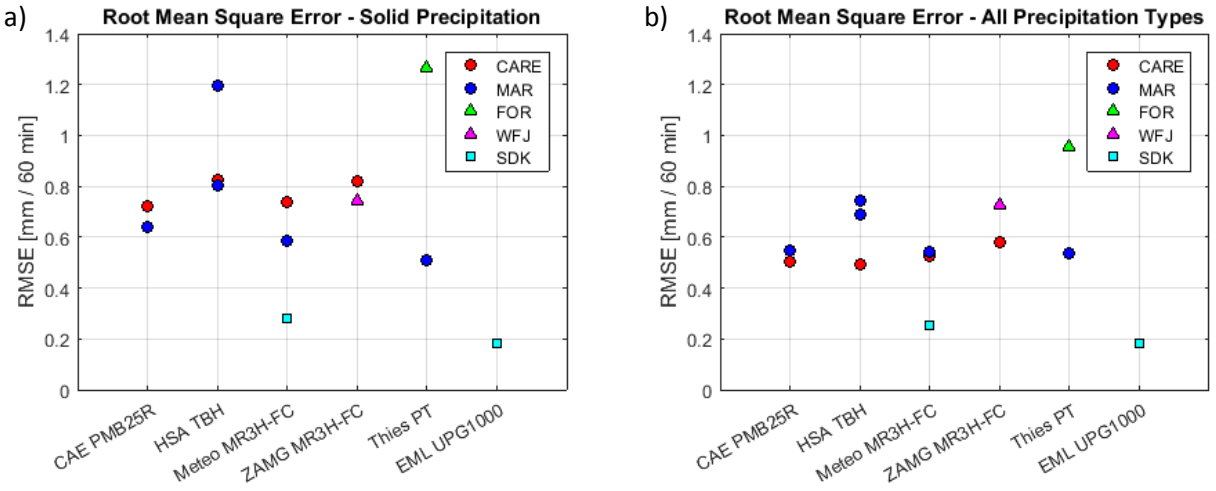


Figure 4.41. Root mean square error calculated for 60-minute YY events in (a) snow and (b) all precipitation types. Gauges at sites in continental, alpine, and northern boreal climate zones are represented by circles, triangles, and squares, respectively.

For 60-minute events in all precipitation types (Figure 4.41b), the RMSE is between ~ 0.5 mm/60 minutes and 0.75 mm/60 minutes for test gauges in continental climates, between ~ 0.75 mm/60 minutes and 0.95 mm/60 minutes for test gauges in alpine climates, and between ~ 0.2 mm/60 minutes and 0.3 mm/60 minutes for test gauges in northern boreal climates. These error values include losses due to wind speed, evaporation from heating, and differences related to gauge siting, configuration, and/or gauge operation. These RMSE values, which cover the full range of conditions experienced over two winter seasons, can be considered to represent the absolute magnitude of uncertainty for each tipping bucket gauge under test relative to the reference configuration at each site, and in each climate regime.

Differences in the RMSE values for solid precipitation and all precipitation types for a given test gauge/site are affected by the relative proportions of solid, liquid, and mixed-phase precipitation events. The similarity of RMSE values for the test gauge at Weissfluhjoch in Figure 4.41a and Figure 4.41b reflects the predominance of solid precipitation events (Table 4.8). The similarity of RMSE values in Figures 14a and 14b for the test gauge at Sodankylä, on the other hand, is believed to result from the low mean wind speeds during precipitation events, which are similar for all precipitation types and for snow, only (Table 4.7 and Table 4.9, respectively). The variability of RMSE results for test gauges at the other sites between Figure 4.41a and Figure 4.41b indicates a larger proportion of mixed and liquid precipitation events at these sites. The relative proportions of events with different precipitation types also influences the overall catch efficiencies observed for the test gauges at each site, discussed in the following section.

4.1.2.3.3.3 Overall catch efficiency

The long-term performance of each test gauge with respect to its ability to report accumulated precipitation of all types (liquid, mixed, solid) relative to the reference was assessed using the overall catch efficiency (Figure 4.42). The overall catch efficiency was calculated using the total accumulation

reported by the test gauge and reference during all 60-minute intervals in which either gauge detected precipitation. Hence, the assessment is not predicated on detection within a given 60-minute interval, mitigating the influence of response delays. The overall catch ratio is therefore less reflective of performance over operational time scales (e.g. 1 hour) and more reflective of longer-term or seasonal performance.

The overall catch efficiency for heated tipping bucket test gauges in continental climates varies from approximately 0.6 to 0.72. Of particular note, the overall catch efficiencies for the HSA TBH gauges at CARE and Marshall are all within about 0.67 to 0.72, indicating seasonal performance similar to the other test gauges. The TBH heating configuration requires less power relative to the other gauges and is evidently effective for long-term reporting of accumulated precipitation, despite the variable detection statistics presented for this gauge in Section 4.1.2.3.2.

For heated tipping bucket test gauges in alpine regions, the overall catch efficiency varies from ~ 0.48 for the ZAMG MR3H-FC at Weissfluhjoch, where the precipitation was predominantly solid (and hence subject to greater wind-induced undercatch; see phase breakdown in Table 4.8), to ~ 0.8 for the Thies Precipitation Transmitter at Formigal, where liquid and mixed-phase precipitation events were more prevalent (Table 4.8). For the unshielded Meteoservis MR3H-FC test gauge at Sodankylä, the overall catch efficiency is ~ 0.58; this value is lower than may be expected for a sheltered site characterized by low winds (Table 4.7) but can perhaps be attributed to the high heating power causing evaporation during the characteristically light snow events (mean accumulation of 0.58 mm; Table 4.9).

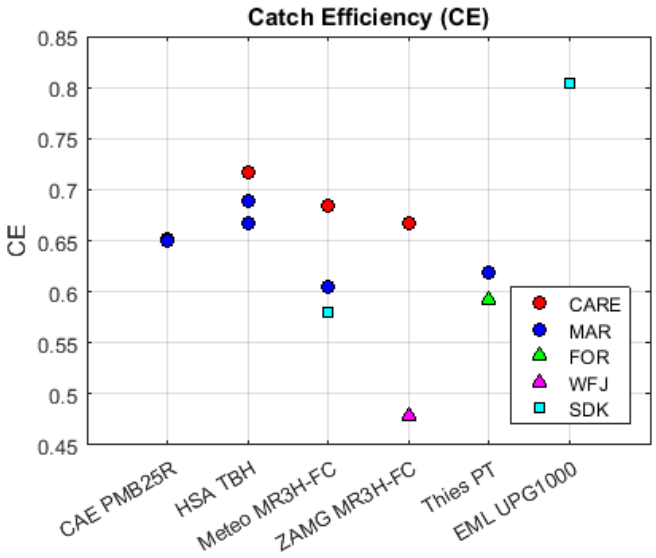


Figure 4.42. Overall catch efficiency calculated for all 60-minute YY, YN, and NY events in all precipitation types. Gauges at sites in continental, alpine, and sub-arctic/northern boreal climate zones are represented by circles, triangles, and squares, respectively.

The overall catch efficiency for all unshielded tipping bucket gauges under test, in all climate zones, is between approximately 0.5 and 0.7. This is an interesting result, given that the numbers and relative proportions of liquid, mixed, and solid precipitation events varied by site (Table 4.8), influencing the

overall catch efficiency (i.e. sites with a larger proportion of liquid events are likely to have higher catch efficiencies relative to sites with a larger proportion of solid precipitation events). One may expect the catch efficiencies to be higher at Sodankylä because of the low mean wind speeds. Indeed, the shielded EML UPG1000 test gauge at Sodankylä shows the highest overall catch efficiency of approximately 0.8, but this value may be inflated by false reports from the blowing of snow accumulated on the horizontal shield components into the gauge orifice. The lower heating power of the EML UPG1000 may be a better match for the lighter snow events in this low-wind, northern boreal environment relative to the Meteoservis MR3H-FC, which has higher heating power and showed a lower overall catch efficiency of ~0.58; however, it is difficult to separate the contributions from heating and shielding when assessing differences in overall catch efficiency.

4.1.2.3.4 *Characterization of response delays*

Response delays for heated tipping bucket gauges relative to the onset of precipitation, as identified by the corresponding reference configurations, were determined using tipping bucket comparison events, as detailed in Section 3.6.1.4.4. PDFs of response delays for each tipping bucket test gauge (at all applicable sites) and for each test site (for all tipping bucket gauges tested) are provided in Figure 4.43 and Figure 4.44, respectively. The TBCEs cover all precipitation conditions during the test periods (one or two winter seasons; see Table 4.7), and hence, the PDFs can be considered to be representative of delays in winter conditions at each test site.

The results for all applicable sites are included for each test gauge in Figure 4.43 for comparison. Note that the Thies Precipitation Transmitter results in Figure 4.43 represent two different models, with different tip resolutions and heating configurations, and cannot be compared directly. In general, the delays with the greatest frequency of occurrence for all test gauges are within 30 to 35 minutes, with the peaks in the distributions typically between about 20 and 35 minutes. The interpretation of observed differences in the PDFs between or among sites (for a given test gauge) is not straightforward, as the specific conditions (including the relative numbers of liquid, mixed, and solid events), the ice crystal densities and shapes, gauge siting and configuration can all differ from site to site.

The same delay results are presented for each test site in Figure 4.44. Here, the comparison of test gauges at a given site carries more weight, as the gauges were subject to the same conditions (noting that differences may still have occurred due to gauge siting and the spatial distribution of precipitation). The same general trends are observed as identified above with respect to Figure 4.43, with differences in the peaks of the PDFs at CARE, Marshall, and Sodankylä – that is, the sites with multiple test gauges – likely resulting from differences in heating power and specific configuration between or among the test gauges. At Formigal, the peaks in the PDF are less pronounced, with little variation in the frequencies of different delay times up to about 55 minutes. This could be related to the characteristically large event accumulations (Table 4.7) and relatively low heating power of the heated tipping bucket gauge being tested (49 W; see Table 4.6).

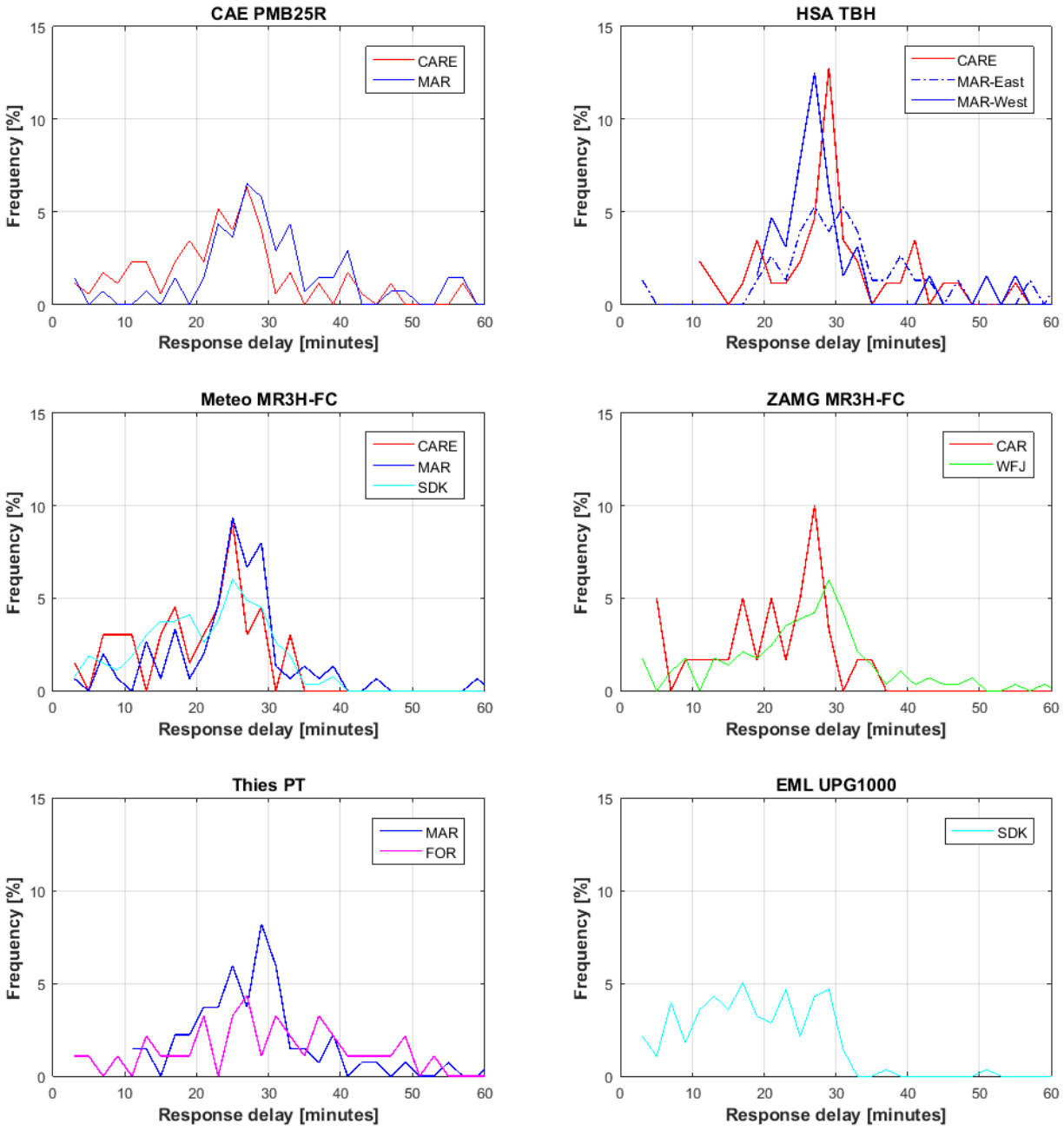


Figure 4.43. Probability density functions of response delays for each heated tipping bucket gauge at applicable test sites.

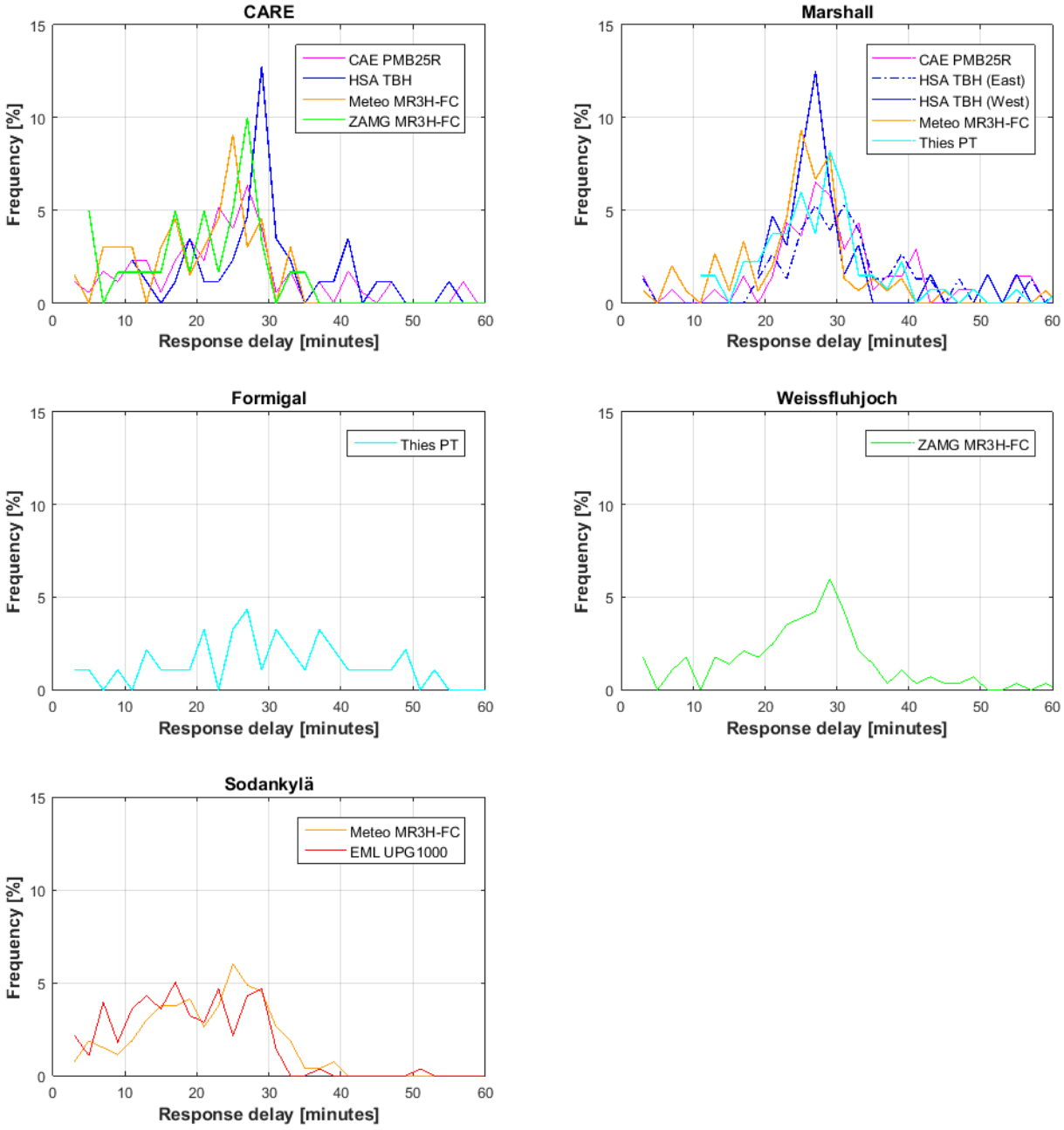


Figure 4.44. Probability density functions of response delays for all heated tipping bucket tested at each site.

4.1.2.3.4.1 Influence of snowfall intensity on response delays

Response delays are plotted as a function of the mean precipitation intensity reported by the reference during the delay period for each tipping bucket gauge (at all applicable sites) and each test site (for all test gauges) in Figure 4.45 and Figure 4.46, respectively. Data are presented only for snowfall events, identified as those during which the maximum temperature during the delay period did not exceed -2 °C. For all test gauges for which the heating application is determined by temperature – specifically, the CAE PMB25R, Meteoservis MR3H-FC, ZAMG MR3H-FC, Thies Precipitation Transmitter, and EML UPG1000 – response delays are typically within 30 minutes for mean reference intensities under 0.4 mm/hour. For mean reference intensities above 0.4 mm/hour,

the delay times show more variability, within 30 minutes for some events, while exceeding 150 minutes for others. At higher snowfall intensities (> 0.4 mm/hour), more of the energy applied to the heaters is required to melt the precipitation. This may extend the delay times, as less energy is available to increase or maintain the temperature, potentially reducing the efficacy of heating/melting later in the event. The duration of events may also play a role, but is beyond the scope of the present assessment. Note that some events with delays longer than 30 minutes are observed for snowfall intensities < 0.4 mm/hour; these may represent cases/events with variable precipitation intensity (some periods of higher intensity, some periods with no precipitation) over a longer duration.

The site-specific results in Figure 4.46 highlight the marked differences in the results for the test gauges at Formigal and Sodankylä. At Formigal, there are relatively few events meeting the criteria for snow, and the low heating power for the model of Thies Precipitation Transmitter tested (49 W; see Table 4.6) may not have been sufficient for the heavier, wet snow observed at this site (Table 4.9). At Sodankylä, the majority of snow events have intensities under 0.6 mm/hour, and the observed response delays do not exceed about 40 minutes. For the specific gauges tested, response delays appear to be less significant for the low-intensity snowfall events at this northern boreal site.

4.1.2.4 Alternative assessment approach: peak-to-peak (P2P)

The temporal separation of heated tipping bucket gauge reports from those of the reference configuration employing a weighing gauge complicates the assessment of tipping bucket gauge performance. An alternative approach, which reflects the different response mechanism of heated tipping bucket gauges, considers assessment intervals of variable duration, based on the time between successive tipping bucket gauge reports. This approach is referred to as the 'peak-to-peak,' or P2P approach, and is presented in Section 4.1.2.4.1. The P2P approach and results are presented for demonstration purposes only; the WMO-SPICE recommendations for heated tipping bucket gauges (Section 4.1.2.6) are based on the results for assessment intervals of fixed duration, consistent with the assessment approach for other gauge types (weighing gauges, non-catchment instruments) in the intercomparison, and with how tipping bucket gauges are typically used in operational settings.

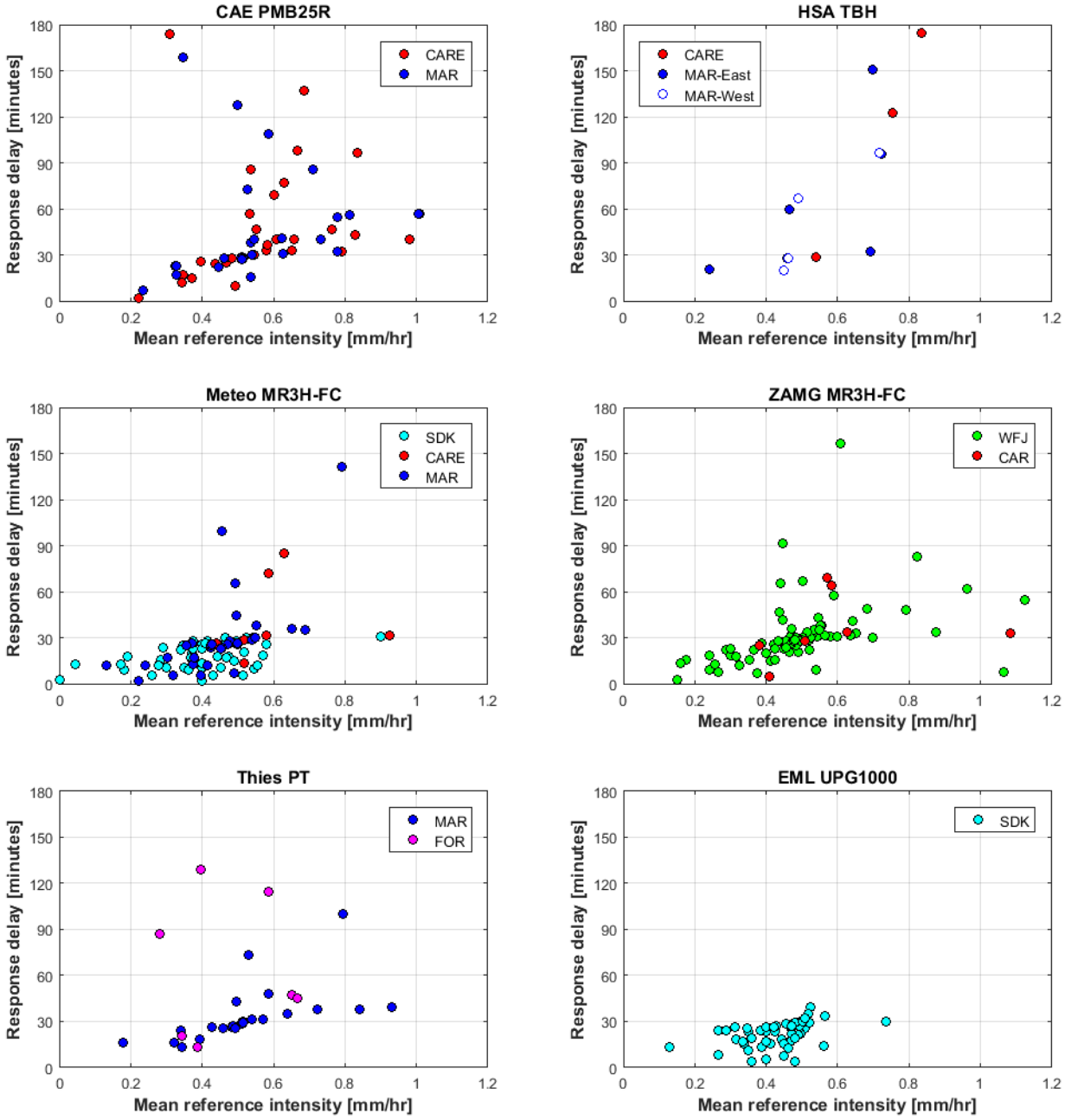


Figure 4.45. Response delays for each heated tipping bucket gauge tested as a function of the mean intensity of precipitation reported by the reference during the period from the onset of precipitation to the first tipping bucket response. Data are presented only for snow events, for which the maximum temperature during the delay period did not exceed -2 °C.

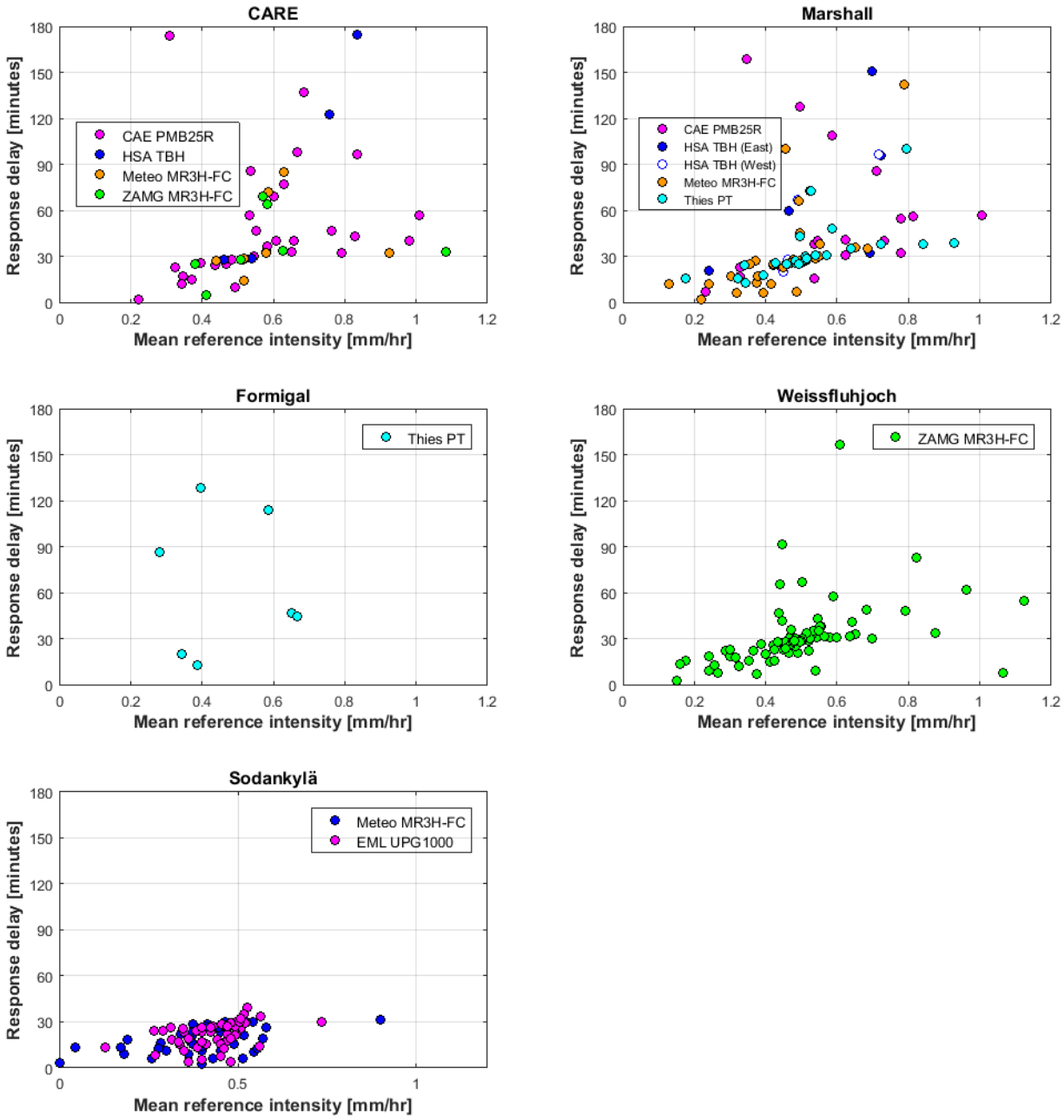


Figure 4.46. Response delays for tipping bucket gauges tested at each site as a function of the mean intensity of precipitation reported by the reference during the period from the onset of precipitation to the first tipping bucket response. Data are presented only for snow events, for which the maximum temperature during the delay period did not exceed -2 °C.

4.1.2.4.1 *P2P approach*

To constrain the P2P assessment to periods of precipitation (as determined by the reference configuration), the P2P periods were identified within each TBCE considered in the response time assessment (Section 4.1.2.3.4; more details in Section 3.6.1.4.4). The requirement of at least 180 minutes without precipitation between successive TBCEs provides additional time for tipping bucket gauges to respond to all precipitation reported by the reference configuration within the TBCE. In some cases, the tipping bucket response extends beyond one minute, resulting in accumulation

reports over two or more consecutive minutes; in these cases, the last consecutive minute during which the tipping bucket reports precipitation is taken as the P2P start/end point.

An example case is provided in Figure 4.47 for the CAE PMB25R gauge under test at CARE relative to the DFAR (Geonor T-200B3 600 mm in DFIR-shield) for a TBCE in February, 2015. In this particular case, the tipping bucket gauge did not respond to the precipitation reported by the reference at the end of the event; the assessment is therefore limited to the four P2P periods between about 04:00 and 12:00 UTC. The periods from the onset of the event to the first tipping bucket response are not considered in the P2P assessment, as the tipping bucket response will be influenced by the conditions experienced since the last tip (sometime before the TBCE) and the bucket or funnel may already contain precipitation which could impact the assessment.

Compiling P2P data from TBCEs over the duration of formal tests, the catch efficiency during P2P periods for solid precipitation (maximum temperature ≤ -2 °C over P2P period) is plotted as a function of the corresponding mean wind speed at gauge height for all test gauges at CARE in Figure 4.48. Note that the CAE PMB25R and HSA TBH data are derived from measurements over both test seasons, while the Meteoservis MR3H-FC and ZAMG MR3H-FC data are derived from measurements over the second season only. The y-axis is limited to catch efficiency values less than 2; P2P periods with larger catch efficiencies are present, but an investigation of their causes is beyond the scope of this assessment, which is included primarily for demonstration purposes.

To more clearly illustrate the catch efficiency – wind speed relationships, the same data are presented as a box and whisker plot in Figure 4.49. These P2P results can be compared against those for 60-minute intervals during which the reference and test gauges both reported precipitation. These are presented in the same format in Figure 4.38. The overall trends for changes in catch efficiency with increasing wind speed differ between the plots, as do the number of cases/assessment intervals in each wind-speed bin, for each test gauge. Hence, the catch efficiency for different test gauges can vary depending on the assessment approach. The assessment approach used should therefore be selected based on the specific application, as the catch efficiency provides the basis for adjustment functions, which can be applied to compensate for wind-induced undercatch in solid precipitation conditions.

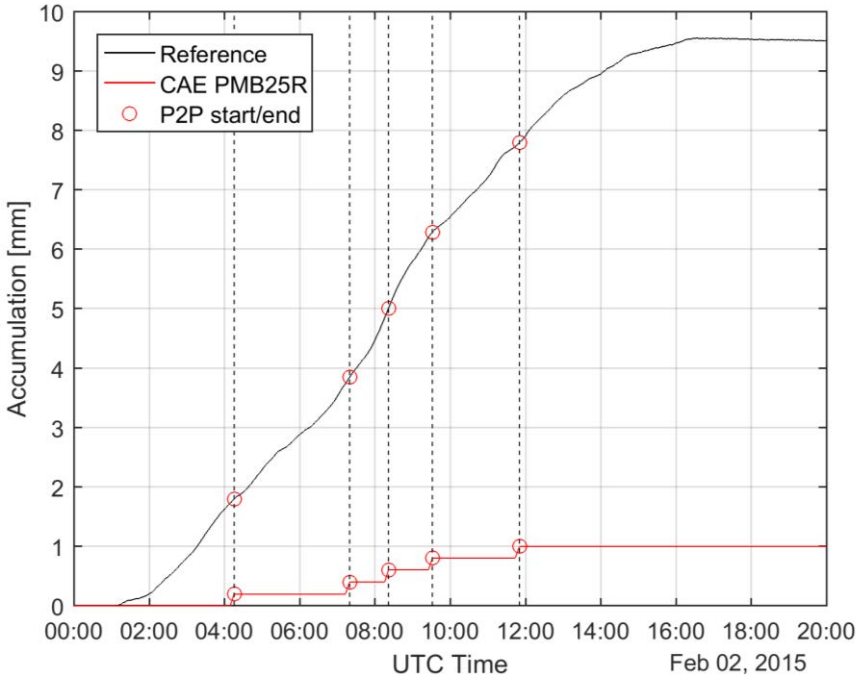


Figure 4.47. Time series of accumulated precipitation reported by CAE PMB25R test gauge and R2 reference configuration at CARE for a TBCE on Feb. 2, 2015. The P2P periods are contained within subsequent pairs of dashed lines, with the red circle markers on the time series marking the start and end points.

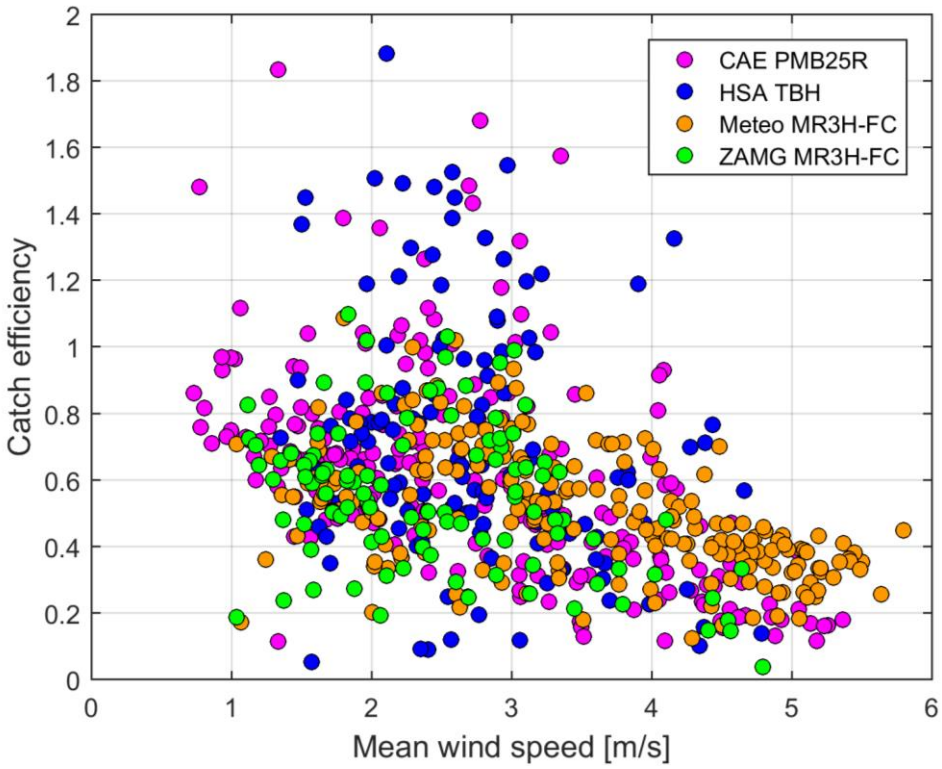


Figure 4.48. Catch efficiency as a function of mean gauge-height wind speed for all P2P cases in solid precipitation (maximum temperature ≤ -2 °C) for all heated tipping bucket test gauges at CARE over the duration of formal tests.

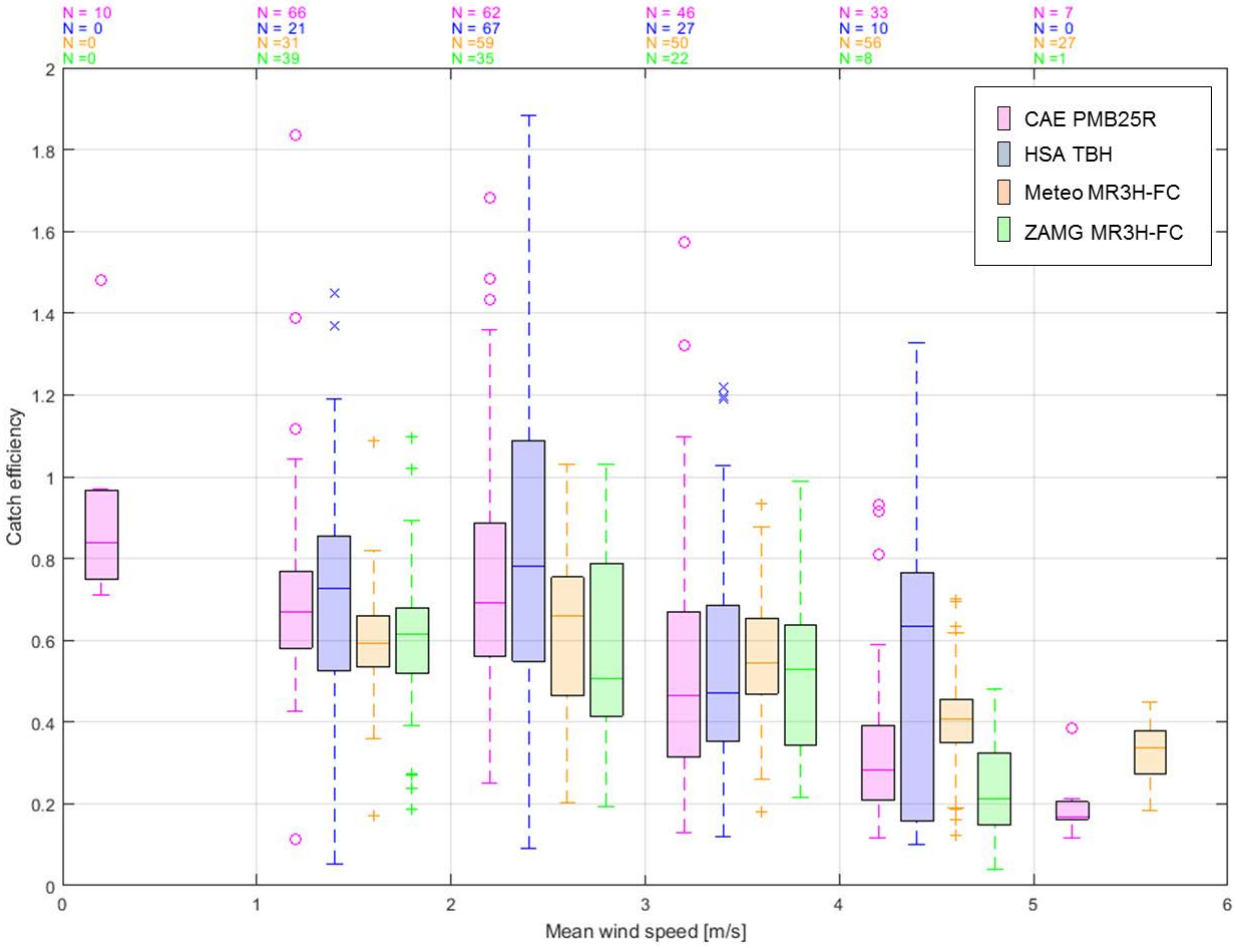


Figure 4.49. Box and whisker plot of catch efficiency as a function of mean gauge-height wind speed for all P2P cases in solid precipitation (max T ≤ -2 °C) for all tipping bucket test gauges at CARE over the duration of formal tests. The number of events in each wind-speed bin, for each gauge tested, is indicated above the plot.

4.1.2.5 Adjustment functions

The results presented in Section 4.1.2.3.3.1 illustrate how the catch efficiency of heated tipping bucket gauges decreases with increasing wind speed for snowfall events. Wind effects can therefore lead to the under-reporting of precipitation accumulation, with implications for meteorological operations, climate records, and hydrological modeling. The development of adjustment functions, or transfer functions, to adjust or "correct" reported precipitation amounts for wind effects are therefore of significant value to these and other domains. The importance of transfer functions for tipping bucket gauges is underscored by the broad use of these gauges operationally, worldwide (Nitu and Wong, 2010).

Buisán et al. (2017) developed a transfer function for the heated tipping bucket gauge used in the Spanish operational network. This function was applied to data from snowfall events at various sites in the network and used to demonstrate specific regions/environments where the underestimation

of snowfall may be more significant. In this study, the transfer function was derived from measurements at a single site over two successive winter seasons. Additional work is required to extend the development of transfer functions to different heated tipping bucket gauge and shield configurations in different environments.

Kochendorfer et al. (2017a, 2017b) developed transfer functions for weighing gauges at different test sites, but also consolidated the data from multiple test sites for the development of "universal" transfer functions. The universal transfer functions were broadly applicable to data from sites in different climate regimes (though site-specific biases were still observed in the results) and found to perform similarly to, or better than, the functions developed for specific gauges and sites. A similar approach will be tested on heated tipping bucket gauge measurements from WMO-SPICE in future investigations.

Transfer functions are predicated on the catch efficiency relative to a site reference configuration. The WMO-SPICE R2 field reference configuration has been used in the studies noted above and is recommended as a site reference configuration for transfer function development. The wind speed employed should be measured at gauge height, or at the standard wind measurement height for a given operational network. Ambient temperature reports are used in WMO-SPICE for precipitation phase identification, but other methods (present weather sensors, observer reports) may be employed.

4.1.2.6 Recommendations

Drawing upon the collective experience of WMO-SPICE site teams with the heated tipping bucket gauges tested, and the results and interpretation presented in this section (for all test gauges) and in the Instrument Performance Reports (for each specific gauge tested; see Annex 6), a set of recommendations has been developed to guide individuals or organizations with interest in employing heated tipping bucket gauges for the measurement of precipitation. These recommendations are grouped into four categories – gauge selection, gauge configuration and siting, adjustment functions and required ancillary measurements, and operational use – and presented in the following sections.

4.1.2.6.1 *Gauge selection*

The selection of a heated tipping bucket gauge should be based on the suitability of the gauge and its configuration to the specific application and site conditions, as well as the availability of site infrastructure and resources. The range of expected conditions for a site, specifically temperature and intensity of precipitation, should fall within the limits stated by the gauge manufacturer. Similarly, the infrastructure at a site must ensure sufficient power for the operation and heating of a gauge; long cords should be avoided, as they may be subject to voltage drops and less effective heating. Sites requiring regular (e.g. hourly) gauge reports should employ gauges for which heating is triggered by temperature. These have shown better statistics with respect to the detection of precipitation over 60-minute periods than gauges for which heating is triggered only when snow is present in the funnel (Section 4.1.2.3.2). However, installations intended for remote or longer term reporting (e.g. seasonal) may benefit from gauges with the latter heating configuration, which have lower power requirements. These gauges have shown more variable detection statistics relative to the reference configuration over 60-minute periods (Section 4.1.2.3.2), but have also shown comparable overall performance to gauges with other heating configurations in terms of the total seasonal precipitation amounts reported (Section 4.1.2.3.3.3).

4.1.2.6.2 *Gauge configuration and siting*

Gauge heating power should not exceed site limitations (Section 4.1.2.6.1), and the specific heating configuration and power should be tailored to match the expected precipitation conditions. Some heating configurations are configurable, in that the temperature thresholds or set temperatures at which specific components are maintained can be modified, allowing users to modify heating to better suit their site conditions. In other cases, however, there is limited or no flexibility with respect to heating. This can potentially pose a problem, as default heating configurations may be guided by testing in the laboratory or in specific sets of conditions or climate regimes, and may not perform as expected in all conditions. It is recommended that prospective users consult with the gauge manufacturer(s) in the interest of finding the most appropriate heating configuration for their specific conditions. For gauge manufacturers, it is recommended that heating be configurable to optimize power consumption and/or to tailor heating for different conditions or applications, and that the related algorithm(s) are documented for the information of instrument and data users.

The assessment of heated tipping bucket gauges in shielded configurations in WMO-SPICE is limited to a single gauge and site, which makes it difficult to make broad recommendations regarding shielding. The benefits of shielding are demonstrated in separate work by Buisán et al. (2016) who compared precipitation observations from a shielded and unshielded tipping bucket gauge over a winter season at the Formigal site in Spain. Shielding is therefore recommended for tipping bucket gauge installations, where feasible. Ideally, the shield should be mounted separately from the gauge post to avoid the potential for wind-induced vibration to affect the tipping mechanism (causing false tips); however, this may not be possible in all environments or installations. Further, the specific shield configuration should not include horizontal (flat) components (like the shield depicted in Figure 4.29) upon which precipitation can accumulate and subsequently be blown into the gauge orifice.

It is important that the tipping bucket gauge outlet ports (see Figure 4.29) are not obstructed by the gauge post, mounting hardware, or other components below the gauge. Precipitation flowing out of the gauge bottom can collect and freeze on the surfaces of such obstructions, posing an outlet blockage risk. Inlet blockage can also occur, and gauge siting should give consideration to the proximity of potential blocking items or debris (e.g. needles from coniferous trees); however, some blockage may be unavoidable due to local conditions or wildlife (e.g. blowing sand, bird droppings, bird nests).

4.1.2.6.3 *Adjustment functions and required ancillary measurements*

Adjustment functions for heated tipping bucket gauges (Section 4.1.2.5) should be applied, where possible, to compensate for the wind-induced undercatch of snowfall. Depending on the expertise and resources available at a given site, adjustment functions may be derived following the approach of Buisán et al. (2017) or Kochendorfer et al. (2017a, 2017b). Suitable gauge- and/or region-specific adjustment functions, or functions applicable to a broader range of gauges and conditions (e.g. the "universal" transfer functions derived by Kochendorfer et al., 2017a), may be investigated in subsequent publications. The application of adjustment functions requires ancillary measurements of wind speed, ideally at gauge height, and temperature. The derivation of adjustment functions requires these ancillary measurements in addition to measurements from a suitable reference configuration, such as the WMO-SPICE R2 field working reference configuration.

4.1.2.6.4 *Operational use*

The use of heated tipping bucket gauges operationally is complicated by response delays, which are inherent to the principles of gauge operation and heating. Delay times are generally shorter than operational reporting intervals (Section 4.1.2.3.4); however, the specific timing of precipitation may lead to missed or false reports within a given reporting interval. The characterization of response delays in WMO-SPICE focused primarily on winter precipitation; the impact of delays is expected to be less significant in other seasons (more liquid precipitation). For all-season reporting of precipitation in operational settings, heated tipping bucket gauges may present a lower-cost alternative or complement to precipitation observations from weighing gauges, which generally respond in real-time to incident precipitation. If a heated tipping bucket gauge is used as the primary precipitation gauge at an observing site, the use of an appropriate adjustment function is recommended to compensate for wind-induced undercatch, if available.

4.1.3 Emerging technologies

Authors: Audrey Reverdin, Yves-Alain Roulet, Rodica Nitu, Michael Earle

4.1.3.1 Introduction

New technologies that measure precipitation without capturing the particles (solid or liquid) in a container (so called '"non-catchment" type instruments, or NCIs) are increasingly being used operationally. The measurements performed by these instruments are based on the principles of light scattering, microwave backscatter, or mass and heat transfer.

One of the SPICE objectives is to evaluate and characterize the ability of these instruments to serve as alternatives to the traditional gauges (tipping bucket or weighing gauges) for the measurement of solid precipitation quantity. While several studies have reported the performance of such instruments in measuring liquid precipitation (WMO/TD-No. 1504, Tokay et al. 2014, Lanzinger et al. 2006), their ability to adequately assess and report snowfall amounts is still to be demonstrated.

Seven different models of instruments employing emerging technologies were submitted by manufacturers and represent the SUT assessed during SPICE. They are listed in Table 4.10 along with the sites on which they were tested, for a total of 11 tested sensors. The three different instrument types tested cover a wide range of current emerging technologies and enabled the assessment of the performance of each measurement principle.

Table 4.10. Emerging technology instruments submitted by manufacturers.

Instrument Model	Instrument Type	Host SPICE sites
Thies Laser Precipitation Monitor shielded	Disdrometer	Marshall, Weissfluhjoch
OTT Parsivel^2 disdrometer	Disdrometer	Sodankylä
Campbell Scientific PWS100	Present weather sensor	Haukeliseter, Marshall
Vaisala FS11P (FS11/PWD32 combination)	Present weather sensor	Sodankylä
Vaisala PWD 33	Present weather sensor	Sodankylä
Vaisala PWD 52	Present weather sensor	Sodankylä
Yankee TPS3100 Hotplate	Evaporative plate	Marshall, Haukeliseter, Sodankylä

The instruments were tested at four SPICE sites, illustrated in Figure 4.50, each representing a different climate regime: Haukeliseter (alpine with high winds); Marshall (dry continental); Sodankylä (northern boreal with low wind); and Weissfluhjoch (alpine with complex topography). Note that the two Thies LPM sensors used a wind shield, as illustrated in Figure 4.51, while all other test instruments were unshielded.

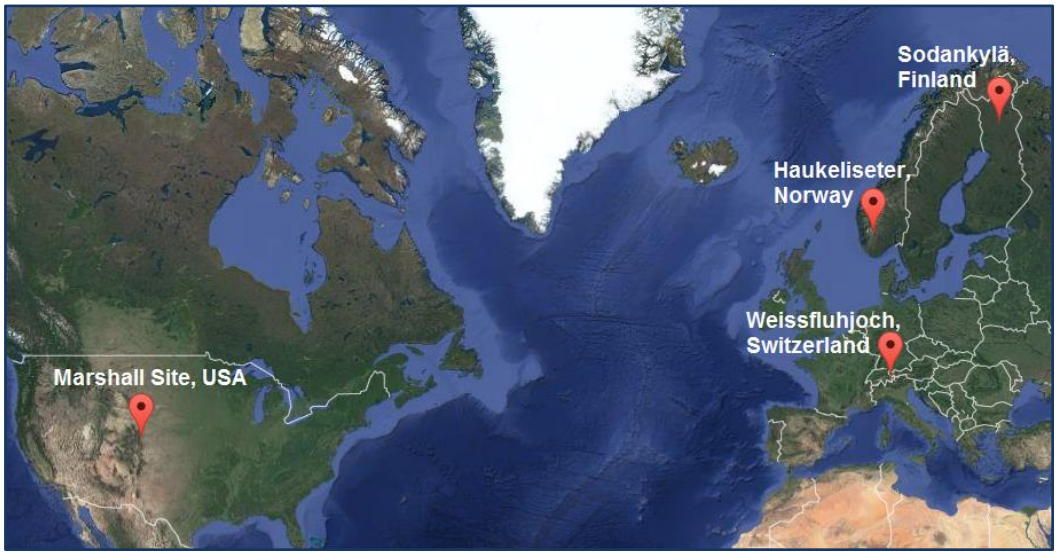


Figure 4.50. Sites hosting emerging technologies under test in WMO-SPICE.

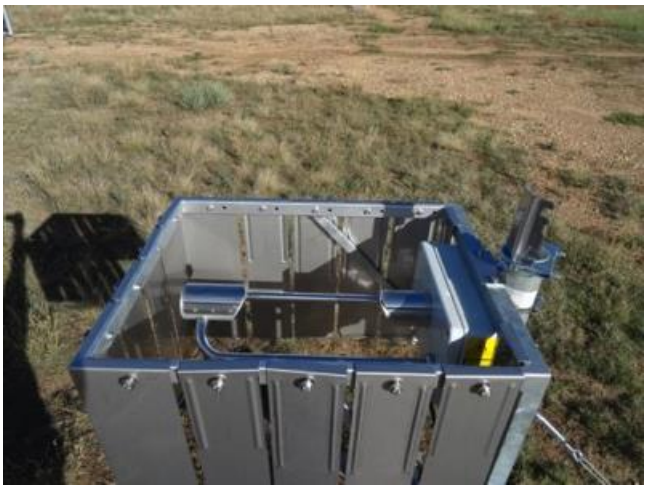


Figure 4.51. Thies LPM sensor at the Marshall site with wind shield.

A detailed assessment of the performance of each individual sensor, at each test site, is provided in the corresponding instrument performance reports (Annex 6). The following sections consolidate the key findings from these reports, both by sensor and by site, and use these findings to make recommendations regarding the field implementation of NCI. To establish context for the interpretation of results, related discussion, and recommendations, an overview of NCI sensor operation is provided.

4.1.3.2 Operating principle of NCI sensors

A primary use of NCI sensors is to report precipitation type. Therefore, most of these instruments report not only precipitation quantity or intensity, but also provide additional information such as weather type (SYNOP or METAR codes), particle type, particle size distribution, and fall-speed velocity, thus increasing the number of possible application areas for this type of instrument. For instance, present weather sensors that are based on the forward- and back-scattering of radiation by particles are widely used at airports to report weather type and visibility. The multiple output parameters and low cost of these instruments makes them attractive in terms of cost/benefit ratio. Their use will, however, depend on specific user needs, and it is therefore important to firstly understand what different operating principles are available.

4.1.3.2.1 *Forward scatter/backscatter present weather sensors*

The present weather sensors/detectors (PWS/PWD) measure light scattering caused by precipitation particles falling through a small, well defined volume of the atmosphere. The scatter of a light source is observed under a fixed angle (Figure 4.52) and gives information on the particles' size. Depending on the sensor model, additional measurements can be performed, such as water content (using a capacitive sensor), fall speed, and/or air temperature. The combination of this information allows for the determination of particle type (e.g. hail, snow, drizzle, rain) through proprietary internal algorithms.

Present weather sensors are often used at airports, and contribute to report weather information included in METAR messages.

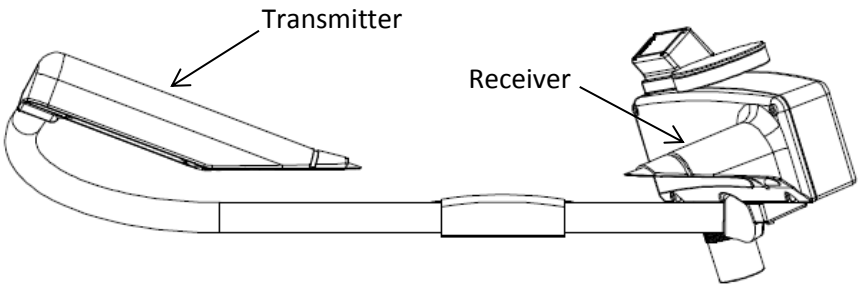


Figure 4.52. Schematic example of a present weather sensor (Vaisala PWD33 EPI User Manual).

4.1.3.2.2 *Disdrometers*

Disdrometers measure particle size and fall velocity, producing raw information in the form of a 2-D matrix (typically around 20 bins for each particle size and fall speed classification). As depicted in Figure 4.53, the instrument consists of a laser diode emitter (typical wavelength between 650 and 850 nm) and a receiver with an optical path of about 20-30 cm. In the absence of precipitation, the receiver produces a constant power signal output. Particles crossing the laser beam will reduce (attenuate) the received laser beam intensity, and hence the power output, which is proportional to the size of the particle. The duration of this reduction correlates with the particle fall speed. Combining these two pieces of information allows for the determination of particle type using a schematic concept such as that depicted in Figure 4.54 (Löffler-Mang and Joss, 2000). This processing is carried out by proprietary internal algorithms.

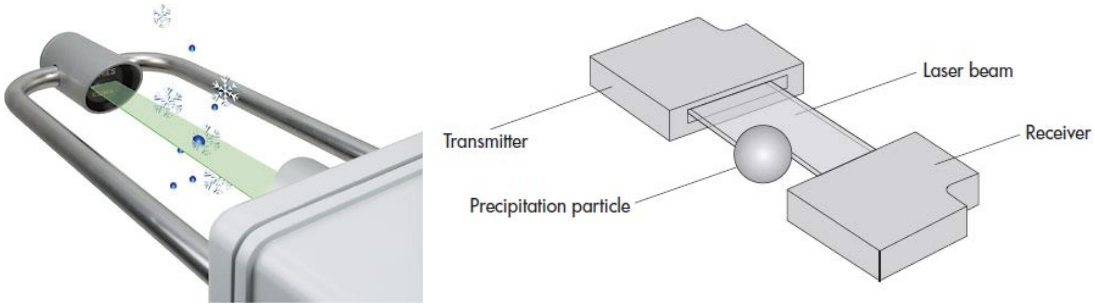


Figure 4.53. Disdrometer operating principle examples from (left to right) the Thies LPM and OTT Parsivel^2 User Manuals, respectively. Particles falling through a transmitted laser beam induce a decrease in the signal captured by the receiver.

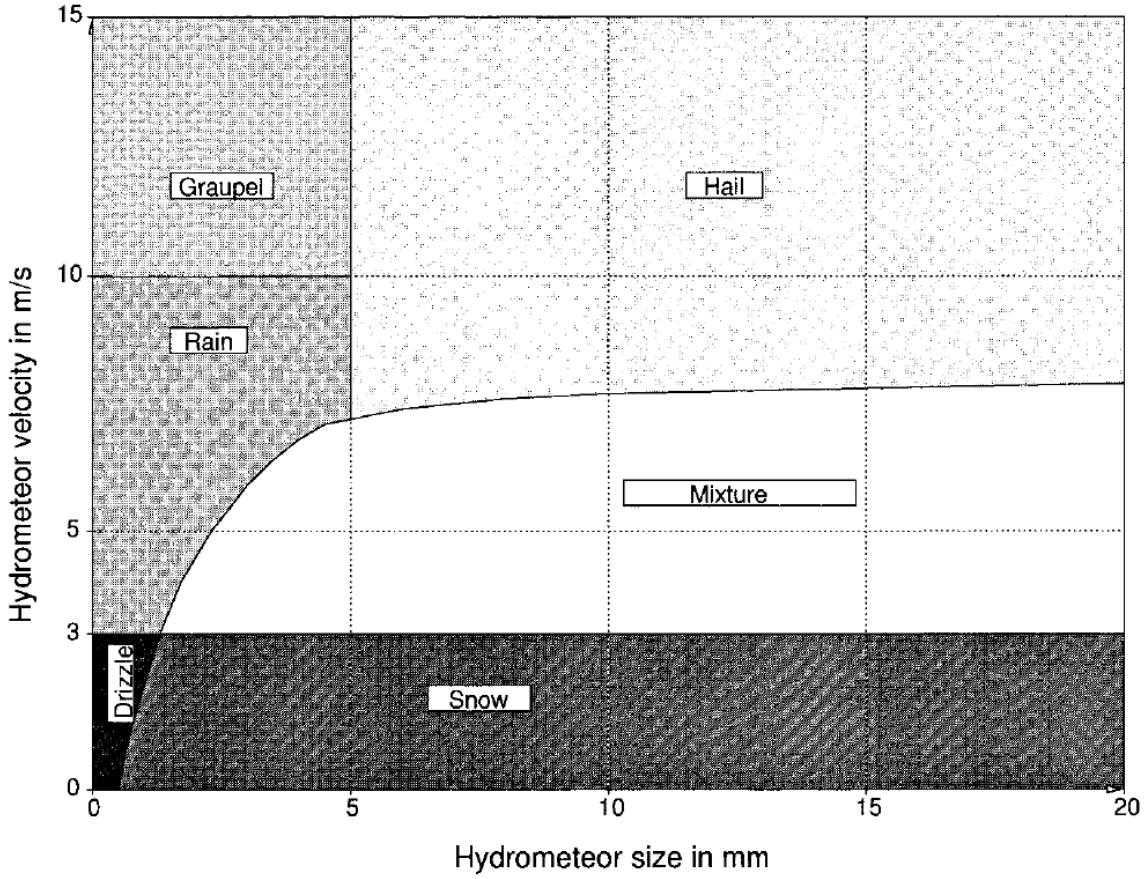


Figure 4.54. Theoretical curves for particle type discrimination as a function of hydrometeor size (X-axis) and fall velocity (Y-axis), from Löffler-Mang and Joss, 2000.

Like forwardscatter/backscatter present weather sensors, these sensors can generally be used to determine precipitation type (WMO-No.8, 2008). The performance compared to human observers is similar to that obtained with scatter sensors (Bloemink and Lanzinger, 2005). The measurement of solid precipitation accumulation with disdrometers is also influenced by configuration (shielding) and specific weather conditions (Reverdin et al., 2016).

4.1.3.2.3 *Evaporative plate (or hotplate)*

Author: Roy Rasmussen

Unlike PWD or disdrometers, the hotplate instrument (Rasmussen et al. 2011) directly measures the precipitation rate and is unique in its measurement principle in that it is based on measuring the power needed to keep a plate at constant temperature. A drawing of the evaporative plate is given in Figure 4.55.

The hotplate consists of two identical 13 cm diameter plates, one facing upward and one facing downward, one on top of the other and separated by an insulator. Both plates are heated to a constant temperature, typically about 75 °C. The top plate is the precipitation sensor, and the bottom plate is used to account for wind and other noise. The temperature of each plate is maintained at a constant level by a feedback circuit. If precipitation or wind cools the plate, the feedback circuit sends current to thermistors in each plate in order to maintain constant

temperature. The amount of current used to maintain constant plate temperature is proportional to the combined cooling associated with the melting and evaporation of snow and wind cooling. If the power of the top plate on which snow falls is subtracted from the power of the bottom plate, which is only subject to wind cooling, the power difference is proportional to the precipitation rate through the latent heat of melting and evaporation, if the particles are snow, and through the latent heat of evaporation, if the particles are rain.

Figure 4.55. Drawing of the TPS 3100 Hotplate sensor, as tested in WMO-SPICE, from the Hotplate User Manual.

Advantages of the hotplate technology are:

- Aerodynamic design leads to a relatively small disturbance of the upstream flow.
- Not sensitive to vibrations of the instrument.
- Spiders and birds do not produce webs or nests on the unit due to the high heat.
- Natural de-icing due to the high temperature of the plates.
- Estimate one minute liquid-equivalent rates without needing to know snow density.

Challenges for the hotplate:

- Need to account for undercatch at high wind speeds due to particles bouncing off the plate.
- Cannot measure one minute liquid-equivalent rates above 50 mm/hour.
- Higher measurement threshold needed to overcome noise at high wind speeds.
- Potential heat plume induced by high heat may deflect particles.

4.1.3.3 Potential limitations of NCIs for precipitation measurements

While the factors and the processes influencing measurements performed by conventional precipitation gauges (tipping bucket or weighing gauges) are well known and documented (e.g. see Sections 4.1.1 and 4.1.2), this is not the case for non-catchment-type instruments. Physical processes such as evaporation, sublimation, and hygroscopic effects (for particles collected from the ambient air), etc., are not relevant, and wind induced errors remain to be studied and understood. The shapes of these instruments are completely different than traditional gauges and are, generally, anisotropic, suggesting that any wind-induced errors could show directional dependencies. The Hotplate is circular, eliminating most wind direction dependencies.

Other processes can impact precipitation measurements made by non-catchment instruments, including:

- Droplets splashing on the instrument arms (depending on the design) that may artificially increase the number of particles (especially true for rain or wet snow).
- Superposition of hydrometeors along the beam that may corrupt the particle count and drop size distribution, hence the derived accumulation.
- Assumptions on hydrometeor shape and density that may lead to over/underestimation of the equivalent water content (accumulation).
- Internal temperature measurement (if not ventilated) that may affect the decision algorithm for precipitation type.
- Power outages or data transmission issues that lead to a loss of precipitation information (compared to the collection of precipitation in the bucket of a weighing gauge for instance, where the information can be retrieved by subtracting total bucket weights).

The SPICE analyses focus mainly on the ability of such instruments to replace traditional precipitation gauges for the accurate measurement of solid precipitation amount. Thus, assessing the impact of such processes (as listed above) was set as a secondary objective within SPICE, which has been only partially investigated. Some aspects are presented in more detail in the IPRs (see Annex 6).

Some preliminary studies have been performed to assess other types of information delivered by these instruments, such as precipitation type and particle size/fall velocity distribution. The algorithms for deriving precipitation type and quantity are typically not shared by the manufacturers, and it is often unclear how the final parameters are derived. These algorithms require assumptions regarding particle density and fall speed. A key limitation of optical instruments is their inability to know the density of each snow particle crossing the measurement volume, which leads to uncertainty in the equivalent water content. Further, the assumption that the speed of a particle is equal to terminal fall velocity for some sensors (which don't determine the direction of movement) may lead to the incorrect classification of precipitation particles (e.g. ice instead of snow) in cases of strong winds or turbulence with high horizontal speed components, and thus to an incorrect estimation of liquid water content. Further work and enhanced collaboration with the manufacturers are needed to fully understand the operation of these instruments and their processing methods, especially with regard to solid precipitation measurements.

4.1.3.4 Assessment approach

The assessment approach for WMO-SPICE is predicated on 30-minute precipitation events as identified by the R2 reference configuration, comprising a DFAR and precipitation detector. A precipitation event is defined as a 30-minute period during which the weighing gauge in the DFAR accumulates ≥ 0.25 mm of precipitation and the precipitation detector reports the occurrence of precipitation for at least 18 minutes.

Consolidated results are provided for each NCI type (at all applicable sites) and for each test site (all applicable test sensors) in Section 4.1.3.5. Results include skill scores representing the ability of NCIs to detect precipitation relative to the reference configuration, as well as root mean square errors and catch efficiencies, which quantify the performance relative to the reference configuration in terms of reporting accumulated solid precipitation. Accumulation reports and catch efficiencies are also considered on a per-event basis (relative to the reference configuration) as a function of wind speed

for snow events. Snow events are identified as those for which the maximum reported temperature does not exceed -2°C during the event.

4.1.3.5 Results

4.1.3.5.1 *Characterization of precipitation events*

To provide context for the interpretation of results from test sensors at different sites, in different climate zones, mean precipitation event characteristics are provided for 30-minute events in all precipitation types (snow, mixed, rain) and for snow events only, in Table 4.11 and Table 4.12, respectively. The number of events is also given for both seasons of operation of the sensors (except the Hotplate at Haukeliseter, operated during the second season only). Note that differences in the environmental conditions experienced at each site do not explain all observed differences in results. Differences in instrument siting and configuration can impact instrument performance significantly, and it is difficult to separate such contributions to errors/uncertainty from those of the environment and the operation of, and internal processing specific to, different instrument types. Further, while the mean parameter values provide perspective on how precipitation events varied from site to site, discussion of differences among sites must give consideration to the variability of event conditions at each site, as reflected by the standard deviation values in Table 4.11 and Table 4.12.

The results in Table 4.11 indicate that the site showing the lowest mean wind speed during events is Sodankylä, which is consistent with its situation – a sheltered site (clearing in a forest) characterized by low wind speed. On the other end, Haukeliseter shows the highest mean wind speed during events at almost 6 m/s, consistent with its designation as a high-wind alpine site. Marshall and Weissfluhjoch both recorded similar mean wind speed conditions during precipitation events. The precipitation events at Sodankylä, in addition to showing the lowest wind speeds, were also the ones with the smallest mean accumulation and with the highest mean temperature, when considering events from all precipitation types (snow, mixed, rain). The alpine Weissfluhjoch site, situated at an altitude of 2,540 m above sea level, is the site with the lowest mean event temperature, close to -7 °C, while precipitation events at the other sites were all around -2 °C. The continental Marshall site had the largest mean event accumulation, all precipitation types included, with a value slightly larger than for the two alpine sites.

During snow events only (Table 4.12), the tendencies are similar as those for all precipitation types, with some exceptions for mean event temperature and accumulation. For mean event temperature, all the sites showed lower values, as expected during snow conditions, with Weissfluhjoch still showing the coldest mean event temperature. The highest mean temperature value is observed for Haukeliseter; however, its value of almost -5 °C is close to that observed for Sodankylä, which showed the lowest mean event temperature in all precipitation types. For mean event accumulation, Sodankylä shows the lowest value, again similar to the results for all precipitation types. In general, the relative magnitudes of mean accumulation during 30-minute precipitation events by climate zone were as follows: alpine > continental > northern boreal.

Table 4.11. Precipitation event characteristics at all NCI test sites during WMO-SPICE. Mean event parameters are provided, with uncertainty represented by the standard deviation. Results presented are for events identified by the R2 reference configuration. Red and blue values indicate, respectively, the highest and lowest parameter values among the sites.

Site	Climate zone	Mean event accumulation [mm]	Mean event temperature [°C]	Mean event wind speed [m/s]
Haukeliseter	Alpine (high wind)	0.61 ± 0.40	-2.5 ± 3.2	5.7 ± 3.5
Marshall	Continental	0.77 ± 0.63	-2.2 ± 5.3	2.9 ± 1.5
Sodankylä	Northern Boreal (low wind)	0.45 ± 0.24	-1.9 ± 4.0	1.5 ± 0.6
Weissfluhjoch	Alpine	0.71 ± 0.52	-6.7 ± 4.5	2.8 ± 2.1

Table 4.12. Snow event characteristics at all NCI test sites during WMO-SPICE. Mean event parameters are provided, with uncertainty represented by the standard deviation. Results presented are for events identified by the R2 reference configuration. Red and blue values indicate, respectively, the highest and lowest value among the sites.

Site	*Climate zone*	*Mean event accumulation [mm]*	*Mean event temperature [°C]*	*Mean event wind speed [m/s]*
Haukeliseter	Alpine (high wind)	0.59 ± 0.35	-4.7 ± 2.9	5.6 ± 3.8
Marshall	Continental	0.54 ± 0.31	-6.6 ± 3.4	3.0 ± 1.4
Sodankylä	Northern Boreal (low wind)	0.43 ± 0.22	-4.9 ± 3.0	1.6 ± 0.6
Weissfluhjoch	Alpine	0.68 ± 0.49	-8.0 ± 3.6	3.0 ± 2.2

4.1.3.5.2 *Detection of precipitation relative to reference*

The detection of precipitation by NCIs under test relative to corresponding site reference configurations was assessed using the following skill scores: probability of detection, false alarm rate, bias, and Heidke Skill Score. The test sensors, due to their high sensitivity, were considered to detect precipitation whenever they reported > 0 mm during an assessment interval (30 minutes), while the reference was considered to detect precipitation when the reference weighing gauge reported ≥ 0.25 mm accumulated precipitation and the reference precipitation detector reported ≥ 18 minutes of precipitation occuring during an assessment interval. On the other hand, test sensors were considered to have reported no precipitation when they reported 0 mm over 30 minutes, while the reference was considered to have reported no precipitation when the weighing gauge reported < 0.1 mm (to account for the noise level of reference measurements) and the precipitation detector reported 0 minutes of precipitation occurrence. There are, therefore, four potential detection scenarios: the reference and NCI both detected precipitation (YY cases); the reference detected precipitation, but the NCI did not (YN cases); the reference did not detect precipitation, but the NCI

did (NY cases); and neither the reference, nor the NCI, detected precipitation (NN cases). The number of events in each detection scenario, for each test sensor, over the duration of experiments, was used to calculate the skill scores, as detailed in Section 3.6.1.3.1.

The reference detection criteria have been selected to identify precipitation events with a high degree of confidence; however, light and/or sporadic precipitation events are not well represented. This has implications for the skill-score assessment, as a proportion of the NY cases may correspond to intervals during which the NCI responds to incident precipitation that does not meet the reference detection criteria. Indeed, the characterization of NY cases in the Instrument Performance Reports (Annex 6) indicates the occurrence of NCI responses to light precipitation events, with reference accumulation below 0.1 mm and the precipitation detector reporting zero minutes of precipitation during a 30-minute assessment interval. A detailed investigation of NCI responses to light precipitation is not considered in the present assessment, but relevant results and discussion are provided in each IPR to assist with the interpretation of skill score results, as required.

Skill scores for all test sensors, at all sites, are compiled in Figure 4.56. These scores represent the ability of the NCI to detect precipitation relative to the reference configuration, over the full range of environmental conditions and precipitation types experienced at each site, within 30-minute periods, and using the specified detection thresholds.

The POD represents the percentage of precipitation events identified by the reference that are also identified by the NCI, with an ideal value of 100%. Sensors from all sites, except Haukeliseter, have very high POD values (between 96 to 100%). Haukeliseter shows POD values around 75%. The high winds at this site (Hotplate) and a maintenance issue (PWS100) likely account for these differences (see the corresponding IPRs). However, the generally high POD values relative to the reference for all NCI types indicate their high sensitivity for the detection of precipitation.

The FAR, which represents the total number of precipitation events detected by the test instrument that are not detected by the reference (ideal value of 0%), is lower than 10% for almost all sensors, except for those located at Sodankylä and the Thies LPM at Marshall, which showed FAR values of around 60%. Among these six instruments with high FAR values, two have experienced issues with the identification of precipitation (details are given in the corresponding IPRs), either due to the probable creation of heat plumes (Hotplate in Sodankylä), or because of a potential internal issue of the sensor when high temperature and relative humidity gradients occur during spring seasons (Parsivel^2 at Sodankylä). The other four instruments showed greater sensitivity to light precipitation than the reference itself (especially greater than the reference precipitation detector). Indeed, in these two sites, Marshall and Sodankylä, the precipitation detector for the second season was changed from a capacitive sensor to an optical sensor, according to SPICE IOC requirements. This new reference precipitation detector showed less sensitivity than both the instrument under test and the "old" capacitive sensor during light precipitation (see corresponding IPR). The optical precipitation detector missed some light precipitation events that were reported by the NCI under test. This led to a certain number of NY events that were apparent light precipitation events. The number of NY events mainly influences the FAR score, but also the bias and HSS scores (see definition of skill scores in Section 3.6.1.3.1), as shown in Figure 4.56. The skill scores recalculated using the capacitive sensor for both seasons (see details in corresponding IPRs) are improved by 20% for the FAR, 80% to 90% for the bias, and 20% for the HSS score. These results indicate that most of

the NCIs tested within SPICE behave similarly well with regard to their ability to detect the presence of precipitation, especially for light intensities.

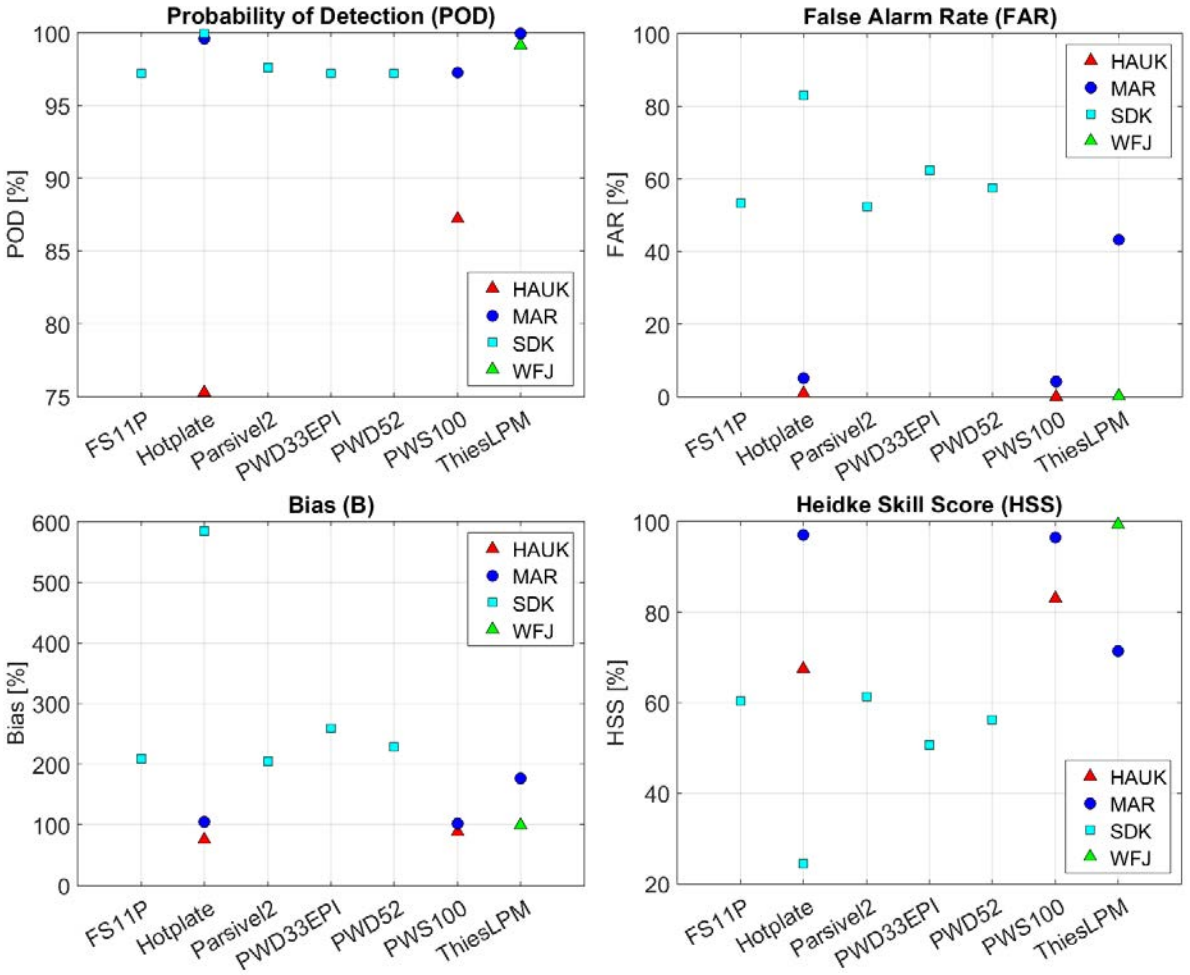


Figure 4.56. Skill scores representing the performance of each NCI tested relative to the reference with respect to the detection of precipitation. Scores are calculated using 30-minute events in all precipitation types. The legends indicate the different sites: "HAUK" for Haukeliseter, "MAR" for Marshal, "SDK" for Sodankylä, and "WFJ" for Weissfluhjoch. Sensors at sites in alpine, continental, and northern boreal climate zones are represented by triangles, circles, and squares, respectively.

4.1.3.5.3 *Reporting accumulated precipitation relative to reference*

The catch efficiency, or catch ratio, is the ratio of accumulated precipitation reported by a SUT relative to that reported by the reference configuration over a specified time interval. As noted earlier, the influence of wind speed on catch efficiency for solid precipitation is not as straightforward for NCIs as for conventional catchment-type gauges, due to the different shapes and measurement principles of NCIs. To better understand wind speed impacts on solid precipitation measurements by NCIs, the catch efficiency of solid precipitation is assessed as a function of the mean wind speed for 30-minute assessment intervals in Section 4.1.3.5.3.1. The assessment is presented first by NCI model, and then by site.

The catch efficiency is a useful indicator of NCI performance relative to the reference, but does not provide information about the magnitude of accumulated precipitation reported by each sensor. The root mean square error, however, represents the absolute difference in reported accumulation between the test sensor and the reference for each 30-minute interval. An assessment of RMSE results for solid precipitation is provided for all NCIs tested, at all sites, in Section 4.1.3.5.3.2.

Finally, the overall catch ratios for solid precipitation are presented for each test NCI, at each site, in Section 4.1.3.5.3.3 .

4.1.3.5.3.1 Wind effects on catch efficiency

The catch efficiency of each NCI tested (at all applicable sites) and at each site (with multiple tested NCI) is assessed as a function of the mean wind speed for 30-minute assessment intervals in Sections 4.1.3.5.3.1.1 (by sensor type) and 4.1.3.5.3.1.2 (by site), respectively. This assessment is limited to snow events during which the maximum temperature did not exceed -2 °C over a given 30-minute assessment interval. A similar assessment is presented for all precipitation types (rain, mixed, snow) in the Instrument Performance Reports (Annex 6).

4.1.3.5.3.1.1 Results by sensor type

Box and whisker plots of catch efficiency as a function of mean wind speed are plotted for each NCI tested on multiple sites (where available) in Figure 4.57 to Figure 4.59. The results for instruments tested at only one site, namely the Parsivel^2, FS11P, PWD33EPI and PWD52, are provided in Figure 4.62. The boxes in each plot represent the range of values between the 25th percentile (lower quartile; bottom of box) and 75th percentile (upper quartile; top of box) referred to as the interquartile range. The median value is indicated by the horizontal line across the box. The whiskers below/above the box indicate the lowest/highest values. Outlying points are indicated by markers above or below the whiskers.

In each plot, results are presented in 1 m/s bins, with sensors at different sites represented by different colors. In general, it is difficult to make attributions for observed differences in the catch efficiency vs. wind speed relationship for the same sensor at multiple sites. These differences are likely the combined result of differences in environmental conditions (mean event accumulation, temperature, predominant ice crystal type for solid precipitation) and differences in sensor configuration and siting. Accordingly, the identification and diagnosis of differences in results between or among sites is not a focal point of this work; rather, the focus is on identifying general trends for sensors of a given instrument type.

The Hotplate (Figure 4.57) shows median catch efficiency values between about 0.9 and 1.3 for mean wind speeds within 7 m/s at all three sites at which it was tested (Haukeliseter, Marshall, Sodankylä). For mean wind speeds above 7 m/s, the median catch efficiency values vary more broadly, within about 0.5 and 1.3, and show larger spread of values (dispersion) in individual wind speed bins. Aside from the observed increase in dispersion, the median catch efficiency values for the Hotplate do not show any clear trend with increasing mean wind speed. This can likely be explained by the fact that the measurement of the bottom plate takes the wind speed into account when deriving the precipitation amount.

The median catch efficiencies for the PWS100 sensors tested at Haukeliseter and Marshall both show apparent decreases, from about 1.3 to 0.5, with increasing mean wind speed up to 5 m/s (Figure 4.58). The median catch efficiencies increase to about 0.8 to 0.9 at higher mean wind speeds up to 7

m/s. The dispersion of catch efficiency values is low over this mean wind speed range. For mean wind speeds above 7 m/s, the dispersion of catch efficiency values within bins increases and median catch efficiency values vary more broadly, from 0.4 to 3.4, with no clear tendency.

The median catch efficiency for the Thies LPM sensor at Marshall shows a marked decrease from 0.8 to 0.3 with increasing mean wind speed up to 5 m/s (Figure 4.59). The sensor tested at Weissfluhjoch shows a more gradual decrease from about 0.55 to 0.3 over the same mean wind speed range. For sensors at both sites, the median catch efficiency stabilizes at around 0.3 to 0.4 for mean wind speeds above 4 m/s. As discussed in the corresponding IPR (Annex 6), the fact that the shield was very close to the sensor (see Figure 4.51) and possibly shadowed the measurement area can likely explain the undercatch increasing with higher mean wind speed. The dispersion of events in wind-speed bins shows an opposing trend as compared to the other NCI types, decreasing with mean wind speed at both sites and remaining low for mean wind speeds as high as 9 m/s.

The remaining sensors were tested only at Sodankylä (Figure 4.62). The PWD33 EPI, PWD52 and FS11P are three Vaisala sensors with similar measurement principles (see Section 4.1.3.2.1) compared to the other NCI types. From Figure 4.62, it can be noted that the three sensors show similar catch efficiency performance related to wind speed; however, Sodankylä was the site with the lowest mean event wind speeds, not exceeding 4 m/s, precluding conclusive statements regarding sensor behavior in windier conditions. The median catch efficiency values generally vary between about 0.75 and 1.25 for mean wind speeds up to 3 m/s, with values as high as 1.5 for the Hotplate at mean wind speeds < 1 m/s. There is an apparent tendency for the median catch efficiency to decrease at mean wind speeds greater than 3 m/s; however, the three events observed under these conditions prevent one from drawing any robust conclusions. The dispersion of events within mean wind speed generally increases with increasing wind speed up to 3 m/s for each sensor type.

The Parsivel^2 tested at Sodankylä (Figure 4.62) employs a different measurement principle (laser disdrometer, see Section 4.1.3.2.2) relative to the other sensors tested at the site, and shows increasing median catch efficiency from about 0.9 to 1.8 with increasing mean wind speed up to 3 m/s.

4.1.3.5.3.1.2 Results by site

By comparing catch efficiency results for NCIs installed at the same site, it is possible to compare the performance of different sensor types under the same conditions, recognizing that some differences in the spatial distribution of precipitation and/or the configuration of specific instruments may still exist. Results are presented in Figure 4.60 to Figure 4.62 for Haukeliseter, Marshall, and Sodankylä, respectively, each with multiple NCIs being tested. Results for Weissfluhjoch, which only hosted the Thies LPM, are available in Figure 4.59.

The NCIs tested at Haukeliseter employed different measurement principles and showed different trends with respect to median catch efficiency (Figure 4.60). The median catch efficiency for the PWS100 decreased with increasing wind speed up to 5 m/s and then increased for mean wind speeds up to 7 m/s. The values for the Hotplate fluctuated within about 1 and 1.3 for mean wind speeds up to 7 m/s. Above 7 m/s, there were no apparent catch efficiency trends for either sensor, and the dispersion of catch efficiency values within individual bins increased. Given that the Hotplate sensor is designed to account for wind speed effects, differences in performance are to be expected; these may be exacerbated by the characteristically high wind speeds experienced at Haukeliseter.

Three different NCIs were tested at Marshall. The median catch efficiency vs. wind speed results are compared in Figure 4.61 and highlight the differences between shielded and unshielded sensors. The median catch efficiency of the shielded Thies LPM decreases with increasing mean wind speed up to about 5 m/s and levels off at values around 0.25, whereas the values for the PWS100 and Hotplate generally remain within the 0.75 to 1.25 range over the full range of mean wind speeds tested.

Sodankylä hosted five different NCIs under test (Figure 4.62). It can be noted that, even in low wind-speed conditions (< 4 m/s during the whole campaign), there are differences in the median catch efficiency trends among the sensors, resulting from differences in the measurement principles (see Table 4.10). The Parsivel^2 (disdrometer) shows increasing median catch efficiency with increasing mean wind speed, the Hotplate (heat/mass transfer) shows an opposing trend, and the Vaisala present weather sensors show median catch efficiency values within about 0.57 and 1 over the wind speed range tested.

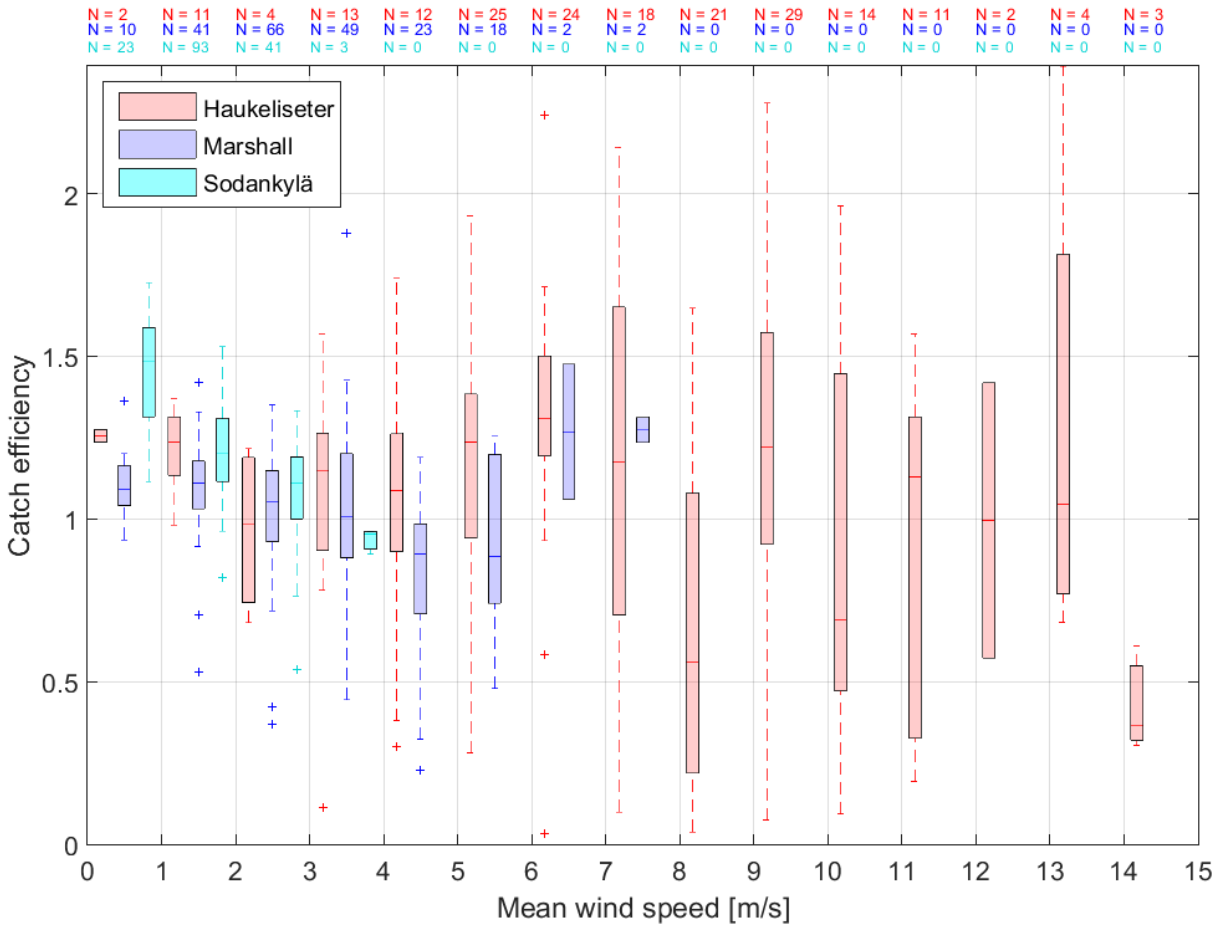


Figure 4.57. Box and whisker plot of catch efficiency as a function of mean wind speed for the Hotplate sensors relative to site reference configurations for 30-min snow events (T_{max} ≤ -2 °C). The number of events in each wind-speed bin, and for each test sensor, is indicated on top of the plot.

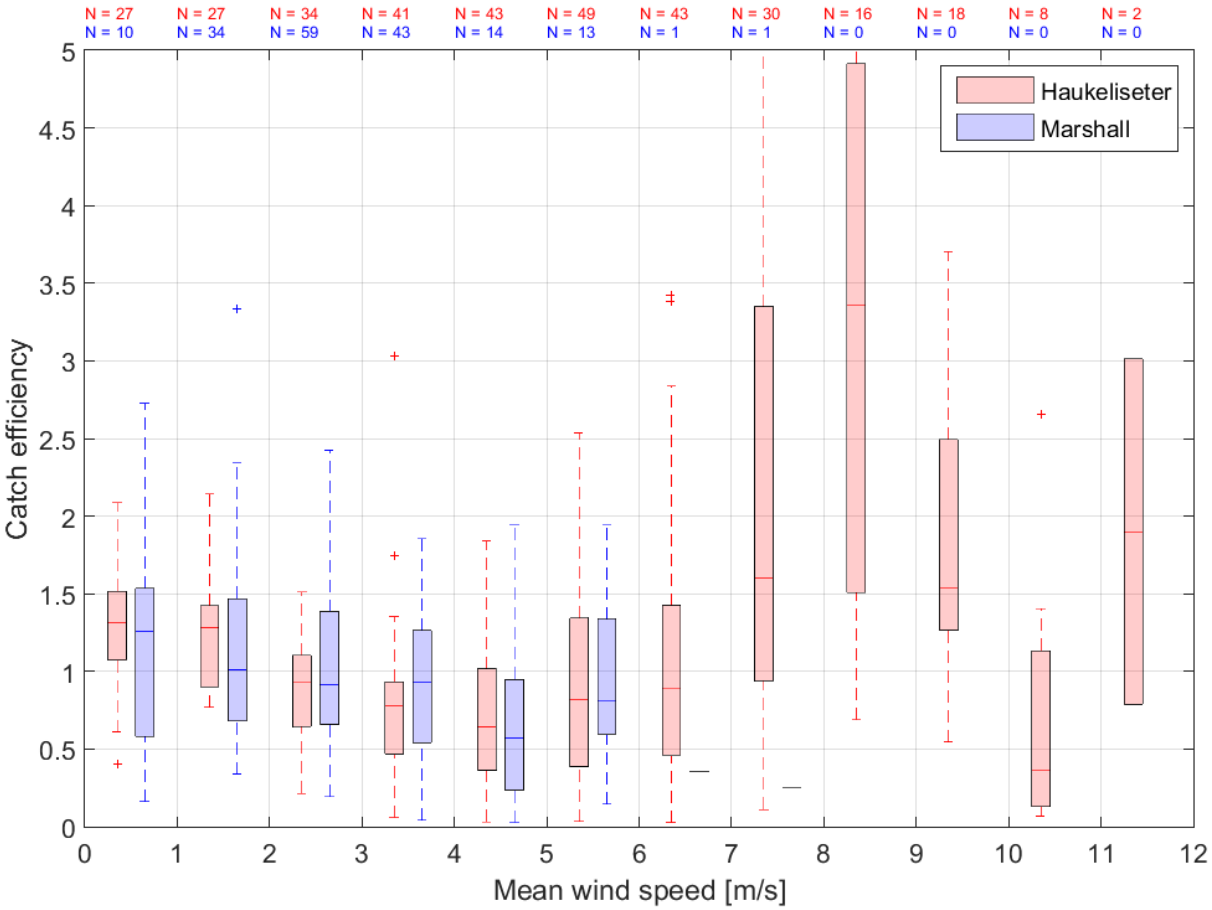


Figure 4.58. Box and whisker plot of catch efficiency as a function of mean wind speed for the PWS100 sensors relative to site reference configurations for 30-min snow events (T_{max} ≤ -2 °C). The number of events in each wind-speed bin, and for each test sensor, is indicated on top of the plot.

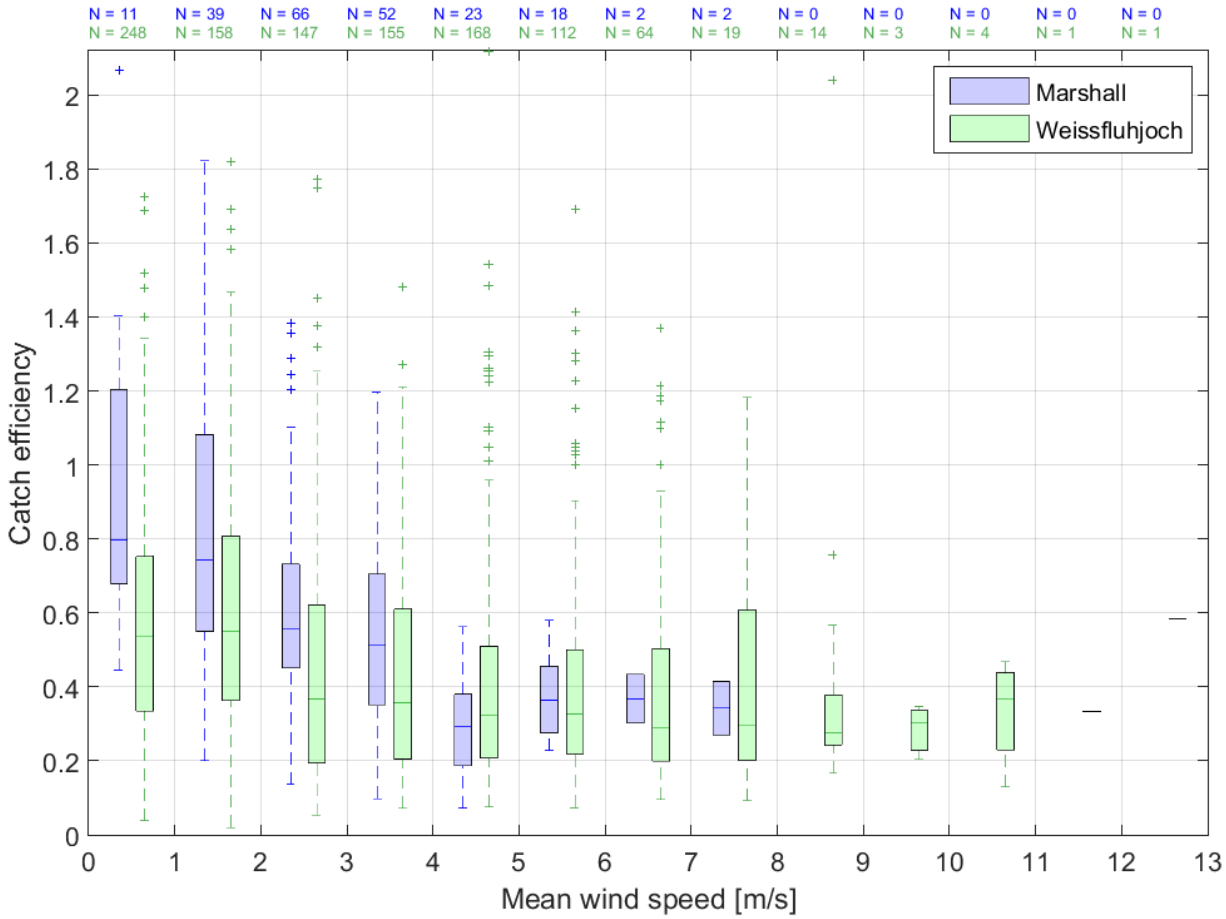


Figure 4.59. Box and whisker plot of catch efficiency as a function of mean wind speed for the shielded Thies LPM sensors relative to site reference configurations for 30-min snow events ($T_{max} \leq$ -2 °C). The number of events in each wind-speed bin, and for each test sensor, is indicated on top of the plot.

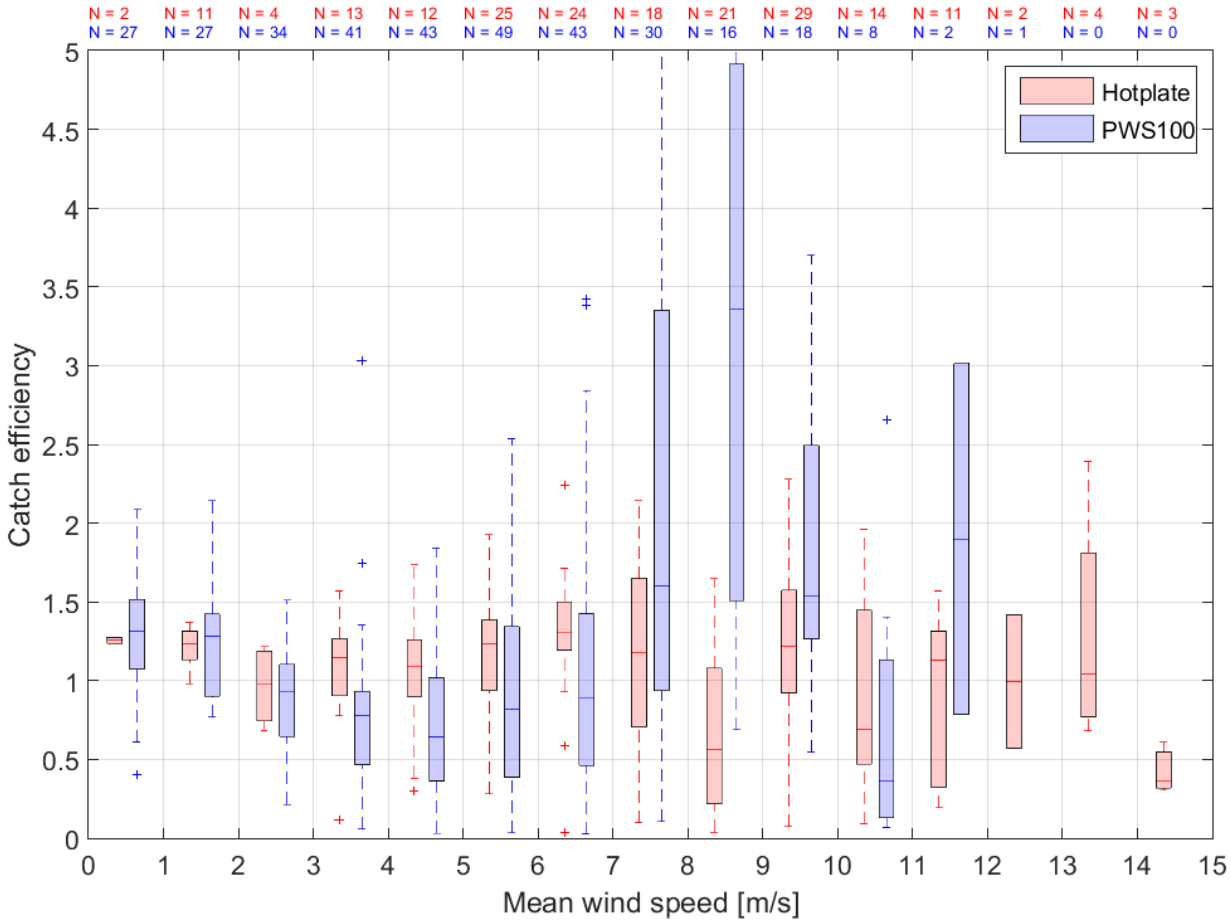


Figure 4.60. Box and whisker plot of catch efficiency as a function of mean wind speed for NCIs under test relative to site reference configuration at Haukeliseter for 30-minute snow events (T_{max} ≤ -2 °C). The number of events in each wind-speed bin, and for each gauge under test, is indicated on top of the plot.

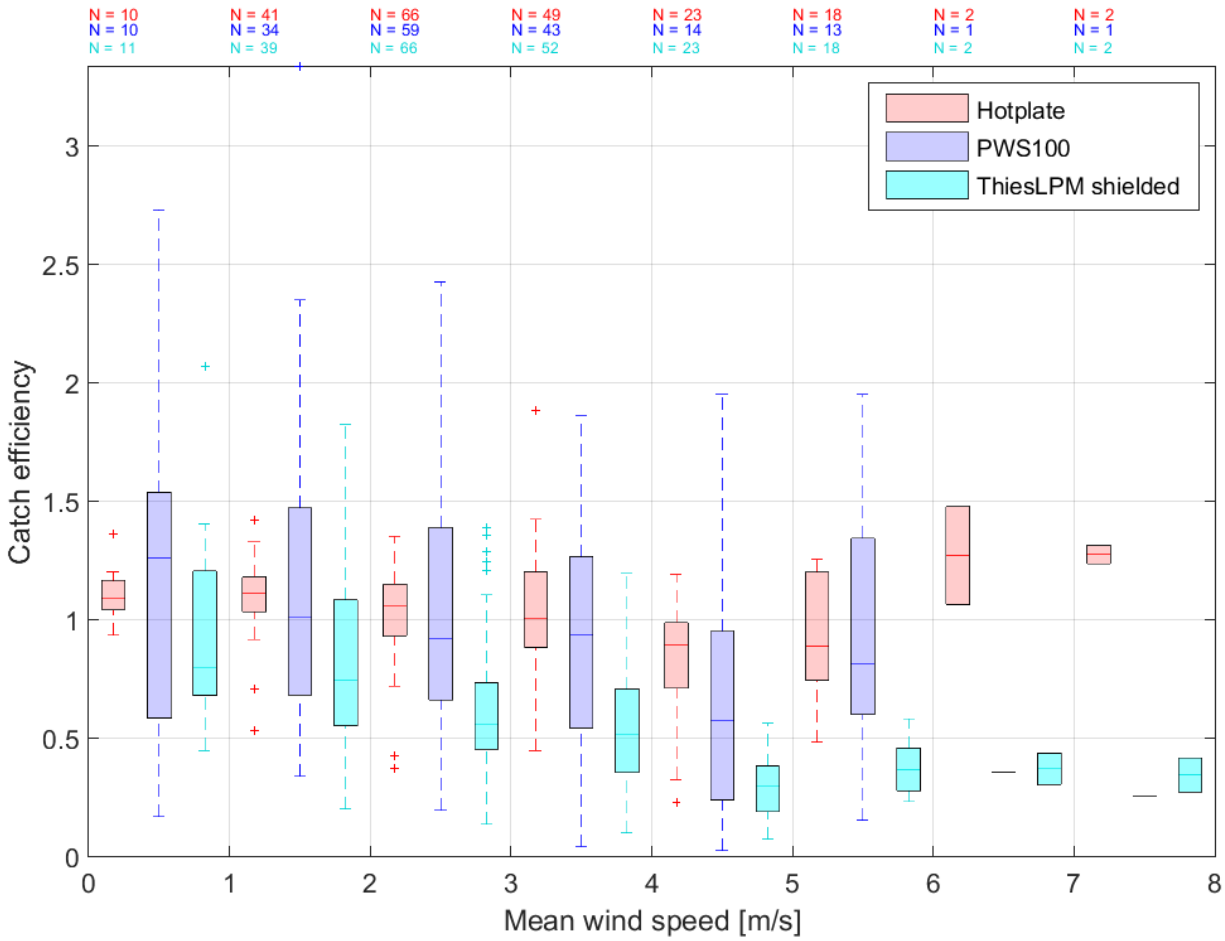


Figure 4.61. Box and whisker plot of catch efficiency as a function of mean wind speed for NCIs under test relative to site reference configuration at Marshall for 30-minute snow events (T_{max} ≤ -2 °C). The number of events in each wind-speed bin, and for each gauge under test, is indicated on top of the plot.

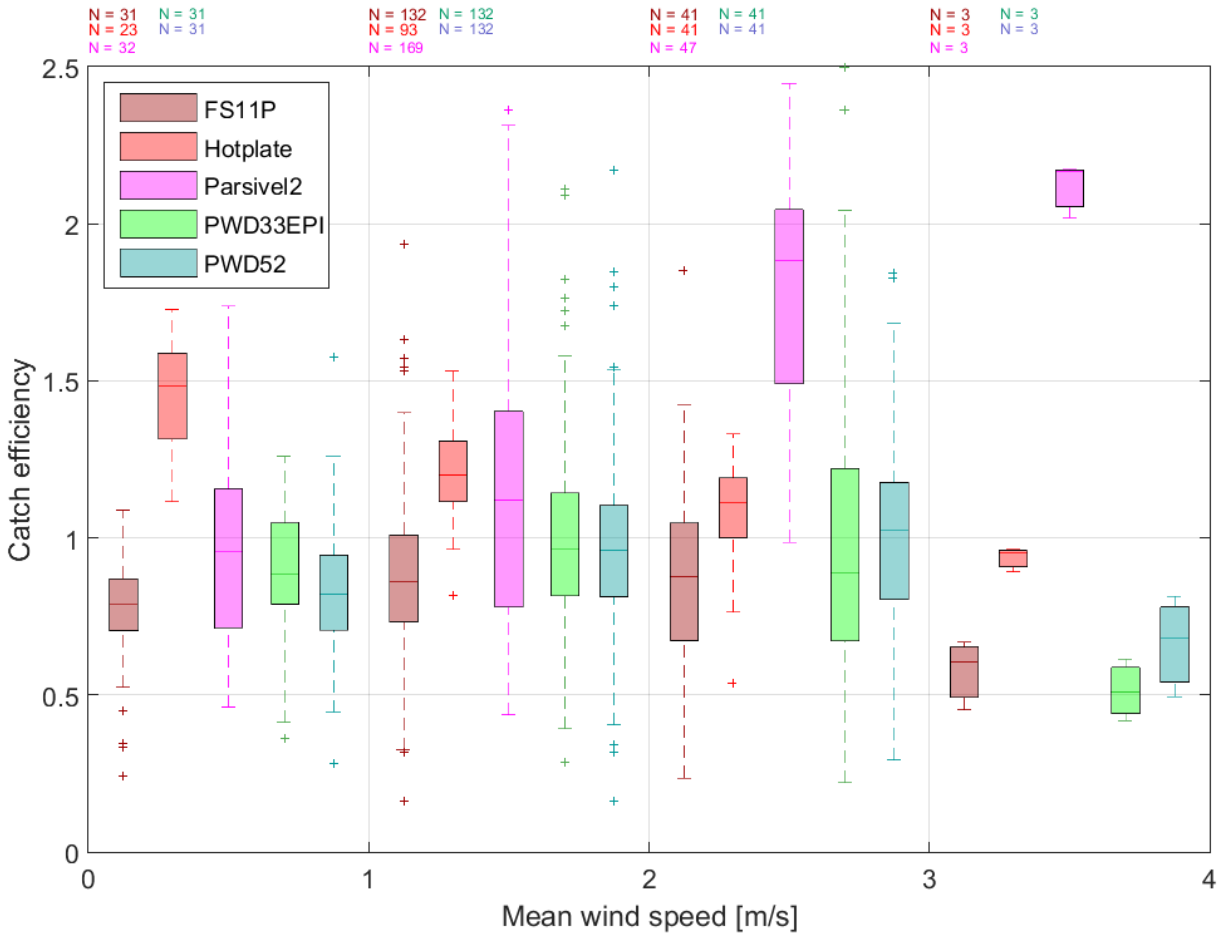


Figure 4.62. Box and whisker plot of catch efficiency as a function of mean wind speed for NCIs under test relative to site reference configuration at Sodankylä for 30-minute snow events (T_{max} ≤ -2 °C). The number of events in each wind-speed bin, and for each gauge under test, is indicated on top of the plot.

4.1.3.5.3.2 Root mean square error

The root mean square error was calculated for 30-minute events during which the NCI under test and reference both reported precipitation (Section B). The RMSE for each instrument, at each site, is shown in Figure 4.63 for solid precipitation events.

Instruments tested in Haukeliseter and Weissfluhjoch, both alpine sites (triangles in Figure 4.63), show the highest RMSE, ranging from 0.3 to 0.8 mm, slightly higher than the instruments tested at Marshall under continental climate (circles), which ranged from 0.1 to 0.3 mm. The instruments tested under a northern boreal climate at Sodankylä (squares) show the lowest RMSE, with values ranging between 0.1 and 0.24 mm. This indicates that the wind probably accounts for the larger part of the RMSE, since higher wind speed is expected to increase the variability among the measurements by accelerating the snowflakes or giving them a strong horizontal trajectory which is wrongly interpreted as the actual fall speed. This leads to an over- or underestimation of the liquid water content by the sensors internal algorithms.

Users should be aware of the potential error magnitude on solid precipitation measurement, which will be proportional to the mean wind values experienced at their site.

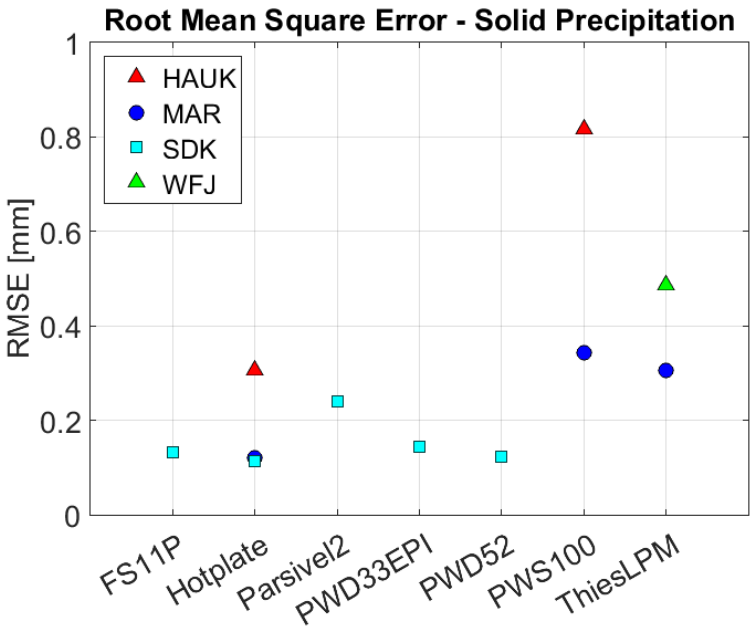


Figure 4.63. Root mean square error calculated for 30-minute YY solid precipitation events. Sensors at sites in alpine, continental, and northern boreal climate zones are represented by triangles, circles, and squares, respectively.

4.1.3.5.3.3 Overall catch efficiency

The long-term performance of each NCI tested, with respect to its ability to report solid accumulated precipitation relative to the reference, was assessed using the overall catch efficiency (Figure 4.64). The overall catch efficiency was calculated using the total accumulation reported by the test and reference instruments during all 30-minute intervals in which both detected precipitation (YY cases). Hence, the assessment is not predicated on detection within a given 30-minute interval. The overall catch ratio is, therefore, less reflective of performance over operational time scales (e.g. tens of minutes or 1 hour) and more reflective of longer-term or seasonal performance.

With the exception of the shielded Thies LPM, all NCIs tested show overall catch efficiencies ranging from 0.85 to 1.3. The Thies LPM, as tested in SPICE, underestimates solid precipitation relative to the other sensors tested, as demonstrated by the lower overall catch efficiencies of 0.5 to 0.6. As mentioned above, the shield may have negatively impacted the instrument's performance by shadowing the measurement area.

Considering the results for different climate regimes in Figure 4.64, and excluding the results for the shielded Thies LPM sensors, it appears that sensors in alpine climates tend to overestimate precipitation amount relative to the reference configuration (catch efficiency > 1), while those in continental climates report amounts similar to the reference (catch efficiency ≈ 1). A wider variety of sensors was tested at the northern boreal site of Sodankylä, and the results vary from over-estimating (Hotplate, Parsivel^2) to under-estimating (FS11P) precipitation amount relative to the sensor, with two sensors (PWD33EPI and PWD52) reporting similar amounts as the reference configuration.

Overall, and without considering the particular case of the shielded Thies LPM discussed earlier, the non-catchment instruments show consistent performance for the derivation of solid precipitation amounts over longer timescales (months to seasons), with errors in the total accumulation of +/- 15-20%. This conclusion is predicated on robust operational functioning of the sensor, i.e. with minimal or no interruption, which was generally the case during SPICE (more details can be found in the IPRs). Any precipitation falling during instrument outages will not be measured and reported, impacting accumulation totals (unlike weighing gauges, for which precipitation is still collected during instrument outages).

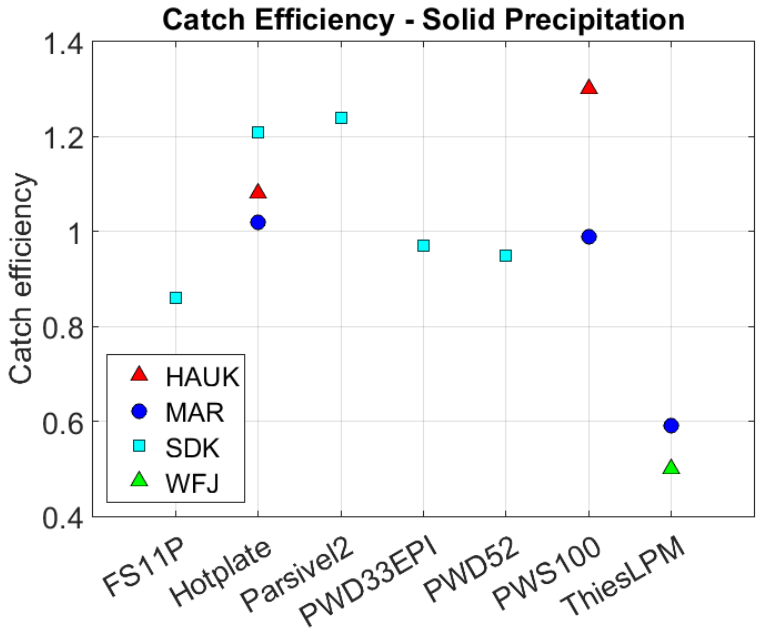


Figure 4.64. Overall catch efficiency calculated for all 30-minute YY solid precipitation events. Sensors at sites in alpine, continental, and northern boreal climate zones are represented by triangles, circles, and squares, respectively. Note that the configuration for the Thies LPM used a shield provided by the manufacturer.

4.1.3.5.3.4 Assumption of snowflake density

Excluding the Hotplate, non-catchment instruments do not measure the mass of snowflakes, and are therefore unable to determine their density. It is generally assumed that 1 cm of new snow corresponds to 1 mm of water content; this is referred to as the 10:1 rule. This assumption is generally true, on average (i.e. over long time periods, consistent with the results presented in Figure 4.64). However, on an event basis (e.g. 30-minute intervals), there is significant variability in particle density.

Figure 4.65 shows the catch efficiency reported by NCIs with different measurement principles (each computed relative to the corresponding site reference configuration) as a function of mean wind speed during 30-minute events from both seasons. The marked variability of results is attributed, in large part, to differences in water content estimation. As an example, aggregates (low-density snowflakes) will be seen as large-diameter particles by disdrometers, which, with the 10:1 ratio rule,

will overestimate the water content (catch efficiency > 1). The opposite can also happen, with small and dense particles (e.g. rimed snowflakes, hail), leading to an underestimation of the water content (catch efficiency < 1). All situations in between these two extreme examples for which the 10:1 rule would not exactly apply are reflected in the dispersion of the catch efficiency values, ranging from 0.1 to 2, even under identical wind conditions.

These results suggest that for NCIs, the main issue impacting the ability of the instruments to report a reliable and accurate precipitation amount is not wind speed, as for tipping bucket and weighing gauges, but the assumption of snowflake density. In this context, it is important to note that the algorithms implemented by the different manufacturers and associated snow density estimates/assumptions are proprietary information.

The Hotplate results are also presented in Figure 4.65. This instrument measures the water content of each snowflake directly (see section 4.1.3.2.3). In this case, the observed scatter in the catch efficiency results is expected to be due to wind rather than to incorrect density estimation. The internal algorithm taking the wind speed into account for the derivation of the water content seems to produce higher scatter at wind speeds > 8 m/s, as explained in the corresponding IPR (Annex 6).

Overall, it can be concluded that NCIs are less suitable for the derivation of solid precipitation amounts over near real-time periods (e.g. 30-minute intervals) relative to longer time intervals, due to differences between the derived/assumed densities of particles and their actual densities. The exception appears to be the Hotplate, which measures particle density; however, the Hotplate results are impacted more siginificantly by wind speed.

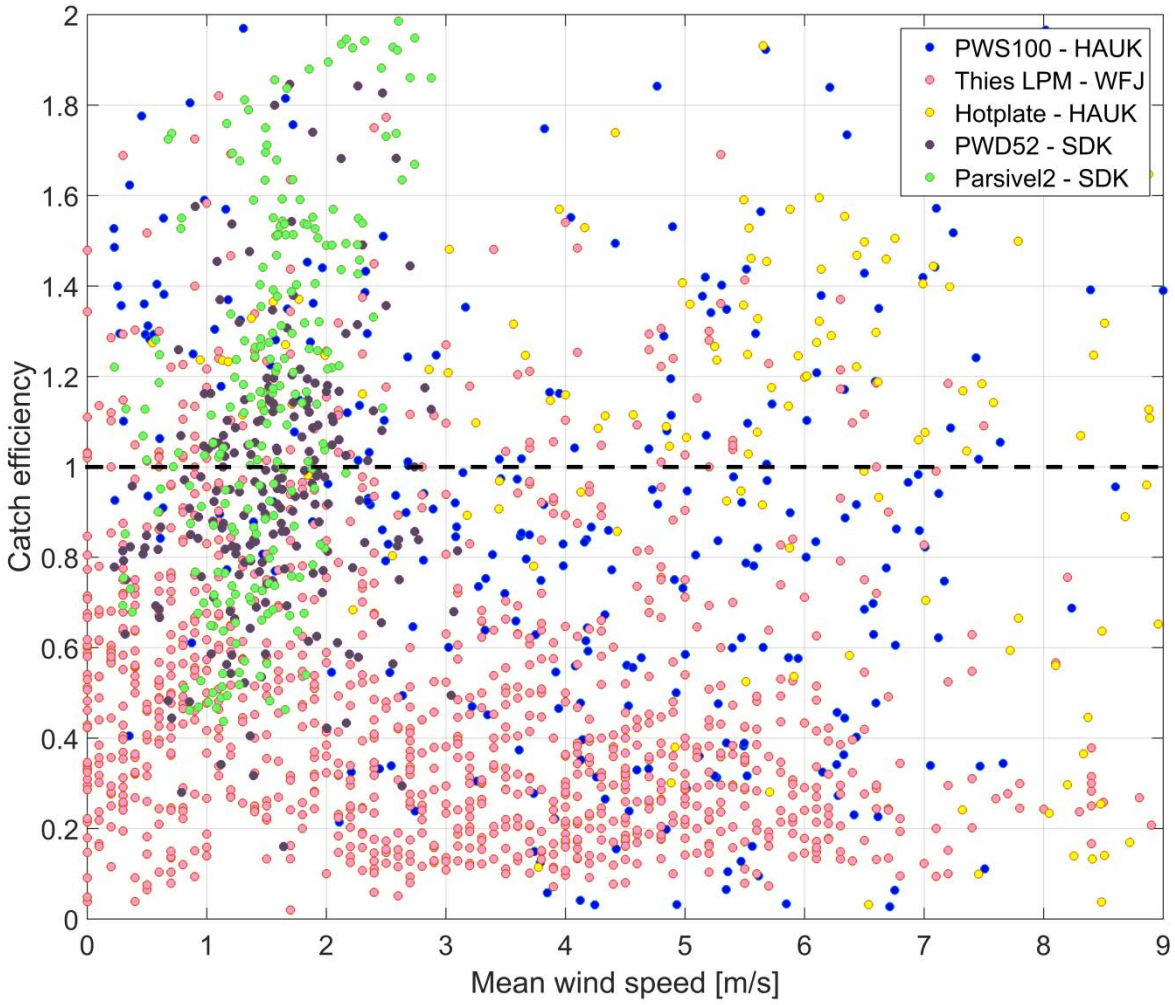


Figure 4.65. Scatter plot of catch efficiency as a function of mean wind speed for NCIs with different measurement principles tested in WMO-SPICE. Catch efficiency values are computed relative to the reference configuration at a given site. Each point represents a 30-min YY event in snow conditions (T_{max} ≤ -2 °C). The dashed black line indicates the ideal case, in which the NCI and reference report the same accumulation amount (CE = 1).

4.1.3.6 Recommendations

Drawing upon the collective experience of WMO-SPICE site teams with the NCIs under test, and the results and interpretation presented in this section (for all tested sensors) and in the Instrument Performance Reports (for each specific sensor under test; see Annex 6), a set of recommendations has been developed to guide individuals or organizations with interest in employing NCIs for the measurement of solid precipitation. These recommendations are grouped into four categories – sensor selection, sensor configuration and siting, adjustment functions and required ancillary measurements, and operational use – and are presented in the following sections.

4.1.3.7 Sensor type selection

The selection of a non-catchment-type instrument should be based on user needs in terms of use of the data, as well as on the site-specific climatic conditions. The selection will also depend on the infrastructure and resources available on site. As a matter of course, the range of expected

conditions for a site, specifically temperature and precipitation intensity, should fall within the limits stated by the sensor's manufacturer.

The non-catchment-type instruments tested in SPICE covered a wide range of different measurement principles. Present weather sensors, disdrometers and evaporative plates were all showing good results in their ability to detect precipitation (for all precipitation types) compared to the reference (see corresponding sections in the IPRs in Annex 6). Their high sensitivity allows for the reporting of very light precipitation events that were not seen by the reference gauge. They are therefore all recommended for the detection of precipitation, especially if the required accuracy is such that the light events must be covered (keeping in mind that the precipitation amount related to these light events may not necessarily be reliable).

For most NCIs tested, the catch efficiency under snow conditions over long periods of time, e.g. a season, showed similar results with +/- 15-20% differences with the reference accumulation. On the other hand, the high scatter observed on an event basis (30-minutes in this study) for present weather sensors and disdrometers indicate that these sensors are not well-suited for the derivation of reliable solid precipitation amounts over near real-time periods. The evaoporator plate, however, is able to measure solid precipitation in real-time for wind speeds up to 8 m/s (Section 4.1.3.5.3.1.1). For disdrometers and present weather sensors, the measurement variability is related to the sensor's internal algorithm and its assumptions of snow density.

Based on the above, the measurement of precipitation during high winds (> 8 m/s) is a challenge for nearly all current instruments.

Based on this analysis, wind-induced errors in catch efficiency are less relevant for NCIs than for traditional catchment-type gauges (tipping bucket and weighing gauges). Some potential trends could be found for the Thies LPM (suspected to be due to its shield) and the Parsivel^2 (Section 4.1.3.5.3.1.1), but these are still to be confirmed, for other configurations and windier sites, respectively. It was, however, shown that more variability in catch efficiency results for NCIs are observed at higher wind speeds.

The evaporative plate reported a significant number of false precipitation events under very low wind conditions (of the order of 1 m/s). This false alarms are attributed to a heat plume effect (more details in the corresponding IPR). The PWD33 EPI, Parsivel^2, and the PWS100 also experienced specific issues leading to false precipitation reports and/or amounts (more details in the IPRs). These reflect the performance of one instrument at one site, and are not considered to be general trends.

The ability of the NCIs to detect and report precipitation type has not been assessed within this project, nor has the information coming from the raw data (particle size and fall velocity distribution), as they were beyond the scope of SPICE. Nevertheless, the under- or overcatch observed for optical sensors can likely be traced to an erroneous precipitation type determination, leading to a faulty liquid water content estimation, as was believed to be the case for the PWD33 EPI sensor (see corresponding IPR). The internal algorithms for deriving accumulation, intensity, and precipitation type are based on the raw data matrix (particle size/fall velocity distribution), and could also use an internal measurement of air temperature. The behavior of these instruments would be more understandable if the algorithm (or at least the physical assumptions made) was shared with the scientific community.

4.1.3.7.1 *Sensor configuration and siting*

The only NCI configuration tested with a shield was the Thies LPM sensor, which showed significant undercatch, increasing with increasing wind speed, consistently on the two test sites (Marshall and Weissfluhjoch). This shield was suspected to shadow the measurement area of the sensor, especially for solid precipitation which can fall with a higher incident angle, due to wind.

It has been shown that the impact of wind on NCI performance is lower than for conventional precipitation gauges. This is expected, as the measurement principles and physical shapes of NCIs are different. A wind direction impact was observed in some instances (see wind roses in IPRs); however, no robust conclusions could be drawn, as it was observed for one unit at one site and it could either be due to the sensor orientation, or to the siting of the sensor on site. An extensive study, specifically focused on the orientation of NCIs, e.g. two collocated sensors with different orientations, would allow for stronger conclusions to be drawn with respect to wind direction impacts (not done in SPICE).

Wind affects the terminal fall velocity of snowflakes, which is used for the derivation of precipitation amount. The shield tested along with the Thies LPM produced significant undercatch. Nevertheless, benefits from a shield around the sensor should be further investigated (not assessed in SPICE).

4.1.3.7.2 *Adjustment functions and required ancillary measurements*

As shown in the analysis above, no clear dependency of the catch efficiency on wind speed (as is the case for tipping bucket and weighing gauges) was observed for unshielded test configurations. Consequently, no adjustment functions could be derived for NCIs. Instead, more work is required with respect to the determination of particle density and corresponding estimation of particle liquid content.

4.1.3.7.3 *Operational use*

The infrastructure at a site must ensure sufficient power for the operation of the sensor. The evaporative plate, for instance, requires particularly high power consumption to maintain the high temperature of the plates.

Sites with low wind conditions may experience snow accumulation on the device structure, as observed for some disdrometers and present weather sensors at Sodankylä (more details in the IPRs). Appropriate sensor heating, design of the mounting structure, or site visits to remove the accumulated snow may be required. The heat released by the evaporative plate will prevent snow accumulation on the sensor. For sites characterized by high wind speed, the snow accumulation might be naturally removed from the structure.

Overall, the experience reported by site managers concerning NCI operation was positive. All sensors were found to be easy to install and run, and simple to maintain (specific advantages and drawbacks can be found in the IPRs). They were generally considered to be operationally reliable, with no major breaks in the data. They require minimal, if any, maintenance. The latter consists mainly of a simple cleaning of the lenses or the measurement surfaces every few months.

4.1.4 Snow on the ground

Authors: Craig Smith, Samuel Morin, Anna Kontu

4.1.4.1 Summary of Instrument Performance Reports

The objectives of the snow-on-the-ground IPRs were to provide a summary of the technical capabilities of the sensor, to specify how and where the sensor was installed for the intercomparison, to summarize the data quality metrics, and to describe to the potential user the experiences and recommendations from the SPICE community on best practices for using the instruments.

IPRs are available for the following snow-depth sensors:

- Campbell Scientific SR50A(TH)
- Felix Technologies SL300
- Lufft/Jenoptik SHM30
- Sommer USH-8
- Dimetix FLS-CH 10

IPRs are available for the following snow-water-equivalent sensors:

- Campbell Scientific CS725
- Sommer SSG1000

The reports show the intercomparisons of sensors under test from Sodankylä (SR50ATH, SHM30, SL300, USH-8, CS725, and SSG1000), CARE (SR50AT, SHM30, SL300 and USH-8), Col de Porte (SHM30, SR50ATH, FLS-CH 10), Caribou Creek (CS725) and Weissfluhjoch (SHM30). With some exceptions, the data included in the intercomparisons and summarized in the IPRs were collected over the 2013/14 and 2014/15 winter seasons.

The methodology for the instrument intercomparisons shown in the IPRs is outlined in Section 3.6.2 and a discussion of the SoG references appears in Section 0. It was decided that, for intercomparing the SUT with a reference, the snow-depth SUT be compared to both the manual reference (which was either visual or photographed observations of snow stakes, generally observed daily) and the automated sensor reference (a mean of all snow-depth sensors, either at the site, or in the case of CARE, at each pedestal grouping, every 1-minute). For SWE, the reference was a bulk-density measurement, typically made every two weeks at a location in proximity to the SUT. It was generally observed from all SPICE sites that the spatial variability of snow depth or SWE was responsible for a large fraction of the difference between sensors and references, even if all possible efforts were made to reduce this effect as much as possible. In some cases, the distances between the manual reference measurement and the SUT were too great given the spatial variability at the site (e.g. Col de Porte) resulting in offsets between the manual reference and the SUT. There are also instances where experiment design has clearly, although inadvertently, impacted the intercomparison. This was, for example, the case at CARE, where mounding was regularly observed under the targets, most likely due to the snow stakes at the corners of each of those targets.

The following statements can be made regarding the SUT evaluation shown in the IPRs:

- All snow-depth instruments performed well compared to the references, considering spatial variability and the known bias of manual measurements. The spatial variability of snow depth at the intercomparison sites is examined in Section 4.1.4.2 and should be considered when intercomparing the SUT and reference statistics.

- The SL300 snow-depth sensor had reliability issues at Sodankylä that may be site dependent, resulting from the relatively harsh conditions at this site. The same instrument functioned satisfactorily at CARE, where minimum temperatures are higher during the winter season (see Table 2.2).
- The SWE sensors showed mixed results, with certain advantages and disadvantages for each of the two measurement principles (the SSG1000 load cell vs. the CS725 passive gamma sensors).
- The SSG1000 SWE sensor performed well compared to the reference, but had reliability issues related to moisture affecting the electronics during snow melt, which needed to be rectified by both the site manager and the instrument manufacturer. This is addressed further in Section 4.1.4.4. Some bridging events were likely at Sodankylä, but the snow conditions were not conducive to these occurrences.
- The CS725 SWE sensor showed a negative bias compared to the reference, which is likely related to a combination of the soil type at the intercomparison sites, changes in soil moisture after calibration, the measurement principle, and difficulties in making accurate manual measurements in difficult conditions. This intercomparison is summarized in greater detail in Section 4.1.4.4.
- Some instruments, such as the USH-8, have relatively large horizontal surfaces that can collect snow during heavier snowfall events at low wind speeds. This is explored further in Section 4.2.6.1. Although there was no evidence during SPICE of negative impacts on the instrument's ability to make an accurate measurement, there is still potential for this to occur. Snow accumulated on the sensor occasionally falls onto the target area, but the impact on the measured snow depth is undetermined. The problem can be exacerbated by a horizontal mounting of the infrastructure.

4.1.4.2 Spatial and interannual variability of SoG during SPICE

Unlike the measurement of falling precipitation, snow on the ground is subject to melting, metamorphism, and redistribution during and after a precipitation event. Snow depth and SWE can be much more variable in space and in time than falling precipitation. This makes intercomparisons difficult; large differences between measurements are more likely to be the result of spatial variability than differences in the measurement technique or sensor configuration. Some spatial and interannual variability metrics are described for four of the participating SoG sites that have spatial measurements of snow depth or SWE, to put the SoG intercomparisons into context.

4.1.4.2.1 *CARE*

The snow-depth measurements at the CARE site are situated in an exposed area. The site itself is located on a topographical rise which increases the exposure. Although average wind speeds are not extremely high (3.7 m/s), snow redistribution due to wind readily occurs, which means that caution is required when comparing measurements made by the sensor under test to reference measurements obtained at some distance away.

The manual snow-depth measurements at CARE are made daily via visual observations of 62 snow stakes distributed throughout the intercomparison field. Table 4.13 summarizes the seasonal conditions at the site for both the 2013/14 and 2014/15 measurement seasons. The average snow depth was higher in 2013/14, with the timing of maximum depth and melt out date later than in 2014/15. Figure 4.66 confirms this very different seasonal time series of mean snow depth. Note that

the 2014/15 season was substantially shorter, but also included fewer manual observations due to lack of observations on the weekends during this season.

Table 4.13. Maximum snow depth, timing of maximum snow depth, and date that the site was snow free for each of the two intercomparison seasons at CARE.

	2013/14	2014/15
Max snow depth*	49.0 cm (average: 39.0 cm)	47.0 cm (average: 21.3 cm)
Time of max snow depth	Jan 6, 2014	Dec 11, 2014
Time of first seasonal snowfall	Dec 14, 2013	Dec 3, 2014
Melt out date**	Apr 14, 2014	Mar 18, 2015

*Values in parentheses represent the maximum average depth, while the values outside of the parentheses represent the maximum depth measured among the individual snow stakes.

**Definition: One week after snow was last measured at the end of the season.

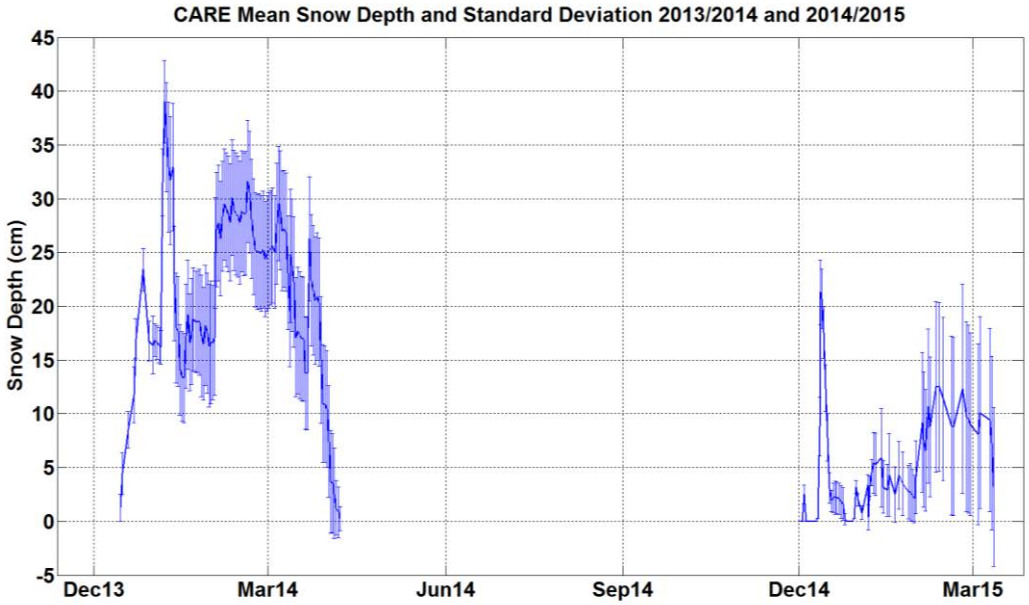


Figure 4.66. Mean and standard deviation of daily snow depth at CARE for the 2013/14 and 2014/15 seasons.

Figure 4.67 shows the daily deviation from the mean for each snow stake, for each season. Corresponding with higher snow depths, 2013/14 also exhibits a greater range in snow depths with stake deviation from the mean, typically ranging from -10 to 10 cm, with some stakes deviating by as much as 23 cm. The range of deviation in 2014/15 is largely confined to -5 to 5 cm, although some individual stakes deviated by as much as 35 cm in mid- to late-season. The coefficients of variation

(COV), defined as the ratio of the standard deviation to the mean and expressed as a percentage, is a standardized measure of dispersion and can be used for quantifying the spatial variability at the sites and for comparing the spatial variability amongst sites. The COV for 2013/14 is generally under 40% and is typically around 25%. The COV for 2014/15 is generally around 80-90%, and occasionally exceeds 100% through mid-season, which is substantially higher than 2013/14. Even though the range in measurements is higher in 2013/14 than 2014/15, the mean is also substantially higher, resulting in a lower COV.

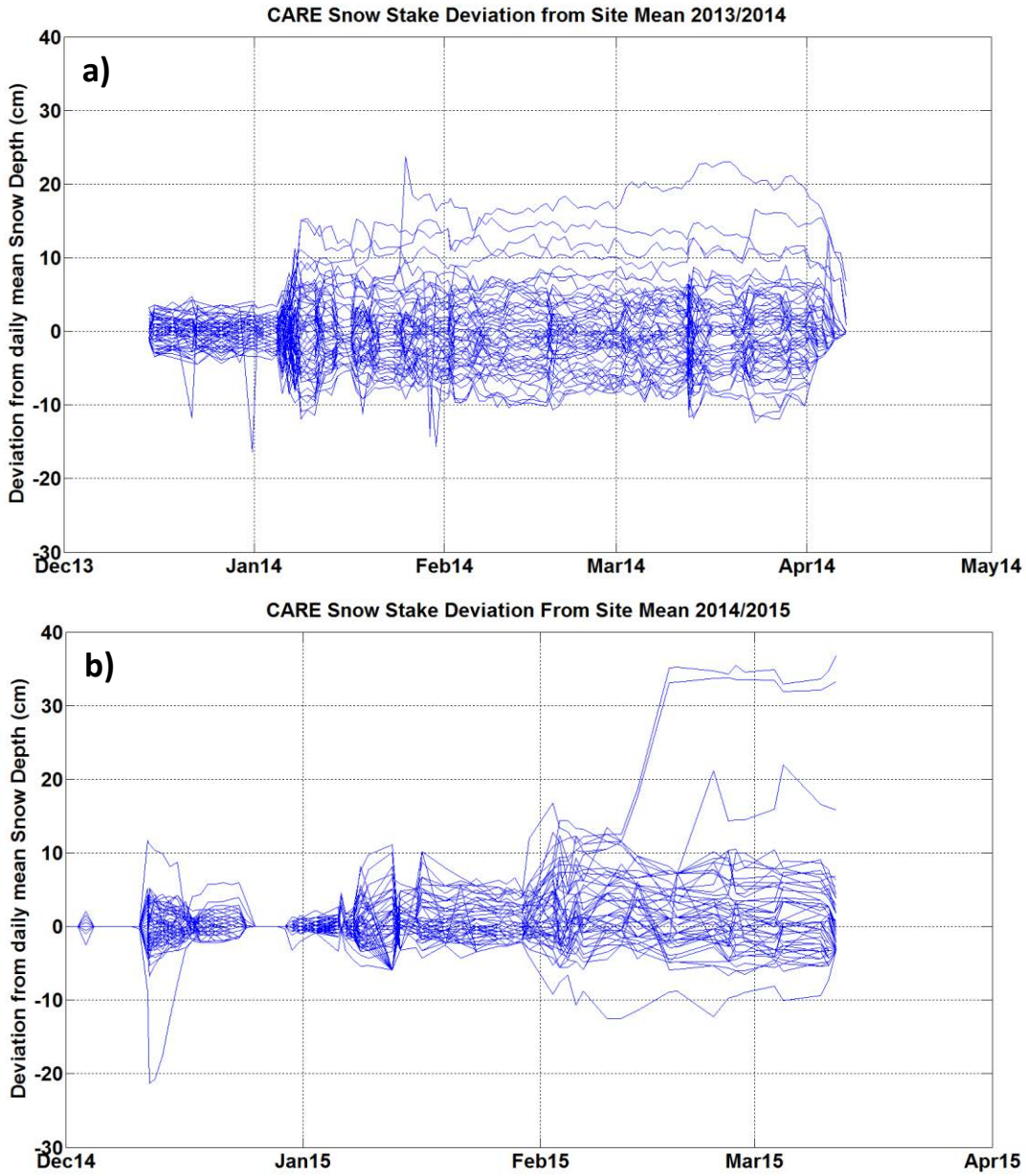


Figure 4.67. Snow stake daily deviation from the site mean for all 62 snow stakes at CARE for a) 2013/14 and b) 2014/15.

Figure 4.67 provides an overview of the variability in snow depth across the CARE site. The focus of the SPICE intercomparison is on the measurements made on pedestals 11A, 12A, and 20 (see site layout in Annex 4); the following figures show the variability of the manual measurements at these pedestals.

From Figure 4.68 (2013/14) and Figure 4.69 (2014/15), the deviation of any of the individual snow stakes from the mean is generally less than +/- 5 cm and only occassionally (once or twice in a season) exceeds this. During 2014/15 (low snow depth year), the individual snowfall events are quite discernable, and it is apparent when drifting occurs across the target area, resulting in one or two snow stake measurements being considerably different than the other corners of the target.

Another notable feature of the manual snowfall measurements at CARE is the mounding of snow over the target area (see Section 4.1.4.1). Because of the design of the snow stakes, the measured depth of snow in the centre of the target is usually higher than at the corners. As an example, the snow-depth time series for the SR50A on Pedestal 20 is shown along with the daily snow-depth measurements at each corner of the SR50A target in Figure 4.70.

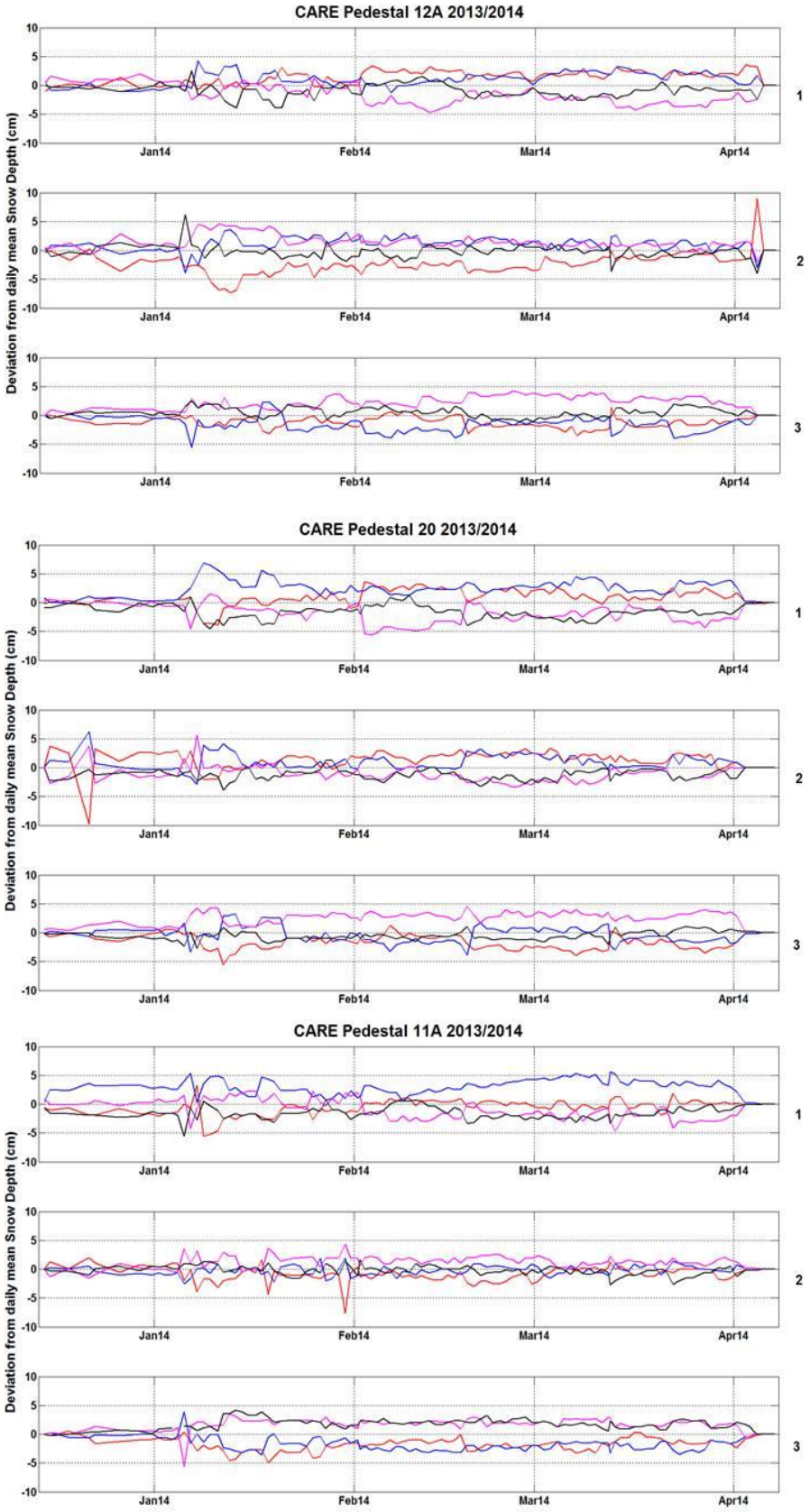


Figure 4.68. Daily deviation from the pedestal mean for snow stakes at each corner of targets 1 (top), 2 (middle) and 3 (bottom) for each pedestal at CARE over the 2013/14 season.

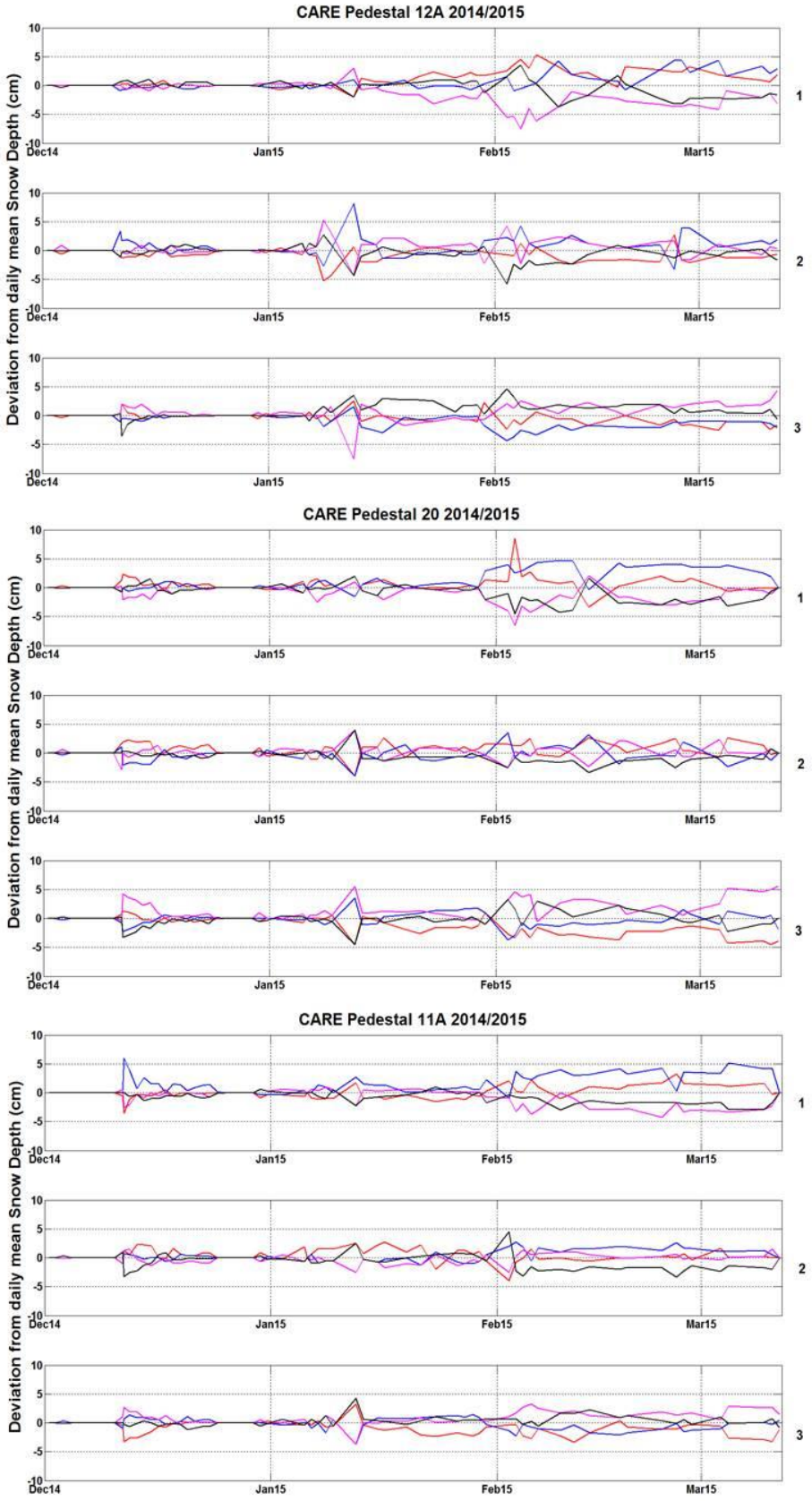


Figure 4.69. Daily deviation from the pedestal mean for snow stakes at each corner of targets 1 (top), 2 (middle) and 3 (bottom) for each pedestal at CARE over the 2014/15 season.

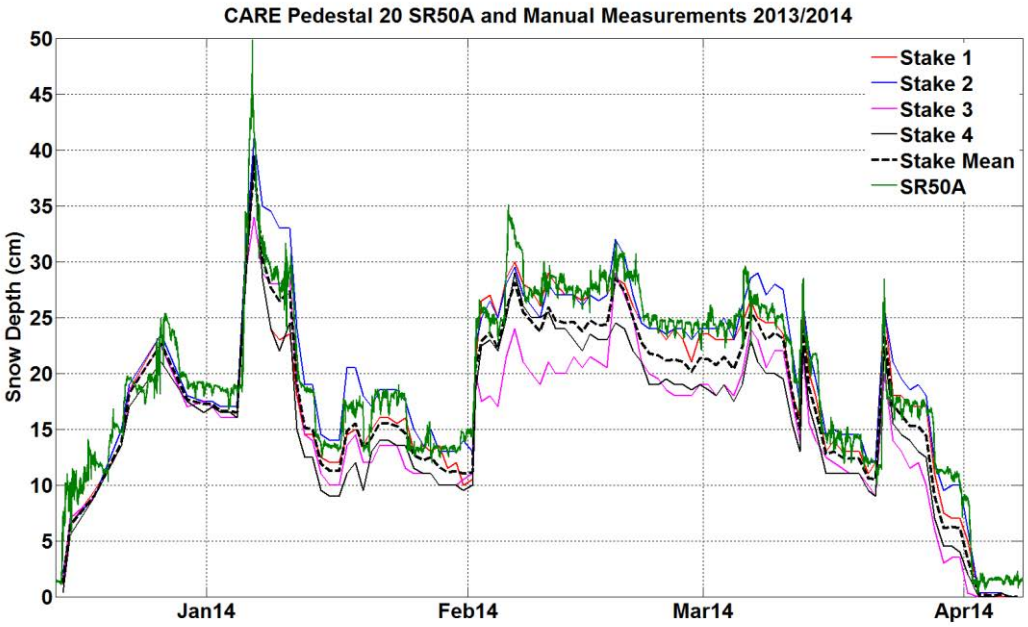


Figure 4.70. Time series of the daily manual snow stake measurements and 1-minute SR50A measurements on pedestal 20 at CARE for 2013/14. The stake measurements are at the corners of the target with the SR50A measuring the snow in the middle of the target.

4.1.4.2.2 *Sodankylä*

Sodankylä is a very sheltered site with average wind speeds less than 3 m/s. Snow redistribution at the site is small and spatial variability is relatively low.

As noted in Section 3.1.4.3.6, the manual snow measurements at Sodankylä are made daily using photographs of four snow stakes at different locations on the intercomparison field. Table 4.14 indicates maximum snow depth, timing of maximum snow depth, and date that the site was snow-free, for each of the intercomparison seasons. Maximum snow depth was slightly above 80 cm for each season. The snowpack in 2013/14 accumulated and melted later than in 2014/15.

Table 4.14. Maximum snow depth, timing of maximum snow depth, and date that the site was snow free for each of the two intercomparison seasons at Sodankylä.

	2013/14	**2014/15**
Max snow depth*	82 cm (average: 79.75 cm)	84 cm (average: 82.75 cm)
Time of max snow depth	Mar 24, 2014	Apr 1, 2015
Time of first seasonal snowfall	Oct 30, 2013	Sep 24, 2014
Time of snow-free site**	May 20, 2014	May 11, 2015

*Values in parentheses represent the maximum average depth while the values outside of the parentheses represent the maximum depth measured among the individual snow stakes.

**Definition: One week after snow was last measured at the end of the season.

As a method of assessing the daily and seasonal spatial variability in snow depth at the site, time series of daily average snow depth and standard deviation (SD) are plotted in Figure 4.71.

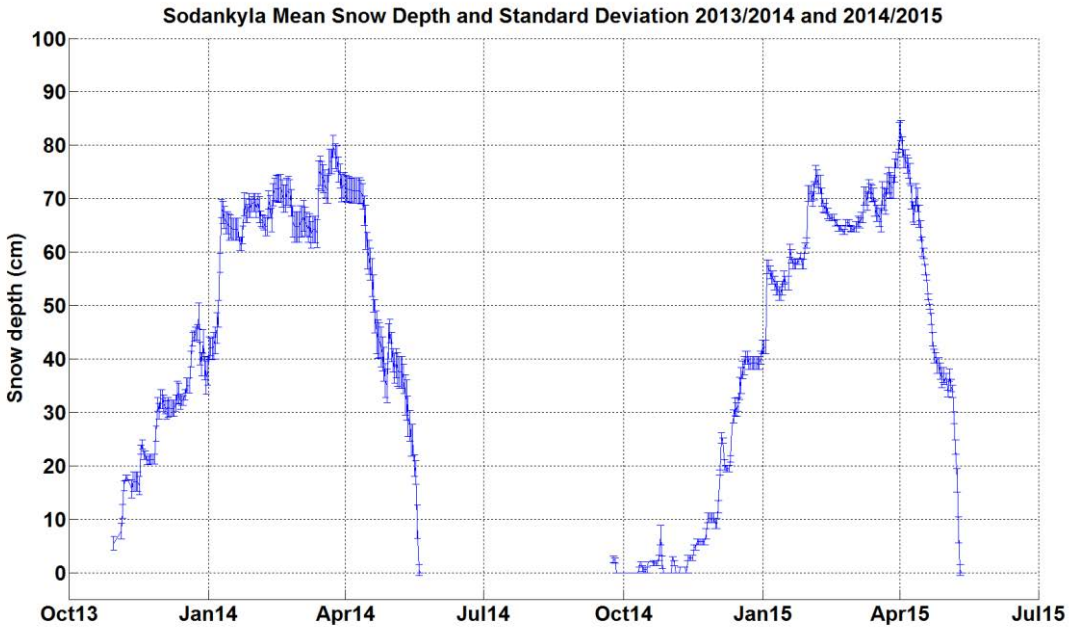


Figure 4.71. Mean and standard deviation of daily snow depth at Sodankylä for the 2013/14 and 2014/15 seasons.

Although the time series of daily snow depth look similar between the two seasons, it appears that snow depth during the 2013/14 season varies more with time and in space. Figure 4.72 shows the daily deviation from the site mean for each of the four snow stakes for 2013/14 (Figure 4.72a) and 2014/15 (Figure 4.72b). The range of values for each snow stake is higher in 2013/14 than it is in 2014/15. The range of deviation varies from about -2 cm to 4 cm in 2014/15 while the range in 2013/14 is from about -4.5 cm to 5 cm. However, considering that the distance between the snow stakes varies from 20 to 46 m, the snow depth is still rather homogenous. The COV is generally under 6% and occasionally approaches 7.5%. Through the middle of the season, the COV is lower for 2014/15 than it is for 2013/14, suggesting lower spatial variability.

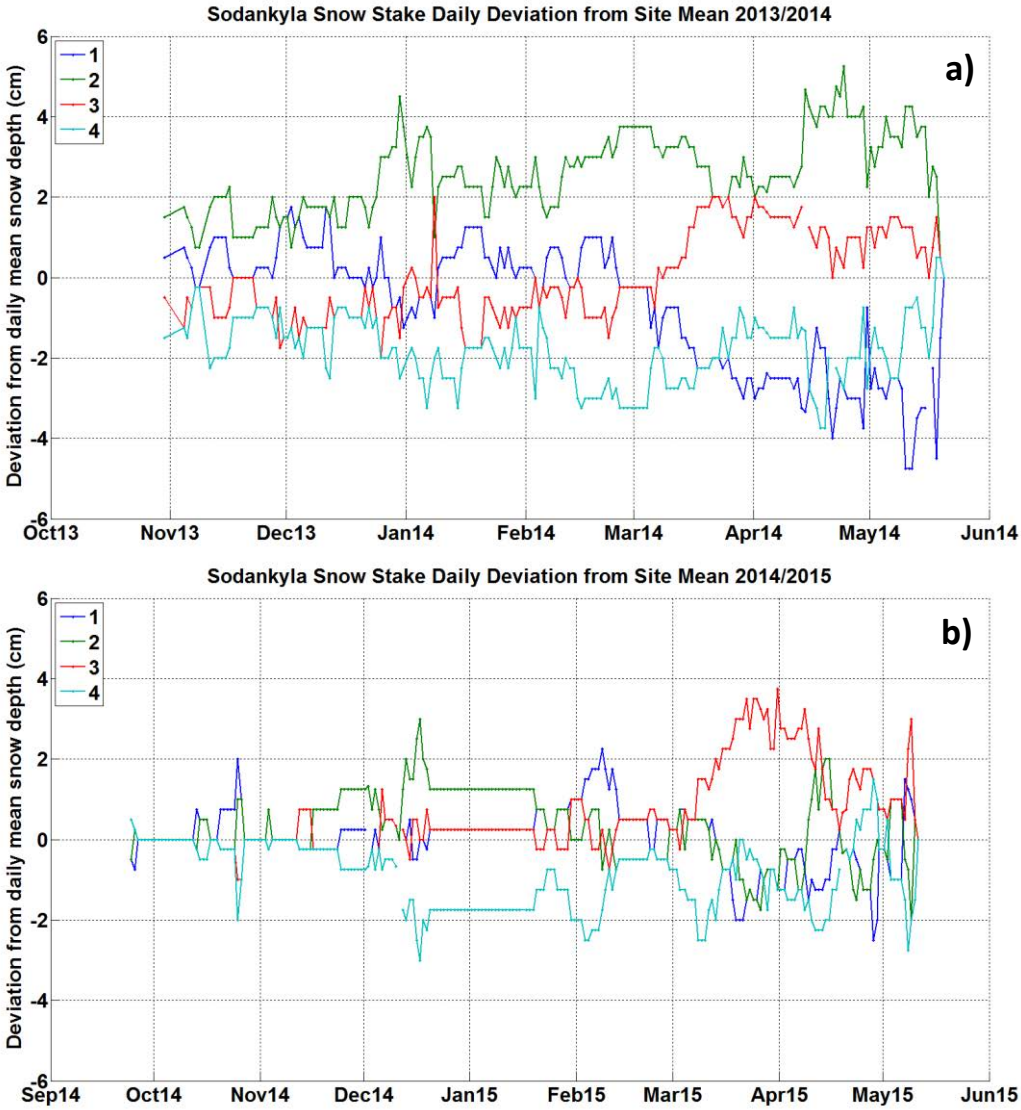


Figure 4.72. Snow stake daily deviation from the site mean for Sodanklyä for a) 2013/14 and b) 2014/15. See corresponding evolution of the daily mean snow depth (Figure 4.71).

4.1.4.2.3 *Col de Porte*

Col de Porte is a relatively sheltered alpine site that receives an abundance of snow. The topography varies across the site and, combined with occasionally high winds, results in some local-scale variability across the measurement site.

The manual snow measurement at Col de Porte is performed via three snow stakes in the instrument compound that are measured weekly during the winter. From Table 4.15 and Figure 4.73, maximum snow depth was larger and occurred later in 2013/14, with the earliest recorded snow depth also occurring earlier in 2013/14.

Table 4.15. Maximum snow depth, timing of maximum snow depth, and date that the site was snow free for each of the two intercomparison seasons at Col de Porte.

	2013/14	2014/15
Max snow depth*	166 cm (average: 164.0 cm)	151 cm (average: 148.0 cm)
Time of max snow depth	Mar 6, 2014	Feb 4, 2015
Time of first seasonal snowfall	Dec 5, 2013	Dec 30, 2014
Time of snow-free site**	Apr 23, 2014**	Apr 22, 2015**

*Values in parentheses represent the maximum average depth, while the values outside of the parentheses represent the maximum depth measured among the individual snow stakes.

**Definition: One week after snow was last measured at the end of the season.

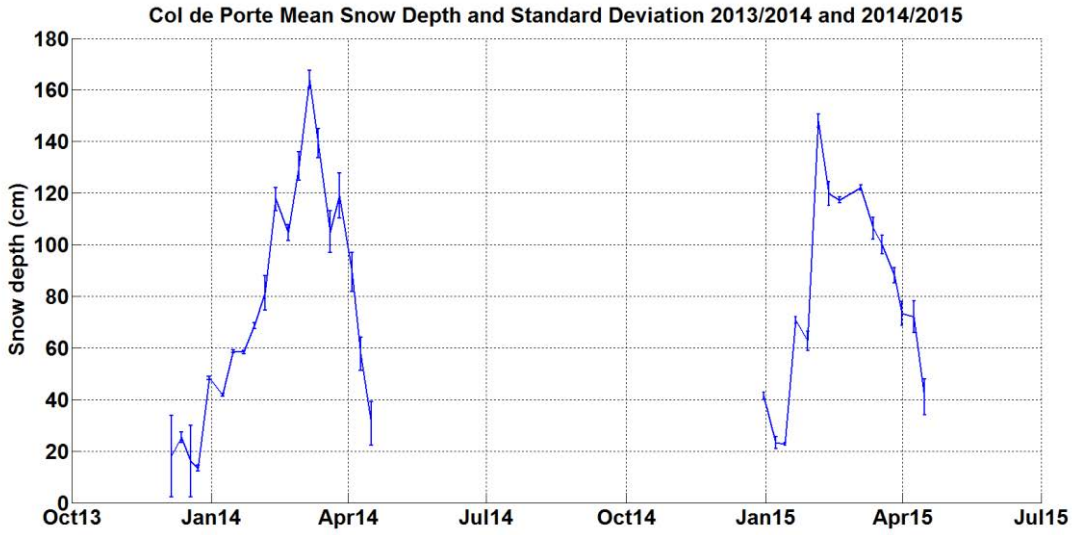


Figure 4.73. Mean and standard deviation of weekly snow depth at Col de Porte for the 2013/14 and 2014/15 measurement seasons.

Figure 4.74 shows the daily deviation from the site mean for each of the three stakes. Along with a deeper snow pack, 2013/14 also shows a higher range in depths at the site, with stakes varying up to +/- 10 cm from the mean (with season start measurements varying by as much as 18 cm from the mean). Generally, the 2014/15 deviations range from -5 cm to +5 cm with values as high as 8 cm at the end of the season.

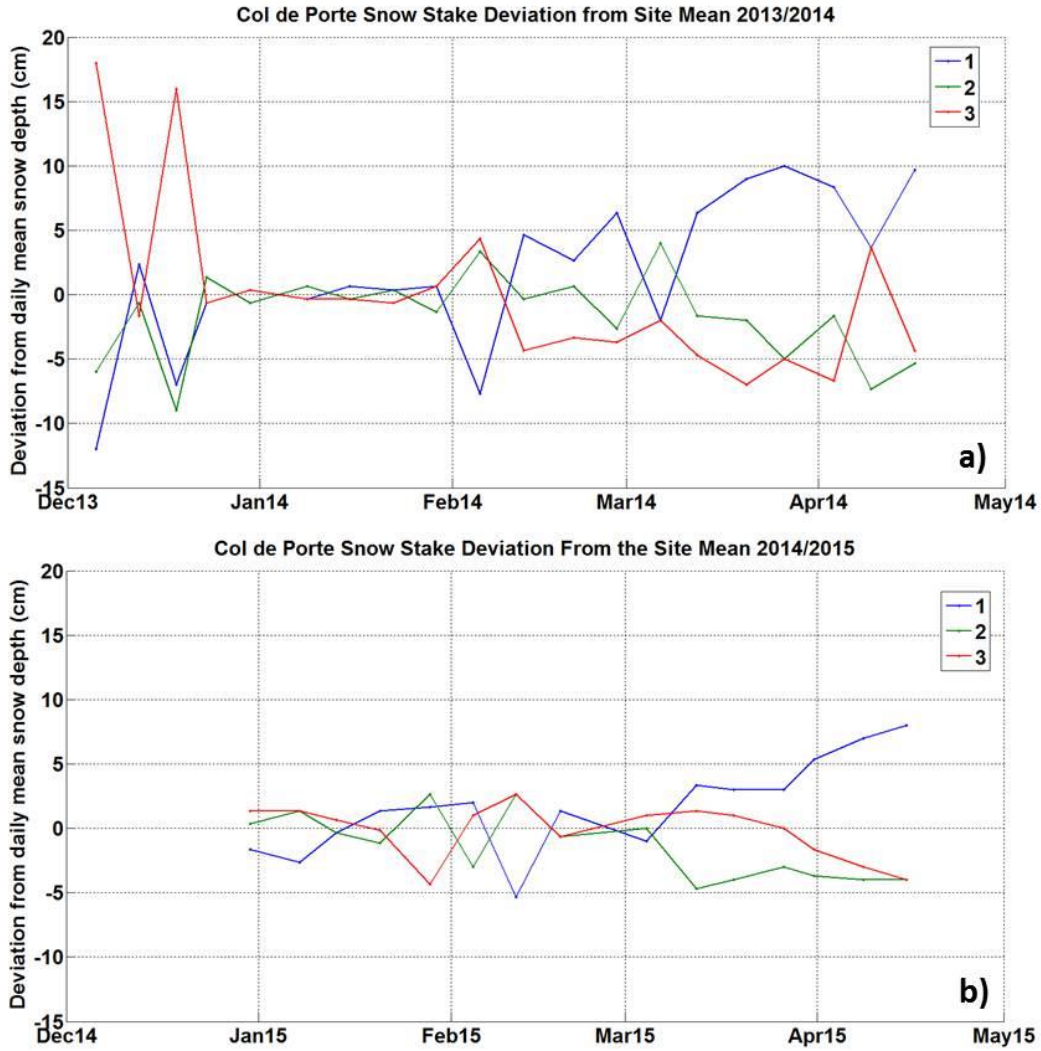


Figure 4.74. Snow stake deviation from the site mean for Col de Porte for a) 2013/14 and b) 2014/15. See corresponding evolution of the daily mean snow depth (Figure 4.73).

The COV for Col de Porte is quite similar to Sodankylä, approaching 10% during the middle of the season, and generally under 6% for much of the season (with the exception of season start and season end). COV is relatively low because of the high mean snow depths relative to the standard deviation.

4.1.4.2.4 *Caribou Creek*

The Caribou Creek snow measurements are made across a clearing surrounded by a young forest canopy (2 to 3 m in height). There is some wind redistribution in the clearing, with deposition favored along the edges of the clearing. The clearing and the surrounding vegetation also produce differential melting due to relative exposures which contributes to the high spatial variability at the site, especially at the beginning and end of the winter periods.

There are no snow-depth instrument intercomparisons at Caribou Creek, but there are SWE instrument intercomparisons. The following analysis shows the spatial and inter-annual variability of

snow depth at the site, which is required to put the SWE intercomparisons into context. Snow depths were measured concurrently with the SWE snow course every two weeks during the winter periods.

The snow course at Caribou Creek is described in the site layout in Figure 4.75. The course consists of five bulk-density SWE samples, with approximately five snow stakes between each SWE measurement (for a total of 25). The observations were made approximately every 2 weeks during the winter season.

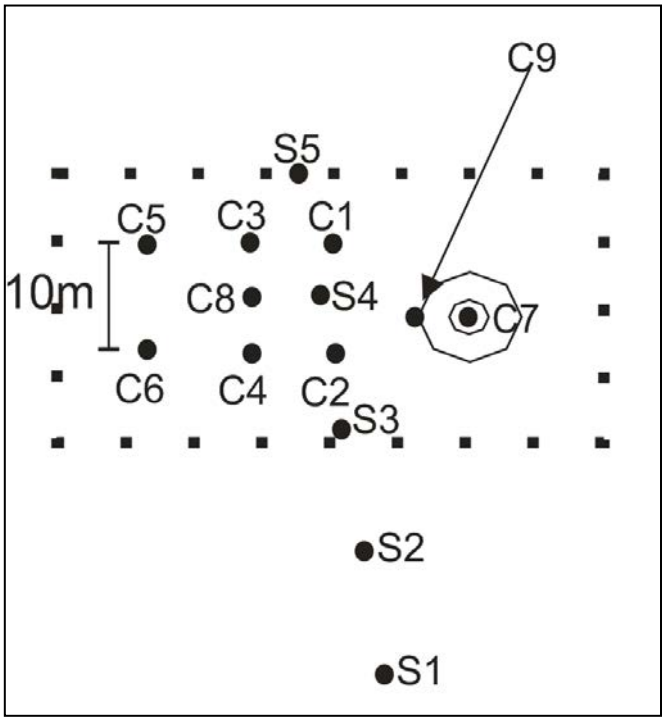


Figure 4.75. The five point snow course at Caribou Creek (S1 through S5) in relation to the intercomparison field clearing (inside area delineated by the dotted lines) and the instrument pedestals (C1 through C9). S1 and S2 are positioned in the Jack Pine regrowth, S3 and S5 are on the edge of the clearing, and S4 is centred in the clearing. C7 is the R2 reference configuration.

Table 4.16 summarizes the snowfall characteristics for the two measurement seasons at Caribou Creek. Due to a warmer and drier winter in 2014/15, the maximum snow depth is lower, but the season lengths are very similar. This can also be seen in Figure 4.76, which shows the time series of mean manual snow-depth measurements with the site standard deviation plotted as error bars. The maximum snow depth shown in Table 4.16 is derived from the automated sensor installed at C8 in the intercomparison field (Figure 4.75) and is considerably less than the snow course mean as shown in Figure 4.76. This is because the manual snow depth mean includes the observations in the treed and transitional areas, which tend to collect more snow than in the clearing where the intercomparison field is located.

Table 4.16. Maximum snow depth, timing of maximum snow depth, and estimated date that the site was snow free for both intercomparison seasons at Caribou Creek (derived from the automated snow-depth sensor in the intercomparison field).

	2013/14	2014/15
Max snow depth	56 cm	41 cm
Time of max snow depth	Apr 6, 2014	Feb 15, 2015
Time of first seasonal snowfall	Nov 5, 2013*	Oct 27, 2014
Time of snow-free site	May 7, 2014*	May 6, 2015*

*Approximate date.

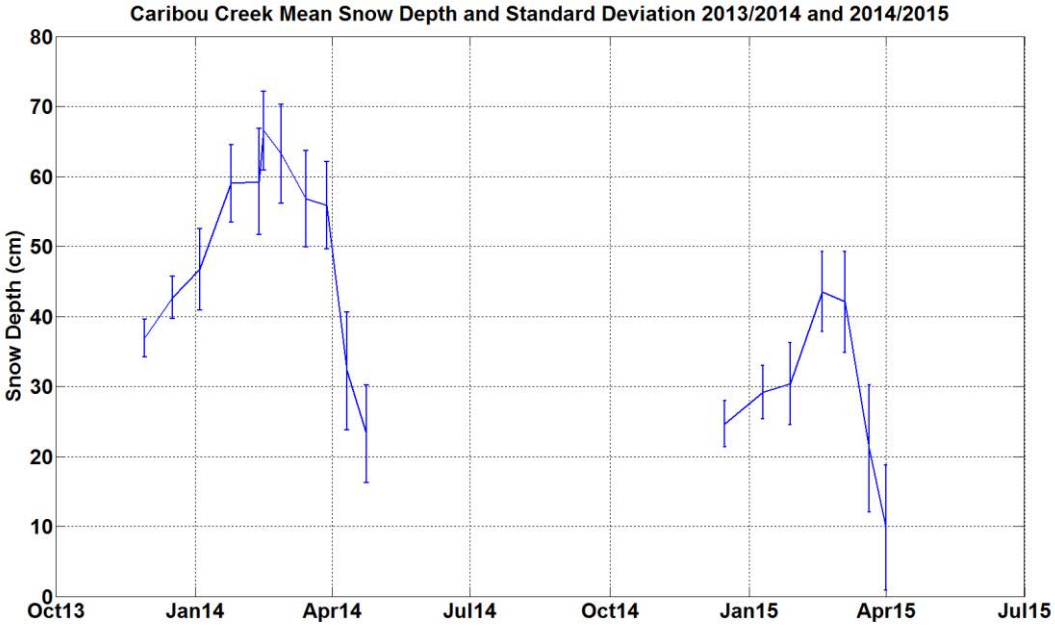


Figure 4.76. Mean and standard deviation of bi-weekly manual snow-depth measurements at Caribou Creek for the 2013/14 and 2014/15 measurement seasons.

Figure 4.77 shows how the manual snow depths measured at the five SWE sampling points (S1 through S5) deviate from the site mean throughout the two SPICE seasons. Overall, the individual stakes are usually within +/- 10 cm of the site mean, but with maximum snow depths of 56 cm and 41 cm for the two respective seasons, this variability is quite high, with the COV exceeding 90% during melting in the 2014/15 season. Generally, the COV for 2013/14 only exceeds 30% at season end and is less than 15% at season maximum snow depth. The COV for mid-season of 2014/15 is under 19%. Keeping in mind the high variability at the site, the S1 and S2 stakes in the trees on the south side of the intercomparison field tend to have higher than average snow depths, while S3 on the edge of the clearing generally measures close to the site average (with exceptions). S4 in the clearing is usually

lower than the site average. S5, in the transition zone on the north side of the clearing, is generally at or slightly above the site mean.

The spatial variability of SWE follows much the same pattern as for snow depth, as shown in Figure 4.78. Following this, the COV for SWE are also very similar to the values for snow depth.

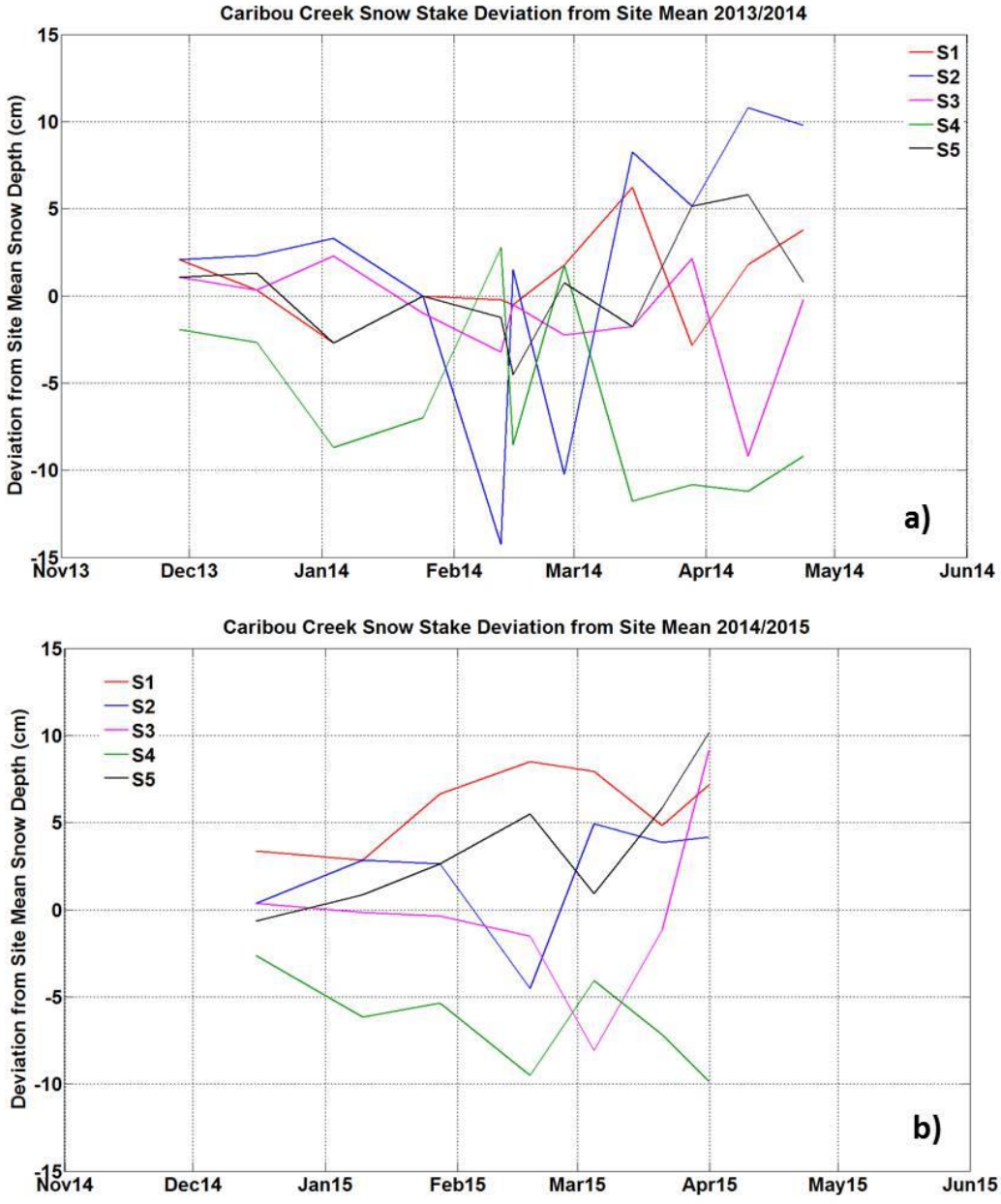


Figure 4.77. Snow stake deviation from the site mean for Caribou Creek for a) 2013/14 and b) 2014/15. See corresponding evolution of the daily mean snow depth (Figure 4.76).

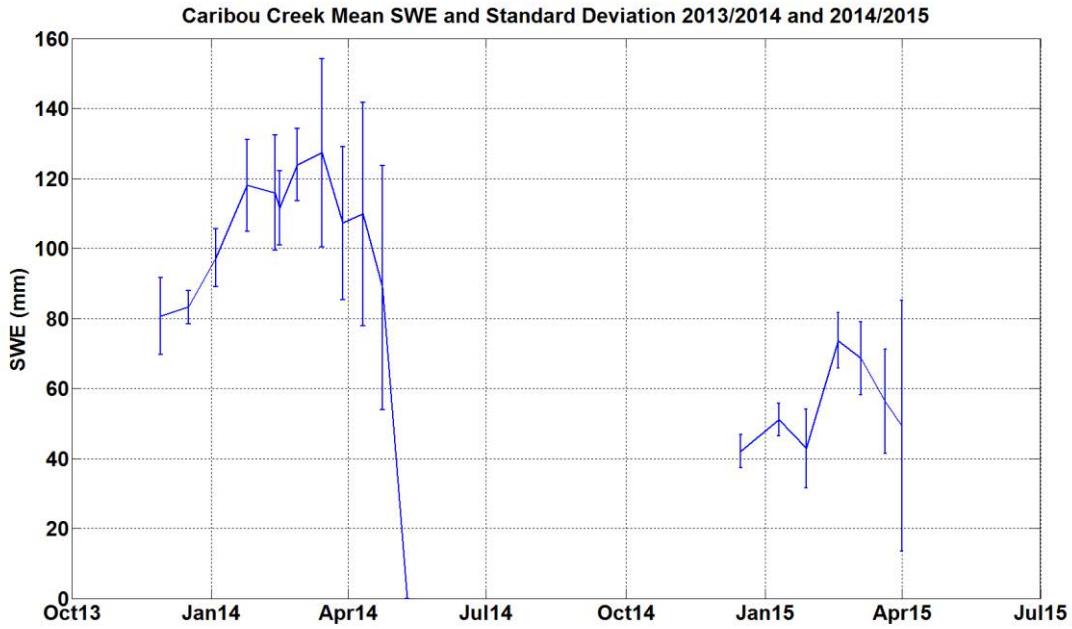


Figure 4.78. Mean and standard deviation of bi-weekly manual SWE measurements at Caribou Creek for the 2013/14 and 2014/15 seasons.

4.1.4.3 Assessing snow-depth instrument diagnostics

Authors: Craig Smith, Samuel Buisán, Anna Kontu, Lauren Arnold, Javier Alastrué, José Luis Collado, Yves Lejeune, Jean-Michel Panel

Two snow-depth instruments tested during SPICE had the capability of outputting measurement or signal diagnostics that could potentially be used to assess the quality of the measurement, or to extend the capabilities of the sensor. The SR50A, the sonic ranging instrument manufactured by Campbell Scientific, outputs a measurement quality number that is a proprietary echo processing calculation related to the return sonic signal reflected from the target. The SHM30, an optical range finding instrument manufactured by Lufft/Jenoptik, outputs a signal strength value for each measurement that represents the strength of the return optical beam reflected from the target. The following analysis is an assessment of how the measurement environment influences these diagnostic outputs in various climate regimes and measurement situations, and provides some guidance on how to use these diagnostics to assess the data quality or extend the capability of the sensor.

4.1.4.3.1 *SR50A quality numbers*

According to the Campbell Scientific manual for the SR50A, the quality number output can be used as an indication of the measurement certainty. Table 4.17 provides the possible ranges of quality numbers and the uncertainty associated with each range. The manufacturer suggests that some of the causes of high uncertainty numbers are: 1) a sensor not perpendicular to the target; 2) a target that is small or a poor reflector of sound; and 3) a rough or uneven target surface. During the SPICE intercomparisons, the impact of distance to target, wind speed (and potentially blowing snow), temperature, humidity, and snowfall rates on the SR50A quality numbers were assessed using data from Formigal, Col de Porte, Sodankylä, and CARE.

Table 4.17. Campbell Scientific SR50A quality number output description, as per user manual.

Quality Number Range	Range Description
0	Not able to read distance
152 – 210	Good measurement
210 - 300	Reduced echo signal strength
300-600	High measurement uncertainty

Although the analysis differed slightly by site, the frequencies of occurrence of the quality number categories listed in Table 4.17 were assessed for each site, with each site having a different distance to the target (higher for alpine sites to accommodate deeper snowpacks). At some windier sites, the impact of wind speed was examined. At all sites, the impact of the occurrence and rate of precipitation (as defined by the SEDS and SNEDS, defined in Section 3.4) on the frequency distribution of the quality numbers was examined. For the precipitation analysis, the quality numbers obtained each minute were averaged over the same 30-minute periods used in the SEDS and SNEDS to provide an overall estimate of the quality of the SR50A measurement during each 30-minute period.

The results of the frequency distribution assessment are provided in Table 4.18. Overall, the percentages of "Good Measurement" quality numbers were high, varying between 71% at Formigal to 98% at CARE, with the majority of the remaining measurements falling into the "Reduced echo signal strength" category. The percentage of "High measurement uncertainty" was very low at all sites, with only Formigal exceeding 2%.

Table 4.18. Frequency distribution (percentage of total measurements) of SR50A quality numbers at Formigal, Col de Porte, Sodankylä, and CARE.

Measurement Quality	% of Total Measurements			
	Formigal	Col de Porte	Sodankylä*	CARE**
Good Measurement	71%	87%	92%	98%
Reduced Echo Signal Strength	27%	12%	7%	2%
High Uncertainty	2%	1%	1%	0%
Not able to Read	0%	0%	0%	0%

*Mean value for two sensors over two seasons

**Mean value for three sensors over two seasons

4.1.4.3.1.1 Formigal

At Formigal, the quality number analysis used data collected between December 2014 and November 2015. The installation height of the instrument was 4.2 m above the ground (Figure 4.79) with a maximum snow depth of 270 cm occurring in mid-February of 2015 (Figure 4.80). The target area under the SR50A was natural ground consisting of rock and grass. The 2014/15 accumulation period was characterized by several large snowfall events between late-December and mid-February with an event in early February resulting in an accumulation of nearly 170 cm of fresh snow.

Figure 4.79. Installation of the SR50A sensor at the Formigal site.

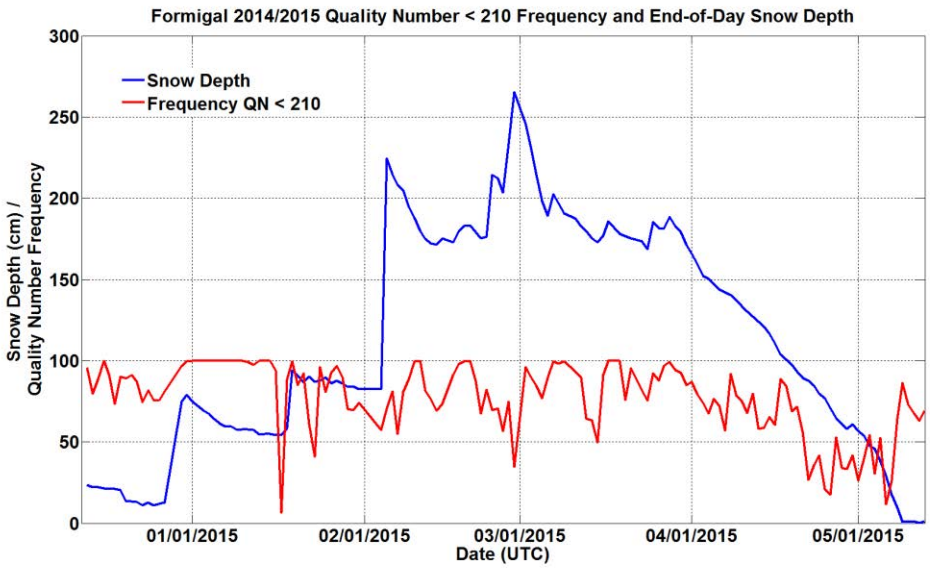


Figure 4.80. SR50A measured snow depth (blue) and percentage of "Good Measurement" quality numbers (quality numbers of 150 to 210; red) during the 2014/15 season at Formigal.

With Formigal having the lowest percentage of "Good measurements", it was necessary to determine the seasonal distribution of the quality numbers. Figure 4.80 shows both the end-of-day snow depth and the daily percentage of "Good measurements". The figure shows some periodic drops in the frequency during large snowfall events (i.e. mid-January) but a more systematic decrease in frequency during rapid melting in mid-April and early May. As a result, the percentage of "Good measurement" quality numbers in April and May are 59% and 47% respectively, a considerable drop from the mean of 71%. This drop in the percentage of "Good measurement" numbers corresponded with an increase in the percentage of "Reduced echo signal strength" quality numbers in April and May to 39% and 52%, respectively (as compared to the mean of 27%). It is difficult to pinpoint the exact cause of this distribution, but one can speculate that it is a result of the degradation of the surface of this deep snowpack as it progresses through the melt period with the distance to the target increasing with melt.

Because of the greater distance to the target, speculation is that the quality output of the SR50A at Formigal is more sensitive to the occurrence of snowfall. Figure 4.81 shows the breakdown of the frequency distribution of quality numbers during precipitation (as determined by the SEDS) and when no precipitation is occurring (as determined by the SNEDS).

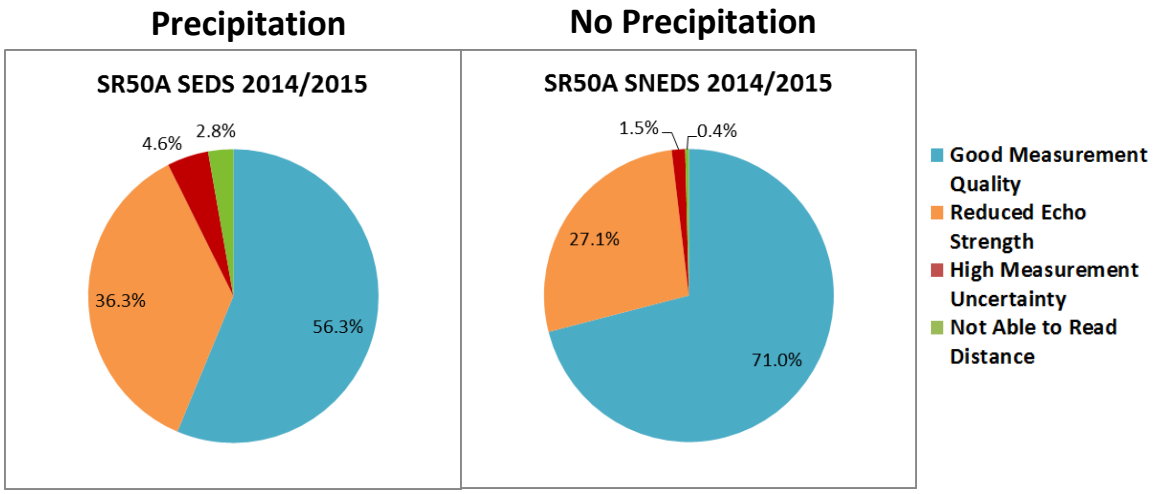


Figure 4.81. Frequency distribution of measurement quality categories for the SR50A at Formigal during the 2014/15 winter season for precipitating events defined by the SEDS and non-precipitating events defined by the SNEDS.

The measurement quality for the SR50A at Formigal decreases with the occurrence of precipitation. This was not unexpected, and is likely due to the presence of hydrometeors in the path of the sonic signal or decreased density of the target surface due to new snow, both exacerbated by the relatively long distance between the sensor and the target surface. Table 4.19 shows that under more intense snowfall events, where the snow depth changes by more than 2 cm in a 30 minute period, the frequency distribution of the quality numbers is shifted further toward higher uncertainty. A more in-depth analysis of what is occurring during heavy precipitation is required.

Table 4.19. Frequency distribution of the SR50A quality numbers at Formigal during periods in which the snow depth increased by > 2 cm during over 30 minutes.

Measurement Quality During Snowfall > 2 cm / 30 min	% of Total Measurements
Good	40.8%
Reduced Echo	49.0%
High Uncertainty	10.2%

Besides the periodic decreases in the percentage of "Good measurements" associated with heavy snowfall events, there are spurious spikes that appear on days with no increase in snow depth. Some examples occur in the third week of January, mid-February, and mid-March, and appear to be associated with high relative humidity (> 90%) but only light precipitation. These events are likely rain or mixed precipitation events which could be impacting the condition of the surface target and reducing the certainty of the measurement.

One of the speculated reasons for decreased measurement quality was increased wind speed combined with the greater distance to the target. Wind could potentially distort the sonic signal over these greater distances and potentially reduce the signal quality. High wind speeds could also result in blowing snow, which may also decrease signal quality. However, no significant correlations were found between the quality numbers and wind speed (analysis not shown here).

4.1.4.3.1.2 Col de Porte

A SR50ATH sensor was installed at Col de Porte in January 2014 at a height of 4 m above the surface. This analysis covers data collected from November 2014 through April 2015. Like Formigal, the distance from the instrument to the target is relatively large. The target under the SR50ATH at this site is mowed natural grass. The snow depth time series for the 2014/15 winter is shown in Figure 4.82 and exceeds 160 cm in early February of 2015.

The frequency distribution of the Col de Porte SR50ATH quality numbers (Table 4.18) suggests that like Formigal, the higher installation height may have an impact on the number of "Good measurements". However, unlike Formigal, Figure 4.82 does not show the same systematic decrease in measurement quality during spring melting. The reason for the difference between the two sites is not known, and requires further investigation.

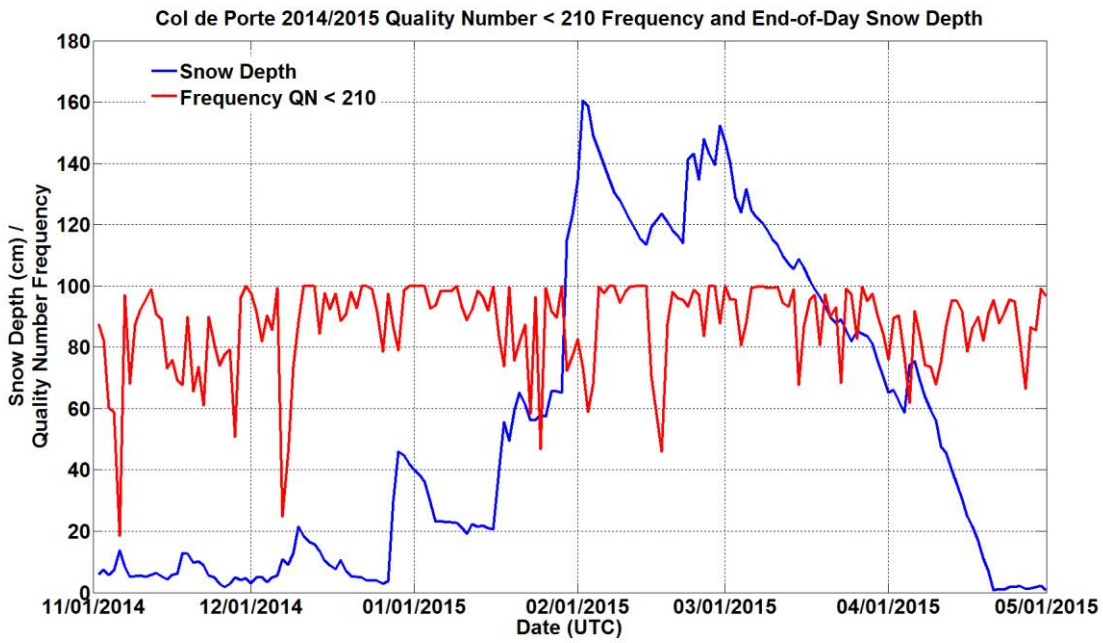


Figure 4.82. SR50A measured snow depth (blue) and percentage of "Good Measurement" quality numbers (quality numbers of 150 to 210; red) during the 2014/15 season at Col de Porte.

The impact of precipitation on the frequency distribution of the quality numbers is indicated in Figure 4.83. The SEDS and SNEDS data for Col de Porte were prepared differently than for S2 SPICE sites because the site does not host an R2 reference, but does have a reliable quality controlled data set of hourly precipitation totals provided by the site manager (see Annex VII of the IOC-5 Final Report, Sodankylä). As with the SEDS and the SNEDS at the other sites in this analysis, the occurrence (or non-occurrence) of precipitation was cross-referenced with the SR50ATH quality numbers. Since the SEDS uses a 30-minute minimum threshold of 0.25 mm to define a precipitation event, a 0.5 mm threshold was used to define an event in this 60 minute dataset.

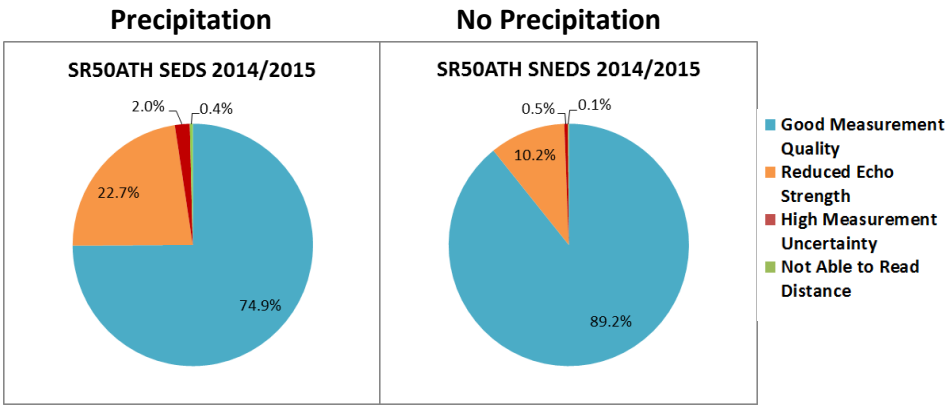


Figure 4.83. Frequency distribution of measurement quality categories for the SR50ATH at Col de Porte during the 2014/15 winter season for precipitating and non-precipitating events in a 60-minute precipitation data set provided by the site host.

The differences in the frequency distributions shown in Figure 4.83 indicate that the instrument at this site is also influenced by the occurrence of precipitation. The decrease in “Good measurements” quality numbers corresponds with an increase in the frequency of measurements with “Reduced Echo” and a small increase in “High Uncertainty” measurements.

Figure 4.82 shows the correspondence between changes in snow depth as measured by the sensor at the end of each day and the corresponding daily percentage of “Good measurement” quality numbers. Rapid drops in the frequency of “Good” measurements are often associated with relatively large increases in snow depth, in a very similar manner as observed for Formigal. This is supported by the results in Table 4.20, which show the frequency distribution of quality numbers during snowfall events when the sensor registers an increase > 2 cm during a 30-minute period. There is an apparent reduction in measurement quality during heavy snowfall, and as at Formigal, this is most likely related to the large distance between the sensor and the target. No relationships were found between wind speed and changes in the quality numbers (analysis not shown here).

Table 4.20. Frequency distribution of the SR50A quality numbers at Col de Porte during periods in which the snow depth increases by > 2 cm over 30 minutes.

Measurement Quality During Snowfall > 2 cm / 30 min	% of Total Measurements
Good	48.1%
Reduced Echo	41.4%
High Uncertainty	10.5%

4.1.4.3.1.3 Sodankylä

Two SR50ATH sensors were installed at the Sodankylä test site on two different pedestals in October 2013. Analysis is based on winter data collected through June 2015. The sensors were installed at a height of 2 m with an expected snow depth anticipated to approach 1 m. The targets under the SR50ATH sensors are 2 x 2.5 m artificial grass mats mounted flush and level to the ground.

In general, the sensor reported “Good measurement” quality over 90% of the time (Table 4.18). The frequency of “Reduced Echo” reports was less than 10%. The frequency of “High Uncertainty” reports was small (1%) and the frequency of “Not able to read” reports was negligible (0%).

Using the SEDS and SNEDS to indicate the occurrence or non-occurrence of precipitation, the impact of precipitation on the quality numbers was examined and the frequency distribution breakdown by season and by sensor is shown in Figure 4.84. In general, the frequency of “Good measurement” quality numbers was lower during precipitation events. Concurrently, the frequency of “Reduced Echo” and “High Uncertainty” measurements increased, with the largest increases in frequency observed for “Reduced Echo” quality numbers. However, when compared to the reduced measurement quality during precipitation at the alpine sites, the reduction at Sodankylä was considerably less (8% vs. 15% reduction in “Good measurement” numbers). As with the alpine sites, there was a further reduction in measurement quality with increased precipitation rates (not shown here), but the impact was less for the sensors Sodankylä relative to those at the alpine sites.

Wind speeds and mounting heights at Sodankylä were not sufficiently high to warrant further analysis.

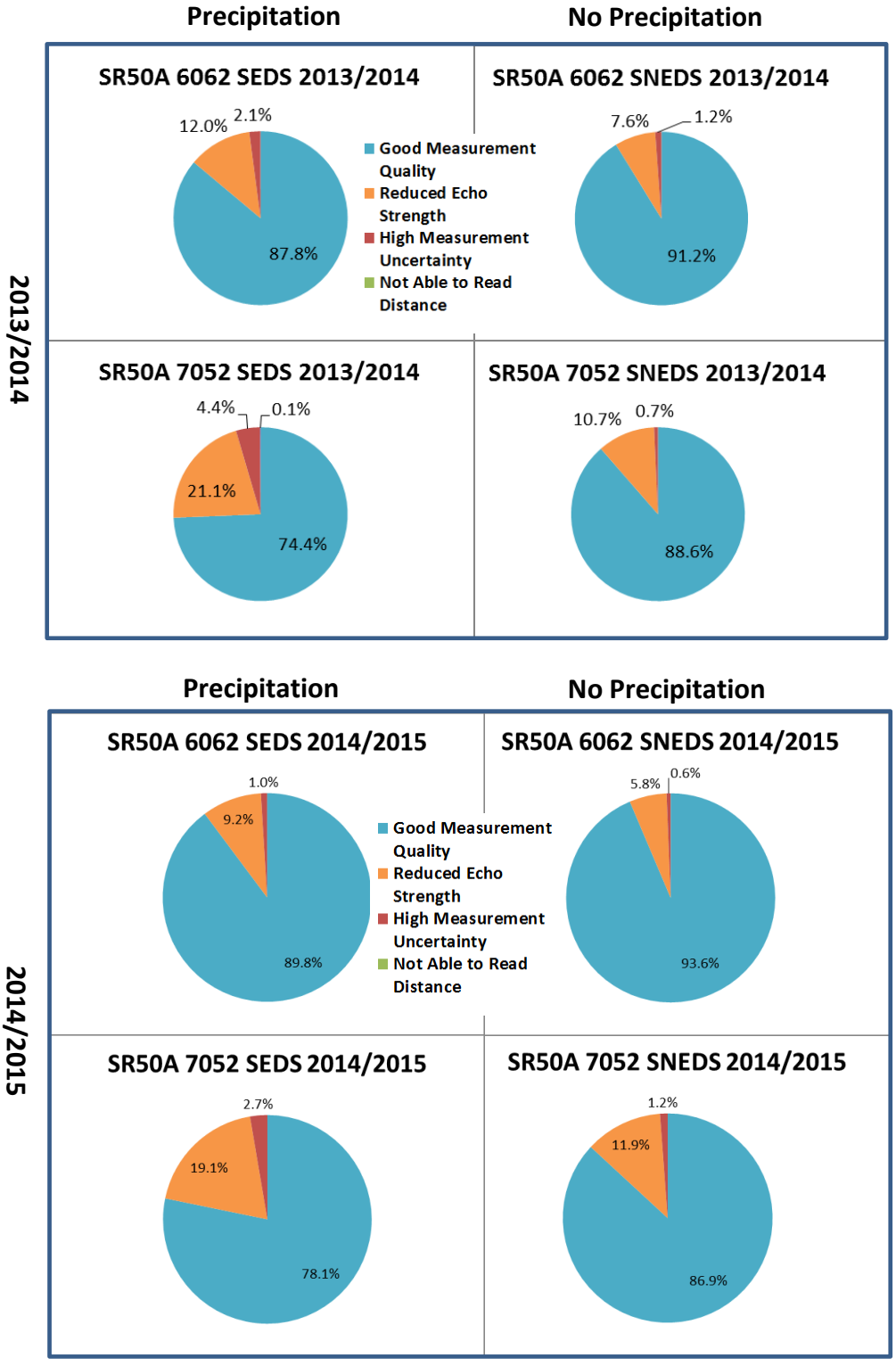


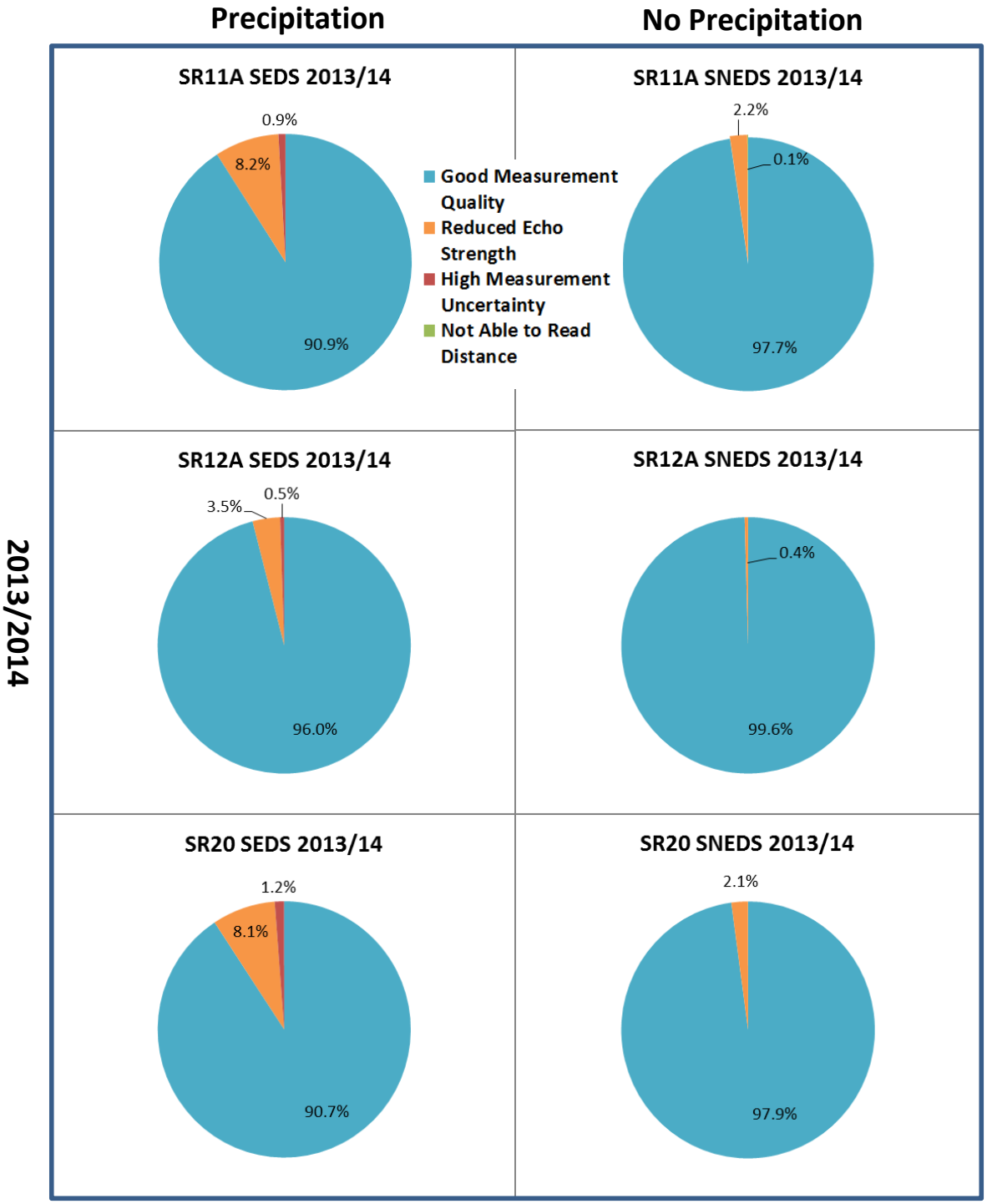
Figure 4.84. Frequency distribution of measurement quality categories for SR50A sensors over the 2013/14 and 2014/15 measurement seasons at Sodankylä. Precipitating events are defined by the SEDS and non-precipitating events are defined by the SNEDS.

4.1.4.3.1.4 CARE

Three SR50A sensors were installed on separate pedestals at CARE in November 2013; analysis is based on data collected through to April 2015. The sensor heights at this site were 1.5 m, with expected snow depths within 0.5 m. Targets under these sensors were 1.2 m x 1.2 m gray, textured, and perforated plastic sheets mounted flush and level to the ground.

Table 4.18 shows that a very large (98%) of the measurements taken at CARE were classified as "Good measurements" with only a small percentage (2%) showing "Reduced Echo" quality numbers. Less than 0.5% of the measurements had "High Uncertainty" quality numbers.

Figure 4.85 shows the reduction in measurement quality with the occurrence of precipitation as defined by the SEDS and SNEDS. As with the other sites, the percentage of data categorized as "Good measurements" is reduced during precipitation, about 3% on average (higher in 2013/14 than in 2014/15). This reduction is less than at Sodankylä and considerably less than at the alpine sites. Also, unlike Sodankylä and the alpine sites, the frequency of "Good measurements" does not appear to decrease further with increased precipitation rates (not shown).


Precipitation
No Precipitation
2013/2014
SR11A SEDS 2013/14
0.9%
8.2%
90.9%
SR11A SNEDS 2013/14
2.2%
0.1%
97.7%
Good Measurement Quality
Reduced Echo Strength
High Measurement Uncertainty
Not Able to Read Distance
SR12A SEDS 2013/14
3.5%
0.5%
96.0%
SR12A SNEDS 2013/14
0.4%
99.6%
SR20 SEDS 2013/14
1.2%
8.1%
90.7%
SR20 SNEDS 2013/14
2.1%
97.9%

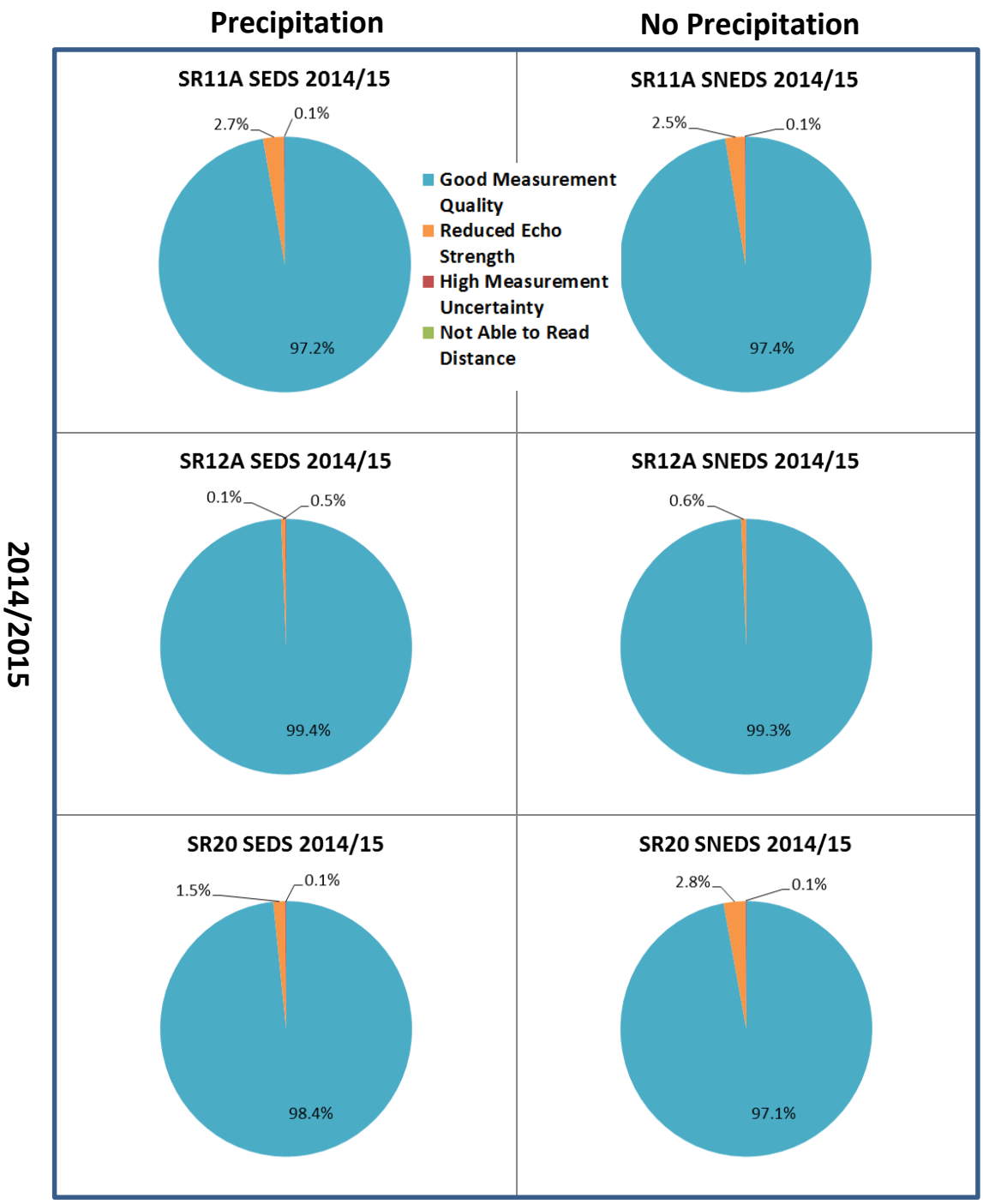


Figure 4.85. Frequency distribution of measurement quality categories for three SR50A sensors at CARE over two seasons (2013/14 top, 2014/15 bottom) for precipitating events defined by the SEDS and non-precipitating events defined by the SNEDS.

The influence of increased wind speeds during the SR50A measurement seems to be negligible (analysis not shown). Using a 3 m/s threshold, the frequency of "Good measurements" is nearly identical for wind speeds (as measured at 2 m) above and below the threshold. Wind-speed thresholds of 5 m/s and 8 m/s were also explored with no negative impact on the quality numbers.

4.1.4.3.1.5 Linking SR50A quality numbers to data quality control

The analysis presented in Sections 4.1.4.3.1.1 to 4.1.4.3.1.4 provides some indication as to what influences the quality numbers generated by the SR50A/TH for each measurement. However, it is unclear how to make use of these numbers for the purpose of quality control. To examine the utility of using these metrics for quality control, the instrument quality number output was cross referenced with the flags generated by the SPICE quality control process (as described in Section 3.3.3). In theory, there should be significant overlap between measurements that the sensor indicates have "Reduced Echo" and "High Uncertainty" with measurements flagged as erroneous or suspicious by the SPICE quality control process.

For each SR50A/TH sensor at each site (Sodankylä, CARE and Formigal), contingency tables were generated to cross-reference the instrument quality numbers with the SPICE QC flags. Contingency tables show the frequency of occurrence when the quality numbers agree and disagree with the QC flags.

Using all of the available data for each sensor, Table 4.21 shows when the SR50A/TH "Good measurement" quality number data agrees or disagrees with the SPICE QC "Good" data. A large percentage of the data (< 72%) is flagged as "Good" by both processes and only a very small percentage (< 1%) is flagged as "Not Good/Suspicious" by both processes. This suggests that there are large discrepancies between the two processes. The largest discrepancy is at Formigal, where 27% of the data has a "Not Good" quality number, but is not being flagged as "Suspicious" or "Erroneous" by the QC process. Conversely, the percentage of data output as "Good" by the sensor and then identified as "Suspicious" by the QC process is less than 0.3% (usually considerably less).

Table 4.22 shows the agreement between the quality numbers and QC process when the SR50A/TH reports a "Good" quality number. The agreement between "Good" quality number and data flagged as "Good" by the QC process is very high. With the exception of Formigal at 99.6%, the frequency of agreement is higher than 99.8%. Table 4.23 and Table 4.24 show the frequency of agreement when the SR50A/TH reports "Reduced Echo" and High Uncertainty" respectively. Table 4.23 shows that less than 1% of the data output by the SR50A/TH with "Reduced Echo" is flagged as suspicious by the data QC process. This percentage starts to increase when the sensor outputs data as "High Uncertainty" such that up to 11.4% of the data output as "High Uncertainty" is flagged as suspicious by the data QC. However, the frequency of this is generally below 4%.

Table 4.21. Contingency table describing the agreement between the SR50A/TH quality number output (Good and Not Good) and the SPICE QC process (Good or Suspicious).

Site/Sensor		QC Flag "Good"	QC Flag "Suspicious"
Sodankylä 6062	SR50A/TH "Good"	93.68%	0.00%
	SR50A/TH "Not Good"*	6.29%	0.03%
Sodankylä 7052	SR50A/TH "Good"	89.14%	0.16%
	SR50A/TH "Not Good"*	10.55%	0.15%
Formigal	SR50A/TH "Good"	72.06%	0.26%
	SR50A/TH "Not Good"*	26.68%	1.00%
CARE 11A	SR50A/TH "Good"	97.49%	0.03%
	SR50A/TH "Not Good"*	2.44%	0.04%
CARE 12A	SR50A/TH "Good"	99.19%	0.06%
	SR50A/TH "Not Good"*	0.74%	0.01%
CARE 20	SR50A/TH "Good"	97.08%	0.04%
	SR50A/TH "Not Good"*	2.84%	0.04%

*"Not Good" refers to SR50A/TH quality numbers that indicate either a reduced echo or high uncertainty.

Table 4.22. Contingency table for data output as "Good" by the SR50A/TH and the agreement with the SPICE QC process (Good or Suspicious).

Site/Sensor		QC Flag "Good"	QC Flag "Suspicious"
Sodankylä 6062	SR50A/TH "Good"	100.00%	0.00%
Sodankylä 7052	SR50A/TH "Good"	99.82%	018%
Formigal	SR50A/TH "Good"	99.64%	0.36%
CARE 11A	SR50A/TH "Good"	99.97%	0.03%
CARE 12A	SR50A/TH "Good"	99.94%	0.06%
CARE 20	SR50A/TH "Good"	99.96%	0.04%

Table 4.23. Contingency table for data output as "Reduced Echo" by the SR50A/TH and the agreement with the SPICE QC process (Good or Suspicious).

Site/Sensor		QC Flag "Good"	QC Flag "Suspicious"
Sodankylä 6062	SR50A/TH "Reduced Echo"	99.68%	0.32%
Sodankylä 7052	SR50A/TH "Reduced Echo"	99.08%	0.92%
Formigal	SR50A/TH "Reduced Echo"	98.68%	1.32%
CARE 11A	SR50A/TH "Reduced Echo"	99.01%	0.99%
CARE 12A	SR50A/TH "Reduced Echo"	99.14%	0.83%
CARE 20	SR50A/TH "Reduced Echo"	99.33%	0.67%

Table 4.24. Contingency table for data output as "High Uncertainty" by the SR50A/TH and the agreement with the SPICE QC process (Good or Suspicious).

Site/Sensor		QC Flag "Good"	QC Flag "Suspicious"
Sodankylä 6062	SR50A/TH "High Uncertainty"	99.04%	0.96%
Sodankylä 7052	SR50A/TH "High Uncertainty"	95.95%	4.05%
Formigal	SR50A/TH "High Uncertainty"	88.61%	11.39%
CARE 11A	SR50A/TH "High Uncertainty"	97.27%	2.72%
CARE 12A	SR50A/TH "High Uncertainty"	99.04%	0.96%
CARE 20	SR50A/TH "High Uncertainty"	97.09%	2.91%

4.1.4.3.1.6 Summary and Conclusions

At the Alpine sites with high instrument installation heights, the sensor quality numbers indicate increased measurement uncertainty during the occurrence of precipitation, which is exacerbated by high precipitation rates. This increase in uncertainty may result from snow falling into the sensor's signal path, or as the manufacturer states in the SR50A/TH manual, decreased density of the target snow at the surface. Formigal shows a decrease in measurement quality during spring melt, which could be an indication of a rough and uneven surface target that develops during rapid snow melt. At Sodankylä, there is some degradation in signal quality with the occurrence of precipitation (and increased precipitation rate) but the impact is not as significant as that observed for Formigal and Col de Porte. At CARE, there is some degradation in signal quality during precipitation, but this appears to be less of a factor here than at Sodankylä, with very little change due to precipitation rate. The impact of precipitation on the quality numbers at Sodankylä and CARE is lower relative to the Alpine sites, most likely due to the sensors being installed closer to the target.

Based on this analysis, it is evident that the sensor's internal algorithms and the resulting quality numbers indicate greater uncertainty in the measurements in some situations. In order to make recommendations on how to use the sensor quality numbers in a quality control process, the frequency distributions of the output categories were cross-referenced with the SPICE quality control process. In total, most of the data (> 72%) were flagged as "Good" by both processes. A very large percentage (> 99%) of the data output as "Good" by the sensor was also identified as "Good" by the QC process. There is also a large proportion of the data output as "Reduced Echo" by the sensor that was still flagged as "Good" by the QC process (> 98%). Finally, more than 88% of the data output as "High Uncertainty" by the sensor was still flagged as "Good" by the QC process. In fact, with the exception of Formigal, this percentage was over 96%. This suggests that even though data output by the sensor is identified as "Reduced Echo" or "High Uncertainty", that data is not necessarily erroneous. If a user was relying on the quality numbers alone, a substantial amount of data that would pass a QC procedure similar to that used for SPICE would be omitted from the data set. Following this, if a user only employed the quality numbers for quality control, only a very small percentage (< 1%) of erroneous data would be included in the quality controlled data set. Based on this analysis, it would be advisable for an instrument user to only use the quality numbers for guidance when developing a quality control process. It would be highly desirable if instrument manufacturers could provide more guidance on how to best use these instrument derived quality numbers to improve the assessment of measurement quality. More analysis is recommended on this topic using the SPICE SR50A data set to explore the impact of other quality control protocols on snow-depth measurements and data quality.

4.1.4.3.2 *The SHM30 signal strength output*

The purpose of examining the behavior of the SHM30 signal strength output is not necessarily to improve data quality control (as with the SR50A/TH quality numbers), but rather to attempt to extend the sensor's capability of detecting the first snow on the bare ground or target and for determining when the ground or target is snow free. According to the manufacturer, signal strength depends on target brightness, distance to the target, and sensor temperature. Although perhaps intended as a diagnostic tool, de Haij (2011) demonstrated the utility of using the signal strength output as an indicator for new snow on a bare target and used this information as a quality control parameter to remove false alarms from the snow-depth observations. This prompted SPICE to explore this utility and report on some of the noteworthy results. It should be noted, however, that the firmware version of the sensor impacts the behavior of the signal strength output. The firmware

versions for Sodankylä, CARE, and Col de Porte were 9.07, 9.08, and 9.06 respectively. According to the manufacturer, the biggest impact of firmware updates would have occurred prior to version 9.05. All firmware versions employed in SPICE were newer than version 9.05, and so are not expected to impact significantly the signal strength output.

The SHM30 sensors for SPICE are installed at Col de Porte (1 sensor), Weissfluhjoch (1 sensor), Sodankylä (1 sensor) and CARE (3 sensors). The surface targets at Col de Porte and Weissfluhjoch are natural ground, consisting of rock and natural grass. The targets at CARE and Sodankylä are the same as described previously; textured plastic at CARE and artificial turf at Sodankylä. Because each of these targets has different optical properties, it is expected that the signal strength output will vary depending on both the target and the surface conditions. This analysis focuses on the behavior of the signal strength output as surface conditions change during the transition seasons.

4.1.4.3.2.1 Sodankylä

The SHM30 at Sodankylä targets a small area on the artificial turf target, as shown in Figure 4.86. During the snow free period in the fall of 2013, the average signal strength output from the sensor was 0.89. This increased to 1.03 after final snow melt in 2014 and increased again to an average of 1.74 during the snow free period leading up to snowfall in the fall of 2014. The average then dropped to 1.33 following melt in the spring of 2015.

Figure 4.87 shows the time series of both the signal strength and snow depth at 1-minute resolution for the 2013 transition season. Unfortunately, web camera photos of this target were not available until November. However, jumps in signal strength are observed that correspond with small increases in snow depth, indicating that a small change in surface optical properties produces a significant response in the signal strength. The chances that the laser beam is significantly reflected/refracted by hydrometeors in its path are very low. Hence, this potential disturbance of the signal can be neglected. This can be seen in Figure 4.88, which shows the behavior of the output for the first snowfalls of the 2013/14 winter.

Figure 4.86. The SHM30 and artificial turf target during the snow free period at Sodankylä.

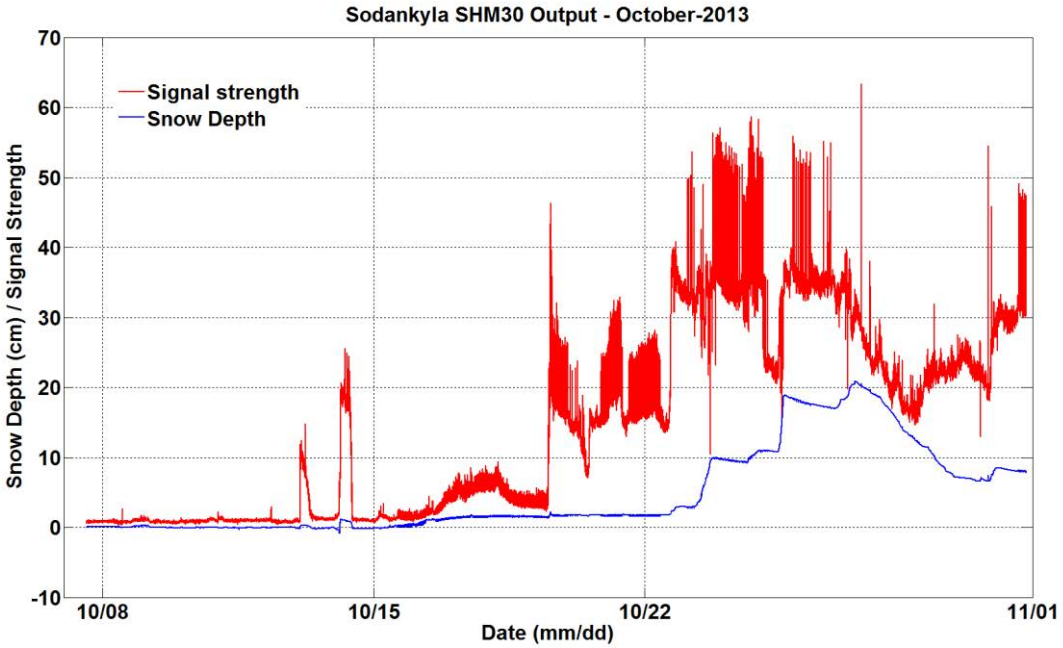


Figure 4.87. SHM30 output of signal strength and snow depth at Sodankylä for October 2013.

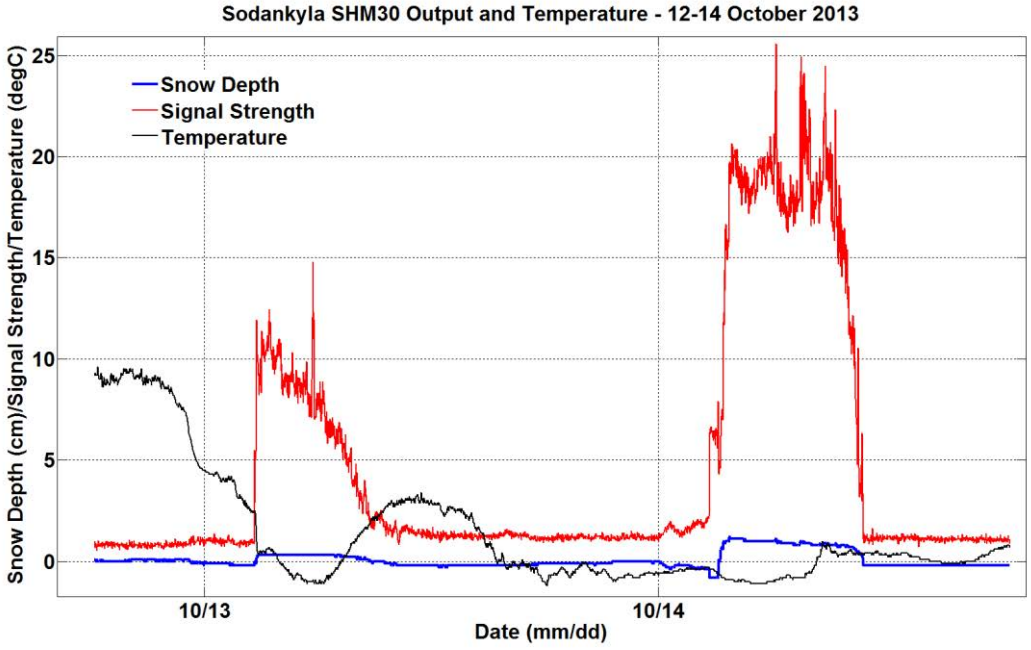


Figure 4.88. SHM30 snow depth and signal strength with air temperature for the first snowfall in October of 2013 at Sodankylä.

There are no webcam photos of the SHM30 target for October 13-14, but photos of the surrounding area show light snowfall (i.e. resulting in changing the ground optical properties, but less than a measurable mount) on the ground during this period corresponding with the jump in signal strength early on October 13 and then again early October 14 (Figure 4.88). As temperature increases later in the day on October 13, the snow melts off of the target and there is a corresponding decrease in the signal strength output as the target becomes snow free under the sensor. The same occurs after melting on the following day. These changes in signal strength can result from very small changes in snow depth (< 1 cm).

Similarly, corresponding with the first snowfall in 2014 (Figure 4.89), we see the rapid increase in signal strength related to small increases in snow depth during a mixed precipitation event (as shown by the present weather sensor output in Figure 4.90) as a result of a snow and mixed precipitation event on October 10 through October 12. The series of webcam photos (Figure 4.91) show the changes in surface optical properties due to the early morning snowfall on October 11 (Figure 4.91a), the subsequent melting that occurs through the day of October 11 (Figure 4.91b), the new snow that occurs later in the day on October 11 and into October 12 (Figure 4.91c), and the following melt (Figure 4.91d). Again, changes in the actual snow depth are small, but small changes related to new snowfall and subsequent melting result in large changes in the SHM30 signal strength.

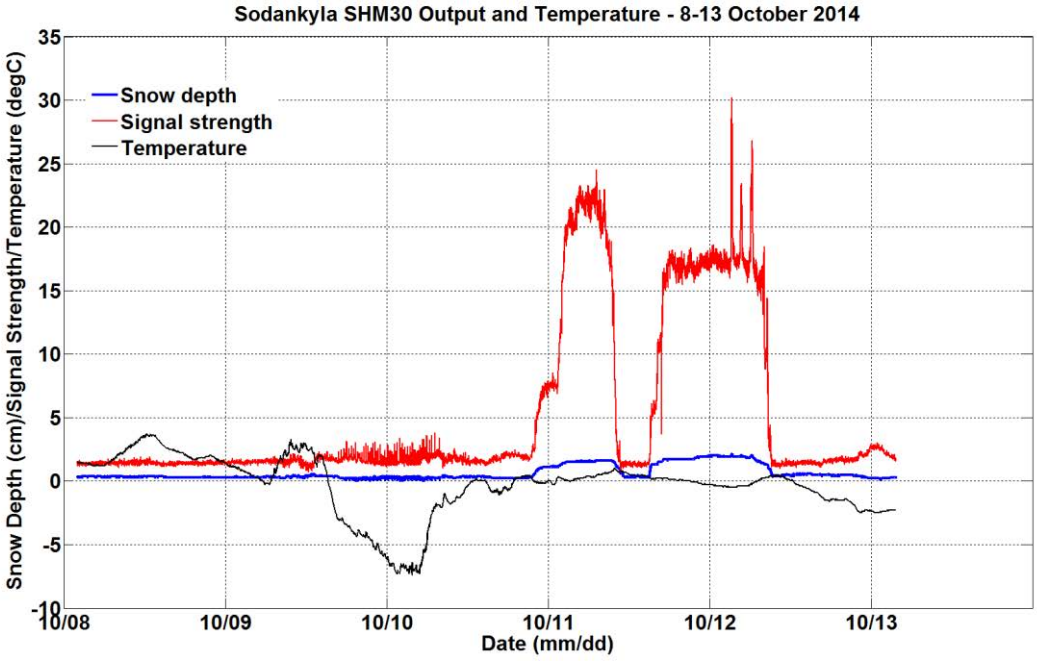


Figure 4.89. SHM30 snow depth and signal strength with air temperature for the first snowfall in October of 2014 at Sodankylä.

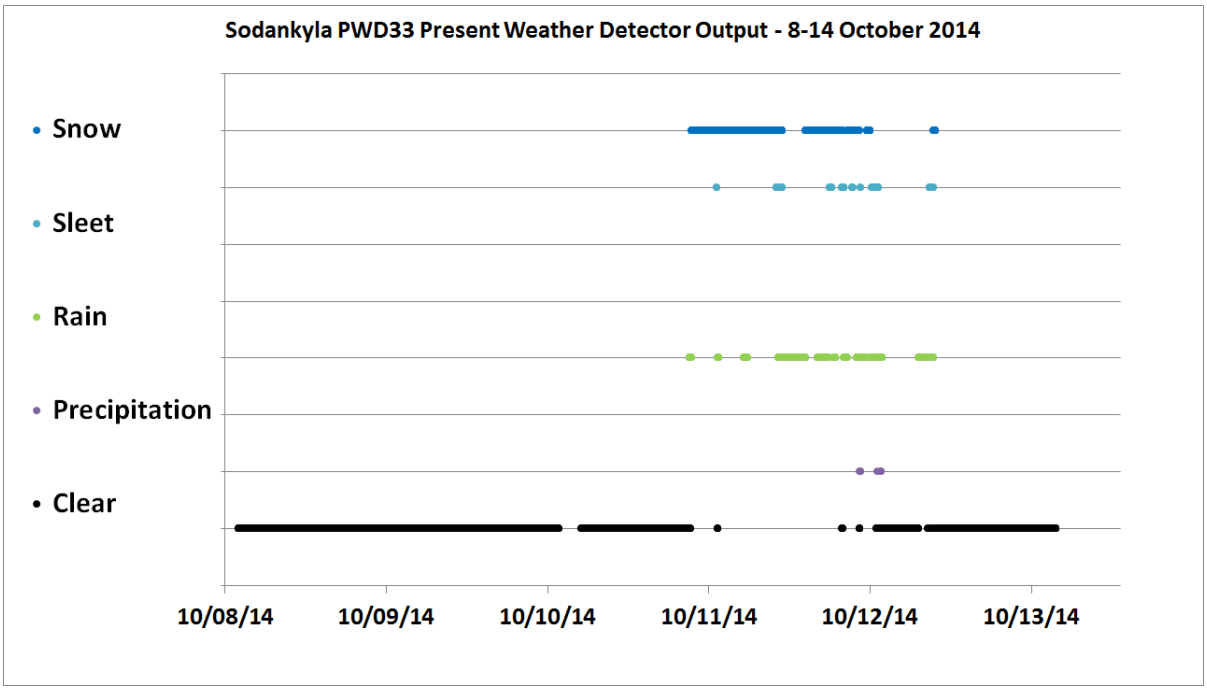


Figure 4.90. Precipitation type as observed by a PWD33 present weather detector during a precipitation event at Sodankylä, October 10 to 12, 2014.

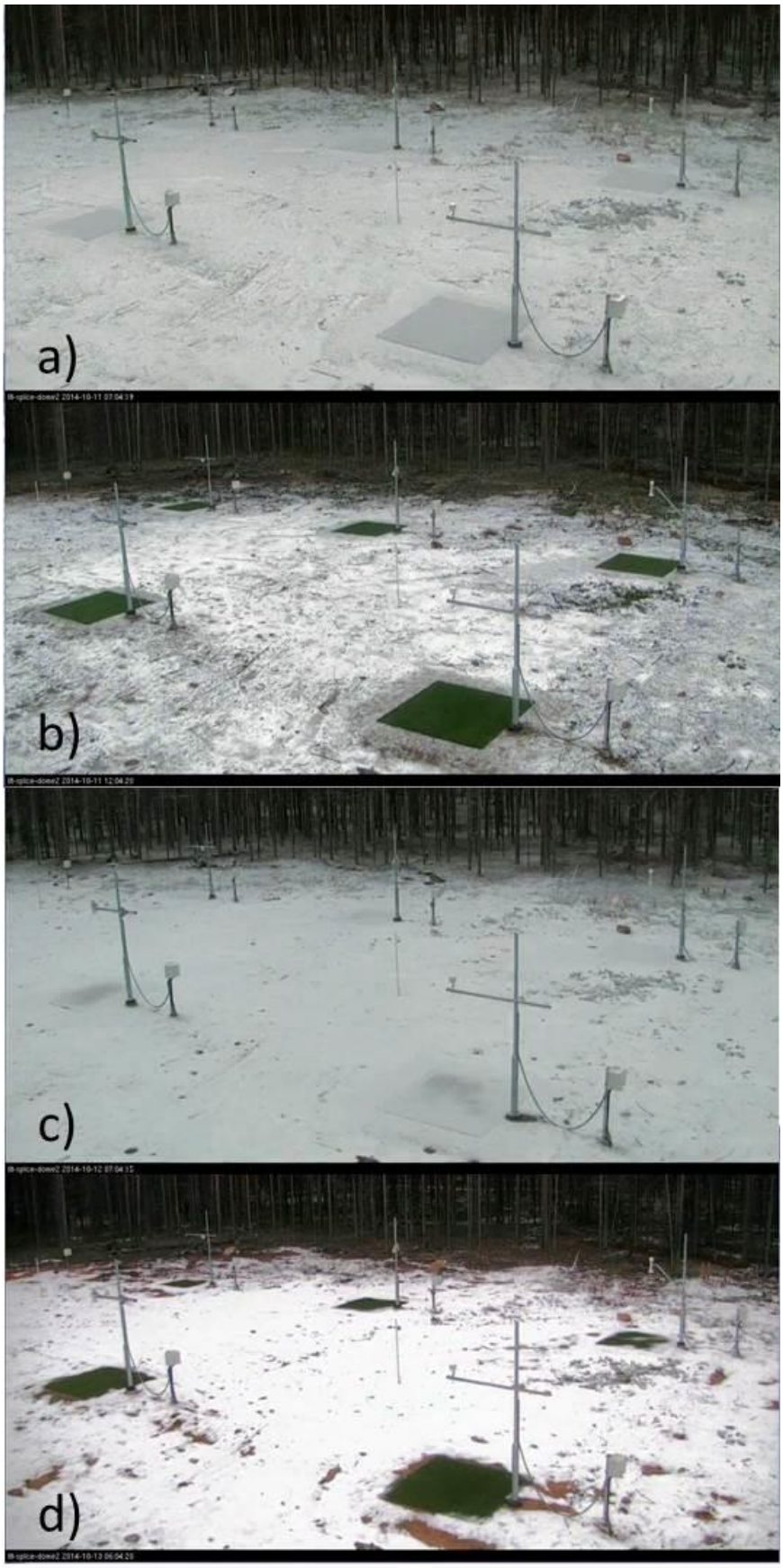


Figure 4.91. Webcam photos of the Sodankylä snow-depth targets in October 2014 a) Oct 11, 0700 UTC, b) Oct 11, 1200 UTC, c) Oct 12, 0700 UTC, and d) Oct 13, 0600 UTC.

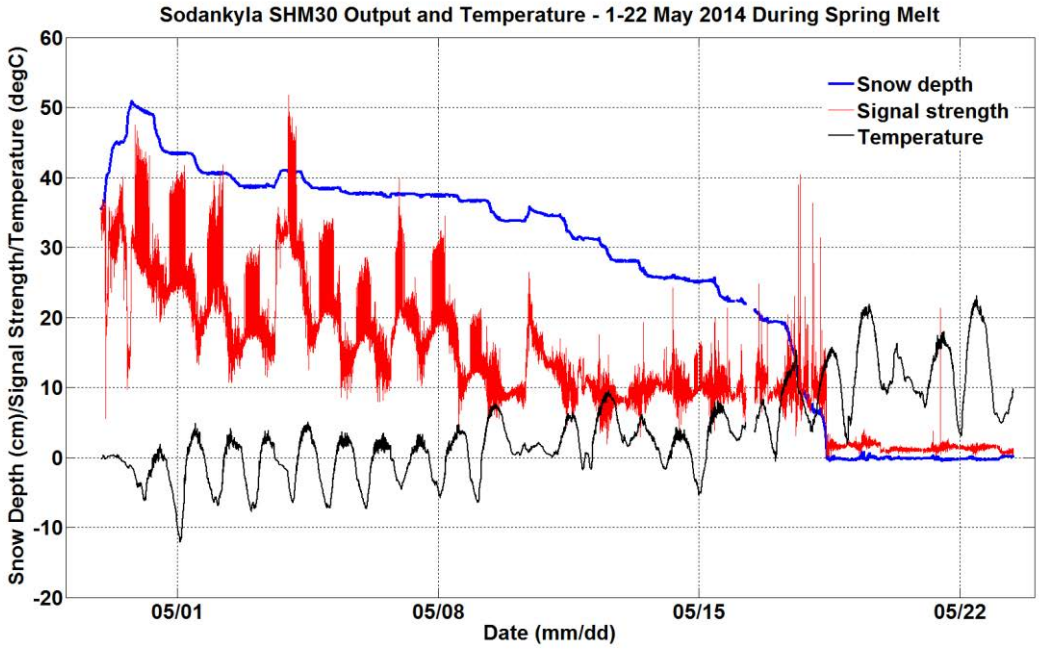


Figure 4.92. SHM30 snow depth and signal strength with air temperature for the seasonal snow melt period in May 2014 at Sodankylä.

The signal strength output from the SHM30 during seasonal melt is useful for determining when the target area is snow free, or similar to fall, when spring snowfall intermittently covers the target area. At Sodankylä, it is apparent from the signal strength response when the target is snow free. Figure 4.92 illustrates this response through the melt period from May 1, 2014, through May 22, 2014, and Figure 4.93 for the melt period from May 1, 2014, through June 1, 2015.

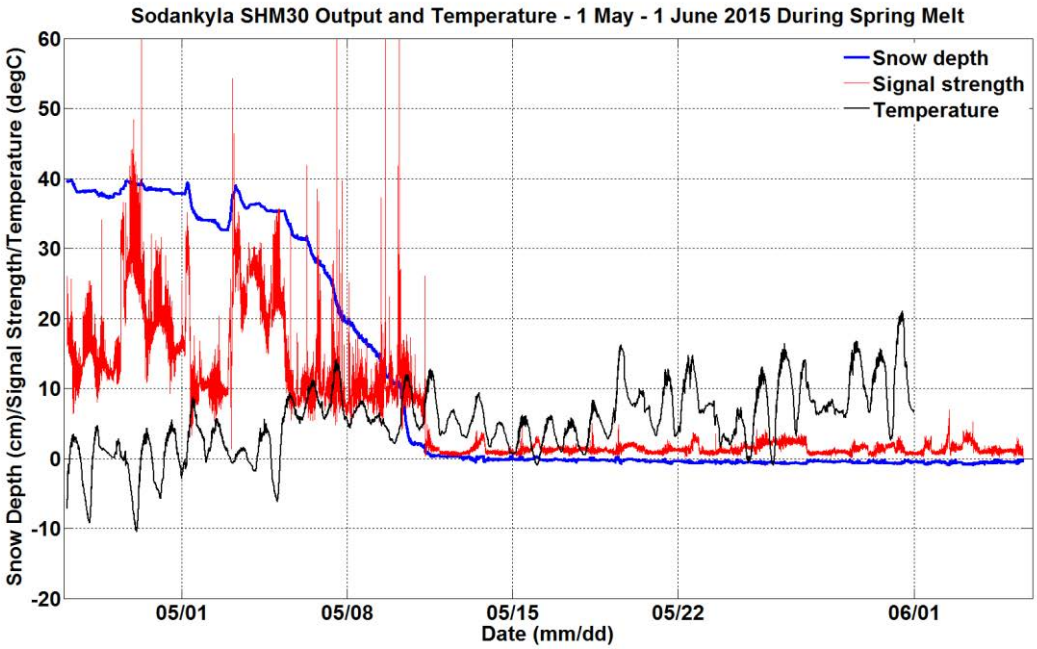


Figure 4.93. SHM30 snow depth and signal strength with air temperature for the seasonal snow melt period in May 2015 at Sodankylä.

Signal strength during the 2014 melt gradually decreases as the snow pack melts, becoming relatively stable at around 10 before dropping rapidly to a value less than 2 when the target is snow free by 10:15 UTC on May 18. The same thing happens in 2015, with the target becoming snow free on May 11, with the rest of the site following a few days afterwards. This progression is shown by the webcam photos during melt (Figure 4.94).

Figure 4.94. Webcam photos of the Sodankylä snow-depth targets during the 2015 melt on a) May 11, 0500 UTC, b) May 11, 1000 UTC.

Another notable result is the increase in the noise present in the signal strength output with the activation of the sensor heater. A good example is seen in Figure 4.89 where the heater activates with a temperature drop mid-day on October 9, resulting in increased noise in the signal. There also appears to be a pronounced diurnal trend in the signal strength output that happens during seasonal melt. This is prominent and well-illustrated in Figure 4.92. This increased noise may reduce the capability of this sensor model (with this version of firmware) to identify new snow on a bare surface.

4.1.4.3.2.2 CARE

The SHM30 at CARE targets a small area on a grey textured plastic platform as shown in Figure 4.95. The optical properties of the plastic targets are different than for the artificial turf targets used at Sodankylä. This results in a much higher signal strength output during snow free periods. The

instruments at CARE were not reporting at the beginning of the 2013/14 winter season, making analysis of the first snowfall of the season impossible. Following this, the signal strength did not behave in the same way at CARE as it did at Sodankylä. Figure 4.96 shows the time series of snow depth, signal strength and temperature during the melt period in late March of 2014. Signal strength starts at about 20 on March 26 and appears to drop during melt to a minimum of under 5 by March 29. However, webcam photos (Figure 4.97) and sensor measured snow depths (Figure 4.96) show that there is still snow on the targets during this time. When the target appears to be finally snow free early on April 5, the signal strength jumps from 5 to about 12.

Figure 4.95. Photo of the snow-free target under the snow-depth sensors at CARE. The area outlined in red shows an SHM30 sensor and its target.

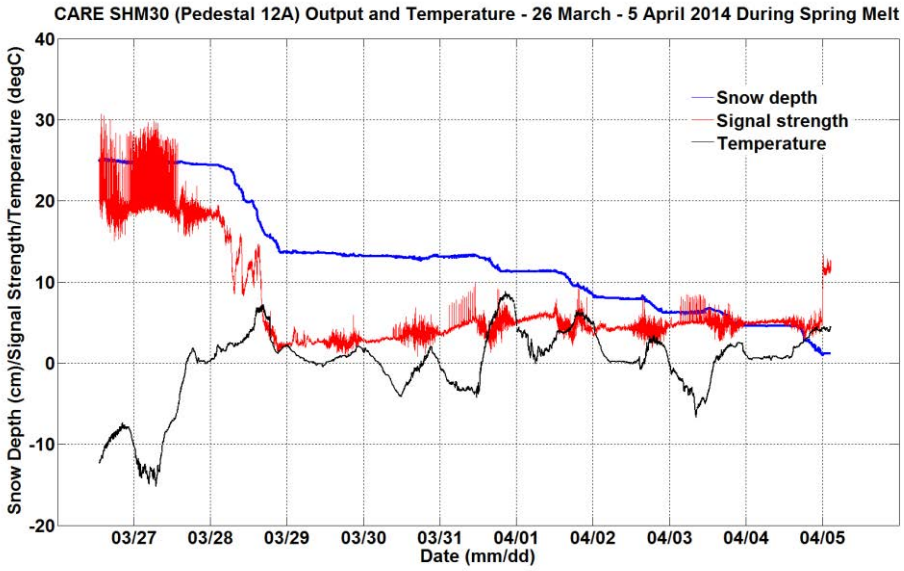


Figure 4.96. SHM30 snow depth and signal strength with air temperature for the seasonal snow melt period in May 2014 at CARE.

Figure 4.97. Webcam photo of pedestal 12A at CARE, March 28, 2014, 2200 UTC.

The reaction of the signal strength output to the first snowfall of the 2014/15 season also differs from what was observed at Sodankylä. Figure 4.98 shows the behavior of the sensor for the event that occurred on November 14, 2014. With no snow on the target, the baseline signal strength is between 12 and 15 (consistent with the final spring values shown in Figure 4.96) and then drops slowly to under 10 during the 8 cm snowfall event.

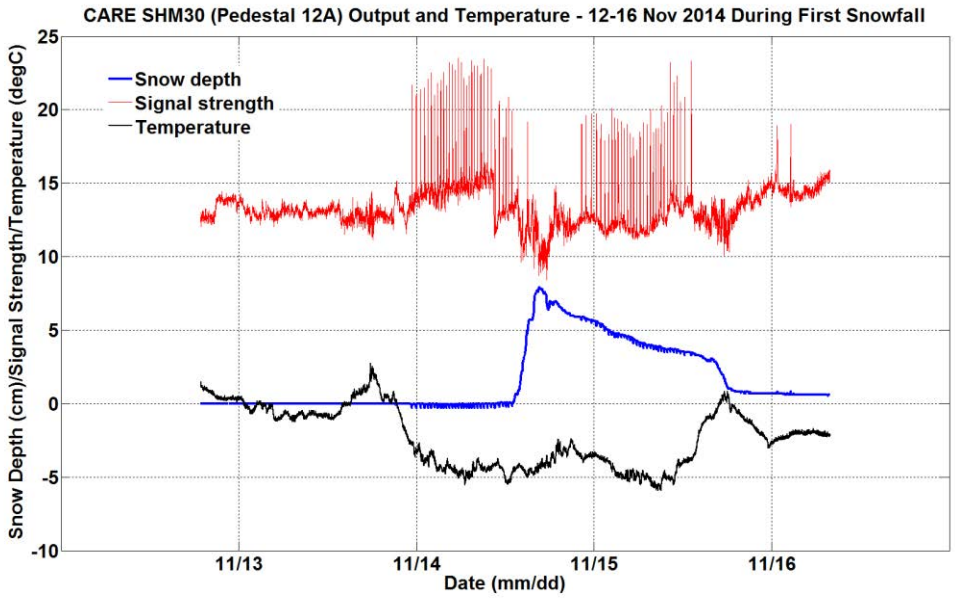


Figure 4.98. SHM30 snow depth and signal strength with air temperature for the first seasonal snowfall in November 2014 at CARE.

The behavior of the sensor in the spring of 2015 is consistent with the spring of 2014. Figure 4.99 shows the behavior of the sensor as the snow disappears from the target in 2015. The signal strength starts high (15 to 20), as it did in the spring of 2014, drops as the snow is melting, and then appears to jump to a value between 12 and 15 when the target is snow free, due to the changes in the optical properties of the target. Note that the snow depth zero value seems to have drifted, as the instrument shows 2-3 cm of snow into April when the target is bare by 2200 UTC on March 22 (Figure 4.100b). Figure 4.100a shows a light dusting of snow or frost on the target early in the day of March 22, which may be the cause of the spikes in signal strength through that day in Figure 4.99.

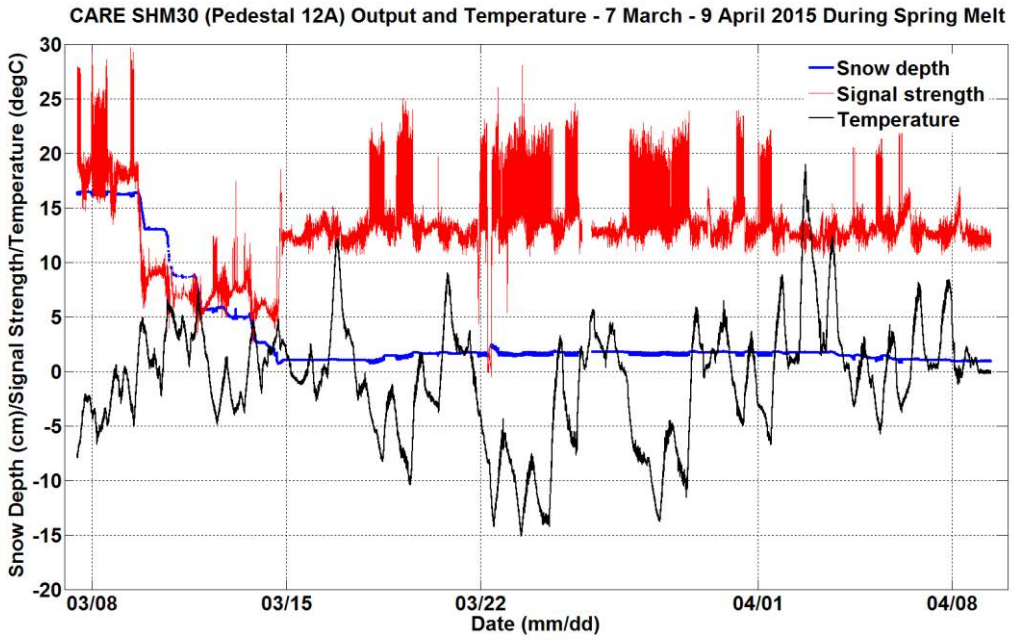


Figure 4.99. SHM30 snow depth and signal strength with air temperature for the spring melt period in March/April 2015 at CARE.

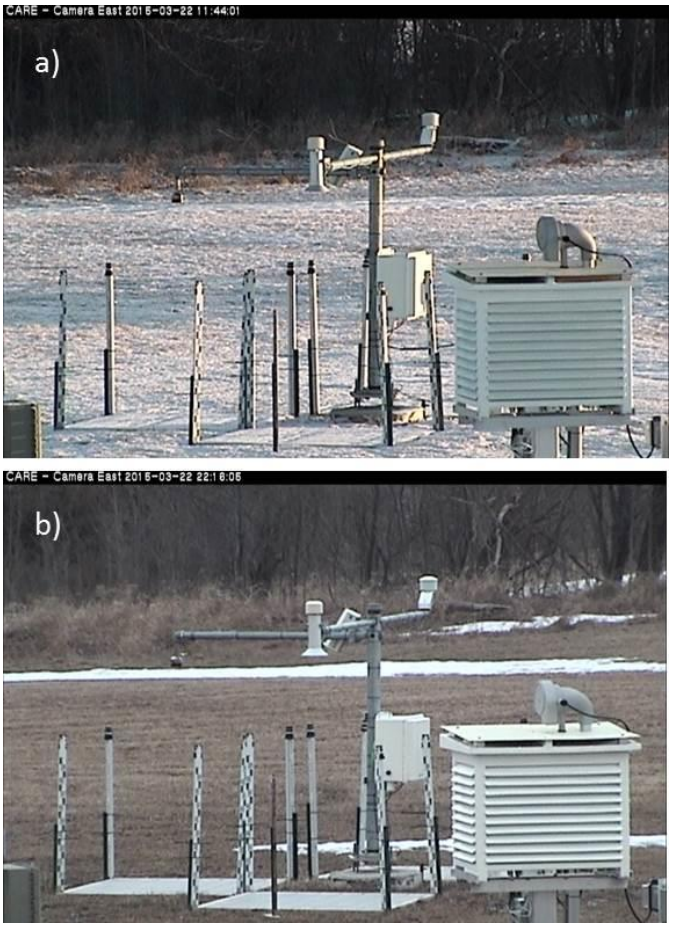


Figure 4.100. Webcam photos of pedestal 12A at CARE on March 22, 2015 at a) 1200 UTC and b) 2200 UTC.

4.1.4.3.2.3 Col de Porte

The SHM30 installation at Col de Porte is on a much higher mount (4 m) than at CARE and Sodankylä due to the deeper alpine snow packs. The other difference at this site is that the sensor target is natural grass, as shown shortly after becoming snow free in Figure 4.101. It is speculated to behave more like the artificial turf at Sodankylä than the plastic targets at CARE when it comes to the signal strength output. This appears to have been the case, as demonstrated in Figure 4.102. Further, there was less noise in the signal strength output relative to the other sensors tested, since the sensor at Col de Porte was not heated.

Figure 4.101. Photo of the target area of the SHM30 at Col de Porte in April 2015 during snow melt.

Figure 4.102 shows the first snowfall event at Col de Porte on November 4-6, 2014. The bare target signal strength is slightly less than 1 (comparable to Sodankylä) and jumps to about 6 with the occurrence of new snow, returning to baseline values after melt. The response to new snow on this natural surface is not as large as that observed for Sodankylä, but is larger and more discernible than the response to new snow on the plastic targets at CARE. Note that the zero-snow-depth offset is about 1.5 cm too high, resulting in a negative bare target snow depth.

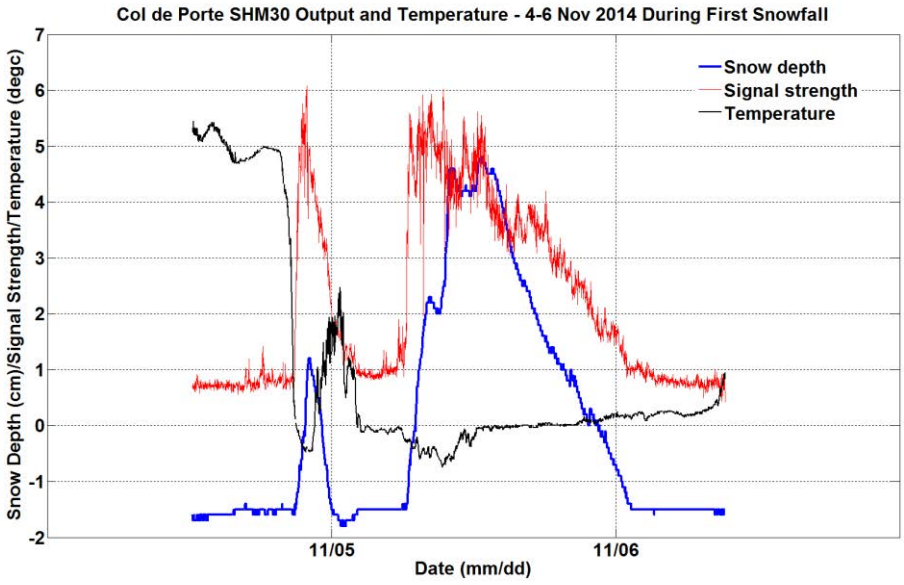


Figure 4.102. SHM30 snow depth and signal strength with air temperature for the first snowfall at Col de Porte in March 2014.

Examination of the evolution of the signal strength at the end of seasonal melt (Figure 4.103) shows a very subtle drop in signal strength when the target under the sensor is snow free. This occurs very early on April 20 and could be indistinguishable from other signal noise without ancillary information. Late day webcam photos of the target area on April 19 show a small hole in the snow where the SHM30 optical beam measures the surface target (Figure 4.104).

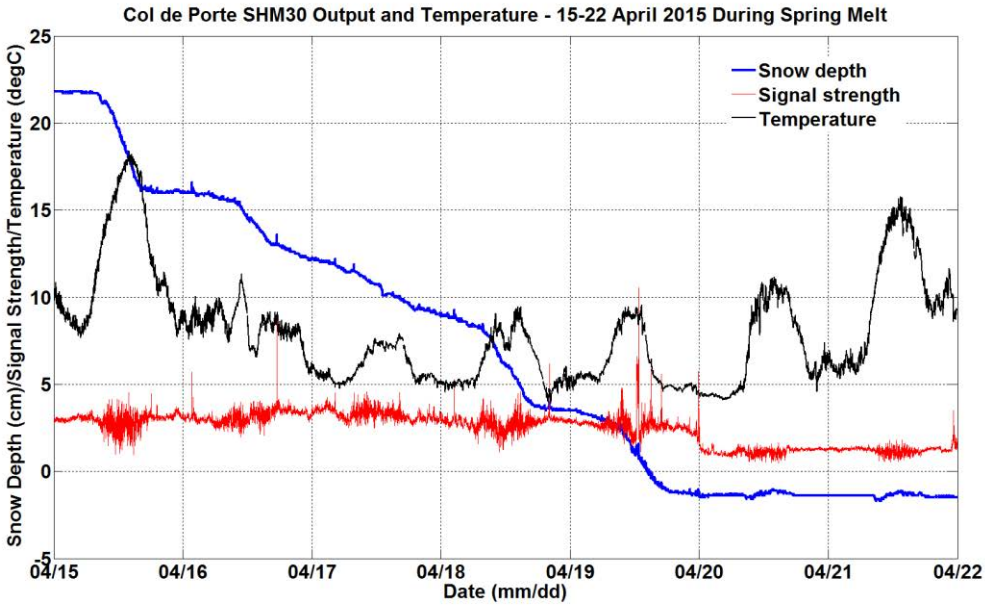


Figure 4.103. SHM30 snow depth and signal strength with air temperature during spring melt in April 2015 at Col de Porte.

Figure 4.104. Target area under the SHM30 (inside red circle) at Col de Porte, 1845 UTC on April 19, shortly before signal output drops to indicate a snow free target.

4.1.4.3.2.4 Conclusions and recommendations

An examination of the signal strength output from the SHM30 during the transition seasons at the SPICE sites demonstrated the capability of the sensor to identify light snow events on the bare target, even before snow depth became measurable within the uncertainty of the sensor. The capability to determine the time at which the target area is bare after snow melt was also explored. Results varied by site and by target type, and were likely influenced by the different firmware versions of the sensors tested.

- The artificial turf targets at Sodankylä created a surface with a high optical contrast between snow covered and snow free, resulting in rather distinct changes in the signal strength output. Signal strength jumped from a low baseline value during the first snowfall on the target and dropped from a higher value to the baseline value when the target was snow free.
- The light grey plastic targets at CARE produced a surface with optical properties very different than for the targets at Sodankylä, resulting in different signal-strength output behavior. This output was much more erratic, with baseline levels higher on the bare target than for a snow-covered target.
- Col de Porte, using natural grass as a target, had a signal strength response similar to Sodankylä, although more subtle, especially at the end of the season.
- The observations indicated that the SHM30 signal strength output could be useful for identifying the time of the first snow of the season and when the target is snow free at the end of the season, even when the snow depth is smaller than the measurable resolution of the sensor. Following this, the output could also be used for data QC to eliminate false accumulations due to the growth of grass under the sensor or from a shifting artificial target prior to snow accumulation.
- The grey plastic targets used at CARE resulted in a less distinct signal-strength response during the transition from bare target to new snow (and back to bare target) relative to that observed for the artificial or natural turf used at Sodankylä and Col de Porte. To use this capability in the sensor, the optical properties of the target need to contrast with those of a fresh snow-covered surface. Natural colors, such as browns, greens, or dark greys, provide a similar signal strength return as bare ground.
- The capability of the sensor to detect new snow was enhanced by the manufacturer in newer firmware releases (newer than 9.08, which was the most up-to-date version tested during SPICE). From the SPICE analysis, it is apparent that for older firmware versions (9.08 and older) it is important to understand how the sensor signal-strength output behaves at individual sites to reliably use this capability. This may not be necessary with newer versions of the sensor3.

4.1.4.4 SWE intercomparison summary

An intercomparison of the SWE sensors under test in SPICE (the CS725 and the SSG100) was undertaken using data collected at Sodankylä, Weissfluhjoch and Caribou Creek over the 2013/14 and 2014/15 measurement seasons. Additonal data from a non-SPICE site at Fortress Mountain, near Kananaskis, Alberta, Canada, were included (2013/14). This intercomparison is described in greater detail in Smith et al. (2017), and the following summarizes those results. The objective of this

[3] Please note that for versions 9.09 and newer, the SHM30 comes with a calibrated signal strength, which purportedly gives a more comparable indication whether there is snow under the sensor. The SHM30 sensors used in SPICE were all configured with older firmware versions.

intercomparison was to inform users of the best way to use these instruments and of any potential measurement issues that may influence data interpretation.

During the intercomparison, measurements from SWE instruments were compared to the manual reference measurements and cross referenced with ancillary measurements of air temperature, soil moisture, and soil temperature (at Caribou Creek) to try to determine causality for some of the bias seen in the results. Intercomparison results for the CS725 showed that it overestimates SWE, on average, by 30% and 35% at Sodankylä and Caribou Creek, respectively. The time series for Sodankylä and Caribou Creek are shown in Figure 4.105 and Figure 4.106, respectively. Correlations were higher at Sodankylä, with r^2 values ranging from 0.92 to 0.99, than they were at Caribou Creek, with r^2 ranging from 0.55 to 0.90. The difference in correlations was attributed to smaller sample size, higher spatial variability of SWE, and ice layers in the snowpack at Caribou Creek, which make the manual sampling more difficult and prone to error. Offsets were generally higher at Caribou Creek, potentially indicative of an inaccurate soil-moisture calibration of the instrument, a change in soil moisture relative to the calibration prior to or after the soil freezing, or systematic sampling errors in the manual SWE measurement due to a more complex snowpack.

Correlations at Fortress Mountain (non-SPICE site) were also quite high, with an r^2 of 0.94 and a mean negative bias of less than 5%. The time series are shown in Figure 4.107. The agreement between the CS725 and the manual SWE measurements were generally better at Caribou Creek and Sodankylä (both with sandy soil) prior to the start of seasonal melt than they were during the melt period. Seasonal melt appeared to have no significant impact on the agreement at Fortress Mountain, perhaps due to saturated frozen soils that restrict infiltration and a mild slope that promotes runoff of meltwater from the site.

The SSG1000 at Sodankylä compared quite well with the manual SWE measurements, showing a mean negative bias less than 11% and r^2 ranging from 0.84 to 0.99. Outliers are likely due to increased spatial variability in site SWE during melt or to errors associated with manually sampling the melting snowpack.

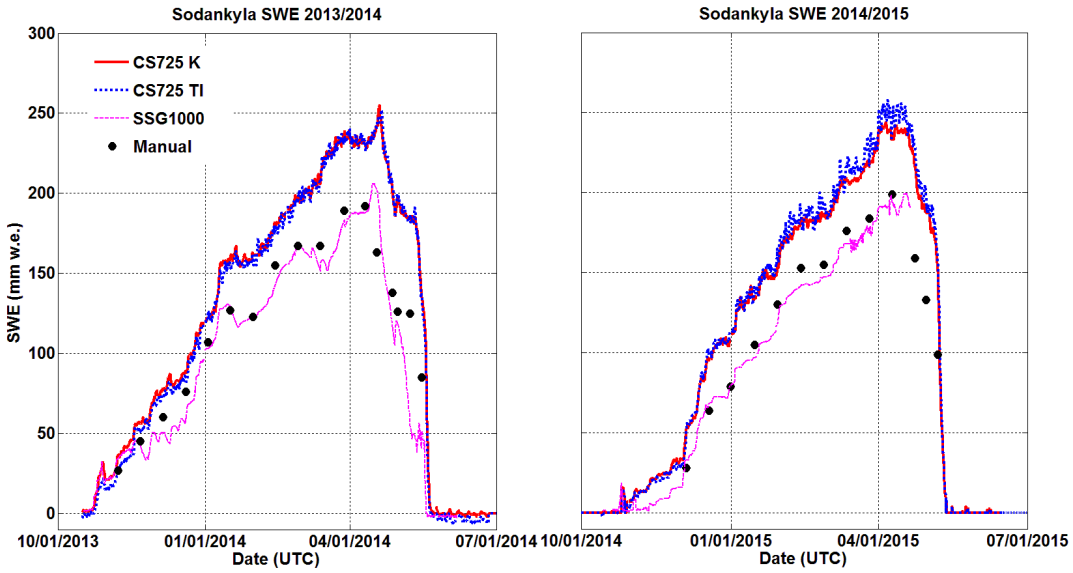


Figure 4.105. Time series of the SWE sensors (potassium or CS725 K output in solid red, thalium or CS725 Tl output in dashed blue, and SSG1000 output in dotted magenta) and manual SWE measurements at Sodankylä for the 2013/14 (left) and 2014/15 (right) seasons.

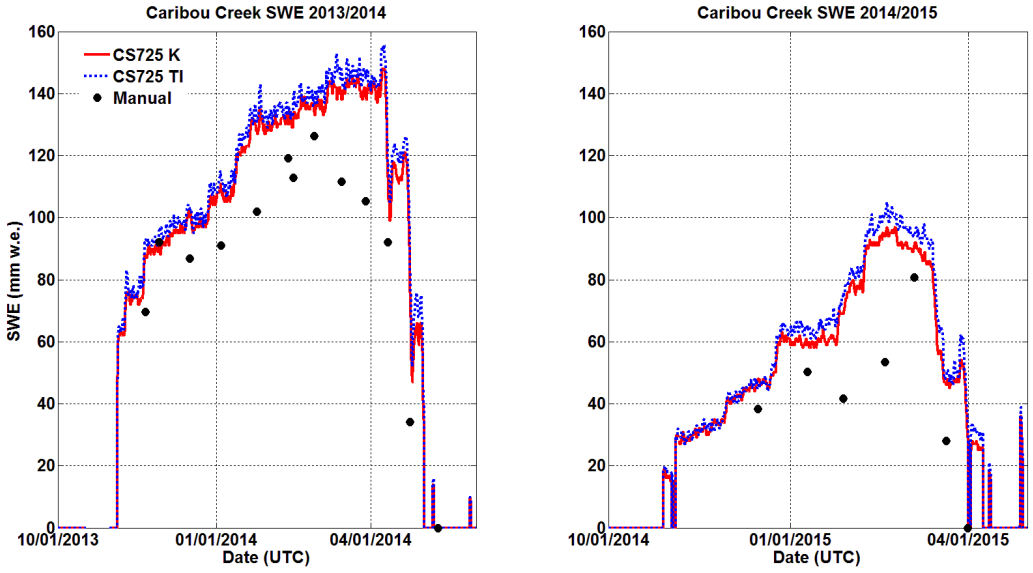


Figure 4.106. Time series of the SWE sensors (potassium or CS725 K output in solid red and thalium or CS725 Tl output in dashed blue) and manual SWE measurements at Caribou Creek for the 2013/14 (left) and 2014/15 (right) seasons.

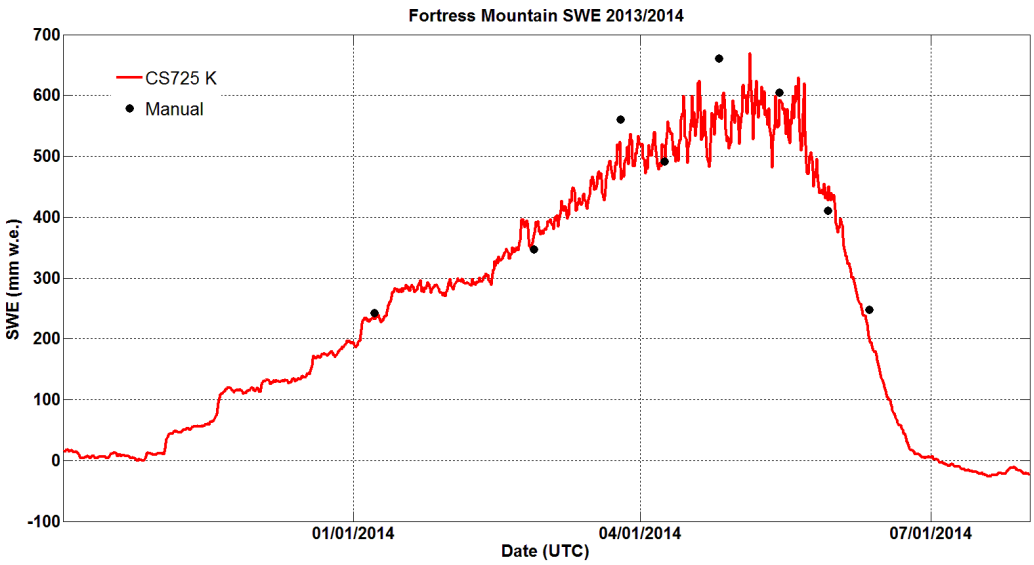


Figure 4.107. Time series of the SWE sensor and manual SWE measurements at Fortress Mountain for the 2013/14 season.

The SSG1000 at Weissfluhjoch also compared quite well to the manual snow pit-derived SWE with a combined r^2 of 0.96 for both seasons. The intercomparison between the manual and SSG1000 measurements, shown in Figure 4.108, was very close until mid-season, when the two measurements began to deviate from each other. It is believed that this resulted from snow bridging, which caused further accumulation of SWE at the site to not be measured by the weighing plate of the instrument.

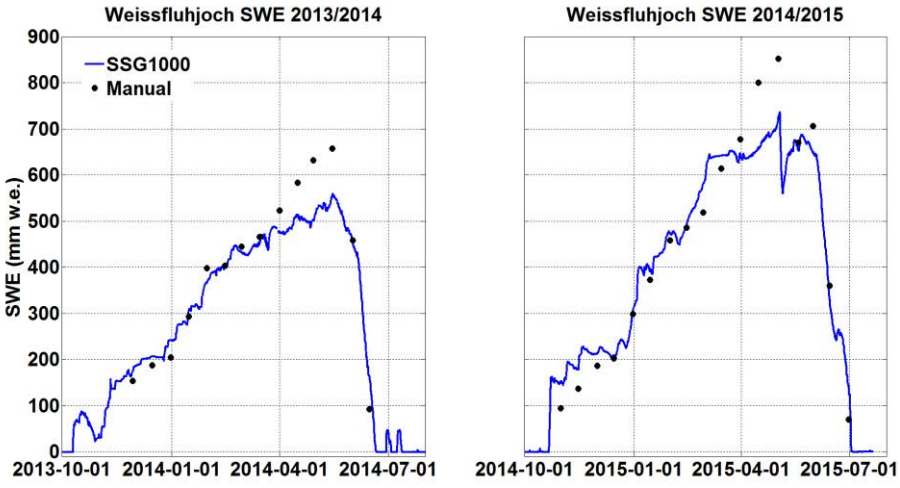


Figure 4.108. Time series of the SSG1000 and manual snow pit SWE measurements at Weissfluhjoch for the 2013/14 (left) and 2014/15 (right) winter periods.

An intercomparison of the SSG1000 with the CS725 at Sodankylä shows a linear relationship between the two instruments that deviates substantially during the melt period (Figure 4.109). Because of

technical issues, only the melt period for 2013/14 is available. At the onset of melt, the CS725 deviates substantially from both the SSG1000 and the manual measurements. This deviation most likely results from infiltration of melt water into the sandy soils at this site. While this meltwater drains away from the SSG1000 platform, the water is still available in the soil to attenuate the gamma radiation signal and, therefore, the CS725 still interprets this water as snow.

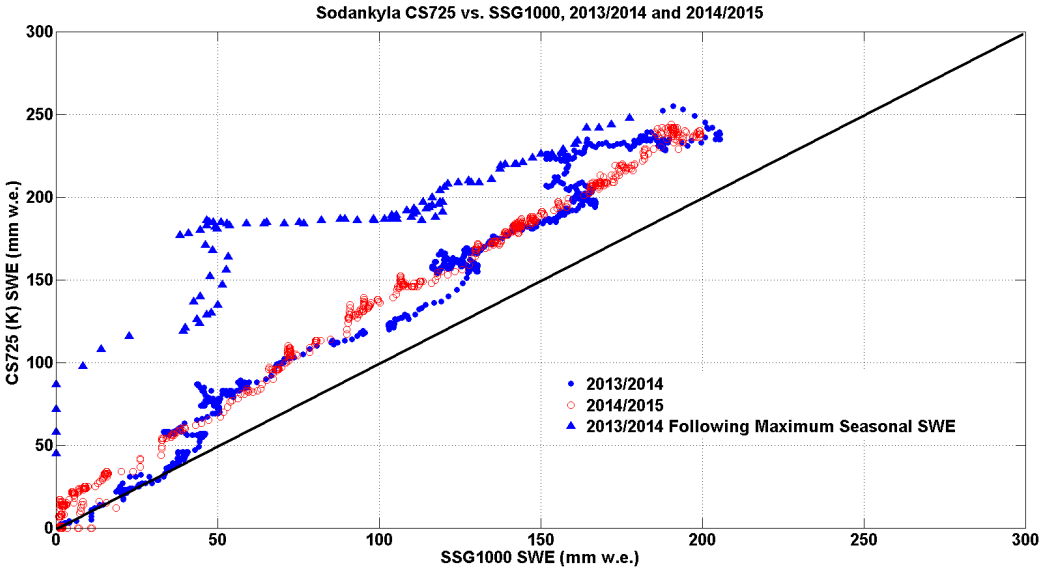


Figure 4.109. CS725 vs. SSG1000 SWE for the 2013/14 (blue circles and triangles) and 2014/15 (red unfilled circles) seasons at Sodankylä. The black line indicates the 1:1 line. Blue triangles mark the 2013/14 season after maximum seasonal SWE has been reached and melt has begun.

Some potential causality for the bias between the CS725 and manual SWE measurements was explored by qualitatively cross-referencing the occurrence and magnitude of the bias with air temperature, in an attempt to assess the impact of mid- and late-season snowmelt. At both Sodankylä and Caribou Creek, the bias seemed to increase substantially after the temperature rose above the freezing point over the course of the winter, although not all increases in the bias can be attributed to this. The bias, however, did not seem to increase substantially again at the onset of seasonal melt in March and April. There could be two possible mechanisms for this. The first mechanism could be the creation of ice layers in the snowpack as a result of freeze/thaw cycles that wouldn't impact the CS725 measurement, but would influence errors and bias in the manual SWE measurements. The second mechanism could be the formation of basal ice layers or infiltration of meltwater into the frozen, sandy soil. This basal ice layer would be difficult to sample accurately with a snow tube and could, therefore, result in an underestimation in the manual SWE observation.

At Caribou Creek, soil moisture/temperature data were used to qualitatively assess the impact of soil moisture change on CS725 measurements at the beginning of a winter season, when precipitation generally transitions from rain to rain/snow to snow. In theory, a change in soil moisture leading up to the soil freezing (freezing locks the moisture in place for the season) could impact the sensor's ability to assess the first snowfall event, and potentially perpetuate an offset throughout the season. Even though the CS725 at Caribou Creek seemed to correctly time the SWE accumulations and melt

during the fall transition in 2014, measured soil moisture did fluctuate during these events, as the soil was not frozen. However, it is difficult to ascertain if this caused the SWE overestimation for these events or, more significantly, if the change in soil moisture at transition resulted in the offset for the remainder of the season. During seasonal melting in March 2015 at Caribou Creek, a large increase in soil moisture related to the thawing of the surface and infiltration of meltwater was observed. This infiltration most likely contributed to the increase in bias of the CS725 sensor relative to the manual SWE measurements made in mid- and late-March 2015.

The conclusions of the SWE intercomparison are as follows:

- The manual measurements of SWE are not free of error, and experience has shown that bulk-density sampling in a snowpack containing ice layers or during melt using a snow tube is difficult and inherently prone to errors. Therefore, caution is required when using a bulk-density SWE measurement as a reference.
- Consideration must be given to the impact of spatial variability of the SWE being sampled and the relative footprints of the SUT.
- In this intercomparison, both sensors showed good agreement with the manual reference measurements; however, the potential for the CS725 to overestimate SWE measurements compared to the reference was identified. This could lead to misinterpretation, especially when deployed over sandy soils and during melting conditions when basal ice-layer formation or infiltration to soils can occur.
- Because of the CS725 measurement principle, it does not differentiate between changes in soil moisture (before freeze up in the fall and after thaw in the spring) or basal ice layers from changes in the SWE of the snowpack. Under these conditions, the CS725 can overestimate the SWE.
- Co-locating SWE instruments with ancillary measurements of soil moisture and temperature and snow depth could be helpful to guide the user in interpreting the dataset.

4.1.4.5 Emerging and alternative technologies for measuring snow on the ground

The WMO-SPICE project assessed performance factors for automated snow-depth measurements using ultrasonic (USH-8, SR50ATH, SL300) and laser (SHM30, FLS-CH 10) rangers, and for SWE measurements using either attenuation of terrestrial gamma radiation from the surface (CS725) or load cells (SSG1000). All of these instruments perform scalar measurements of snow-pack properties (depth or SWE) with various spatial scales of integration ranging from 1 mm^2 to a few square meters, depending on sensor installation height and settings. These measurements generally qualify as “point” measurements from a practical and data-reporting point of view (i.e. one measurement datapoint reported per instrument and per reporting time period).

Several alternative approaches have been developed over the past few decades to measure the two main snow properties (depth and SWE) at a point or distributed scale. Some of these measurements were provided by WMO-SPICE participants and implemented at some WMO-SPICE sites. This is particularly the case for the Republic of Korea Gochang site, which featured an alternative ultrasonic sensor (KMA04), a sensor based on optical assessment of snow depth using automated picture analysis (FX-D1), and two laser sensors with multiple point measurements (SDMS-100 and JH-20).

Given the spatial heterogeneity of snow cover at a variety of scales, the representative nature of a single point measurement remains questionable, especially in mountainous or windy environments (e.g. Grunewald et al., 2013). Measurement approaches capturing the spatial variation of snow depth

and SWE, or the scaling of these properties with respect to area, are increasingly being considered for research applications, and could be considered for implementation in operational networks if the instrument development and testing are sufficiently mature.

This section reviews existing methods used to measure snow depth and SWE. The scope of this brief review is restricted to stand-alone ground-based in situ instruments, keeping within the scope of WMO-SPICE. As defined by Kinar and Pomeroy (2015b), these are stationary devices that are noninvasive and rely on empirical or non-empirical relationships utilizing active or passive measurement techniques. Methods based on remotely sensed information (e.g. Nolin, 2010) or external sources of information from a snowpack evolution model are not covered here. For each of the methods reviewed below, their main expected strengths, weaknesses and considerations are provided. The review is not meant to be exhaustive, but rather seeks to capture existing and emerging technological approaches that may eventually develop for use in operational networks and could be considered in future intercomparisons. This review draws upon existing review papers (e.g. Kinar and Pomeroy, 2015b) and extends this information within the scope of the WMO-SPICE project.

4.1.4.5.1 *Point scale snow depth*

Beyond the technologies considered in the context of WMO-SPICE (laser, ultrasonic, and image analysis from cameras), there are currently no emerging or alternative technologies for snow-depth measurements. However, it is worth mentioning that research efforts have been focused on miniaturizing and decreasing manufacturing costs of existing sensors to allow bulk deployment of smaller units at potentially remote sites or to complement existing baseline sites that use established sensors. The strengths of such systems would lie in their low initial and deployment cost and low power consumption. Potential weaknesses would have to be addressed following detailed assessments of the performance of such systems once on the market. Considerations will necessarily include the sensitivity to environmental conditions, the quality of the ancillary measurements (e.g. temperature measurements for ultrasonic corrections), the robustness of such devices under potentially harsh conditions, and the implications for calibration, maintenance and lifecycle management.

4.1.4.5.2 *Point scale SWE*

The two SWE instruments assessed as part of the SPICE intercomparison were the CS725, a passive gamma-radiation sensor, and the SSG1000, a load-cell measurement of overlying mass. However, there are several emerging technologies for measuring SWE in situ that were not assessed during SPICE. These include technologies that use the propagation of electromagnetic waves, the propagation of sound, the attenuation of cosmic rays, and a combination of GPS and ground-penetrating radar installed under the snow pack. These systems are all considerably more technologically advanced than the snow pillow that was first used in the 1960s for making automated measurements of SWE (Beaumont, 1965) and are still being used today in networks such as SNOTEL in the United States (Serreze et al., 1999).

SWE can be derived from dielectric measurements of snow properties. An example of this measurement technique is the Sommer Snowpack Analyzer (SPA) system (Niang et al., 2003). This device, installed before the snow-accumulation season, uses flat coaxial cable antennas, one installed horizontally and another at an angle to the ground surface. As snow accumulates around the wires, electromagnetic waves are sent through the antenna that interact with the surrounding snowpack. This interaction is used to measure (dry and wet) density and liquid water content of the snow. An ancillary measurement of snow depth using a ranging sensor permits the calculation of SWE. The

system was evaluated by Egli et al. (2009) who showed an RMSE of 65 mm water equivalent (with maximum snowpack SWE approaching 600 mm w.e.) when comparing with manual snow tube measurements. One potential advantage of this system is that it allows the measurement of more than one property (i.e. SWE and depth) using one instrument. A potential disadvantage could arise from the cables that extrude into the snow pack and may create preferential flow pathways for melt water and thus lead to measurement errors.

Kodama (1979) and Paquet and Laval (2006) have developed and implemented instrumentation to measure SWE using attenuation of cosmic rays by the snow. This measurement device is currently applied in an operational network in France and Spain (Gottardi et al., 2013). The sensor is buried in the ground, thereby minimizing snowpack and environmental disturbances; however, this requires site preparation prior to the snowfall season. The cosmic ray counting device (3He tube and photomultiplier) has minimal power consumption, which makes it appropriate to deploy at remote locations. The main weakness of the system lies in the dwindling availability of 3He, leading to recent skyrocketing manufacturing costs of the photomultiplier tube. This system is currently not available on the retail market.

Kinar and Pomeroy (2015a) have recently introduced an automated system designed to measure bulk snow density, liquid water content, and temperature based on the propagation of acoustic waves through the snow. This device is called the System for Acoustic Sensing of Snow (SAS2). The acoustic system builds on previous research conducted on the acoustic properties of snow, and can be used to measure SWE. A one-layer model of the snowpack was proposed based on the Biot-Stoll theory of sound wave propagation through snow. Unlike previous research, this is a model that does not consider the snowpack to be a rigid-frame medium and does not rely on the assumption that sound waves from an acoustic source must be of sufficiently low frequency. Ancillary measurements of snow depth can be used to constrain the system outputs, particularly when the bottom of the snowpack cannot be identified due to high attenuation of sound waves in the snow. Evaluation of the system performance was carried out at several sites in the Rocky Mountains of Western Canada. Using 452 field measurements, Kinar and Pomeroy (2015a) estimated the mean bias and root mean square deviation (RMSD) of the SWE measurements to be 9 mm w.e. and 47 mm w.e. (where maximum SWE exceeded 400 mm w.e), respectively, as compared to ESC30 snow tube measurements of SWE. Strengths of the senor include the capability of making non-invasive measurements of the snow pack that can be used to determine SWE and internal snow properties. Weaknesses include the fact that power consumption of the prototype is high and, therefore, on-site AC power may be required for robust operation. Moreover, the sensor may not directly yield information that can be used in operational networks, especially if snow depth is not available for identifying the bottom of the snowpack when the maximum penetration depth occurs. The prototype system is currently not available on the market as a retail instrument. In order to evaluate the capability of this instrument for operational use, it would be beneficial to test this instrument under various environmental contexts, not only in terms of scientific performance, but also for evaluation of operational constraints including maintenance and instrumental lifecycle.

Schmid et al. (2015) reported on a sensing device comprised of upward-looking, ground-penetrating radar (upGPR) and a low-cost GPS system deployed beneath the snow pack to simultaneously measure snow-depth, density, liquid water content, and SWE. Separate from the SPICE experiments, the system was evaluated at Weissfluhjoch, Switzerland, during the 2012/13 and 2013/14 snow seasons using other standard in situ measurements at this site, and was found to agree well with

those measurements. Indeed the RMSD compared to continuous SWE data recorded with a snow pillow ranged from 30 to 50 mm w.e. and compared to manual snow-pit measurements ranged from 40 and 80 mm w.e. depending on the snow season considered (where maximum SWE approached 800 mm w.e.). Strengths include the fact that this sensor combination is buried in the ground, thereby minimizing snowpack and environmental disturbances. Therefore, this technique might also be useful, for example, at avalanche-prone slopes. Another advantage is that more than one snow parameter can be derived by the sensor combination. Weaknesses include the fact that it requires on-site power and in its current state may not directly yield information which could be used in operational networks. The prototype system is currently not commercially available. In order to evaluate its capability for operational use, it would be beneficial to test this instrument in different environmental conditions to evaluate its performance and operational considerations (maintenance, lifecycle).

4.1.4.5.3 *Spatially integrated snow depth*

Scalar measurements of snow depth that are representative of large areas can be derived from Global Navigation Satellite System (GNSS) signals following the seminal work of Larson et al. (2009) and following work by Jacobson (2010), Rodriguez-Alvarez (2012), Jin and Najibi (2014), and others. The GNSS Interferometric Reflectometry Method was implemented by Boniface et al. (2014) at 100 geodetic sites in the USA with a stated RMSE of 15 cm with respect to established methods of measuring snow depth. The measurement footprint of the reflectometry method is on the order of 1000 m^2 for shallow snow conditions, but this can vary as a function of the topography, elevation angle of the GNSS satellite (which varies during the day), and on the actual snow depth. The main strength of this approach is the capability to use existing geodetic networks for snow-depth measurements without the need to deploy new instrumentation. However, GNSS sites have generally not been chosen according to meteorological/snow on the ground considerations, which may hamper the representativeness and the usefulness of many of existing geodetic sites. An additional strength of the approach is the fact that it provides an integrated measurement rather than a point measurement, which should be able to better capture representative snow conditions around the site. However, the fact that the footprint area varies depending on the time of the day and the snow conditions can make the data difficult to interpret and analyze in combination with other snow-depth measurements from more traditional networks. In order to be usable operationally, such systems should be capable of providing not only snow depth, but also metrics representing the characteristics of the footprint area (size, variability within the footprint etc.). These should be made available as metadata along with archived snow-depth measurements.

4.1.4.5.4 *Spatially integrated SWE*

Spatially integrated measurements of SWE can be performed using a cosmic-ray soil-moisture probe (CRP) which was primarily developed for measuring average soil water content at the landscape scale (Zreda et al., 2008) but has also been demonstrated to be able to measure SWE (Desilets et al., 2010). Sigouin and Si (2016) present a performance analysis of the CRP at a site near Saskatoon, Saskatchewan, Canada, representative of a North American prairie and agricultural environment. The results of this study showed that the CRP can provide SWE measurements within a 300 m radius considered to be the footprint of the device. The strength of this approach is that it can potentially be used with existing CRP networks (e.g. the COSMOS network in the US, primarily dedicated to ground water content monitoring; Zreda et al., 2012). However, such CRP network sites have generally not been chosen according to meteorological/snow on the ground considerations which may hamper the representativeness and usefulness of existing CRP sites. The additional strength of

the approach is that it provides an integrated measurement rather than a point measurement, which may better capture representative snow conditions around the site. However, the footprint area may depend on the site considered (ground geological characteristics and topography) and should be assessed in detail (Sigouin and Si, 2016). Furthermore, according to Sigouin and Si (2016), the use of CRP for SWE estimates is limited by groundwater variations, which occur, in particular, at the beginning of the snow ablation period.

4.1.4.5.5 *Spatially distributed snow depth*

Several approaches based on laser ranging systems have been developed to measure more than one point (which is the case with fixed laser-based systems tested within WMO-SPICE). This ranges from a small number of points (less than 10) to larger numbers, making it possible to map snow depth over variable surface areas. Such laser ranging systems operate either in the visible or in the near-infrared spectral regions. However, measurements using those wavelengths potentially make this technique weather dependent.

The SDMS-100 and JH-20 sensors tested at the Gochang test site in the Republic of Korea during the WMO-SPICE campaign both measured multi-point snow depths. The JH-20 uses three different lasers at different viewing angles and the SDM-100 uses multiple lasers at different viewing angles.

Picard et al. (2016) have developed a prototype measurement system consisting of a lasermeter mounted on a two-axis stage which can run unattended and can scan approx. 200,000 points over an area of 100-200 m^2 in 4 hours. Based on measurements performed at Col de Porte, French Alps, and Dome C in Antarctica, the precision of single point measurements and long-term stability were evaluated to be about 1 cm and the accuracy to be 5 cm or better. The spatial variability in the scanned area reached 7-10 cm (RMSE) at both sites, which means that the number of measurements can be considered sufficient to average out the spatial variability and yield precise mean snow depths. The strength of this system lies in its cost-effectiveness and its robustness, and the fact that it can provide not only a mean snow-depth value but also quantitative estimates of the variability of the scanned area. Weaknesses include the fact that it requires on-site power and can be very data-intensive. The prototype system is currently not available on the market as a retail instrument.

Recently, there has been significant application of ground-based laser scanning devices for snow-pack measurement (e.g. Prokop et al., 2008, Deems et al., 2013). In most cases, periodic snow-depth maps are produced from one or several landscape vantage points in various environments (Mott et al., 2010, Vionnet et al., 2015, Filhol et al., 2015). Kaasalainen et al. (2010) describe the use of such instruments on mobile platforms. Adams et al. (2013) have recently reported a successful deployment of a terrestrial laser scanner capable of performing continuous measurements. This was achieved by inserting a commercially-available terrestrial laser scanner into a dedicated enclosure that was then deployed for field measurements. The main strength of terrestrial laser scanning systems is that they provide quantitative estimates of the snow depth variations of a scanned area which can cover several km^2 (depending on terrain and instrument configuration). Their main weakness is that they require significant resources to operate at a field site, including power and qualified field personnel to operate the system. Moreover, dedicated data management protocols are required to handle the large amount of data that can be generated. Such instruments are also characterized by high purchasing costs (on the order of 100,000 € minimum). However, the technology is progressing, systems costs continue to drop, and the accuracy and precision continues to improve.

It is worth mentioning here that differential GNSS processing is used to measure snow depth on ski slopes using receivers installed on grooming machines (Söderström et al., 2013). This requires that a high-resolution digital elevation model of bare ground is available. Radar-based systems are also used to indicate snow depth on ski slopes.

4.1.4.5.6 *Conclusions*

The WMO-SPICE project evaluated the operational performances of several commercially available sensors that measure snow depth and SWE. Recent years have seen the emergence and consolidation of various technological approaches to measure snow properties in the field, including snow depth and SWE, either directly or as a byproduct of sensors primarily targeting internal properties of snow. This section collated a list of such instruments, along with their performances as stated in the corresponding publications. If such devices, currently belonging mostly to the research community, are to be made available to a broader user community including operational networks, it would be highly desirable to carry out extended testing of "packaged" versions of the instruments, involving not only research groups but also field testing sites operated by national operational organizations. This would help to evaluate and improve system operation, allow for expeditious calibration of sensors and devices, and encourage the development of new sensors and technologies that further the scientific goal of understanding, measuring and modeling changes in the seasonal snowpack. Reference to or mention of specific products, processes or services in the above section are for identification only and do not constitute or imply a recommendation or endorsement by SPICE. Most were not tested in the field by as part of the WMO-SPICE project.

4.2 Challenges impacting the measurement of snow

4.2.1 Capping

Authors: Samuel Buisán, Javier Alastrué, José Luís Collado, Mareile Wolff, Mike Earle, Yves-Alain Roulet

4.2.1.1 Introduction

Capping is the process by which the orifice of a precipitation gauge is blocked, partially or totally, by snow and/or ice. This can induce errors of various types: no recording of precipitation during an event, partial recording of an event, or time-delayed recording of events when the snow/ice collected at the orifice falls into the bucket. Capping is very difficult to detect automatically from gauge data and without personnel on-site or webcam pictures.

This source of error is unique to the measurement of snow and needs to be considered for either manual or automated snow measurements with nearly any type of instrument. The most common approach to alleviate this problem is to heat the instrument. Note that the hotplate sensor is not subject to capping unless the precipitation rate exceeds the rate of snow melting on the sensor. This would require rates higher than 50 mm/hr, which to our knowledge has never happened for snowfall.

Rasmussen et al. (2001 and 2012) addressed the importance of preventing the accumulation of snowfall on automatic precipitation gauges and established some recommendations for near-real time applications. Colli et al. (2013) tested the SPICE reference sensors in a wind-free laboratory and focused mainly on quantifying the heating related by varying the environmental temperature and the snowall intensity by using the NCAR snow machine. However, it was not possible to consider other environmental factors such as wind, solar radiation, and temperature gradients, as well as other site features such as elevation, configuration of sensors, and exposure.

The SPICE test sites are characterized by different weather conditions, representative of different climate regimes. For this reason, during the experiment, partial or complete capping events were only detected at some sites with conditions favoring this effect, such as wet snow, high snowfall rates, and light winds. In addition, a number of sensors already include heating, often preventing capping from occurring, especially at sites with lower snowfall rates and colder temperatures. This section examines capping using events reported at different SPICE sites to determine the main factors enhancing capping probability, to develop procedures to detect this issue, and finally to provide some recommendations to avoid this problem. For these purposes, SPICE reference weighing gauges are used, but it should be stated that this issue has also been observed on other gauges (tipping buckets and weighing gauges under test).

The SPICE data quality control procedure applied to data from instruments at all sites included measures to detect potential capping events (Section 3.3.2). These events were flagged for further assessment. The procedure to detect potential partial or complete capping events in the SPICE QC procedure was as follows:

- The accumulation value at each minute *i* was assessed relative to the value at minute *i-1*; if the difference exceeded specified erroneous or suspect thresholds, the value at minute *i* was flagged as a jump.
- For 1-minute data from weighing gauges, the suspect and erroneous jump thresholds were 20 mm and 30 mm, respectively. These thresholds can be adapted to detect smaller jumps depending on the instruments, sampling resolution, and operational requirements.
- If more than 120 consecutive minutes were recorded for which the data after minute *i* had a suspect or erroneous jump relative to the value of minute *i-1*, a baseline shift was identified.
- Considering the potential for the baseline shift to result from capping, the period of time 1 hour before and 1 hour after the baseline shift was flagged for manual assessment/intervention.

An example of the application of the algorithm is shown in Figure 4.110. This approach has been shown to be useful for the automatic detection of capping events at SPICE sites such as Formigal and Haukeliseter, demonstrating its potential for operational use. The use of web cameras monitoring the instruments is also an effective method to detect or verify suspected capping events.

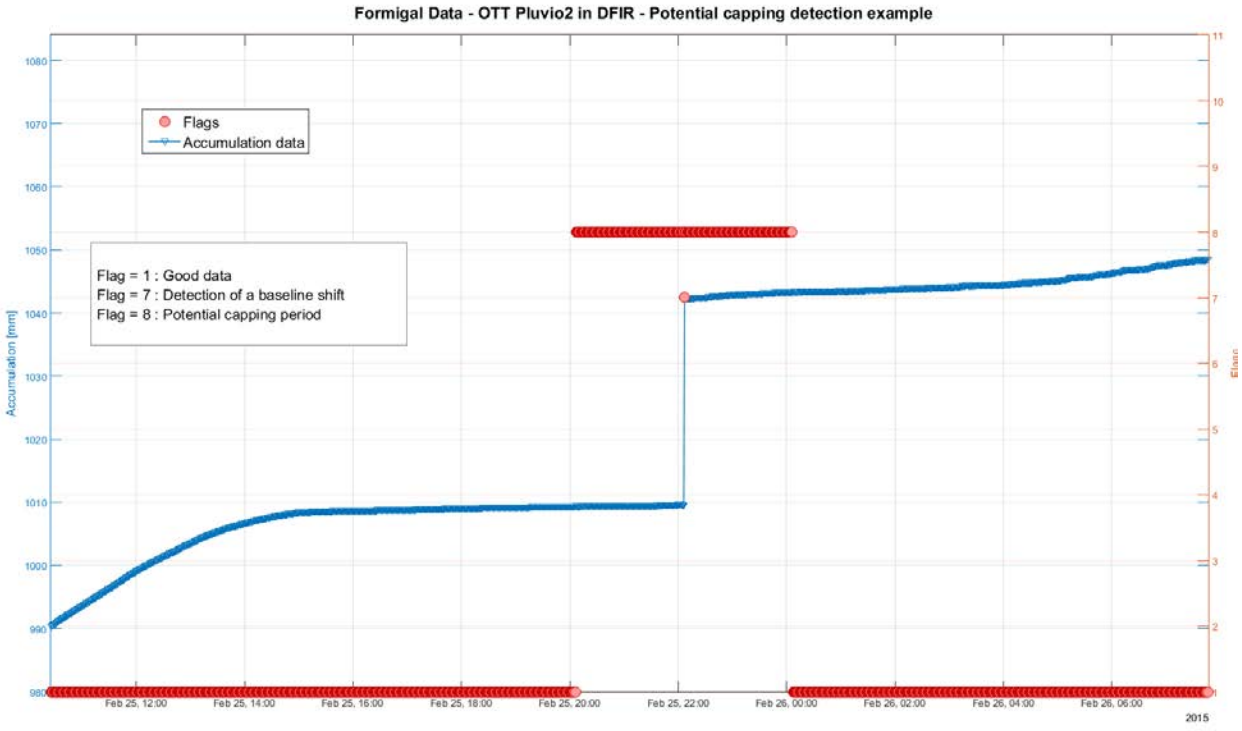


Figure 4.110. Detection of snow capping at the Formigal site using SPICE algorithm.

4.2.1.2 Case studies of capping

4.2.1.2.1 *Weissfluhjoch, March 2014*

High snowfall intensities can produce capping or partial capping, especially under light winds or for shielded gauge configurations. Figure 4.111 shows this effect at Weissfluhjoch, where the wind removed faster the accumulated snow on the unshielded Pluvio^2 (Figure 4.111c) than that on the single-Alter-shielded Pluvio^2 (Figure 4.111b) and single-Alter-shielded Pluvio^2 in the DFAR (Figure 4.111a).

Figure 4.111. Snow accumulation on heated Pluvio^2 in a DFAR (left), in a single-Alter shield (middle) and in an unshielded configuration (right) during a snowfall episode on March 2, 2014 at Weissfluhjoch.

The DFIR-fence is designed to mitigate horizontal wind effects on the measurement of precipitation. The results in Figure 4.111c illustrate that during high-intensity snowfall events, this reduction of wind speed within the DFIR-fence allows snow to accumulate more easily on the instrument orifice.

4.2.1.2.2 *Formigal, February 2015*

Figure 4.114 shows the time series of accumulated precipitation during a snowfall event at Formigal. During this heavy snowfall episode (> 75 mm/day) the heated Pluvio^2 within the DFIR-fence experienced severe capping (Figure 4.112).

Figure 4.112. Severe capping of Pluvio^2 in DFIR, Formigal, February 2, 2015.

During the same episode, snow accumulated on the Thies LPM in the DFIR-fence, but did not obstruct the sample area (Figure 4.112). The snow accumulated on the arms of the sensor was not likely to be blown into the sample area and detected by the laser. (Note: Aside from capping events, the accumulated precipitation of the LPM and Pluvio^2 in the DFAR at Formigal showed good agreement, with differences of less than 10% during snowfall episodes over the whole winter season). According to the accumulated precipitation measured by the LPM, the Pluvio^2 gauge underestimated the total event accumulation by 110 mm, or 55%. Over the following days, the accumulated snow melted and fell into the bucket of the Pluvio^2 in the DFAR; however, the recorded quantity was only 31 mm.

The unshielded Pluvio^2 was not impacted by capping during this episode (Figure 4.113). This is attributed to the greater influence of wind in the absence of a shield. This positive performance with respect to capping is offset by negative performance with respect to catching and reporting precipitation. The unshielded Pluvio^2 was characterized by undercatch of approximately 50% relative to the DFAR under weather conditions characterized by temperatures between -3 °C and -6 °C and average wind speeds of 4 m/s.

The snow-depth sensor indicates the time when the snowfall during this episode ended (Figure 4.114). This information, in combination with the accumulated data from the LPM and the unshielded gauge, could potentially be used to develop algorithms based on the correlation between instruments before the capping started to infer the total accumulated precipitation of the capped instrument.

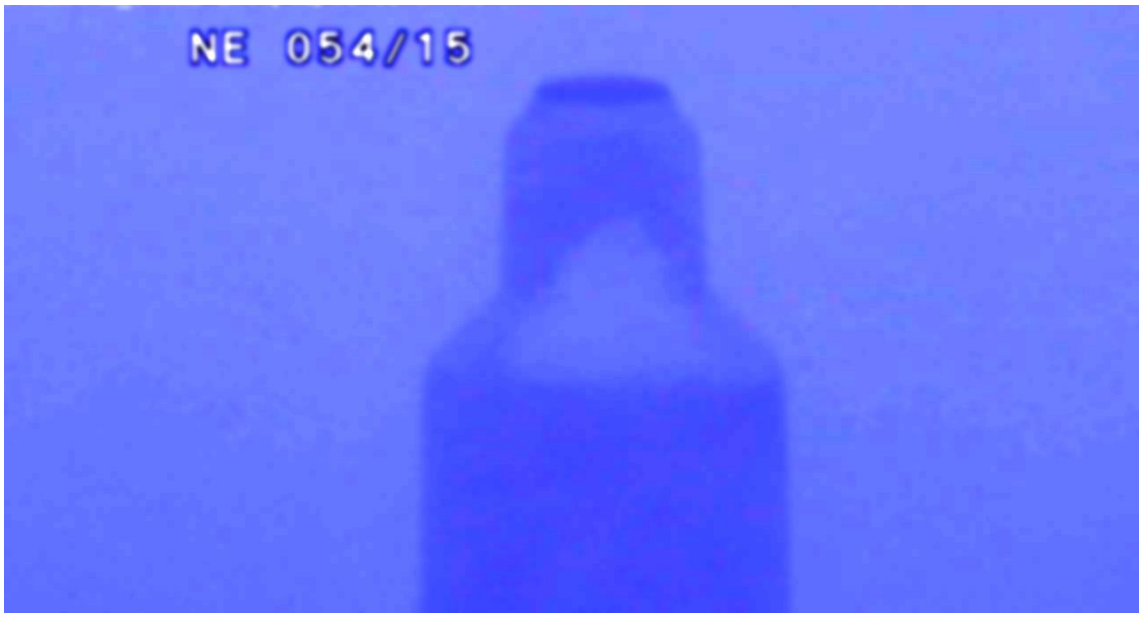


Figure 4.113. No capping for the unshielded Pluvio^2 gauge at Formigal on February 2, 2015.

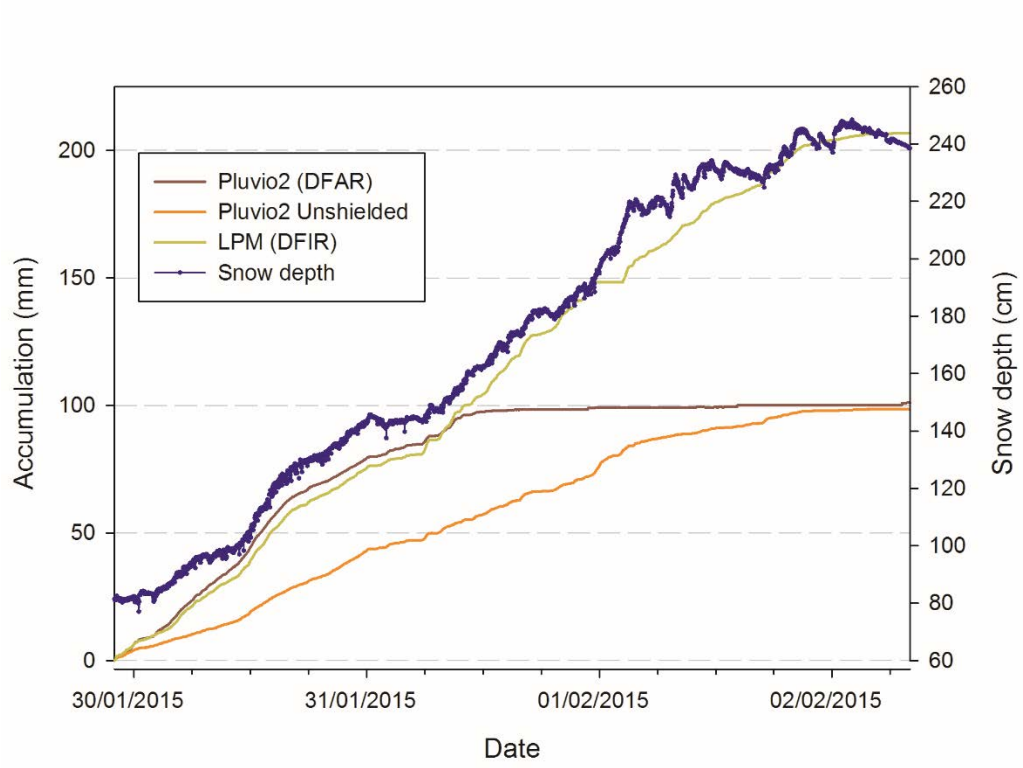


Figure 4.114. Time series of accumulated precipitation and snow depth during snowfall episode between January 30 and February 2, 2015 at Formigal.

Heated tipping bucket gauges are also susceptible to capping. Figure 4.115 shows a Thies tipping bucket gauge during the same heavy snowfall event at Formigal. The heater is not able to melt the

incident snowfall fast enough, causing the funnel to become filled with snow and preventing subsequent precipitation from being collected.

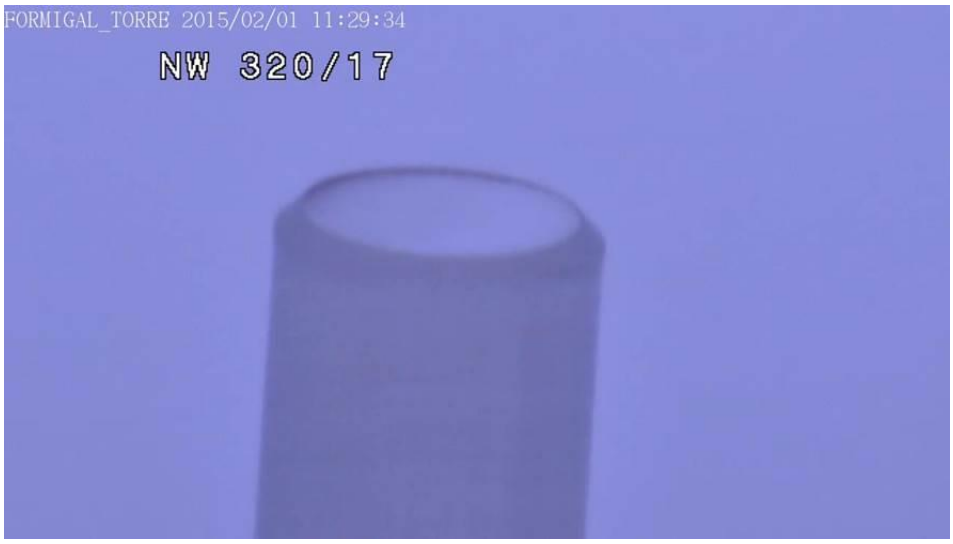


Figure 4.115. Capping observed in Thies tipping bucket, Formigal site on February 1, 2015.

4.2.1.2.3 *Sodankylä, January 2105*

Instruments located in forest clearings are typically less impacted by wind, given the sheltering effects of the surrounding trees. In these cases, wind is not an effective removal mechanism for accumulated snow, making heating the only effective removal mechanism. Figure 4.116 shows snow accumulated on the DFAR and other gauges at Sodankylä. In these low wind conditions (mean wind speeds are typically below 4 m/s), the buildup of snow is continuous over the season until about Februaury. This situation is also favoured by limited sunlight over this period.

Figure 4.116. Snow accretion on instruments and various structures (DFIR, SA shield) at Sodankylä.

Figure 4.117 shows the buildup of snow on unshielded and single-Alter-shielded Pluvio^2 gauges at Sodankylä on January 15, 2015. Snow builds up on the shoulders of the unshielded Pluvio^2, but not around the orifice rim, because it is heated (left). This situation presents a clear risk of future capping, since new snow could easily cover the orifice, building on the snow already accumulated on the shoulders. Figure 4.117 also shows snow accretion on the blades and rim of the single-Alter shield, which will impact its performance. Under low wind conditions, the only solution is to remove the snow manually. In general, however, for sites characterized by mean wind speeds of 2 m/s or less, the use of windshield is less critical than for sites characterized by higher winds.

Figure 4.117. Snow capping on unshielded Pluvio^2 (left) and on single-Alter-shielded Pluvio^2 before (middle) and after (right) manual snow removal on January 15, 2015 at Sodankylä.

4.2.1.2.4 *Joetsu, January 2014*

Figure 4.118 shows the time series of accumulated precipitation during a snowfall event at the Joetsu site, Japan. A baseline shift is detected on January 12 at 0900 LT, likely indicating snow from capping falling into the bucket (referred to as a "dump"). The Geonor gauge inside the DFIR was unheated for this event. When sidewall heating was applied to the Geonor (heating applied to external surface of inlet), no capping or dumping was observed at this site.

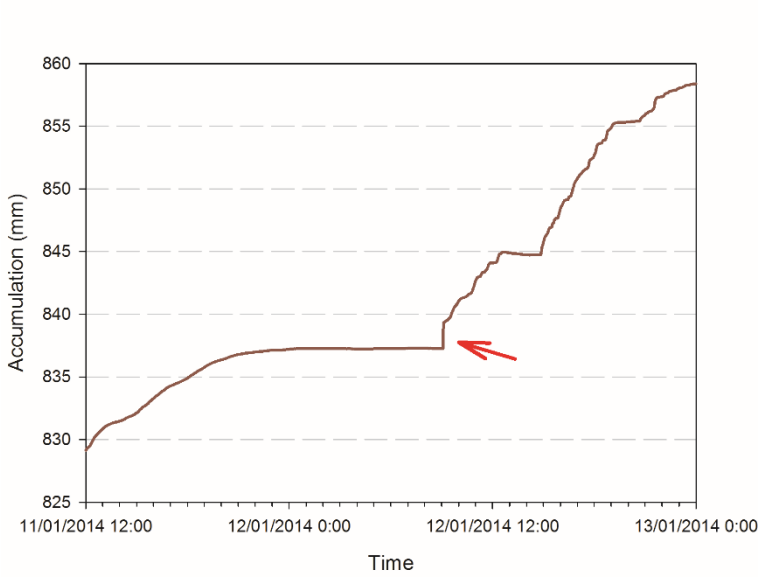


Figure 4.118. Time series of accumulated precipitation for the Geonor gauge in the DFAR at Joetsu during a snowfall episode from January 11 to 13, 2014. The red arrow shows a dump of precipitation when snow accumulated on the orifice during capping falls into the bucket.

Camera images of this episode show the temporal evolution of capping (Figure 4.119). The snow cap grew until it collapsed under its own weight.

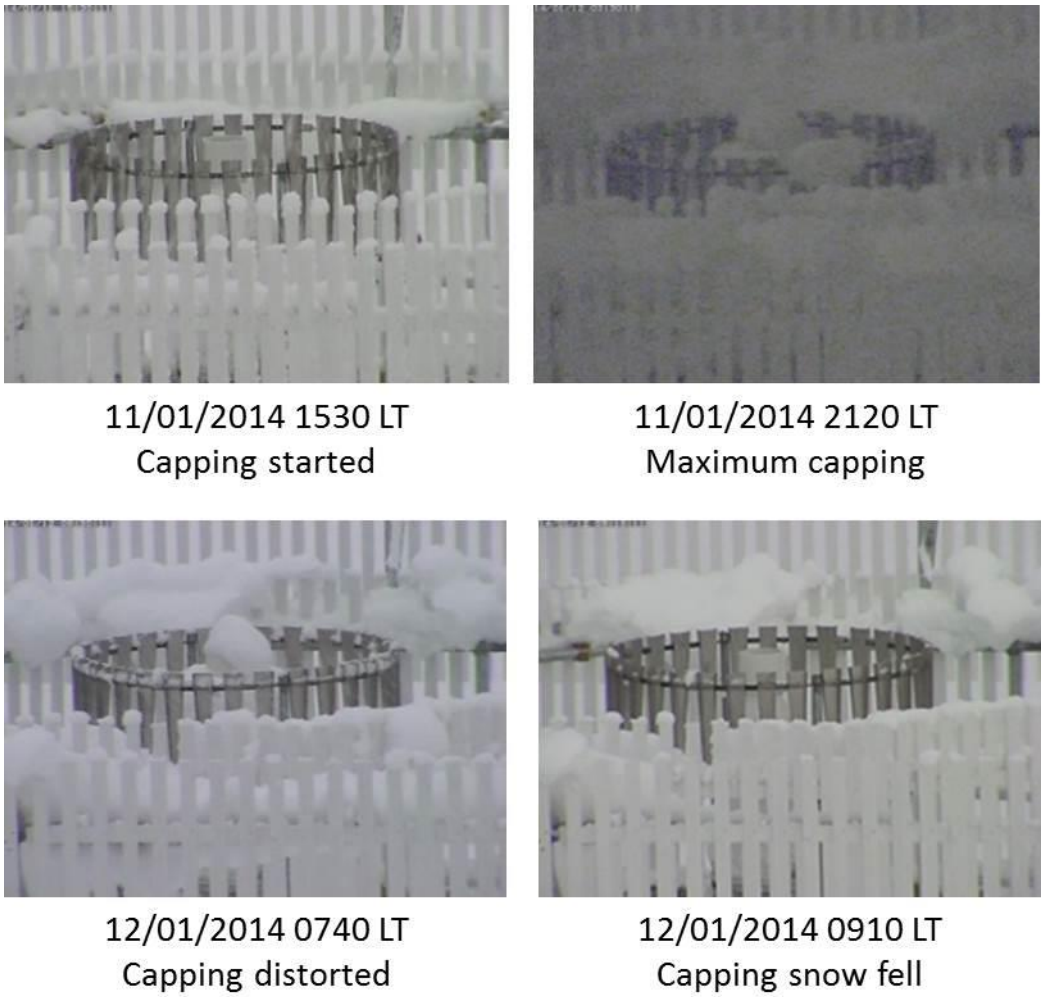


Figure 4.119. Snow capping on the Geonor gauge in the DFAR at Joetsu, January 11 to 12, 2014.

4.2.1.2.5 *Haukeliseter, December 2014*

Figure 4.120 shows the time series of accumulated precipitation over 10 days at Haukeliseter in December, 2014, during which multiple snowfall events occurred. The unheated Geonor in the DFAR was capped, whereas the other heated Geonor gauges, shielded and unshielded, outside of the DFAR did not show any capping. Weather conditions during this period of time were characterized by cold temperatures, -4 °C on average, which helped to maintain the capping. Also, low solar radiation at this latitude in December prevented snow melting that could have removed the capping. Again, capping only occurred at this site when the Geonor gauges were unheated.

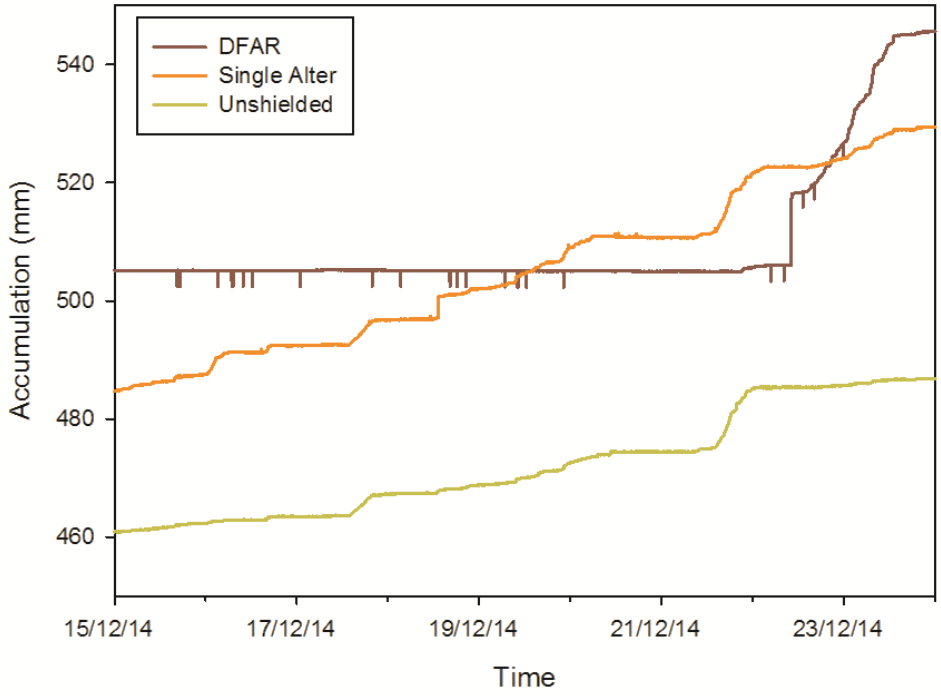


Figure 4.120. Time series of accumulated precipitation during snowfall episodes, December 15 to 24, 2014 at Haukeliseter

Camera images show the time series of capping on the unheated gauge in the DFAR (Figure 4.121). The snow cap grew around the orifice, taking advantage of snow already accumulated on the shoulders. During the second season, in which heating was applied to this gauge, no capping or dumping events were observed.

Figure 4.121. Snow capping on Geonor at Haukeliseter site, December 5 to 10, 2014.

4.2.1.2.6 *CARE, December 2014*

Figure 4.122 shows the accumulation of snow on double-Alter-shielded gauges at the CARE site in December, 2014. Snow accumulated on the heated Pluvio^2 (left) to a greater extent than the heated Geonor gauge (middle). For the Pluvio^2, only the ring rim was heated, whereas for the Geonor, an extended portion of the gauge inlet (collar) was heated. The unheated Geonor (right) shows snow accretion on both the orifice and collar (right).

Figure 4.122. Accretion of snow on heated Pluvio^2 (left), heated Geonor (center) and unheated Geonor (right) at CARE in December, 2014. All gauges are in double-Alter shield configurations.

4.2.1.2.7 *Marshall, April 2015*

Figure 4.123 shows the time series of accumulated precipitation for the Pluvio^2 gauge in the DFAR and an unshielded Pluvio^2 at the Marshall site for a snowfall episode in April, 2015. The gauge in the DFAR was heated and the unshielded gauge was unheated. Weather conditions were characterized by high precipitation rates, average wind speeds of 3 m/s to 4 m/s (not shown), and by temperatures close to 0 °C; typical conditions for wet snow. Note that the dump in the unheated gauge data occurs in association with a rise in temperature above 0 °C, while the heated gauge shows a linearly increasing accumulation, suggesting that the unheated gauge was capped, but not the heated gauge.

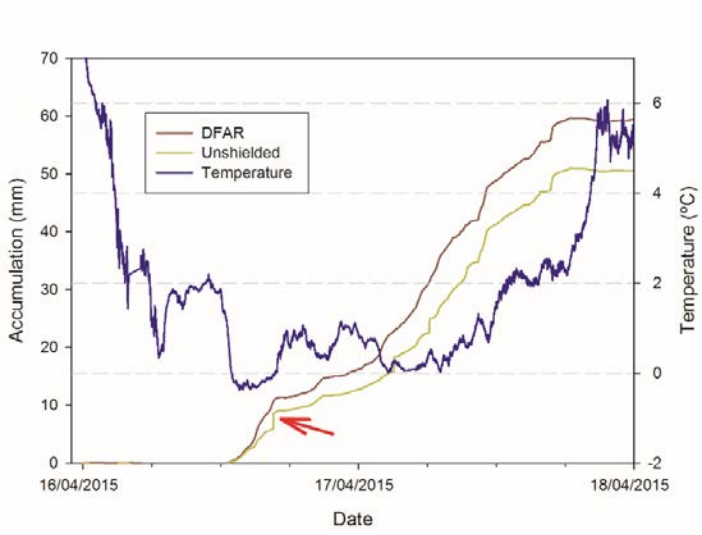


Figure 4.123. Time series of accumulated precipitation during a snowfall episode at Marshall from April 16 to 18, 2014. The red arrow shows an apparent dump in the unheated gauge accumulation, suggesting that this gauge was capped.

The impact of dumps can be significant if the data are used to develop transfer functions. Dumps represent time-delayed responses to precipitation that has already fallen, and will, therefore, increase artificially the reported accumulation during a given interval. This impact is demonstrated in Figure 4.124, which shows catch ratios of the unshielded gauge accumulation to that of the DFAR for 30-minute periods during the snowfall episode in Figure 4.123. The outlying values, with catch ratios > 2, represent 30-minute periods during which the unshielded and unheated gauge reported more than twice as much precipitation as the reference configuration, and are attributed to dumps.

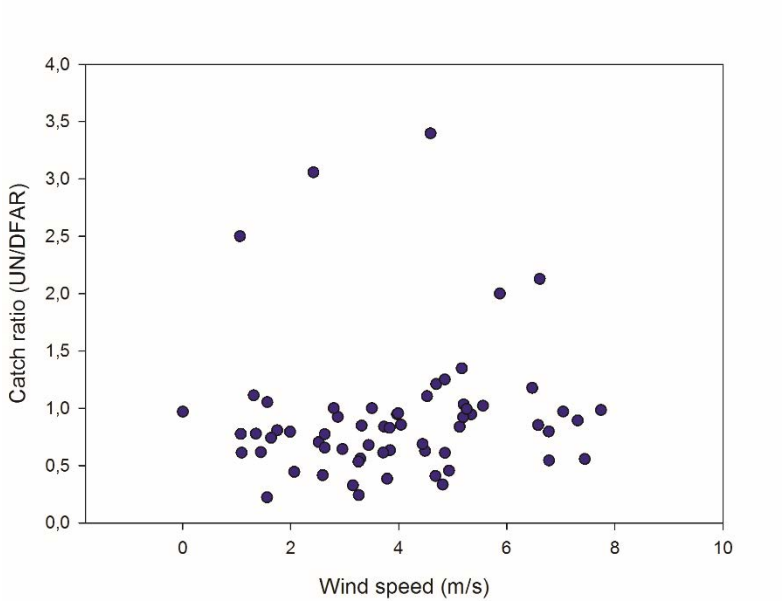


Figure 4.124. Catch ratio of an unshielded gauge (unheated) versus the reference measurement in a heated DFAR as a function of wind speed for 30-minute periods during the snowfall episode at Marshall from April 16 to 18, 2014.

4.2.1.2.8 *Formigal, November 2015*

Figure 4.125 shows the Pluvio^2 in the DFAR at the Formigal site no November 25, 2015. When rim heating is off (left), snow partially covers the orifice; when rim heating is turned on (right), the cap is melted and the orifice is clear.

Figure 4.125. Pluvio^2 in the DFAR at Formigal with rim heating off (left) and after turning on the rim heating (right) on November 25, 2015.

Figure 4.126 shows that the ring temperature of the Pluvio^2 had a fixed value of 4.99 °C between 0000 and 0820 LT, indicating a problem with the heater. A remote command was sent to Pluvio^2 andthe rim heating was turned on again; this melted the accumulated snow, which partially fell into the bucket and partially on the shoulder a few minutes later. It is a challenge to assess if the final measured quantity corresponds to the effective snowfall. Nevertheless, this action allowed the gauge to operate and measure correctly (i.e. without obstruction) during the rest of the snowfall episode.

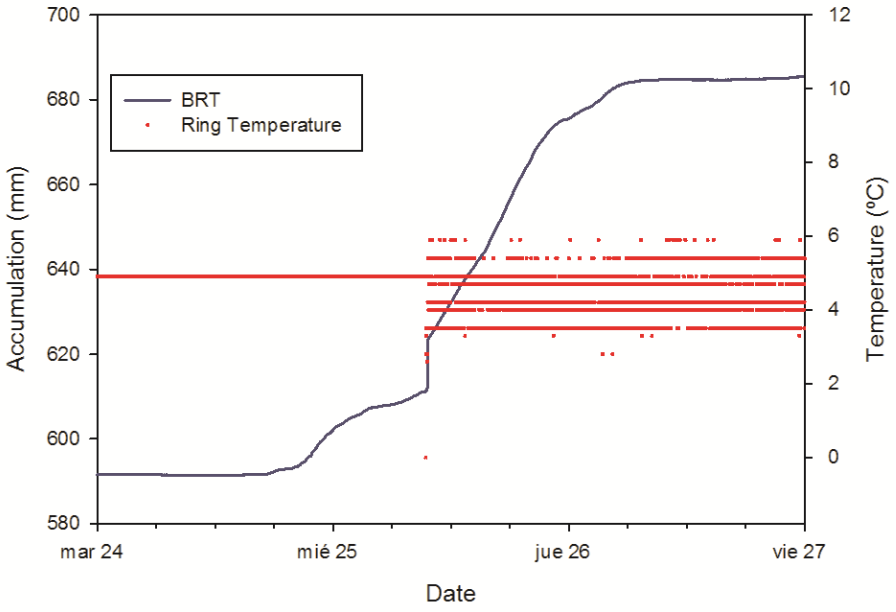


Figure 4.126. Time series of accumulated precipitation and temperature of orifice ring during snowfall episode from November 24 to 27, 2015 at Formigal.

4.2.1.3 The importance of raw data for the detection of capping

Depending on the instrument used, some filters or algorithms may be applied to the data. In the case of the Geonor, all measurements are recorded as raw data. For the Pluvio^2, however, multiple data outputs are available, employing different algorithms to report precipitation parameters in real-time (RT), or with a delay for processing (non-real-time, NRT). An overview of the Pluvio^2 outputs is provided in Section 3.1.3.7.1.

Figure 4.127 shows the time series of accumulation for the Pluvio^2 gauge in the DFAR (left) and a single-Alter-shielded Pluvio^2 at Formigal. The Accumulated Total NRT parameter of the Pluvio^2, which filters the data to avoid unrealistic measurements, and the Bucket RT, which is the equivalent of a raw data for Pluvio^2, are shown for each gauge configuration. In the DFAR data, the baseline shift in the Bucket RT data, indicating a potential capping event, is filtered out in the Accumulated Total NRT data. This demonstrates that capping episodes in some situations may not be detected in filtered data outputs, and illustrates the importance of retrieving and using raw data for the detection of capping events.

The single-Alter-shielded Pluvio^2 data shows some suspect baseline shifts in Figure 4.127; however, the magnitude of these shifts is smaller, and so they were not filtered out in the Accumulated Total NRT data. In this case, the algorithm developed within SPICE detected the capping.

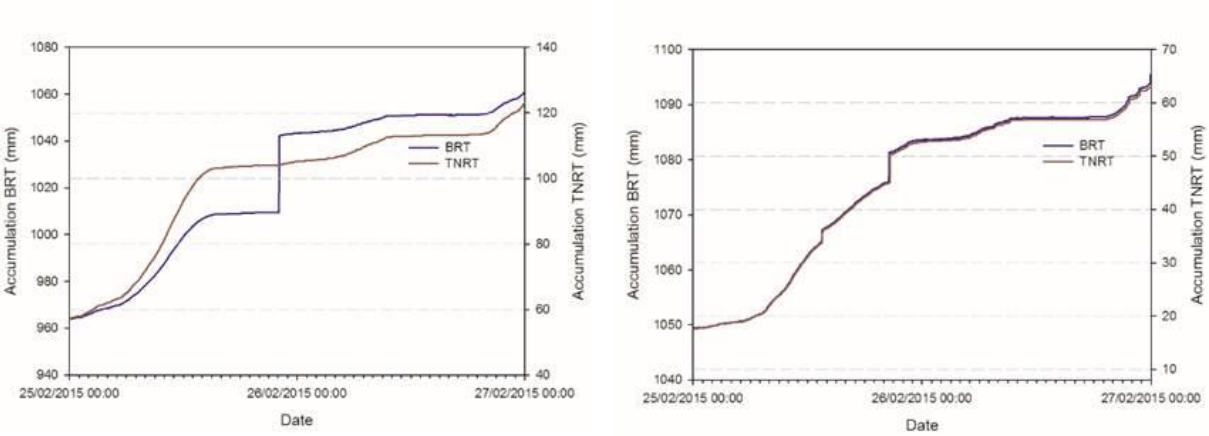


Figure 4.127. Time series of Bucket RT (BRT) and Accumulated Total NRT (TNRT) accumulation from Pluvio^2 gauges in the DFAR (left) and in a single-Alter shield (right) during a snowfall episode at Formigal from February 25 to 27, 2015.

4.2.1.4 Partial capping

One of the most challenging issues in the measurement of solid precipitation using precipitation gauges is to detect partial capping. The accretion of snow around the orifice reduces the opening area, without capping it completely, thus allowing new snow to be measured, but producing erroneous data (underestimation of precipitation accumulation) with a certain degree of uncertainty.

4.2.1.4.1 *Weissfluhjoch, November2014*

Figure 4.128 shows an episode of partial capping at the Weissfluhjoch site that affected the Pluvio^2 gauges in DFIR-shielded, single-Alter-shielded, and unshielded configurations.

Figure 4.128. Partial capping of Pluvio^2 gauges in (left) DFIR-fence, (middle) single-Alter shield, and (right) unshielded configuration on November 6, 2014 at Weissfluhjoch.

Figure 4.129 shows that all gauges measured precipitation; however, based on the webcam images that show partial capping, an underestimation of the accumulation can be expected, particularly for the gauges in the DFAR and single-Alter. Additional heating of the upper orifice is required to alleviate this condition.

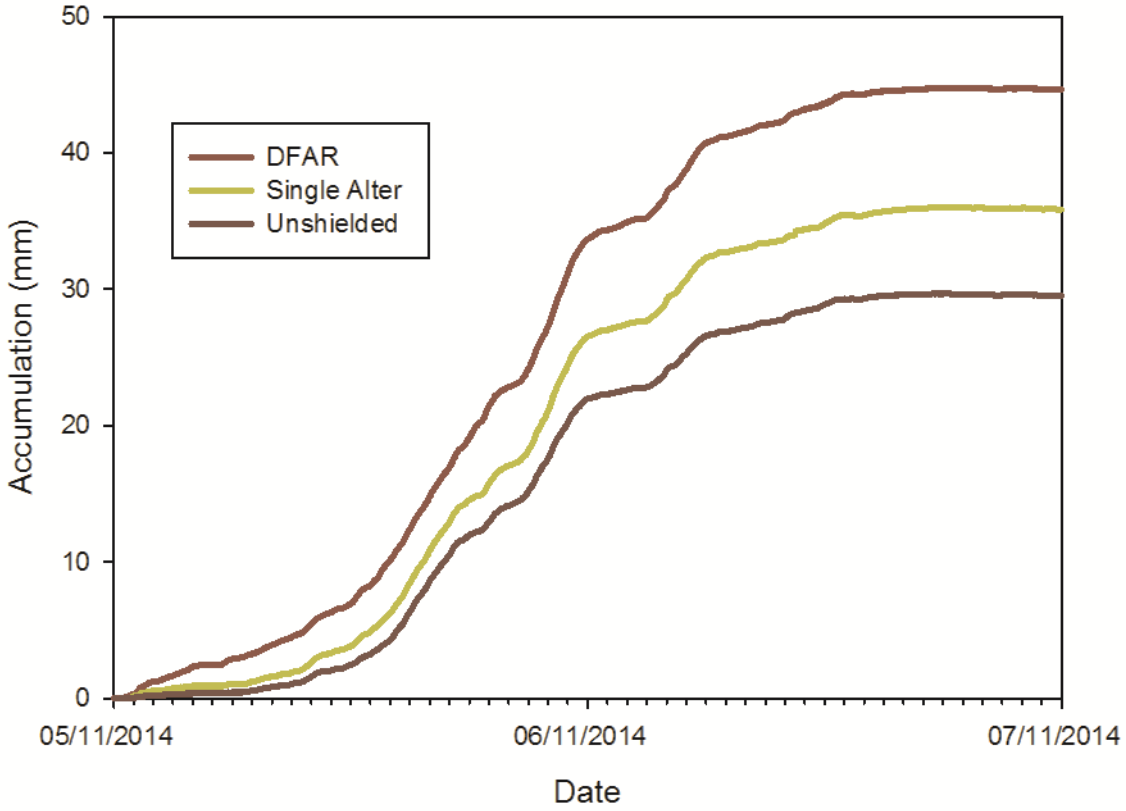


Figure 4.129. Time series of accumulated precipitation during a snowfall episode from November 5 to 7, 2014 at Weissfluhjoch.

4.2.1.4.2 *Haukeliseter, January 2015*

Figure 4.130 shows webcam images of a heated Geonor gauge in the DFAR at Haukeliseter that experienced a partial capping event. Snow accumulated on the shoulders of the gauge (January 12, 2015), which in turn served as a platform for fresh snow to accumulate and partially cover the gauge (January 14, 2015). The heating was useful to melt the snow close to the heated surface of the neck of the gauge and partially remove the snow (January 15). The low sunlight radiation at this site contributed to enhance the capping.

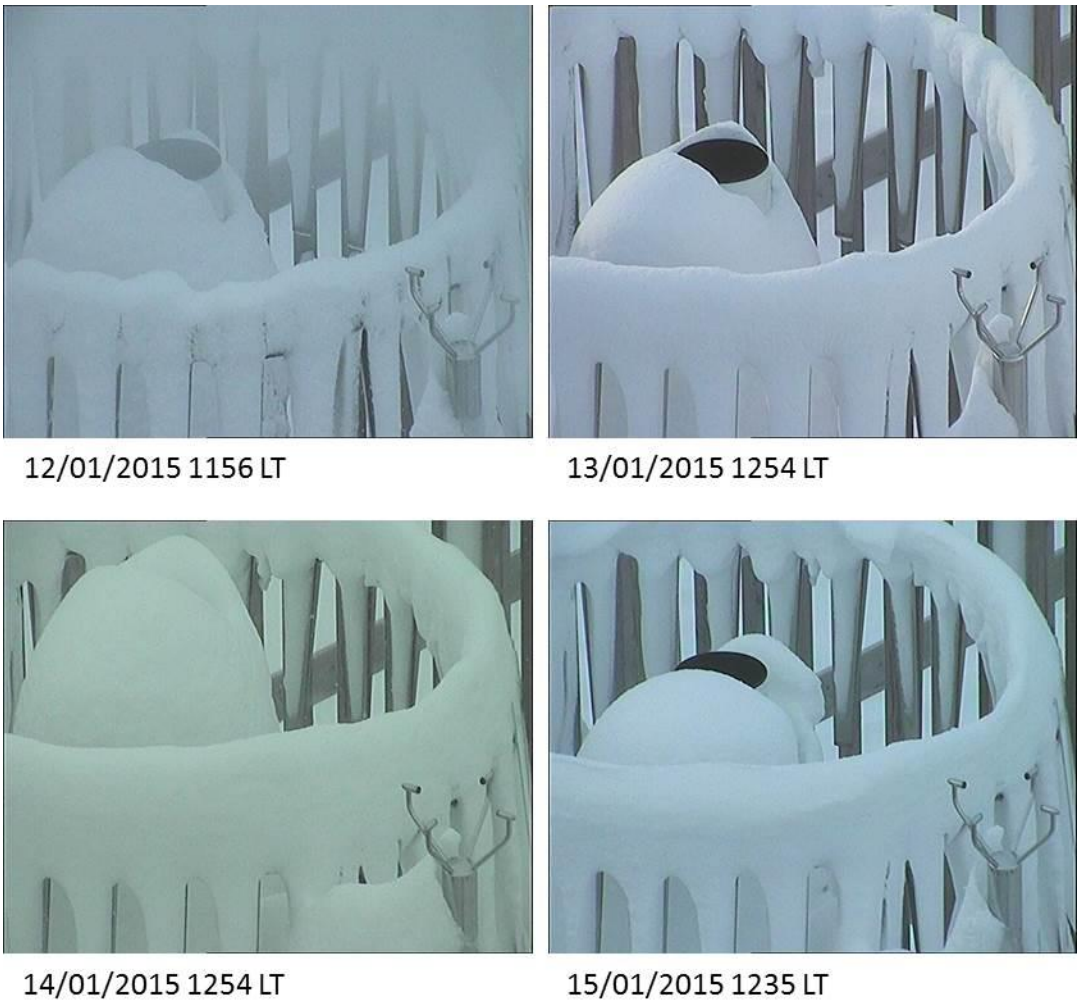


Figure 4.130. Evolution of a partial snow capping event for a Geonor gauge at Haukeliseter over four consecutive days.

Figure 4.131 shows the time series of accumulated precipitation during the first two days of the partial capping event. The snowfall was characterized by high winds, resulting in significant undercatch for the single-Alter-shielded gauge relative to the gauge in the DFAR. However, some dumps are observed in the accumulation of the DFAR, the first one on January 14 around 0100 LT and the second one around 1200 LT. Both are likely associated with snow falling into the bucket. This example demonstrates the difficulty of distinguishing between partial and complete capping in accumulation time series.

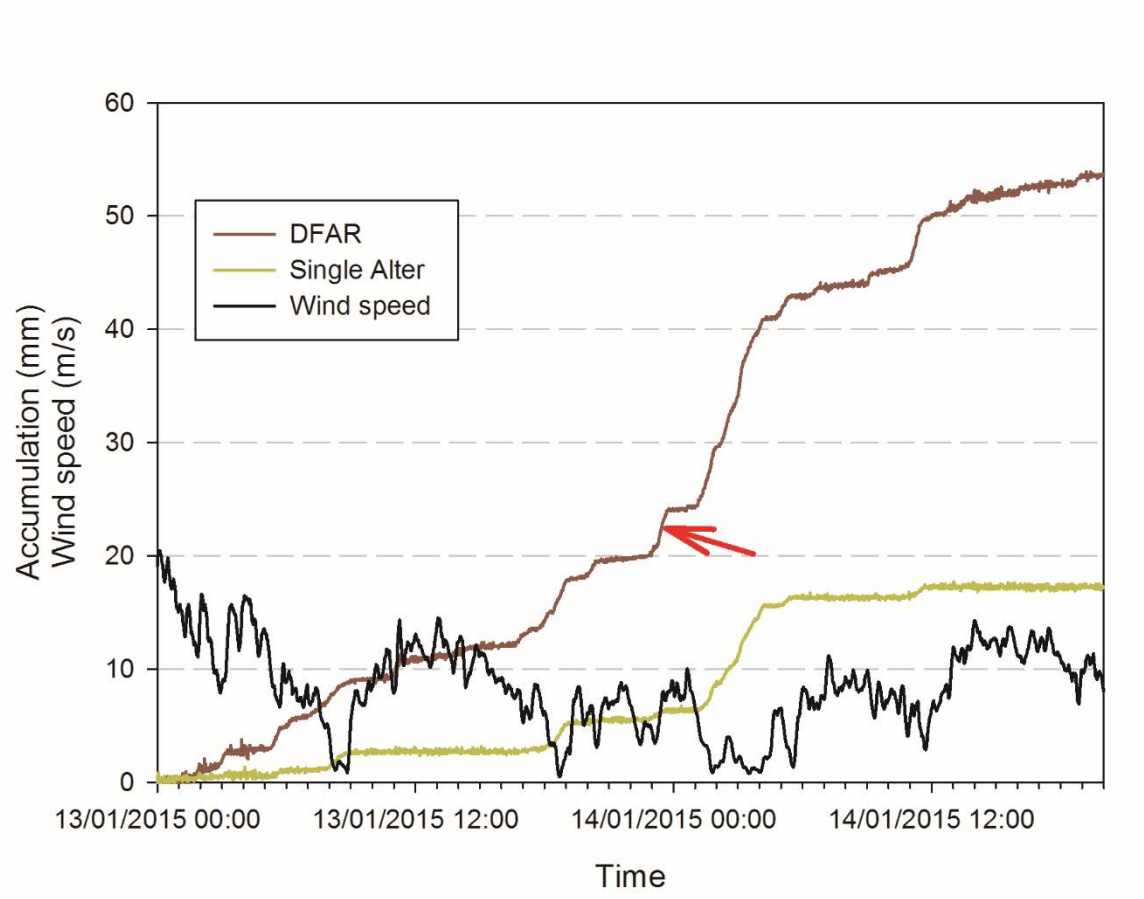
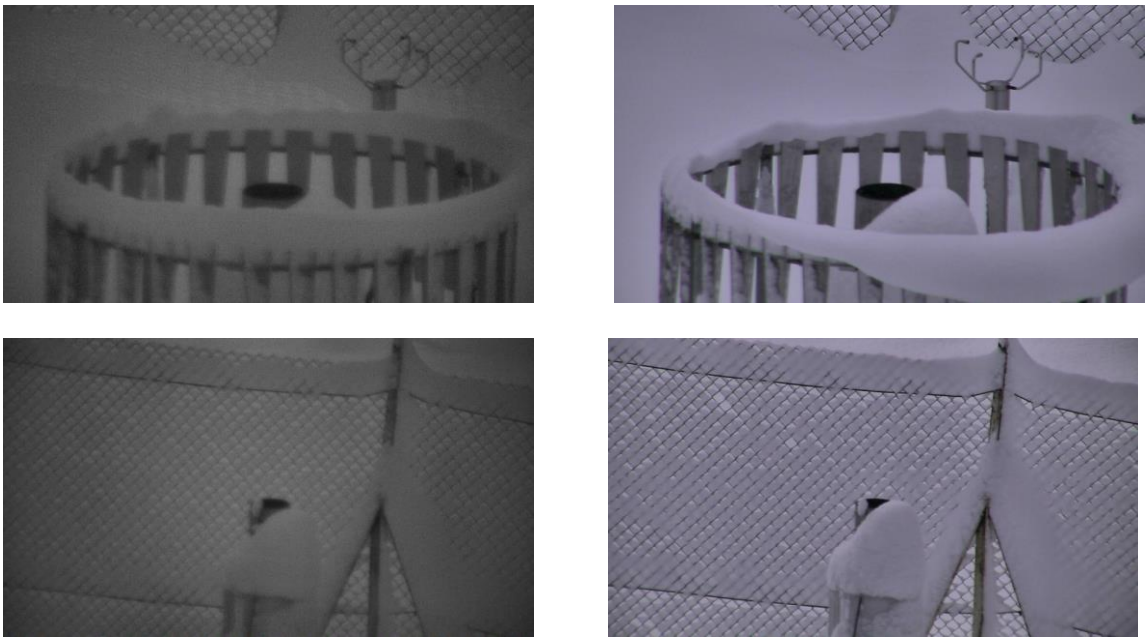


Figure 4.131. Time series of accumulated precipitation for Geonor gauges in the DFAR and a single-Alter shield at Haukeliseter during a snowfall event in January, 2015. Apparent dumps (red arrow) are superimposed on the accumulation time series for the gauge in the DFAR.

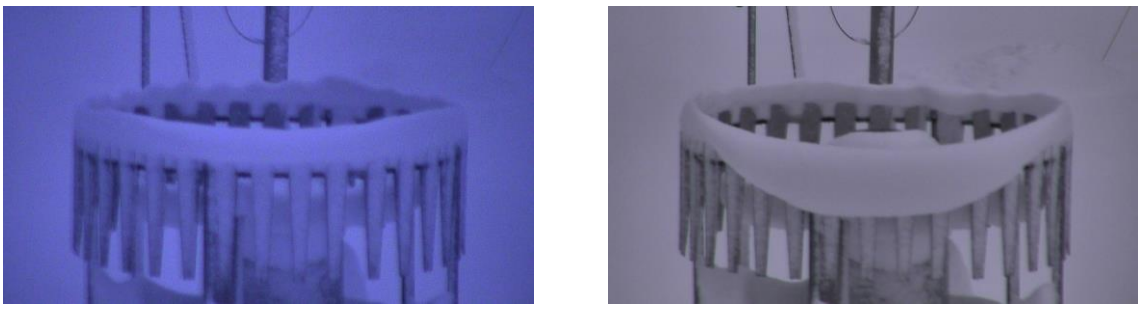
4.2.1.4.3 *Col de Porte, January 2015*

Figure 4.132 shows webcam images of two heated Geonors in single-Alter and unshielded configurations during a partial capping event at Col de Porte in January 2015. Figure 4.133 shows a single-Alter-shielded Pluvio^2 gauge (orifice area of 400 cm2) during the same episode. The impact of capping appears to be more significant for the gauge, which is attributed to two factors: first, the Pluvio^2 has a smaller heated area (rim only) relative to the Geonor gauges; and second, the close proximity of the shoulders on the gauge housing of the 400 cm^2 Pluvio^2 to the orifice.



17/01/2015 6h38m 17/01/2015 12h38m

Figure 4.132. Evolution of a partial snow cap on single-Alter-shielded (top) and unshielded (bottom) Geonor gauges at Col de Porte in January 2015.



17/01/2015 6h54m 17/01/2015 11h24m

Figure 4.133. Evolution of a partial snow cap on a single-Alter-shielded Pluvio^2 with 400 cm^2 orifice area at Col de Porte in January 2015.

Figure 4.134 shows that this episode was characterized by a fast transition from rain and mixed precipitation to snow, as confirmed by the precipitation detectors at this site (present weather data not shown). The snowfall was associated with high-intensity precipitation rates (more than 40 mm in less than 12 hours), temperatures near 0 °C (wet snow), and wind speeds lower than 2 m/s during most of the episode. These weather conditions favoured a capping event. Heating melted the snow around the gauge collar for the Geonors, but was insufficient to remove the cap on the Pluvio^2. After the event, the low temperatures and subsequent precipitation helped to maintain and enhance the capping.

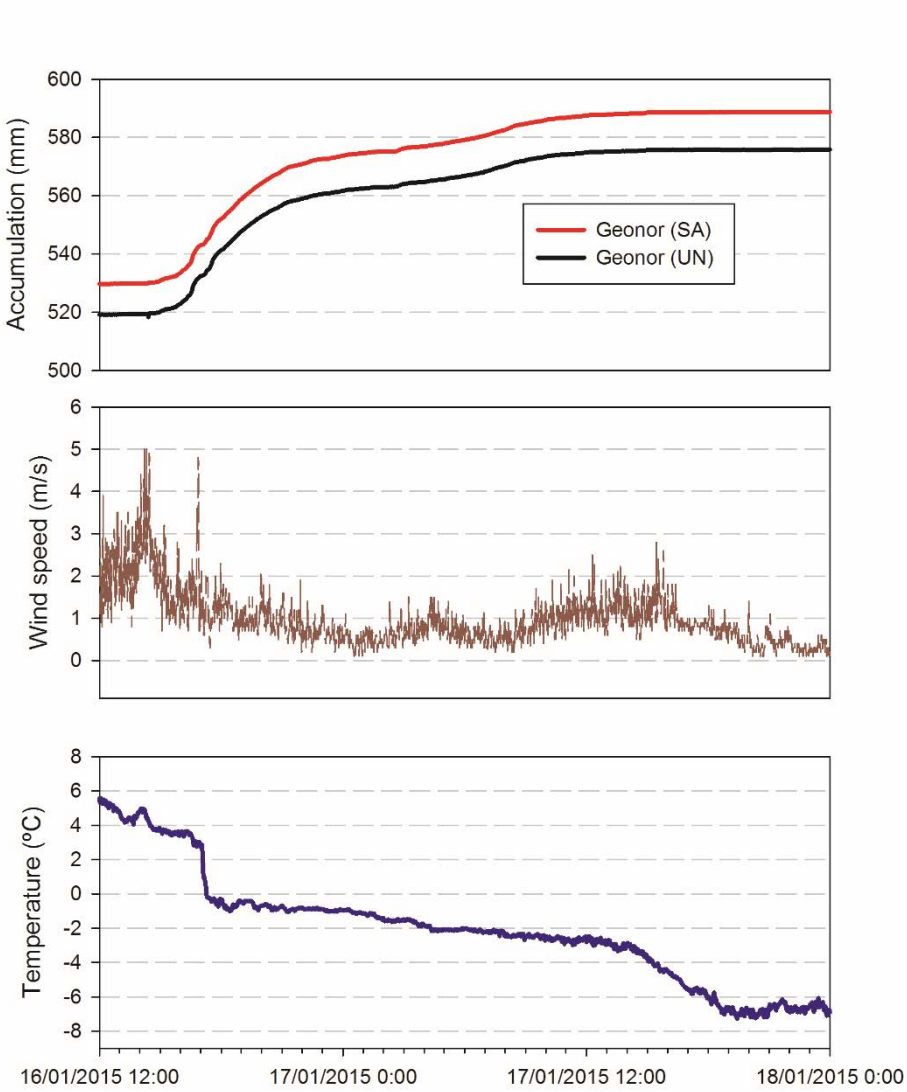


Figure 4.134. Time series of temperature, wind speed and accumulated precipitation for single-Alter-shielded and unshielded Geonor gauges for the precipitation event leading to partial capping at Col de Porte in January 2015.

4.2.1.5 Conclusions

The occurrence of capping is not due to a single factor, but is typically a combination of multiple phenomena and configuration issues. Table 4.25 summarizes the main factors associated with capping events as observed and detected during SPICE. In some sites, weather conditions were the main factor (i.e. Formigal); in others, the exposure was critical (i.e. Sodankylä); but for all cases, configuration and design of the gauge played a role. Capping is a particularly important issue when sufficient power is not available to heat surfaces of the gauge exposed to snowfall. The Pluvio^2 gauge suffered from capping more often than the Geonor gauge, due to heating only along the rim of the gauge, whereas the Geonor was heated throughout the inlet (both inside and outside). If the Geonor is not heated inside the orifice, snow can melt from the upper part of the orifice and freeze on the lower part, leading to undercatch and potentially a complete blockage of the inlet. Thus, it is imperative to heat the entire inlet of Geonor gauges.

Capping can also be observed under singular weather conditions such as situations where a sudden decrease in temperature occurs during a precipitation event, and freezes mixed precipitation onto the gauge. Usually, and despite of heating, the ice accretion will remain on the gauge and grow. These cases may be specific to the Pluvio^2 gauge since the slanted sidewall is not heated, allowing snow to grow from that surface upwards. In contrast, the Geonor gauge is heated throughout the entire upper sidewall and very little snow accumulation was observed with that gauge during SPICE, even at the Joetsu Japanese site that experienced heavy, wet snow.

Table 4.25. Summary of relevant factors affecting capping

		Favoring Capping	Favoring No capping
Weather conditions			
	Wind	Calm conditions (< 2 m/s)	Windy conditions (> 2m/s)
	Precipitation rate	High	Light
	Precipitation phase	Fast transition from mixed snow to snow or wet snow	Dry snow
	Temperature	Near 0 °C condition	Extremely cold
Location			
	Sunlight	Low radiation	High radiation (melting)
	Exposure	Sheltered (i.e. forests clearings)	
Installation			
	Power supply for heating	Not available	Available
	Wind shield and configuration	DFIR	Unshielded or shields with moving parts
Gauge design			
	Heaters	Not available	
	Shoulders	Existing and low inclination	

4.2.1.6 Recommendations

Up until now, capping has only been partially addressed in official documents, since the previous solid precipitation measurement intercomparison from the 1990s considered only manual measurements. It has been demonstrated that unattended automated systems can be impacted by capping and recommendations must be put forward to manufacturers and users in order to recognize and address this problem. Among the recommendations are:

- Heating is the most effective way to minimize/prevent the gauge from capping (e.g. Figure 4.135a). The larger heated surface area of Geonor gauges relative to Pluvio^2 gauges, as tested, was shown to be more effective at mitigating the occurrence and influence of capping (see Section 4.2.2). At sites characterized by heavy, wet snowfall, such as Joetsu, experience has demonstrated that additional heating of the gauge exterior is required to prevent capping.

- Webcams should be used to detect partial or complete capping events, especially for stations where the quality of data is critical (i.e. airports).
- Rim heating alone is often not sufficient to eliminate capping. Heating the entire upper orifice of the gauge (both inside and outside) is recommended under all conditions.
- The temperature of the orifice should not exceed 2 °C, if possible, to avoid the generation of a heat plume under low wind conditions that can cause undercatch.
- Non-catchment-instruments (precipitation detectors) co-located with catchment gauges can be used to detect complete capping events.
- In locations characterized by heavy and wet snow events (coastal climates), and in the absence of power resources, sheltered areas should be avoided for gauge siting due to the increased likelihood of capping (e.g. Figure 4.135b). However, although more windy locations can help to remove the accumulated snow, the undercatch of precipitation due to wind-induced errors will typically be enhanced.
- Snow-depth sensors can help to determine the end of a snowfall episode and can be used to develop algorithms to retrieve correct precipitation of a capped instrument in combination with the accumulated precipitation and detection of precipitation by other instruments (if available).
- The design of the gauge is important; shoulders favour the build-up of snow. Future designs should consider separate heating of the shoulders (e.g. thermal blanket as shown in Figure 4.135c).
- Filtered or processed gauge data must be used with caution, since it can “smooth” or remove indications of capping. Raw data should be used to detect potential capping events, when available.
- Transfer function adjustments should not be applied to data impacted by capping or subsequent dumps, as the calculated catch ratios may be affected.

Figure 4.135. a) Capping in a location with limited power resources (Switzerland), b) Huge capping in a forested area in British Columbia (Canada), c) Thermal blanket in laboratory testing (Meteoswiss).

4.2.2 Heating

Authors: Jose Luis Collado, Samuel Buisán, Jeffery Hoover, Michael Earle, Javier Alastrué, Audrey Reverdin, Craig Smith

4.2.2.1 Introduction

Heating of instruments measuring solid precipitation is of key importance to avoid capping and also to provide a more timely response to snowfall events if this precipitation must to be melted to be measured.

However, even when the heating is controlled, some issues such as evaporation can occur. In this section we will analyze the heating control in reference gauges and the impact of heating on other instruments tested during SPICE.

4.2.2.2 Heating of weighing gauges

All reference weighing gauges in WMO-SPICE were heated to maintain the gauge orifice at a fixed temperature, typically between +2 °C and +4 °C. These thresholds are based on on experience collected during the pre-SPICE winter season (2012/13) and are documented in the IOC-4 Final Report, Davos. Specific details of the heating configuration and control algorithm for each type of weighing gauge employed in reference configurations are provided in the following sections.

4.2.2.2.1 *Geonor T-200B3 reference gauges*

For Geonor gauges, heating is not included as an integrated option from the manufacturer and, therefore, has to be provided externally. Two methods were used during the experiment: the Canadian heating configuration (Figure 4.136) and the configuration based on the experience gained in the Climate Reference Network (CRN) of the US National Oceanographic and Atmospheric Administration (NOAA).

The Canadian heating configuration employs two Geonor T-200B inlet heaters fixed to the Geonor orifice on the exterior of the gauge chimney, between the rim and the shoulder cylinder, and on the outside of the cylinder within the gauge housing, where it extends down towards the collecting bucket (Figure 4.137).

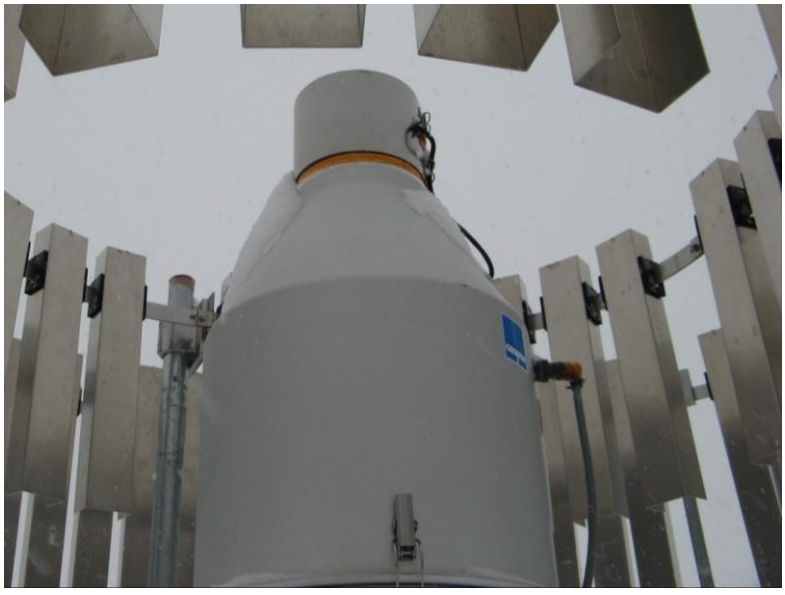


Figure 4.136. Geonor T-200B3 precipitation gauge with Canadian heating configuration.

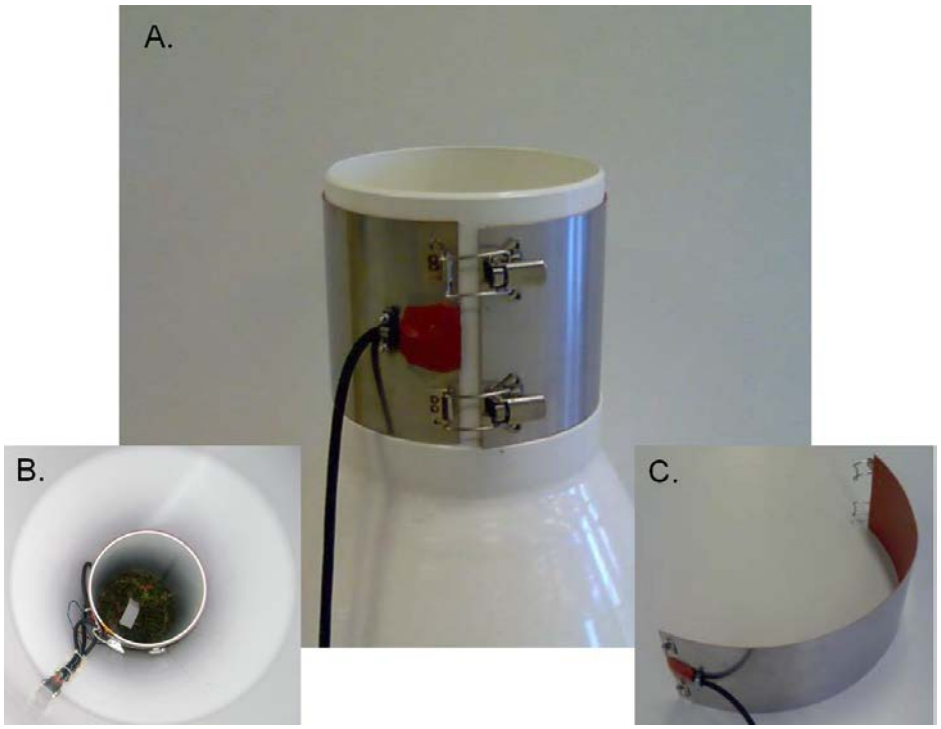


Figure 4.137. Detailed view of heater configuration of Geonor T200-B3 with Canadian heating, including (A) upper heater, (B) lower heater (bottom view), and (C) inlet heater (not used here).

The ambient temperature is measured using a YSI44212 temperature sensor mounted near the Geonor gauge. The rim temperature is monitored at the top and bottom of the Geonor inlet using two thermistors (YSI44003), as shown in Figure 4.138.

The heating of the upper and lower heaters is controlled separately using the ambient and rim temperature. The heating algorithm is activated when the ambient temperature is ≤ +2 °C and the rim temperature is ≤ +2 °C. The algorithm operates to maintain the rim temperature at +2 °C. Each heater is individually powered using a DC power supply. If the power supply is 24 VDC, the heater resistance is approximately 11.5 Ω (each heater) and the power for each heater is approximately 50 W. An optional 48 VDC power supply will produce ~200 W per heater (not currently in use).

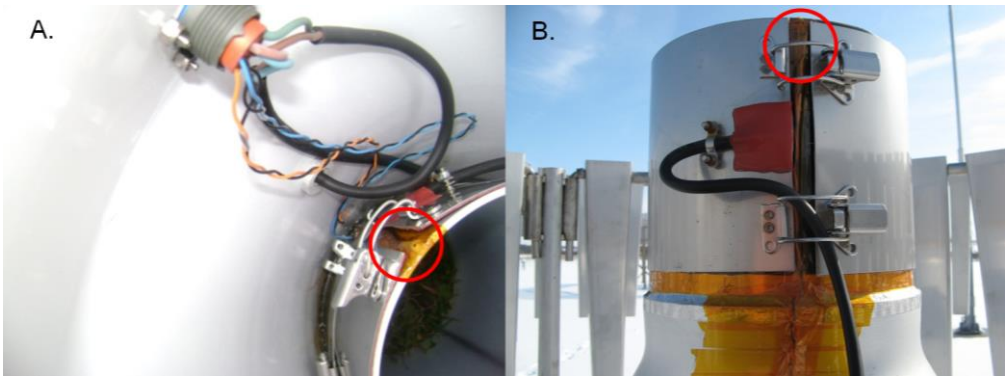


Figure 4.138. Geonor T-200B3 Canadian heating, YSI44003RC (A) lower, (B) upper thermistor.

The CRN heating configuration for Geonor T-200B3 gauges (Figure 4.139) employs two heaters (Minco, Model HR23937) that are fixed to the Geonor orifice on the exterior of the gauge chimney, between the orifice and the shoulder of the gauge, and on the outside of the chimney cylinder, inside the gauge cover (see Figure 4.140). The physical positioning of heaters is similar to that of the Canadian heating configuration (Figure 4.136 and Figure 4.137).

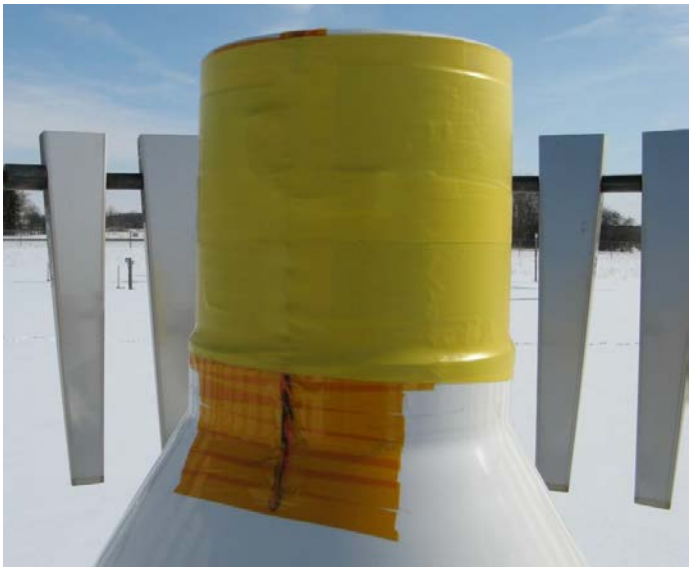


Figure 4.139. Geonor precipitation gauge with CRN heating configuration (exterior view).

The ambient temperature is measured using a YSI44212 temperature sensor mounted near the Geonor gauge. The rim temperature is monitored at the top of the Geonor orifice using the YSI44003 thermistor as shown in Figure 4.141. The temperature is monitored at the bottom of the Geonor orifice for information purposes only. (*Note*: This bottom thermistor is not required for CRN heating.)

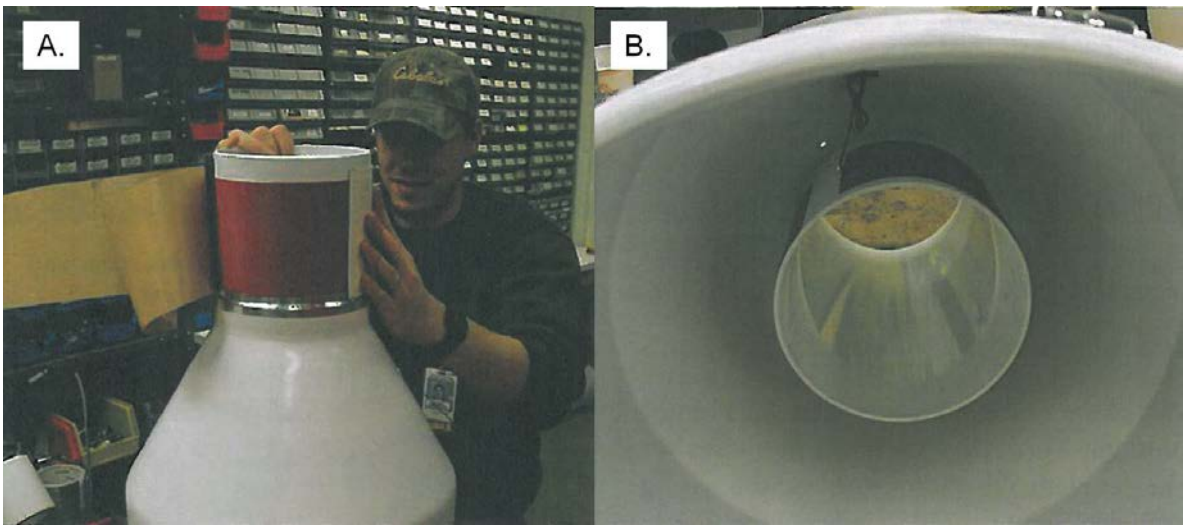


Figure 4.140. Heater configuration for Geonor T-200B3, CRN approach, including (A) the upper heater and (B) the lower heater (NOAA Technical Note NCDC No. USCRN-04-1).

The heating status of both heaters is controlled using the ambient temperature and upper-rim temperature. Similar to the Canadian heating algorithm, the CRN algorithm turns the heating on when the ambient temperature is ≤ +2 °C and the upper-rim temperature is ≤ +2 °C. Both heaters are powered in parallel using a 12 VDC power supply, for which the heater resistance is approximately 5.1 Ω (each heater) and the heater power is approximately 28.2 W (each heater).

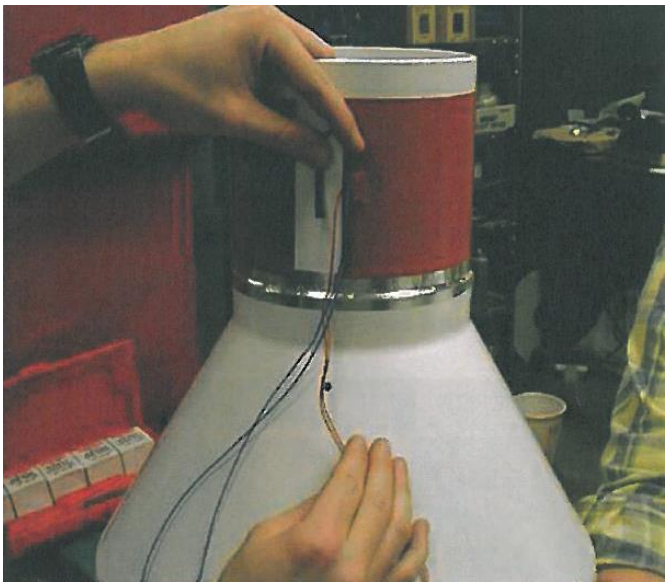


Figure 4.141. Geonor T-200B CRN heating, YSI44003RC thermistor (Hall, M.E., May E., 2004).

In some locations, such as the Haukeliseter site, additional heating power was required to keep the gauge orifice snow-free (see Figure 4.142). To achieve this, the heating cables were connected in serial order to allow for 50 W per element (using 24 VDC).

Figure 4.142. Influence of heating on the DFIR Geonor at Haukeliseter, January 5, 2014.

An example illustrating the monitoring of rim and air temperatures associated with heating at the Bratt's Lake site is shown in Figure 4.143. The heater configuration used at Bratt's Lake was based on previous guidelines; however, a lower ambient-air-temperature threshold of -30 °C was employed. This was done to prevent the heaters from operating at very low temperatures, which would result in unnecessary stress on the 12 V power supplies that were only rated to -20 °C.

At Bratt's Lake and at several of the colder, windier sites (e.g. Caribou Creek), it was observed that the rim temperature could not be maintained at the desired +2 °C when air temperatures dropped below the -10 °C threshold. This is demonstrated in Figure 4.143, which shows the Bratt's Lake DFAR (Geonor) rim temperature when the air temperature was above -10 °C and below -10 °C. Although the heater seemed to work well over the temperature range between +2 and -5 °C, the power was insufficient to operate the heaters below that range, and the rim temperatures dropped below the set point temperature of +2 °C. A solution was implemented prior to the 2014/15 winter: the input voltage to the heaters was doubled from 12 to 24 V, thereby quadrupling the power to the heaters. This was combined with adding heavier gauge electrical wire between the power supplies and the heaters. This increased the efficacy of the heaters in cold and windy conditions, but the heaters still could not maintain the desired +2 °C temperature when the ambient temperature dropped below -5 °C. The heaters did, however, succeed at keeping the rim temperature above the ambient air temperature, which may be beneficial in keeping the rim free of snow and frost. Even if the use of 24 V improved the performance of the gauge, the same features as in Figure 4.143 still appear. An upgrade of the heaters to AC power should be tested in the future (not performed in SPICE).

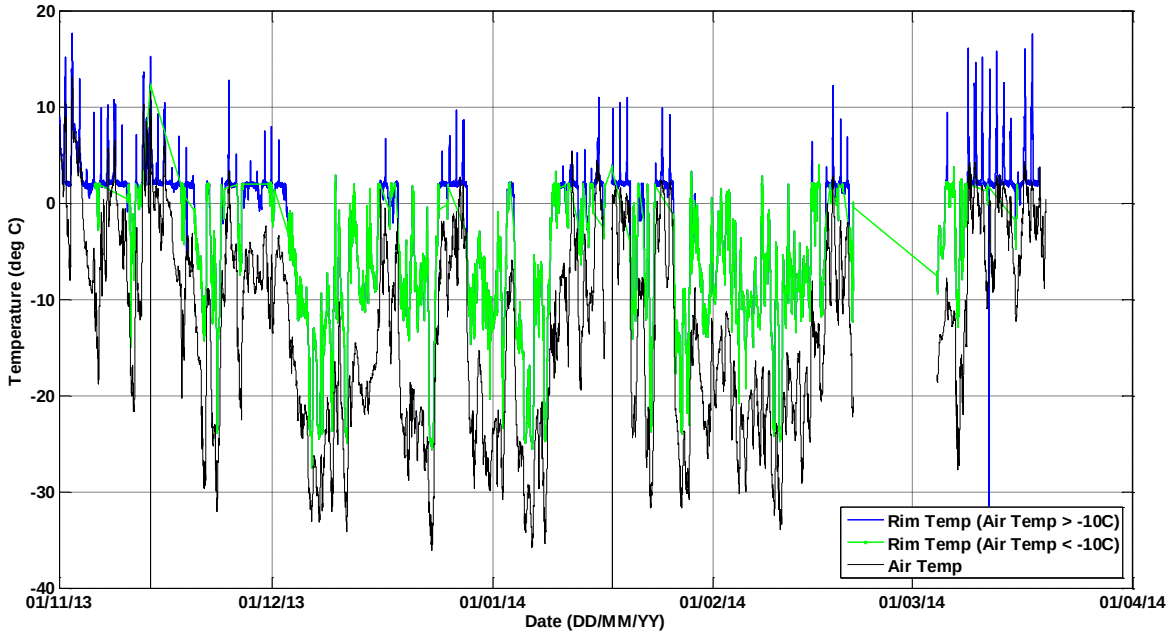


Figure 4.143. DFAR (Geonor T-200B3 gauge) rim and air temperatures at Bratt's Lake during the 2013/14 winter collection period.

4.2.2.2.2 *OTT Pluvio² reference gauges*

The heating configuration of the OTT Pluvio² (Figure 4.144) is based on the standard configuration provided by the manufacturer in which heating is applied to the rim (grey region in Figure 4.145).

Heating is controlled by the load cell and rim temperatures, which are both measured using integrated temperature sensors. Heating can be activated when the load cell temperature is ≥ -40 °C and ≤ +8 °C and the rim temperature is ≤ +2 °C because snow and ice do not typically build up outside this temperature range. (See the operating instructions for Pluvio^2 for details.) The heating power is variable within the active range to achieve the desired rim set temperature, between +2 °C and + 4 °C, as decided during the IOC-4 meeting.

Figure 4.144. OTT Pluvio^2 precipitation gauge with standard heating configuration.

The rim heater is powered using a separate OTT 24 VDC power supply. Using the 24 VDC power supply, the heater resistance is approximately 10.9 Ω (rim heater) and the heater power is approximately 53 W (rim heater).

Figure 4.145. Illustration of OTT Pluvio^2 200 cm^2 gauge. The orifice ring rim region is indicated by the grey shading; this is the heated part of the housing. (OTT Hydrometrie: Operating Instructions OTT Pluvio^2 precipitation gauge).

Figure 4.146 shows an OTT Pluvio^2 gauge at the Sodankylä site with the standard rim-heating configuration and algorithm. It is evident that heating was effective at keeping snow from collecting on the rim. However, snow accumulation on the shoulders of the gauge increases the risk of capping and can generate false precipitation reports if the collected snow is blown into the orifice. This demonstrates that, in this case, the heated surface area is too small to melt all of the snow collecting on the gauge exterior. It was concluded that the combination of a smaller heating surface and pronounced shoulders on the gauge housing increases the risk for snow buildup, which in turn increases the risk of capping and related measurement impacts (see Section 4.2.1.1). This is likely less of an issue for sites characterized by high winds or colder, drier snow, at which the potential for snow buildup is reduced.

Figure 4.146. Buildup of snow on the shoulders of an OTT Pluvio^2 gauge at Sodankylä employing rim heating and the OTT heating algorithm.

An example illustrating the operation of the Pluvio^2 heating algorithm during a snow episode at the Formigal site is provided in Figure 4.147. The rim temperature was set to +4 °C. At the beginning of this snowfall episode, the rim temperature was above 15 °C because it was exposed to sunlight. However, a rapid decrease in temperature followed by precipitation resulted in the activation of heating to maintain the temperature of the orifice in the vicinity of the set point during the episode.

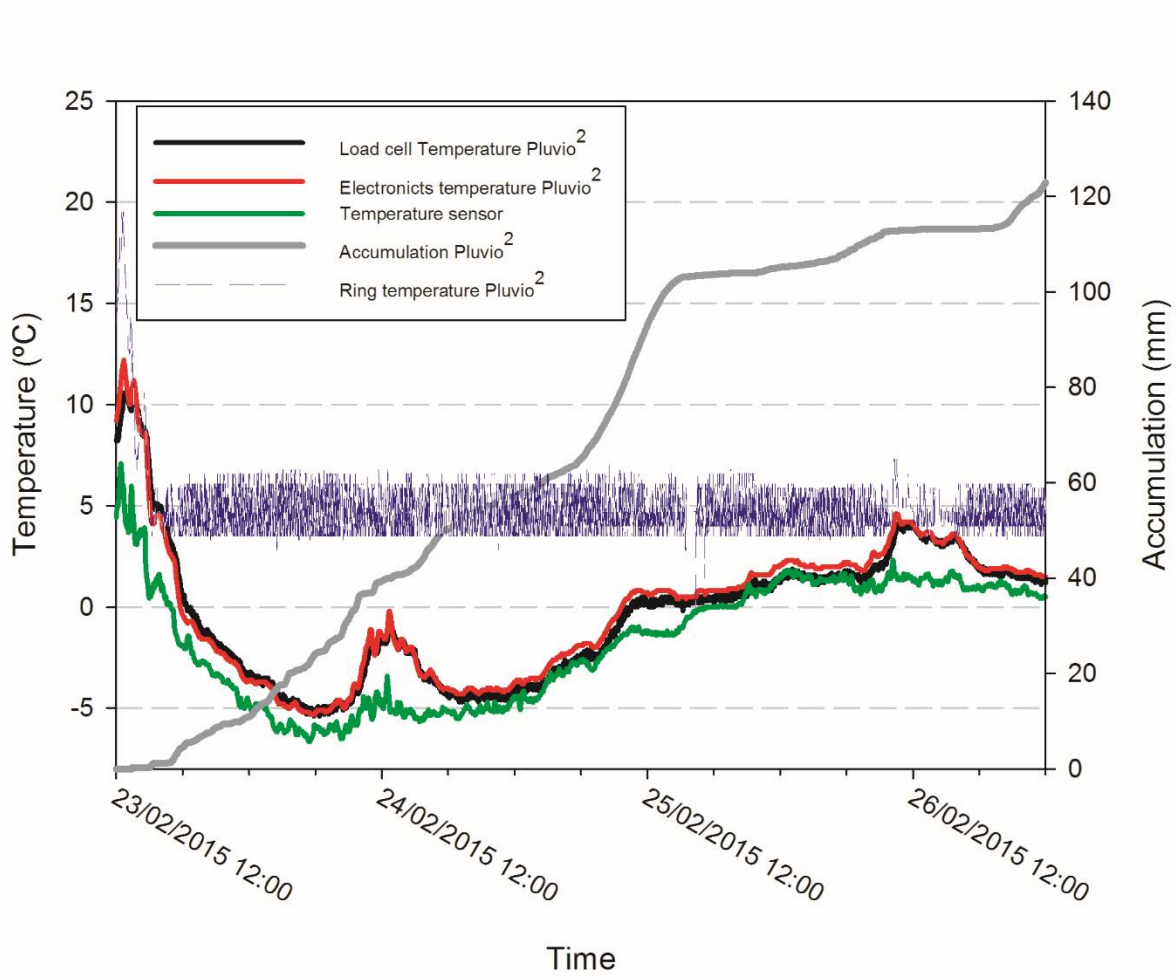


Figure 4.147. Application of OTT Pluvio^2 rim-heating algorithm during a snowfall episode at Formigal in February 2015.

4.2.2.2.3 *Other weighing gauges*

Details of specific heating configurations are available in the site commissioning reports (available at http://www.wmo.int/pages/prog/www/IMOP/intercomparisons/SPICE/SPICE.html) and in the respective instrument manuals. Key features of the heating configurations for the weighing gauge models under test are outlined in the following paragraphs.

The Meteoservis MRW500, MPS TRwS405, OTT Pluvio^2, and Sutron TPG gauges tested all employed heating of the orifice ring (collar), while the Belfort AEPG 600 test gauge heated the interior of the orifice tube. The specific temperature range over which heating is applied can be configured by the user (except the Meteoservis MRW500), with some gauges applying heat only when precipitation is detected within a specified temperature range (Belfort AEPG 600, optional; Meteoservis MRW500, default; Sutron TPG, optional). In the case of the MRW500, a short period of intense shock heating is applied to remove any snow/ice accumulated on the orifice. For the Belfort AEPG 600, MPS TRwS405, and Sutron TPG gauges, the heating power can be adjusted by changing the voltage supplied to the heater(s).

The Geonor T-200B3 and T-200B3MD gauges were unheated as supplied by the manufacturer. All T-200B3 gauges under test employed the CRN heating configuration (Section 4.2.2.2.1), with the

exception of the unheated gauges at Caribou Creek, where electrical interference noise from heater operation was found to impact gauge performance. The T-200B3MD gauges at Bratt's Lake, Caribou Creek, and Marshall were all unheated as provided by the manufacturer; at CARE, the Canadian Geonor heating configuration was used (Section 4.2.2.2.1). At Weissfluhjoch, a Geonor-provided heater was installed on the exterior of the inlet tube at the top of the gauge housing, and heat was applied for 10 minutes out of every hour, as recommended by Geonor.

4.2.2.3 Heating of tipping bucket gauges

The primary function of heaters for tipping bucket gauges is to melt solid precipitation collected by the gauge so that it can pass into the bucket assembly and be measured. In addition to funnel heating for the purpose of melting, the bucket assembly and/or outlet ports are typically heated to prevent re-freezing of precipitation and potential gauge blockage. Even with heating, gauge performance is impacted in solid precipitation conditions, as the time required for melting can result in delays in the reporting of precipitation relative to the reference configurations. These delays increase the potential for missed and false reports over operational time scales. Heating can also cause evaporative and wetting losses, which reduce reported precipitation totals.

Details regarding the specific heating configurations of TB gauges under test and the influence of heating on gauge performance are provided in Section 4.1.2 and in the Instrument Performance Reports in Annex 6. Heating power and configuration vary by gauge type, with varying degrees of user configurability. The selection of a heated TB gauge for a given site must consider the power available, typical conditions at the site, and the application (i.e. near real-time for operations, daily/seasonal totals for water budgets).

4.2.2.4 Heating of non-catchment-type instruments and evaporative plate

Heating of non-catchment-type instruments is used primarily to prevent icing or snow accumulation on the optical components of the instruments. The Thies LPM disdrometer, for instance, applies heating to glass panes on both heads that is used year-round to prevent icing or condensation on the optical lenses. Extended housing heating is available for locations experiencing extreme weather conditions (e.g. high mountains) to prevent against malfunction due to icing and packing of snow. The OTT Parsivel^2 disdrometer has a temperature sensor that identifies when the ambient temperature is below 10 °C to apply heat to maintain the head temperatures at 10 °C or until the maximum current consumption is reached.

Present weather sensors, such as Campbell PWS100 or Vaisala PWD sensors, have hood heaters that are activated below a set ambient temperature between 2 °C and 10 °C. They also have automatic dew heaters inside the optics units to keep the lenses clear. The Vaisala PWD33 EPI sensor has an additional impact sensor heater (on the impact precipitation device), triggered by ambient temperature, to prevent ice and snow accumulation on its surface.

4.2.2.5 Heating of snow-depth sensors

Some ultrasonic snow-depth sensors can be heated to prevent the sensor from being blocked by accumulated snow or frost. During SPICE, both the Sommer USH-8 and the Campbell Scientific SR50A sensors were tested with heaters. The benefits of instrument heating (primarily to de-ice the ultrasonic membrane, but also in some occasion to avoid snow build up on the top of the sensor) are discussed in more detail in Section 4.2.6.1, as well as the merit of heating the mounting infrastructure to prevent snow build up.

Optical snow-depth sensors are heated to extend their operation to colder temperatures. The heater in the Lufft SHM30 sensor is designed to keep the laser diode at an operating temperature > -10 °C. The purpose, according to the manufacturer, is to increase the lifespan of the diode and not to clear snow from the sensor. As noted in Section 4.1.4.3.2, there appears to be an impact of the heater switching on noise levels in the signal strength output from the sensor. The impact of the heating on the snow-depth output appears to be negligible. Snow does accumulate on the SHM30 sensor body during heavy snowfall events in light wind conditions (Section 4.2.6.1). Although it was never shown to hamper the snow-depth measurement, heating of the sensor body could be considered to prevent such an occurrence.

4.2.2.6 Discussion and conclusions

The previous sections have presented heating primarily as a means of mitigating instrument-performance limitations resulting from low-temperature operation. In the absence of heating, snow can build up on the gauge assembly, presenting an avenue for full or partial blockage (precipitation cannot enter gauge orifice) and false reports (snow accumulated on gauge assembly is blown into orifice). It is possible for accumulated snow to be removed manually, but this presents a challenge for unattended and remote sites. For tipping bucket gauges, heating is critical to enable the measurement of solid precipitation. For snow-depth sensors, the sensing elements are heated to prevent snow/frost buildup and enable measurement.

Negative performance effects resulting from heating have also been noted. Evaporative losses of incident precipitation are a key consideration. Another important consideration is chimney effects in which buoyant plumes of warm air generated by heating can alter the flow field of hydrometeors around an instrument. Using lower set-point temperatures and/or heating powers may help to mitigate the potential for negative heating effects, but may also limit the utility of heating. One potential solution is to employ short periods of high temperature/high power heating at regular intervals (e.g. once per day) to remove accumulated precipitation; indeed, this shock-heating approach has been adopted by some instrument manufacturers. More work is needed to examine the potential negative impacts of heating (i.e. reduced catch efficiency) that could be caused by the plume or from hydrometeors evaporating or submimating on the warm inlet before dropping into the collector.

These effects underscore the importance of selecting an instrument with a heating configuration suitable for the expected range of environmental conditions at a given site. The installation of temperature sensors is recommended to monitor instrument heating (e.g. at the orifice of a weighing or tipping bucket gauge) and assess potential over- or under-heating for a given instrument and configuration in a given environment. Instruments with user-configurable heating algorithms are an asset in that the approach employed can be adjusted based on experience. In all cases, it is recommended that instrument users consult with manufacturers in the interests of finding an instrument with heating suitable not only to their site conditions and installation requirements/limitations, but also to their comfort level in terms of hardware modification, programming, and/or field testing.

4.2.3 Antifreeze and oil for weighing gauges

Authors: Jeffery Hoover, Michael Earle, Floor van den Heuvel

The basic principle of a weighing gauge implies that the collected precipitation will remain in the gauge's bucket. Several physical phenomena can impact the quality of the measurement.

Evaporation will decrease the water content and will be registered by the gauge as “negative precipitation”, which must be considered for the assessment of precipitation amounts (e.g. by filtering out in data processing). At low temperatures, water in the bucket can turn into ice and cause inhomogeneity of the content or the gauge to fill with unmelted snow, which in turn will impact the weight measurement.

Weighing gauges used to measure solid precipitation should always be used with antifreeze mixtures. The use of an oil layer on top of the antifreeze is recommended to prevent evaporation of the antifreeze and bucket contents. Mayo (1971) explains that antifreeze solutions used in precipitation gauges should be self-mixing as soon as ice comes in contact with the solution. This ensures fast melting of ice entering the gauge or forming within it. A solution is considered self-mixing if its density is between 0.796 and 1.116 kg/l at 20 °C.

The proper use of antifreeze and oil is important for measurement quality. Antifreeze is often a mixture of either ethylene or propylene glycol and methanol. If the mixture is not saturated with water, it will act as a hygroscopic agent absorbing water from the air in or above the bucket. This will increase artificially, and possibly significantly, the amount of water in the bucket and the corresponding precipitation amounts reported. (See the report on the laboratory experiment for oil and antifreeze testing in Annex 7).

The use of oil is a delicate topic. The layer must have the right thickness to allow new precipitation to fall through and mix with the rest of the bucket content even at cold temperatures, but must be thick enough to cover the entire water surface to prevent the bucket contents from evaporating. The oil must stay in its liquid state at all temperatures. Also, the type of oil and antifreeze is critical, since it can have an important impact on the environment in case of spills.

4.2.3.1 Recommendations

Geonor recommends a 40/60 mixture of either ethylene glycol and methanol or propylene glycol and methanol. A table provided by Geonor provides the amounts required for different minimum temperatures. (See Annex 7) To be effective at the same temperature, slightly more of the propylene glycol mixture must be used relative to the ethylene glycol mixture. Propylene glycol is less toxic than ethylene glycol; however, the methanol recommended for both mixtures is toxic. The mixture requires an oil layer to prevent evaporation, especially of the methanol. Each national operational service uses its own solution; for instance, Canada and the USA are using the propylene glycol mixture suitable for their temperatures, Norway is using the ethylene glycol recommendation, and New Zealand is using a 50/50 mixture of propylene glycol and methylated spirits.

OTT Hydromet recommends using Powercool for the Pluvio^2. Powercool is hygroscopic and must be mixed with water to prevent the antifreeze from absorbing water from the environment. An oil film is not recommended, since the gauge firmware is able to filter evaporation out of the dataset. Here again, different solutions are used among operational services; for instance, Switzerland is not using Powercool for its Pluvio^2 gauges, but rather a propylene glycol and water mixture that has a freezing point range from -40 °C to -12 °C (for a filled bucket). Note that the self-mixing abilities of Powercool or propylene glycol/water mixtures are not well characterized, and require further assessment.

The specific antifreeze/oil solutions for different networks or countries vary depending on the range of temperatures experienced and national regulations for the use and/or disposal of materials. It is expected that the different mixtures and volumes discussed in this section will not affect the response obtained by the gauge in terms of measurement accuracy, such that results from different

sites remain comparable. Based on the experience gained by participating members during the intercomparioson, the following general recommendations for the use of antifreeze and oil in weighing gauges have been developed:

- An oil film is recommended for all gauges and antifreeze mixturers to prevent evaporation. This will also mechanically minimize the hygroscopic effect of the Powercool or propylene glycol/water mixture. The amount of initial water could be reduced when using oil as a top layer (thus increasing the bucket capacity), since the oil layer would prevent hygroscopic accumulation.
- The use of synthetic oil is recommended; several countries have reported that experiments with less dangerous oils were not successful. However, synthetic oil must be used with great care (in particular when emptying a gauge) since oil spills can have a dramatic impact on the environment. Careful handling and disposing of waste in accordance with local regulations is mandatory.

4.2.4 Precipitation measurements in areas with high winds and/or complex terrain

Author: Mareile Wolff

Precipitation measurements in areas characterized by high winds can present additional challenges. First, the precipitation loss of many gauges is dependent on wind speed. Thus, the precipitation loss in high-wind areas can be severe. Second, high winds can cause vibrations on the gauge structures, adding to the measurement noise. Third, blowing snow events may occur more often, which can enhance artificially the reported precipitation amounts.

Accurate wind-speed measurements are crucial for the evaluation of gauge undercatch for snowfall. Terrain complexity adds to the problem, as it is not possible to measure wind speed at exactly the same spot where the precipitation measurement occurs. Thus, some homogeneity of the wind field needs to be assumed for the application of transfer functions, which adjust accumulation data for the effects of wind speed (and temperature/precipitation type).

This section describes challenges associated with high winds and complex terrain, demonstrates the magnitude of their impacts on precipitation measurements, and provides recommendations for how to mitigate those impacts.

4.2.4.1 Measurement noise due to high winds

In general, raw data from weighing gauges show some level of noise, which can be differentiated between high frequency and low frequency noise. Temperature dependencies of the weighing gauge sensor (vibrating wire or load cell) typically produce low frequency noise, usually in phase with temperature changes during the day. High frequency noise is often due to electromagnetic disturbances in the area. Proper grounding is a necessity, and cables should be shielded and as short as possible. Instruments should also be shielded to prevent electromagnetic disturbances caused by other instruments or equipment. Additionally, mechanical noise can occur due to vibrations of the mountings and instruments.

Figure 4.148. to Figure 4.150. illustrate high frequency instrument noise using data from the Haukeliseter site in Norway between April 21 and 24, 2015. The data included an extended period with very high wind speeds (>10 m/s) and almost no precipitation.

The noise is illustrated by the difference between each measurement (here 6-second data) and the 30-minute average of the bucket weight. The small amount of precipitation detected at the end of

the period was well below the noise level for these 30-minute averages and is therefore negligible. This approach minimizes the influence of low frequency noise, which has a period on the order of a day. Values of two times the standard deviation of the differences are considered to be representative of the noise levels within each 30-minute period. The resulting values are at the upper end of the deviations experienced over that period. The magnitude of noise appears to increase at wind speeds of 8 m/s and higher (marked in red in Figure 4.148. to Figure 4.150.).

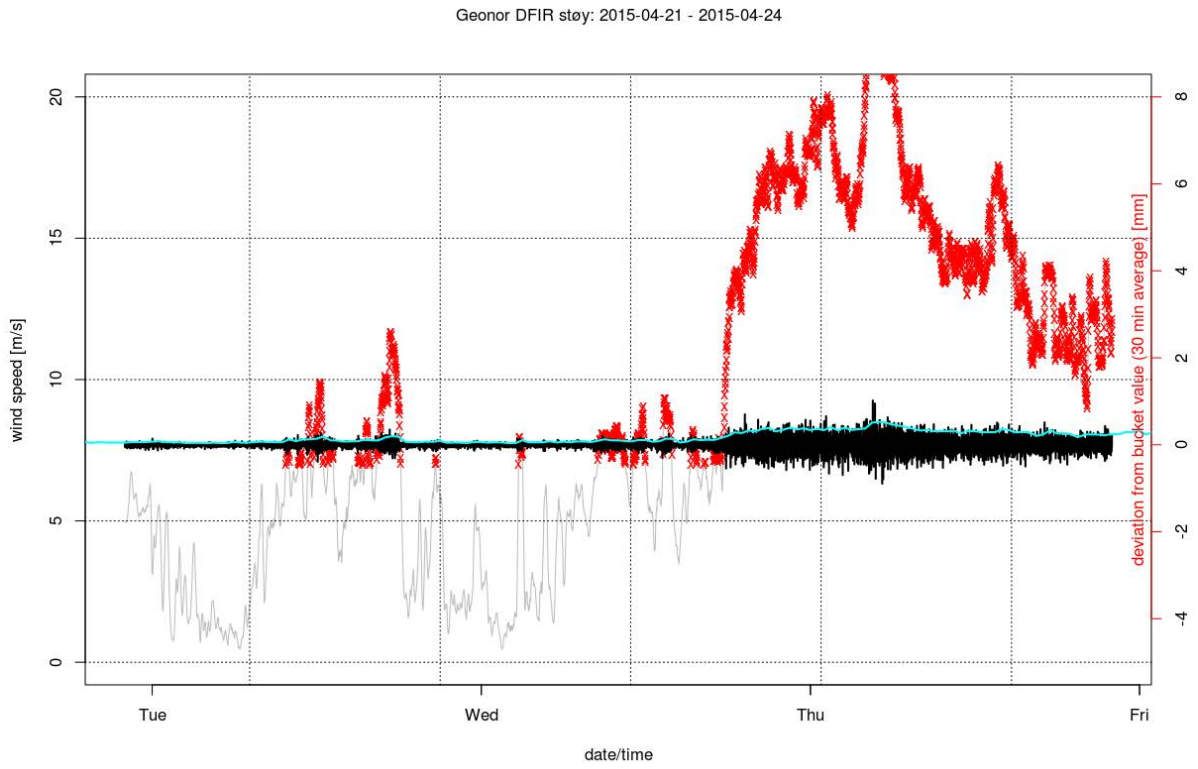


Figure 4.148. Noise analysis for the Geonor in the DFAR at Haukeliseter, Norway, April 21-24, 2015. The noise is shown as the difference between each datapoint (6-second, three-wire averaged bucket content in mm) and the 30-minute average (black, right axis). The cyan curve represents two times the standard deviation of these differences and reflects the order of magnitude of the noise during that period. Additionally, the wind speed (left axis, gray curve) is shown for the same period of time. Wind speeds greater than 8 m/s are marked in red.

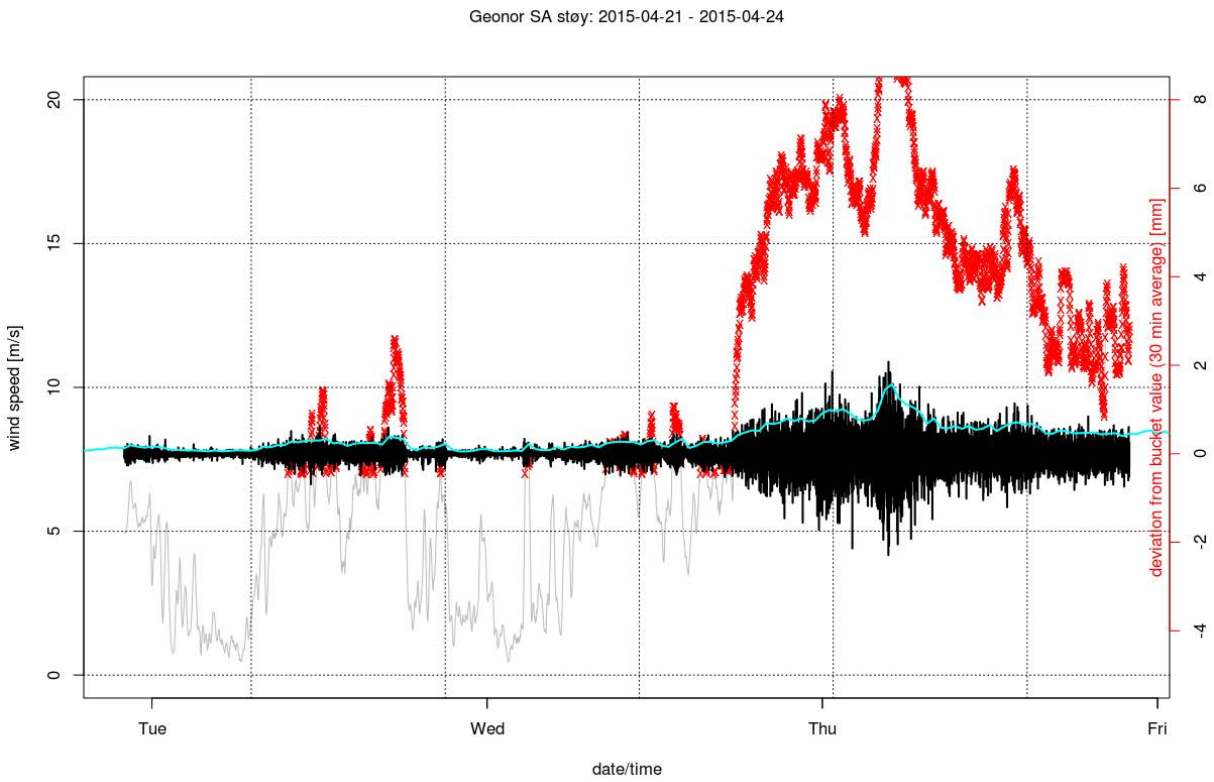


Figure 4.149. Noise analysis for the Geonor in single-Alter shield at Haukeliseter, Norway, April 21-24, 2015. The noise is shown as the difference between each datapoint (6-second, three-wire averaged bucket content in mm) and the 30-minute average (black, right axis). The cyan curve represents two times the standard deviation of these differences and reflects the order of magnitude of the noise during that period. Additionally, the wind speed (left axis, gray curve) is shown for the same period of time. Wind speeds greater than 8 m/s are marked in red.

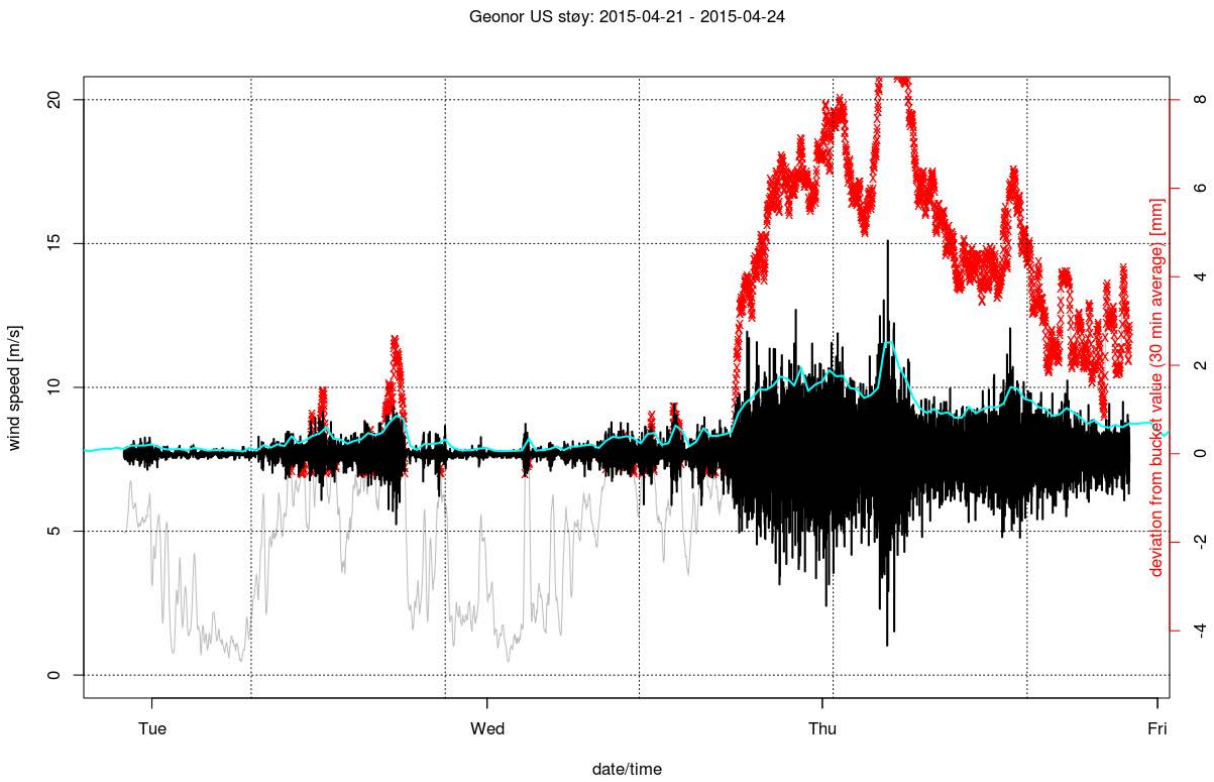


Figure 4.150. Noise analysis for the unshielded Geonor at Haukeliseter, Norway, April 21-24, 2015. The noise is shown as the difference between each datapoint (6-second, three-wire averaged bucket content in mm) and the 30-minute average (black, right axis). The cyan curve represents two times the standard deviation of these differences and reflects the order of magnitude of the noise during that period. Additionally, the wind speed (left axis, gray curve) is shown for the same period of time. Wind speeds greater than 8 m/s are marked in red.

The magnitude of noise is largest for the unshielded gauge and smallest for the gauge in the DFAR. The data demonstrate that the SA shield effectively reduces the magnitude of noise due to wind speed relative to the unshielded configuration. The SA shield at Haukeliseter is mounted directly to the gauge post; mounting the shield separate from the post would likely reduce further the magnitude of noise. Indeed, some national weather services employ wind shields mounted separately from gauge posts (e.g. Canada).

In Figure 4.151. , the dependency of the 30-minute noise levels (again, twice the standard deviation of noise within 30 minute periods) on the average wind speed over the same time periods is shown. Precipitation events with accumulation > 0.25 mm are removed, as those values represent accumulation, rather than noise, within 30-minute periods.

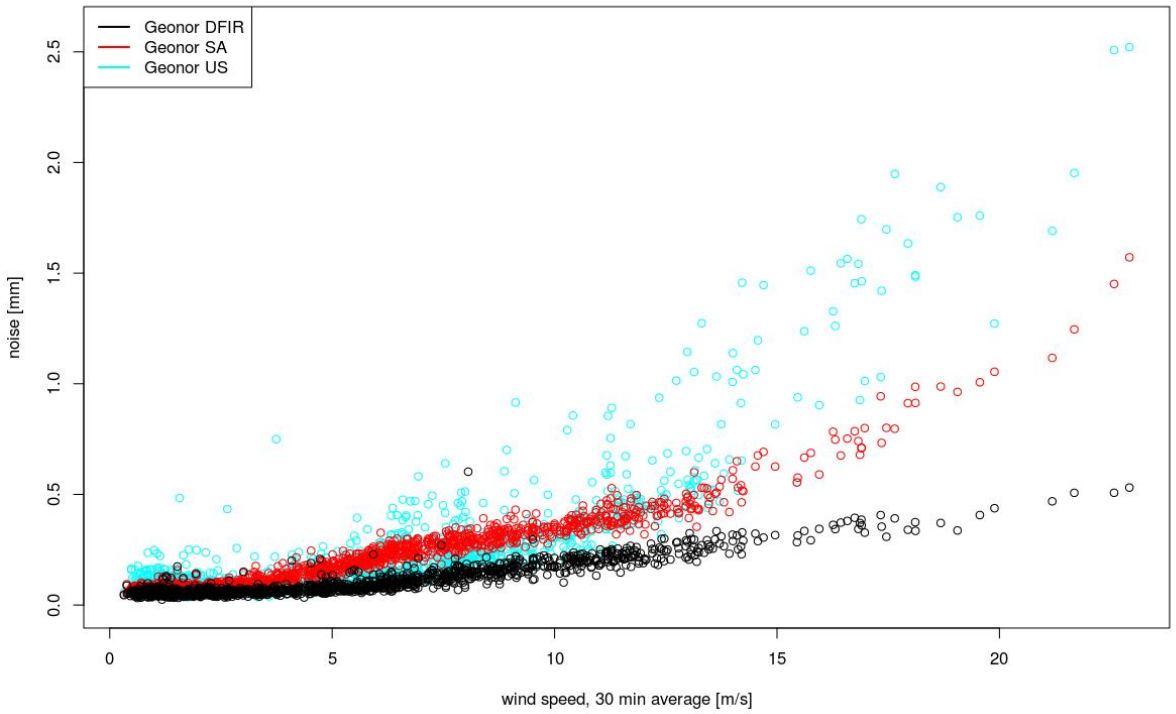


Figure 4.151. Noise level (twice the standard deviation for 30-min periods) of 6-sec data for Geonor gauges in DFAR (black), single-Alter (red), and unshielded (cyan) configurations as a function of wind speed (at 10 m, 30-min average). Data from Haukeliseter, Norway, April 21-24, 2015.

For all gauge configurations, the noise levels increase at higher mean wind speeds. The unshielded gauge shows the highest noise levels relative to the other configurations at a given wind speed, while the gauge in the DFIR-fence shows the lowest noise levels. At extreme wind speeds > 20 m/s, the unshielded gauge shows noise levels as high as 2.5 mm, while the DFIR-shielded gauge noise levels remain within about 0.5 mm.

Averaging the measurements can reduce the effects of high frequency noise. Figure 4.152 shows a similar noise level plot as Figure 4.151. , but for 1-minute averages of the 6-second data. The data are subject to no further quality control or processing beyond the averaging. Indeed, averaging reduces the magnitude of noise for each gauge configuration. The same relative magnitude of noise, with unshielded > single-Alter > DFIR-shielded, is observed as for the 6-second data.

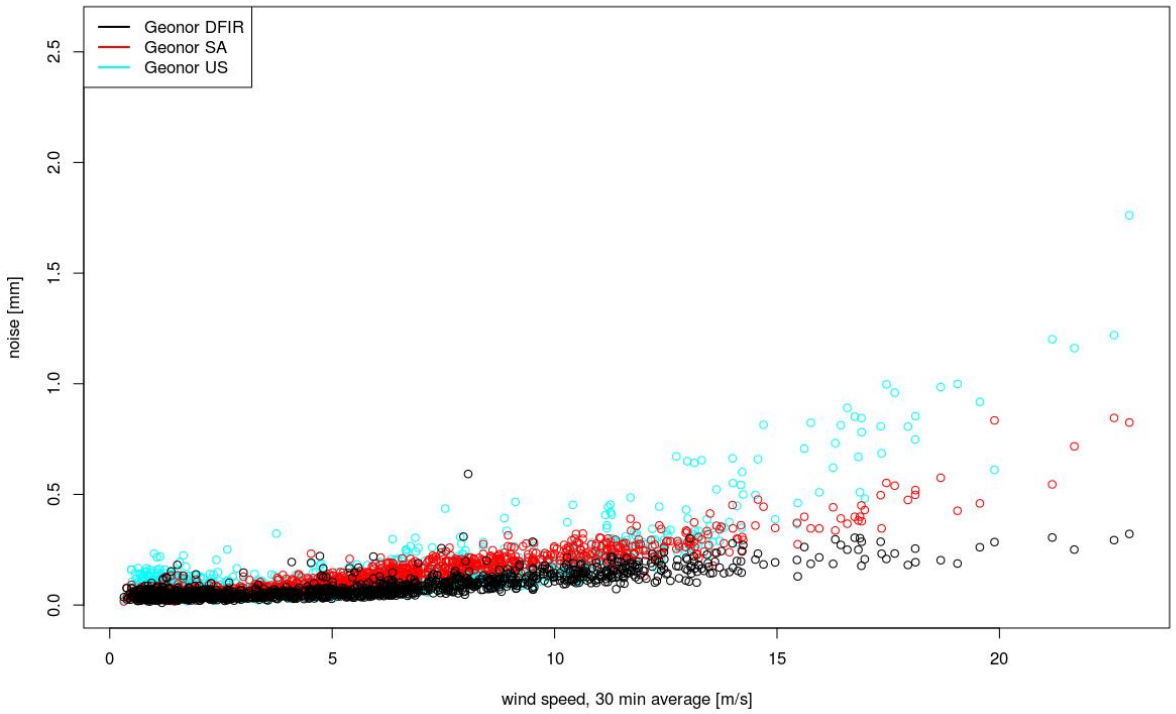


Figure 4.152. Noise level (twice the standard deviation for 30-minute periods) of 1 minute unfiltered data for Geonor gauges in DFAR (black), single-Alter (red), and unshielded (cyan) configurations as a function of wind speed (at 10 m, 30-minute average). Data are from Haukeliseter, Norway, April 21-24, 2015.

The quality-control procedures for SPICE include a Gaussian filter of the reference data (see Section 3.4) to reduce the influence of high frequency noise. In Figure 4.153, a Gaussian filter has been applied to the 1-minute average data from Figure 4.152. The noise levels are reduced further for each configuration, while again, the same relative magnitude of noise is maintained (unshielded > single-Alter > DFIR-shielded). For the single-Alter-shielded and unshielded gauges, the dependence of noise level on wind speed is not significant below about 5 m/s. For the gauge inside the DFIR-fence, this dependence is not significant for wind speeds below 10 m/s.

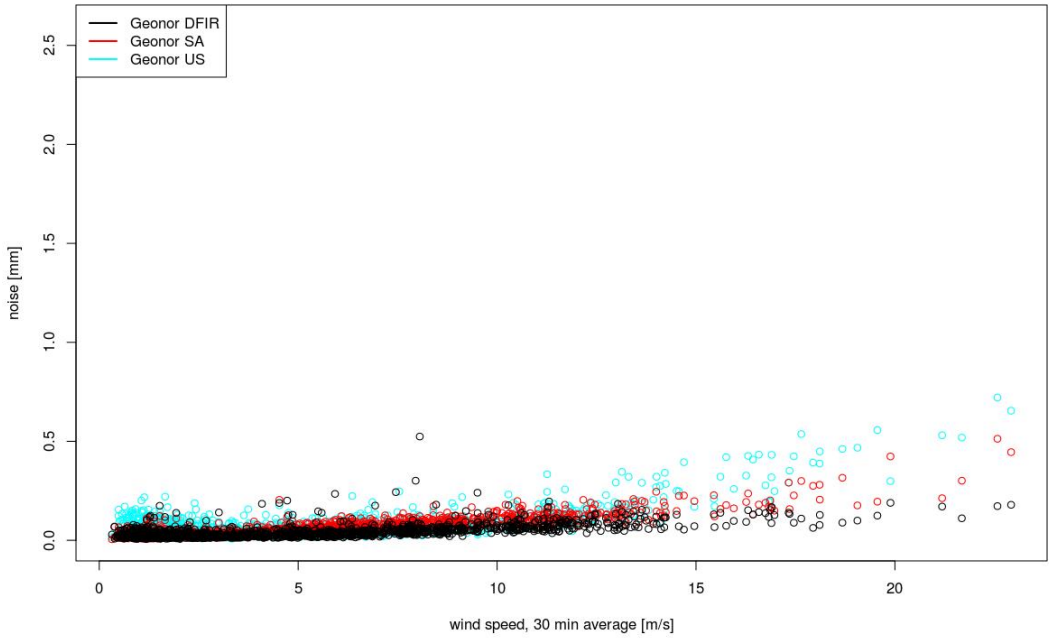


Figure 4.153. Noise level (twice the standard deviation for 30-minute periods) of quality controlled and Gaussian filtered 1-minute data for Geonor gauges in DFAR (black), single-Alter (red), and unshielded (cyan) configurations as a function of wind speed (at 10 m, 30-minute average). Data are from Haukeliseter, Norway, April 21-24, 2015.

The Haukeliseter data were characterized by high noise levels that could not be reduced significantly during the SPICE period. The instruments were properly grounded and shielded cables were used. Due to the layout of the site, very long cable lengths were required, which might be one of the reasons for the observed noise. At the same time, extended periods with very high winds added to the problem, as they increased further the noise levels.

At sites characterized by lower noise levels, the magnitude of mechanical noise and its dependency on wind speed are also lower, even if slightly increased for higher wind speeds. For example, the 30-minute noise levels for quality-controlled 1-minute Geonor data for unshielded, single-Alter shielded, and DFIR-shielded configurations at Marshall for January 2015 are shown in Figure 4.154. . Noise levels are about a factor of 10 smaller than at Haukeliseter and, consequently, the observed increases at higher wind speeds were low.

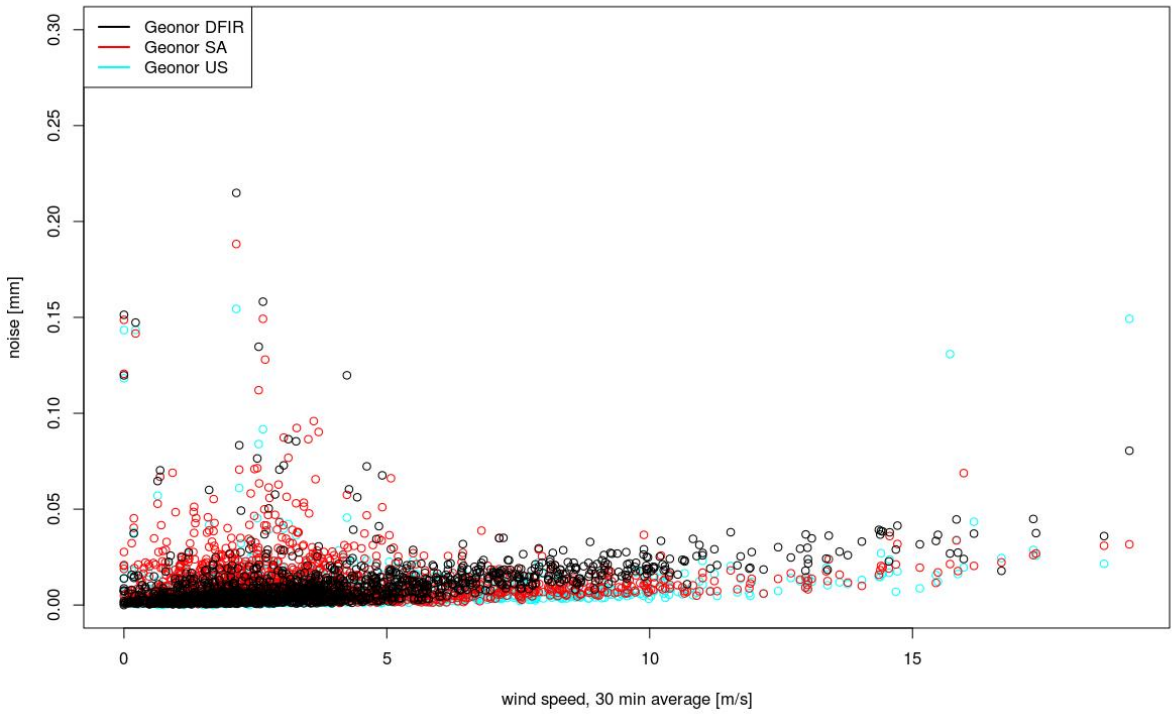


Figure 4.154. Noise level (twice the standard deviation for 30-minute periods) of quality controlled and Gaussian filtered 1-minute data for Geonor gauges in DFAR (black), single-Alter (red), and unshielded (cyan) configurations as a function of wind speed (at 10 m, 30-minute average), Marshall. Data are from Marshall, USA, January 2015.

Averaging and filtering are effective ways of reducing the impact of high frequency noise on precipitation measurements. These methods also help to reduce mechanical noise at higher wind speeds. The most effective way to reduce noise, however, is to reduce its influence before any digital filtering is applied. Proper grounding, shielded cables and other electromagnetic shielding or filtering are recommended. The use of digital interfaces (i.e. SDI-12) for data transmission from the instrument has been shown to reduce electrical interference when transmission lengths are long (more than a few metres). Further, a solid and stable mounting of the instrument, use of wind shields and, if possible, separate mounting of the gauge and shield assembly will all help to reduce mechanical noise caused by wind.

4.2.4.2 DFAR/DFIR under high wind conditions

As mentioned in Section 3.2.1, the bush gauge is considered to be the best available measurement (closest to truth) for solid precipitation. Because of the obvious limitations to “build” and maintain a similar bush gauge at different stations in various climate zones, the DFIR was designated as the working reference for the Solid Precipitation Measurement intercomparison study during 1986 to 1993 (WMO/TD –No. 872, 1998). Within WMO-SPICE, the DFIR continued to be the manual working reference R1. Additionally, the automated adaptation, the DFAR, was accepted as the automated working reference R2 for WMO-SPICE.

Following earlier studies by Golubev (1985), Golubev (1989), Metcalfe and Goodison (1992) and Yang (1995) on the relationship between the bush gauge and DFIR, Yang (2014) re-examined and updated the relationship. Results of this comparison are also summarized in section 3.2.1. An automated bush gauge configuration was installed at the Canadian host site at Caribou Creek in 2012, and comparisons between automated gauges within the bush and a neighboring DFAR were performed (see Section 3.2.1).

On average, the bush gauge caught approximately 5% to 6% more than the DFIR at Valdai (Section 3.2.1 and Yang (2014)) and between 7% and 11% less than the DFAR at Caribou Creek (Section 3.2.1). For the Valdai data, the undercatch of the DFIR increased slightly with increasing wind speed, whereas no significant dependency on wind speed was observed for the DFAR data from Caribou Creek.

No measurements at wind speeds higher than 9 m/s are avaialable in the Valdai and Caribou Creek datasets considered above, precluding any statements regarding the undercatch of a gauge inside a double fence configuration at higher wind speeds.

Based on the following reasoning, a simple extrapolation of the DFIR/DFAR bush gauge transfer functions toward higher wind speeds is not recommended.

- Wind-speed measurements from inside and outside the DFIR construction at Haukeliseter indicate that the wind speed is effectively slowed, irrespective of the incident speed. Figure 4.155. shows wind speeds (1-minute average) for a 30-day period, measured both inside and outside the DFIR-fence. On average, over the whole period, the wind speeds measured inside the DFIR-fence were 22% lower than the corresponding values measured outside the fence. The 1-minute averages of wind speeds measured inside the DFIR-fence rarely exceed 5 m/s, except for a short period of time when the outside wind-speeds exceeded 20 m/s.
- Further, catch-efficiency plots of weighing gauges, which include measurement data for the entire range of wind speeds, show a halt in their observed decrease for wind speeds greater than 6 to 8 m/s, (see Figure 4.11 to Figure 4.13). This indicates that an extrapolation of the low- to moderate- wind relationship of the catch efficiency may lead to incorrect values at higher wind speeds.

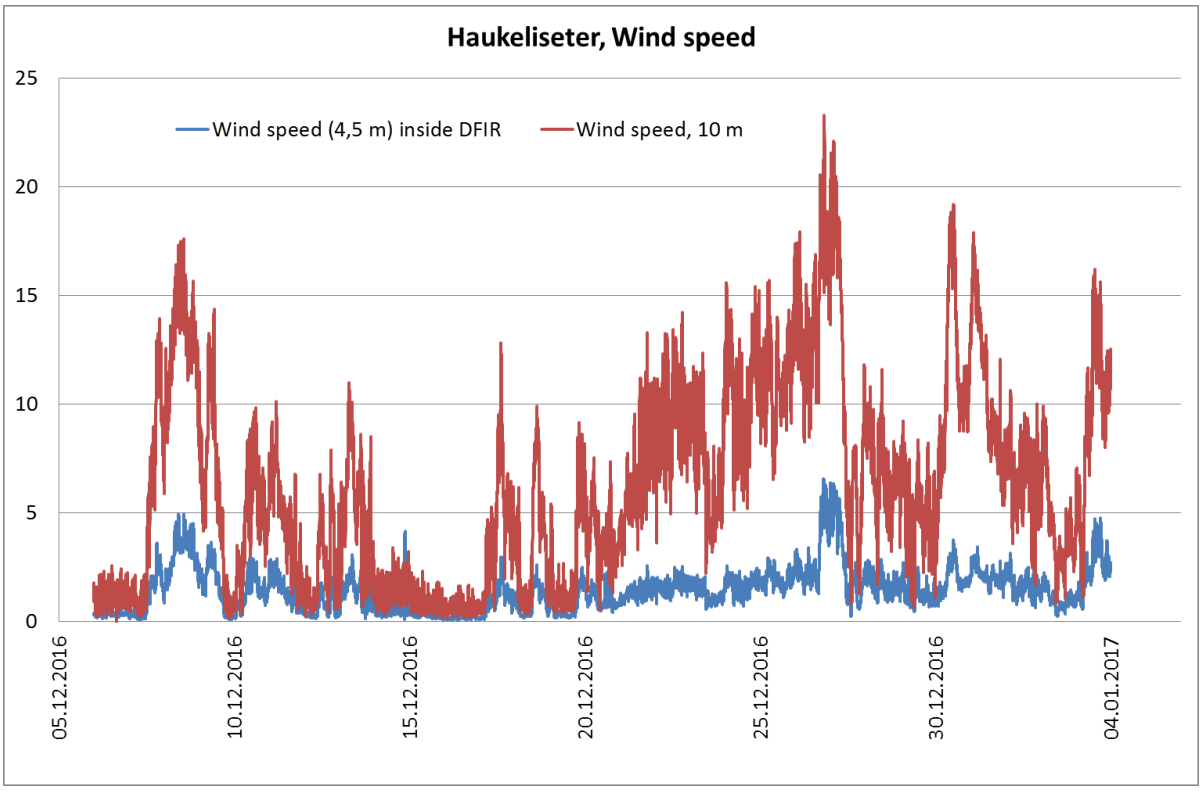


Figure 4.155. Wind-speed measurements at Haukeliseter for a 30-day period, starting December 6, 2016. The red curve shows the 10-m standard wind measurements, while the blue curve shows wind speeds as measured inside the DFIR fences.

As a result, all transfer functions in this report are developed to adjust for the difference between the gauge being tested and the working reference (DFAR). This is in contrast to Goodison et al. (1998), where all transfer functions included the adjustment to the bush-gauge reference.

Applying the recommended WMO-SPICE transfer functions to precipitation measurements from a given gauge/shield configuration will, therefore, result in a precipitation amount similar to what would be measured by a DFAR. Therefore, in most cases, it would underestimate the measurements of a bush gauge according to the Valdai intercomparisons.

On average, the undercatch can be up to 10% for wind speeds under 9 m/s (based on data for Valdai). For wind speeds between 10 m/s and 20 m/s, the undercatch can only be estimated and is believed to be within 20%.

Exactly how well solid precipitation measurements of the DFAR field reference represent the ground truth at high wind-speeds remains a knowledge gap. Further studies are necessary to quantify the high-wind undercatch of a DFAR. Improvements can be achieved by more direct comparison measurements, including a significant amount of high wind cases, or by the application of computational fluid dynamics (CFD) methods to simulate the relationship.

4.2.4.3 Blowing snow

Blowing snow events are observed frequently in areas with high winds but remain difficult to detect automatically; human observers are the most reliable detection method. Blowing snow events can occur with or without snowfall. Blowing snow occurs with varying vertical extent, from snow just

above the ground to tens of metres high. If the level of blowing snow reaches the orifice of a precipitation gauge, it will be reported as accumulated precipitation; thus, gauges mounted at different heights may be impacted differently by blowing snow. Further, the local wind patterns caused by the small-scale topography, construction, or vegetation at the site can also affect the distribution of blowing snow.

If blowing snow develops, it is not solely dependent on the wind speed and its local distribution, but also on the kind of snow on the ground. Newly fallen, dry snow is more easily transported than icy or wet snow layers. Wind-speed thresholds for blowing snow occurrence are, therefore, only approximations, and are not absolute indications for the occurrence of blowing snow.

Goodison et al. (1998) treated blowing snow events as separate events, which was possible because of the availability of manual observations at all sites. During WMO-SPICE, which had a focus on automated measurements, manual observations were not available on all sites. The application of general wind-speed thresholds for the occurrence and/or vertical extent of blowing snow would not capture the individual aspects of the different sites and the meteorological situation in each case, and would, therefore, provide only a rough indication of blowing snow events.

Blowing snow observations are often classified by their vertical extent, i.e. under or over “eye height”. See a description of blowing-snow event statistics for Haukeliseter by Wolff et al. (2010). Naturally, the higher-reaching blowing-snow events are more extreme events, and occur less often than blowing-snow events closer to the surface. In areas with a high probability for blowing-snow events, a higher installation of the gauges is recommended. At Haukeliseter and Weissfluhjoch, gauges were mounted at 4.5 m and 3.5 m, respectively (also taking into account the total snow pack).

In WMO-SPICE, the large amount of complementary and ancillary data made it possible to identify blowing-snow events. For example, Figure 4.156. which shows the snow-regime catch ratio versus wind speed for the 1500 mm capacity Geonor gauge at Bratt’s Lake. About 10 datapoints around 10 ±1 m/s wind speed are clearly separated from the main data set, which shows the typical decrease of catch ratio with increasing wind speed. Those datapoints, which have a higher catch ratio than expected for the measured wind speed, occurred on two days during which a nearby manned station reported significantly reduced visibility and blowing snow. This suggests that the unusually high catch ratio values are caused by additional accumulation from blowing snow.

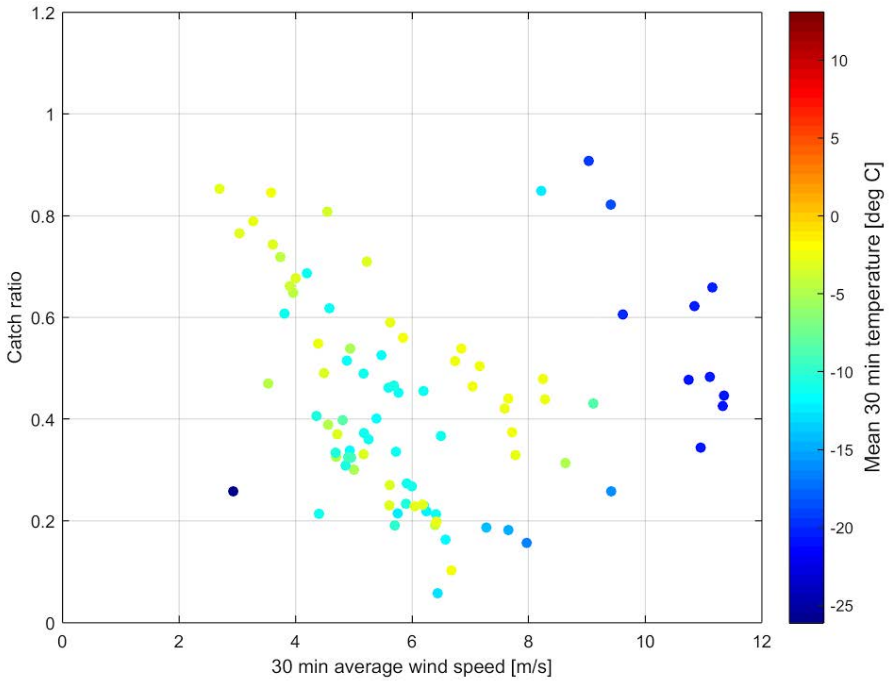


Figure 4.156. Catch ratio (SA/DFIR) as a function of mean wind speed for 1500 mm capacity Geonor gauge at Bratts Lake for snow events, only. The high catch ratios observed at wind speeds between 9 m/s and 11 m/s wind speeds were recorded during a blowing-snow event.

At an operational site without a manual observer, blowing-snow events will, in most cases, be undetected. These events can introduce additional errors to the precipitation data set due to the extra accumulation. If applying a transfer function, that error will be exacerbated by overcorrecting the precipitation amount.

It is, therefore, crucial to have information about the statistical occurrence of blowing-snow events and their typical vertical extent at a given site. A simple technique to avoid or minimize those errors in the data set is to mount the precipitation gauge at a level above the expected vertical extent of blowing snow. This might also be necessary for areas with large annual snow accumulation in order to ensure enough vertical clearance for the gauge (i.e. to avoid instances of gauges being buried by snow).

Mounting precipitation gauges at higher levels is common practice for the precipitation sensor networks operated by the Norwegian Hydropower Companies, often located in mountainous areas and used to calculate the precipitation over a catchment connected to a water reservoir for hydropower production.

Further work on developing instruments for the automatic detection of blowing snow is highly encouraged.

4.2.4.4 Measuring snow in complex terrain

Goodison et al. (1998) emphasized the importance of homogeneity of the test sites. In addition to gauge design and configuration, the comparability of measurements by gauges at different locations on a site may be influenced by characteristics of the site that may result in an uneven distribution of

the precipitation. This is an important consideration for the analysis of data within WMO-SPICE and their usage for the development of transfer functions.

By extension, the homogeneity of the site plays a major role when applying transfer functions to adjust the measured precipitation. It is rarely possible, and not recommended, to have the wind sensor and the precipitation gauge co-located. It is, therefore, necessary to make sure that the wind measurements are representative of the conditions at the gauge to a very high level so that transfer functions can be applied to the measured precipitation.

Typically, a homogeneous site is characterized by:

- relatively flat terrain;
- a large enough area that instruments can be spaced from each other at a distance that minimizes mutual influences (i.e. disturbing the general wind flow);
- a large distance to obstacles that may cause wind effects.

Further, nearby instruments should be placed along a line perpendicular to prevailing storm winds to further minimize any disturbances of the wind fields.

In reality, this is not always possible. A few sites within WMO-SPICE were situated in complex terrain, which may have resulted in a non-evenly distributed wind field. For the derivation of transfer functions, data influenced by known shadowing effects of installations or surrounding topography were not used.

At sites where two (or more) wind sensors were distributed around the site, a simple homogeneity test can be performed. Figure 4.157. and Figure 4.158 show scatter plots from wind sensors at Haukeliseter and Weissfluhjoch, respectively. The homogeneity of wind conditions at Haukeliseter and inhomogeneity of wind conditions at Weissfluhjoch are evident. Data from precipitation gauges at each site are also provided to illustrate the influence of wind homogeneity/inhomogeneity on the variability of precipitation measurements.

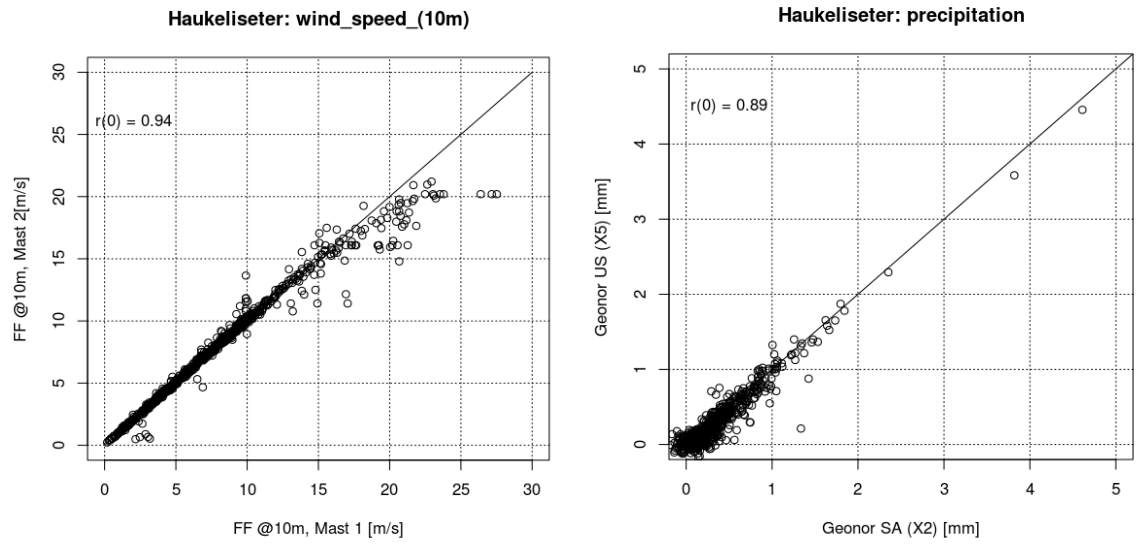


Figure 4.157. Scatter plot comparing two different sensors for wind speed (left) and precipitation (right) at Haukeliseter.

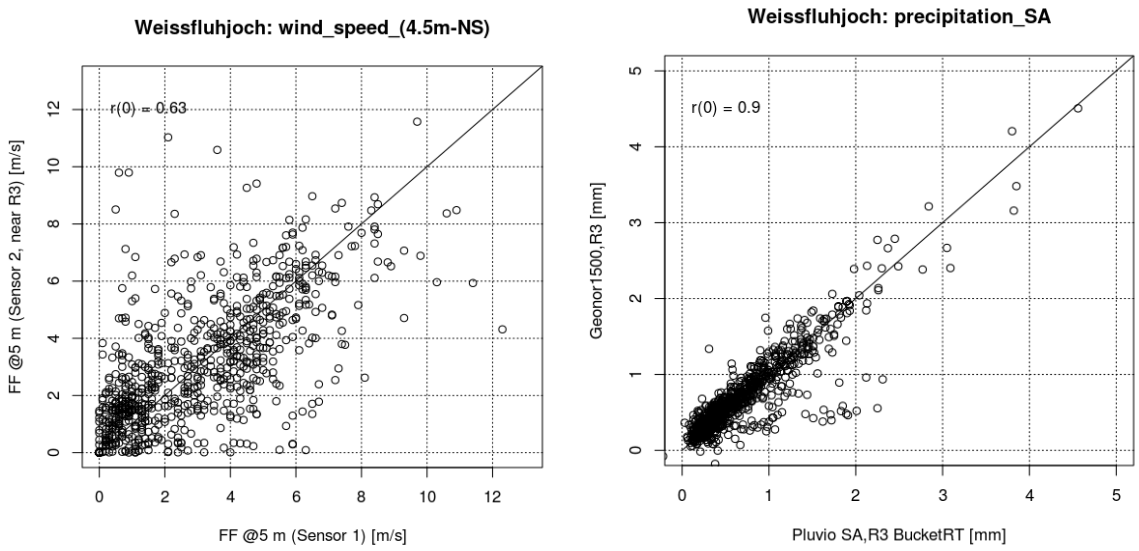


Figure 4.158. Scatter plot comparing two different sensors for wind speed (left) and precipitation (right) at Weissfluhjoch.

4.2.5 Detectability of light precipitation measurements

Author: Eva Mekis

Assessing the detectability of light precipitation and the quality of light-precipitation measurements is important, as precipitation events occur with characteristically low intensities in many geographic areas, particularly those at high latitudes. Historically, light precipitation was included in terms of "trace" precipitation reports for manual gauges, and represented events with non-significant accumulation (i.e. below the gauge sensitivity or minimum measurable amount) (Patterson, 1930; Woo and Steer, 1979; Metcalfe et al., 1994; Mekis and Hogg, 1999; Mekis 2005; Mekis and Vincent, 2011). The International Meteorological Vocabulary (IMV, WMO, 1992) defines "trace precipitation" as accumulated amounts of less than 0.1 mm. For automated precipitation measurements, the "trace" designation is no longer applicable, as measurements can be taken at higher temporal resolutions and with higher sensitivity, depending on the specific gauge and configuration employed. For this reason, the "light precipitation" terminology was adopted. This section presents a comparative analysis of light precipitation events as reported by selected gauge configurations at three WMO-SPICE test sites.

4.2.5.1 Data and methodology

To be consistent with the assessment approach used elsewhere in WMO-SPICE, accumulated precipitation totals observed within specified time periods were used. Any precipitation amount greater than 0 mm and smaller than 0.25 mm from the reference gauge will be referred to as light precipitation. A light precipitation event (LPE) can differ depending on the assessment interval (time period of accumulation). A light event can become a regular event in a longer period by a few minutes of additional accumulation. As a starting point, three time periods for LPEs were studied. The 30-minute LPE period was selected to be comparable with precipitation events considered elsewhere in this report. In operational practices, most often the 1 hour interval is applied and,

historically, the manual observations are taken over 6 hour- or 12-hour periods; so 60-minute and 360-minute LPE periods were also considered.

The detection of light precipitation depends on several factors, including the weather conditions (local climate, wind, temperature, etc.), installation characteristics (gauge type, shield, heater, etc.), and site characteristics (proximity to trees, buildings, slope, etc.).

For automated gauges, the natural fluctuation of accumulation values around zero (noise) can be mistakenly interpreted as light precipitation. To mitigate this issue, a minimum reference gauge accumulation value of 0.1 mm was chosen for LPEs for all time intervals. The 0.1 mm threshold is in agreement with the noise level detected for reference gauges during non-precipitating events (see Section 3.4.3, defining non-precipitation events, SNEDS). For all LPEs selected, the reference accumulation was greater than or equal to 0.1 mm and smaller than 0.25 mm.

4.2.5.1.1 *Event-based analysis*

The Site Light Event Datasets (SLEDS) were used for the derivation of the LPE assessment data sets (for further details of the SLEDS, see Section 3.4.3). The objective was to identify the detectability of LPEs using different gauges and wind shields. The reference configuration at all sites considered in this assessment is the R2 configuration, comprising an automated gauge within a DFIR-fence. Where and when available, observations from present weather sensors, which typically have higher sensitivity than automated precipitation gauges, were used for independent identification and verification of LPEs.

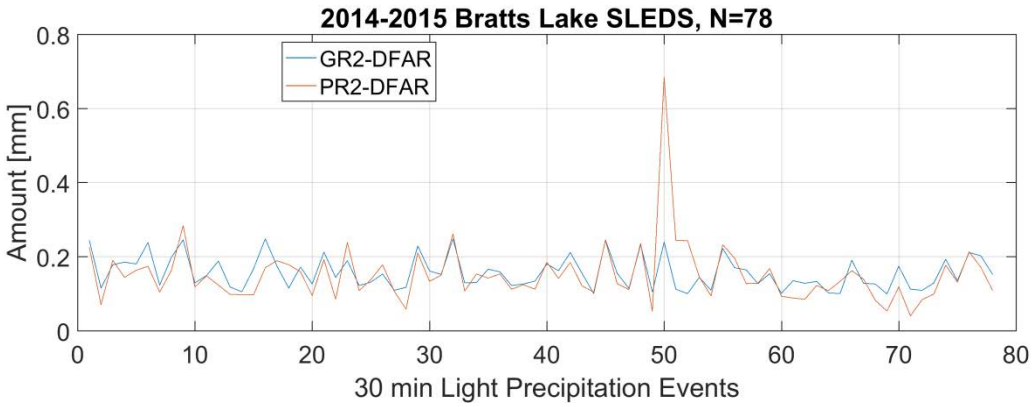
The light precipitation may not happen throughout the selected time interval (event duration). Based on the present weather sensor output, three cases were considered, during which the sensor signaled the occurrence of precipitation:

- at least 15 minutes within the interval
- at least 10 minutes within the interval
- a minimum of 1 minute within the interval (least strict condition)

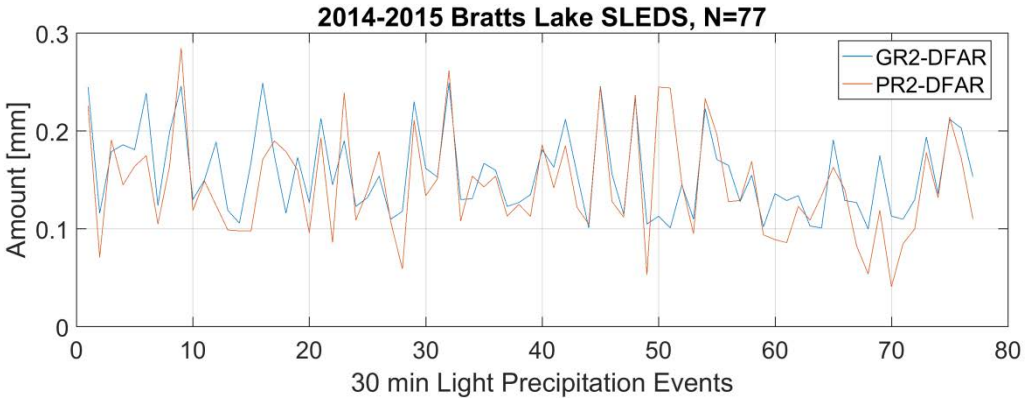
To summarize, all LPE events were selected when the R2 reference measured 0.1 mm to 0.25 mm precipitation, and the disdrometer detected 15, 10 or 1 minutes of precipitation events within the event duration of 30, 60 or 360 minutes.

After applying data quality control procedures (both automatic and manual; see Section 3.3.2), some outlying events remained in the dataset. As an example, a blowing-snow event was identified based on the scatter plot comparison of two gauges at Bratt's Lake on February 14, 2015. (See Figure 4.159.). The Pluvio^2 recorded much more snow due to the blowing airflow around the gauge compared to the Geonor. The occurrence of blowing snow was confirmed using nearby airport observations and the event was removed from the dataset.

a)



b)



c)

d)

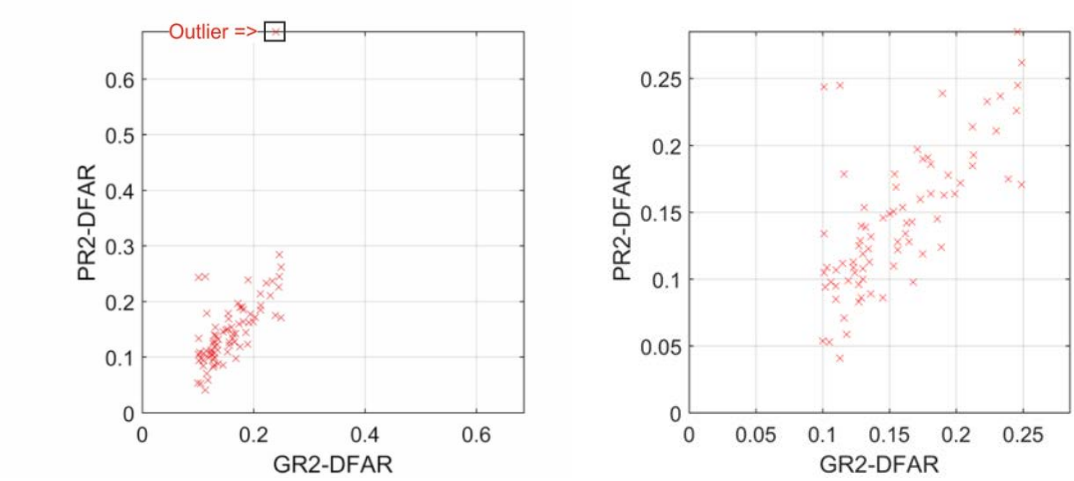


Figure 4.159. Outlier identification for Geonor (GR2-DFAR) and Pluvio^2 (PR2-DFAR) gauges in DFAR configurations at Bratt's Lake. Event accumulations are plotted a) chronologically and c) as scatter plots. Accumulations after outlier removal are plotted b) chronologically and d) as scatter plots.

The LPE data sets were separated based on mean temperature conditions into rain ($T > 2$ °C), mixed precipitation (-2 °C $\leq T \leq 2$ °C), and snow ($T < -2$ °C) events.

4.2.5.1.2 *Test statistics*

Two different types of statistics were used to assess the precipitation observations: detection and comparative statistics.

(a) Detection statistics are used to assess if the tested gauge observed any precipitation when the reference gauge observed light precipitation in the 0.1 mm - 0.25 mm range within a given time period:

i. Probability of detection: POD = YY / (YY+NY)

Where YY indicates the number of periods during which the reference observed light precipitation and the test gauge reported precipitation, and NN indicates the number of periods during which the reference did not observe light precipitation and the test gauge reported precipitation.

(b) Comparative statistics were used to compare the precipitation amounts reported by the reference and test gauges over a given time period, and include the following:

i. Systematic difference (*d*): the mean of the sum of differences in the measurements by the reference and test gauges (the smaller the value the better the match):

$$d = \frac{1}{N}\sum_{i=1}^{N}(x_{ai} - x_{bi})$$

Where:

- x_{ai} = i^{th} measurement reported by the reference system
- x_{bi} = i^{th} measurement reported by the system being tested
- N = number of data samples used for the evaluation

ii. Operational comparability (*C*) is the root mean square of the difference between simultaneous readings from two systems (reference and test gauges) measuring the same quantity in the same environment.

$$C = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{ai} - x_{bi})^2}$$

iii. Correlation coefficient (r_{xy}; CORR): a measure that determines the degree to which time-series for two systems (*x* and *y*) are correlated.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

4.2.5.1.3 *Application of transfer functions*

The applicability of transfer functions to the LPEs was also tested. The catch-efficiency computation was conducted for the SLEDS using the following exponential general equation (Kochendorfer et al., 2017):

$$CE\ (T,U) = e^{(-a \times U)} \times \left(1 - (\tan^{-1}(b \times T) + c)\right)$$

Where:

- *U* is the wind speed at gauge height;
- *T* is air temperature;
- *a*, *b* and *c* are coefficients fit to the data.

The transfer function coefficients were determined using the SEDS precipitation events, with accumulation precipitation ≥ 0.25 mm reported by reference configuration, from multiple sites ("universal" transfer functions). In the present analysis, the same equation was applied to the SLEDS events. For comparison purposes, each catch-efficiency graph contains the (tested gauge/reference gauge) ratio, the fitted line to the points using the general equation above and the coefficients for the corresponding "universal" transfer functions obtained from SEDS events analysis and provided in Table 4.26.

Table 4.26. Universal transfer function coefficients for single-Alter shielded and unshielded gauge configurations, for snow and mixed precipitation at gauge height wind speed (Kochendorfer et al., 2017).

SHIELD type	a	b	c
Single Alter - snow	0.728	0.230	0.336
Unshielded - snow	0.860	0.371	0.229
Single Alter - mixed	0.668	0.132	0.339
Unshielded - mixed	0.641	0.236	0.356

4.2.5.2 Site descriptions

Three sites were selected for case studies, representing three different climate regions. All of the selected weighing gauges were R2 or R3 installations, and all were heated. Available present weather sensors (disdrometers) were also included in the study. For the location of the individual instruments within the test sites, please see site layouts in Annex 4. Two different R2 reference gauges were installed at the test sites considered in this assessment; namely, Geonor T-200B3 gauges at CARE and Bratt's Lake, and OTT Pluvio^2 gauges at Bratt's Lake and Sodankylä (note that the R2 configuration with the Geonor T-200B3 gauge served as the primary reference for Bratt's Lake).

The CARE (CR) site in Ontario, Canada, represents a continental climate with lake-effect snow squalls. The average annual total precipitation is 430.2 mm and the average wind speed is 3.7 m/s (see Table 2.2). The list of gauges included in this study is provided in Table 4.27.

Table 4.27. CARE (Canada) instruments included in SLEDS analysis.

Gauge	Type	Shield	Class	Heated	Name
Geonor	T-200B 600mm	DFAR	R2	Heated	GR2-DFAR
Geonor	T-200B 600mm	SA	R3	Heated	GR3-SA
Geonor	T-200B 600mm	UN	R3	Heated	GR3-UN
OTT Pluvio^2	200 cm^2, 1500mm	SA	R3	Heated	PR3-SA
Thies LPM	Disdrometer	UN	-	-	THIES-UN
Vaisala	HMP155	-	-	-	Temp
Vaisala	NWS425	-	-	-	Wind

The Bratt's Lake (BL) site in Saskatchewan, Canada, represents a continental environment with flat terrain (prairie) and is characterized by blowing snow, dry snow, and windy conditions. The average annual total precipitation is 205.5 mm and the average wind speed is 5.3 m/s (see Table 2.2). The list of gauges included in this study is summarized in Table 4.28.

Table 4.28. Bratts Lake (Canada) instruments included in SLEDS analysis.

Gauge	Type	Shield	Class	Heated	Name
Geonor	T-200B 600mm	DFAR	R2	Heated	GR2-DFAR
Geonor	T-200B 600mm	SA	R3	Heated	GR3-SA
OTT Pluvio^2	200 cm^2, 1500mm	DFAR	R2	Heated	PR2-DFAR
Vaisala	HMP45C	-	-	-	Temp
Younge	ACM	-	-	-	Wind

The Sodankylä site in Finland is a sheltered, northern boreal site characterized by light precipitation events and low wind speeds. The average annual total precipitation is 527 mm, and the average wind speed is as low as 2.7 m/s (see Table 2.2). The list of gauges included in this study at Sodankylä is summarized in Table 4.29.

Table 4.29. Sodankylä (Finland) instruments included in SLEDS analysis.

Gauge	Type	Shield	Class	Heated	Name
OTT Pluvio^2	200 cm^2, 1500mm	DFAR	R2	Heated	PR2-DFAR
OTT Pluvio^2	200 cm^2, 1500mm	SA	R3	Heated	PR3-SA
OTT Pluvio^2	200 cm^2, 1500mm	UN	R3	Heated	PR3-UN
OTT Parsivel 2	Disdrometer	UN	-	Heated	PARS-UN
Vaisala	HMP155	-	-	-	Temp
Thies 2D	2062	-	-	-	Wind

4.2.5.3 Results

4.2.5.3.1 *30/60/360 minute LPE comparison*

As mentioned in Section 4.2.5.1, three different light precipitation event durations were considered in the analysis. The test gauges were compared to the DFAR reference gauge for all LPEs over the 2014/15 winter period with a minimum of 15 precipitating minutes from the disdrometer signal (see Table 4.30 to Table 4.32). The results are color-coded; the best results are indicated by red text, while the worst statistics are indicated by blue text. The number of qualified light precipitation events (identified by "Events" in the table) decreases for longer LPE durations, since some of the light events were qualified as "normal events" (accumulation ≥ 0.25 mm) over longer intervals.

The probability of detection is consistently high for all periods and gauge configurations, with the exception of the 360-minute Pluvio^2 (PR3-SA) gauge at CARE. The POD only tells us whether the tested instrument identified the light precipitation event or not. Further details of how the accumulation reports from the test and reference gauges compare can be obtained from the systematic difference (d), operational comparability (C) and correlation values (CORR).

The systematic difference values are always less than than 0.1 mm, with the smallest values for all three time intervals observed for DFIR-shielded gauges in snow, and the next smallest values for single-Alter-shielded Geonor and Pluvio^2 gauges in mixed precipitation. The operational comparability values show similar trends, since the two parameters are related. The highest correlation values were observed for the sheltered Sodankylä site and 30-minute intervals. The 360-minute correlation values could not be evaluated due to the limited number of LPEs with this duration. The correlation coefficients for single-Alter-shielded gauges were always higher than for unshielded gauges, indicating improved agreement with accumulation values reported by the reference configuration for the shielded gauges relative to unshielded gauges.

Overall, the results were similar for the three durations investigated. The 30-minute SLEDS were selected for further analysis.

Table 4.30. 30-min LPE statistics. Colors for each parameter are defined as follows: POD = 100 red and < 50 blue; d < 0.02 red, > 0.2 blue; C < 0.05 red and > 0.2 blue; CORR > 0.8 red and < 0.3 blue.

DUR	SITE	TYPE	REF	SUT	Events	YY	NY	POD	d [mm]	C [mm]	CORR
30	BL	SNOW	GR2-DFIR	GR3-SA	77	77	0	100	0.09	0.10	0.29
30	BL	MIXED	GR2-DFIR	GR3-SA	23	23	0	100	0.08	0.08	0.65
30	BL	SNOW	GR2-DFIR	PR2-DFIR	78	78	0	100	0.01	0.06	0.72
30	BL	MIXED	GR2-DFIR	PR2-DFIR	23	22	1	95.7	0.03	0.06	0.59
30	CR	SNOW	GR2-DFIR	GR3-SA	102	102	0	100	0.05	0.06	0.65
30	CR	MIXED	GR2-DFIR	GR3-SA	44	44	0	100	0.02	0.03	0.85
30	CR	SNOW	GR2-DFIR	GR3-UN	102	102	0	100	0.09	0.10	0.42
30	CR	MIXED	GR2-DFIR	GR3-UN	44	44	0	100	0.03	0.05	0.74
30	CR	SNOW	GR2-DFIR	PR3-SA	102	97	5	95.1	0.06	0.08	0.53
30	CR	MIXED	GR2-DFIR	PR3-SA	44	44	0	100	0.02	0.04	0.71
30	CR	SNOW	GR2-DFIR	THIES-UN	102	97	5	95.1	-0.05	0.24	0.31
30	CR	MIXED	GR2-DFIR	THIES-UN	44	39	5	88.6	-0.03	0.10	0.39
30	SD	SNOW	PR2-DFIR	PR3-SA	193	193	0	100	0.01	0.02	0.94
30	SD	MIXED	PR2-DFIR	PR3-SA	198	198	0	100	0.02	0.02	0.91
30	SD	SNOW	PR2-DFIR	PR3-UN	193	193	0	100	0.03	0.04	0.83
30	SD	MIXED	PR2-DFIR	PR3-UN	198	198	0	100	0.04	0.05	0.82
30	SD	SNOW	PR2-DFIR	PARS-UN	193	193	0	100	-0.37	0.44	0.50
30	SD	MIXED	PR2-DFIR	PARS-UN	198	198	0	100	-0.33	0.39	0.57

Table 4.31. 60-minute LPE statistics, with colours for each parameter defined using the same rules as in Table 4.30.

DUR	SITE	TYPE	REF	SUT	Events	YY	NY	POD	d [mm]	C [mm]	CORR
60	BL	SNOW	GR2-DFIR	GR3-SA	54	54	0	100	0.09	0.10	0.48
60	BL	MIXED	GR2-DFIR	GR3-SA	16	16	0	100	0.09	0.10	0.45
60	BL	SNOW	GR2-DFIR	PR2-DFIR	54	54	0	100	0.01	0.06	0.56
60	BL	MIXED	GR2-DFIR	PR2-DFIR	16	15	1	93.8	0.04	0.06	0.47
60	CR	SNOW	GR2-DFIR	GR3-SA	81	81	0	100	0.05	0.06	0.59
60	CR	MIXED	GR2-DFIR	GR3-SA	26	26	0	100	0.03	0.04	0.88
60	CR	SNOW	GR2-DFIR	GR3-UN	81	79	2	97.5	0.09	0.10	0.37
60	CR	MIXED	GR2-DFIR	GR3-UN	26	26	0	100	0.05	0.06	0.6
60	CR	SNOW	GR2-DFIR	PR3-SA	81	72	9	88.9	0.08	0.09	0.45
60	CR	MIXED	GR2-DFIR	PR3-SA	26	26	0	100	0.03	0.05	0.82
60	CR	SNOW	GR2-DFIR	THIES-UN	81	75	6	92.6	-0.06	0.22	0.02
60	CR	MIXED	GR2-DFIR	THIES-UN	26	24	2	92.3	-0.04	0.11	0.6
60	SD	SNOW	PR2-DFIR	PR3-SA	80	80	0	100	0.01	0.02	0.93
60	SD	MIXED	PR2-DFIR	PR3-SA	125	125	0	100	0.02	0.03	0.88
60	SD	SNOW	PR2-DFIR	PR3-UN	80	80	0	100	0.04	0.04	0.79
60	SD	MIXED	PR2-DFIR	PR3-UN	125	125	0	100	0.05	0.06	0.72
60	SD	SNOW	PR2-DFIR	PARS-UN	80	80	0	100	-0.33	0.41	0.47
60	SD	MIXED	PR2-DFIR	PARS-UN	125	125	0	100	-0.31	0.38	0.49

Table 4.32. 360-minute LPE statistics, with colours for each parameter defined using the same rules as in Table 4.30.

DUR	SITE	TYPE	REF	SUT	Events	YY	NY	POD	d [mm]	C [mm]	CORR
360	BL	SNOW	GR2-DFIR	GR3-SA	20	17	3	85.0	0.08	0.10	0.23
360	BL	MIXED	GR2-DFIR	GR3-SA	2	1	1	50.0	0.16	0.17	-
360	BL	SNOW	GR2-DFIR	PR2-DFIR	20	18	2	90.0	-0.01	0.12	0.19
360	BL	MIXED	GR2-DFIR	PR2-DFIR	2	2	0	100	-0.07	0.13	-
360	CR	SNOW	GR2-DFIR	GR3-SA	24	24	0	100	0.06	0.07	0.67
360	CR	MIXED	GR2-DFIR	GR3-SA	8	8	0	100	0.06	0.07	0.81
360	CR	SNOW	GR2-DFIR	GR3-UN	24	20	4	83.3	0.09	0.11	0.06
360	CR	MIXED	GR2-DFIR	GR3-UN	8	7	1	87.5	0.10	0.10	0.79
360	CR	SNOW	GR2-DFIR	PR3-SA	24	12	12	50	0.11	0.13	0.27
360	CR	MIXED	GR2-DFIR	PR3-SA	8	5	3	62.5	0.10	0.13	0.25
360	CR	SNOW	GR2-DFIR	THIES-UN	24	22	2	91.7	-0.11	0.30	0.42
360	CR	MIXED	GR2-DFIR	THIES-UN	8	7	1	87.5	-0.13	0.21	0.02
360	SD	SNOW	PR2-DFIR	PR3-SA	5	5	0	100	0.02	0.04	-
360	SD	MIXED	PR2-DFIR	PR3-SA	15	15	0	100	0.02	0.03	0.65
360	SD	SNOW	PR2-DFIR	PR3-UN	5	5	0	100	0.06	0.07	-
360	SD	MIXED	PR2-DFIR	PR3-UN	15	15	0	100	0.06	0.07	0.41
360	SD	SNOW	PR2-DFIR	PARS-UN	5	5	0	100	-0.27	0.43	-
360	SD	MIXED	PR2-DFIR	PARS-UN	15	15	0	100	-0.39	0.58	0.06

All test statistics were computed for the two winter seasons and for additional instruments, shields, and configurations. As well, statistics were computed using both 10-minute and 1-minute (less strict) requirements for precipitation detector signal within LPEs. These additional results were computed for reference, and are not presented here.

4.2.5.3.2 *Case studies*

Light Precipitation Event datasets for selected gauges, configurations, and test sites were considered in four separate case studies with the following objectives:

a) To evaluate the effect of local climate;
b) To evaluate shield performance for a given gauge type;
c) To compare the performance of different test gauges in the same configuration, at the same site;
d) To evaluate the performance of disdrometers.

Additionally, the applicability of the universal transfer functions (TF) obtained from SEDS analysis (Kochendorfer et al., 2017) to the LPE data was tested. For each gauge/shield combination, the universal coefficients in Table 4.26 were fit to the data and compared with the optimal line fitted directly to the SLEDS events catch-efficiency data (same functional form with unique coefficients).

To address the objectives, eight gauge-shield combinations were used at the three test sites. Case study analysis results are presented for 30-minute light-precipitation snow events during the 2014/15 winter season in Figure 4.160. to Figure 4.167. Each figure includes four subplots:

a) The light precipitation event accumulations for the test gauge and reference configurations, in order of occurrence;
b) The difference between the accumulation reported by the reference and test gauge configurations for each LPE, in order of occurrence;
c) A scatter-plot comparison of reference and test gauge accumulations for each event;
d) A scatter plot of catch efficiency (test gauge accumulation/reference gauge accumulation) as a function of mean wind speed at gauge height for each LPE, colour-coded by air temperature. A curve is fit to the data and compared with the curve computed using the appropriate universal transfer function coefficients.

Relevant results in support of each of the above objectives are detailed in the subsections below. All comparative statistics used in the discussion are taken from Table 4.30.

4.2.5.3.2.1 Evaluating the effect of local climate

To evaluate the effect of local climate, identical gauge/shield configurations were compared at different test sites. First, the single-Alter shielded Geonor gauge (GR3-SA) configurations installed at CARE and Bratt's Lake were compared. The results are plotted in Figure 4.160. and Figure 4.161. , respectively. The single-Alter shielded gauge at CARE showed higher correlation values, lower operational comparability, and smaller systematic differences relative to the same configuration at Bratt's Lake (CORR = 0.65, C = 0.06, and d = 0.05 at CARE; CORR = 0.29, C = 0.1, and d = 0.09 at Bratt's Lake). These differences are attributed to the characteristically higher wind speeds at Bratt's Lake.

Next, the single-Alter shielded Pluvio^2 gauge (PR3-SA) configuration installed at Sodankylä (northern boreal climate) was compared with that installed at CARE (Figure 4.162 and Figure 4.163). In terms of the agreement between test gauge and reference gauge accumulation reports, the configuration at Sodankylä showed higher correlation coefficients, lower operational comparability, and smaller systematic differences relative to that at CARE (CORR = 0.94, C = 0.02, and d = 0.01 at Sodankylä; CORR = 0.53, C = 0.08, and d = 0.06 at CARE). These differences likely result from the characteristically lower wind speeds at Sodankylä relative to those at CARE.

Additional analysis was conducted for LPEs in the mixed precipitation regime (-2 °C ≤ T ≤ 2 °C) at all sites. Based on the results in Table 4.30, both the correlation and operational comparability between reference configurations and single-Alter-shielded Geonor configurations at Bratt's Lake and CARE were found to improve (increase and decrease, respectively) for mixed-phase LPEs relative to snow events. Similar improvements were observed in mixed-phase precipitation relative to snow for the single-Alter-shielded Pluvio^2 gauge at CARE, while the results for the configuration at Sodankylä were similar to those for snow. The improvements in parameters observed for gauges at CARE and Bratt's Lake can likely be attributed to improvements in catch efficiency resulting from the larger overall density for mixed precipitation relative to snow, which makes hydrometeors less prone to deflection away from the gauge orifice at a given wind speed. The results for Sodankylä are impacted less by phase, likely due to the very low wind speeds, for which the effects of undercatch are reduced.

Based on visual comparison of Figure 4.160. Figure 4.162. and Figure 4.163, the universal lines fit to data using the transfer function coefficients in Table 4.26 compare well with the lines fit directly to the light precipitation event data. The universal lines appear to overestimate catch efficiency at higher wind speeds at CARE and Bratt's Lake, and to underestimate catch efficiency at higher wind speeds experienced at Sodankylä.

4.2.5.3.2.2 Evaluating shield performance

To evaluate the performance of the shield around a given gauge, the single-Alter (SA) shielded and unshielded (UN) gauges were compared at the same site. For the Geonor gauges at CARE, the SA configuration showed higher correlation values, lower operational comparability values, and smaller systematic differences relative to the unshielded configuration (Figure 4.160 and Figure 4.164; CORR = 0.65, C = 0.06, and d = 0.05 for SA; CORR = 0.42, C = 0.1, and d = 0.09 for UN). Similar trends were observed for the Pluvio^2 gauges at Sodankylä (Figure 4.163 and Figure 4.165; CORR = 0.94, C = 0.02, and d = 0.01 for SA; CORR = 0.83, C = 0.04, and d = 0.03 for UN). At both sites, the operational comparability values and systematic differences were lower for mixed precipitation events relative to snow events, which can be attributed to differences in hydrometeor density (see Section 4.2.5.3.2.1).

The universal transfer functions underestimate catch efficiency at higher wind speeds for the gauges at CARE, and to greater extent for the unshielded Geonor gauge relative to the shielded gauge. The universal transfer functions underestimate catch efficiency at higher wind speed for both gauge configurations at Sodankylä, and to a similar extent for both configurations. These site differences are attributed to differences in characteristic wind conditions.

4.2.5.3.2.3 Comparing performance for different test gauges

To compare the performance of the different test gauge types, the performance of Geonor and Pluvio^2 gauges at the same site was evaluated. Of the three test sites considered, only CARE had single-Alter shielded installations of both Geonor and Pluvio^2 gauges (Figure 4.160 and Figure 4.162). Based on this very limited sample, the correlation coefficient for the Geonor was higher than for the Pluvio^2 (CORR = 0.65 for Geonor; CORR = 0.53 for Pluvio^2). The operational comparability and systematic difference values were comparable for both gauges, but slightly lower for the Geonor (C = 0.06 and d = 0.05 for Geonor; C = 0.08 and d = 0.06 for Pluvio^2). The probability of detection was also higher for the Geonor gauge relative to the Pluvio^2 (POD = 100 for Geonor; POD = 94.9 for Pluvio^2). It is possible that this comparison is influenced by the fact that the reference gauge type at CARE was a Geonor, and hence that the comparison could be improved for gauges with the same operating principle, data processing, and housing geometry.

For both gauge types, the detectability improved for mixed precipitation relative to snow – the correlation coefficient increased, and operational comparability and systematic differences decreased for mixed precipitation (Table 4.30). Comparing the catch efficiencies of the SA-shielded Geonor and Pluvio^2 gauges at CARE, lower variability of light precipitation catch efficiency values and improved agreement between the Universal and Fitted transfer functions are observed for the SA-shielded Geonor gauge (see Figure 4.160. and Figure 4.162.). This improved agreement likely results from the test and reference gauges being of the same type, as discussed above.

4.2.5.3.2.4 Evaluating the performance of disdrometers

The performance of disdrometers in terms of reporting precipitation accumulation (relative to a reference configuration) during light precipitation events was evaluated using test gauges at CARE (Thies LPM) and at Sodankylä (OTT Parsivel^2). The precipitation amount is often overestimated relative to the reference configuration using disdrometers (Figure 4.166 and Figure 4.167). The event detection capability of disdrometers could not be assessed, since the disdrometer signal was used within the event selection procedure (the samples are not independent). The disdrometer results for mixed precipitation were very similar to the snow events (see Table 4.30). No transfer functions were available for the disdrometers tested. The catch efficiency results in Figure 4.166 and Figure 4.167

show very different trends with respect to mean speed than those for the weighing gauges in Figure 4.160 to Figure 4.165.

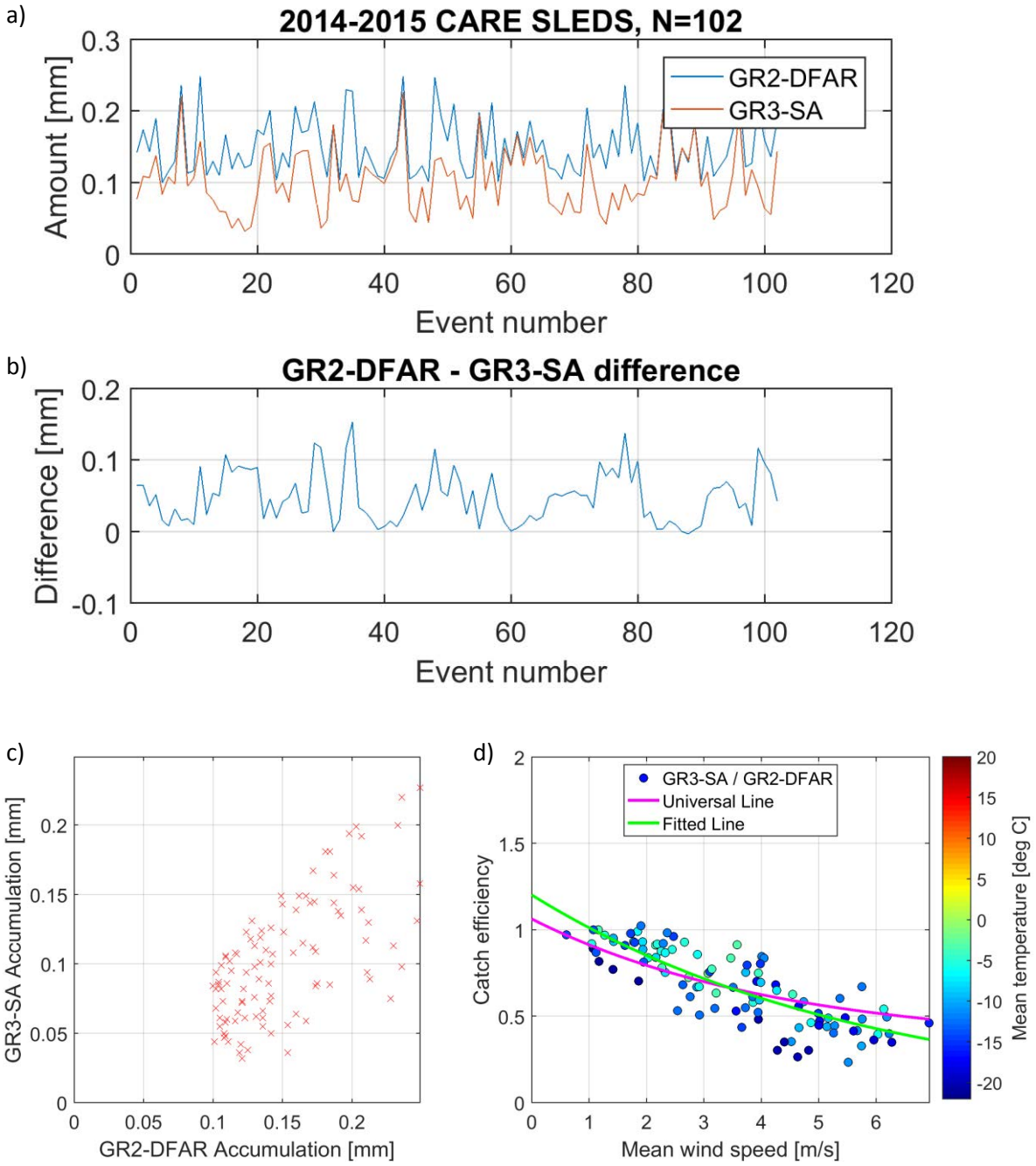


Figure 4.160. Single-Alter shielded Geonor (GR3-SA) gauge analysis results at CARE for 30-minute light-precipitation snow events: a) GR3-SA gauge and reference event amounts, plotted in order of occurrence; b) reference minus GR3-SA event accumulations, in order of occurrence; c) scatter-plot comparison events; d) catch efficiency as a function of mean gauge-height wind speed all events, colour-coded by temperature, with universal and fitted transfer function lines.

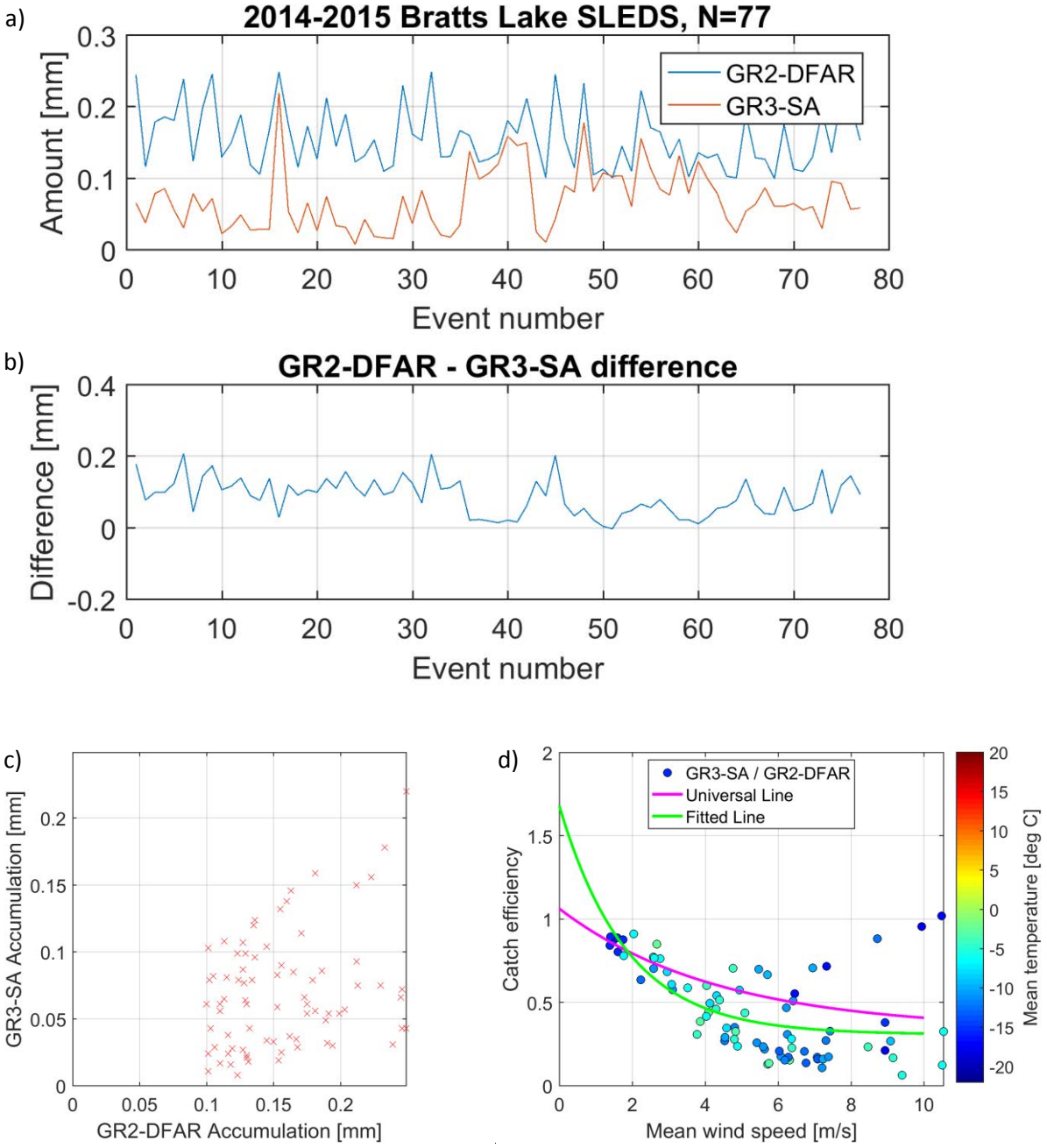


Figure 4.161. Single-Alter shielded Geonor (GR3-SA) gauge analysis results at Bratt's Lake for 30-minute light-precipitation snow events: a) GR3-SA gauge and reference event amounts, plotted in order of occurrence; b) reference minus GR3-SA event accumulations, in order of occurrence; c) scatter-plot comparison events; d) catch efficiency as a function of mean gauge-height wind speed all events, colour-coded by temperature, with universal and fitted transfer function lines.

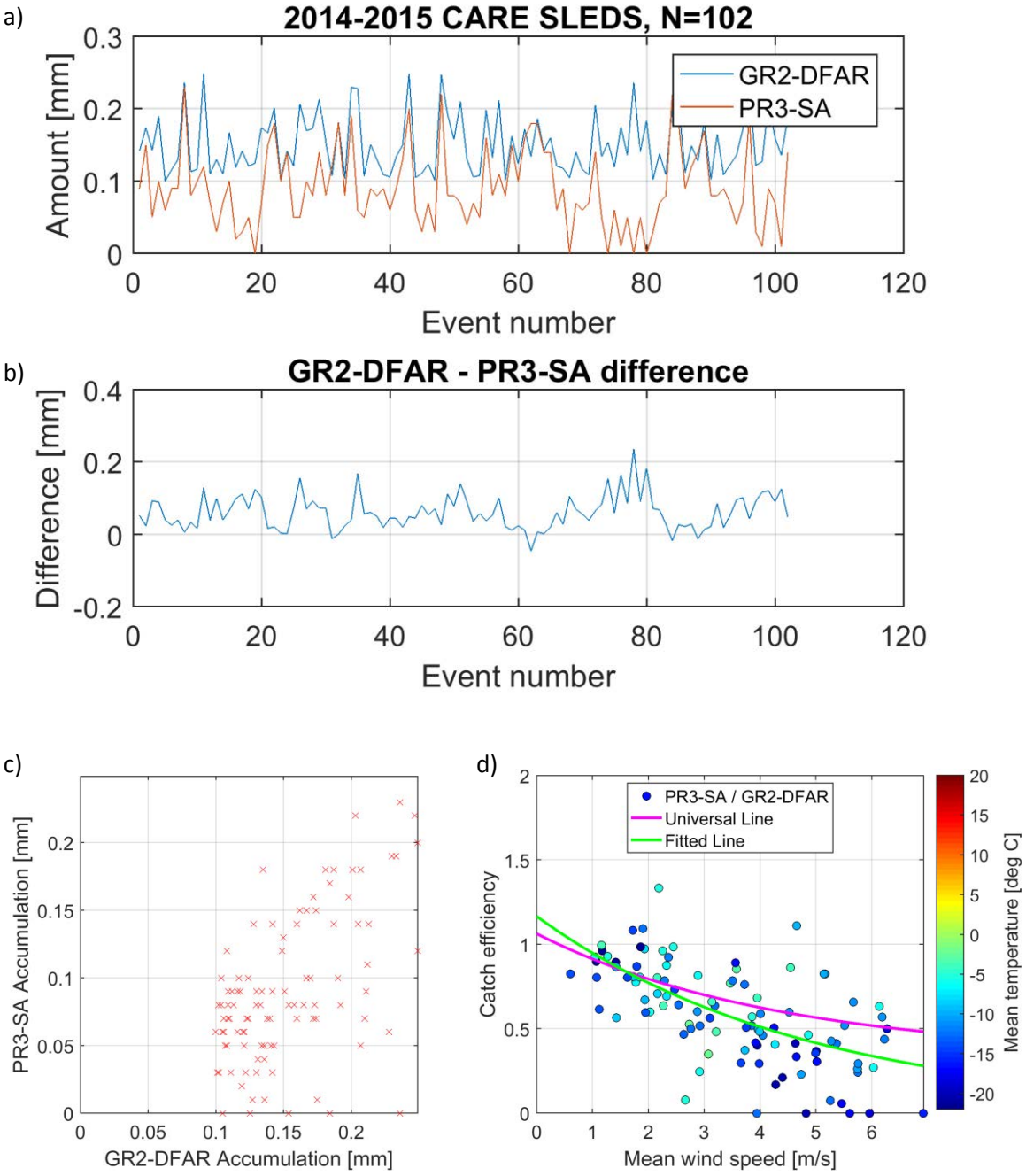


Figure 4.162. Single-Alter shielded Pluvio^2 (PR3-SA) gauge analysis results at CARE for 30-minute light-precipitation snow events: a) PR3-SA gauge and reference event amounts, plotted in order of occurrence; b) reference minus PR3-SA event accumulations, in order of occurrence; c) scatter-plot comparison events; d) catch efficiency as a function of mean gauge-height wind speed all events, colour-coded by temperature, with universal and fitted transfer function lines.

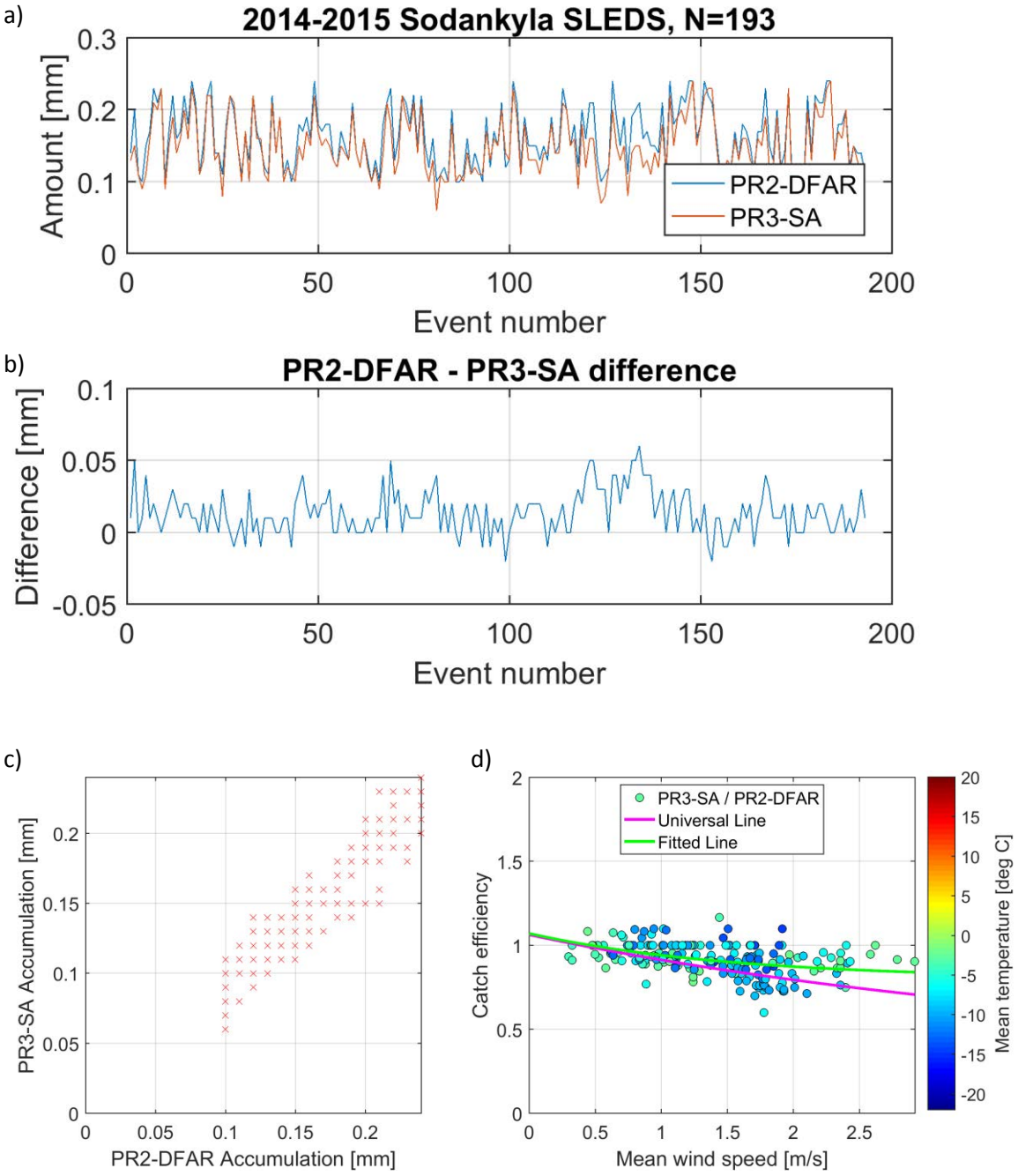


Figure 4.163. Single-Alter shielded Pluvio^2 (PR3-SA) gauge analysis results at Sodankylä for 30-minute light-precipitation snow events: a) PR3-SA gauge and reference event amounts, plotted in order of occurrence; b) reference minus PR3-SA event accumulations, in order of occurrence; c) scatter-plot comparison events; d) catch efficiency as a function of mean gauge-height wind speed all events, colour-coded by temperature, with universal and fitted transfer function lines.

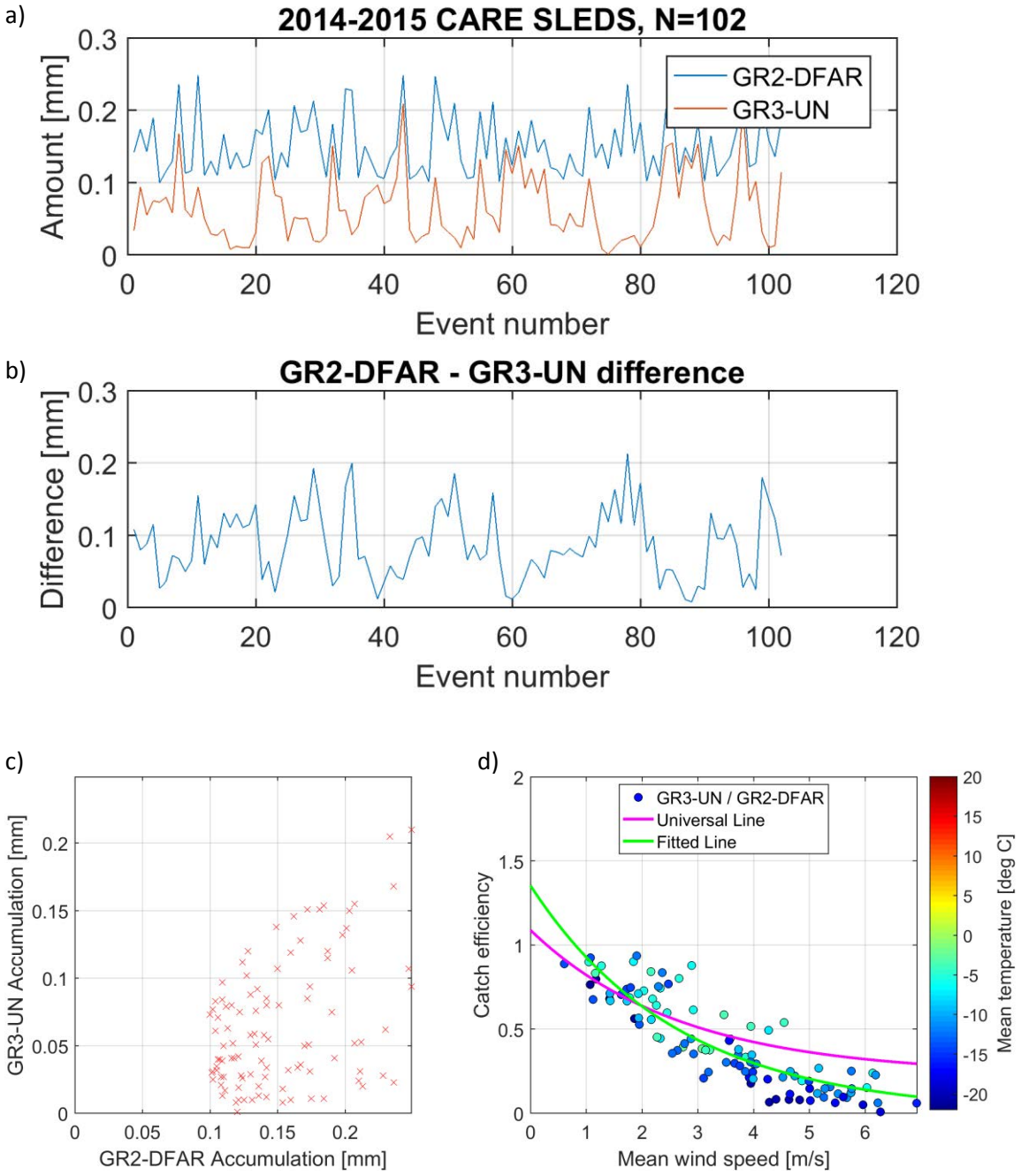


Figure 4.164. Unshielded Geonor (GR3-UN) gauge analysis results at CARE for 30-minute light-precipitation snow events: a) GR3-UN gauge and reference event amounts, plotted in order of occurrence; b) reference minus GR3-UN event accumulations, in order of occurrence; c) scatter-plot comparison events; d) catch efficiency as a function of mean gauge-height wind speed all events, colour-coded by temperature, with universal and fitted transfer function lines.

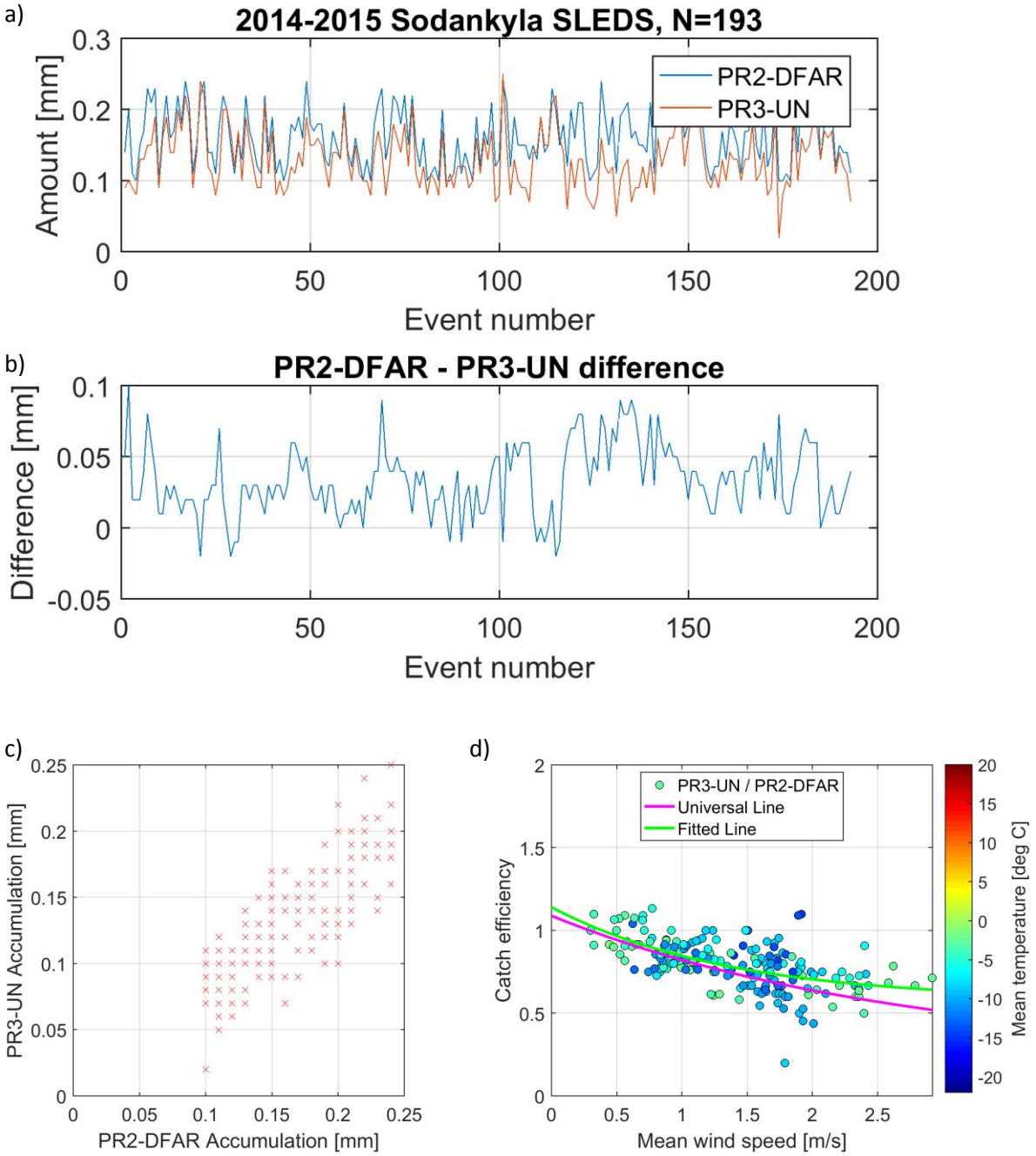


Figure 4.165. Unshielded Pluvio^2 (PR3-UN) gauge analysis results at Sodankylä for 30-minute light-precipitation snow events. a) PR3-UN gauge and reference event amounts, plotted in order of occurrence; b) reference minus PR3-UN event accumulations, in order of occurrence; c) scatter-plot comparison events; d) catch efficiency as a function of mean gauge-height wind speed all events, colour-coded by temperature, with universal and fitted transfer function lines.

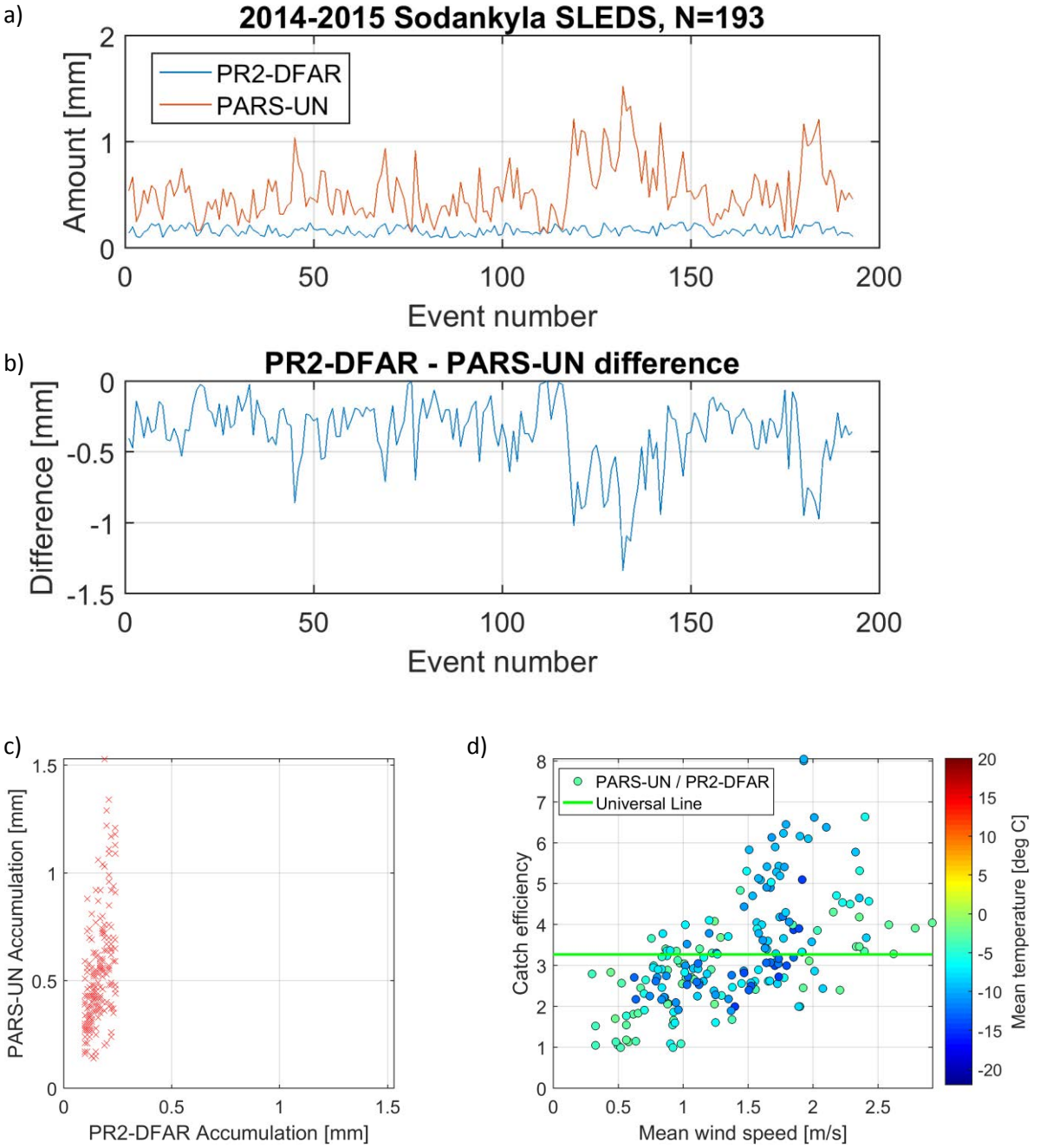


Figure 4.166. Unshielded Parsivel^2 (PARS-UN) sensor analysis results at Sodankylä for 30-minute light-precipitation snow events: a) PARS-UN and reference event amounts, plotted in order of occurrence; b) reference minus PARS-UN event accumulations, in order of occurrence; c) scatter-plot comparison events; d) catch efficiency as a function of mean gauge-height wind speed all events, colour-coded by temperature, with universal and fitted transfer function lines.

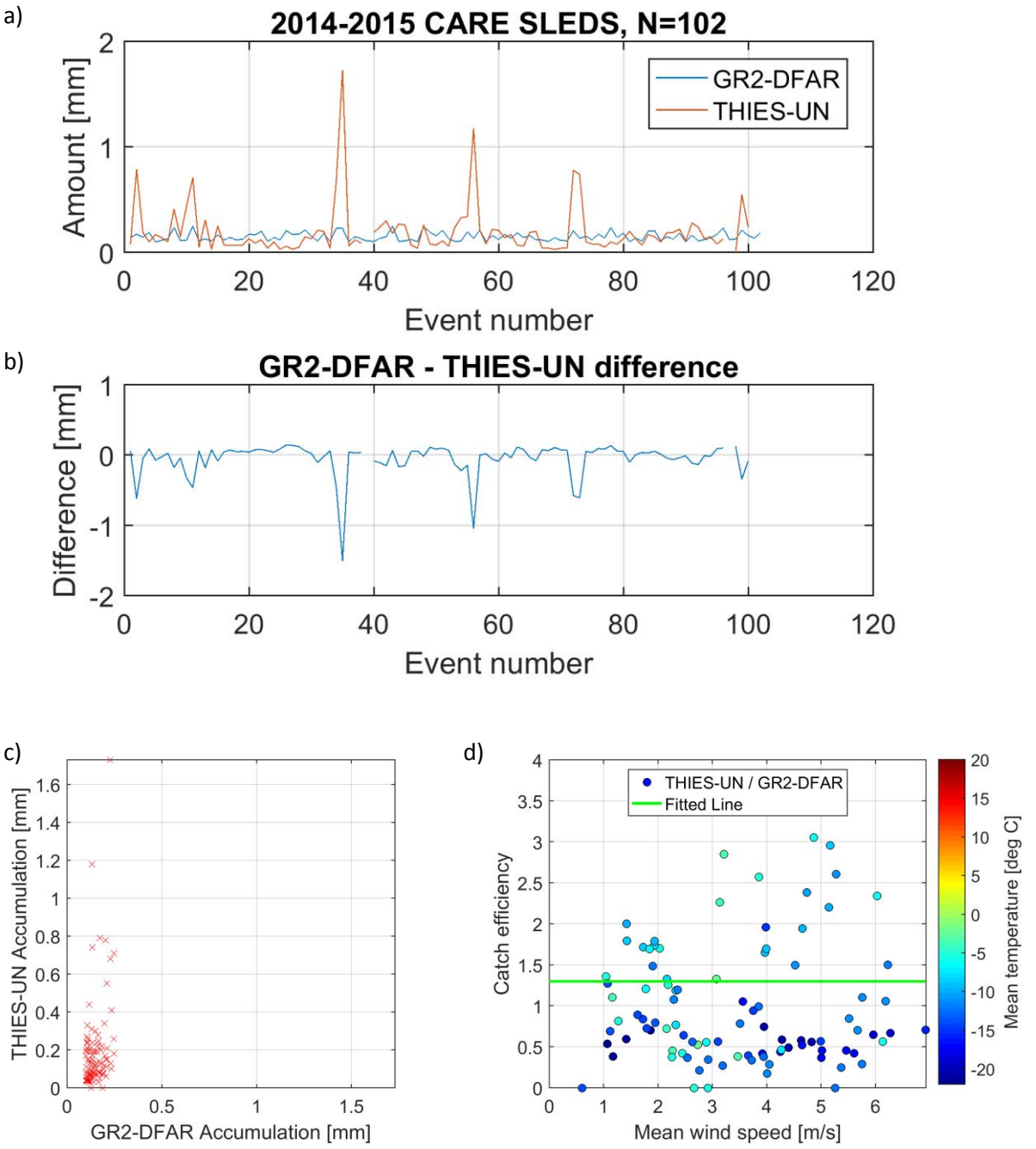


Figure 4.167. Unshielded Thies LPM (THIES-UN) sensor analysis results at CARE for 30-minute light-precipitation snow events: a) THIES-UN and reference event amounts, plotted in order of occurrence; b) reference minus THIES-UN event accumulations, in order of occurrence; c) scatter-plot comparison events; d) catch efficiency as a function of mean gauge-height wind speed all events, colour-coded by temperature, with universal and fitted transfer function lines.

4.2.5.3.3 *Correlation*

Separate correlation analysis was completed between the test gauges/sensors considered in the case studies and the corresponding reference configurations (Geonor or Pluvio^2 in DFAR) for both 30- and 60-minute periods. Figure 4.168. shows large differences among the different test gauges/sensors and sites. The highest correlation values were observed at Sodankylä, between the reference gauge

(Pluvio^2 in DFAR) and the single-Alter shielded and unshielded Pluvio^2 gauges, mainly due to the characteristically low wind-speed conditions at the site. Comparing the 30- and 60-minute intervals, higher correlation values were detected for the SA Geonor at Bratt's Lake, unshielded Pluvio^2 and Parsivel^2 at Sodankylä, and Thies LPM at CARE for the 60-minute periods relative to the 30-minute periods. For the gauges at CARE (and not the Thies LPM disdrometer), higher correlations were found for 30-minute intervals relative to 60-minute intervals.

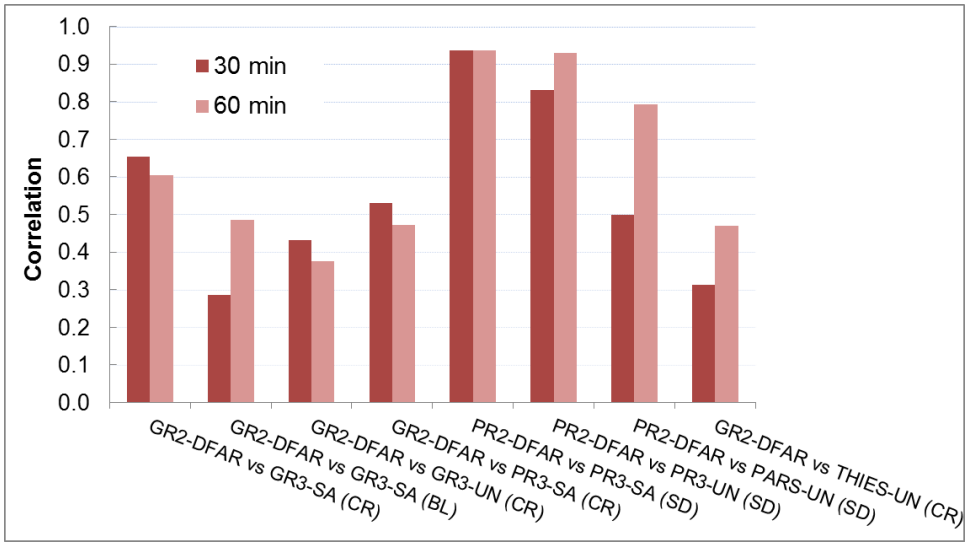


Figure 4.168. Correlation comparison of 30- and 60-min light-precipitation snow events for eight gauge/sensor and wind shield configurations at three locations (CARE; Bratt's Lake; Sodankylä). All events were taken from the 2014/15 season and had a minimum of 15 minutes during which the disdrometer in the reference configuration indicated the occurrence of precipitation.

4.2.5.3.4 *Probability of detection*

Similarly, the probability of detection was also analyzed for each of the gauges/sensors considered in the case studies relative to the corresponding reference configurations and for both 30-minute and 60-minute periods. All test gauges and sensors considered, at all three test sites, show high detection skill for light snow precipitation events, with POD values of approximately 90% or higher (Figure 4.169.). It is important to note here that the "noise zone" was avoided by excluding the DFAR reports < 0.1 mm from the analysis (the tested gauges had no restrictions). Detection in this context means that the tested gauge was only required to measure any precipitation when the reference gauge measured precipitation within the 0.1 - 0.25 mm range. The 60-minute POD values are similar to those for 30-minute assessment intervals.

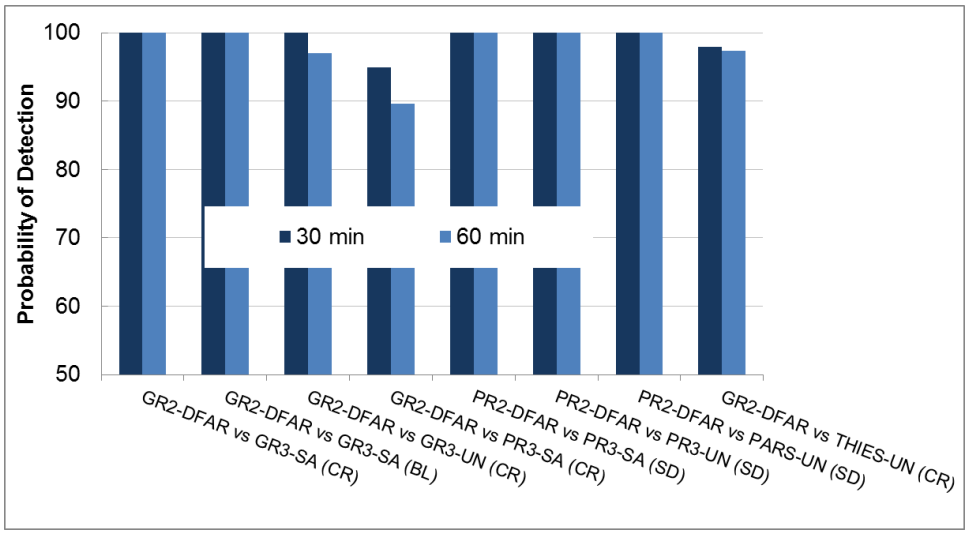


Figure 4.169. Probability of Detection for 30- and 60-minute light-precipitation snow events during which the disdrometer indicates a minimum of 15 minutes of precipitation occurrence over the 2014/15 winter season.

4.2.5.4 Summary and conclusions

The objective of this study was to assess the detectability and quality of light-precipitation observations within the SPICE experiment using various gauge/shield combinations at test sites with different characteristic climate conditions. Further, the applicability of universal transfer functions derived using larger precipitation events (≥ 0.25 mm) to light precipitation events was also evaluated.

In the current report, light-precipitation events were defined as events for which less than 0.25 mm precipitation was observed within a given time period (typically 30 minutes, but 60 and 360 minutes were also examined). For accumulation values < 0.1 mm, it is difficult to distinguish between noise and precipitation in weighing gauge data. For that reason the lower limit for reference gauge accumulation was set to 0.1 mm for light-precipitation events.

The Light-Precipitation Event data set used in the study originated from the quality-controlled SLEDS database. Identification of any remaining data issues (e.g. outiers) was possible by comparison with reference gauge data. Ancillary observations, namely weather reports from nearby observing stations, were used to corroborate the occurrence of suspected blowing-snow events.

The local climate characteristics play an important role in the detection and measurement of light precipitation. This was determined by comparison of identical gauge/shield configurations at different sites. The characteristic wind speeds at different sites were identified as the main factor affecting the agreement between the test and reference configurations. Single-Alter shielded gauges at Sodankylä (lowest wind speeds) showed the best agreement with the reference configuration, as indicated by the comparison statistics considered. Based on the same parameters, the agreement with the reference decreased for the single-Alter shielded gauge at CARE (intermediate wind speeds), and to a greater extent for the same gauge configuration at Bratt's Lake (highest wind speeds). The probability of detection exceeded (85%) for most cases. For mixed precipitation, the comparison statistics improved for most cases relative to snow.

For the shield comparison, different shield configurations (single-Alter and unshielded) were assessed for the same gauge type at the same site. For light precipitation, the single-Alter shielded

gauges showed better agreement with the references relative to the unshielded gauges, as indicated by the comparison statistics. The differences in performance between the single-Alter-shielded and unshielded gauges decreased for mixed precipitation relative to snow.

For the gauge comparison, the performance of Geonor and Pluvio^2 gauges at the same site and in the same shield configuration were evaluated. For this component of the assessment, only CARE had the parallel single-Alter-shielded installations of both test gauge types. The Geonor gauge showed better agreement with the reference configuration relative to the Pluvio^2; however, the reference configuration also employed a Geonor gauge (same principle of operation, data processing, and housing geometry), which may impact the comparison. For both gauge types, comparison statistics improved for mixed precipitation relative to snow.

The disdrometers were found to over-report precipitation relative to the corresponding reference configurations. Their ability to detect precipitation could not be assessed, as their data are used in the procedure for identifying reference gauge events.

In general, the universal transfer function derived using larger precipitation events fit the light-precipitation event data well, with varying degrees of over- or under-estimation depending on the specific wind speed, shield configuration, and site characteristics.

4.2.6 Challenges for the measurement of snow on the ground

Authors: Craig Smith, Anna Kontu, Timo Laine, Antti Poikonen, Leena Leppänen, Henna-Reetta Hannula, Samuel Morin, Lauren Arnold

4.2.6.1 Impact of sensor and mount design on the collection of snow

The design of the sensor and mounting infrastructure can impact the sensor's capability to make a measurement or influence the snow pack it is measuring. This can be especially true in locations that receive heavy snowfall in low-wind situations. At the Sodankylä SPICE site, for example, the instruments and mounting structures are often covered in snow. After a heavy snowfall event during non-windy conditions, several centimeters of fresh snow can sit on top of the instrument, or even encase the instrument on occasion, and last for days until the snow either falls off or is cleared manually. During SPICE, instruments that became packed with snow were left for 24 hours to clear naturally before they were cleared with human intervention. This provided many opportunities to photograph and document occurrences, especially at Sodankylä.

During heavy snowfall at the Sodankylä site, snow can stick to, and accumulate on, any horizontal surface, whether that surface is part of a sensor or the sensor's mounting infrastructure. Figure 4.170 shows this occurring after a heavy snowfall event on January 27, 2014. Of note here is that the heated SR50ATH in Figure 4.170 remains clear of snow; however, the buildup of snow near and below the sensor as a result of the infrastructure could potentially impact the sensor's capability to make a measurement, even if the sensor itself is heated. The USH-8 pictured in Figure 4.170 also employs heating to de-ice the ultrasonic membrane, but does not prevent snow from building up on the sensor housing.

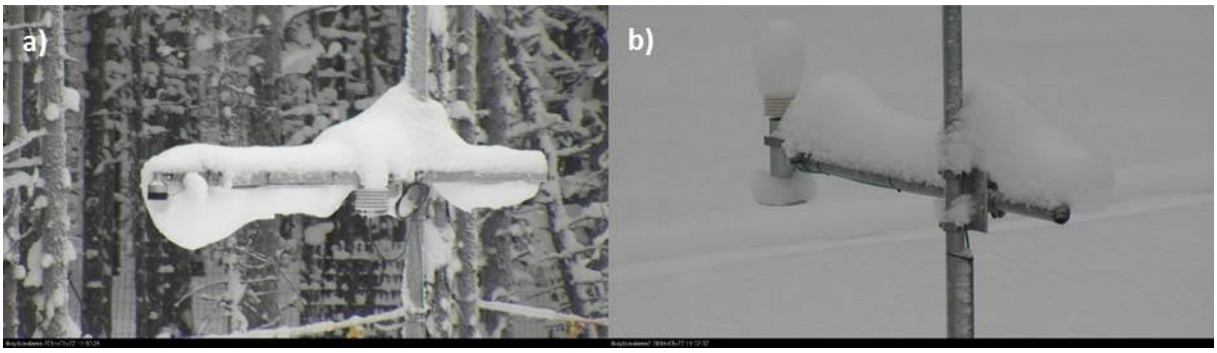


Figure 4.170. Snow sticking to, and accumulating on, a) heated SR50ATH on an unheated horizontal boom, and b) heated USH-8 on an unheated horizontal boom at Sodankylä following a snowfall event on January 27, 2014.

Snow falling from the sensor and the infrastructure, whether naturally or from manual clearing, can fall onto the target area, as illustrated in Figure 4.171. This has the potential to impact the depth measurement either positively, through increased accumulation, or negatively, through compaction. Compaction from falling snow could also potentially change the density of the snow under the sensor and impact melt rates in spring. Analysis was undertaken to try to quantify the impact of snow dropping off the sensor and the mounting arm. The timing of snow drops was derived from the hourly webcam photos and cross-referenced with the instrument data. Occasionally, the timing of the snow drop could be matched with an increase in snow depth under the sensor, but often the snow depth decreased. Because of the noise in the sensor signal and natural variability in the snow depth, it was difficult to ascertain the true impact of snow dropping from the sensor on the measurement of snow depth.

Figure 4.171. The impact of snow falling off horizontal unheated instrument booms at Sodankylä following a heavy snowfall event. The fallen snow in the target area is circled in red.

One solution to preventing snow accumulation on the mounting infrastructure is to install the mounting boom at a 45° angle and apply heating to melt or sublimate snow as it contacts the boom. The heat source, in this case, is a self-regulating heating cable inserted inside the hollow boom. The

angled boom not only inhibits accumulation, but may also prevent meltwater from dripping onto the target area and impacting the measurement and snowpack composition. Figure 4.172 shows the impacts of heating and the angled boom on snow accumulation during the same January 27, 2014 snowfall event as shown in Figure 4.170.

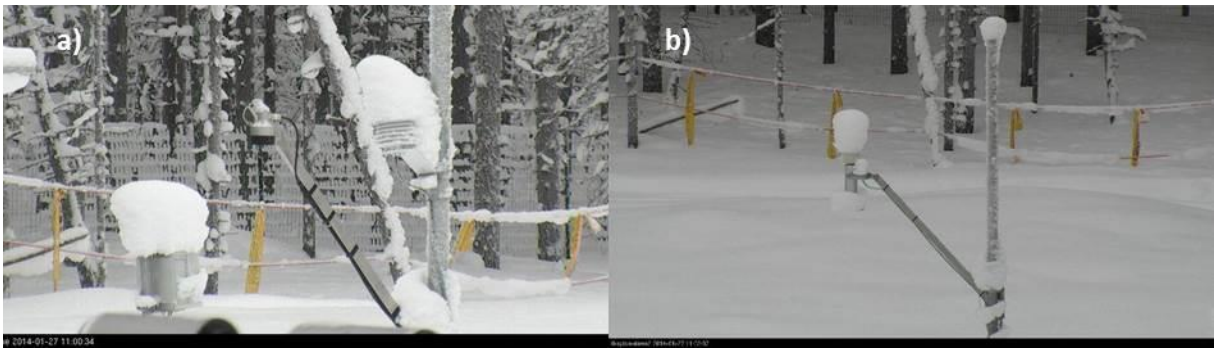


Figure 4.172. Photos of the heated angled booms and their impact on preventing snow from accumulating on the mounting infrastructure for a) SR50ATH and a heated angled boom at position 60:62 and b) USH-8 and a heated angled boom at position 70:32 at Sodankylä following the January 27, 2014 snowfall event (same event as shown inFigure 4.170).

Snow accumulation on snow-depth sensors is not limited to the sonic sensors and their mounting booms, and can also impact the SHM30 optical sensor, as shown in Figure 4.173. Even though the laser diode in the SHM30 is heated, the body of the instrument is not, making it susceptible to snow accumulation. Although there were no documented events of the SHM30 becoming completely encased in snow and impacting the measurements, this could potentially happen (e.g. after heavy snow events). One of the advantages of the SHM30 measurement principle and mount, however, is that the target area is not immediately below the sensor and mounting infrastructure, such that snow falling off the sensor will not fall onto the target area and impact the measurement.

Figure 4.173. The SHM30 at location 70:42 at the Sodankylä SPICE site following a heavy snowfall event on February 9, 2014.

4.2.6.2 Snow-depth target assessment

The purposes of a prepared surface target underneath a snow-depth sensor are:

- to provide a level and stable surface for snow to accumulate and melt (as naturally as possible) while being measured by the snow-depth sensor; and
- to provide a reflective surface (for either a sonic pulse or an optical beam) for maximum precision and reliability from the snow-depth sensor when the actual snow depth is small, inconsistent, or zero.

According to Campbell Scientific, the manufacturer of the SR50A, measurement quality can be reduced if the target is “small and reflects little sound”, “rough or uneven” or a “poor reflector of sound”. Sommer, the manufacturer of the USH-8 sensor, considers a flat target area free of obstruction and with a surface preparation of gravel or crushed rock to be representative of the natural terrain, thereby preventing inhomogeneity in accumulation or melt with the surrounding surface. Felix, the manufacturer of the SL300, suggests that the user be cautious about obstacles obstructing the beam between the intended target and the instrument. For optical sensors, Jenoptik/Lufft (manufacturer of the SHM30) does not make any recommendations about the installation of targets, but does show how target color and the distance to the target impacts the sensor signal strength measurement.

There were three types of targets used during the SPICE field intercomparison of snow-depth sensors. CARE and Caribou Creek used a perforated grey plastic target, Sodankylä used green artificial turf, and Col de Porte and Weissfluhjoch used a natural surface of mown grass and bare ground.

The plastic target at CARE consisted of a 1.2 m x 1.2 m textured and perforated sheet of grey plastic that was installed inside a wooden frame flush with the surface of the ground (Figure 4.174). The plastic material used is commercially sourced and typically used in the construction of boat docks, so the material is very durable. Landscape fabric overlaid with sand is used to prevent vegetation growth under the platform. In theory, the textured surface replicates the natural ground such that accumulating snow will not just blow off the surface, as it would if the plastic sheet was completely smooth. Perforations allow for more rapid drainage during melt. Each of the three instrument pedestals at CARE has three targets under four snow-depth sensors as shown in Figure 4.175**.** The Caribou Creek target (Figure 4.176) was similar, but due to the presence of tree roots at this site, the target was simply set on top of the surface, rather than installed on a recessed frame. The disadvantage of this technique is that the target is perhaps less stable than at CARE, and the exposed edge could be a barrier to blowing snow. Also, the target at Caribou Creek was not perforated, requiring melt water to drain off the edges of the target rather than through the target. Neither of these differences appeared to cause significant issues with this installation or impact the analysis.

Figure 4.174. Installation of the plastic targets at CARE prior to the first intercomparison season.

Figure 4.175. Completed target installation at one of the three snow depth measurement pedestals at CARE.

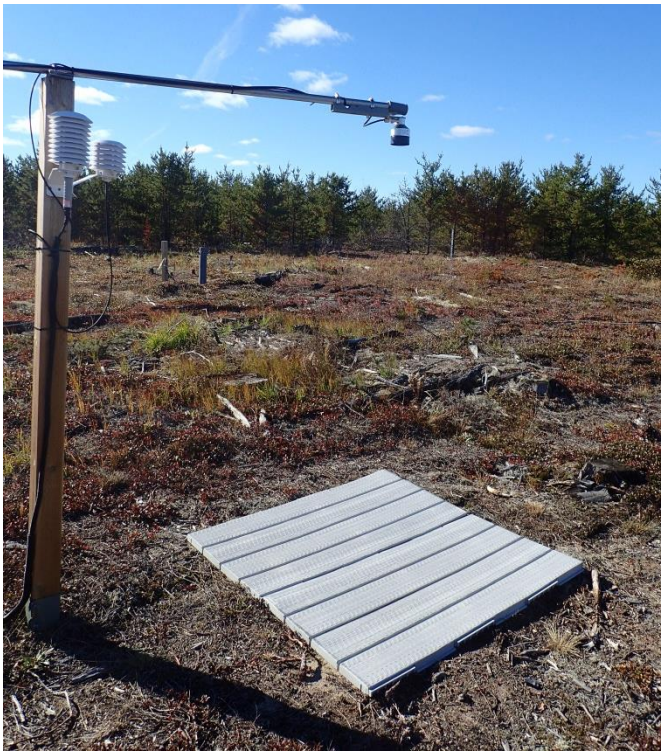


Figure 4.176. Plastic target at the Caribou Creek site.

The artificial targets at Sodankylä measured 2 m x 2.5 m and had a surface constructed from artificial turf called Limonta T8/20. The targets were green in color and constructed on a wooded frame on top of landscape fabric to prevent vegetation growth. As with the targets at CARE, the intention was to have the surface of the target flush with the level of the natural ground. Figure 4.177 shows the target during construction at an FMI observational site in Finland. Figure 4.178 shows the targets in the Sodankylä SPICE intercomparison field.

FMI has used this target method for several years, and have found that the construction provides an even, vegetation-free surface for the snow-depth measurement. An analysis of the zero-depth drift (see Section 4.2.6.3) shows that the construction is quite stable during the season and from year to year. The turf was selected to be similar to a natural grass surface but, as is evident in Figure 4.178, the surface is quite different from the surrounding landscape in the intercomparison field. Although the artificial turf should, in theory, catch and hold snow as would natural grass, the construction may have inadvertent issues related to snow melt.

a) The measurement site before construction.

b) The surface soil is removed.

c) Landscape fabric is placed on the ground. A sandbox is built of 2'' x 2'' wood.

d) The wooden frame is filled with sand.

e) Another layer of landscape fabric is placed over the sand.

f) The artificial turf covers the box.

Figure 4.177. Construction of the artificial surface target for measuring snow depth in Finland.

Figure 4.178. Artifical targets in the SPICE intercomparison field at Sodankylä.

Using the natural surface beneath the sensor appears to be the simplest technique for a surface target, but actually requires more maintenance than the artificial targets. If, for example, the natural surface is grass, it needs to be mowed to a consistent height before the first snowfall of the season. Failing to do this could result in false snow-depth reporting and inconsistencies in the zero-snow-depth distance (due to grass growth) from season to season. Two sites employed natural targets in SPICE, namely Col de Porte (Figure 4.179a) and Weissfluhjoch (Figure 4.179b). For these two sites, the target area is representative of the surrounding area, thus behaving exactly the same for snowfall accumulation and melt as the surrounding area.

It is difficult to compare the performance of targets at different sites due to the fact that the sites have different wind regimes, snowfall properties, radiation balances, etc. Therefore, the comments and recommendations that follow are based largely on a qualitative assessment of the target behavior at each of the test sites.

Figure 4.179. Natural ground targets under the snow-depth sensors at a) Col de Porte (mown grass), and b) Weissfluhjoch (grass and rock).

4.2.6.2.1 *Target impact on pre-snow and first snow-depth measurements*

The specific surface target used, whether natural or artificial, impacts significantly the measurements that the sensor makes before and during the first snow accumulation on the bare target. The measurements obtained from the bare target are typically the source of the zero snowfall distance, which is used to derive snow depth during the measurement season. A stable target, resulting in a stable zero-snow-depth measurement, will increase the quality of the accumulating snow-depth measurement for the entire season. If the target is poor (unstable, unleveled, or a poor reflector of sound/light), an accurate zero-depth measurement may not be possible, which will, in turn, impact the accuracy of the snow-depth measurements all season. Following this, a good target will allow the sensor to function as the manufacturer intended and maximize the likelihood of detecting and accurately measuring the first new accumulations of snow on the target. The intention here is not to assess the sensitivity of the sensor, but rather to isolate the impact of the target, recognizing that the target is only one potential source of measurement uncertainty.

When investigating the bare target measurements receding snowfall, the timing of the snowfall was identified and snow depth was plotted for five days prior to that event. This was done for one sonic sensor and one optical sensor at each site and for each target type, where available. For the sake of intercomparison between sites, the SR50A and the SHM30 were selected, as they are the most widely-used snow-depth sensors under test. Figure 4.180 shows the SR50A depths preceding snowfall at CARE (Figure 4.180a), Sodankylä (Figure 4.180b), and Col de Porte (Figure 4.180c), representing artificial plastic, artificial turf, and natural targets, respectively. Figure 4.181 shows the SHM30 snow depths preceding snowfall at CARE (Figure 4.181a), Sodankylä (Figure 4.181b), and Weissfluhjoch (Figure 4.180c), representing artificial plastic, artificial turf, and natural targets, respectively. Prior to the snowfall events at each of these sites, and for each target type, three-day standard deviations of the bare target distances were calculated, which are shown in Table 4.33.

From the data shown in Figure 4.180 and Figure 4.181, there are several important considerations. First, the bare target snow depth is not always zero and depends on the zero-depth distance entered by the user during installation or during post-processing. A non-zero offset is irrelevant in this analysis, but would be corrected by post-processing if deemed by the user to be significant to snow-depth calculations. The other consideration, as implied earlier, is that the noise seen in the data is not necessarily a reflection of the target performance or the instrument performance and could be related to the mounting structure or other external factors.

All other things being equal, the bare target measurements made by the SR50A over the plastic and artificial turf targets (Figure 4.180a and Figure 4.180b) are relatively stable and exhibit a standard deviation (SD) around the mean of 0.13 and 0.20 cm, respectively. These are well within the accuracy of ± 1 cm for this instrument as specified by the manufacturer. Figure 4.180c shows the bare target measurement by the SR50A over mown grass. In this case, the SD increases to 1.0 cm, with a specified accuracy of 0.4% of the distance to the target, or ± 1.6 cm. This increased variability of the bare target measurement could be caused by the sonic sensor interpreting the returned signal from a non-solid surface (grass mown to a height of 2 or 3 cm). The sensor will generally report distances to the tallest object in the field of view, which could vary on a grass surface. However, this sensor at Col de Porte is installed at a height of 4 m (the other two sites are at 2 m), which may be a factor in the increased noise observed in the measurements from this sensor at this site. The optical sensor, shown in Figure 4.181, behaves somewhat differently over the four target types (the sonic sensors show larger SD values).

Table 4.33. Standard deviations of the bare target measurements for various target types, sensors and sites.

Target Type	Sensor	Site	Bare Target SD [cm]	Sensor Accuracy* [cm]
Plastic	SR50A	CARE	0.13	±1.0
	SHM30	CARE	0.04	±0.5
Artificial Turf	SR50A	Sodankylä	0.20	±1.0
	SHM30	Sodankylä	0.06	±0.5
Mown Grass	SR50A	Col de Porte	1.00	±1.6
	SHM30	Col de Porte	0.06	±0.5
Bare Ground/Rock	SHM30	Weissfluhjoch	0.07	±0.5

*as specified by the manufacturer

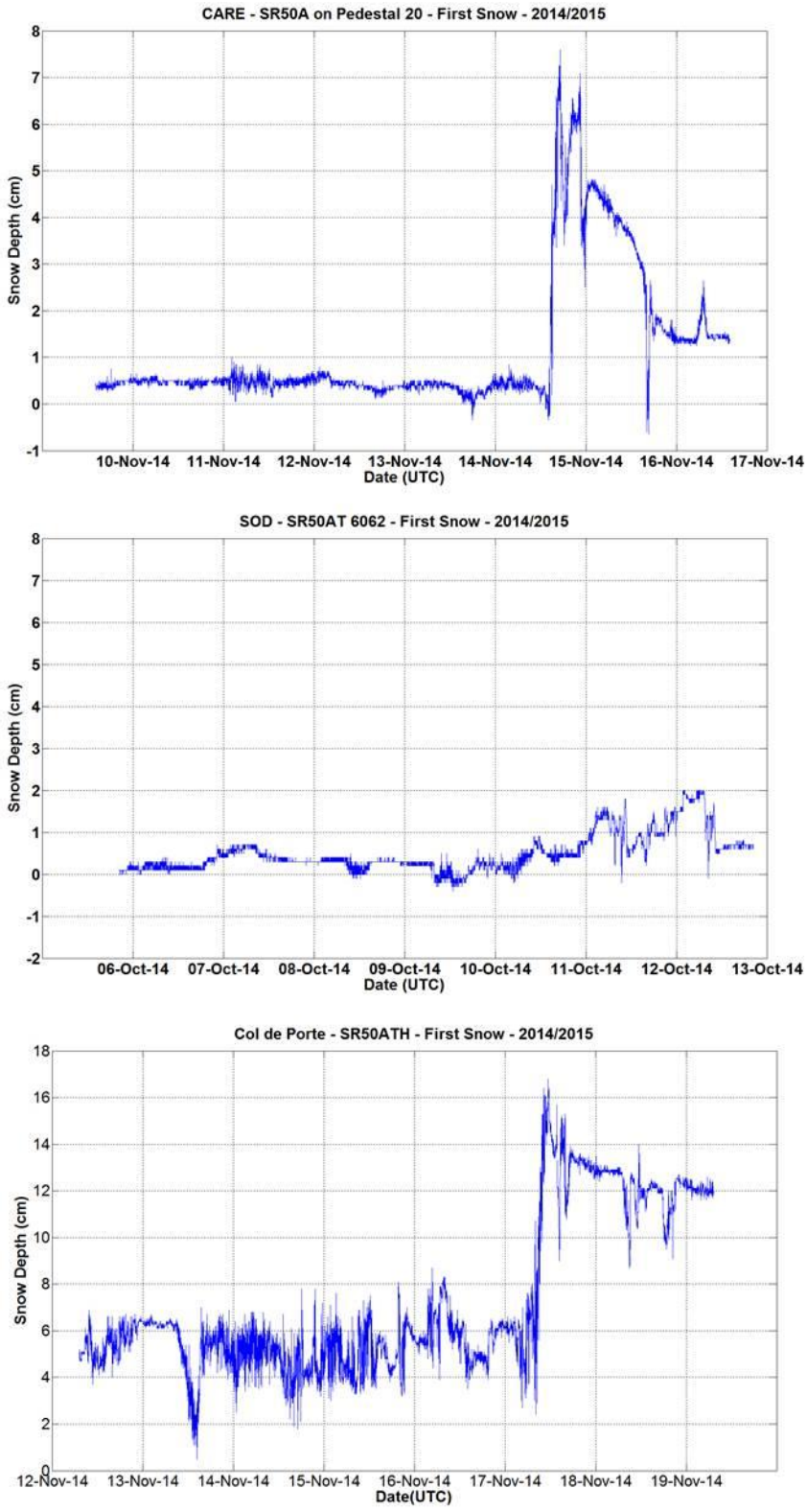


Figure 4.180. Measurements of bare target snow depths preceding a snow event on the bare target as measured by the SR50A at CARE (top, plastic target), Sodankylä (middle, artificial turf), and Col de Porte (bottom, mown grass).

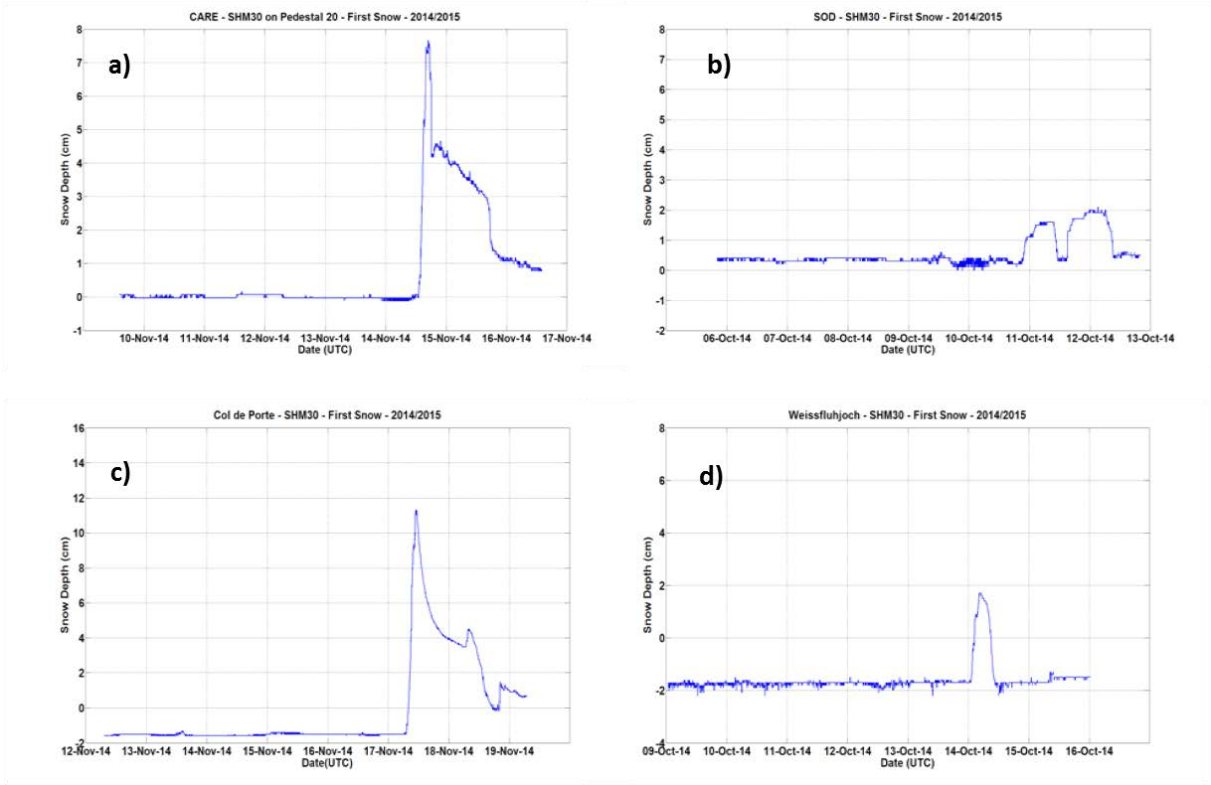


Figure 4.181. Measurements of bare target snow depths preceding a snow event on the bare target as measured by the SHM30 at a) CARE (plastic target), b) Sodankylä (artificial turf), c) Col de Porte (mown grass), and d) Weissfluhjoch (bare ground/rocks).

Overall, the SD values for the sonic sensors in Table 4.33 are higher than those for the optical sensor for all target types, largely because of the different measurement principles of the sensors. Because of limitations with respect to the measurement principle for sonic sensors, lower accuracy values are stated by the manufacturers for sonic sensors relative to optical sensors.

As noted above, the SD values in Table 4.33 indicate that noise in the bare target signal from the optical sensor is less than that for the sonic sensors. This corresponds with higher manufacturer specified accuracy for the optical sensor. Both artificial targets (Figure 4.181a and Figure 4.181b) appear to work well with the optical sensor. The SD values for the bare target measurements over natural targets (Figure 4.181c and Figure 4.181d) are comparable to those for the artificial targets. The first snowfall on the bare targets is detectable for all target types.

It is difficult to draw clear conclusions from this assessment, as only one target type is tested at each site, and each site experiences different environmental conditions. Based on the results available, it appears that the sonic sensors are more susceptible to noise related to natural grass targets. The optical sensors appear to be less reliant upon the surface target to produce a reliable zero-snow-depth distance.

4.2.6.2.2 *Catching snowfall on a bare target*

Artificial targets provide a level and stable surface for snow to accumulate and have the potential to improve the measurements made by some sensor types relative to natural targets. However, one of the disadvantages of artificial targets results from their impact on surface homogeneity; that is, the

target surface is no longer representative of the surrounding area. Besides the potential for changing the energy balance of the surface, an artificial target could also change the texture of the surface and/or produce an edge effect in the measurement area. Each of these may impact the way the target collects snow, especially when the target starts out bare, and can affect the representativeness of snow measurements at a given site. For example, a smooth target could cause falling snow to blow off of the target and be deposited in the natural area away from the target and the sensor's field of view, while a target with too much texture or an edge could have the opposite effect, resulting in the deposition of more snow than the surrounding environment.

The artificial plastic targets used at CARE have a textured surface, designed to prevent slips and falls when the material is used as a surface for boat docks. In theory, this textured surface should help to catch falling snow and prevent that snow from blowing away from the target area, much in the same way as shortly mown grass or a gravel pad would. Site photographs from CARE show mixed results. The webcam at the site wasn't aimed at the snow-depth targets, so this qualitative analysis relies on photos taken during site visits. Also, some of the CARE photos are from a non-SPICE experiment to test the plastic targets for operational use.

Figure 4.182 shows a photo of one of the instrument pedestals at CARE on November 12, 2013, after a snowfall event of approximately 9-10 cm. From the photo, the snow depth appears relatively uniform around the pedestal, with the outline of the targets just barely visible. Although difficult to quantify, it appears that the snow across the target areas is similarly uniform as around the pedestal.

Figure 4.182. Distribution of snow at CARE on November 12, 2013, after a snowfall event on bare ground and bare targets.

Photos of the experimental targets located away from the sensors under test show occasions in which the snow depth on the targets after an event is not as uniform as the surrounding area. Figure 4.183 shows a photo of a plastic target on December 17, 2014, after a snowfall event at CARE. Here, the target shows less snow than the areas surrounding it, with the exception of the ground directly behind the target, which has even less snow. In this case, it looks like snow has blown off of the target, resulting in less snow on the target than in surrounding areas. Edge effects from the target installation may also be impacting snow deposition/scouring around the target area. Another photo in the same area taken on December 20, 2014 (Figure 4.184) shows the target nearly completely scoured of snow while the surrounding ground is still mostly covered.

Figure 4.183. Experimental plastic target at CARE taken on December 17, 2014, after a snowfall event on the bare ground and targets. Photos obtained during a pre-SPICE target assessment project (Courtesy of Jeffery Hoover, Environment and Climate Change Canada).

Figure 4.184. Experimental plastic target at CARE taken on December 20, 2014, several days after a snowfall event on the bare ground and targets.

From the qualitative analysis of the photos at CARE, the performance of the textured plastic target with respect to representing the surrounding area during the first snow is somewhat mixed. It appears that the presence of wind, resulting in blowing snow during and after a snowfall event on a bare target, can remove snow from the target, which would cause the sensor to underestimate snow depth. However, once the snow depth around the sensor is more substantial (e.g. 10 cm), the target may be less susceptible to scouring and the depth measurements may be more representative.

The artificial turf targets used at Sodankylä were designed to mimic natural grass, and it was expected that they would catch and hold snow on the target as it fell, in much the same way as natural grass. Photos from the site during and after snowfall on the bare targets (Figure 4.185) show that this appears to be the case. The snow sticks to the turf and should therefore start to accumulate in a natural manner. However, wind speeds at the Sodankylä site were lower in comparison to CARE, so it is difficult to say from this study whether or not the targets would also perform satisfactorily in windy conditions.

Figure 4.185. Artificial turf targets at Sodankylä during or shortly after a snowfall event on the bare targets October 11, 2014 (observed depths for this event are shown in Figure 4.180b and Figure 4.181b).

4.2.6.2.3 *Snow melt from artificial targets*

A key advantage of artificial targets is the provision of a level and stable platform upon which snow can accumulate. However, the construction of an artificial target and the related changes to the natural ground has the potential to change the energy balance of the area under the sensor. This could potentially result in melt rates that differ from the natural ground, and in turn, in unrepresentative measurements of snow depth during melt.

Hourly web cam photos of the grey plastic target and the surrounding intercomparison field are available for one melt season at Caribou Creek. Here, the plastic target is surrounded by a sandy area

with surface growth of lichens and short grasses, with exposed roots and stumps left over from when the site was cleared of trees. As the snow melts at this site, areas where the stumps and vegetation become exposed will melt out first, creating an inhomogeneous surface and vastly impacting the energy balance. The time lapse photos of Caribou Creek taken around April 30, 2014 (Figure 4.186) show areas opening in the snow pack where these stumps and vegetation are exposed. The plastic target starts becoming clear adjacent to some exposed areas, but doesn't clear immediately. The target area is not the first or the last area to become snow free (see Figure 4.187), suggesting that this particular target type is suitable for sites with inhomogeneous surface radiative properties. More work is required on the behavior of targets in areas with inhomogeneous surface cover. The use of hourly or daily photometry would aid in the interpretation of representative depths in inhomogeneous areas.

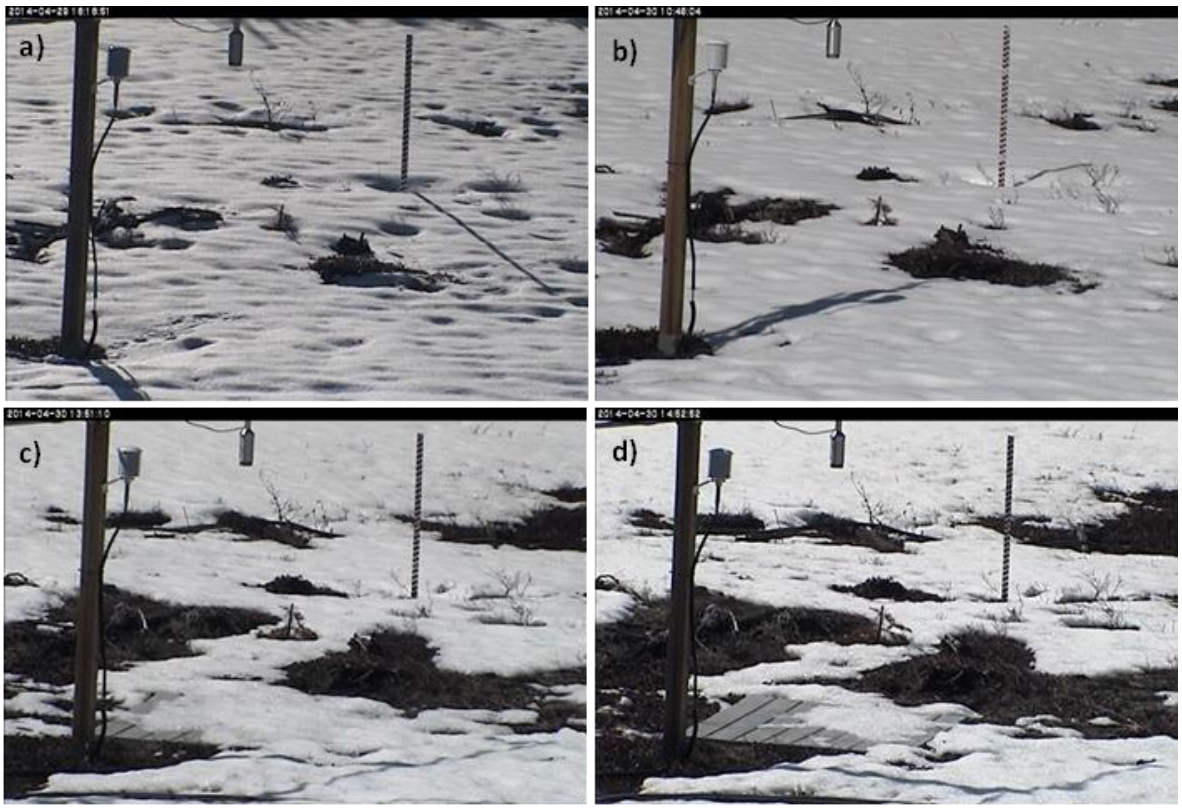


Figure 4.186. Time lapse photographs of the snow melt around the artificial plastic target at the Caribou Creek SPICE site: a) April 29, 2014, 1818 UTC; b) April 30, 2014, 1048 UTC; c) April 30, 2014, 1351 UTC; d) April 30, 2014, 1452 UTC.

Figure 4.187. Wide angle view of the Caribou Creek intercomparison field (snow-depth target in the center) on April 30, 2014, at 1517 UTC, shortly after the photo shown in Figure 4.186d.

The behavior of the artificial turf targets during melt, as tested at Sodankylä, was not ideal. The targets in the Sodankylä intercomparison field (as shown in Figure 4.178) appear to have different surface radiative properties than the surrounding area; this is not surprising, given the contrast between the dark green turf and the lighter surrounding soil. Snow on the targets appears to melt much faster than that on the surrounding surface. This is evident in the photo in Figure 4.188, which was taken approximately 5 hours after the photos in Figure 4.185, following a snowfall event. This differential melting also occurs during spring melt, when the dark green targets melt much more rapidly than the surrounding landscape in the intercomparison field, as shown in Figure 4.189. This is one disadvantage of this target selection for a bare sand surface; performance may vary for other surface types with more similar radiative properties, such as mown grass.

Figure 4.188. Bare targets in the Sodankylä intercomparison field taken on October 11, 2014, 1104 UTC, 5 hours after the photos shown in Figure 4.185.

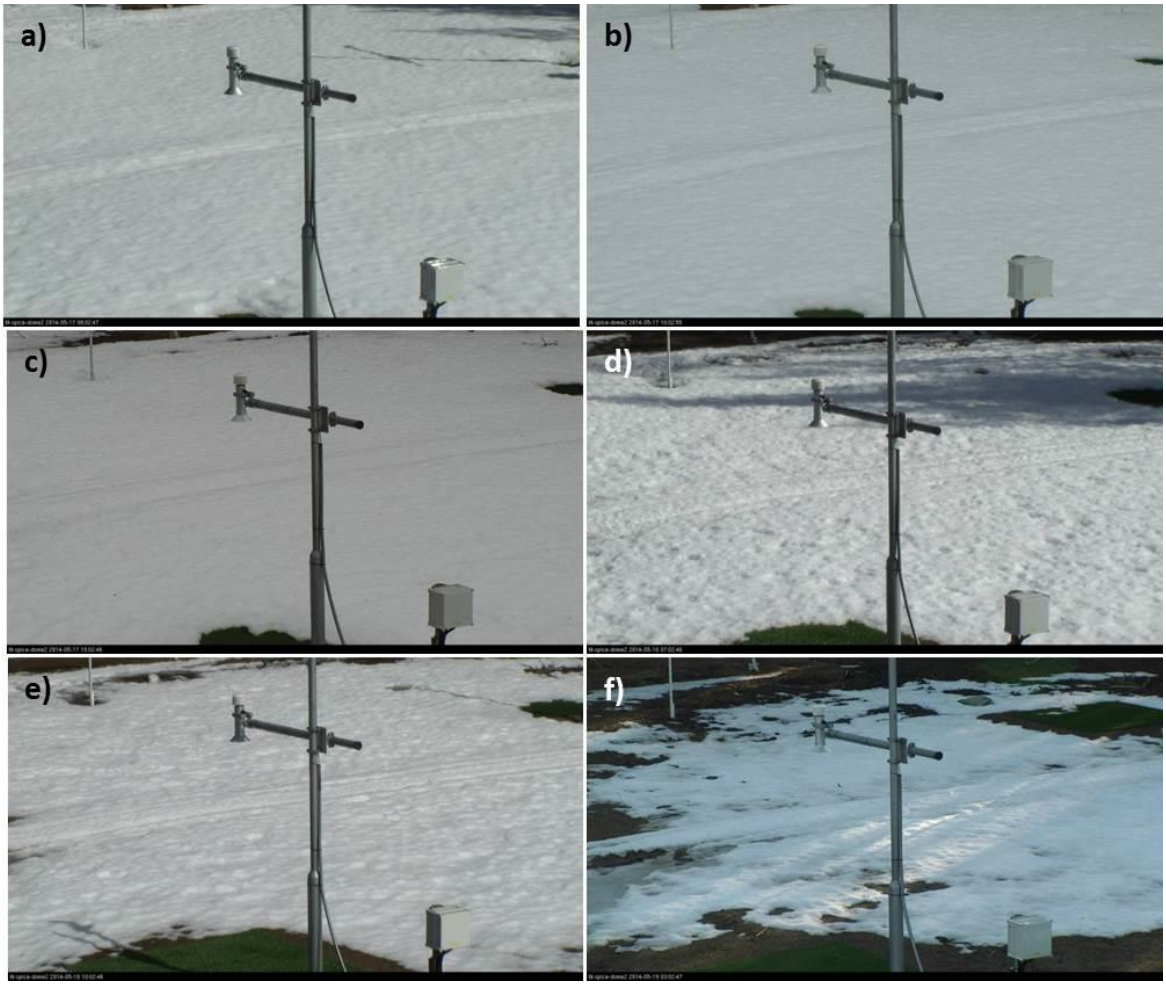


Figure 4.189. Time lapse photos of the target area under the USH-8 at Sodankylä during spring melt: a) May 17, 2015, 0800 UTC; b) May 17, 2015, 1000 UTC; c) May 17, 2015, 1502 UTC; d) May 18, 2015, 0702 UTC; e) May 18, 2015, 1002 UTC; and f) May 19, 2015, 0302 UTC.

4.2.6.3 Assessment of zero-snow-depth drift

Most snow-depth sensors don't measure snow depth, but rather, the distance from an above-placed sensor to a target below. This is true for both the sonic and optical sensors tested during SPICE. Snow depth is then derived as the difference between the distance to the target and the distance to the snow free surface. One potential issue with this method is a change in the distance to the snow-free target before or during the accumulation and measurement period. Ideally, the zero-snow-depth (ZSD) should be the same at the end of the accumulation season as it is at the beginning. A shift in this distance could be caused by a change in the mounting infrastructure for the sensor, a settling or heaving of the target area relative to the sensor, or both. This is referred to here as "zero-snow-depth drift". Whether a snow-depth sensor is installed for operational or research purposes, it should be recognized by the users of the sensor and the data that ZSD drift is a potential source of uncertainty in the derived snow-depth estimates. A check for ZSD drift should be completed by the user at the end of each season and any adjustment to the ZSD should be made before the beginning of the next accumulation period.

Prior to start of the SPICE measurement periods, most of the snow-depth sensors were new installations. Instrument mounts were constructed and snow-depth targets designed and installed. For this reason, some ZSD drift was expected to occur during the measurement periods, largely related to the installation of the new targets (see Section 4.2.6.2 for a description of the target installations). This significance of the ZSD drift was expected to be dependent on the design of the target, the installation method, the substrate over which the target was installed, and the seasonal snow load.

Before the beginning of the SPICE observation periods, the site managers estimated the ZSD value for each sensor by either manually measuring the distance between the sensor and the target or using the sensor output distance to the snow free surface. The assumption is that this value will be constant throughout the observation period. To test this assumption for each intercomparison season and for each sensor, a snow free period was selected just prior to the first seasonal snowfall and just after the final seasonal melts. These periods were selected visually after plotting the time series for the entire season. The length of these selected snow free periods varied between sites and instruments due to the need to capture variations in the ZSD values or because of missing or incomplete data at the beginning or end of the accumulation periods. Generally, the length of this period was approximately three days, comprising over 4000 1-minute measurements. The time stamps of these periods were recorded and the mean and standard deviation of the ZSD were calculated. The ZSD drift was then calculated from the difference between the spring (post melt) and the autumn (pre accumulation) averages. A positive ZSD drift indicates that the target height in the spring was higher relative to the sensor, which could be a result of frost heave. A negative ZSD drift indicates that the target in the spring is lower relative to the sensor, which could be a result of settling.

A summary of ZSD values for the snow depth sensors tested in SPICE is provided in

. If settling of the new target was the cause of the ZSD drift, it is expected that the drift would be greatest for the first season of operation and diminish in subsequent seasons. In heavier clay soils, frost heave of the target could occur during any season, and the sensor may experience both settling after installation and subsequent frost heave. The differences between season 1 and season 2, where available, are also shown in Table 4.34.

Table 4.34. Summary of the zero-snow-depth drift values and differences between the 2013/14 and 2014/15 seasons.

Site	Instrument/ Pedestal	ZSD Drift (Spring-Fall) [cm]		Difference between 2013/14 and 2014/15 seasons [cm]
		2013/14	2014/15	(2013/14)-(2014/15)
CAR	SR50A 12A	0.84	-1.03	1.87
	SR50A 11A	0.83	-0.22	1.05
	SR50A 20	0.53	-0.52	1.05
	USH-8 12A	0.96	0.01	0.95
	USH-8 11A	1.17	-0.24	1.41
	USH-8 20	0.98	0.94	0.04
	SL300 12A	1.20	1.21	-0.01
	SL300 11A	1.09	-0.17	1.26
	SL300 20	-	0.34	-
	SHM30 12A	-	0.61	-
	SHM30 11A	-	-0.13	-
	SHM30 20	-	0.49	-
SOD	USH-8 #1	0.00	-0.01	0.01
	USH-8 #2	0.00	-0.47	0.47
	SHM30	-0.08	-0.30	0.22
	SR50A 6062	-0.22	-0.51	0.29
	SR50A 7052	-0.27	-0.73	0.46
	USH-8 #1	0.00	-0.01	0.01
CDP	SHM30	-	0.23	-
	SR50A	-	-3.73	-

In general, the ZSD drift for most of the snow-depth sensors tested during SPICE was low. At CARE in 2013/14, the average drift was under 1 cm, which is within the measurement uncertainty of the sonic sensors. The average (using the absolute values of ZSD drift) for that same sub-set of sensors in 2014/15 was approximately 0.5 cm. Combining both seasons, only 5 of the 20 ZSD drift numbers exceeded 1 cm. For 2013/14 at CARE, all of the ZSD drift values were positive, indicating that the distance between the sensor and the target decreased during the accumulation period. This suggests

that the target platform heaved over the course of the first winter. Overall, the ZSD drift values for sensors at CARE were lower for 2014/15. In fact, several negative values were observed, indicating that the distance to the target actually increased over the course of the winter, perhaps due to settling of the target back into the soil.

The targets at Sodankylä appeared to be relatively stable over the course of both seasons; a maximum change in ZSD of -0.73 cm was observed, but values were generally less than 0.5 cm. The average absolute change over both seasons was less than 0.5 cm. The greater degree of stability at Sodankylä is most likely due to the sandy substrate under the targets, which is less susceptible to further compaction or frost heave.

The ZSD drift at Col de Porte was low for the SHM30, but quite high for the SR50A. The ZSD drift for the SR50A was negative, indicating that the target was lower in the spring relative to the sensor as compared with the fall. Because the surface is relatively stable and undisturbed, the change in the SR50A ZSD could be related to the growth of grass under the sensor before the first snowfall in the fall, and the compaction of this grass throughout the winter.

A seasonal assessment of the ZSD is recommended, and is perhaps more important if using artificial targets in heavier (clay) soils, as these tend to be less stable than natural or sandy substrates. If the ZSD drift approximates sensor accuracy (e.g. +/- 1 cm for the SR50A), no adjustment should be made. If the ZSD is offset but this offset is approximately equal for season start and season end, then the entire snow-depth dataset for that season should be adjusted accordingly. Adjusting the data if the ZSD is unequal from season start to season end is difficult. Without knowing the rate of drift, a linear adjustment is likely the only potential correction; but unfortunately, the drift is likely not linear. If the drift is due to compaction and settling of the target, the rate of drift may be related to total snow load on the target and/or substrate moisture, but this has not been demonstrated explicitly. It is best to recognize that drift has occurred and factor this drift into the total uncertainty of the measurement.

4.2.6.4 Air temperature correction for ultrasonic snow-depth measurements

Ultrasonic snow-depth measurements rely on the time delay between the emission of an ultrasonic pulse by the sensor and the reception of the signal reflected by the snow surface. Because of the temperature dependency of the velocity of sound in air, this measurement technique requires a correction for the temperature of the air in the column beneath the sensor. This section specifically addresses different methods to correct ultrasonic measurements for air temperature effects.

The influence of the air temperature profile between the snow surface and the snow-depth sensor is difficult to quantify. Strong temperature gradients can develop over the surface of snow, as the snow surface is generally colder than the air above due to radiative cooling effects (Andreas, 2002). The location of a temperature sensor between the surface and sensor will, therefore, influence the correction. Further, the impact of errors in air temperature measurements, which can be specifically enhanced over a snow covered surface (Huwald et al., 2009), must also be considered.

Considering an ultrasonic sensor located at an altitude h_{tot} above ground, and a snowpack with a snow depth h_s, the distance d between the sensor and the snow surface is related to h_{tot} and h_s by (A1):

$$h_s = h_{tot} - d$$

Figure 4.190 shows an overview of the geometrical configuration of the ultrasonic sensor considered in this analysis.

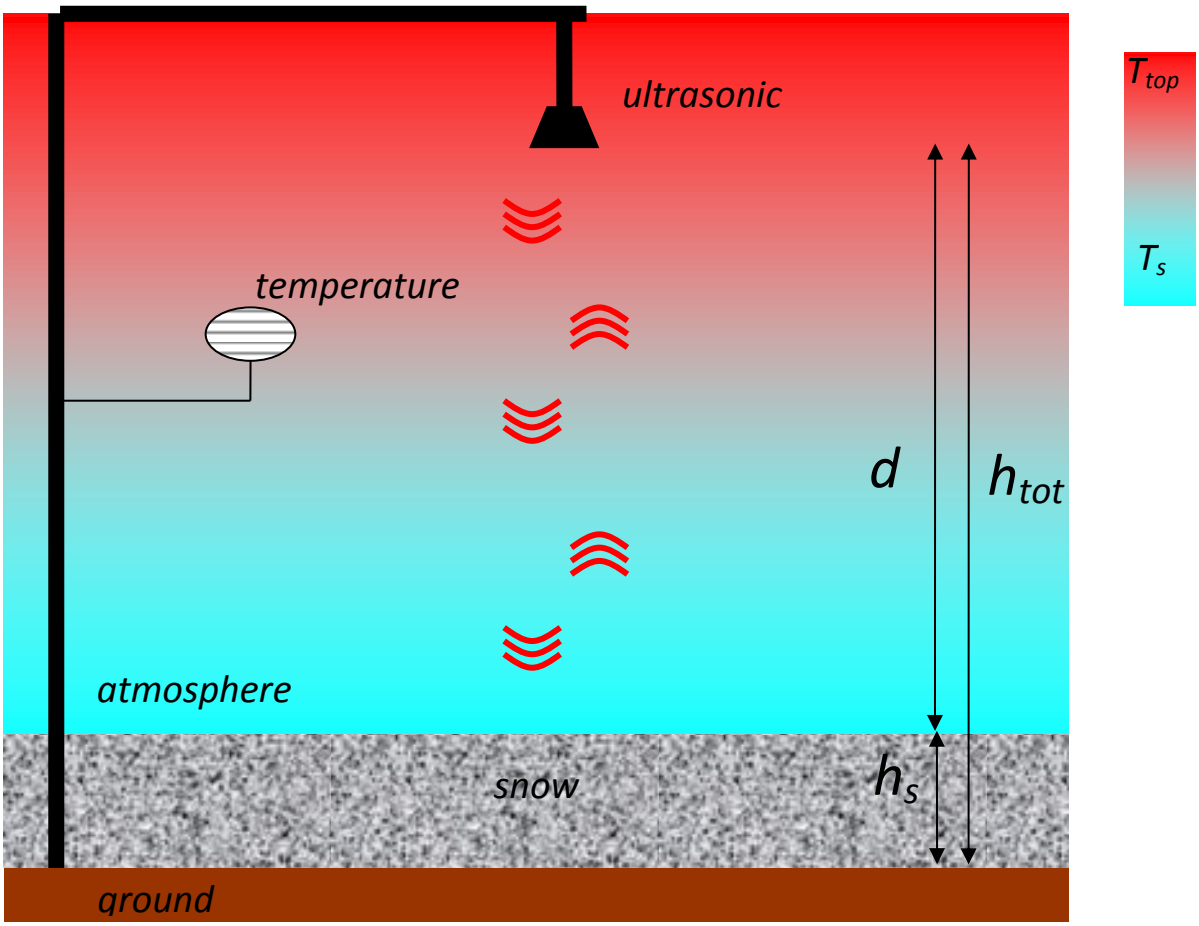


Figure 4.190. Overview of the geometrical configuration of an ultrasonic sensor.

Ultrasonic measurements of h_s are possible through the integration of the following equation, which derives from (A1), where c is the speed of sound in the atmosphere (A2):

$$h_s = h_{tot} - \int_0^{\Delta t/2} c\, dt$$

Here, Δt is the time difference between the emission of the ultrasonic pulse and the reception of its reflection back to the sensor.

It is necessary to account for the fact that the speed of sound in the atmosphere depends on temperature (A3):

$$c(T) = c_0 \left(\frac{T}{T_0}\right)^{1/2}$$

where T is temperature and T_0 is a reference temperature for which the velocity of sound c is known $c(T_0) = c_0$. T_0 is usually set to 0 °C.

Combining equations (A2) and (A3) leads to (A4):

$$h_s = h_{tot} - \left(\frac{c_0}{T_0^{1/2}}\right) \int_0^{\Delta T/2} T^{1/2}\, dt$$

Accounting for temperature variations of c is generally carried out using a single temperature, assumed to be representative of the conditions for the measurements, referred to here as T^*. The estimate of h_s obtained using this single temperature value T^* is referred to as h_s^*. Solving (A4) in the case where T is assumed to be constant in time and equal to T^* gives (A5):

$$h_s^* = h_{tot} - c_0 \left(\frac{T^*}{T_0}\right)^{1/2} \left(\frac{\Delta t}{2}\right)$$

The previous equation is typically used for practical implementation of the ultrasonic measurements of snow depth, using measured values of Δt and T^* and known values of h_{tot}, c_0 and T_0. In fact, the actual value of Δt, the return travel time of the snow pulse, depends on the actual profile of air temperature above snow (A6):

$$\Delta t = 2 \int_0^d \frac{dz}{z}$$

Combining (A1) with (A5) and (A6), this leads to an alternative expression for the "true" value of h_s (A7):

$$h_s = h_{tot} - (T^*)^{1/2} \int_0^d \frac{dz}{T(z)^{1/2}}$$

Note that in case of a uniform temperature profile above snow equal to T^*, $h_s^* = h_s$.

The difference between the estimated value of h_s (h_s^*) and the "true" h_s^* value is defined as Δh_s. In case of a non-uniform temperature profile above snow, Δh_s is given by (A8):

$$\Delta h_s = h_s - (T^*)^{1/2} \int_0^d \frac{dz}{T(z)^{1/2}}$$

In case of a uniform temperature profile T above snow, Δh_s is given by (A9):

$$\Delta h_s = h_s \left(1 - \left(\frac{T^*}{T}\right)^{1/2}\right)$$

Here again, $\Delta hs = 0$ if the uniform value for T is equal to T^*.

It is evident from (A8) and (A9) that the factors most affecting the value of Δh_s are the shape of the temperature profile, the choice of the correction temperature T^*, and the vertical distance d

between the ultrasonic sensor and the snow surface. This equation can be solved for various vertical profiles of temperature between the snow surface and the ultrasonic sensor height. The following discussion explores the impact of these variables on the error of snow-depth measurements, based on theoretical case studies and field investigations. Increasing the distance between the sensor and the surface potentially results in larger temperature correction values.

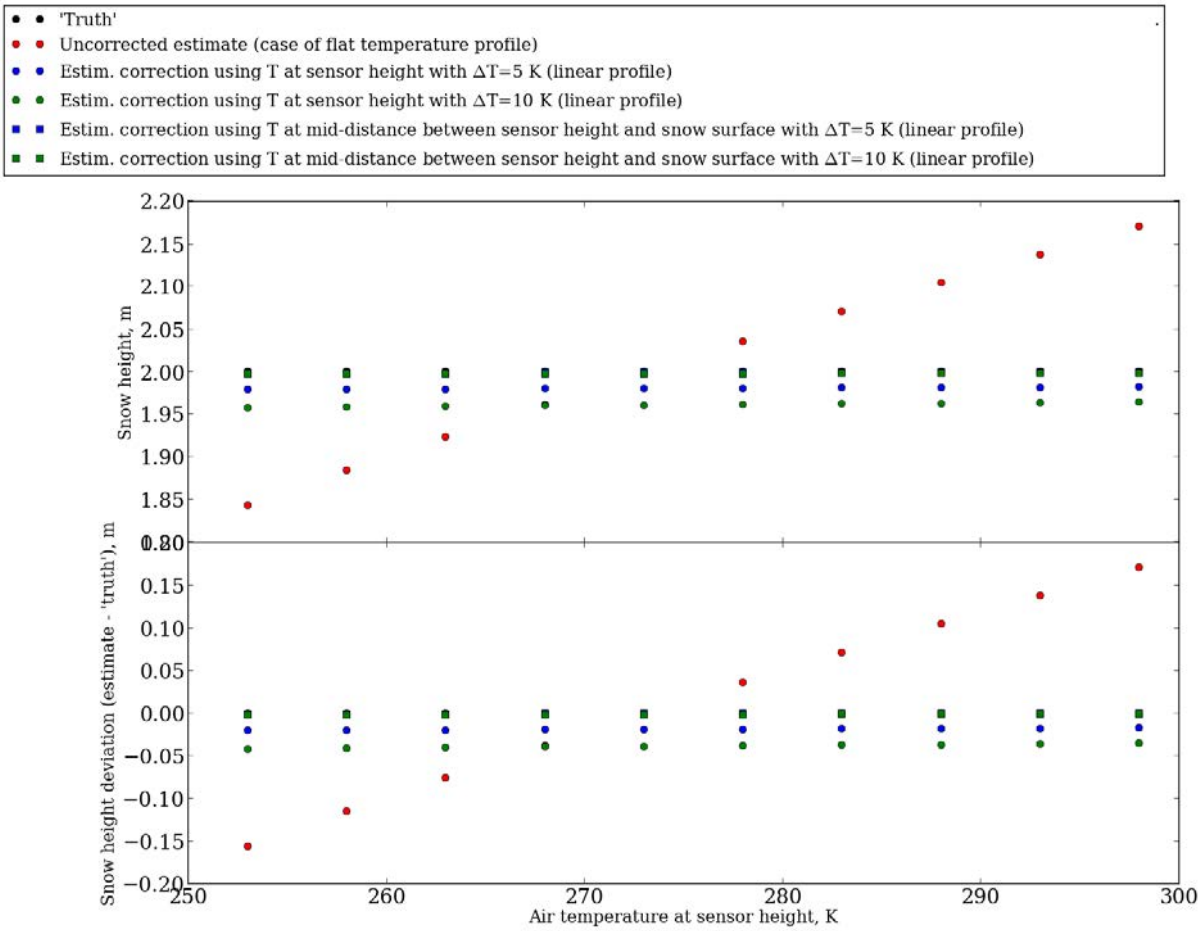


Figure 4.191. Snow depth (top) and deviation between estimated and true snow depth (bottom) as a function of air temperature at sensor height, for various temperature profiles and configurations of the temperature sensor (located at the height of the ultrasonic sensor or at mid-distance between snow surface and ultrasonic sensor height).

Figure 4.191 shows an example of the deviation that can be expected between true and measured snow depth for a configuration in which the distance *d* between the sensor and the snow surface is 4 m, and the snow depth h_s is 2 m. This graph primarily shows that, without any temperature correction and assuming a constant temperature field between the snow surface and the sensor height, measurement errors on the order of several centimeters can be encountered. The magnitude of errors depends on the difference between the actual temperature and the reference temperature used to estimate the speed of sound (typically 0 °C).

Assuming that there is a linear temperature profile, with a 5 °C temperature difference between the snow surface and the ultrasonic sensor height (lower temperature at the snow surface), and using

the temperature measurement at the ultrasonic sensor height results in a snow-depth underestimation of 2 cm (blue circles in Figure 4.191). Indeed, this corresponds to a case in which the actual air temperature is generally lower than at the measurement point, hence leading to an overestimation of the distance between the sensor and the snow surface. This corresponds to a 0.5% error on the distance estimate between the sensor and the snow surface. Using the temperature found at mid-distance between the sensor and the snow surface leads to a deviation of snow depth smaller than 1 cm (blue squares in Figure 4.191).

Assuming a linear profile with a 10 °C temperature difference between the snow surface and the sensor leads to larger measurement errors (4 cm) using temperature at the ultrasonic sensor height (1% error on the distance estimation between the sensor and the snow surface), and negligible (less than 1 cm) error using temperature measured at mid-distance between the ultrasonic sensor height and the snow surface. In both cases (5 and 10 °C differences), the differences found hardly depend on the temperature at ultrasonic sensor height.

The same calculations were also performed by adding a 2 °C bias to the temperature measurement at sensor height. Such biases can be generated by using an improper radiation shield for the temperature sensor, particularly over snow surfaces (e.g. Huwald et al., 2009). With such a bias, the measurement error reaches 6 cm, which corresponds to a 1.5% error on the distance between the ultrasonic sensor and the snow surface.

The analysis demonstrates that temperature corrections are absolutely required for accurate measurements of snow depth using an ultrasonic sensor. Taking into account the fact that the temperature field may not be uniform between the ultrasonic sensor height (where temperature measurements are most often performed) and the snow surface, and that it could exhibit a temperature gradient, indicates that measurement errors can result from the use of a single temperature measurement to estimate air temperature. Errors are minimized when the temperature measurement is performed at mid-distance between the ultrasonic sensor and the snow surface, but it is recognized that this configuration is generally not feasible when measuring deep snow packs. In these situations, the mid-distance location changes substantially with snow depth and it is impractical to automatically adjust the height of the temperature sensor. If the sensor is installed at mid-distance between sensor height and the ground surface, in order to minimize errors under shallow snow conditions, there may be cases where the temperature sensor becomes snow-covered during the winter season, making any correction virtually impossible. However, even when the temperature sensor is located at the height of the ultrasonic sensor, maximum errors remain within a few centimeters at most.

The trade-off between the potential snow-depth measurement errors due to the placement of the temperature sensor at the same height as the ultrasonic sensor on the one hand, and detrimental practical implications of attempting to place it at mid-distance between the ultrasonic sensor and the snow surface on the other hand, corroborates the choice generally made to install the temperature sensor at the same height as the ultrasonic sensor, which corresponds to the manufacturer's instructions. However, significant errors can be made if the temperature sensor used for the correction deviates from the actual air temperature, which can in particular be driven by the use of an improper radiation shield.

Figure 4.192 shows field observations of snow-depth measured using an ultrasonic sensor (Campbell Scientific SR50A) and a laser sensor (Jenoptik SHM30), along with ventilated temperature measurements at sensor height (4 m above ground) and at snow surface (using a Campbell Scientific

IR120 instrument) at the Col de Porte SPICE site. Measurements were obtained at a time when the snow depth was on the order of 1 m. All records are shown at 1-minute time resolution. The best agreement between the laser (unaffected by air temperature) and ultrasonic sensor is found when the ventilated air temperature measurement carried out at the ultrasonic sensor height is used. The largest deviation is found when using raw output from the ultrasonic sensor. When the temperature used for the correction is computed as the mean between air temperature at ultrasonic sensor height and the snow surface, results show higher deviations between the ultrasonic and the laser sensor snow depth. This is most likely due to the fact that the temperature profile above snow is not linear (usually, stronger gradients are found near the snow surface; Andreas, 2002), and taking the arithmetic mean between snow surface and sensor-level temperature places too much weight on the lower (usually colder) part of the atmospheric column between the snow surface and the sensor height. This is exacerbated when there is a large temperature difference between the snow surface and the height of the ultrasonic sensor. This example illustrates that attempting to refine temperature corrections of ultrasonic measurements can ultimately be detrimental to the quality of results. Using ventilated measurements at ultrasonic sensor height is thus both a relatively simple and practical solution to reduce errors in snow depth related to the temperature correction.

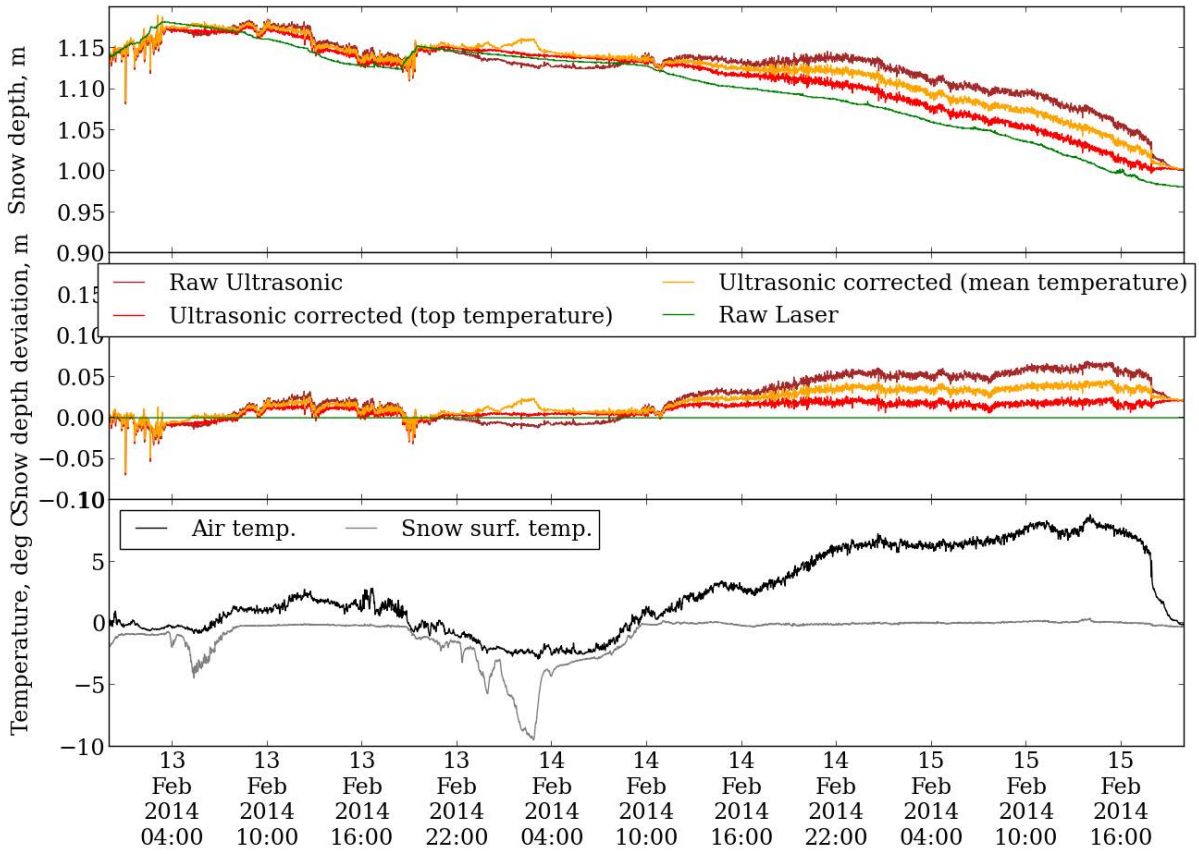


Figure 4.192. Comparison of snow depth (top), snow depth deviation between ultrasonic and laser sensor (middle), and air temperature and snow surface temperature (bottom) at Col de Porte from February 13 to 15, 2014.

4.3 Linking changes in snow depth to precipitation

Authors: Samuel Morin, Craig Smith, Anna Kontu, Yves Lejeune

4.3.1 Introduction

One of the SPICE snow-on-ground objectives is to test the relationship between the change in snow depth as measured by an automated sensor and the total precipitation amount as measured by the SPICE reference configuration. In general, snowfall leads to an increase of snow mass on the ground, and it follows that it could be attempted to use snow on the ground amount (or depth) variations to infer the precipitation amount in the absence of a precipitation gauge capable of measuring snowfall. Data collected during SPICE can be used to investigate the feasibility of this approach in more detail, by examining the relationships between snow depth and reference precipitation observations at different SPICE sites.

The estimation of snow precipitation amounts from variations of snow depth on the ground is complicated by several factors. First, snow depth and SWE can both be affected by non-precipitation processes, such as wind redistribution, sublimation, and snow melt, which lead to snow mass variations irrespective of any precipitation. Second, the depth of the snow on the ground as measured by an automated sensor during a precipitation event is also impacted by compaction of layers underlying the freshly fallen snow, as well as by densification of the fresh snow itself. Third, the wide range of density values observed for freshly-fallen snow (from < 50 kg/m^3 to > 200 kg/m^3; see e.g. Essery et al., 2013) complicates the determination of a direct relationship between the thickness of freshly-fallen snow and the corresponding precipitation amount.

Relationships are examined for 60 minute accumulation periods of snowfall at CARE, Sodankylä, and Col de Porte. For CARE and Sodankylä, precipitation as measured by the SPICE R2 reference was extracted from the 60 minute SEDS (as defined in Section 3.4) which incorporates a minimum precipitation threshold of 0.25 mm and requires independent confirmation of precipitation occurrence via a precipitation detector for at least 60% of the period. As Col de Porte does not have an R2 reference, 60 minute precipitation amounts were derived from a site-specific manually quality-controlled data set from a shielded gauge, which was adjusted for wind undercatch (Morin et al., 2012). The change in snow depth was determined from the difference in reported snow depth at the start and end of each 60 minute period. For this analysis, snow depths at CARE and Sodankylä are derived from the Jenoptik/Lufft SHM30 sensor measurements, while snow depths at Col de Porte are derived from a Dimetix FLS-CH 10 sensor. A disdrometer is used at all three sites to enable the refinement of relationships by precipitation type.

4.3.2 Results

The following results for the three sites indicate how the relationship between the hourly reported snow depth and precipitation amount is impacted by the site conditions. The regression statistics are summarized for each site in Table 4.35. Figure 4.193 shows the comparison between measured reference gauge snowfall and the corresponding change in snow depth for CARE. Comparing 165 hourly snowfall events over the two seasons, a significant amount of scatter can be seen in the relationship, yielding an R^2 value of 0.28. Figure 4.194 shows the same comparison for 442 hourly snowfall events at Sodankylä over the 2013/14 seasons. The R^2 for this site is higher than for CARE at 0.58, with less scatter in the data apparent in Figure 4.194. The 60-minute results from Col de Porte are illustrated in Figure 4.195 for the 2014/15 season. As with Sodankylä and CARE, scatter is significant and the linear regression statistical indicators indicate weak linear correlation.

Table 4.35. Summary of the regression statistics for the relationships between 60 minutes precipitation accumulation (in mm) as measured by the reference gauge and change in snow depth (in cm) as measured by the automated snow-depth sensor.

Site	Slope	Intercept	R^2	N
CARE	0.31	0.44	0.28	165
Sodankylä	0.57	0.26	0.58	442
Col de Porte	0.88	0.49	0.59	522

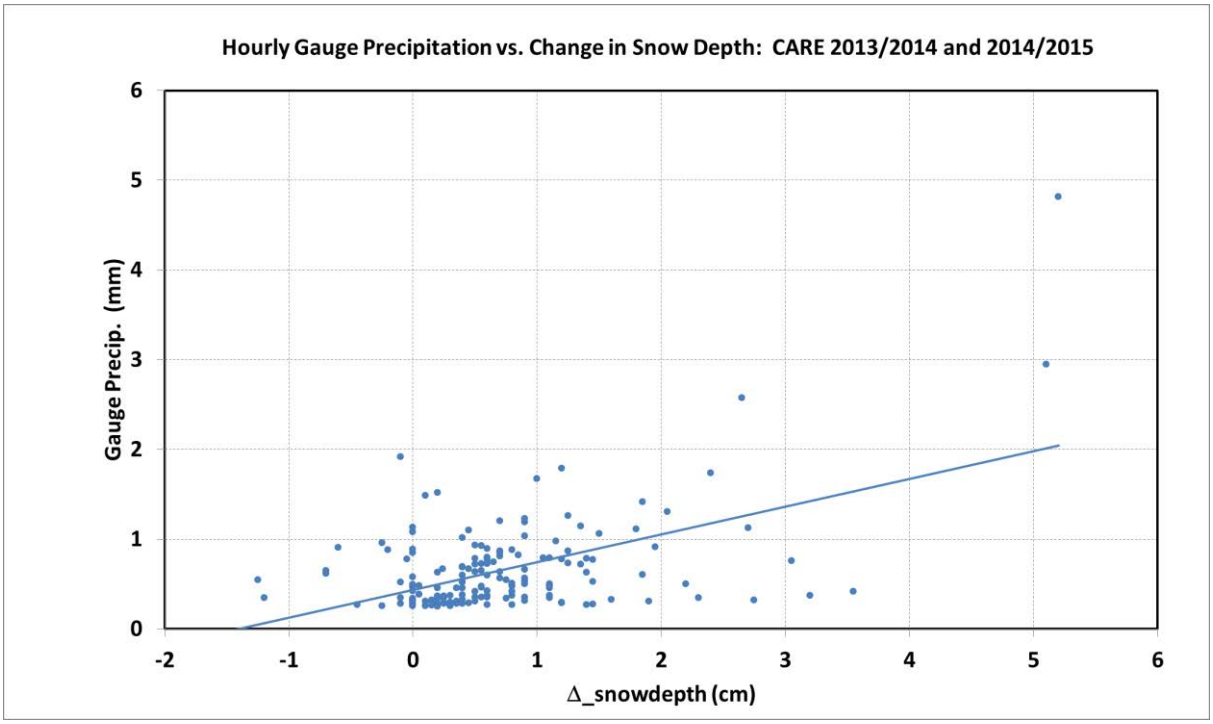


Figure 4.193. Relationship between hourly precipitation as measured by the reference (R2) gauge and the change in snow depth at CARE for the 2013/14 and 2014/15 winter seasons.

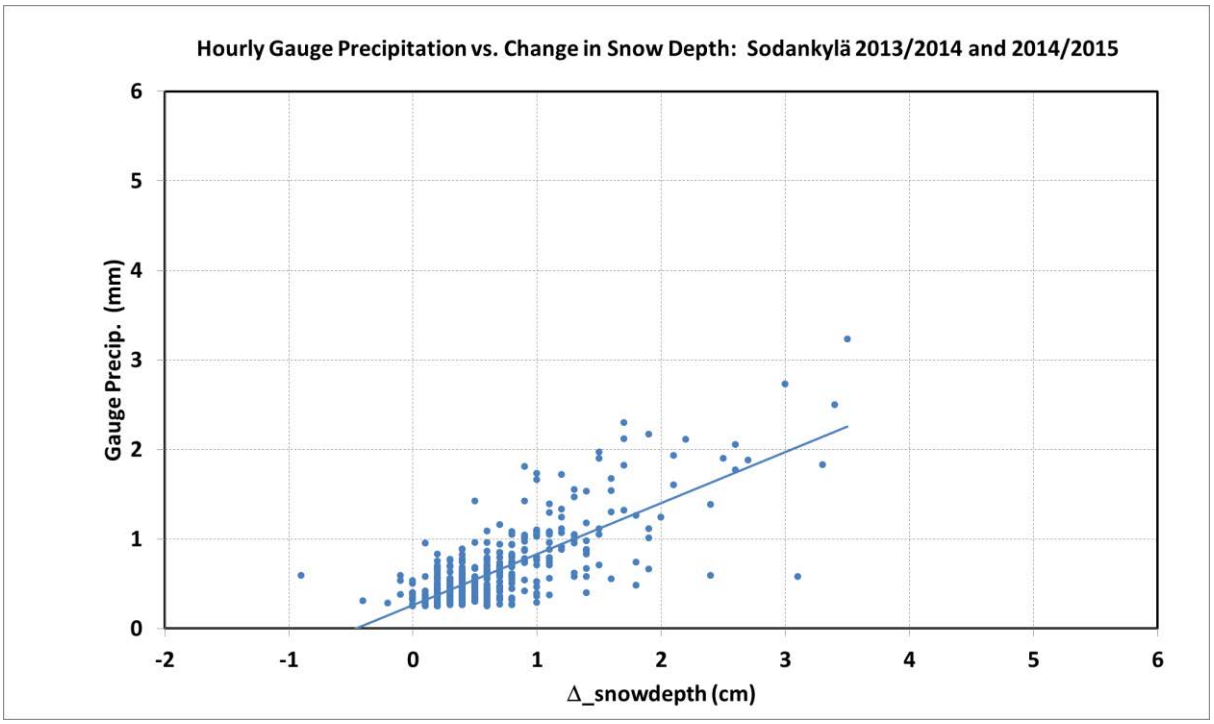


Figure 4.194. Relationship between hourly precipitation as measured by the reference (R2) gauge and the change in snow depth at Sodankylä for the 2013/14 and 2014/15 winter seasons.

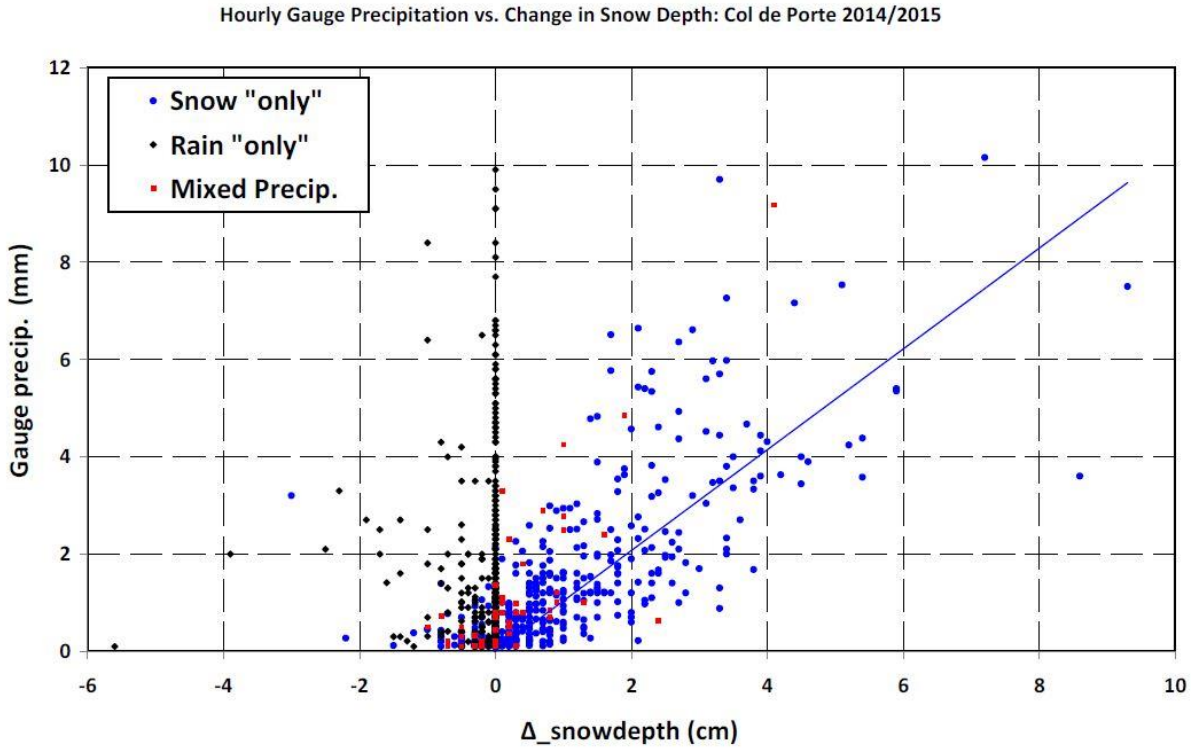


Figure 4.195. Relationship between hourly precipitation as measured by a wind-adjusted site gauge and the change in snow depth at Col de Porte for winter 2014/15. The hourly snow type discrimination at Col de Porte is determined based on temperature thresholds and ancillary information such as relative humidity, snow albedo, snow-depth fluctuations (Morin et al., 2012).

The Col de Porte dataset makes it possible to extend the analysis to time periods covering entire precipitation events in a continuous way. In this case, a precipitation event is defined as continual precipitation with an accumulation of at least 5 mm and delineated by a break in precipitation of at least 5 hours. For the 2014/15 season, the result is 13 snowfall events varying in length between 7 and 58 hours. The scatter plot is shown in Figure 4.196. By amalgamating precipitation into longer periods, the scatter in the relationship (taking into account only snow precipitation events) is reduced substantially, such that the R^2 value increases to 0.93 with 13 events considered.

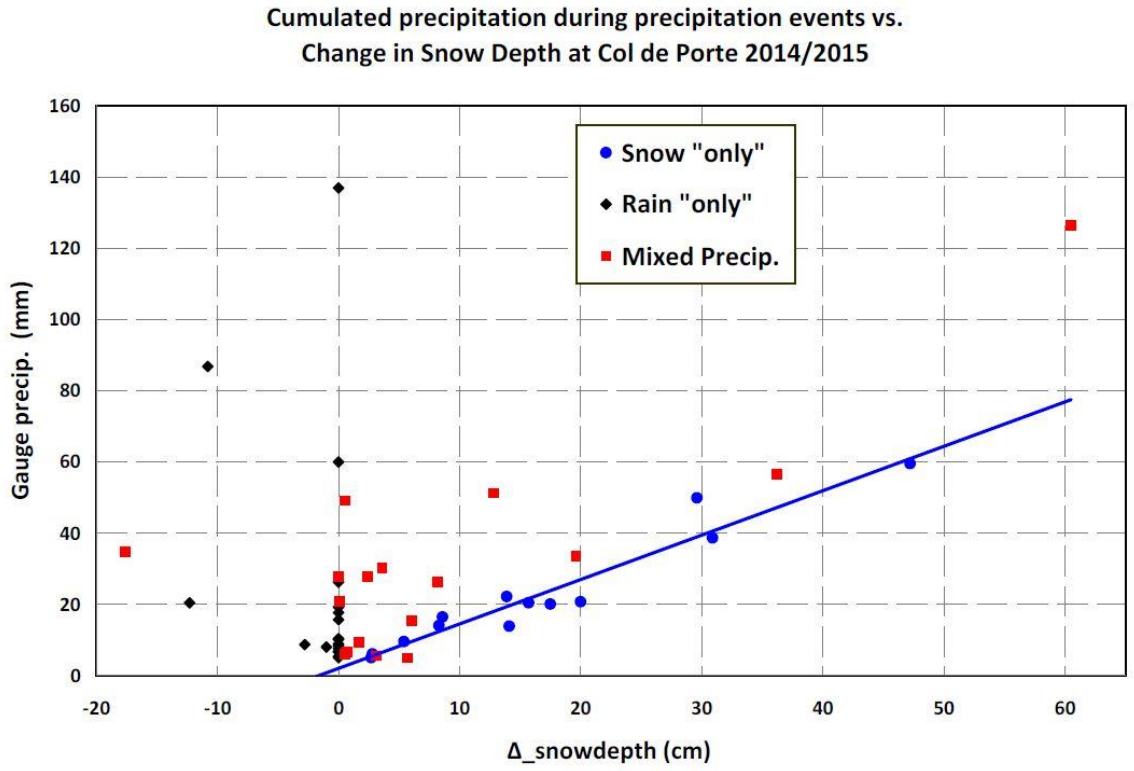


Figure 4.196. Relationship between event-based precipitation amounts (with durations between 7 and 58 hours) and the change in snow depth at Col de Porte, 2014/15 winter season.

4.3.3 Summary and conclusions

The large amount of scatter in the hourly relationships between gauge measured precipitation and the corresponding change in snow depth was not unexpected given the complexity and dynamic nature of the relationship between incoming precipitation, snow depth, and snow density. The data gathered during the WMO-SPICE experiment at sites in different climate conditions shows that there is no unambiguous relationship between snow-depth variation and precipitation amount at the hourly time scale.

Data from Col de Porte suggests that the relationship between gauge snow-only precipitation and change in snow depth becomes tighter when aggregated to longer time periods. Although this analysis shows promise, it needs to be explored further and at other SPICE sites. For the most part, these results suggest that it is generally not possible to use variations of snow depth to infer precipitation amounts with a level of accuracy equivalent to, or better than, precipitation gauge measurements.

4.4 Use of visibility to estimate snowfall intensity

Author: Roy Rasmussen

The standard relationship between snowfall intensity and visibility used by many national weather services (1/4 mile or less visibility corresponds to heavy snowfall intensity, between 5/16 and 5/8 mile corresponds to moderate intensity, and greater than 5/8 mile corresponds to light intensity) does not always provide the correct indication of actual liquid equivalent snowfall rate because of the variations in snow type and the differences in the nature of visibility targets during day and night (Rasmussen et al. 1999). This false indication may have been a factor in previous ground-deicing accidents in which light snow intensity was reported based on visibility, when in fact the actual measured liquid equivalent snowfall rate was moderate to heavy.

The poor relationship between snowfall rate and visibility was determined by an observational and theoretical study by Rasmussen et al. (1999). The observational data were collected at the Marshall Field site in Boulder, Colorado, USA, and included simultaneous liquid equivalent snowfall rates from a weighing gauge in a DFAR, crystal types, and both automated and manual visibility measurements. Theoretical relationships between liquid equivalent snowfall rate and visibility were derived for 27 crystal types, and for ‘‘dry’’ and ‘‘wet’’ aggregated snowflakes. Both the observations and theory showed that the relationship between liquid equivalent snowfall rate and visibility depends on the crystal type, the degree of riming, the degree of aggregation, and the degree of wetness of the crystals, leading to a large variation in the relationship between visibility and snowfall rate.

Typical variations in visibility for a given liquid equivalent snowfall rate ranged from a factor of three to a factor of 10, depending on the storm. This relationship was shown to have a wide degree of scatter from storm to storm and also during a given storm. The main cause for this scatter was the large variation in cross-sectional area to mass ratio and terminal velocity for natural snow particles.

It was also shown that the visibility at night can be over a factor of two greater than the visibility during the day for the same atmospheric extinction coefficient. Since snowfall intensity is defined by the U.S. National Weather Service using visibility, this day/night difference in visibility results in a change in snowfall intensity category caused by only whether it is day or night. For instance, a moderate snowfall intensity during the day will change to a light snowfall intensity at night, and a heavy snowfall intensity during the day will change to a moderate snowfall intensity at night, for the same atmospheric extinction coefficient.

Based on the above, it is not recommended that the liquid equivalent snowfall rate be estimated using visibility.

5. REFERENCES

Adams, M. S., E. Gleirscher, T. Gigele and R. Fromm, 2013: Automated terrestrial laser scanner measurements of small-scale snow avalanches, Proceedings of the International Snow Science Workshop Grenoble – Chamonix Mont-Blanc – 7-11 October 2013.

Andreas, E., 2002: Parameterizing scalar transfer over snow and ice: a review. *Journal of Hydrometeorology*., 3, 417–432.

ASTM D4430-00, 2015: *Standard Practice for Determining the Operational Comparability of Meteorological Measurements*, ASTM International, West Conshohocken, PA, http://dx.doi.org/10.1520/D4430-00R15

Atmospheric Measurement Techniques, Special Issue 78: The World Meteorological Organization Solid Precipitation InterComparison Experiment (WMO-SPICE) and its applications (AMT/ESSD/HESS/TC inter-journal SI). Editor(s): M. E. Earle, S. Morin, R. M. Rasmussen, M. A. Wolff, and D. Yang.

Baker, B., R. Buckner, W. Collins, and M. Phillips, 2005: Calculation of USCRN precipitation from Geonor weighing precipitation gauge. NOAA Technical Note NCDC No. USCRN-05-1, Asheville, NC.Beaumont, R.T.: Mt. Hood pressure pillow snow gage, J. Appl. Meteor., 4, 626–631, 2005.

Beaumont, R.T., 1965: Mt. Hood pressure pillow snow gage, *Journal of Applied Meteorology and Climatology,* 4, 626–631, 1965.

Bloemink, H.I. and E. Lanzinger, 2005: *Precipitation type from the Thies disdrometer.* Paper presented at the WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (TECO-2005) (Bucharest, Romania). Instruments and Observing Methods Report No. 82, WMO-TD No. 1265, Geneva.

Boniface, K., J. J. Braun, J. L. McCreight, and F. G. Nievinski, 2014: Comparison of snow data assimilation system with GPS reflectometry snow depth in the western United States. *Journal of Hydrological Processes*, 29, 2425–2437, doi:10.1002/hyp.10346.

Buisán, S., J. Collado, and J. Alastrue, 2016: The impact of a windshield in a tipping bucket rain gauge on the reduction of losses in precipitation measurements during snowfall events. *EGU General Assembly Conference Abstracts*, 18, S. 7055.

Buisán, S., M. Earle, J. Collado, J. Kochendorfer, J. Alastrué, M. Wolff, et al. 2017: Assessment of snowfall accumulation underestimation by tipping bucket gauges in the Spanish operational network. *Atmospheric Measurement Techniques*, 10(3), 1079-1091.

C-SPICE – Precipitation Gauge Heating Summary, May 17, 2013.

CARE Commissioning Report Revised - Appendix A: Proof of Performance (POP) Forms, 2013 October 29.

Colli, M., 2013: Denoising filters for snowfall rate measurements. An evaluation of the filters accuracy basing on artificial step-shaped snowfall rate data set. Report for the WMO SPICE Data Analysis Team meeting, Friday 31th May, 2013.

Colli, M., M., S. Landolt, E. Vuerich, L.G. Lanza and R.M. Rasmussen, 2013: Snow Machine testing of SPICE snow gauges. Report for the IV International Organizing Committee of WMO SPICE, Davos, Switzerland, 17 - 21 June 2013.

Colli, M., 2014: Assessing the accuracy of precipitation gauges: A CFD approach to model wind induced errors.: PhD Thesis, University of Genoa, 206 pp.

Colli, M., R. M. Rasmussen, L. Lanza, and J. Theriault, 2015: An improved trajectory model to evaluate the collection performance of snow gauges, J. Appl. Met. and Clim., 54, 1826 – 1836.

Colli, M., L. Lanza, R.M. Rasmussen, and J. Theriault, 2016: The collection efficiency of shielded and unshielded precipitation gauges. Part I: CFD airflow modeling, J. Hydromet., 17, 231-243.

Colli, M., L. Lanza, R. M. Rasmussen, and J. Theriault, 2016: The collection efficiency of shielded and unshielded precipitation gauges. Part II: Modeling particle trajectories, J. Hydromet., 17, 245 – 255.

de Haij, Marijn, 2011: Field test of the Jenoptik SHM30 laser snow-depth sensor. *Technical report TR-325*, Koninklijk Nederlands Meteorologisch Instituut.

Desilets, D., Zreda, M., and Ferré, T. P. A., 2010: Nature's neutron probe: Land surface hydrology at an elusive scale with cosmic rays. *Journal of Water Resources Research*, 46, 1-7

Duchon, C.E. 2004. Field measurements of temperature sensitivity in Geonor vibrating-wire transducers. *Bulletin of the American Meteorological Society*, pp. 5935-5940.

Duchon, C.E. 2008. Using vibrating-wire technology for precipitation measurements. In S.C. Michaelides (Ed.), *Precipitation: Advances in measurement, estimation and prediction* (pp. 33-58). Berlin: Springer.

Earle, M and P. Raczynski, 2013 Feb.6: Manual observation procedures at CARE, Observing Systems and Engineering, Meteorological Service of Canada, V4.

Egli, L., T. Jonas, and R. Meister, 2009: Comparison of different automatic methods for estimating snow water equivalent, *Cold Regions Science and Technology*, 57(2–3), 107–115, doi:10.1016/j.coldregions.2009.02.008.

Essery, R., S. Morin, Y. Lejeune, and C.B. Ménard, 2013: A comparison of 1701 snow models using observations from an alpine site. *Journal of Advances in Water Resources*, 55, 131-148.

Farnes, P.F., B.E. Goodison, N.R Peterson and R.P. Richards, 1983: *Metrication Of Manual Snow Sampling Equipment.* Final Report Western Snow Conference, Spokane, Washington 106 p.

Fischer, A.P., 2011: The measurement factors in estimating snowfall derived from snow cover surfaces using acoustic snow depth sensors. JAMC, 50, 681-699.

Fierz, C., Armstrong, R.L., Durand, Y., Etchevers, P., Greene, E., McClung, D.M., Nishimura, K., Satyawali, P.K. and Sokratov, S.A., 2009: *The International Classification for Seasonal Snow on the Ground*. IHP-VII Technical Documents in Hydrology N°83, IACS Contribution N°1, UNESCO-IHP, Paris.

Filhol, S., and M. Sturm, 2015: Snow bedforms: A review, new data, and a formation model, *Journal of Geophysical Research. Earth Surface*, 120, 1645–1669, doi:10.1002/2015JF003529.

Fischer, A.P., 2011: The measurement factors in estimating snowfall derived from snow cover surfaces using acoustic snow-depth sensors. *Journal of Applied Meteorology and Climatology*, 50, 681-699.

Geonor T-200B Series Precipitation Gauge Instruction Manual, 600mm, 1000mm, and 1500mm capacity options, Campbell Scientific (Canada) Corp. September 2011, Rev. 10.7.

Golubev, V.S., 1979: An investigation of the double shielding fences effect on the Tretyakov gauge measurements (in Russian), Trans. State Hydrol. Inst, Leningrad (St. Petersburg), vol. 258, p. 91-101.

Golubev, V.S., 1985: *On the problem of actual precipitation measurement at the observation site.* Proceedfings of the International Workshop on the Correction of Precipitation Measurements, WMO/TD No. 104, WMO, Geneva, pp. 60-64.

Golubev, V.S., 1986: *On the problem of standard condition for precipitation gauge installation*. Proceedings of the International Workshop on the Correction of Precipitation Measurements, WMO/TD No. 104, Geneva, pp. 57-59.

Golubev, V.S., 1989: *Assessment of accuracy characteristics of the reference precipitation gauge with a double-fence shelter*. Final report of the Fourth Session of the International Organizing Committee for the WMO Solid Precipitation Measurement Intercomparison, St. Moritz, Switzerland, WMO, Geneva, 22-29.

Goodison et al, 1981: Measurement and Data Analysis. In D.M. Gray and D.H. Hale (Eds.), *Handbook of Snow: Principles, Processes, Management & Use*, Toronto: Pergamon Press.

Goodison B., Louie P.Y.T. and Yang D., 1998: WMO Solid Precipitation Measurement Intercomparison, Final Report, WMO IOM Report No. 67, WMO/TD – No. 872.

Gottardi, F., P. Carrier, E. Paquet, M.T. Laval, J. Gailhard and R. Garçon, 2013: A decade of snow water equivalent monitoring in the French mountain ranges, Proceedings of the International Snow Science Workshop Grenoble, – Chamonix Mont-Blanc – 7-11 October, 2013.

Griesel, S., M. Theel, K. Schubotz, and E. Lanzinger, 2012: Field Intercomparison and laboratory tests of precipitation detectors; P3-12, TECO, Brussels.

Grunewald, T., Stotter, J., Pomeroy, J. W., Dadic, R., Moreno Banos, I., Marturia, J., Spross, M., Hopkinson, C., Burlando, P., and Lehning, M., 2013: Statistical modeling of the snow-depth distribution in open alpine terrain. *Journal of Hydrology and Earth System Sciences*, 17, 3005–3021, doi:10.5194/hess-17-3005-2013.

Haji, M. J. de, 2011: *Field test of the Jenoptik SHM30 laser snow-depth sensor*. KNMI Technical Report No. 325, KNMI, De, Bilt, The Netherlands.

Hall, M.E., E. May, 2004: *Inlet Heater for USCRN Weighing Precipitation Gauge*. NOAA Technical Note NCDC No. USCRN-04-01).

Hamon, W.R., 1973: Computing Actual Precipitation. In: Distribution of Precipitation in Mountainous Areas. Vol. 1, p.159-174. World Meteorological Organization No. 326.

Hanson, C.L., F.B. Pierson, and G.L. Johnson, 2004 : Dual-gauge system for measuring precipitation: Historical development and use. *Journal of Hydrologic Engineering*, 9(5), 350-359.

Huwald, H., C.W. Higgins, M.O., Boldi, E. Bou-Zeid, M. Lehning, and M.B. Parlange, 2009: Albedo effect on radiative errors in air temperature measurements. *Journal of Water Resources Research*, 45, doi:10.1029/2008WR007600.

Ishizaka, M., M. Motoyoshi, S. Nakai, T. Shiina, T. Kumakura and K. Muramoto, 2013: A new method for identifying the main type of solid hydrometeor contributing to snowfall from measured size-fallspeed relationship, Journal of the Meteorological Society of Japan, 91, 747-762.

Jacobson, M.D., 2010: Inferring Snow Water Equivalent for a Snow-Covered Ground Reflector Using GPS Multipath Signals, Remote Sensing, 2(10), 2426-2441, doi:10.3390/rs2102426.

Jin, S., Najibi, N., 2014: Sensing snow height and surface temperature variations in Greenland from GPS reflected signals. *Journal of Advances in Space Research*, 53(11), 1623-1633, doi:10.1016/j.asr.2014.03.005.

Joint Committee for Guides in Metrology (JCGM), 2008: International vocabulary of metrology – basic and general concepts and associated terms, JCGM 200.

Kaasalainen, S., H. Kaartinen, A. Kukko, K. Anttila, and A. Krooks, 2011: Brief communication "Application of mobile laser scanning in snow cover profiling", *The Cryosphere*, 5, 135-138, doi:10.5194/tc-5-135-2011.

Kinar, N. J., and J. W. Pomeroy, 2015: SAS2: the system for acoustic sensing of snow. *Journal of Hydrological Processes*, 29, 4032–4050. doi: 10.1002/hyp.10535.

Kinar, N. J., 2015b: Measurement of the physical properties of the snowpack, *Journal of Reviews of Geophysics*, 53, doi:10.1002/2015RG000481.

Koch, F., M. Prasch, L. Schmid, J. Schweizer, and W. Mauser, 2014: Measuring Snow Liquid Water Content with Low-Cost GPS Receivers. *Sensors*, 20975-20999, doi: 10.3390/s141120975.

Kochendorfer, J., R. Nitu, M. Wolff, E. Mekis, R. Rasmussen, B. Baker, et al., 2017a: Analysis of single-Alter-shielded and unshielded measurements of mixed and solid precipitation from WMO-SPICE. *Journal of Hydrology and Earth System Sciences*, 21(7), 3525-3542.

Kochendorfer, J., Rasmussen, R., Wolff, M., Baker, B., Hall, M., Meyers, T., et al., 2017b: The quantification and correction of wind-induced precipitation measurement errors. *Journal of Hydrology and Earth System Sciences*, 21(4), 1973-1989.

Kochendorfer, J., R. Nitu, M. Wolff, E. Mekis, R. Rasmussen, B. Baker, et al., 2018: Testing and Development of Transfer Functions for Weighing Precipitation Gauges in WMO-SPICE. *Journal of Hydrology and Earth System Sciences Discussions,* 22, 1437–1452.

Kodama, M., K. Nakai, S. Kawasaki, and M. Wada, M., 1979: An application of cosmic-ray neutron measurements to the determination of the snow water equivalent. *Journal of Hydrology,* 41, 85-92, doi:10.1016/0022-1694(79)90107-0.

Lanza, L., M. Leroy, C. Alexandropoulos, L. Stagi, and W. Wauben, 2005: *Laboratory intercomparison of rainfall intensity gauges*. World Meteorological Organization, Geneva.

Lanzinger, E., M. Theel, and H. Windolph, 2006: *Rainfall amount and intensity measured by the THIES laser precipitation monitor*. Paper presented at the WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (TECO-2006) (Geneva, Switzerland), Instruments and Observing Methods Report No. 94, WMO-TD No.1354, Geneva.

Larson, K. M., E. D. Gutmann, V. U. Zavorotny, J. J. Braun, M. W. Williams, and F. G. Nievinski, 2009: Can we measure snow depth with GPS receivers? *Journal of Geophysical Research Letters*, 36(17), L17502, doi:10.1029/2009GL039430.

Lee, J.-E., Lee, G. W., Earle, M., and Nitu, R., 2015: Uncertainty analysis for evaluating the accuracy of snow-depth measurements, *Journal of Hydrology and Earth System Sciences Discussions,* https://doi.org/10.5194/hessd-12-4157-2015.

Leeper, R., and J. Kochendorfer, 2015a: Evaporation from weighing precipitation gauges: impacts on automated gauge measurements and quality assurance methods. *Journal of Atmospheric Measurement Techniques*, 8(6), 2291-2300.

Leeper, R., M. Palecki, and E. Davis, 2015b: Methods to Calculate Precipitation from Weighing-Bucket Gauges with Redundant Depth Measurements. *Journal of Atmospheric and Oceanic Technology*, 32(6), 1179-1190.

Leeper, R., J. Rennie, and M. Palecki, 2015c: Observational Perspectives from US Climate Reference Network (USCRN) and Cooperative Observer Program (COOP) Network: Temperature and Precipitation Comparison. *Journal of Atmospheric and Oceanic Technology*, 32(4), 703-721.

Löffler-Mang, M. and J. Joss, 2000: An Optical Disdrometer for Measuring Size and Velocity of Hydrometeors. *Journal of Atmospheric and Oceanic Technology*, Vol. 17, p. 130-139.

Marshall, H.-P., and G. Koh, 2008: FMCW radars for snow research. *Journal of Cold Regions Science and Technology*, 52(2), 118–131, doi:10.1016/j.coldregions.2007.04.008.

Mayo, L. R., 1971: Self-Mixing Antifreeze Solution for Precipitation Gages. *Journal of Applied Meteorology*, 11(2), pp. 400-404.

Mekis E. and W.D. Hogg, 1999: Rehabilitation and analysis of Canadian daily precipitation time series. *Journal of Atmosphere-Ocean*, 37, 53-85.

Mekis E., 2005: J3.7 *Adjustments for trace measurements in Canada*. 15th Conference on Applied Climatology, Savannah, Georgia, USA, 20-24 June 2005.

Mekis E., and L.A. Vincent, 2011: An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. *Journal of Atmosphere-Ocean*, 49(2), 163-177.

Metcalfe, J.R. and B.E. Goodison, 1992: Automation of winter precipitation measurements: The Canadian experience. *Proceedings of the WMO Technical Conference on Instruments and Methods of Observation*, pp. 11-15.

Metcalfe, J.R., S. Ishida and B.E. Goodison, 1994: *A corrected precipitation archive for the Northwest Territories. Environment Canada - Mackenzie Basin Impact Study*, Interim Report #2, 110-117.

Meteorological Division, Department of Transport – Canada, 1947: Manual of standard procedures and practices for weather observing and reporting, short title MANOBS – Effective 1st January, 1947.

Meteorological Division, Department of Transport – Canada, 1949: Manual of standard procedures and practices for weather observing and reporting, MANOBS 2nd Edition – Effective 1st January, 1949.

Meteorological Division, Department of Transport – Canada, 1951: Manual of standard procedures and practices for weather observing and reporting, MANOBS 3rd Edition – Effective 1st October, 1951.

Mohamed, R. and J. Hoover, 2013: CARE Geonor T200B Precipitation Gauge Field Verification, July 15-26, 2013.

Morin, S., Y. Lejeune, B. Lesaffre, J.M. Panel, D. Poncet, P. David, M. Sudul, 2012: An 18-yr long (1993–2011) snow and meteorological dataset from a mid-altitude mountain site (Col de Porte, France, 1325 m alt.) for driving and evaluating snowpack models, *Journal of Earth System Science Data,* 4, 13–21, www.earth-syst-sci-data.net/4/13/2012/doi:10.5194/essd-4-13-2012.

Mott, R., M. Schirmer, M. Bavay, T. Grünewald, and M. Lehning, 2010: Understanding snow-transport processes shaping the mountain snow-cover, *The Cryosphere*, 4, 545-559, doi:10.5194/tc-4-545-2010.

National Aeronautics and Space Administration, 2010: *Earth Observing System Data and Information System (EOSDIS): Data Processing Levels*. Retrieved from https://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products/

Nemeth, K., 2008: *OTT Pluvio²: Weighing Precipitation Gauge and Advances in Precipitation Measurement Technology*. Proceedings of the CIMO TECO Conference held at St-Petersburg, Russian Federation, 27-29 November 2008.

Niang, M., M. Bernier, Y. Gauthier, G. Fortin, E. Van Bochove, M. Stacheder, and A. Brandelik, 2003: On the validation of snow densities derived from SNOWPOWER probes in a temperate snow cover in eastern Canada: First results. *Proceedings of Eastern Snow Conference at Sherbrooke*, Québec, Canada, 9-11 June 2003.

Nolin, A. W., 2010: Recent advances in remote sensing of seasonal snow. *Journal of Glaciology,* 56(200), 1141-1150.

Ohaus Explorer Analytical and Precision Balance (S/N 8033021016), Calibration Data Report by M & L Testing Equipment (1995) Inc. on January 17, 2012 and July 22, 2013.

Orlov, G.I., 1946: On the question of the measurement of solid precipitation (in Russian), Trans. Main. Geophys. Observ. Leningrad (St. Petersburg), Vol. 81, ser. 1, p. 5-18.

OTT Hydrometrie: Operating Instructions OTT Pluvio^2 precipitation gauge - OTT Hydromet. http://www.ott.com.

Pan, X., D. Yang, Y. Li, A. Barr, W. Helgason, M. Hayashi, P. Marsh, J. Pomeroy, and R.J. Janowicz, 2016: Bias corrections of precipitation measurements across experimental sites in different ecoclimatic regions of western Canada, *The Cryosphere*, 10, 2347-2360, https://doi.org/10.5194/tc-10-2347-2016.

Paquet, E. and Laval, M.T., 2006: Operation feedback and prospects of EDF Cosmic-Ray Snow Sensors. *La Houille Blanche,* 2, 113-119, doi:10.1051/lhb:200602015.

Patterson, J. 1930 : Instructions to Observers in the Meteorological Service of Canada. Ottawa, ON: Department of Marine.

Peel, M.C., B.L. Finlayson, and T.A. McMahon, 2007: Updated world map of the Köppen-Geiger climate classification, *Hydrology and Earth System Sciences*, 11, 1633-1644, doi: 10.5194/hess-11-1633-2007.

Picard, G., L. Arnaud, J.M. Panel and S. Morin: Design of a scanning laser meter for monitoring the spatio-temporal evolution of snow depth and its application in the Alps and in Antarctica. *The Cryosphere*, in press.

Pikounis, M., E. Baltas, and M. Mimikou, 2002: Temporal sampling and bucket volume effects on tipping bucket measuring accuracy. *Journal of Meteorology*, 27(273), 335-348.

Rasmussen, R.M., J. Vivekanandan, J. Cole, B. Myers and C. Masters, 1999: The estimation of snowfall rate using visibility. J. Appl. Meteor., 38(10), 1542-1563.

Rasmussen, R.M., M. Dixon, F. Hage, J. Cole, C. Wade, J. Tuttle, S. McGettigan, T. Carty, L. Stevenson, W. Fellner, S. Knight, E. Karplus, and N. Rehak, 2001: Weather Support to Deicing Decision Making (WSDDM): A winter weather nowcasting system. Bull. Amer. Meteor. Soc., 82, 579 - 595.

Rasmussen, R.M., J. Hallett, R. Purcell, S. Landolt and J. Cole, 2011: The Hotplate Precipitation Gauge. J. Appl. Met.. and Clim., 28, 148 – 163.

Rasmussen, R.M., B. Baker, J. Kochendorfer, T. Myers, S. Landolt, A. Fisher, J. Black, J. Theriault, P. Kucera, D.J. Gochis, C. Smith, R. Nitu, M. Hall, S. Cristanelli and E. Gutmann, 2012: How Well Are We Measuring Snow? The NOAA/FAA/NCAR Winter Precipitation Test Bed: *Bulletin of the American Meteorological Society,* 93, 811-829.

Reverdin, A., S. Buisán, Y.-A. Roulet, J.L. Collado, and J. Alastrue, 2016: Intercomparison of snowfall measurements using disdrometers in two mountainous environments: Weissfluhjoch (Switzerland) and Formigal -Sarrios (Spain), *Proceedings of the CIMO-TECO Conference, Madrid*, Spain, 8-11 October 2016.

Rodriguez-Alvarez, N., Aguasca, A., Valencia, E., Bosch-Lluis, X. 2012: Snow thickness monitoring using GNSS measurements, *Journal of IEEE Geoscience and Remote Sensing Letters*, 9(6), 1109-1113, doi: 10.1109/LGRS.2012.2190379.

Ryan, W. A. and N. J. Doesken, 2007: *Ultrasonic snow-depth sensors for National Weather Service Snow Measurements in the U.S.: Evaluation of Operational Readiness.* 14th Symposium on Meteorological Observation and Instrumentation, San Antonio, Texas.

Ryan, W. A., N. J. Doesken, and S. R. Fassnacht, 2008: Evaluation of Ultrasonic Snow-depth Sensors for U.S. Snow Measurements. *Journal of Atmospheric and Oceanic Technology,* 25, 667-684.

Savina, M., B. Schäppi, P. Molnar, P. Burlando, and B. Sevruk, B., 2012: Comparison of a tipping-bucket and electronic weighing precipitation gage for snowfall. *Atmospheric Research*, 103, 45-51.

Schmid, L., F. Koch, A. Heilig, M. Prasch, O. Eisen, W. Mauser, and J. Schweizer, 2015: A novel sensor combination (upGPR-GPS) to continuously and nondestructively derive snow cover properties. *Journal of Geophysical Research Letters,* 42, 3397-3405, doi:10.1002/2015GL063732.

Serreze, M.C., M.P. Clark, R.L. Armstrong, D.A. McGinnis, and R.S. Pulwarty, 1999: Characteristics of the western United States snowpack from snowpack telemetry(SNOTEL) data. *Journal of Water Resources Research*, 35(7), 2145-2160.

Sevruk, B., 1994. Spatial and temporal inhomogeneity of global precipitation data. In *Global Precipitations and Climate Change*, pp. 219-230, Berlin, Heidelberg: Springer.

Sheppard, B., P. Joe, 2000: Automated Precipitation Detection and Typing in Winter: A Two-Year Study. *Journal of Atmospheric and Oceanic Technology*, Vol. 17: 1493–1507.

Sigouin, M. J. P. and B.C. Si, 2016: Calibration of a non-invasive cosmic-ray probe for wide area snow water equivalent measurement, *The Cryosphere*, 10, 1181-1190, doi:10.5194/tc-10-1181-2016.

Simone Griesel, Manfred Theel, Karsten Schubotz, and Eckhard Lanzinger, 2012: Field Intercomparison and laboratory tests of precipitation detectors; *P3-12, Proceedings of the CIMO-TECO Conference, Brussels*, Belgium, 8-16 October 2012.

Smith, C.D., 2009: The relationships between snowfall catch efficiency and wind speed for the Geonor T-200B precipitation gauge utilizing various wind shield configurations. *Proceedings of 77th Western Snow Conference*, Canmore, 115-120.

Smith, C.D., and D. Yang, 2010*:* An assessment of the Geonor T-200B used with a large octagonal double fence wind shield as an automated reference for the gauge measurement of solid precipitation. *90th AMS Annual Meeting/15th SMOI*, Atlanta, https://ams.confex.com/ams/pdfpapers/158669.pdf

Smith, C.D., A. Kontu, R. Laffin, and J.W. Pomeroy, 2017: An assessment of two automated snow water equivalent instruments during the WMO Solid Precipitation Intercomparison Experiment. *The Cryosphere*, 11, 101-116, https://doi.org/10.5194/tc-11-101-2017.

Söderström, P., S. Knutsson, T. Olofsson, T. Edeskär, S. Mårtensson, and A. Zeinali, 2013: Snow grooming using machine guidance for piste management processes: Case study: Ormberget ski piste. Luleå: Luleå tekniska universitet, 20 pp.

Thériault, J. M., R. Rasmussen, K. Ikeda, and S. Landolt, 2012: Dependence of snow gauge collection efficiency on snowflake characteristics. *Journal of Applied Meteorology and Climatology,* 51, 745-762.

Thériault, J. M., R. Rasmussen, E. Petro , M. Colli, and L. Lanza, 2015: Impact of wind direction, wind speed and particle characteristics on the collection efficiency of the Double Fence Intercomparison, *J. Appl. Met. and Clim.*, 54, 1918 – 1930.

Thériault, J. M., R. Rasmussen, M. Colli, and L. Lanza, 2016: Examination of the Catch Efficiency of Snow Gauges based on the Observed Characteristics of Snow, *18th Symposium on Meteorological Observations and Instrumentation*, AMS Annual Meeting, New Orleans.

Thom, A. S., 1975: Momentum, mass and heat exchange of plant communities, Vegetation and the Atmosphere, Vol. 1, edited by: Monteith, J. L., Academic Press.

Tokay, A., D.B. Wolff and W. A. Petersen, 2014: Evaluation of the New Version of the Laser-Optical Disdrometer, OTT Parsivel^2. *Journal of Atmospheric and Oceanic Technology*, Vol. 31, p. 1276-1288.

U.S. National Weather Service, 1989: *Cooperative station observations*. NWS Observing Handbook 2, 83 pp. [Available from NWS/NOAA, 1325 East-West Highway, Silver Spring, MD 20910.]

Vionnet, V., E. Martin, V. Masson, G. Guyomarc'h, F. Naaim-Bouvet, A. Prokop, Y. Durand, and C. Lac, 2014: Simulation of wind-induced snow transport and sublimation in alpine terrain using a fully coupled snowpack/atmosphere model. *Cryosphere*, 8, 395-415, doi:10.5194/tc-8-395-2014.

Vuerich, E., C. Monesi, L. Lanza, L. Stagi, and E. Lanzinger, 2009: WMO field intercomparison of rainfall intensity gauges. World Meteorological Organization, Geneva.

Wagner, W. and A. Pruß, 2002: The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use. *Journal of Physical and Chemical Reference Data*, 31(2), 387–535

Watson, S., C. Smith, M. Lassi, and J. Misfeldt, 2007: An evaluation of the effectiveness of the double-Alter wind shield for increasing the catch efficiency of the Geonor T-200B precipitation gauge. *Canadian Meteorological and Oceanographic Society Bulletin*, 36(5), 168-175.

Wild, H., 1885: Einfluss der Qualität und Aufstellung auf die Angaben der Regenmesser. *Kaiserliche Akademie der Wissenschaften*, St. Petersburg. Repertorium für Meteorologie, 9, p. 1-23.

Wolff, M. et al., 2010: A new test site for wind correction of precipitation measurements at a mountain plateau in southern Norway*, Proceedings of WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation*. 30th Aug - 1st Sep 2010, Helsinki, Finland.

Wolff, M., K. Isaksen, R. Braekkan, E. Alfnes, A. Petersen-Overleir, and E. Ruud, 2013: Measurements of wind-induced loss of solid precipitation: description of a Norwegian field study, *Hydrology Research*, 44, 35-43, 10.2166/nh.2012.166.

Wolff, M., K. Isaksen, A. Petersen-Øverleir, K. Ødemark, T. Reitan, and R. Brækkan, 2015: Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: results of a Norwegian field study, Hydrol. Earth Syst. Sci., 19, 951-967, https://doi.org/10.5194/hess-19-951-2015, 2015.

Woo, M. K., and P. Steer, 1979: Measurement of trace rainfall at a high Arctic site. Arctic, 32 (1), 80–84.

World Meteorological Organization, 1992: International Meteorological Vocabulary 2nd ed. World Meteorological Organization, (WMO-No.182), Geneva, Switzerland.

——, 1998: WMO Intercomparison of Present Weather Sensors/Systems: Final Report (Canada and France, 1993–1995) (M. Leroy, C. Bellevaux and J.P. Jacob). Instruments and Observing Methods Report No. 73, WMO/TD‑No. 887, Geneva.

——, 2008: *Guide to Hydrological Practices*, Volume 1, (WMO-No. 168), Geneva.

——, 2009: *WMO field intercomparison of rainfall intensity gauges*. Report No. 99. WMO/TD-No. 1504, Geneva.

——, 2010: *CIMO survey on national summaries of methods and instruments for solid precipitation measurement at automatic weather stations.* Report No. 102. WMO/TD-No. 1544, O), Geneva.

——, 2011: *WMO/CIMO Joint meeting of CIMO Expert Team on Instrument Intercomparisons and International Organizing Committee for the WMO Solid Precipitation Intercomparison Experiment, First Session, Final Report*, Geneva, Switzerland, 5-7 October 2011.

——, 2012: *International Organizing Committee for the WMO Solid Precipitation Intercomparison Experiment: Second Session, Final Report*, Boulder, United States, 11 - 15 June 2012.

——, 2012a: *International Organizing Committee for the WMO Solid Precipitation Intercomparison Experiment: Third Session, Final Report,* Brussels, Belgium, 15 October 2012.

——, 2013: *International Organizing Committee for the WMO Solid Precipitation Intercomparison Experiment: Fourth Session, Final Report,* Davos, Switzerland, 17 - 21 June 2013.

——, 2014: *International Organizing Committee for the WMO Solid Precipitation Intercomparison Experiment: Fifth Session, Final Report,* Sodankylä, Finland, 19 - 23 May 2014.

——, 2014: *Guide to Meteorological Instruments and Methods of Observation,* (WMO-No.8), 2014 edition, Geneva.

——, 2015: *International Organizing Committee for the WMO Solid Precipitation Intercomparison Experiment: Sixth Session, Final Report,* Zaragoza, Spain, 18 – 22 May 2015.

——, 2017: *WMO Space Programme: From data to products*. Retrieved from http://www.wmo.int/pages/prog/sat/dataandproducts_en.php.

Yang, D., J.R. Metcalfe, B.E. Goodison, and E. Mekis, 1993: *True Snowfall: An evaluation of the Double Fence Intercomparison Reference Gauge*. Proceedings 50th Eastern Snow Conference/61st Western Snow Conference, Quebec City, 105-111.

Yang, D., B.E. Goodison, J.R. Metcalfe, V.S. Golubev, E. Elomaa, T. Gunther, R. Bates, T. Pangburn, C. Hanson, D. Emerson, V. Copaciu, and J. Milkovic, 1995: Accuracy of Tretyakov precipitation gauge: result of WMO intercomparison. *Journal of Hydrological Processes* 9 (8), 877–895.

Yang, D., B.E. Goodison, C.S. Benson, and S. Ishida, 1998: Adjustment of daily precipitation at 10 climate stations in Alaska: application of World Meteorological Organization intercomparison results. *Journal of Water Resources Research* 34 (2), 241–256.

Yang D, B.E. Goodison, J.R. Metcalfe V.S. Golubev, R. Bates, T. Pangburn, and C.L. Hanson, 1998: Accuracy of NWS 8-inch standard non-recording precipitation gauge: result of WMO intercomparison. *Journal of Atmospheric and Oceanic Technology* 15(2): 54–68.

Yang, D., and A. Simonenko, 2013: Comparison of winter precipitation measurements by six Tretyakov gauges at the Valdai experimental site. *Journal of Atmosphere-Ocean*, p. 1-15.

Yang, D., 2014: Double Fence Intercomparison Reference (DFIR) vs. Bush Gauge for "true" snowfall measurement. *Journal of Hydrology*, Vol. 509, p. 94-100.

Zreda, M., D. Desilets, T.P.A. Ferré, and R.L. Scott, 2008: Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. *Journal of Geophysical Research Letters,* 35, 1-5.

Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., and Rosolem, R., 2012: COSMOS: the COsmic-ray Soil Moisture Observing System. *Journal of Hydrology and Earth System Sciences*, 16, 4079-4099, doi:10.5194/hess-16-4079-2012.

Zweifel, A. and B. Sevruk, 2002: *Comparative accuracy of solid precipitation measurement using heated recording gauges in the Alps*. WCRP Workshop on Determination of Solid Precipitation in Cold Climate Regions. Fairbanks, Alaska.

6. ANNEXES

1. Objectives and Deliverables
2. Questionnaire templates
   - *2.1. Questionnaire to manufacturers for submission of precipitation gauges*
   - *2.2. Questionnaire to manufacturers for submission of snow on the ground measurement systems*
   - *2.3. Questionnaire to site participants for site description*
3. Protocols templates
   - *3.1. SPICE Commissioning Protocol*
   - *3.2. SPICE Data Protocol*
4. Site description
   - *4.1. Australia, Guthega Dam*
   - *4.2. Canada, Bratt's Lake*
   - *4.3. Canada, CARE (Center for Atmospheric Research and Experiments)*
   - *4.4. Canada, Caribou Creek*
   - *4.5. Chile, Tapado*
   - *4.6. Finland, Sodankylä*
   - *4.7. France, Col de Porte*
   - *4.8. Italy, Forni Glacier*
   - *4.9. Japan, Joetsu*
   - *4.10. Japan, Rikubetsu*
   - *4.11. Korea (Republic of), Gochang*
   - *4.12. Nepal, AWS1 Pyramid*
   - *4.13. New Zealand, Mueller Hut*
   - *4.14. Norway, Haukeliseter*
   - *4.15. Poland, Hala Gasienicowa*
   - *4.16. Russian Federation, Valday*
   - *4.17. Russian Federation, Volga*
   - *4.18. Spain, Aramon-Formigal*
   - *4.19. Switzerland, Weissfluhjoch*
   - *4.20. USA, Marshall*
5. Sensors under test allocation
   - *5.1. Selection and inclusion of instruments*
   - *5.2. Sensors under test allocation lists (by instrument and by site)*
6. Instrument performance reports
   - *6.1. Weighing gauges*
     - *6.1.1. Belfort AEPG 600*
     - *6.1.2. Geonor T-200B3 600*
     - *6.1.3. Geonor T-200B3MD 1500*
     - *6.1.4. Meteoservis MRW500*
     - *6.1.5. MPS TRwS405*
     - *6.1.6. OTT Pluvio^2*
     - *6.1.7. Sutron TPG*
   - *6.2. Tipping buckets*
     - *6.2.1. CAE PMB25R*
     - *6.2.2. EML UPG1000*

WMO-Solid Precipitation Intercomparison Experiment (WMO-SPICE)

Mission Statement and Objectives

1. Mission Statement

To recommend appropriate automated field reference system(s) for the unattended measurement of solid precipitation in a range of cold climates and seasons, and to provide guidance on the performance of modern automated systems for measuring: (i) total precipitation amount in cold climates for all seasons, especially when the precipitation is solid, (ii) snowfall (height of new fallen snow), and (iii) snow depth.

To understand and document the differences between an automatic field reference system and different automatic systems, and between automatic and manual measurements of solid precipitation using equally exposed/shielded gauges, including their siting and configuration.

2. Scope and Definition

Building on the results and recommendations of previous intercomparisons, the WMO Solid Precipitation Intercomparison Experiment (SPICE) will focus on the performance of modern automated sensors measuring solid precipitation. SPICE will investigate and report the measurement and reporting of the following parameters:

With highest priority:

- Precipitation amount, over various time periods (minutes, hours, days, season), as a function of precipitation phase (liquid, solid, mixed).
- Snow on the ground (snow depth); as snow depth measurements are closely tied to snowfall measurements, the intercomparison will address the linkages between them.

With lower priority:

- Solid and mixed precipitation intensity.

As a key outcome, recommendations will be made to WMO Members, WMO programs, manufacturers and the scientific community, on the ability to accurately measure solid precipitation, on the use of automatic instruments, and the improvements possible. The results of the experiment will inform those Members that wish to automate their manual observations.

An important aspect of this project will be to ensure that all available remotely sensed precipitation data is collected and included as part of the intercomparison data base. However, analysis of these data is beyond the scope of this intercomparison. The results of this intercomparison can later contribute to improved spatial and temporal estimates of precipitation.

3. Background

Solid precipitation is one of the more complex parameters to be observed and measured by automatic sensors and systems. The measurement of precipitation has been the subject of a multitude of studies,

but there have been limited coordinated assessments of the ability and reliability of automatic sensors to accurately measure solid precipitation. The WMO Solid Precipitation Measurement Intercomparison (WMO/TD-No. 872, 1998) focused on the instruments in use in national networks at the time of the intercomparison, primarily manual methods of observation. The assessment of automatic sensors/systems for snow depth and snowfall measurement was not a central part of the study, and no intercomparison stations were included in the Arctic or Antarctic.

Since then, an increasing percentage of precipitation data used in a variety of applications have been obtained using automatic instruments and stations, including the measurement of snow depth, and many new applications (e.g., climate change, nowcasting, water supply, complex terrain, avalanche warnings, etc) have emerged. At the same time, many of the new techniques used for the measurement of solid precipitation are of non-catchment type, e.g. light scattering, microwave backscatter, mass and heat transfer, etc.

Additionally, during the development of proposals for satellite sensors to measure solid precipitation, the issue of validation and calibration of such products using in-situ measurements (network or reference stations) identified the availability of reliable measurements of solid precipitation at automatic stations as a key input in assessing measurements in cold climates.

The modern data processing capabilities, data management and data assimilation techniques provide the means for better assessment and error analysis.

4. Intercomparison Objectives

WMO-SPICE will focus on the following key objectives:

I. Recommend appropriate automated field reference system(s) for the unattended measurement of solid precipitation. Define and validate one or more field references using automatic instruments for each parameter being investigated, over a range of temporal resolutions (e.g. from daily to minutes).

II. Assess/characterize automatic systems (both the hardware and the associated processing) used in operational applications for the measurement of Solid Precipitation (i.e. gauges as “black boxes”):
   a. Assess the ability of operational automatic systems to robustly perform over a range of operating conditions;
   b. Derive adjustments to be applied to measurements from operational automatic systems, as a function of variables available at an operational site: e.g., wind, temp, RH;
   c. Make recommendations on the required ancillary data to enable the derivation of adjustments to be applied to data from operational sites on a regular basis, potentially, in real-time or near real-time;
   d. Assess operational data processing and data quality management techniques;
   e. Assess the minimum practicable temporal resolution for reporting a valid solid precipitation measurement (amount, snowfall, and snow depth on the ground);
   f. Evaluate the ability to detect and measure trace to light precipitation.

III. Provide recommendations on best practices and configurations for measurement systems in operational environments:
   a. On the exposure and siting specific to various types of instruments;
   b. On the optimal gauge and shield combination for each type of measurement, for different collection conditions/climates (e.g., arctic, prairie, coastal snows, windy, mixed conditions);
   c. On instrument specific operational aspects, specific to cold conditions: use of heating, use of antifreeze (evaluation based on its hygroscopic properties and composition to meet operational requirements);
   d. On instruments and their power management requirements needed to provide valid measurements in harsh environments;
   e. On the use of visibility to estimate snowfall intensity
   f. On appropriate target(s) under snow depth measuring sensors;
   g. Consideration will be given to the needs of remote locations, in particular those with power and/or communications limitations.

IV. Assess the achievable uncertainty of the measurement systems evaluated during SPICE and their ability to effectively accurately report solid precipitation.
   a. Assess the sensitivity, uncertainty, bias, repeatability, and response time of operational and emerging automatic systems;
   b. Assess and report on the sources and magnitude of errors including instrument (sensor), exposure (shielding), environment (temperature, wind, microphysics, snow particle and snow fall density), data collection and associated processing algorithms with respect to sampling, averaging, filtering, and reporting.

V. Evaluate new and emerging technology for the measurement of solid precipitation (e.g. non-catchment type), and their potential for use in operational applications.

VI. Configure and collect a comprehensive data set for further data mining or for specific applications (e.g., radar- and/or satellite-based snowfall estimation). Enable additional studies on the homogenization of automatic/manual observations and the traceability of automated measurements to manual measurements.

5. Deliverables

WMO-SPICE will provide reports on the intermediate and final results of the experiment covering the following aspects:

   a. Recommendations of automatic field references systems, for the unattended measurement of the parameters evaluated;
   b. Characterization of the performance of existing, new, and emerging technologies measuring solid precipitation, and their configurations, addressing the objectives of the intercomparison.;
   c. A comprehensive data set for legacy use, for further data mining.
   d. Update of relevant chapters of the CIMO Guide (WMO No 8) and potential publications of WMO/ISO standards (under the WMO-ISO agreement, 2009).
   e. Guidance to Members on transition to automation from manual observations of solid precipitation measurements;
   f. Recommendations made to manufacturers on instrument requirements and improvements.

6. Potential Candidate list of Instruments and Configurations

The experiment may include many instrument types, models and configurations identified as currently operational, as summarized in WMO CIMO IOM 102, Survey on National Summaries of Methods and Instruments for Solid Precipitation Measurement at Automatic Weather Stations, http://www.wmo.int/pages/prog/www/IMOP/publications-IOM-series.html. In addition, known emerging technologies may be included, based on the recommendations of WMO Members.

The following is a list of instrument types and configurations, as identified in IOM 102, following the 2008 CIMO Survey:

- Weighing Gauges, Tipping Buckets, other storage gauges;
- Instruments employing emerging technologies e.g. laser, particle disdrometers, hot plate, spinning arm, vertically pointing radar, optical gauges, acoustic, precipitation video imaging, video camera.
- Wind shields: type: (e.g. Alter, Nipher, Tretyakov, Wyoming, Belfort, wood), and configurations (single, double, mixed type, small DFIR);
- Gauges equipped with heating in various configurations;
- Emerging trends: low-cost sensors, with (potential for) wide use.

7. Duration of the Intercomparison

Each intercomparison site will be operated for a minimum of two winter seasons.

WORLD METEOROLOGICAL ORGANIZATION

QUESTIONNAIRE on Solid Precipitation Gauges

Addressed to potential participants applying for participation in the

WMO Solid Precipitation Intercomparison Experiment (SPICE)

To be performed at multiple sites, 2012 - 2014

Note: please complete a separate questionnaire for each type of Sensor /System.
If necessary, attach additional pages.

Electronic version of the Questionnaire is available at:
http://www.wmo.int/web/www/IMOP/intercomparisons.html

1.	**Name of institution/company applying for participation**		
	You are applying for participation as ☐ WMO Member	☐ Company	
	Are you an HMEI member (www.hydrometeoindustry.org)?	☐ Yes	☐ No
	Address		

2.	**Contact Person for the Intercomparison**	
	Surname	First name
	Tel.:	Fax:
	E-mail:	Other:
	Alternative contact person	
	Surname	First name
	Tel.:	Fax:
	E-mail:	Other:

3.	**Name and address of the manufacturer** *(if different from no.2 above)*
	Address

4.	**General qualification for participation**	
4.1	Is your instrument capable to measure solid precipitation? ☐ YES ☐ No	*NOTE: If your answer to any of the questions 4.1 or 4.2 is "No" your instrument is not qualified to participate in SPICE*
4.2	Is your instrument making automatic measurements? ☐ YES ☐ No	
	Operating temperature range: °C	
	Total number of operational installations: In case of new instruments: Number of intended installations in 2012/2013:	

4.	**General qualification for participation**
	Could you provide one ☐ or two ☐ or how many identical instruments max. ? *(Provision of at least two identical instruments is preferred – one as backup. Given the fact that it is a multi site experiment an increased number would allow testing in different climatic regimes. The International Organizing Committee will decide on the appropriate test site based on instrument properties, site climatology and logistic aspects.)*

5.	**Instrument specifications**	
	Instrument manufacturer	
	Instrument name	Model/Type
	Catchment type instrument ☐	Non-Catchment type instrument ☐
	Principle of operation: Tipping bucket ☐ Weighing gauge ☐ Drop counter ☐ Optical disdrometer ☐ Other ☐ please describe:	
	Orifice / measurement area cm² or Measurement volume cm³	
	Which parameter(s) does the sensor report? Precipitation Intensity (PI) ☐ Precipitation Amount (PA) ☐ Time of Tipping ☐ Others	
	Measurement range for PI: from mm/h to mm/h	
	Resolution for PI *(measured or calculated over a period of <u>one</u> minute):* mm/h.	
	Measurement uncertainty for daily accumulation: Measurement uncertainty for PI:	
	What is the time interval over which the measurements are averaged? Averaging time for PI: seconds Averaging time for PA: seconds	
	Delay time for Precipitation Intensity (PI) measurement*:* minutes.	
	Internal update cycle for the calculation of a new measurement value*:* s. Please provide a procedure for synchronization of the instrument's calculation cycle with the clock of the data acquisition system. *(You can refer to another document if available).*	
	Values are internally linearized / corrected ☐ External correction necessary ☐ please provide algorithm	
	Precipitation accumulation limit *(e.g. due to full bucket of weighing gauge):* mm	

6.	**Sensor Interface**		
	Serial Digital Output	RS485 ☐ RS422 ☐	Ethernet ☐
	Data Protocol	SDI-12 ☐ proprietary ASCII ☐ Other: ☐ Please refer to manual	
	Analogue output	*Pulse* ☐ *Reed Relay* ☐ Current ☐	Voltage ☐
	Other ☐ Please describe		
	Note: If any other digital serial interface than Ethernet or RS485/422 is provided the participating institution/company should submit an appropriate converter (1 piece) with RS485/422 output.		

7.	**Information for field installation**	
	Notes on the power supply: if necessary converters must be provided for available voltages at the Intercomparison sites	
	Nominal power supply voltage: V	Maximum total power consumption W
	Heating of the instrument Which parts of the instrument are heated and at what heating power? heated by W heated by W heated by W Total maximum heating power: W	
	Max. cable length for power supply m	max. cable for signal cable m
	Notes on the cable lengths: Cable lengths for power supply and signal cable should be at least 4 m.	
	Notes on the amount of space for installation: there will be an area of 50 cm x 50 cm on a separate concrete foundation for each instrument.	
	Overall dimensions of the instrument, in cm Length **x** Width **x** Height cm	Total weight kg
	Is the instrument in operational use equipped with a windshield? Yes ☐ No ☐ *Note: In case of Yes you have to provide a windshield as used operationally in your network.* Type of windshield: *(please add a picture)* Overall dimensions: cm X cm X cm	
	Will you provide a mast that is suitable for the expected snow depth at the site? Yes ☐ No ☐ *Note: This is a multi-site Intercomparison. Depending on technical criteria and availability of space your instrument(s) will be proposed for installation at one or more sites.*	
	Dimensions: Length x Width x Height (in cm); and Weight (in kg) of main elements	
	Part id L x W x H kg	
	Part id L **x** W **x** H kg	
	Part id L **x** W **x** H kg	
	Part id L **x** W **x** H kg	
	Part id L **x** W **x** H kg	
	Requirements for installation (e.g. special mast, foundation etc.):	
	Will an expert give assistance with the field installation? Yes ☐ No ☐	
	Will an installation tools kit be provided? Yes ☐ No ☐	
	Any special tools required for the installation? Yes ☐ No ☐ Please describe	
	Special fixtures required for the installation? Yes ☐ No ☐ Please describe	
	Siting restrictions (e.g. clearance)?	

8.	**Operation and Maintenance**
	Average maintenance period: Maintenance item:

8.	**Operation and Maintenance**
	Any other special requirements to ensure proper operation? Yes ☐ No ☐ Please specify

9.	**Calibration**
	Note: Your instrument has to be calibrated before shipment.
	Calibration reference used:
	Recommended calibration intervals
	Calibration procedure

10.	**Shipment information**	
	This section has to be filled out only if shipment is necessary. *Note: (indicate information for <u>one</u> instrument only!)*	
	Approx. commercial value Euro	Total weight of consignment kg
	Number of boxes	Overall volume of boxes m^3
	Overall dimension, in cm (i.e. for storage purposes) Length **x** Width **x** Height cm	
	Other information concerning shipping	

11.	**Documentation**
	Appropriate documentation including detailed • installation instructions and manuals • operation and maintenance manuals • calibration procedures will be requested in due course (if possible in electronic format).
	Please note any other relevant information.

12.	**Motivation**
	What are your expectiations from SPICE as an instrument proponent in excess of the objectives that are already stated in SPICE? (see: http://www.wmo.int/pages/prog/www/IMOP/intercomparisons/SPICE/PR-6601-SPICE_en-Annex.pdf)?

Date

Name of person who completed this form

Please return an electronic copy (MS-Word document) of the completed Questionnaire, as soon as possible, but not later than **15 March 2012** to**:**

Dr Isabelle Rüedi
Senior Scientific Officer
OSD/OBS
World Meteorological Organization P.O. Box 2300
CH 1211 Geneva 2, Switzerland

E-mail: iruedi@wmo.int
Tel: +41 22 730 82 78
Fax: +41 22 730 80 21

With copy to:
Mrs Rodica Nitu
Environment Canada
4905 Duffering St.
TORONTO, ON
Canada M3H 5T4

E-mail: rodica.nitu@ec.gc.ca
Tel: +1 416 739 4133
Fax: +1 416 739 5721

WORLD METEOROLOGICAL ORGANIZATION

QUESTIONNAIRE for Snow Depth Gauges

Addressed to potential participants applying for participation in the

WMO Solid Precipitation Intercomparison Experiment (SPICE)
(Field Intercomparison at multiple sites, 2012 – 2014)

Note: please complete a separate questionnaire for each type of Sensor /System.
If necessary, attach additional pages.

Electronic version of the Questionnaire is available at:
http://www.wmo.int/web/www/IMOP/intercomparisons.html

1.	**Name of institution/company applying for participation**		
	You are applying for participation as ☐ WMO Member	☐ Company	
	Are you an HMEI member (www.hydrometeoindustry.org)?	☐ Yes	☐ No
	Address		

2.	**Contact Person for the Intercomparison**	
	Surname	First name
	Tel.:	Fax:
	E-mail:	Other:
	Alternative contact person	
	Surname	First name
	Tel.:	Fax:
	E-mail:	Other:

3.	**Name and address of the manufacturer** *(if different from no.2 above)*
	Address

4.	**General qualification for participation**	
4.1	Is your instrument capable of measuring snow depth automatically? ☐ YES ☐ No	*NOTE: If your answer to any of the questions 4.1 is "No" your instrument is not qualified to participate in SPICE*
	Operating temperature range: from °C to °C	
	Total number of operational installations: In case of new instruments: Number of intended installations in 2012/2013:	
	Could you provide One☐ or Two ☐ or how many identical instruments max. ? *(Provision of at least two identical instruments is preferred for this field intercomparison – one as backup. Given the fact that it is a multi site experiment an increased number would allow testing in different climatic regimes. The International Organizing Committee will decide on the appropriate test site based on instrument properties, site climatology and logistic aspects.)*	

5.	**Instrument specifications**	
	Instrument manufacturer	
	Instrument name	Model/Type
	Principle of operation: Ultrasonic distancemeter ☐ Laser distancemeter ☐ Other ☐ please describe:	
	Measurement area on the ground: Does the actual measurement area on the ground vary with target distance? ☐ YES ☐ No Please explain (e.g. by giving the aperture angle) Minimum distance between sensor and snow surface cm	
	Which parameter(s) does the sensor report? Snow Depth ☐ Others	
	Measurement range: from cm to cm	
	Resolution: cm	
	Measurement uncertainty:	
	What is the time interval over which the measurements are averaged? minutes	
	Delay time for Snow Depth measurements: minutes.	
	Internal update cycle for the calculation of a new measurement value: s. Do you have a procedure for synchronization of the instrument's calculation cycle with the clock of the data acquisition system? *(You can refer to another document if available).*	
	Do the original distance measurements need any corrections (e.g. temperature correction)? ☐ Yes ☐ No Please give details which variables are used for correction and what kind of sensors: If you need input from sensors of the station, please indicate and give details: Are the values internally corrected ☐ Yes ☐ No Algorithm for external correction:	

6.	**Sensor Interface**				
	Serial Digital Output	RS485 ☐	RS422 ☐	Ethernet	☐

6.	**Sensor Interface**	
	Data Protocol	SDI-12 ☐ proprietary ASCII ☐ Other: ☐ Please refer to manual
	Analogue output	Current ☐ Voltage ☐ Other ☐ Please describe
	Note: If any other digital serial interface than Ethernet or RS485/422 is provided the participating institution/company should submit an appropriate converter (1 piece) with RS485/422 output.	

7.	**Information for field installation**	
	Notes on the power supply: if necessary converters must be provided for available voltages at the Intercomparison sites	
	Nominal power supply voltage: V	Maximum total power consumption W
	Heating of the instrument Which parts of the instrument are heated and at what heating power? heated by W heated by W heated by W Total maximum heating power: W	
	Cable length for power supply m for signal cable m	
	Notes on the cable lengths: Cable lengths for power supply and signal cable should be at least 6 m.	
	Notes on the amount of space for installation: there will be an area of 50 cm x 50 cm on a separate concrete foundation for each instrument.	
	Overall dimensions of the instrument, in cm Length **x** Width **x** Height cm	Total weight kg
	Will you provide a mast that is suitable for the expected snow depth at the site? Yes ☐ No ☐ *Note: This is a multi-site Intercomparison. Depending on technical criteria and availability of space your instrument(s) will be proposed for installation at one or more sites.*	
	Dimensions: Length x Width x Height (in cm); and Weight (in kg) of main elements	
	Part id L x W x H kg	
	Part id L **x** W **x** H kg	
	Part id L **x** W **x** H kg	
	Part id L **x** W **x** H kg	
	Part id L **x** W **x** H kg	
	Requirements for installation (e.g. special mast, foundation etc.):	
	Will an expert give assistance with the field installation? Yes ☐ No ☐	
	Will an installation tools kit be provided? Yes ☐ No ☐	
	Any special tools required for the installation? Yes ☐ No ☐ Please describe	
	Special fixtures required for the installation? Yes ☐ No ☐ Please describe	
	Siting restrictions (e.g. clearance)?	

8.	**Operation and Maintenance**
	Average maintenance period: Maintenance item:
	Any other special requirements to ensure proper operation? Yes ☐ No ☐ Please specify

9.	**Calibration**
	Note: Your instrument has to be calibrated before shipment.
	Calibration reference used:
	Recommended calibration intervals
	Calibration procedure

10.	**Shipment information**	
	This section has to be filled out only if shipment is necessary. *Note: (indicate information for <u>one</u> instrument only!)*	
	Approx. commercial value Euro	Total weight of consignment kg
	Number of boxes	Overall volume of boxes m^3
	Overall dimension, in cm (i.e. for storage purposes) Length **x** Width **x** Height cm	
	Other information concerning shipping	

11.	**Documentation**
	Appropriate documentation including detailed • installation instructions and manuals • operation and maintenance manuals • calibration procedures will be requested in due course (if possible in electronic format).
	Please note any other relevant information.

12.	**Motivation**
	What are your expectiations from SPICE as an instrument proponent in excess of the objectives that are already stated in SPICE? (see: http://www.wmo.int/pages/prog/www/IMOP/intercomparisons/SPICE/PR-6601-SPICE_en-Annex.pdf)?

Date

Name of person who completed this form

Please return an electronic copy (MS-Word document) of the completed Questionnaire, as soon as possible, but not later than **15 March 2012** to:

Dr Isabelle Rüedi
Senior Scientific Officer
OSD/OBS
World Meteorological Organization
P.O. Box 2300
CH 1211 Geneva 2, Switzerland

E-mail: iruedi@wmo.int
Tel: +41 22 730 82 78
Fax: +41 22 730 80 21

With copy to:
Mrs Rodica Nitu
Environment Canada
4905 Duffering St.
TORONTO, ON
Canada M3H 5T4

E-mail: rodica.nitu@ec.gc.ca
Tel: +1 416 739 4133
Fax: +1 416 739 5721

WORLD METEOROLOGICAL ORGANIZATION

QUESTIONNAIRE on Intercomparison Sites
Addressed to potential participants applying for participation in the

WMO Solid Precipitation Intercomparison Experiment (SPICE)
To be performed at multiple sites, 2012 - 2014
Note: If necessary, attach additional pages.
Electronic version of the Questionnaire is available at: http://www.wmo.int/web/www/IMOP/intercomparisons.html

IMPORTANT: If you are proposing a site, please ensure that a short letter is sent by your Permanent Representative to the WMO Secretariat, endorsing the proposal of the site and nominating the designated representative (see section 2 below) as point of contact for further discussions related to this project, if this hasn't already been done.

1.	**Name of institution/company applying for participation**	
	Name of the proposed Site/Facility Address	

2.	**Designated Representative for the Intercomparison**	
	Surname	First name
	Tel.:	Fax:
	E-mail:	Other:
	Alternative contact person	
	Surname	First name
	Tel.:	Fax:
	E-mail:	Other:

Note:
The information provided in this document will be used by the SPICE Project Team for deciding the allocation of the instruments and systems proposed by the SPICE participants, in direct correlation with the objectives of the experiment.
The SPICE Project Team may follow up with the designated representative, to gather additional information that will enable the project decisions. Any decision will be validated with the designated representative, before becoming final.

For further details on the scope and objectives of the WMO SPICE, please consults
http://www.wmo.int/pages/prog/www/IMOP/reports/2011/Joint_ET-II_IOC-SPICE.pdf

Section A: General Information:

Name and address of the proposed site/facility	
Will you provide instruments for inclusion in the SPICE intercomparison? If yes, please provide details in Section E.	☐ Yes ☐ No
In addition to your own instruments, does the site have capacity to accommodate instruments proposed for SPICE, by other participants? (e.g. provided directly by the manufacturers or other WMO Members or their representatives)? If yes, please provide details in Section C.	☐ Yes ☐ No
Are you interested in specific SPICE objectives? Would you consider focusing the experiments on your site on selected SPICE objectives? Please refer to the SPICE objectives and provide details in Section G.	☐ Yes ☐ No
Do you have other projects underway on the proposed intercomparison site, which may share/use data from instruments included in SPICE? (e.g. Instruments under test or instruments providing ancillary data). If yes, please provide details in Section G.	☐ Yes ☐ No

Section B: Site location, climate, and general facilities:
Additional information on the site climatology is required shall be provided in Section H of this document.

Latitude:		
Longitude:		
Elevation:		
Attach a topographic/satellite map of the area, if available (preferable in a KMZ file format)		
Proximity of large body of water or forested area (distance to site, surface):		
Proximity of obstacles (type, distance and direction to site, size):		
Type of terrain (provide details, including vegetation type, slope, etc):		
Prevailing climate, in particular during the winter period:		
Access to the site	Proximity to urban areas	
	Accessible year round	☐ Yes ☐ No
	Means of transportation for people and freight	
	Other:	
Status of the proposed site: e.g. Existing site or new; if new, when will it be completed and fully operational?		
Other programs operating on site or in the proximity of the site: (provide overview)		

Proximity to a weather radar system:	☐ Yes ☐ No	
	Distance to radar	
	Radar type	
Personnel available on site, during the experiment	☐ Yes ☐ No	
	If yes, indicate duration per day, type of maintenance performed;	
	If no, indicate the frequency of personnel visits over a season. Also indicate if emergency intervention is possible, and the response time.	
Would additional personnel be available during the intercomparison?	☐ Yes ☐ No	
Other information regarding the host site:		

Section C: Site Facilities Description:

Description of capacity for additional instruments submitted for evaluation under SPICE

Attach the site layout, including schematics, distances, including from the surroundings; mark the pads available for the additional instruments.		
Number of instrument pads :	Total:	
	Utilized for instruments proposed by host, as participants in the intercomparison;	
	Utilised with instruments providing ancillary measurements:	
	Available for additional instruments selected for the intercomparison:	
Distance between the instrument pads:		
Power available at the instrument pads (voltage, number of circuits, limitations):		
Type of data communication available at the instrument pads (number of circuits, uni/bi-directional):		
Data communication/ transmission:	Between instruments and the data acquisition system on site:	
	Between the site and a central location/archive.	
	Type and frequency of data download (real time, daily, monthly, etc)	

[Table continued from previous page]

On site power	AC voltage (e.g. 110VAC/220VAC)	
	DC voltage (e.g. ±12VDC)	
	Limitations	
	Other information	

Site Data Acquisition System:

Provide a detailed description of the data acquisition system used on the proposed site for collecting, storing, transmitting the intercomparison data (from the instruments and systems under test and the ancillary measurements); the description should include a diagram of the system.

Data logging system (hardware, software):	
Platform:	
Type of inputs:	
Capacity for handling analog sensors	
Number of interfaces available for the experiment:	
Capacity for expansion, and back up	
Feasible temporal resolution of the data collection:	
Frequency of data download (real time, daily, etc)	
Available designated central archiving location/facility, for actively storing the intercomparison data set obtained from the proposed site (provide overview)	
Back-up facilities	

Section D: Site reference configuration:

Summary of field references currently available on the proposed host site;

NOTE: the SPICE IOC may suggest changes to existing configurations to meet project objectives.

D1: Reference for the measurement of the total precipitation (precipitation amount)

	Status (E-Existing, P-planned, N/A – not available)	**Description (incl. gauge orifice height, model and SW version, configuration, heating)**
WMO Secondary Field Reference: Double Fence Intercomparison Reference (DFIR) with a manual (Tretyakov) gauge and Tretyakov shield;		
Frequency of manual observations		
Methodology of measurement (weighing, measure snow water content, etc)		
Length of time since operational (archived data)		
Field Reference using an automatic weighing gauge in a DFIR:		
Type of weighing gauge in the DFIR		
Type of shield used with the automatic gauge		

[Table continued from previous page]

	Status (E-Existing, P-planned, N/A – not available)	**Description (incl. gauge orifice height, model and SW version, configuration, heating)**
Frequency and format of reference data		
Length of time since operational (archived data)		
Other field reference for the measurement of total precipitation, currently in use on the site		
Instrument used (type, model, firmware version, configuration, heating)		
Type of shield (include height, diameter)		
Frequency and format of reference data		
Other information regarding the field reference used		
Does the site have capacity for the installation of a field reference of the configuration recommended by the SPICE IOC? (the potential configuration consists of gauge to be identified, in a Double Alter wind shield). Provide description, limitations, etc.		

D2: Reference for the measurement of snow depth

Availability of manual snow depth measurements	☐ Yes ☐ No
Frequency of measurement	
Method of measurement	
Instrument/equipment used (provide description)	
Availability of cameras for snow depth monitoring; provide details	☐ Yes ☐ No
	Description:

Section E: Instruments proposed by the host

Detailed information will be provided for each instrument, using the Instrument Questionnaire, by filling out sections 4, 5, 6, 7*, 8, and 9** (see NOTES).

NOTES:

- (*)Section 7: provide information on heating, if applicable, only. Attach a description of the heating algorithm, if applicable.
- (**) Section 9: provide reference to the calibration procedure, as available (e.g. manufacturer provided, internally defined, etc). The SPICE IOC may request additional information on the calibration procedure, at a later date.

Instrument Type (e.g. Weighing Gauge, Tipping Bucket, Disdrometers, Snow Depth, etc…)	Instrument Make and Model (details on each instrument will be provided in response to the questionnaire on Instruments)	Configuration (shield type, number of transducers, other as applicable)		Number of instruments in the same configuration
		Shield type (if applicable)	Other configuration information	

Section F: Ancillary measurements:

Parameter	**Instruments Used (model, make, configuration, height, number of instruments.**	**Data available (data format)**	**Reporting Interval**	**Status (E-Existing, P-planned, N/A – not available for SPICE)**
Air Temperature				
Relative Humidity				
DewPoint Temperature				
Atmospheric Pressure				
Wind Speed 10 m				
Wind Direction 10 m				
Wind Speed at the gauge orifice (specify height)				
Precipitation Detector (Y/N output)				
Precipitation type				
Ice detection sensor				
Snow Depth				

[Table continued from previous page]

Parameter	**Instruments Used (model, make, configuration, height, number of instruments.**	**Data available (data format)**	**Reporting Interval**	**Status (E-Existing, P-planned, N/A – not available for SPICE)**
Net Radiation				
Visibility				
Web / video / still cameras				
Micro physical (any method e.g. Auto or manual)				
Snow Water Equivalent observations (manual or automatic)				
Manual observations: precipitation type, snow course, snow depth, assessment of snow drift, blowing snow, etc:				
Radar data				
Other data available				

Section G: MOTIVATION:

List the SPICE objectives of primary interest for your organization:	
Indicate whether and which SPICE objectives you would prefer to focus on, for the experiments organized on the proposed site:	
List existing or planned national projects organized in parallel with SPICE on this site, and which may be sharing data with SPICE	
Other information, relevant to SPICE:	

Section H: Site Climatological Information

Site Climatology (provide information for the months when solid precipitation is expected)													
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Annual
Temperature:													
Daily Average (°C)													
Daily Maximum (°C)													
Daily Minimum (°C)													
Extreme Maximum (°C)													
Extreme Minimum (°C)													
PRECIPITATION													
Total average Snowfall (cm)													
Total average Precipitation (mm)													
Extreme Daily Snowfall (cm)													
Extreme Snow Depth (cm)													
Average number of days per period with Maximum Temperature:													
<= 0 °C													
> 0 °C													

[Table continued from previous page]

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Annual
Average number of days per period with Minimum Temperature:													
> 0 °C													
<= 0 °C													
< - 30 °C													
Average number of days per period with snowfall													
trace													
>= 10 cm													
>= 25 cm													
Average number of days per period with precipitation													
>= 0.2 mm													
>= 25 mm													
Wind Speed													
Maximum wind speed (m/s)													
daily mean wind speed (m/s)													

Date	Name of person who completed this form

Please return an electronic copy (MS-Word document) of the completed Questionnaire, as soon as possible, but not later than **15 March 2012** to:	
Dr Isabelle Rüedi E-mail: iruedi@wmo.int Observing System Division Tel: +41 22 730 82 78 OBS Department Fax: +41 22 730 80 21 World Meteorological Organization P.O. Box 2300 CH 1211 Geneva 2, Switzerland	with copy to: Ms Rodica Nitu E-mail: rodica.nitu@ec.gc.ca Environment Canada Tel: +1 416 739 4133 4905 Duffering St. Fax: +1 416 739 5721 TORONTO, ON Canada M3H 5T4

WMO SPICE

SITE COMMISSIONING PROTOCOL

V3.1 (JUL, 23 2013)

TABLE OF CONTENTS

1. ORGANIZATION OF THE DOCUMENT

The Commissioning Protocol is organized into four parts:

1. **The site components,** data transfer and sharing pathways, and project organizational structure are outlined in Section 3;
2. **The site commissioning procedures**, including pre-commissioning activities and the Interaction with the Instrument Providers, Sections 4 to 6;
3. **SPICE Data Archive,** Section 7**.**
4. **Appendix A: the template for the Proof of Performance (POP) Report**, in which all site configuration details and commissioning activities are documented.

Appendix B outlines the SPICE Data Levels and Data Sets, and Appendix C includes a list of acronyms used throughout the document.

The first two sections are intended to provide background information on the commissioning process within the scope of the SPICE project, while the Appendix A contains the forms which are required to be filled out as part of the commissioning of the site. Once completed, these forms become the Commissioning Report.

The SPICE data archive section outlines the requirements regarding the SPICE data levels and datasets and the planned strategy for the archival of SPICE data to a central location(s).

2. PURPOSE AND SCOPE

This document is prepared by the WMO SPICE IOC. It outlines the procedures for post-installation testing and commissioning of the sites participating in the WMO SPICE experiment and documents the responsibilities for each aspect of the commissioning process.

Commissioning of a WMO SPICE site refers to the act of "turning it on" and marking the start of the collection of the "official" observations and measurements from the instruments included in the intercomparison (reference, instruments under test, ancillary measurements), and their archival on the designated Site Data Archive.
For this purpose, each site will designate a location for the Site Data Archive, which must protect the integrity of the intercomparison data.

End-to-end data quality and integrity for each instrument on each SPICE site will be verified before the commissioning can take place. It is essential that:

- Only agreed upon instruments are to be installed, in an accepted and standardized configuration;

- Each component be properly tested, and its performance verified, prior to commissioning;
- The transfer of instrument data to the Site Data Archive is validated and the archive secured.

Various individuals and organizations are referred to in this document as having responsibilities.

- SPICE IOC
- SPICE Project Team
- SPICE Data Analysis Team
- Site Manager
- Site (SPICE) Project Team
- ER refers to the Evaluation Representative, an individual named by the SPICE IOC
- IR, the Installation Representative, is identified by the Site Manager, responsible for the site configuration.
- Instrument Providers

3. CONFIGURATIONS AND ASSOCIATED COMMISSIONING REQUIREMENTS

3.1 SPICE SITE COMPONENTS

The SPICE Components include all or some of the following components:

- Field working reference systems (R3, and where applicable R2, and R1: site-specific)
- Reference measurements for snow on the ground (where applicable)
- Instruments under test provided by the host;
- Instruments under test supplied by the Instrument Providers;
- Ancillary measurements (both required and desired measurements listed):
  - Precipitation occurrence/intensity/size/type
  - Station pressure
  - Temperature/dew point
  - Relative humidity
  - Wind speed/direction (2-D and/or 3-D): different heights;
  - Manual observations
  - Vertical particle velocity
  - Net radiation
  - Snow Water Equivalent (SWE)
  - Icing occurrence
  - Visibility
  - Sky condition

- Derived or modeled ancillary parameters: wet bulb temperature, upper air temperature, snow particle density;
- Photography and video equipment for recording and archival of site conditions;

3.2 COMMUNICATION INTERFACES

The SPICE site teams are led by their respective Site Managers and are responsible to setup and manage an effective data communication system collecting, transmitting and archiving the site dataset, continuously, or at predefined intervals (e.g. daily) on the Site Data Archive.

As stated in the report of the SPICE IOC-2 meeting (Boulder), it is recommended that 6 s data be collected for gauges in reference systems and instruments under test, where possible; alternatively, 10 s or 60 s sample intervals can be used.
The frequency of the collection of ancillary measurements will be similar to that of the instruments under test, to the extent possible.

Data communication for SPICE includes the following components:

- Instrument to data logger (site specific);
- Instrument to a site data acquisition system located on site, site specific;
- Transmission of SPICE data from the site to a designated Site Data Archive;
- Transmission of SPICE data from the Site Data Archive to SPICE Archive(s) (See Section 7);
- Transmission of gauge-specific and requisite ancillary SPICE data to Instrument Providers for review.

The communication components and any future changes that may impact the availability of instruments will be documented. Any change to the configuration will be subject to a period of testing to ensure that the availability of instrument data is not affected. The IOC will review and accept the final configuration.

3.3 SPICE SITE PROJECT TEAM

The Site Manager will document the membership of the SPICE Site Project Team, including the names of the individuals who are engaged in the SPICE experiment on the respective site. This information will include reference to the roles relative to the SPICE experiment.
During the project, the participation in the SPICE Site Project Team could change. The Site Manager will to update the Site Documentation to reflect the changes (people, roles).

4. PRE-COMMISSIONING ACTIVITIES

The pre-commissioning activities are an integral part of the process of ensuring the quality of the experiment. The following sections detail the pre-commissioning activities ensuring that site infrastructure and procedures are properly managed and documented.

4.1 STATION INSTALLATION AND SCHEDULING

The IOC and the Site Managers will develop target dates for the installation and commissioning of each SPICE Site. An Installation Representative will be identified by the Site Manager to manage the installation. Site drawings, instrument siting and installation according to national standards, IOC agreed guidelines, or manufacturer recommendations, and exceptions will be documented as part of the POP Report.

4.2 TESTING OF INSTRUMENTS INCLUDED IN THE INTERCOMPARISON

The testing of instruments is conducted by the SPICE Site Project Team. Based on the results, the Site Manager will determine the readiness of instruments and the site for the formal phase of the experiment.

4.2.1 SITE DOCUMENTATION

Technical documentation for each SPICE component will include, but not limited to, the site layout, instruments details and configuration, data collection (including the data format), number of similar instruments, installation details, maintenance standards.
Specific information on the Site Documentation is provided in Appendix A.

4.2.2 MONITORING OF PERFORMANCE

The Site Manager will establish feasible procedures for monitoring the performance of instruments, identifying problems with the data, and initiating and tracking remedial actions. This may include:

- Review data, diagnostic data, quick view plots, QC reports, etc.
- Establishing Site Journals/Blogs documenting the performance and intervention on the instruments (directly – e.g. snow clearing - or indirectly – e.g. system reset -)

4.2.3 SITE MAINTENANCE

The SPICE Site Manager will ensure that site maintenance is available to limit the periods or data outage.

5. COMMISSIONING ACTIVITIES

The commissioning of a SPICE site is led by the Site Manager. The SPICE POP Report will document the status of the site operation at the start of the intercomparison.

The site commissioning process consists of the following steps:

- Determine the instrument readiness, including;
  - ⇒ Installation and configuration of the instruments participating in the experiment;
  - ⇒ Data integrity confirmation at the Site Data Archive;
- Review and approval of the POP Report by the IOC;
- Agreement on the official start of the experiment on the site.

5.1 DETERMINATION OF SITE READINESS

This sub-section details the activities to be conducted following the installation of instruments, and which are completed prior to the official start of the SPICE experiment on the site.

5.1.1 SITE READINESS EVALUATION

The Site Manager will initiate the evaluation of the SPICE Site and will provide to the IOC adequate notice of the SPICE site commissioning.

The IOC will name a representative (the ER) to conduct the evaluation of the Site Documentation prepared by the Site Manager. The ER will work with the Site Manager on the evaluation of the POP Report.

The site readiness evaluation should be sufficient to ensure proper operation of all instruments and interfaces. The assessments will include:

- Satisfactory performance of the field reference system(s).
- Satisfactory performance of each instrument under test.
- Satisfactory performance of instruments providing ancillary measurements.
- Satisfactory performance of site communication components and interfaces.
- Satisfactory performance of the data transmission to the Site Data Archive;
- Proper functioning of service backup capabilities for that particular site, if available.
- Maintenance capacity.

5.1.2 COMPLETION OF POP REPORT

The SPICE Site POP Report documents the readiness of the site and is approved by the IOC.

The POP Report includes:

- A form for recording station information and configuration, including the site layout;
- A form for documenting the configuration of SPICE field working reference configurations, including both manual and automatic measurements;
- Forms for recording the specifications of instruments under test and instruments used to provide ancillary measurements ;
- Details of tests conducted for instrument data validation;
- Details of tests conducted for end-to-end data validation;
- A checklist for all additional documentation to be recorded and submitted ;
- A table for recording commissioning milestones.

The Site Manager will provide the POP Report to the IOC, for final review.

5.1.3 INVOKING WORKAROUNDS

A workaround is a temporary solution to a system limitation that requires special attention and will be removed eventually. Any workarounds will be documented and included as part of the POP Report. Each work-around will be tracked as an open item until resolved.

5.2 Approval of Site Commissioning

The Site Manager will notify and update the IOC on the organization and completion of the tests outlined in Appendix A. Once all tests results are verified, the IOC and the Site Manager will agree on the start date of the formal experiment on the site.

In case some of the instruments under test are not ready for the start of the experiment as planned (currently Nov. 15, 2012), the experiment could commence in steps, provided that all field references and key ancillary parameters (wind speed and direction, temperature) have been commissioned.

Commissioning of additional instruments would follow as their configurations are finalized; this will allow for their inclusion in the experiment as early as feasible, with no compromise to the data quality. The Data Analysis Team will take into consideration the commissioning data for each instrument.

5.3 Implementation of Approved SPICE Site Commissioning

Upon commissioning, the site will commence the official collection of the SPICE project dataset and ancillary measurements/observations.

6. Interaction with the Instrument Providers

Instrument Providers are responsible for the delivery of their instruments to the SPICE Sites and for supporting the Site Managers in verifying their proper functioning before and during SPICE.

6.1 Pre-Commissioning Activities: Engagement of the Instrument Providers

During the installation, the Site Manager or a representative will engage the Instrument Provider regarding the preparation of their instruments, to ensure the operation within recommended standards.

The Site Manager would confirm with the Instrument Provider the functioning of the instrument prior to the commissioning of the site. This could be done by the sharing of instrument and/or ancillary data and pictures, coordinated site visits, or any other method agreed upon by the two parties.

The Site Manager should be able to indicate in the Commissioning Report the confirmation from the Instrument Provider that the instrument operates as expected.

6.2 Engagement of Instrument Providers during the Experiment

During the experiment, each Instrument Provider will be given access to the unprocessed output from its own instrument(s), and a minimum set of corresponding ancillary data consisting of air

temperature, relative humidity, and wind speed. These data are provided only for ensuring the proper functioning of the instruments, and will neither be reported nor published prior to publication of the SPICE Final Report.

The Site Manager will coordinate the data transfer to the Instrument Provider(s), including such aspects as the frequency, methodology, etc. It is desired that this data transfer is in place prior to the start of the experiment. The Instrument Provider is expected to alert the Site Manager in the event that a malfunction of an instrument is noted, and provide support to the Site Project team (including site visits), if needed, to address the failure.

The Instrument Providers could visit the intercomparison sites, after prior arrangements are made with the Site Manager.

7. SPICE DATA ARCHIVAL

The SPICE Project Team will establish and maintain a SPICE Archive on at least one SPICE designated Server where the Site Intercomparison Datasets and the Input Documentation will be stored. This will facilitate the preparation of data for the individual and comparative data analysis and the preparation of the Final Report. A description of the data levels and datasets for SPICE, as currently defined, is provided in Appendix B.

The National Centre for Atmospheric Research (NCAR), USA, will host the SPICE Archive and provide quick view capabilities of (near) real time data. Options for a second SPICE Archive are being explored by Environment Canada, Canada.

Each Site Manager will work towards preparing the transfer of Level 1 and Level 2a datasets to the SPICE Archive(s). The IOC will provide to the Site Managers the requirements regarding the data transfer to enable the preparation of datasets (format change, setup of data uploads/availability, etc…)

The data transfer between the Site Data Archive and the SPICE Archive is expected to be established and validated within 3 months of the official start of the experiment, and implemented based on site specific conditions and limitations.

APPENDIX A: PROOF OF PERFORMANCE (POP) FORMS

SECTION A1: STATION INFORMATION

Station name	
Reference town	
Station latitude	
Station longitude	
Station elevation in metres	

Insert here a Site Layout indicating the location of SPICE references and all instruments, including distances and the direction of the prevailing winter winds.

Insert here a set of pictures documenting the overall site installation (views from N, E, S, W).

It is suggested to submit here also a horizon / sky view diagram taken with a camera., if available

SECTION A2: SPICE FIELD WORKING REFERENCE SYSTEM CONFIGURATION

Field Reference Type R0

R0 type	MANUAL ☐ AUTOMATIC ☐
Measurement frequency, planned	
Measurement methodology planned (volume, weight, etc)	

Additional information required: Provide details of the planned measurement procedure.

Configuration of the bush

Description of surrounding obstacles (including distance/direction from, height, and type)	
Bush area	
Average height of the bush	
Bush vegetation type	*i.e plant species, deciduos leaves or not, etc.*
Maintenance details	*i.e prune every XX months;*

Collector and shield specifications (manual configuration)

Model	
Inlet area	
Installation height (measured at the top of the collector)	
Number of collectors available for the experiment	
Shield type	

Weighing gauge specifications (automatic configuration)

Make and model	
Serial number	
Firmware version (if applicable)	
Number of transducers (if applicable)	
Height of installation (measured from the top of the gauge)	
Heater configuration and algorithm	

Output data message format	
Frequency of data sampling	

Single Alter shield

According to the SPICE instructions?	☐ Yes ☐ No
Attached to the post of the weighing gauge?	☐ Yes ☐ No
If different, provide details:	

Picture. Field Reference Type R0

Table. Field Calibration of Reference Type R0 (if applicable)

48h Observation Table for Reference Type R0 (Manual) or Plots (Automatic)

Field Reference Type R1 (Manual)

Measurement frequency, planned	
Measurement methodology planned (volume, weight, etc)	

Additional information required: Provide details of the planned measurement procedure.

Configuration of the DFIR fence

Description of surrounding obstacles (including distance/direction from, height, and type)	
Diameter	
Height of the outer fence (measured at the top)	
Height of the inner fence (measured at the top)	
Length of slats	
Width of slats	
Slat material	

Collector and shield specifications

Model	
Inlet area	
Installation height (measured at the top of the collector)	
Number of collectors available for the experiment	
Shield type	

Picture. Field Reference Type R1 (Manual)

Table. Field Calibration of Reference Type R1 (Manual) ??????

48h Observation Table for Reference Type R1 (Manual)

Field Reference Type R2 (Automatic)

Configuration of the DFIR fence

Description of surrounding obstacles (including distance/direction from, height, and type)	
Diameter	
Height of the outer fence (measured at the top)	
Height of the inner fence (measured at the top)	
Length of slats	
Width of slats	
Slat material	

Single Alter shield

According to the SPICE instructions?	☐ Yes ☐ No
Attached to the post of the weighing gauge?	☐ Yes ☐ No
If different, provide details:	

Weighing gauge (WG)

Make and model	
Serial number	
Firmware version (if applicable)	
Number of transducers (if applicable)	
Height of installation (measured from the top of the gauge)	
Heater configuration and algorithm	
Output data message format	
Frequency of data sampling	

Precipitation detector

Make and model	
output data message format	
Data sampling frequency	
Height of installation. *DAT team recommend the following place for an optical precipitation detector or precipitation type sensor inside the DFIR:* • *Inside the inner fence* • *75 cm below the gauge opening, corresponds to half way down the inner fence*	
Location of installation relative to WG in reference system. *DAT team recommend to locate the optical precipitation detector or precipitation type:* • *perpendicular to the main wind direction* • *if possible using two precipitation sensors at different places to account for different wind directions.* • *in the middle between Alter and inner fence*	

Picture. Field Reference Type R2 (Automatic)

Table. Field Calibration of Reference Type R2 (Automatic)

48h Plot. Field Reference Type R2 (Automatic)

Field Reference Type R3 (Automatic)

Presence of a WG with a single Alter shield?	☐ Yes ☐ No
Presence of a WG with no shield?	☐ Yes ☐ No
Description of surrounding obstacles (including distance/direction from, height, and type)	
Distance between WGs (as close as possible, but exceeding minimum distance between gauges for a Class 1 siting configuration (as per WMO guidelines): Generally a flat area within 10m of instrument. This area surrounded by generally open space with a slope of less than 1:3 (19°) that is considered to be representative of the large scale area.	

Weighing gauge (1 of 2)

Make and model	
Serial number	
Firmware version (if applicable)	
Number of transducers (if applicable)	
Height of installation (measured from the top of the gauge)	
Heater configuration and algorithm	
Output data message format	
Frequency of data sampling	

Weighing gauge (2 of 2)

Make and model	
Serial number	
Firmware version (if applicable)	
Number of transducers (if applicable)	
Height of installation (measured from the top of the gauge)	
Heater configuration and algorithm	

Output data message format	
Frequency of data sampling	

Single Alter shield

According to the SPICE instructions?	☐ Yes ☐ No
Attached to the post of the weighing gauge?	☐ Yes ☐ No
If different, provide details:	

Precipitation detector

Make and model	
Data output format	
Data sampling frequency	
Height of installation.	
Location of installation relative to WGs in reference system.	

Pictures. Field Reference Type R3 (Automatic).

Weighing Gauge 1 Weighing Gauge 2

Table. Field Calibration of Reference Type R3 (Automatic) Weighing Gauges 1 and 2

48h Plots. Field Reference Type R3 (Automatic). Weighing Gauges 1 and 2

Field Reference for the Measurement of Snow on the Ground

Method used	
Equipment used	
Frequency of measurement	

Picture. Field Reference for the Measurement of Snow on the Ground

Table. Field Calibration for the Measurement of Snow on the Ground

48h Observation Table. Field Reference for the Measurement of Snow on the Ground

SECTION A3: INSTRUMENT METADATA REPORT

For each instrument under test and each instrument used to provide ancillary measurements, an Instrument Metadata Report should be completed in full and submitted as part of the POP Report.

Instrument Metadata Report

> *IMPORTANT: Please copy this form (as necessary) and complete separately for each instrument under test and each instrument that will be used to provide ancillary measurements during WMO SPICE.*

Instrument Name: _____
Instrument number _____ of _____

Manufacturer	
Model	
Serial number	
Firmware version (if applicable)	

Field configuration

Location on site	
Orientation	
Height (measured at top)	
Shield (if applicable)	
Heating (if applicable)	

Data output

Data communication protocol	
Output data message format (include description of fields)	
Data sampling frequency	

Instrument Picture.

Field calibration (if any).

48h Plot.

SECTION A4: CONFIRMATION OF EXPERIMENT CONFIGURATION

TEST 1: INSTRUMENT CALIBRATION AND CHECKS

The Site Manager will organize the check and calibration of each instrument included in the experiment (as part of the reference, or as an instrument under test). The check sheets and calibration results will be included in the designated areas of Sections A2 and A3.

- The calibration and check of the WG used as part of the reference will be conducted based on the guidelines adopted by the SPICE IOC.
- The calibration and check of the instruments under test will be conducted as specified by the manufacturer prior to the installation on the SPICE site, as well as following the installation in the field.

TEST 2: INSTRUMENT VALIDATION

After the field installation of each instrument (both those that are part of the reference and those that are instruments under test), at the minimum, a **continuous 48 hour data set** of the entire test setup will be stored and examined as an indication of instrument performance**.** The data sets for each instrument included in the intercomparison will be reviewed for data integrity and representativeness, against the predefined data format.

The evaluation of the instrument performance at this stage will be conducted using the 48 hour time series plots provided in Sections A2 and A3. The readiness state of each instrument will be reported in the Instrument Data Validation table below.

Any discrepancies will be investigated, addressed, and documented. Following the resolution of the discrepancies, the 48-hour end to end (e2e) test will be repeated. Notes, plots, logs, will be appended to the POP table of the reference/instrument under test, and the readiness state and date will be updated in the Instrument Data Validation table.

TEST 3: SITE-TO-ARCHIVE TRANSFER VALIDATION

Once the transfer of site data files to the SPICE Data Archive at NCAR has been initiated, compare the site data with those received at the SPICE Data Archive for a 24 hour period to ensure that no errors occurred during archival or transmission.
If any errors occur, log them and following the resolution of the discrepancies, repeat the 24-hour validation test.

When the Test 3 is passed mark the check box YES in the Instrument Data Validation table below (this

means that they have been also validated), with the starting date of the data transfer.

If Test 3 is not passed at the time of the Commissioning Report tick the checkbox NO and provide the expected date.

(Plots, datasets, errors logs, referred to Test 3 are **NOT** included in this document but archived by the site manager if further tests or analysis are required),

IMPORTANT: *Test 2 and Test 3 may be conducted simultaneously, depending on the site configuration.*

Instrument Data Validation

Instrument	Readiness (if Yes, indicate the date)	Data transfer to NCAR archive (**Test 3**) (If the answer is No report the expected date)	Comments
	☐ Yes ☐ No Date:	☐ Yes ☐ No Date:	

SECTION A5: SITE DOCUMENTATION CHECKLIST

A **Site Documentation Checklist** is provided below to track the inclusion of requisite documentation, data plots, and photos in sections A1 to A4.

Site Documentation Checklist

Site information and layout (Section A1)	☐ Included
Complete set of pictures documenting the overall site installation - views from N, E, S, W (Section A1)	☐ Included
Details of manual measurement procedure (Section A2)	☐ Included ☐ Not Applicable
Instrument Metadata Reports for all instruments under test and all instruments used to provide ancillary measurements (Section A3)	☐ Included
Calibration results and check sheets for all instruments (Sections A2, A3)	☐ Included
Instrument data validation:, 48h time series plots (Sections A2, A3)	☐ Included
Instrument data validation table (Section A4)	☐ Included
48h Instrument data validation: discrepancy reports (Section A4)	☐ Included ☐ Not Applicable
Pictures of installations of all reference instruments, instruments under test, and instruments used to provide ancillary measurements (Sections A2, A3)	☐ Included
End-to-end data validation (Section A4; see Instrument data validation table).	☐ Full (all gauges) ☐ Partial (some gauges) ☐ No
SPICE archive end-to-end data validation: discrepancy reports (Section A4)	☐ Yes ☐ No

Details of any workarounds (Sections A2, A3, A4)	☐ Included ☐ Not Applicable

APPENDIX B: SPICE DATA LEVELS AND DATASETS

Details of the different levels of data and associated datasets for SPICE are included below. **The present document addresses only data up to and including Level 2a.** Data of higher levels, and the associated datasets, are tentatively defined here for completeness.

Data Levels:

Level 1 data: are those collected as the output of each individual instrument, which have been converted into geophysical measurements (e.g. weight, mass, intensity), generally with high temporal resolution, and before any significant data quality control has been applied. A **Level 1** dataset contains data from only one instrument at one site.

Level 2a data: are time-synchronized data resulting from the sampling, averaging or some other signal/data processing having been applied to **Level 1** data from an individual instrument in order to separate signal from noise. These data have not been quality controlled, and should be used only for monitoring an instrument's status. A **Level 2a** dataset contains data from only one instrument at one site.

Level 2b data: are time-synchronized **Level 2a** data after a basic data quality control procedure has been applied. Basic data quality flags for validity and quality have been added. Missing records have been created and filled with a missing data quality indicator. A **Level 2b** dataset contains data from only one instrument at one site.

Level 3 data: derived by combining and further processing all **Level 2b** datasets from a site. At this level, advanced and multiple instrument data quality techniques have been applied. A **Level 3** dataset contains data from all instruments at an individual site.

Level 4 data: derived after performing an intercomparison of the **Level 3** data from one or more sites, taking into account snow climatology, wind regimes, temperatures, etc., and where applicable, differences in these from one site to another.

Datasets:

SPICE Site Dataset: A dataset comprising all **Level 1, 2a, 2b and 3** datasets from that Intercomparison Site.

SPICE Intercomparison Dataset: this is the Level 4 dataset that combines the **Level 3** data from all SPICE intercomparison sites. The **Project Team** will develop the **SPICE Intercomparison Dataset** using the Level 3 datasets from each **Intercomparison Site.** It contains summary Level 3 data and intercomparison data for all instruments and all sites.

The SPICE Dataset: The total SPICE dataset including all **SPICE Site Datasets**, **Site Documentation** and **Instrument Documentation** for all participating sites and instruments, the **SPICE Intercomparison Dataset,** and all SPICE analysis and assessment documentation.

APPENDIX C: ACRONYMS AND ABBREVIATIONS

DFIR	Double-Fence Intercomparison Reference
e2e	End-to-end
ER	Evaluating Representative
IOC	International Organizing Committee
IR	Installation Representative
NCAR	National Center for Atmospheric Research (USA)
POP	Proof of Performance
QC	Quality control
R0	Working field reference configuration 0: manual or automatic precipitation gauge in bush
R1	Working field reference configuration 1: manual precipitation gauge in DFIR
R2	Working field reference configuration 2: automatic weighing gauge in DFIR
R3	Working field reference configuration 3: two automatic weighing gauges; one shielded (single-Alter), one unshielded
SPICE	Solid Precipitation Intercomparison Experiment
SWE	Snow water equivalent
WG	Weighing gauge
WMO	World Meteorological Organization

World Meteorological Organization
Organisation météorologique mondiale

Secrétariat
7 *bis*, avenue de la Paix – Case postale 2300 – CH 1211 Genève 2 – Suisse
Tél.: +41 (0) 22 730 81 11 – Fax: +41 (0) 22 730 81 81
wmo@wmo.int – www.wmo.int

TEMPS • CLIMAT • EAU
WEATHER • CLIMATE • WATER

ACCEPTANCE OF SPICE DATA PROTOCOL

I, ... (insert your name), ……..……..(insert your title/function) hereby declare that I and the organization/company, ……..... (insert your org/co name), shall abide by the SPICE Data Protocol as set out below.

Signature: .....
(initials of the person are put on each page of the document)

Date: …… … Place: ……

SPICE DATA PROTOCOL

1. Introduction

1.1 The World Meteorological Organization (WMO) Solid Precipitation Intercomparison Experiment (SPICE) is an international intercomparison project being conducted as part of the work programme of the WMO Commission for Instruments and Methods of Observation (CIMO). A description of SPICE and its objectives is provided at Annex A.

1.2 Achieving the objectives of SPICE will involve the participation of numerous observing sites, and continuous and frequent observations of precipitation, snow depth and ancillary variables over a long period of time, sampled by a number of instruments of different makes supplied by different providers.

1.3 The purpose of this document is to define the protocol governing access to, use of and publication of information regarding the intercomparison sites and instrumentation, the algorithms employed by the instruments, the algorithms used in analysis of the data, the intercomparison data and the results to ensure that all SPICE participants are treated in a fair manner, and to ensure the timely dissemination/publication of SPICE results.

1.4 For clarity, all terms written in bold type in this document are defined in the Glossary provided at Annex B.

2. Project Governance

2.1 Overall project governance is the responsibility of an International Organizing Committee (IOC). The initial membership of the IOC was nominated by the President of CIMO and approved by the Secretary-General of WMO. The IOC membership also includes, as ex-officio members, the Site Manager of each SPICE intercomparison site.

2.2 The IOC is responsible for project governance, organization, overall planning and selection of participants (see Sect. 3.1), including:

- setting of project terms of reference,
- goals and objectives,
- ensuring the scientific integrity of the project,
- taking pragmatic steps to promote the project,
- approval of the project conclusions and output recommendations,
- reviewing the draft Final Report, and
- approving the SPICE Final Report.

2.3 The IOC reports, through its Chair, to CIMO. The IOC is also responsible for the establishment of a SPICE Project Team.

2.4 The SPICE Project Team is responsible for advising the IOC as regard to the detailed technical requirements for SPICE, including data analysis algorithms and methodology to be employed. The Project Team is also responsible for analysis and intercomparison of the data from the different SPICE intercomparison sites, and for drafting the SPICE Final Report.

3. SPICE Participation

3.1 SPICE involves participation in several different roles:

- IOC members;
- Project Team members;
- Intercomparison Sites (represented by Site Managers and their respective Site Teams);
- **Instrument Providers** (instrument manufacturers or WMO Members, and their representatives),
- Other participants, such as experts, computing facilities providers, data analysis contributors, and capacity building observers.

3.2 All SPICE participants (as listed in Sect. 3.1) are selected by the IOC.

3.3 All SPICE participants, and others who are provided future access to SPICE data and information, shall abide by this SPICE Data Protocol.

4. SPICE Project Execution

4.1 Each Intercomparison Site shall nominate a Site Manager and a Site Team. As noted in 2.1, Site Managers will represent their site as ex-officio members of the IOC. Each Site Manager must ensure that all members of their Site Team abide by the SPICE data protocol.

4.2 The IOC will establish guidelines to be followed by each Intercomparison Site for the conduct of SPICE. These will include:

- The configuration (layout) of the Intercomparison Site,
- The configuration, installation, operation and maintenance of instruments (which will be developed in consultation with **Instrument Providers**),

- The data collection setup, data archival and data quality control.

4.3 Each Site Manager, assisted by the respective Site Team, will be responsible for:

- compliance of the site with all intercomparison guidelines established by the IOC,
- securing the data collected from the Intercomparison Site,
- documentation of a site data protocol which is consistent with the SPICE Data Protocol and which governs access to the **SPICE Site Dataset** by the Site Team and others (such as staff of the site's host organization and staff of **Instrument Providers**),
- liaising with the **Instrument Providers**,
- documentation of the system implemented for managing the **SPICE Site Dataset**,
- preparation of the **SPICE Site Dataset**, and
- preparation of a final intercomparison report for that site (the **SPICE Site Final Report)**,

4.4 **Instrument Providers** are responsible for the delivery of their instruments to the intercomparison site, and for supporting the site managers in verifying their proper configuration/functioning before and during SPICE.

5 SPICE Documentation/Information

5.1 Pre-existing analysis methodologies and/or instrument algorithms provided by a SPICE participant will remain the property of that participant and may only be used or published with the prior written permission of the owner.

5.2 New analysis methodologies and/or instrument algorithms provided by a SPICE participant will remain the property of that participant, but will be freely available.

5.3 Intercomparison Sites give permission to WMO to use and publish Site Documentation detailing various aspects of the site, including its instrumentation and data handling system.

5.4 **Instrument Providers** give permission to WMO to use and publish **Instrument Documentation** provided throughout SPICE that describes the instrument(s) proposed, in terms of performance specifications, principle of operation, data format, internal data processing, installation requirement, interfaces and synchronization, unless provided in confidence.

6. SPICE Data, Datasets

6.1 Each Site Team shall collect and prepare its own **SPICE Site Dataset** that shall include both the data from the instruments under test and the ancillary measurements. These data shall be collected, processed and stored according to guidelines adopted by the IOC.

6.2 Each Intercomparison Site shall retain its **SPICE Site Dataset**, its **Site Documentation** and the **Instrument Documentation** from the participating instruments at that site.

6.3 Each **Instrument Provider** will be given access to unprocessed output from its own instrument(s), and a minimum set of corresponding ancillary data consisting of air temperature, relative humidity, and wind speed. These data are provided only for ensuring the proper functioning of the instruments, and shall neither be reported nor published prior to publication of the **SPICE Final Report.**

6.4 Each **SPICE Site Dataset** will be made available by the respective Site Team to the Project Team.

6.5 The Project Team will take all **SPICE Site Datasets** and use them to perform the overall SPICE intercomparison analysis and assessment, to produce the **SPICE Intercomparison Dataset.**

6.6 At the conclusion of SPICE, the Project Team will derive **The SPICE Dataset**.

6.7 After publication of the **SPICE Final Report**, WMO will keep a copy of **The SPICE Dataset** and make it available to whoever may request it, subject to their agreement in writing to abide by this SPICE Data Protocol.

7 Publications and Presentations

In the following, the word "publication" is used for publications as well as for presentations made at conferences (national and international)

7.1 WMO may publish in the **SPICE Final Report** part or all of **The SPICE Dataset.**

7.2 The IOC may develop and approve a set of slides that will be made available to the IOC, the Project Team and Site Teams for general use in presentations on SPICE.

7.3 All reports, presentations and publications using part or all of **The SPICE Dataset**, either before or after the publication of the **SPICE Final Report**, shall acknowledge SPICE as the source of the data. They should also include the general disclaimer: "

> Results presented in this work were obtained as part of the Solid Precipitation Inter-Comparison Experiment (SPICE), conducted on behalf of the World Meteorological Organization (WMO) Commission for Instruments and Methods of Observation (CIMO). The analysis and views described herein are those of the author(s) at this time, and do not necessarily represent the official outcome of WMO SPICE. Mention of commercial companies or products is solely for the purposes of information and assessment within the scope of the present work, and does not constitute an endorsement by the author(s) or WMO.

7.4 Site Team(s) are free to publish results from single- or multiple-site experiments that were underway prior to the commencement of SPICE.

7.5 Site Teams, with the permission of their Site Manager, may analyse their **SPICE Site Datasets** and publish this work**,** prior to the publication of the **SPICE Final Report**, addressing instruments that they own.

7.6 Site Teams, with the permission of their Site Manager, may also publish results of instruments provided to them in the context of SPICE by **Instrument Providers.** However, these **Instrument Providers** shall be invited, through the relevant Site Manager, to provide comments on the planned publication(s) and be given a reasonable time to reply to ensure that the results are fairly reported and correspond to the proper use of the instruments. Site Teams shall consider those comments in finalizing their publication(s).

7.7 Site Teams are encouraged to follow the guidelines provided in Annex C and to share their draft publications with the Project Team.

7.8 Site Managers shall notify the IOC of all reports, presentations and publications made using part or all of **The SPICE Dataset**, to ensure their appropriate inclusion, consideration, and citation in the **SPICE Final Report**.

7.9 Each **Instrument Provider** will be provided with an opportunity to review the analysis and assessment results presented in the draft final report for the instrument(s) it provided, and each will be given a reasonable time to provide comments on the draft final report. Any feedback shall be included in the **SPICE Final Report**.

7.10 All SPICE participants and those subsequently accessing the SPICE Dataset agree to use the data, Final Report and related publications based on SPICE data solely for the purpose of scientific research and development and not in order to make comparative statements to gain commercial advantage.

ANNEX A

SPICE AND ITS OBJECTIVES

Mission statement

To recommend appropriate automated field reference system(s) for the unattended measurement of solid precipitation in a range of cold climates and seasons, and to provide guidance on the performance of modern automated systems for measuring: (i) total precipitation amount in cold climates for all seasons, especially when the precipitation is solid, (ii) snowfall (height of new fallen snow), and (iii) snow depth.

To understand and document the differences between an automatic field reference system and different automatic systems, and between automatic and manual measurements of solid precipitation using equally exposed/shielded gauges, including their siting and configuration.

Scope and Definition

Building on the results and recommendations of previous intercomparisons, the WMO Solid Precipitation Intercomparison Experiment (SPICE) will focus on the performance of modern automated sensors measuring solid precipitation. SPICE will investigate and report the measurement and reporting of the following parameters:

With highest priority:

a. Precipitation amount, over various time periods (minutes, hours, days, season), as a function of precipitation phase (liquid, solid, mixed);
b. Snow on the ground (snow depth); as snow depth measurements are closely tied to snowfall measurements, the intercomparison will address the linkages between them.

With lower priority:

c. Solid and mixed precipitation intensity.

As a key outcome, recommendations will be made to WMO Members, WMO programmes, manufacturers and the scientific community, on the ability to accurately measure solid precipitation, on the use of automatic instruments, and the improvements possible. The results of the experiment will inform those Members that wish to automate their manual observations.

Intercomparison Objectives

WMO-SPICE will focus on the following key objectives:

I. Recommend appropriate automated field reference system(s) for the unattended measurement of solid precipitation. Define and validate one or more field references using automatic instruments for each parameter being investigated, over a range of temporal resolutions (e.g. from daily to minutes).

II. Assess/characterize automatic systems (both the hardware and the associated processing) used in operational applications for the measurement of Solid Precipitation (i.e. gauges as “black boxes”):

a. Assess the ability of operational automatic systems to robustly perform over a range of operating conditions;
b. Derive adjustments to be applied to measurements from operational automatic systems, as a function of variables available at an operational site: e.g., wind, temp, RH;
c. Make recommendations on the required ancillary data to enable the derivation of adjustments to be applied to data from operational sites on a regular basis, potentially, in real-time or near real-time;

d. Assess operational data processing and data quality management techniques;
e. Assess the minimum practicable temporal resolution for reporting a valid solid precipitation measurement (amount, snowfall, and snow depth on the ground);
f. Evaluate the ability to detect and measure trace to light precipitation.

III. Provide recommendations on best practices and configurations for measurement systems in operational environments:

a. On the exposure and siting specific to various types of instruments;
b. On the optimal gauge and shield combination for each type of measurement, for different collection conditions/climates (e.g., arctic, prairie, coastal snows, windy, mixed conditions);
c. On instrument specific operational aspects, specific to cold conditions: use of heating, use of antifreeze (evaluation based on its hygroscopic properties and composition to meet operational requirements);
d. On instruments and their power management requirements needed to provide valid measurements in harsh environments;
e. on the use of visibility to estimate snowfall intensity
f. On appropriate target(s) under snow depth measuring sensors;
g. Consideration will be given to the needs of remote locations, in particular those with power and/or communications limitations.

IV. Assess the achievable uncertainty of the measurement systems evaluated during SPICE and their ability to effectively accurately report solid precipitation.

a. Assess the sensitivity, uncertainty, bias, repeatability, and response time of operational and emerging automatic systems;
b. Assess and report on the sources and magnitude of errors including instrument (sensor), exposure (shielding), environment (temperature, wind, microphysics, snow particle and snow fall density), data collection and associated processing algorithms with respect to sampling, averaging, filtering, and reporting.

V. Evaluate new and emerging technology for the measurement of solid precipitation (e.g. non-catchment type), and their potential for use in operational applications.

VI. Configure and collect a comprehensive data set for further data mining or for specific applications. Enable additional studies on the homogenization of automatic/manual observations and the traceability of automated measurements to manual measurements.

Annex B

GLOSSARY OF TERMS

Datasets:

SPICE Site Dataset: A dataset comprising all level datasets from that Intercomparison Site.

SPICE Intercomparison Dataset: this is the dataset that combines the data from all SPICE intercomparison sites and from all instruments. The **Project Team** will develop the **SPICE Intercomparison Dataset** using the datasets from each **Intercomparison Site** and performing additional analysis on them**.**

The SPICE Dataset: The total **SPICE** dataset including all **SPICE Site Datasets**, **Site Documentation** and **Instrument Documentation** for all participating sites and instruments, the **SPICE Intercomparison Dataset,** and all SPICE analysis and assessment documentation.

SPICE Site Final Report: The final report for SPICE from an Intercomparison Site, derived from all relevant data and information from that site.

SPICE Final Report: The Final Report on SPICE, derived from **The SPICE Dataset.**

Documentation:

Instrument Documentation: Documentation prepared and provided to the IOC by an Instrument Provider, which includes a description of the instrument proposed in terms of performance, principle of operation, data format, installation requirements, interfaces and synchronization.

Instrument Providers: Manufacturers or WMO Members that provide instruments for SPICE but who will not be hosting a SPICE Intercomparison Site(s).

Site Documentation: Documentation prepared and provided to the IOC by an Intercomparison Site, which includes a description of the proposed host site, its location, capacity, the data acquisition system, data acquisition protocol, data archive and data quality control system available to support SPICE.

Annex C

Guidelines for publications:

- All publications shall include the disclaimer provided in the main part of the SPICE Data Protocol.

- Site teams are encouraged to analyse and publish preliminary and partial results of their sites in advance of the Final Report, thus preparing the ground for the SPICE cross-sites analysis.

- Co-authorship of publications is highly encouraged and could include all contributors

- All publications prepared using SPICE data sets, partially of entirely, should include a section describing the configuration of the experiment, the results of which are included in the publication, indicating:
    - Instruments used, whether part of SPICE , and the instrument ownership;
    - The field configuration of instruments in the experiment (siting, windshield, heating, data logger used)
    - Processing of the instrument output into data used for the work presented.
    - Any information, specific to the experiment, and relevant to the work presented.
    - Any exceptions from the recommended practices regarding the use and configuration of the instruments.
    - This information should be reflective of the site participation in SPICE.

- If reference is made to an instrument make and model, the author(s) should:
    - Identify how the instruments is used for meeting the objectives of work presented in the publication;
    - avoid any comparative and generalized statements, as well as broad qualifiers which would be perceived as ranking instruments (e.g. better, worse.);
    - When comparative assessments are required as part of the results presented, the author(s) will indicate the context in which these assessments are made; e.g. a specific application for national purposes, outside the scope of SPICE, studies of a specific feature relevant to the scope of work (heating, shielding, siting, etc), preliminary studies on a SPICE topic, presented from a site only, etc.;
    - Reflect as far as possible the comments of the involved parties in the subproject.

- Authors should avoid making statements that would imply acceptance or rejection of instruments or inferring any conclusions for the Final Report.

Site Description: Guthega Dam, Australia

SECTION 1: Site Location

Country	AUSTRALIA
Site name	Guthega Dam
latitude	36.3773° S
longitude	148.3706° E
elevation (m)	1586

Site Manager: Shane Bilish (shane.bilish@snowyhydro.com.au**)**

Figure 1: site picture

WMO SPICE / Site objectives:

Meteorological observations have been made at this site since 1953. Multiple automated precipitation gauges have been present since 2006 and the site hosts two wind fences based on the DFIR design: one meeting the exact specifications as well as a second with a diameter of 6 m, hereafter known as the Half DFIR. Historically, the primary objective of intercomparisons at the site was to develop wind speed vs catch efficiency relationships for snowfall measured with the ETI NOAH II precipitation gauges in various wind shield configurations; see for example Chubb et al. (2015), DOI: 10.1175/JHM-D-14-0216.1. Geonor T-200B gauges have been installed as reference gauges to allow participation in SPICE.

Site Description

The Guthega Dam weather station, located within Kosciuszko National Park, New South Wales, Australia, was proposed as a SPICE experimental site by Snowy Hydro Limited, the operator of the station. The proposal is endorsed by the Bureau of Meteorology of Australia.

The site is located opposite Guthega village, a winter ski resort, and approximately 40 km from Jindabyne, the nearest town. It lies adjacent to Guthega Pondage, which has a surface area of 17 ha at Full Supply Level and is generally frozen during winter. The station is approximately 10 m higher in elevation than the pondage, which lies at a junction of valleys aligned to the N, NE and SE. Vegetation in the immediate vicinity is predominantly shrubs and grasses. Peaks rise approximately 200 m above the site within 1 km.

Data from this site were used in the Snowy Precipitation Enhancement Research Project (SPERP); see Manton et al. (2011), DOI: 10.1175/2011JAMC2659.1.

SECTION 2: Site pictures and layout

Site pictures

Figure 2. From West

Figure 3. From South

Figure 4. From East

Figure 5. From North

Overview of **Australia** SPICE site **Guthega Dam**

Site Layout

Figure 6. Site Layout of **Guthega Dam**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Guthega Dam** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project.

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer	Project
Geonor T-200BM3 1000mm	**WG**	**R3**	**SA**	**1 min**	**Y**	**3**	**Precipitation**	**Geonor**	**SPICE**
Geonor T-200BM3 1000mm	**WG**	**R3**	**0**	**1 min**	**Y**	**3**	**Precipitation**	**Geonor**	**SPICE**
ETI NOAH II	WG	UT	DFIR	1 min	Y	3	Precipitation	ETI	SITE
ETI NOAH II	WG	UT	1/2 DFIR	1 min	N	3	Precipitation	ETI	SITE
ETI NOAH II	WG	UT	SA	1 min	N	3	Precipitation	ETI	SITE
Thies	OD	R3	DFIR	10 min	Y	3.1	Precipitation occurrence	Thies	SPICE
NRG IceFree3 anemometer		Anc.	0	10 min	Y	3	Wind speed	NRG	
NRG IceFree3 anemometer		Anc.	1/2 DFIR	10 min	Y	3	Wind speed	NRG	
NRG IceFree3 anemometer		Anc.	outside Half DFIR	10 min	Y	3	Wind speed	NRG	
NRG IceFree3 wind vane		Anc.	0	10 min	Y	3	Wind Direction	NRG	
Vaisala WAA252 anemometer		Anc.		10 min	Y	10	Wind speed	Vaisala	
Vaisala WAV252 wind vane		Anc.		10 min	Y	10	Wind Direction	Vaisala	
Vaisala HMP45A T+RH probe		Anc.		10 min	n/a	2.7	Air Temperature and humidity	Vaisala	
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height (m)**	**Parameter**	**Manufacturer**	**Project**

NB: * = instruments operating prior to the successful commissioning of the R3 instruments on 02 April 2014

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: 0 = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Bratt's Lake, Canada

SECTION 1: Site Location

Country	CANADA
Site name	Bratt's Lake
latitude	50.200531° N
longitude	104.711299° W
elevation (m)	585m

Site Manager: Craig Smith (craig.smith2@canada.ca)

Figure 1: site picture

WMO SPICE / Site objectives:

The intercomparison facility has been operating since 2004 and hosts two DFIR wind fences. Historically, the objective of intercomparisons at the site were to develop wind speed vs catch efficiency relationships, for snowfall measured with the Geonor T-200B precipitation gauges in various wind shield configurations

Site Description

Environment Canada proposed the experimental site of Bratt's Lake as a S2 site type, as a result of the availability of two DFIR-fences with automatic gauges. The Bratt's Lake precipitation intercomparison facility is located 36 km south of Regina, Saskatchewan, Canada. The site is in an agricultural landscape with surrounding vegetation typically less than 1m in height. The topographical relief is small, with elevation changing less than 1 m per km.

SECTION 2: Site pictures and layout

Site pictures

Figure 2. North

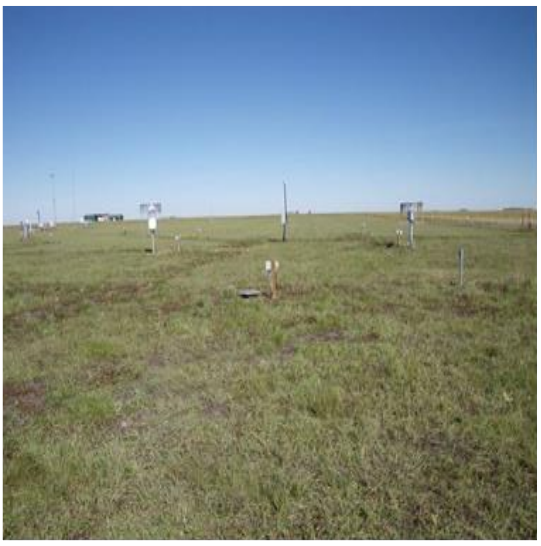


Figure 3. East

Figure 4. Northwest

Figure 5. Southwest

Overview of **Canada** SPICE site **Bratt's Lake**

Site Layout

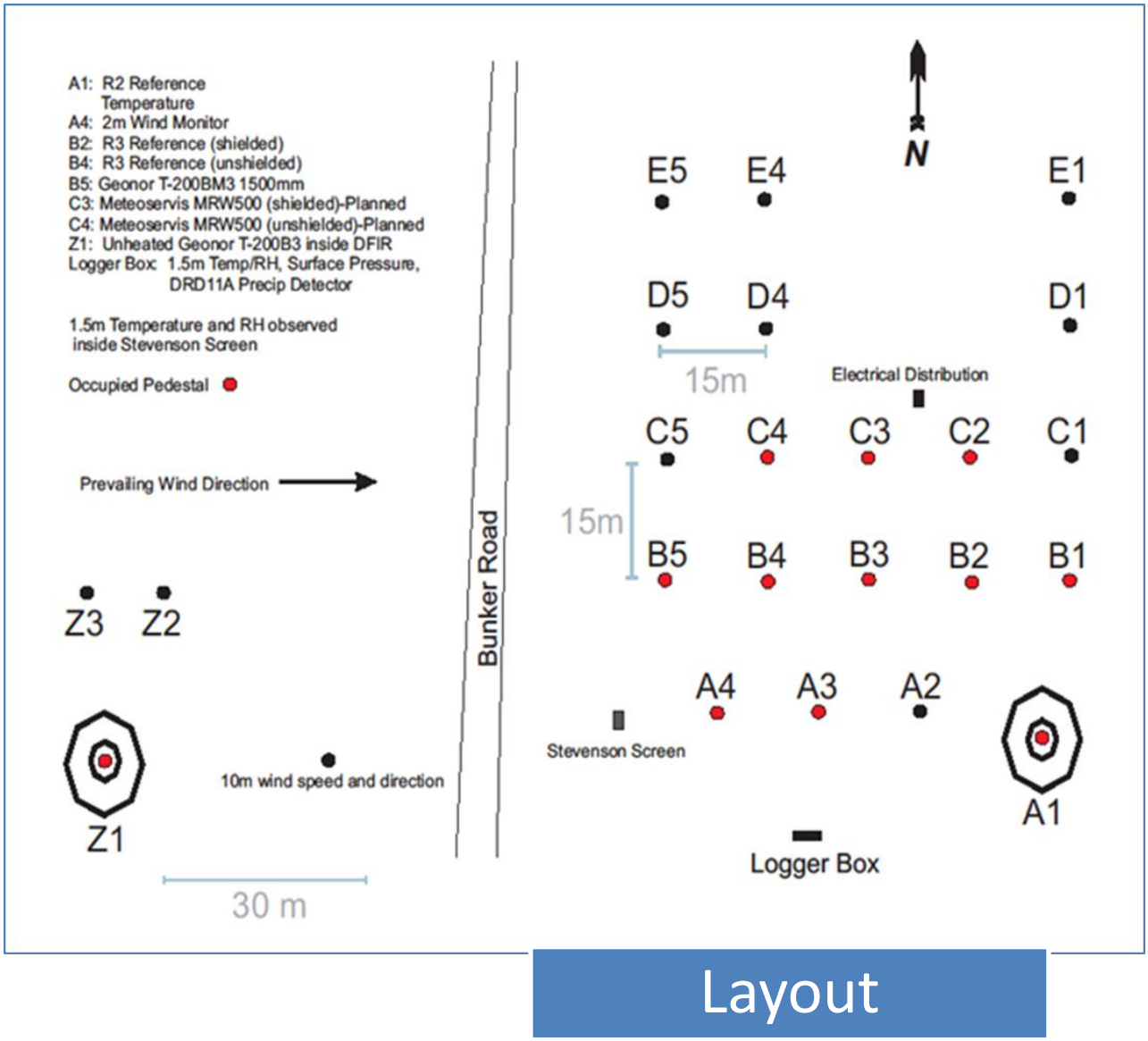


Figure 6. Site Layout of **Bratt's Lake**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Bratt's Lake** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project.

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer	Project
GEONOR T-200B3 600mm (A1)	WG	R2	DFIR	1 min	Y	3	Precipitation	Geonor	SPICE
GEONOR T-200B3 600mm	WG	R2	DFIR	1 min	N	2.9	Precipitation	Geonor	SITE
GEONOR T-200B3 600mm (B2)	WG	R3	Single Alter	1 min	Y	2.2	Precipitation	Geonor	SPICE
GEONOR T-200B3 600mm (B4)	WG	R3	Unshielded	1 min	Y	2.2	Precipitation	Geonor	SPICE
DRD11A Precip. Detector (old)	OD	Anc.	DFIR	1 min	Y	3	Precipitation detection/type	Vaisala	SPICE
Thies Precipitation Sensor (new)	OD	Anc.	DFIR	1 min	Y	?	Precipitation detection/type	Thies	SPICE
Geonor, West (old)	WG	Reference R2-G	DFIR		N		Precipitation	Geonor	SPICE
Pluvio2 on pedestal Z1, West DFIR wind fence (new)	WG	Reference R2-G	DFIR				Precipitation	Ott	SPICE
YES TPS3100 Hotplate	HP	UT	Unshielded	1 min			Precipitation	YES	
Geonor T-200B3 1500mm (B5)	WG	UT	Single Alter	1 min	N	2.2	Precipitation	Geonor	
CS107 (A1)		Anc.		1 min			AirTemp_A1	Campbell	
HMP45C		Anc.		1 min		2	HMP_AirTemp	Vaisala	
HMP45Cy		Anc.		1 min		2	HMP_RelHum	Vaisala	

MetOne inner fence		Anc.	DFIR	1 min			MetOne_Wind Speed_A1	MetOne	
PTB100		Anc.		1 min			Surface Atm. Pressure	Campbell	
RMY5103		Anc.		1 min		2.2	Wind direction	Young	
RMY5103 A4		Anc.		1 min		2.2	Wind speed	Young	
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height**	**Parameter**	**Manufacturer**	**Project**

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: 0 = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: CARE, Canada

SECTION 1: Site Location

Country	CANADA
Site name	CARE
latitude	44° 14' 0.3294" N
longitude	-79° 46' 45.3282" W
elevation (m)	251

Site Manager: Rodica Nitu (rodica.nitu@canada.ca)

SPICE experiments took place at the CARE site (Center for Atmospheric Research and Experiments of the Environment and Climate Change Canada), between Oct. 2013 - Apr. 2014 and Oct. 2014 – Apr. 2015.

Figure 1: CARE Site aerial view (2014)

Site Description

The test facility at the Center for Atmospheric Research and Experiments of the Environment and Climate Change Canada support the national weather monitoring programs with results on the performance of measurement of instruments and methods, prior to network wide deployment. The facility hosts a Reference Climate Station, hosted the NASA GPM Col-Season Precipitation Experiment (GCPEX) in 2012. It also participated in the 1st WMO intercomparison of solid precipitation.
The land area is relatively flat, with good exposure, and the landscape is mostly covered with mosaic vegetation/croplands. The climate is humid continental, with relatively mild winters and moderate amounts of precipitation.

WMO SPICE / Site objectives:

• Define appropriate automated field reference system(s) for the unattended measurement of solid precipitation.
• Characterize automatic systems (both the hardware and the associated processing) used in operational networks, and new and emerging technologies.
• Define best practices and configurations for measurement systems in operational environments.
• Derive adjustment curves to account for gauge undercatch.
• Focus on snow depth measurement and the derivation of snowfall.

SECTION 2: Site pictures and layout

Figure 2: CARE Site: Reference Systems R1 and R2

Figure 3: CARE Site: general view, looking south

Figure 4: CARE Site, General view, looking north

CARE Site Layout

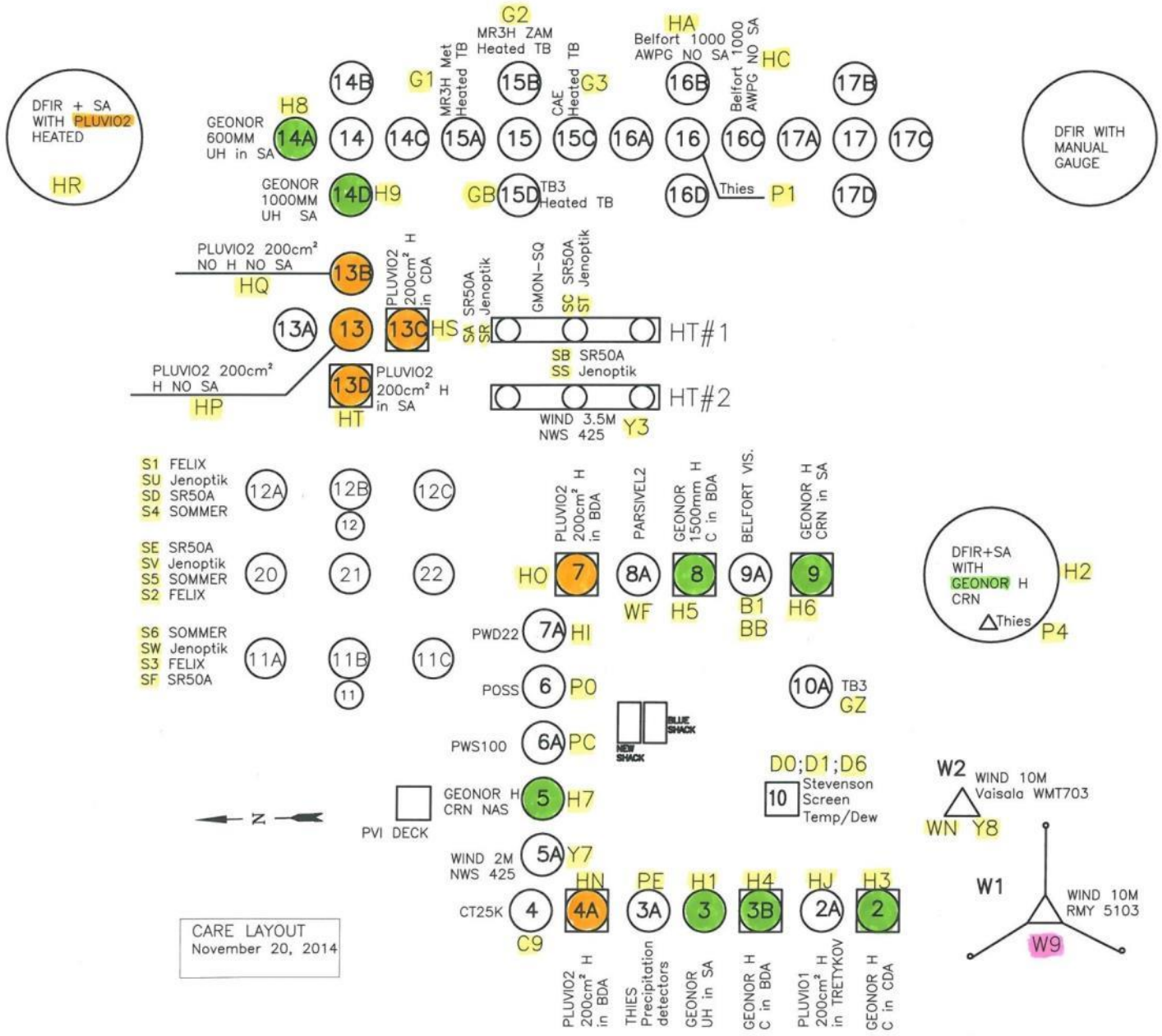


Figure 5. Site Layout of **CARE site (all instruments). The coloured bases show the location of Weighing Gauges type Geonor (green) and Pluvio (orange). Internal codes for data file prefixes are indicated for each instrument, together with basic sensor metadata.**

SECTION 3: Site instrumentation list

Table 1. Instrumentation operational on the CARE site, during the SPICE project, organized by instrument name, type, and manufacturer; by the measured parameter.

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer
Tretyakov gauge	M	R1	DFIR	2 obs/day	N	3	Precipitation	
Geonor T-200B3 600 mm	WG	R2	DFIR	1 min	Y	3	Precipitation	Geonor
Geonor T-200B3 600 mm	WG	R3	Single Alter	1 min	Y	2	Precipitation	Geonor
Geonor T-200B3 600 mm	WG	R3	Unshielded	1 min	Y	2	Precipitation	Geonor
Rain Gauge PMB25R	TB	UT	Unshielded	1 min	Y	1.5	Precipitation	CAE S.p.A.
Heated Tipping Bucket TBH/TBH-LP	TB	UT	Unshielded	6 s	Y	1.5	Precipitation	Hydrological Services Pty. Ltd
MR3H-FC	TB	UT	Unshielded	1 min	Y	1.5	Precipitation	Meteoservis v.o.s.
MR3H-FC ZAMG-version	TB	UT	Unshielded	1 min	Y	1.5	Precipitation	Meteoservis v.o.s.
Pluvio2	WG		Double Alter	6 s	Y	2	Precipitation	OTT Hydromet GmbH
Pluvio2	WG		Double Alter	6 s	Y	2	Precipitation	OTT Hydromet GmbH
Pluvio2	WG		Unshielded	6 s	Y	2	Precipitation	OTT Hydromet GmbH
Pluvio2	WG		Double Alter	6 s	Y	2	Precipitation	OTT Hydromet GmbH
Pluvio2	WG		Single Alter	6 s	Y	2	Precipitation	OTT Hydromet GmbH
Laser Precipitation Monitor	L	Anc.	DFIR	1 min		1,2,3,4 m		THIES Clima
POSS	Microwave radar	Anc.	Unshielded				Precipitation: occurrence, type, accumulation	
R.M. Young	Propeller	Anc.		6 s	N	10	Wind speed	Young

5103 wind monitor	and vane						and direction	
Vaisala HMP155		Anc.		1 min	n/a	1.5	Air temperature, relative humidity	Vaisala
Instrument name	**Type**	**Status**	**Shield**				**Parameter**	**Manufacturer**

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **F** = Frequency; **L** = Laser; **M** = Manual; **C** = Capacitive; **U** = Ultrasonic;

Status: **UT** = Under Test (provided by Manufacturer); **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: 0 = Unshielded; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; SA = Single alter; TRET = Tretyakov; DA = Double Alter

Site Description: Caribou Creek, Canada

SECTION 1: Site Location

Country	CANADA
Site name	Caribou Creek
latitude	53.944737N
longitude	104.649339W
elevation (m)	519 m

Site Manager: Daqing Yang (email: daqing.yang@canada.ca)

Figure 1: Caribou Creek site picture

WMO SPICE / Site objectives:

1. Expand the WMO SPICE concept of experiments to instrument configurations, technologies, environmental conditions, and program objectives specific to the Canadian interests and needs
2. Test and evaluate the DFIR (R2G) and Bush gauge (R0G) references with automatic instruments in Canada
3. Run the site for several years to collect reliable data
4. Link this site/work with the DFIR and Bush gauge evaluation and data collected at Valdai

Site Description

The Caribou Creek intercomparison facility will be located at an Environment Canada research site in the southern boreal forest, approximately 280 km north-east of Saskatoon, Saskatchewan. Predominate vegetation in this region is mature Jack Pine. However, this site was situated in the centre of an area harvested in 2002 with naturally regenerating growth now approaching heights of 2.5 m

Caribou Creek was classified as a S2 site type, as a result of the availability of a DFIR-fence with an automatic gauge. In addition, this site become an S0a reference because of the installation of an automatic gauges, installed in an area with recent jack pine growth maintained at the height of the gauges for the duration of the experiment. This would be a representation of the primary field reference (i.e. the bush gauge) as implemented in Valdai, Russian Federation, but using an automatic gauge instead of a manual gauge.

This facility is unique in that the post-harvest vegetation has grown to the level of 2 to 2.5 m, which is the typical height of a precipitation gauge orifice. The height of the vegetation and the vegetation density make this an ideal site for intercomparisons with a bush shielded gauge. The site will include two bush shielded gauges and a clear-cut area to accommodate a DFIR shielded gauge, as well as up to six additional instruments for intercomparison.

SECTION 2: Site pictures and layout

Site pictures

Figure 2. Looking East

Figure 3. Looking West

Figure 4. Looking North

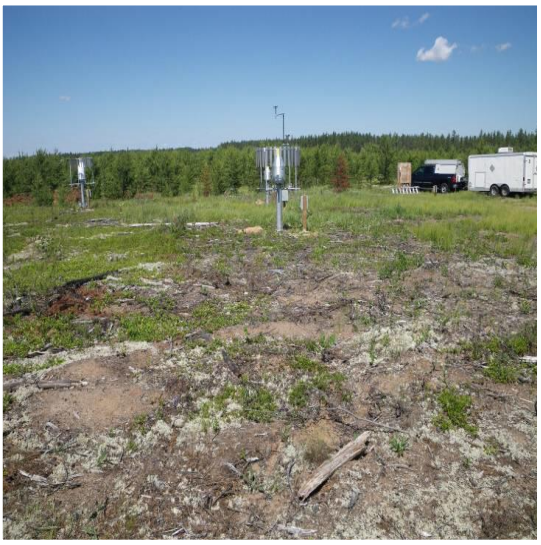


Figure 5. Looking South

Overview of **Canada** SPICE site **Caribou Creek**

Site Layout

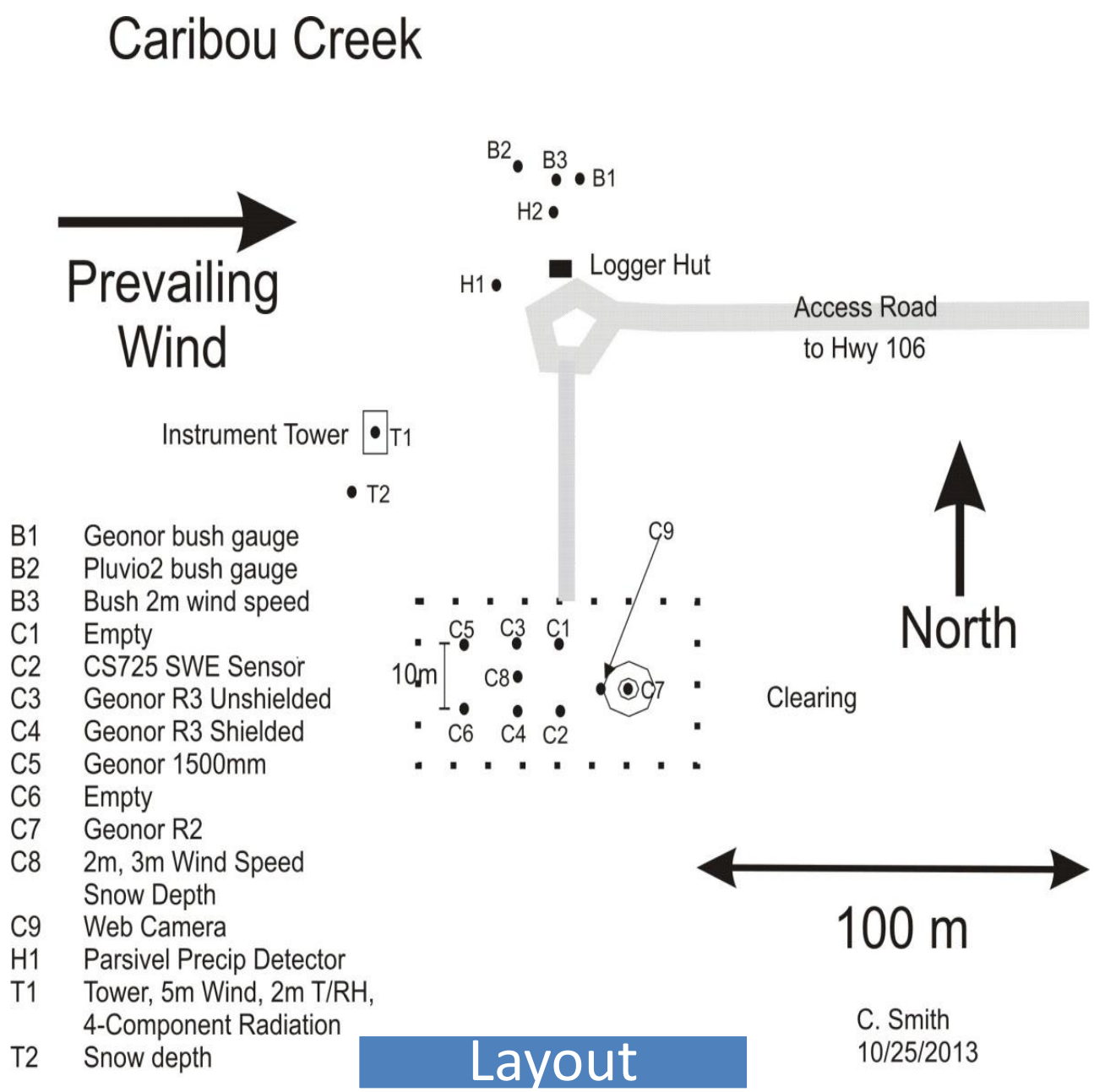


Figure 6. Site Layout of **Caribou Creek**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Caribou Creek** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project.

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height	Parameter	Manufacturer	Project
Pluvio2	WG	**R0a**	**Single Alter in bush**	1 min	Y	2	Precipitation	OTT	SPICE
T-200BM3 1500mm	WG	**R0a**	**Bush**	1 min	Y	2	Precipitation	Geonor	SPICE
Geonor	WG	**R2**	DFIR	20 s	Y	3	Precipitation	Geonor	SITE
Geonor	WG	**R3**	Shielded		Y	2	Precipitation	Geonor	SITE
Geonor	WG	**R3**	Unshielded		Y	2	Precipitation	Geonor	SITE
Thies Precipitation Sensor	OD	Anc.	DFIR				Precipitation type	Thies	SITE
DRD11A Precip Ddetector	OD	Anc.	DFIR	1 min		2.25	Precipitation type	Vaisala?	SITE
Geonor T-200B3 1500mm	WG	UT	Single Alter		N	2.25	Precipitation	Geonor	From manufacturer C5 pedestal
Parsivel disdrometer	OD	UT	Unshielded				Precipitation type	OTT	SITE
CS725 SWE sensor		UT	n/a	6 hour	N	2	Snow Water Equivalent	Campbell	SITE on Pedestal C2
Cup Wheel 2m		Anc.		20 s	N	2	Wind speed	MetOne	at C8
Cup Wheel 3m		Anc.		20 s	N	3	Wind speed	MetOne	at C8
Cup Wheel 2m		Anc.		20 s	N	2	Wind Dir	MetOne	near Bush gauges
Vaisala HMP45A T+RH probe		Anc.		20 s	n/a	1.5	Air Temp. and humidity	Vaisala	Clearing
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height**	**Parameter**	**Manufacturer**	**Project**

Type: WG = Weighing gauge; TB = Tipping bucket; OD = Optical detector

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: 0 = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Tapado, Chile

SECTION 1: Site Location

Country	CHILE
Site name	Tapado
latitude	30º 9' 30" S
longitude	69º 54' 30" W
elevation (m)	4318

Site Manager: Shelley MacDonnell (shelley.macdonell@gmail.com**)**

Figure 1: site picture

WMO SPICE / Site objectives:

The expansion of the network will give a better overview of snow distribution and event size throughout the region.

Site Description

The site is four hour drive from La Serena city; the access is by 4WD vehicles, mules. The road to the site is generally closed in the winter; can be opened on request.

The site is visited monthly during the summer (November - April). In a real emergency, can apply to access the site during the winter, depending on road conditions. No personnel are available on site.

The station belongs to the regional meteorological station network.

The location is unique because it offers the ability to understand snow processes in a high elevation, semi-arid region where snow contributes the bulk of the water available in the wider catchment.

The site was primarily used as part of a glacier research program. Two projects are currently operational:

- Modelling glacier meltwater production in the dry Andes
- Characterization and monitoring of rock glaciers in the Elqui River catchment, and mass balance of the Tapado Glacier

CEAZA (Centro de Estudios Avanzados en Zonas Áridas) has a regional network of meteorological stations. By June 2013, 35 stations should be operational, five of which will be above 2500 m.

The site is approximately 14 km far to the SW large body of water; there are no obstacles within a 100 m radius from the station. There is rising cirque wall 400 m to the south of the station, a debris-covered glacier tongue 400 m to the north, the eastern valley wall is 700 m away, and the western wall is 600 m away. 1700 m to the Tapado Glacier terminus.

The terrain is dry without vegetation, steppe and disparate short mountain vegetation. The immediate slope is close to zero.

SECTION 2: Site pictures and layout

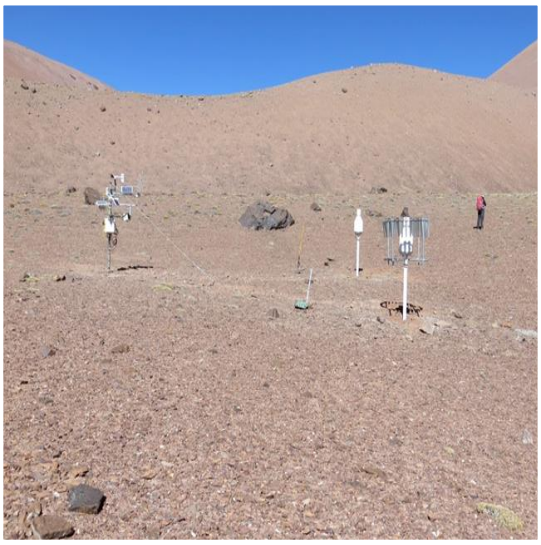
Site pictures

Figure 2. From West

Figure 3. From South

Figure 4. From East

Figure 5. From North



Overview of **CHILE** SPICE site **Tapado**

Site Layout

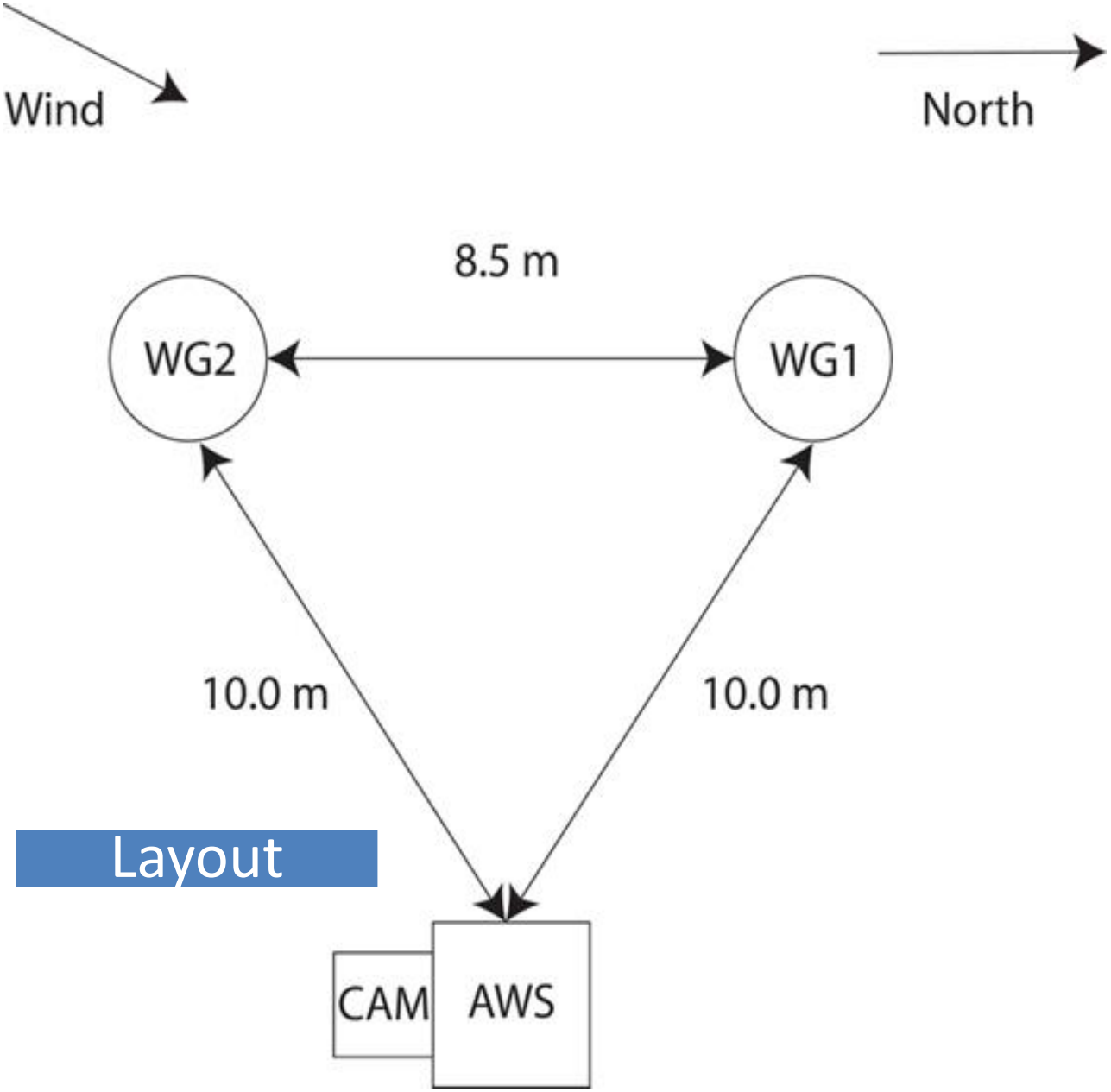


Figure 6. Site Layout of **Tapado**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Tapado** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project.

Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height (m)**	**Parameter**	**Manufacturer**	**Project**
Geonor T-200BM3 1000mm	**WG**	**R3**	**SA**	**10s**	**N**	**2.15**	**Precipitation**	**Geonor**	**SITE**
Geonor T-200BM3 1000mm	**WG**	**R3**	**Un**	**10s**	**N**	**2.15**	**Precipitation**	**Geonor**	**SITE**
CS-106		Anc.		15 min		1.50	Atm. Pressure	Campbell	SITE
HMP45 T+RH probe		Anc.		15 min	n/a	2.10	Air Temp. and hum.	Vaisala	SITE
05103 windmonitor		Anc.		15 min	N	2.50	Wind speed & dir.	Young	SITE
CNR4 radiometer		Anc.		15 min	Y	2.13	In&out Long&short radiation	Kipp&Zonen	SITE
SR50		Anc.		15 min	Y	1.64	Snow depth	Campbell	SITE
Automatic camera		Anc.		1/day			.JPG image	Harbortronics	SITE
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height (m)**	**Parameter**	**Manufacturer**	**Project**

NB: * = instruments operating prior to the successful commissioning of the R3 instruments on 02 April 2014

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: Un = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Sodankylä, Finland

SECTION 1: Site Location

Country	**FINLAND**
Site name	**Sodankylä**
latitude	**67° 21' 59.87"**
longitude	**26° 37' 44.44"**
elevation (m)	**179 m**

Site Manager: Anna Kontu, email: Anna.Kontu@fmi.fi

Figure 1. Overview of Sodankylä SPICE site (13 December 2015)

WMO SPICE / Site objectives:

To find WMO recommendations for automated precipitation and snow measuring instruments. Globally, 15 countries and 20 sites have participated in WMO SPICE.
All the documented Spice objectives are valid and important for FMI, but the most important are:

II. Assess/characterize automatic systems (both the hardware and the associated processing) used in operational applications for the measurement of Solid Precipitation (i.e. gauges as "black boxes"):
a. Assess the ability of operational automatic systems to robustly perform over a range of operating conditions;
d. Assess operational data processing and data quality management techniques;
e. Assess the minimum practicable temporal resolution for reporting a valid solid precipitation measurement (amount, snowfall, and snow depth on the ground);

IV. Assess the achievable uncertainty of the measurement systems included in SPICE and the ability to effectively report solid precipitation:
a. Assess the sensitivity, uncertainty, bias, repeatability, and response time of operational and emerging automatic systems;

V. Evaluate new and emerging technology for the measurement of solid precipitation (e.g. noncatchment type), and their potential for use in operational applications.

Primary focus desired:
Research and development of satellite data calibration and validation methods and instrumentation.
Sodankylä ARC has a strong background in the field of snow research in a harsh arctic environment.
Sodankylä ARC's first priority is snow and hydrology research and their climatological impact. (*From IOC-SPICE-2.pdf*)

Site Description

Sodankylä Arctic Research Centre is validated as WMO CIMO-Testbed site, because of the good research and test facilities in the arctic environment. Sodankylä SPICE site is focused on snow on the ground. Construction of Sodankylä SPICE site was started in 2012 and it was ready in the summer of 2013. 9 manufacturers provided their instruments for the site. WMO SPICE measurement campaign was held during winters 2012-2013, 2013-2014 and 2014-2015. Some of the instruments in Sodankylä SPICE site were still measuring in winter 2015-2016. Locations of the instruments are presented in the layout and more specific details of the instruments are presented in Table 1. WMO SPICE Final Report is expected to be published in September 2016. (*From SPICE_Sodankylä_final.docx)*
Sodankylä is located above the Arctic Circle, and is a supersite for satellite data calibration and validation activities, with several on-going programs. It also contributes to the Global Atmospheric Watch (GAW) and Global Cryosphere Watch (GCW). These programs involve the remote sensing of snow, soil moisture, permafrost and atmospheric constituents.

Also on site is the EU SNAPS - Snow, Ice and Avalanche Applications project which will focus on snow and avalanche services for transport infrastructure. Near-real time snow cover maps will be produced and further developed to become an input to snowdrift and avalanche forecasts aimed at e.g. transport authorities.

The Academy of Finland ClimWater Project on site focuses on development of methodology using remote sensing for water circulation in a Boreal region. In Boreal regions, precipitation (water/snow), temperature, ice coverage of the lakes, wind, frost, soil humidity and flood reserves are the key factors that affect water circulation. (*From IOC-SPICE-2.pdf*)

SECTION 2: Site pictures and layout

Site pictures

Figure 2. Towards South-West

Figure 3. Towards West

Figure 4. Towards North

Figure 5. Towards East-West

Overview of Finland SPICE site of **Sodankylä**

Site Layout

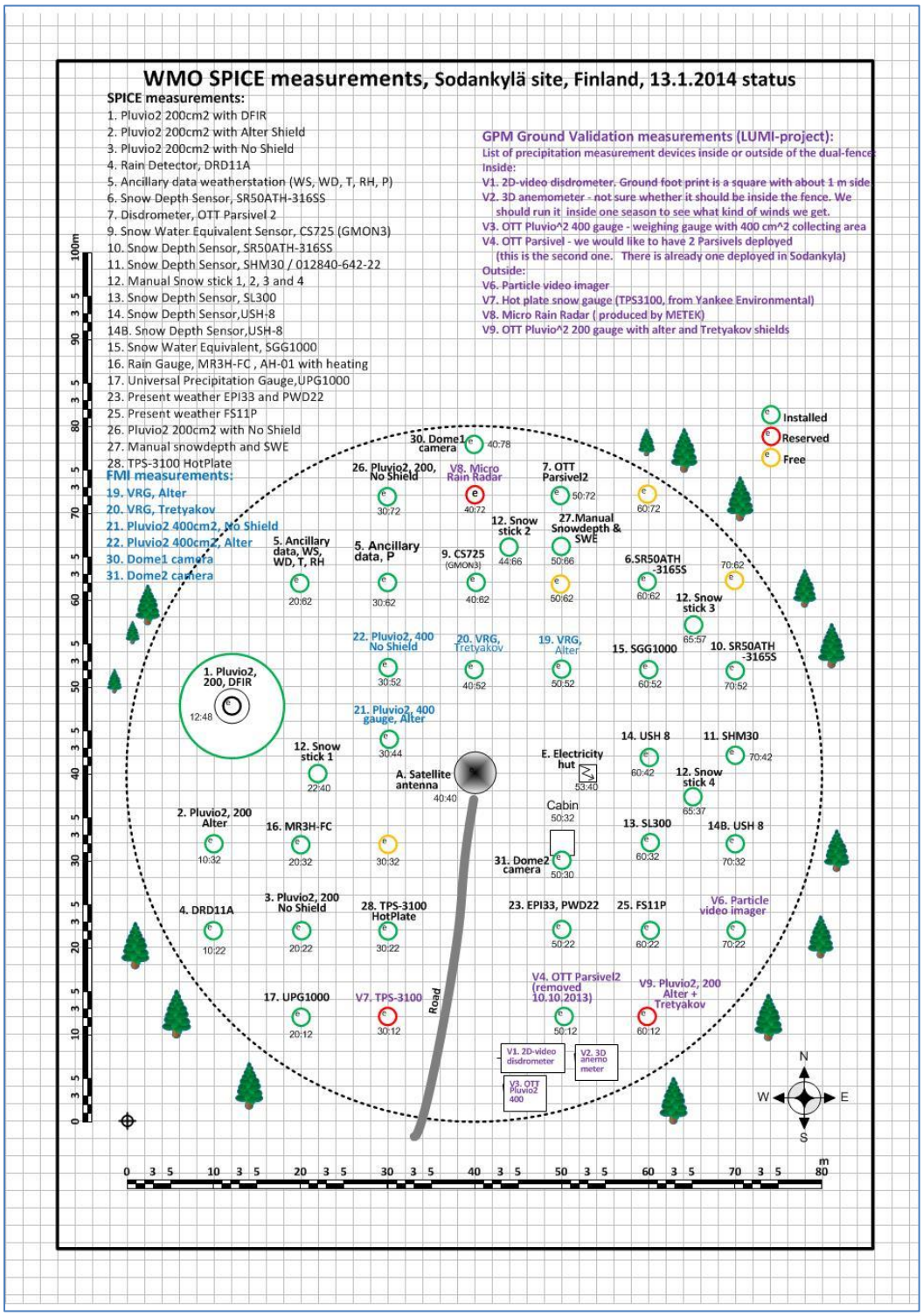


Figure 6. Site Layout of **Sodankylä**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Sodankylä** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project.

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height	Parameter	Manufacturer	Project
Pluvio2 200 cm^2	**WG**	**R2**	**DFIR**	**6 s**	**Y**	**4**	**Precipitation**	**OTT**	**SPICE**
Pluvio2 200 cm^2	**WG**	**R3**	**A**	**6 s**	**Y**	**1.5**	**Precipitation**	**OTT**	**SPICE**
Pluvio2 200 cm^2	**WG**	**R3**	**0**	**6 s**	**Y**	**1.5**	**Precipitation**	**OTT**	**SPICE**
Snow stick	**M**	**R SOG**	**n/a**	**1/day**		**n/a**	**Snow depth**	**FMI**	**SPICE**
UPG1000	TB	UT	Custom	1 min		?	Precipitation	EML	SPICE
MR3H-FC AH-01 heated	TB	UT	0	1 min	Y	2.0	Precipitation	Meteoservis	SPICE
CS725 (GMON3)	SWE	UT	n/a	1 min	Y	1.35	SWE	Campbell	SPICE
SR50 p7052	US	UT	n/a	6h		1.9	Snow depth	Campbell	SPICE
SR50 p6062	US	UT	n/a	1 min	Y	2.0	Snow depth	Campbell	SPICE
TPS-3100	HP	UT	n/a	1 min	Y	2.0	Total precipitation	Yankee Environmental Systems	SPICE
SL300	US	UT	n/a	1 min	Y	2.10	Snow depth	Felix	SPICE
SHM30	Laser	UT	n/a	1 min		1.5	Snow depth	Lufft	SPICE
USH-8 p6042?	US	UT	n/a	1 min		2.0	Snow depth	Sommer	SPICE
USH-8 p7032?	US	UT	n/a	1 min	Y	2.0	Snow depth	Sommer	SPICE
FS11P	OD	UT	n/a	1 min		2.05	Visibility, precipitation type and intensity	Vaisala	SPICE
PWD22 p5022	OD	UT	n/a	1 min		2.05	MOR, visibility, precipitation type and intensity	Vaisala	SPICE
PWD 33 EPI	OD	UT	n/a	15s	?	2.90	Present weather	Vaisala	SPICE
Pluvio2 200 cm^2	WG	UT	0	15s	?	2.7	Precipitation	OTT	FMI
Pluvio2 400 cm^2	WG	UT	0	15s	?	2.7	Precipitation	OTT	FMI
Pluvio2 400 cm^2	WG	UT	A	6s			Precipitation	OTT	

SGG1000	P	UT	n/a	6s			SWE	Sommer	SPICE
Parsivel 2 p1248	OD	Anc.	DFIR	6s			Size and fall velocity of hydrometeors	OTT	FMI
Parsivel 2 p5072	OD	Anc.	0	1 min		0	Size and fall velocity of hydrometeors	OTT	FMI
HMP155 p2062		Anc.	n/a	1 min		2	Air temperature and humidity	Vaisala	FMI
Thies 2D p2062 3.5m		Anc.	n/a	1 min		2.5m (3.5m)	Wind speed	Thies	FMI
Thies 2D p3032 1.5m (near DFIR)		Anc.	n/a	1 min			Wind speed	Thies	FMI
PTB220 p3062		Anc.	n/a	1 min		1	Atmospheric pressure	Vaisala	FMI
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height**	**Parameter**	**Manufacturer**	**Project**

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: 0 = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Col de Porte, France

SECTION 1: Site Location

Country	France
Site name	Col de Porte
latitude	45.30 N
longitude	5.77 W
elevation (m)	1325 m

Site Manager: Samuel Morin (samuel.morin@meteo.fr)

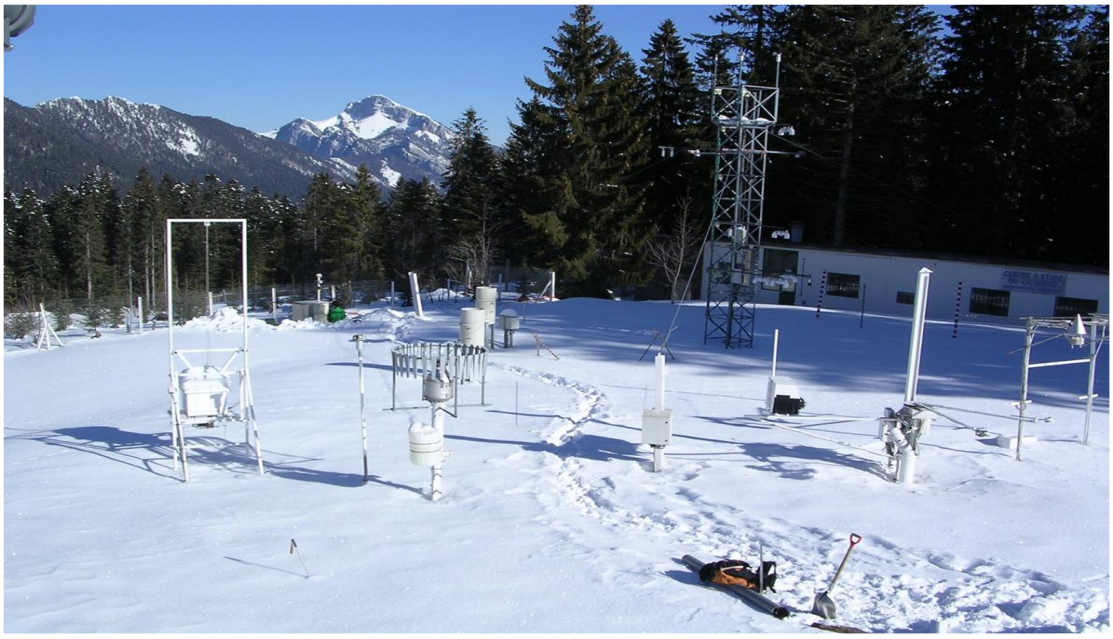


Figure 1: Col de Porte site

WMO SPICE / Site objectives:

MOTIVATION:

- Snow precipitation observations are key for many operational and research applications of Météo-France and its partners, which is the main motivation for participating to WMO-SPICE.
- The Col de Porte site has a long history of snow and meteorological observations in mountain environment.
- It serves as a test bed for meteorological and snow observations systems used in mountain environments by Météo-France (until recently under the direct responsibility of CNRM-GAME/CEN, now formally transferred to the operational observation department of Météo-France (DSO)).
- It is used by CNRM-GAME/CEN teams and local research groups to test instruments and to carry out research projects requiring continuous monitoring of the meteorological and snow conditions.
- Furthermore, the site has been pivotal in the development of detailed snowpack models used for various applications.
- The participation of CDP to the WMO-SPICE initiative corresponds perfectly to the goals and purposes of this site.

SPICE objectives would prefer to focus on: Updated WMO guidelines for operational automatic snow/rain precipitation measurements.

Site Description

The site is located in a subalpine grassy meadow and thus immediately surrounded within a few meters of horizontal distance to a coniferous forest

The total size of the reported area is ~2000m2. The contours of the measurement field are approximately rectangular and large side is north-south oriented, 55 m (N-S) x 36 m (E-W). In the layout distances are approximately reported to scale and indicated in cm. Prevailing wind direction is along the South/North axis.

The contribution of SPICE to the WMO-SPICE projects is undertaken by several academic and nonacademic partners located in Grenoble, France.

SECTION 2: Site pictures and layout

Site pictures

Figure 2. View to North

Figure 3. View to South

Figure 4. View to West

Figure 5. View to East

Overview of **France** SPICE site **Col de Porte**

Site Layout

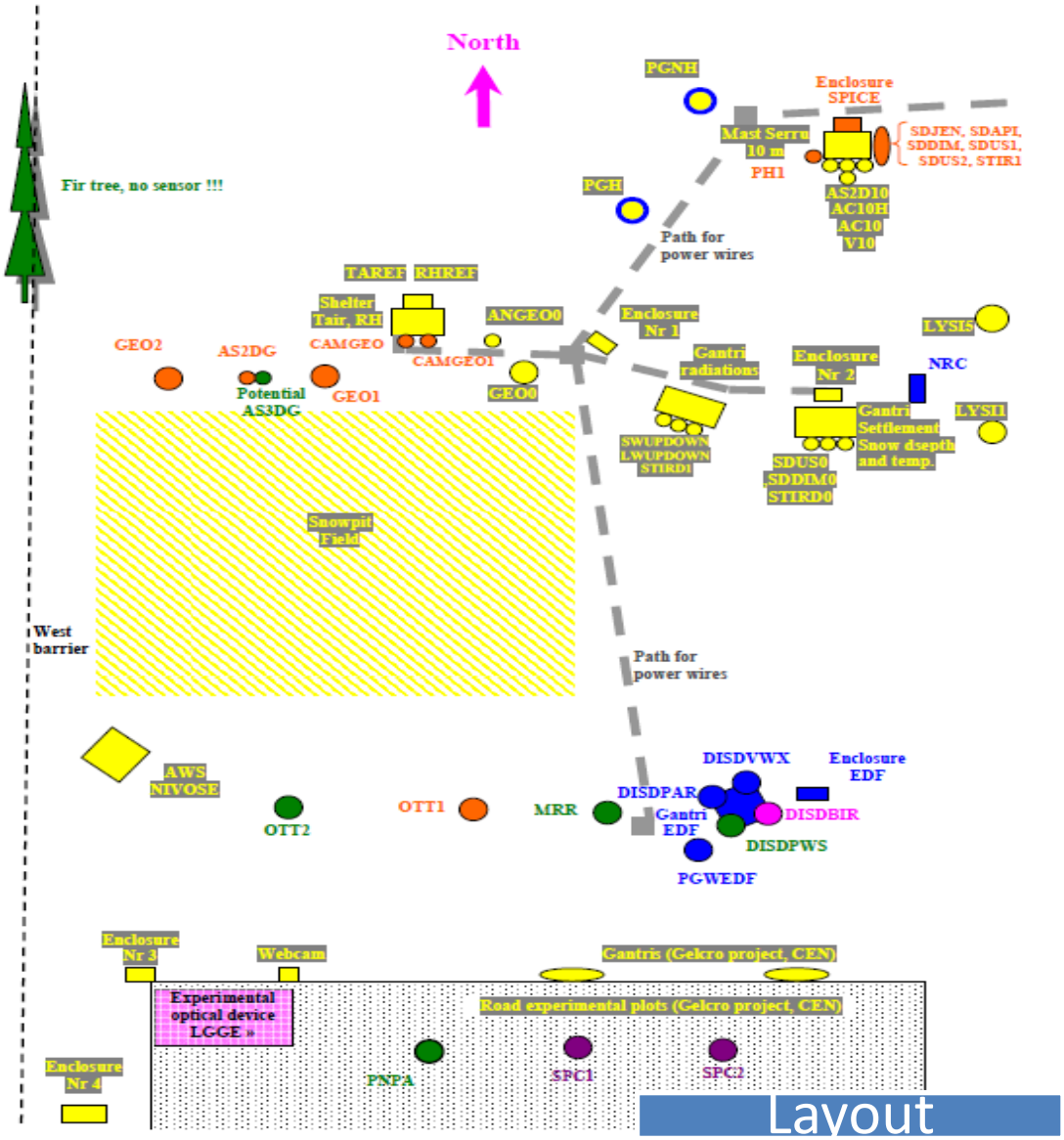


Figure 6. Site Layout of **Col de Porte**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Col de Porte** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project.

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer	Project
R3 Geonor T200-B3 (600 mm)	WG	R3	SA	1min	Y	3.09	Precipitation	Geonor	SPICE
R3 Geonor T200-B3 (600 mm)	WG	R3	Un	1min	Y	3.09	Precipitation	Geonor	SPICE
TLN35-R 20deg angle (corrected)	Laser	UT	n/a	1min	Y	4.00	Snow depth	APICAL	Site
FLS-CH 10 20deg angle (corrected)	Laser	UT	n/a	1min	Y	4.00	Snow Depth	DIMETIX	Site
DIMETIX laser sensor vertical	Laser	UT	n/a	1hour	Y	3.00	Snow Depth	DIMETIX	Site
Jenoptik SHM30 laser sensor 20deg angle (corrected)	Laser	UT	n/a	1min	Y	4.00	Snow Depth	Jenoptik	SPICE
SR50A Ultrasonic	US	UT	n/a	1min	N	4.00	Snow Depth	Campbell	Site
SR50AT	US	UT	n/a	1min	Y	4.00	Snow Depth	Campbell	SPICE
Pluvio2 (750mm), 400 cm2	WG	UT	Un	6min	N	3.10	Precipitation	OTT	Site
PWS100	OD	UT	n/a	6min	Y	3.10	Present weather	Campbell	Site
Geonor T200-B1 (600mm), 1 transducer	WG	UT	Single Alter	1hour	Y	3.10	Precipitation	Geonor	Site
PGWEDF, PG2000, 2000 cm2	TB	UT	Un	1hour	Y	2.75	Precipitation	EDF	Site
PG2000 (North). PGH	TB	UT	Un	1hour	Y	2.75	Precipitation	EDF	Site
PG2000 PGWEDF	TB	UT	Un	1hour	Y	3.10	Precipitation	??	Site

SPC1 Précis-Mécanique	TB	UT	Un	1min	N	1.0	Precipitation	Précis-Mécanique	Site
SPC2 Précis-Mécanique	TB	UT	Un	1min	Y	3.10	Precipitation	Précis-Mécanique	Site
Pluvio2 400cm2 (OTT1)	WG	UT	Single Alter	1min	Y	3.10	Precipitation	OTT	Site
Biral VPF730 disdrometer	OD	UT	n/a	1min	Y	3.00	Precipitation	Biral	Site
PNPA 50cm inlet diam.	Tot	UT	Un	10min	N	3.05	Precipitation	Self-made	Site
Thies ultrasonic 2D	US	Anc.	n/a	1min	Y	3.10	Wind speed and direction	Thies	Site
ANGEO0 Laumonier		Anc.	n/a	1s 1h	Y	3.00	Wind speed	Laumonier	Site
STDAir PT100		Anc.	n/a	1min	n/a	4.00	Air Temp.	--	Site
MET Data		Anc.	n/a	1hour	n/a	--	Air Temp. +Rel. Hum.	--	Site
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height (m)**	**Parameter**	**Manufacturer**	**Project**

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: Un = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Forni Glacier, Italy

SECTION 1: Site Location

Country	ITALY
Site name	Forni Glacier
latitude	46°23’56”N
longitude	10°35’25”E
elevation (m)	2631

Site Manager: Guglielmina Diolaiuti (guglielmina.diolaiuti@unimi.it)

Figure 1: site picture

WMO SPICE / Site objectives:

SPICE objectives of primary interest , Ref.: Spice Annex, February 8th 2013:

- Obj. I: Recommend appropriate automated systems for measurement of solid precipitation.
- Obj. II: Assess automatic systems over a range of operating conditions.
- Obj. IV: Assess uncertainty and sources of error
- Obj. VI: Collect data for specific applications (climatology, hydrology, glaciology)

focus on

- Obj. II: Assess automatic systems over a range of operating conditions.
- Obj. VI: Collect data for specific applications (climatology, hydrology, glaciology)
- List existing or planned national projects organized in parallel with SPICE:
- SHARE, SHARE Stelvio, CEOP-GEWEX.

Site Description

This site, at elevation of 2669 m, has been operational since 2005 as the first Italian permanent supraglacial AWS. The Forni Glacier hosts several experiments, in particular devoted to deepen the glaciological issues. The surrounding mountains reach an elevation of 3200-3700 m (the highest is Palon della Mare, 3703 m, then Mount San Matteo, 3678 m), affecting shading conditions on southern and eastern sectors of the glacier, in the Stelvio National Park

The terrain has a rough topography, glacier melting surface, rock exposures occur on limited areas (e.g.: nunataks and rock outcrops). Vegetation is present down valley (c.a. 1 km far by the AWS site). The glacier is about 3 km long, has a northward down-sloping surface, and stretches over an elevation range of 2600 to 3670 m a.s.l.

The site was proposed at the IOC SPICE 4 2013 by EVK2CNR (Italy) on Forni Glacier (Northern Italian Alps, Valtellina) performing the Snow on the Ground measurements (in particular accumulated snow on the ground, and snow water equivalent). Given that SPICE is a demonstration project of the Global Cryosphere Watch (GCW), it could effectively contribute to GCW objectives, and therefore its inclusion within the experiment was highly welcomed. This site will contribute in the assessment of the Snow on the Ground primarily, and should be considered as SPICE S4 sites.

SECTION 2: Site pictures and layout

Site pictures

Figure 2. from South

Figure 3. Snow pit

Figure 4. From East

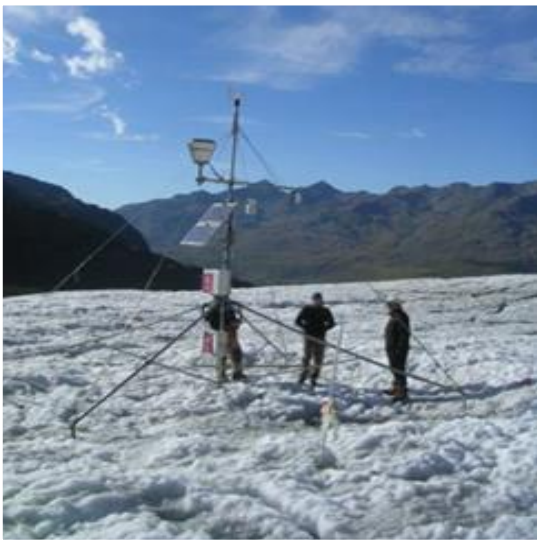


Figure 5. from North-East (after the installation on 6th May 2014)

Overview of **ITALY** SPICE site **Forni glacier**

Site Layout

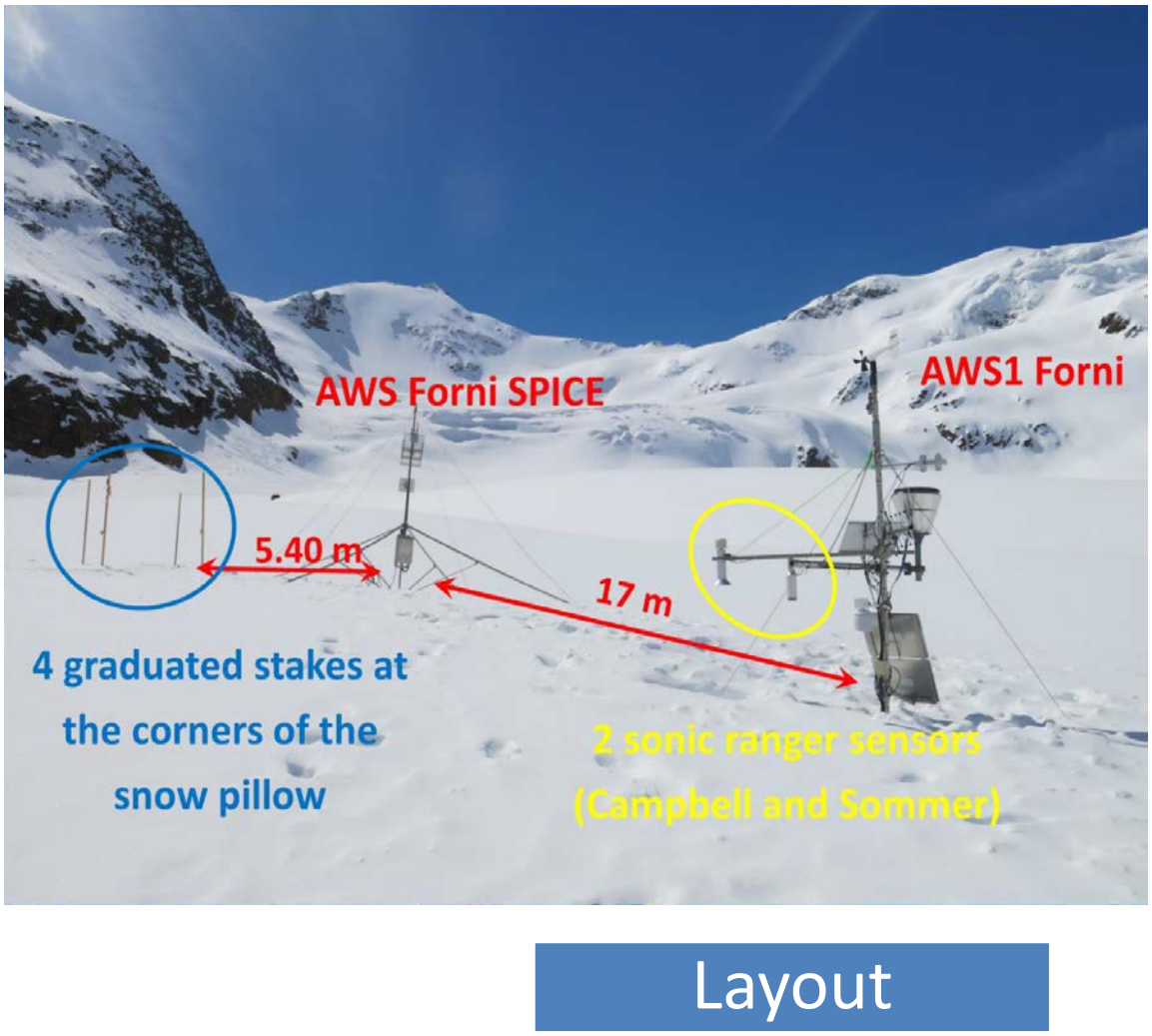


Figure 6. Site Layout of **Forni Glacier**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Forni Glacier** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project.

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer	Project
Snow Pit	**M**	**R SoG**	**n/a**	1/month	n/a	n/a	**Snow density + SWE**	**AINEVA protocol**	**Site**
Snow stakes (x4) + camera	**M**	**R SoG**	**n/a**	1h	n/a	n/a	**Snow depth**		**Site**
Snow weighing tube	**M**	**R SoG**	**n/a**	1/month	n/a	n/a	**Snow density + SWE**	**Enel-Valtecne**	**Site**
Sonic ranger SR50 at 3.17m	US	Anc.	n/a	1h	N	3.17	Snow depth	Campbell	Site
Sonic Ranger USH-8 at 3.15m	US	Anc.	n/a	10 min	N	3.15	Snow depth	Sommer	Site
Snow pillow	Press	Anc.	n/a	10 min	n/a	0	Snow Water Equivalent	Park Mechanical Inc.	Site
Thermo-hygro at 2.6m		Anc.	n/a	30 min	n/a	2.6	Air temperature and humidity	LSI LASTEM	Site
Barometer at 1.5m		Anc.	n/a	1h	n/a	1.5	Atmospheric pressure	LSI LASTEM	Site
Radiometer CNR1		Anc.	n/a	30 min	n/a		Radiation fluxes	Kipp & Zonen	Site
Pluviometer unheated	TB	Anc.	Unshielded	30 min	N	4	Precipitation	LSI LASTEM	Site
Anemometer 05103V		Anc.	n/a	10 min	N	5	Wind speed and direction	Young	Site
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height (m)**	**Parameter**	**Manufacturer**	**Project**

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: 0 = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Joetsu, Japan

SECTION 1: Site Location

Country	**JAPAN**
Site name	**Joetsu**
latitude	**37° 06' 56"**
longitude	**138° 16' 24"**
elevation (m)	**11 m**

Site Manager: Sento Nakai, email: saint@bosai.go.jp

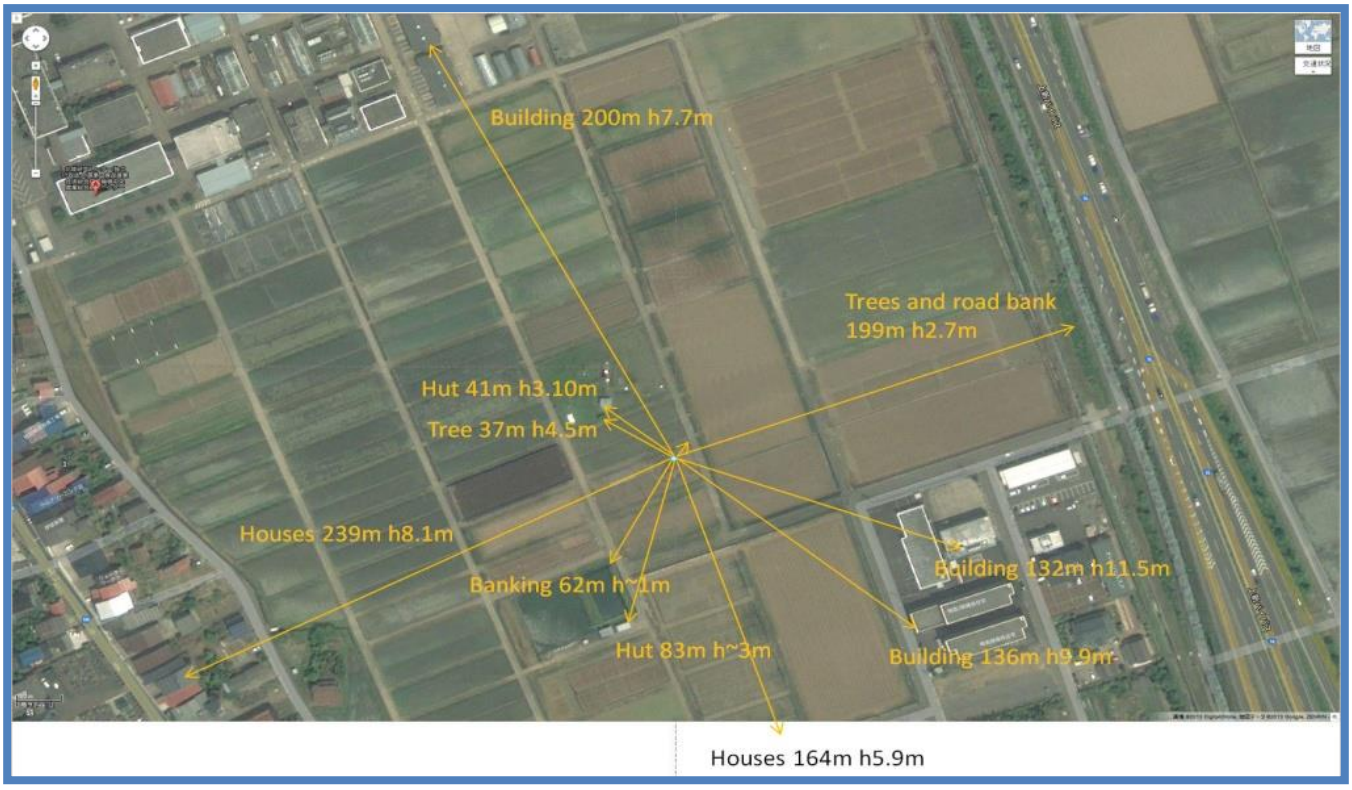


Figure 1. Overview of Joetsu SPICE site

WMO SPICE / Site objectives:

Focus on SPICE objective II. Assess/characterize automatic systems (both the hardware and the associated processing) used in operational applications for the measurement of Solid Precipitation (i.e. gauges as “black boxes”)

Site Description

The Joetsu station is located in the middle of Honshu main island, Joetsu city, Niigata prefecture, Japan (37°06’56’’N, 138°16’23’’E, 10m ASL). The station is about 6 km away from the coast of Sea of Japan, and is surrounded by paddy fields. Due to the winter monsoon wind from the Siberian air mass and the warm current in the Sea of Japan, there is a large amount of precipitation at the site every year. Climate classification of the Joetsu station is Sea of Japan side climate. The area around the Joetsu station is designated in special heavy snowfall zones from the central government of Japan. According to the record from Takada weather station (37°06’24’’N, 138°14’48’’E, 13m ASL) where 2 km apart from the station, the climate normal (1981-2010) of annual mean temperature and precipitation is 13.6 °C and 2755.3 mm, respectively. Snow precipitates from November to April. The climate normal of mean precipitation and maximum snow depth between January and March is 875.3 mm and 122 cm, respectively. The wind speed is not so strong absolutely but is relatively strong in winter. The mean temperature of the coldest month is greater than the freezing point. The ratio of solid precipitation to the winter precipitation varies largely year by year. Thunderstorms and graupel are quite frequent in winter season.

SECTION 2: Site pictures and layout

Site pictures

Figure 2. Geonor T-200B-MD-3-W DFIR

Figure 3. Rain gauges

Figure 4. 2D Video Disdrometer

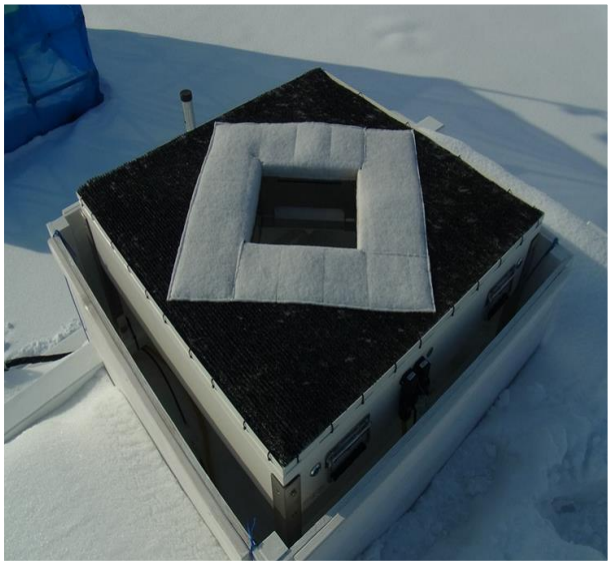


Figure 5. From West

Overview of Japan SPICE site of **Joetsu**

Site Layout

Figure 6. Site Layout of **Joetsu**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Joetsu** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project.

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer	Project
T-200B-MD-3-W	WG	R2	DFIR	1 min	Y	3.5	Precipitation	Geonor	SITE
2D-Video Disdrometer	OD	Anc.	DFIR	1 min	?	3	particle diameter, volume, falling velocity and oblateness	JOANNEUM RESEARCH	SITE
RT-3 Rain Gauge (RT-3)	TBRG	UT	Cylinder ring	1 min	Y	3.5	Precipitation	Ogasawara Keiki Co.	SITE
RT-3 Rain Gauge (RT-3)	TBRG	UT	Unshield	1 min	Y	3.5	Precipitation	Ogasawara Keiki Co.	SITE
RT-3 Rain Gauge (RT-4)	TBRG	UT	Cylinder ring	1 min	Y	3.5	Precipitation	Yokogawa Denshikiki Co.	SITE
RT-3 Rain Gauge (RT-4)	TBRG	UT	Unshield	1 min	Y	3.5	Precipitation	Yokogawa Denshikiki Co.	SITE
Tamura snow-rain intensity meter		UT	Unshield	1 min	Y	3.5	Precipitation and precipitation type	Sanyo Kogyo Co.	SITE
Laser Precipitation Monitor	Laser	UT	n/a	1 min	Y	3.5	Precipitation and precipitation type	Adolf Thies GmbH	SITE
CYG-5103AP		Anc.	n/a	1 min	N	4	Wind speed and direction	Young	SITE
HMP155D	US	Anc.	n/a	1 min	n/a	2.5	Air temperature and humidity	Vaisala	SITE
WS600-UMB compact AWS		Anc.	n/a	1 min	Y	3.5	Air temperature and humidity, prec. intensity, type and quality, air pressure, wind speed+direction	Lufft	SITE
MP-3000A Microwave radiometer		Anc.	n/a	1 min	N	3.5	Vapor/liquid water amount and temperature	Radiometrics Co.	SITE
Micro rain radar		Anc.	n/a	10 s	Y	3.5	Snow depth	METEK Co	SITE

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer	Project

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: 0 = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Rikubetsu, Japan

SECTION 1: Site Location

Country	**JAPAN**
Site name	**Rikubetsu**
latitude	43° 28' 59.2" (43.48311)
longitude	143° 45' 51.8" (143.7644)
elevation (m)	217

Site Manager: Naohiko Hirasawa, email: hira.n@nipr.ac.jp

Figure 1. Overview of **Rikubetsu** SPICE site

WMO SPICE / Site objectives:

Focus on SPICE objective II. Assess/characterize automatic systems (both the hardware and the associated processing) used in operational applications for the measurement of Solid Precipitation (i.e. gauges as “black boxes”)

Site Description

The site is on a bottom basin of a wide valley with the width of about 5-6 km, where farms occupy mainly. The height of the watershed surrounding the basin is approximately several 100 m.

This is one of the coldest areas in Japan. The winter temperature often goes down less than -20°C in early morning. Snowfall amount is not so much and the maximal snow depth is around 50cm through winter. Wind is not so strong, often calm at night. The weather is rather stable but moderate storm comes about 10 times in winter, which contributes a large part of snowfall amount and when the air temperature goes up around 0°C in all time.

SECTION 2: Site pictures and layout

Site pictures

Figure 2. From South

Figure 3. From North

Figure 4. Test Field

Figure 5. Meteorological Field

Overview of Japan SPICE site of **Rikubetsu**

Site Layout

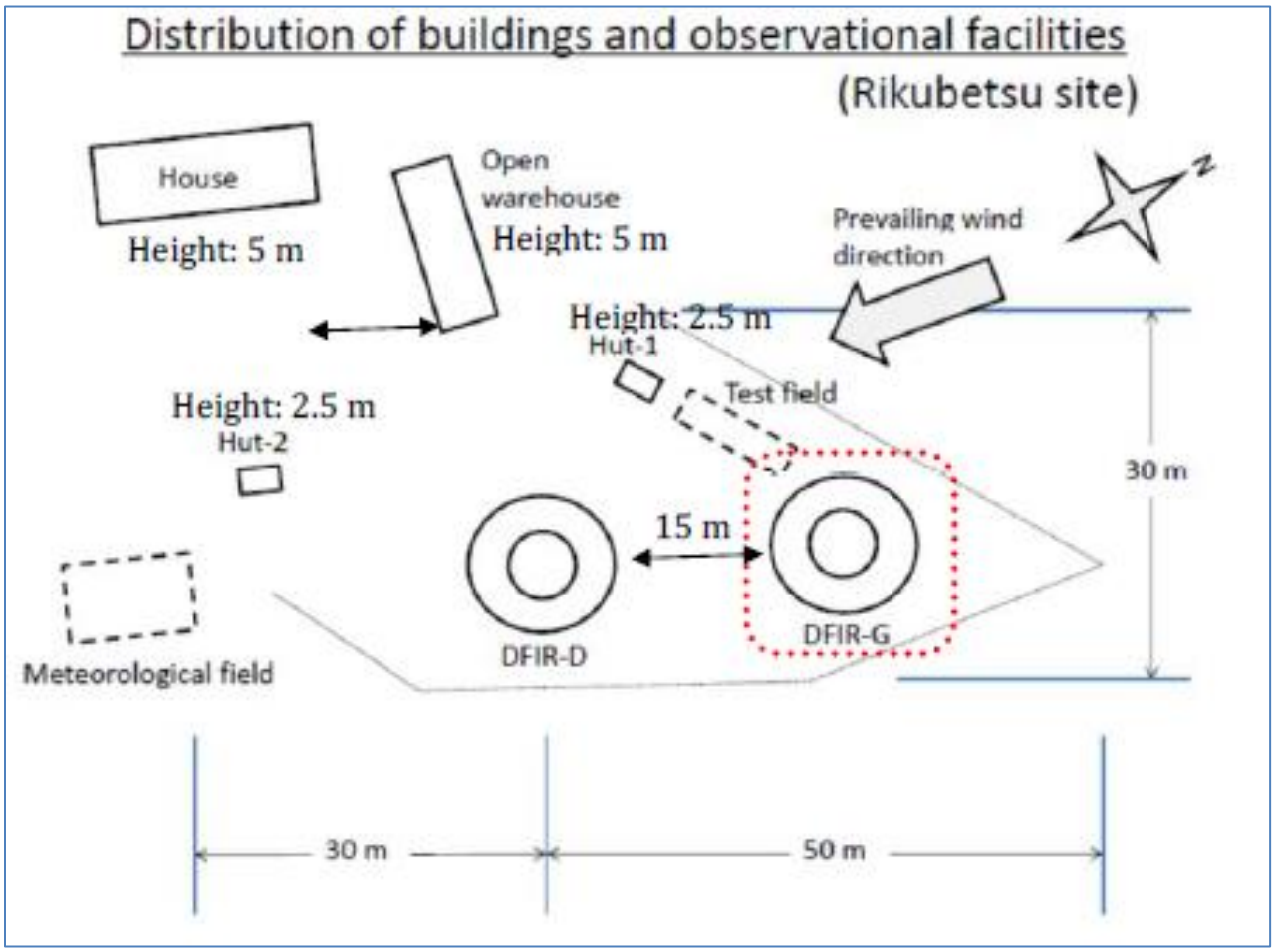


Figure 6. Site Layout of **Rikubetsu**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Rikubetsu** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project.

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer	Project
T-200B-MD-3-W	WG	R2	DFIR	10s	??	3.0	Precipitation	Geonor	SITE
Laser Precipitation Monitor	Laser	UT	DFIR	1 min	Y	3.0	Precipitation and precip. type	Adolf Thies GmbH	SITE
RT-3 Rain Gauge	TB	UT	Cylinder ring	10 min	Y	3.0	Precipitation	Ogasawara Keiki Co.	SITE
RT-4 Rain Gauge	TB	UT	Cylinder ring	10 min	Y	3.0	Precipitation	Ogasawara Keiki Co.	SITE
Laser Precipitation Monitor	Laser	UT	Un	1 min	Y	3.0	Precipitation & precip. type	Adolf Thies GmbH	SITE
Parsivel	OD	UT	n/a	1 min	Y	3.0	Precipitation & precip. type	OTT	SITE
PWS100	L	UT	n/a	1 min	Y	3.0	Precipitation and precip. type	Campbell	SITE
Laser Precipitation Monitor	Laser	UT	n/a	1 min	Y	3.0	Precipitation and precipi. type	Adolf Thies GmbH	SITE
SPC		UT	n/a	1s	N	3.0	Snow Particle Counter	Niigata Denki	SITE
SPC		UT	n/a	1s	N	3.0	Snow Particle Counter	Niigata Denki	SITE
CT25K Ceilometer		Anc.	n/a	--	N	1.0	Ceilometer	Vaisala	SITE
USA1		Anc.	DFIR	1s	Y	3.0	Wind speed and direction	Metek	SITE
AWS WXT530		Anc.	n/a	1 min	N	1.5	Air Temp. & humid., Wind speed + dir., Rainfall	Vaisala	SITE
USA1		Anc.	n/a	1s	Y	3.0	Wind speed and direction	Metek	SITE
??		Anc.	n/a	--	n/a	--	Atm. Press.	??	SITE
05103 anem.		Anc.	n/a	--	n/a	10m	Wind speed & dir.	YOUNG	SITE
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height (m)**	**Parameter**	**Manufacturer**	**Project**

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: Un = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Gochang, Republic of Korea

SECTION 1: Site Location

Country	**Republic of KOREA**
Site name	Gochang
latitude	35.34823068° N
longitude	126.59901788° E
elevation (m)	52

Site Manager: Young-San Park (sanpark@korea.kr)

Figure 1: site picture

WMO SPICE / Site objectives

The SPICE objectives would prefer to focus on A and C:
A) Automation of measurement methods for snow depth and snow water equivalent,
C) Improvement of the quality of Observed data through an accurate measurement

Site Description

Gochang is located in the southwestern region of Korea adjacent to the west coast (19km inland). It is one of the operational observation sites of KMA (Korea Meteorological Administration), and has 5 years of observation history.

- NW of the site, there is a mountain (400 meter height) which is more than 20km distant
- NE of the site, there is a mountain (748 meter height)
- Due to the geographical and topographical characteristics of the site, it has a high probability of heavy snow when CP expands toward the west coast of the Korean peninsula in the winter

Type of terrain:

- Agricultural terrain
- The ground around the site has a gentle slope and is composed of patches of rice paddy and vegetable field

SECTION 2: Site pictures and layout

Site pictures

Figure 2. From South

Figure 3. From North

Figure 4. From East

Figure 5. From West

Overview of the SPICE site in **Gochang**, Republic of Korea

Site Layout

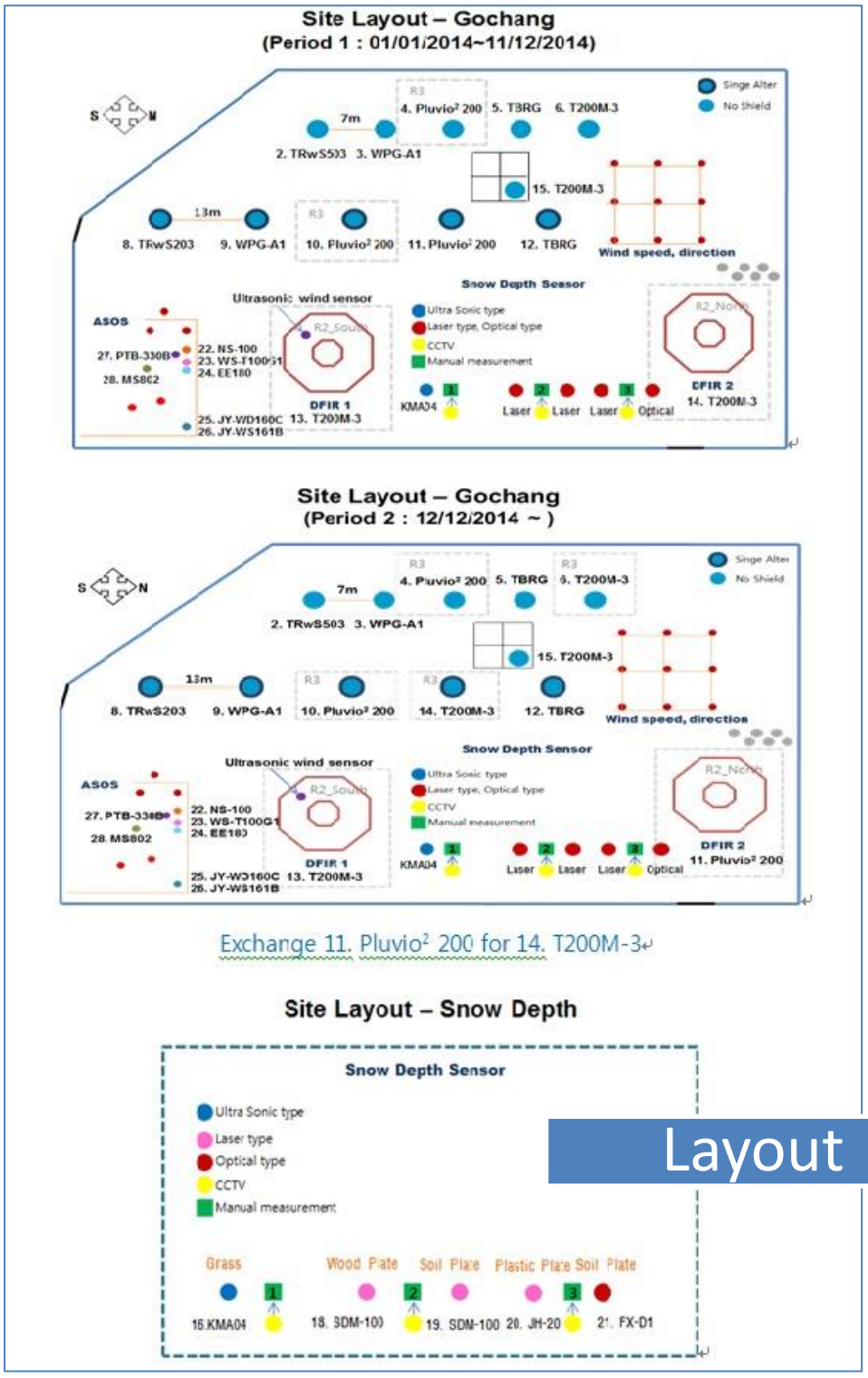


Figure 6. Site Layout of **Gochang**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Gochang** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project.

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer	Project
T200M-3 1000mm base 13	**WG**	**R2**	**SA in DFIR**	**6s**	**N**	**3**	**Precipitation**	**Geonor**	**SITE**
T200M-3 1000mm base 14	**WG**	**R2**	**SA in DFIR**	**6s**	**N**	**3**	**Precipitation**	**Geonor**	**SITE**
Pluvio2 1500mm base 11	**WG**	**R2**	**SA in DFIR**	**6s**	**N**	**2**	**Precipitation**	**OTT**	**SITE**
Pluvio2 1500mm base 10	**WG**	**R3**	**Single alter**	**6s**	**Y**	**2**	**Precipitation**	**OTT**	**MANUF-ACTURER**
Pluvio2 1500mm base 4	**WG**	**R3**	**Unshielded**	**6s**	**Y**	**2**	**Precipitation**	**OTT**	**SITE**
T200M-3 1000mm base 6	**WG**	**R3**	**Unshielded**				**Precipitation**	**Geonor**	**SITE**
T200M-3 1000mm base 14	**WG**	**R3**	**Single alter**				**Precipitation**	**Geonor**	**SITE**
Pluvio2 1500mm base 11	WG	UT	Single alter	1h	n/a		Precipitation	OTT	SITE
ASOS Rain Sensor	Cap	Anc.	n/a	1h	n/a		Precipitation detector	Nippon electric instrument Inc.	SITE
T200M-3 1000mm base 15	WG	UT	Unshielded in PIT	6s	Y	2	Precipitation	Geonor	SITE
JH Multi Laser sensor	Laser	UT	n/a				Snowdepth	Junghan electronics, Inc.	MANUF-ACTURER
SJ Optical sensor	OD	UT	n/a	6s	N	0 (pit)	Snowdepth	Sunjintech	MANUF-ACTURER
MPL1 Multi Point Laser base 18	Laser	UT	n/a	10m	N	2.3	Snowdepth	Weatherpia Co. Ltd.	MANUF-ACTURER

MPL2 Multi Point Laser base 19	Laser	UT	n/a	1m	N	1.2	Snowdepth	Weatherpia Co. Ltd.	MANUF-ACTURER
TBRG 0.5mm heatedbase12	TBRG	UT	Alter	1m	Y	2	Precipitation	Seouljeonggi	SITE
TBRG 0.5mm heatedbase5	TBRG	UT	Unshielded	1m	Y	1.8	Precipitation	Seouljeonggi	SITE
TRwS503 240mm heated	WG	UT	Unshielded	1m	Y	2	Precipitation	MPS	SITE
TRwS203 750mm heated	WG	UT	Alter	1m	Y	2	Precipitation	MPS	SITE
Auto-Emptying - WPG-A1 heated base 3	WG	UT	Unshielded	1m	Y	2	Precipitation	Wellbian System	MANUF-ACTURER
Auto-Emptying - WPG-A1 heated base 9	WG	UT	Tretyakov	1m	Y	2	Precipitation	Wellbian System	MANUF-ACTURER
KMA 04base 16	US	UT	n/a	1m	Y	2	Snowdepth	Kaijo Sonic	SITE
MS802		Anc.	n/a	1m	Y	2	Global radiation	EKO Instr.	SITE
EE180		Anc.	n/a	10m	N	4	Humidity	E+E Elektronik	SITE
PTB-330B		Anc.	n/a	1h		1	Atmospheric pressure	Vaisala	SITE
WS-T100G1		Anc.	n/a	1m		1.6	Temperature	Weatherlink	SITE
JY-WD160C		Anc.	n/a	1m		1.2	Wind direction	Jinyang Industrial Co., Ltd.	SITE
JY-WS161B		Anc.	n/a	1m		1.6	Wind speed	Jinyang Industrial Co., Ltd.	SITE
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height (m)**	**Parameter**	**Manufacturer**	**Project**

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: 0 = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: AWS1 Pyramid, Nepal

SECTION 1: Site Location

Country	NEPAL
Site name	AWS1 Pyramid
latitude	27°57’32.17” N
longitude	86°48’47.23” E
elevation (m)	5050

Site Manager: Elisa Vuillermoz (elisa.vuillermoz@evk2cnr.org)

Figure 1: site picture

WMO SPICE / Site objectives:

SPICE objectives of primary interest , Ref.: Spice Annex, February 8th 2013:

- Obj. I: Recommend appropriate automated systems for measurement of solid precipitation.
- **Obj. II: Assess automatic systems over a range of operating conditions.**
- Obj. IV: Assess uncertainty and sources of error
- Obj. VI: Collect data for specific applications (climatology, hydrology, glaciology)

would prefer to focus on

- Obj. II: Assess automatic systems over a range of operating conditions.
- Obj. VI: Collect data for specific applications (climatology, hydrology, glaciology)

Site Description

The AWS1 Pyramid was installed on September 2000 and data of standard meteorological parameters are available from October 1, 2000. Later, in 2004 the station has been equipped with additional sensors including the ultrasonic snow level measurement. From the installation time, the AWS1 Pyramid is regularly checked both from local technicians and from Italian technicians in order to guarantee the regular operation.
This site is the highest within SPICE at an elevation of 5.050 m asl. The Pyramid International Laboratory-Observatory (Lobuche, SoluKhumbu, Nepal) was established in the framework of the collaboration between Ev-K2-CNR and Nepal Academy of Science & Technology-NAST. The Pyramid site hosts several experiments (SHARE – Stations at High Altitude for Research on the Environment; Next Data; Meteomet), and local personnel are operating also with remote support from the head office.
The terrain is rough topography, high mountain vegetation, mosses, lichens and the prevailing climate is dry winters, warm summers, mainly monsoonal precipitation.
The site was proposed at IOC-SPICE 4 (2013) by EVK2CNR (Italy). Snow on the Ground measurements (in particular accumulated snow on the ground, and snow water equivalent) are currently running. Given that SPICE is a demonstration project of the Global Cryosphere Watch (GCW), the site was accepted to effectively contribute to GCW objectives. The site contributes in the assessment of the Snow on the Ground primarily, and should be considered as SPICE S4 site.

SECTION 2: Site pictures and layout

Site pictures

Figure 2. from North

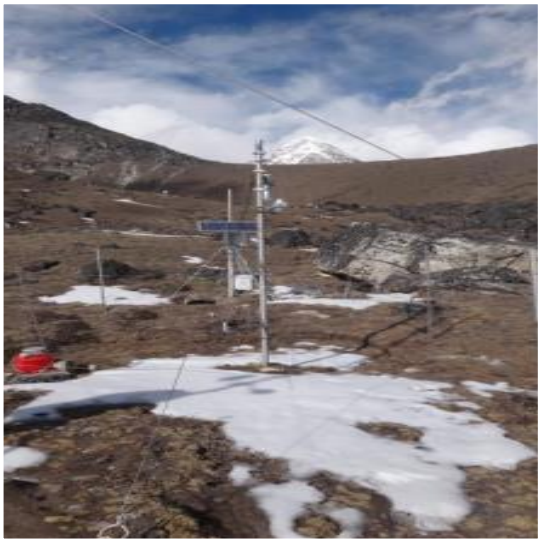


Figure 3. from South

Figure 4. From East

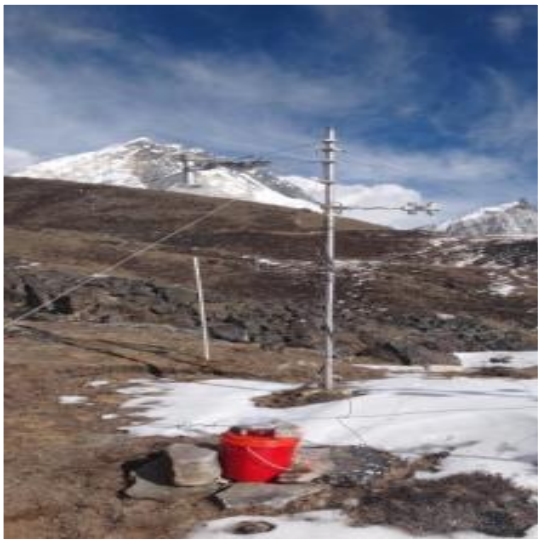


Figure 5. from West

Overview of **NEPAL** SPICE site **Pyramid**

Site Layout

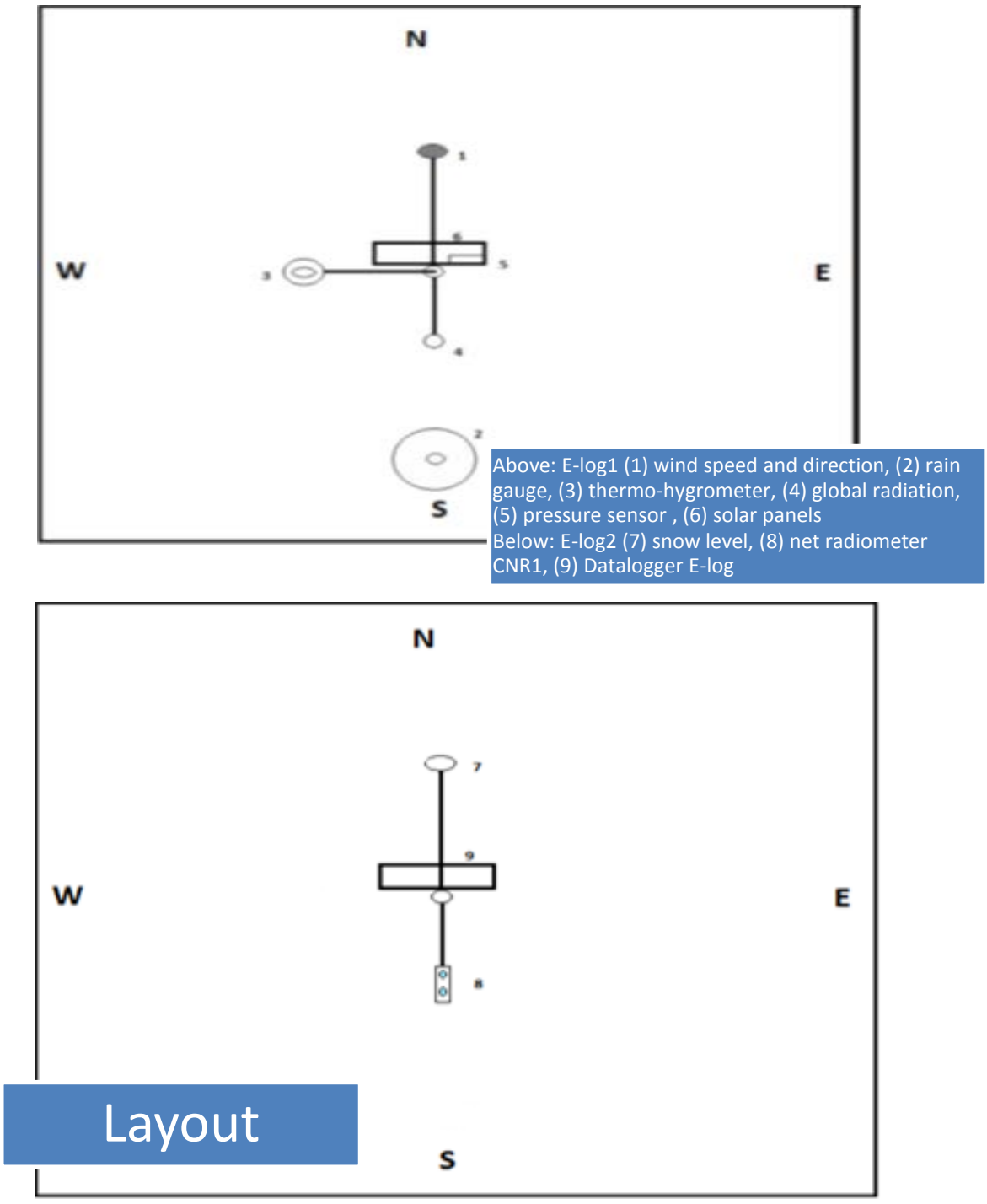


Figure 6. Site Layout of **Pyramid**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Pyramid** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project.

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer	Project
Snow density	**M**	**R SoG**	**n/a**	**Snow event**	**n/a**		**Snow density + SWE**	**AINEVA protocol**	**Site**
Snow stakes (x4)	**M**	**R SoG**	**n/a**	**1h (snow event)**	**n/a**		**Snow depth**		**Site**
USH-8 1 at 2.53m	US	UT site		1m	N	2.53	Snow depth	Sommer	
USH-8 1 at 2.5m	US	UT site		1m	N	2.5	Snow depth	Sommer	
SR50	US	UT site		1m	N		Snow depth	Campbell	Not described in Comm Rep.
DMA570 T+RH at 2m		Anc.	n/a	1m		2	Air temperature and humidity	LSI LASTEM	Site
CX115P Barometer at 2m		Anc.	n/a	1m		2	Atmospheric pressure	LSI LASTEM	Site
DPE260		Anc.		1m			Soil Heat Flux	LSI LASTEM	
Soil Moisture		Anc.		1m			Soil Moisture	LSI LASTEM	
Soil Temp. (-5 ; -20 cm)		Anc.		1m			Soil Temperature	LSI LASTEM	
Radiometer CM6B		Anc.	n/a	1m		2	Global radiation	Kipp & Zonen	Site
Radiometer CNR1		Anc.	n/a	1m		2	Shortwave and longwave Radiation fluxes	Kipp & Zonen	Site
DQA035 Pluviometer unheated at 1.5m	TB	Anc.	Unshielded	1m	N	1.5	Precipitation	LSI LASTEM	Site
Anemometer DNA022 at 5m		Anc.		1m		5	Wind speed and direction	Young	Site

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: 0 = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Mueller Hut, New Zealand

SECTION 1: Site Location

Country	**NEW ZEALAND**
Site name	Mueller Hut
latitude	43.72154 °S
longitude	170.06493 °E
elevation (m)	1818

Site Manager: Christian ZAMMIT (Christian.Zammit@niwa.co.nz)

Figure 1: site picture

WMO SPICE / Site objectives:

All SPICE objectives are of interest but 1, 5 and 6 would be priorities for the site:
1) Recommend appropriate automated field reference system(s) for the unattended measurement of solid precipitation
5) Evaluate new and emerging technology for the measurement of solid precipitation (e.g. non-catchment type), and their potential for use in operational applications.
6) Configure and collect a comprehensive data set for further data mining or for specific applications

Furthermore the Service is seeking contributions from remote SPICE site managers in regards to:

i) configuration of the gauge in remote locations;
ii) power source design;
iii) failure identification and remediation;
iv) testing of new and emerging technologies

Site Description

The National Institute of Water and Atmospheric Research Ltd of New Zealand proposes the Mueller Hut Electronic Weather Station site, which is part of the National Climate Network, for participation in SPICE. The site is located in a remote location at an elevation of 1818 m, above the tree line.
A C- band Hokitika rain radar operated by NZ Met. Service is 133km to North of the station that has no road access. The only suitable means is by helicopter; if not available a half-day tramp is required to reach the site. The inspections are event driven and emergency intervention is possible if warranted and weather conditions permit helicopter access.
Proximity: Tasman Sea 36km to NW, Lake Pukaki 22km to SSE. Station is above tree line. Mueller Hut is to NW of station but sufficiently distant.
The terrain is prevalent rock with snow covered during winter and spring with a typical late melt (November/December).
Data transmission is by Satellite only - currently hourly, but dependent on orbiting satellite (Globalstar).
The Service is planning a project for the operation of a suitable high altitude network of climate stations as part of a great national network.

SECTION 2: Site pictures and layout

Site pictures

Figure 2. Looking North

Figure 3. Looking North-West

Figure 4. Looking South-East

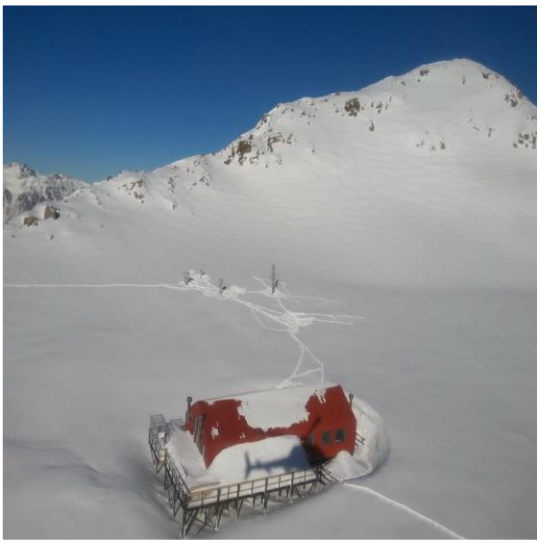


Figure 5. Looking South

Overview of **New Zealand** SPICE **Mueller Hut**

Site Layout

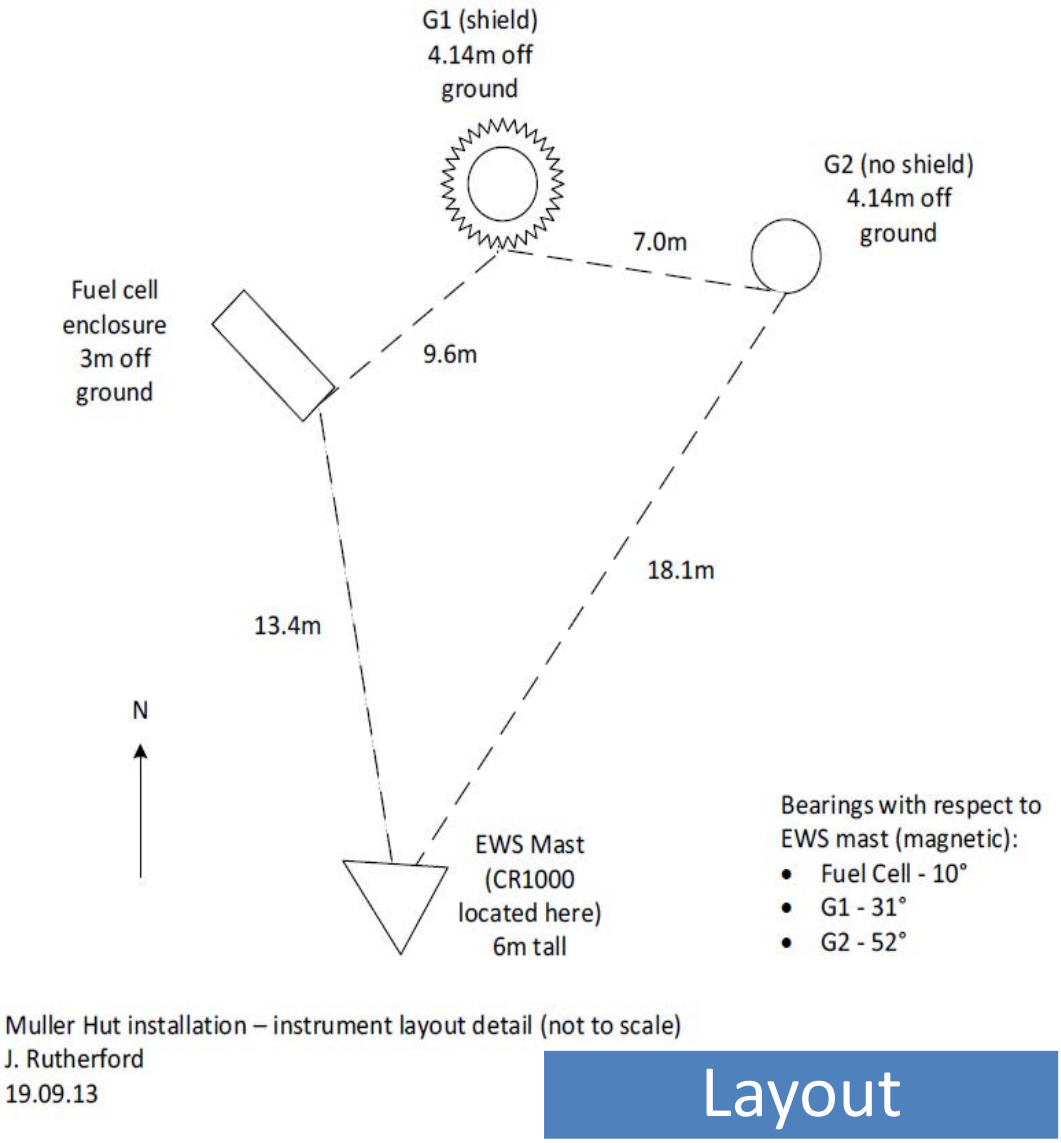


Figure 6. Site Layout of **Mueller Hut**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Mueller Hut** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height	Parameter	Manufacturer	Project
T-200B 3W (1500mm) heated	WG	UT	Single alter	1 hour	Y	4.14	Precipitation	GEONOR	SPICE
T-200B 3W (1000mm)	WG	UT	Unshielded	1 hour	Y	4.14	Precipitation	GEONOR	SITE
HMP45D		Anc.	n/a	1 hour	n/a	4	Air Temperature & humidity	Vaisala	SITE
SR50A		Anc.	n/a	1 hour	n/a	2 (?)	Snow Depth	Campbell	SITE
LI-200		Anc.	n/a	1 hour	n/a	6	Global Solar radiation	LI-COR	SITE
Wind Monitor		Anc.	n/a	1 hour	n/a	6	Wind Speed & Direction		SITE
Tipping bucket	TB	Anc.	Unshielded	1 hour	N	?	Rainfall	?	SITE
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height**	**Parameter**	**Manufacturer**	**Project**

Type: WG = Weighing gauge; TB = Tipping bucket; OD = Optical detector

Status: UT = Under Test; Anc. = Ancillary, RX = Reference (X = 0,0a,1,2,3)

Shield: 0 = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Haukeliseter, Norway

SECTION 1: Site Location

Country	**NORWAY**
Site name	Haukeliseter
latitude	59° 48.71’ N
longitude	7° 12.86’ E
elevation (m)	991

Site Manager: Mareile Astrid Wolff (mareile.wolff@met.no)

Figure 1: site picture

WMO SPICE / Site objectives

The SPICE objectives of primary interest for the Norwegian Meteorological Institute are:
I. Recommend appropriate automated field reference system for the unattended measurement of solid precipitation
IIb. Derive adjustments to be applied to measurements from operational automatic systems, as a function of variables available at an operational site
IId. Assess operational data processing and data quality management techniques
IIe. Assess the minimum practicable temporal resolution for reporting a valid solid precipitation measurement
IV Assess the achievable uncertainty of the measurement systems included in SPICE and the ability to effectively report solid precipitation

Site Description

The site is located on a relative plateau in Southern Norway in the mountains well about the tree limit. The site has been in use since the winter 2010/2011. Natural vegetation are low scrubs. The actual test site is artificially flattened and embraces 4518 m2. The most significant slope is towards European Road E134, 150 m east of the site, which has an altitude of 1005m. There are several water bodies in the area, which freezes completely during the winter period and provide a larger flat area
Distances to the surrounding mountaintops (1200m - 1500m altitude) are generally 2-4 km, the closest is located in north east direction in ca. 1 km distance. The main wind directions are east and west directions, perpendicular to the two rows of instruments. Snow precipitation can be expected from October to May.

SECTION 2: Site pictures and layout

Site pictures

Figure 2. North view

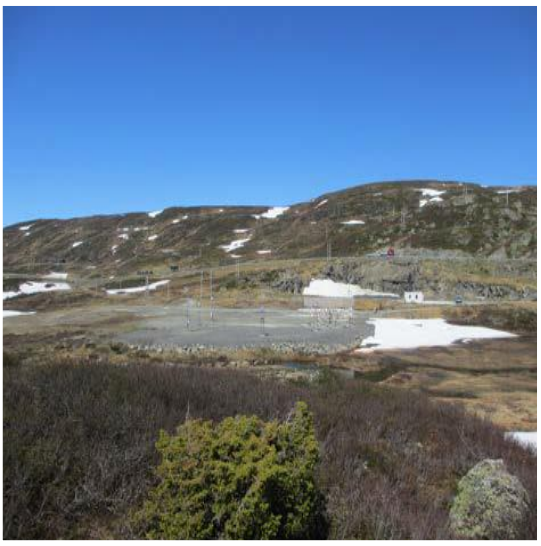


Figure 3. East view

Figure 4. South view

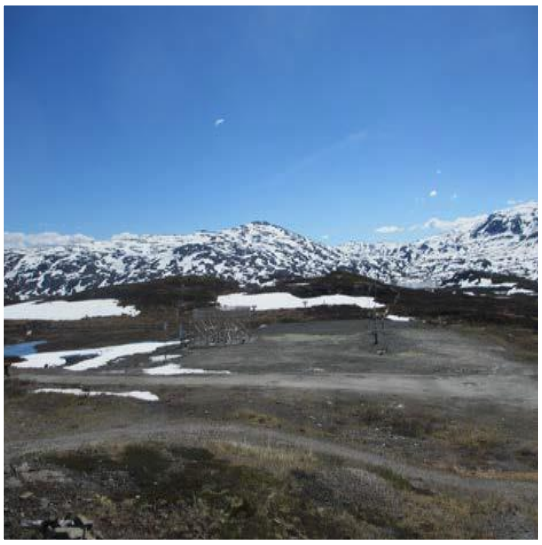


Figure 5. West view

Overview of **Norway** SPICE site **Haukeliseter**

Aerial picture of **Norway** SPICE site **Haukeliseter**

Site Layout

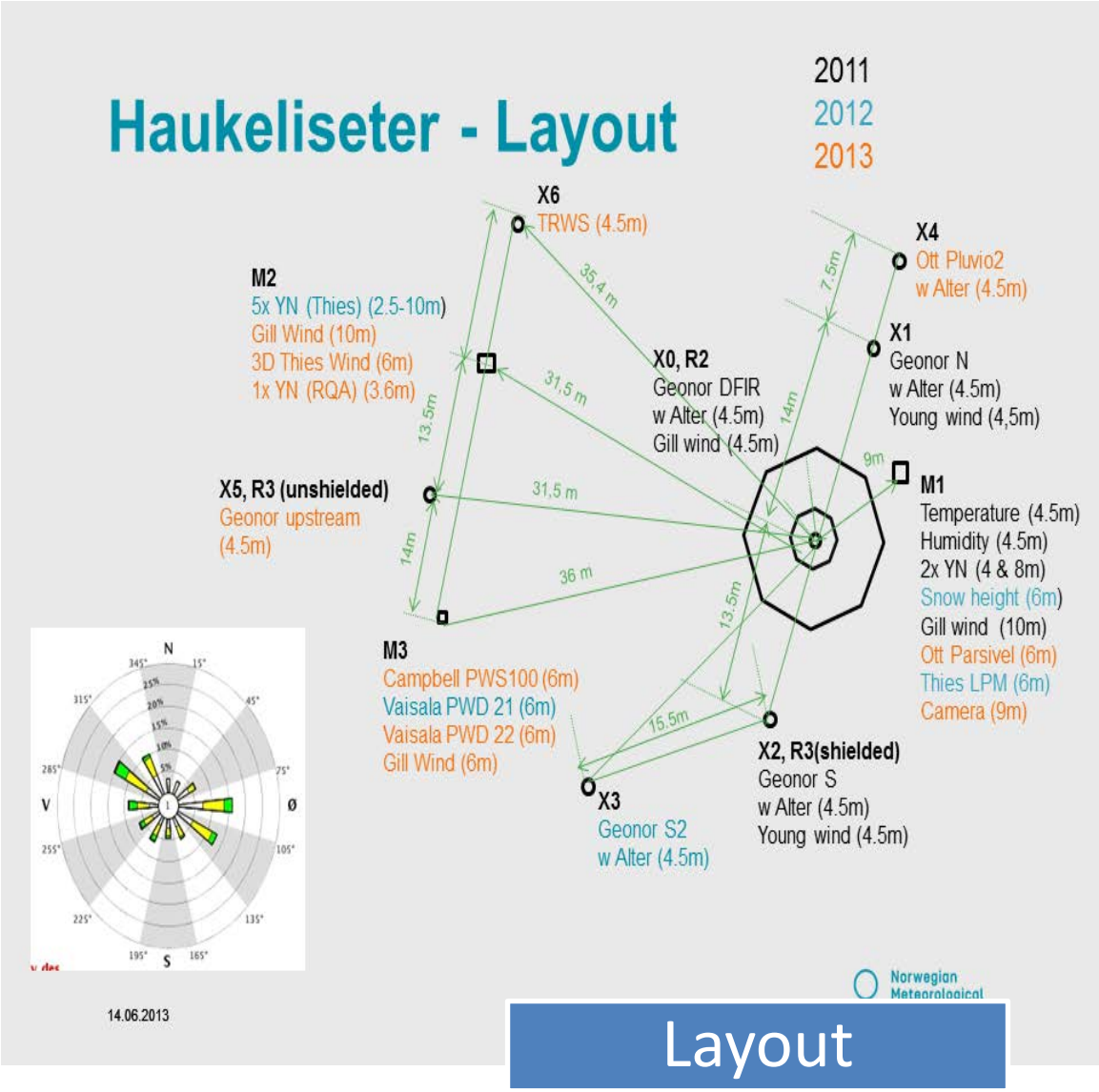


Figure 6. Site Layout of **Haukeliseter**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Haukeliseter** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project.

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer	Project
Geonor TB200 (1000mm)	**WG**	**R2**	**DFIR**	**1 min**	**Y**	**4.55**	**Precipitation**	**Geonor**	**SPICE**
Geonor TB200 (1000mm)	**WG**	**R3**	**Single Alter**	**1 min**	**Y**	**4.55**	**Precipitation**	**Geonor**	**SPICE**
Geonor TB200 (1000mm)	**WG**	**R3**	**Unshield**	**1 min**	**Y**	**4.55**	**Precipitation**	**Geonor**	**SPICE**
Thies LPM	OD	Anc	DFIR	1 min		4.5	Precipitation type	Thies	SPICE
TRWS 405 heated	WG	UT from manufacturer	Unshield	1 min	Y	4.5	Precipitation	MPS Systems	SPICE
PWS100	OD	UT from manufacturer	n/a	1 min	Y	6.0	Present Weather	Campbell	SPICE
Hotplate	HP	UT from manufacturer	n/a	1 min	?	?	Present Weather	Yankee Environmental Systems, Inc	SPICE
Pluvio2	WG	UT	Single Alter	1 min	Y	4.55	Precipitation	OTT	SITE
Geonor TB200 (1000mm) X1	WG	UT	Single Alter	1 min	Y	4.55	Precipitation	Geonor	SITE
Geonor TB200 (1000mm) X3	WG	UT	Single Alter	1 min	Y	4.55	Precipitation	Geonor	SITE
SHM30	Laser	Anc.	n/a	1 min	Y	9.0	Snowdepth	Jenoptik	SITE
HMP155		Anc.		1 min	n/a	4.5	Air temp.	Vaisala	SITE

WMT702 - mast 2		Anc.			Y	10	Wind direction	Vaisala	SITE
WindObserver II - MET mast		Anc.		1 min	Y	10	Wind direction	Gill	SITE
Parsivel (X0)	OD	Anc.			Y	6.0	Present weather	OTT	SITE
PWD 22 (X4)	OD	Anc.		1 min	Y	6.0	Present weather	Vaisala	
LPM (X2)	Laser	Anc.		1 min	Y	6.0	Precip. monitor	Thies	
HMP155		Anc.		1 min	n/a	4.5	Relative humidity	Vaisala	SITE
WindObserver II		Anc.	DFIR	1 min	Y	4.5	wind speed	Gill	SITE
Windmonitor SE, North Geonor (X1)		Anc.		1 min	N	4.5	wind speed	Young	SITE
Windmonitor SE, South Geonor (X2)		Anc.		1 min	N	4.5	wind direction	Young	SITE
Ultrasonic Anemometer 3D, mast 2		Anc.		1 min	Y	4.5	wind speed	Thies	SITE
Precipitation sensor (5.4103.20.041) sensor 0	OD	Anc.		1 min	Y	4.0 from 12.9.2013 (4 mast M1)	precipitation detector	Thies	SITE
Precipitation sensor (5.4103.20.041) sensor 1	OD	Anc.		1 min	Y	8.0	precipitation detector	Thies	SITE
Precipitation sensor	OD	Anc.		1 min	Y	2.25	precipitation detector	Thies	SITE

(5.4103.20.041) sensor 2									
Precipitation sensor (5.4103.20.041) sensor 3	OD	Anc.		1 min	Y	3.55 from 9.9.2014 (9.75)	precipitation detector	Thies	SITE
Precipitation sensor (5.4103.20.041) sensor 4	OD	Anc.		1 min	Y	4.0 from 9.9.2014 (4.55)	precipitation detector	Thies	SITE
Precipitation sensor (5.4103.20.041) sensor 5	OD	Anc.		1 min	Y	7.55	precipitation detector	Thies	SITE
Precipitation sensor (5.4103.20.041) sensor 6	OD	Anc.		1 min	Y	9.75	precipitation detector	Thies	SITE
RQA6.1 sensor 7	Capacitive	Anc.		1 min	Y	4.2 from 9.9.2014 (4.7)	precipitation detector	MPS	SITE
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height (m)**	**Parameter**	**Manufacturer**	**Project**

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: 0 = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Hala Gasienicowa, Poland

SECTION 1: Site Location

Country	POLAND
Site name	Hala Gąsienicowa)
latitude	N 49 14 39
longitude	E 20 00 21
elevation (m)	1520

Site Manager: Maciej Karzynski (Maciej.Karzynski@imgw.pl)

Figure 1: site picture

WMO SPICE / Site objectives:

The Hala Gasienicowa (Poland) site is a suitable candidate for examining the functionality of snow depth instrumentation in a relatively severe environment while still taking advantage of manual observations

Site Description

This is an existing synoptic station, staffed 24h, situates in a natural environment, uncluttered, mountainous area. The Hala Gasienicowa site typically receives large amounts of snow. Given the local conditions at the site and the presence of a well-established manual observation program, the IOC accepted the participation of this site in SPICE, with a primary focus on the measurement of snow on the ground and its relation to snowfall.
The site does not have a DFIR. The measurement of precipitation is achieved using an ASTER TPG-037-H24 rain-gauge. Additionally, precipitation amount is measured using a Hellmann rain gauge, three times per day.
For the measurement of snow on the ground, the site runs daily manual measurements, once per day, using snow stakes, as part of the operational program. No cameras are available on site.

SECTION 2: Site pictures and layout

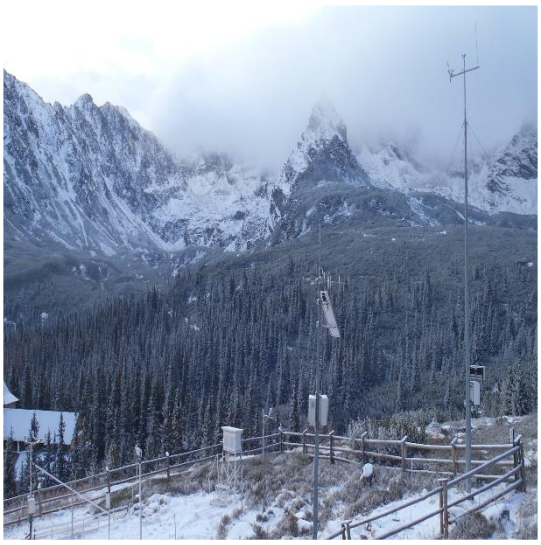
Site pictures

Figure 2.

Figure 3.

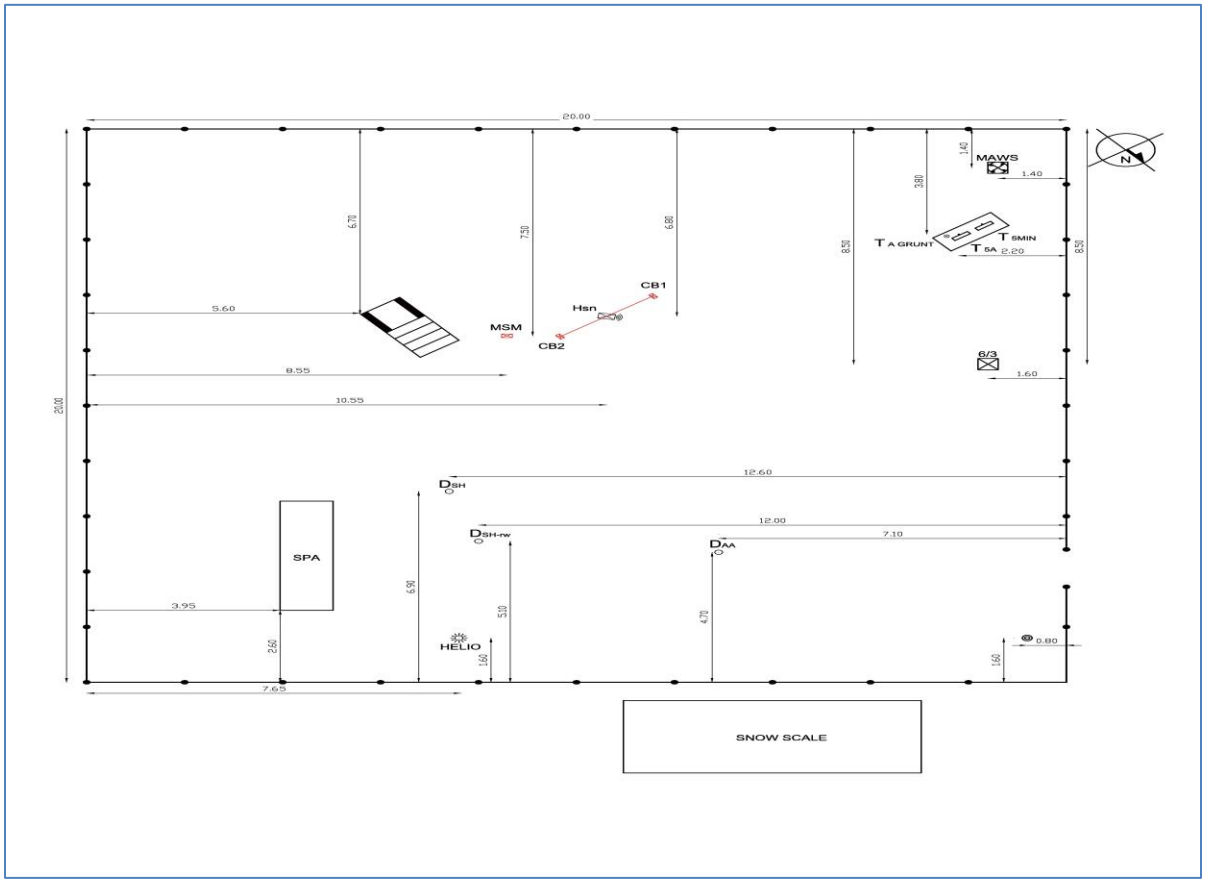
Figure 4.

Figure 5.



Overview of **Poland** SPICE site **Hala Gasienicowa**

Site Layout



Layout

Figure 6. Site Layout of **Hala Gasienicowa**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Hala Gasienicowa** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project.

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height	Parameter	Manufacturer	Project
SR50ATH-316 SS sensor 1 ESE	US	UT	n/a	10 min	Y	3.1	Snow Depth	Campbell	SPICE
SR50ATH-316 SS sensor 2 WNW	US	UT	n/a	10 min	Y	3.1	Snow Depth	Campbell	SPICE
--		Anc.		10 min		3.1	Air temperature	Campbell	SPICE
--		Anc.		10 min		3.1	Air temperature	Campbell	SPICE
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height**	**Parameter**	**Manufacturer**	**Project**

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: 0 = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Valday, Russia

SECTION 1: Site Location

Country	**RUSSIA**
Site name	Valday precipitation polygon
latitude	57º 59´
longitude	33º 15´
elevation (m)	194

Site Manager: Anton Timofeev (timofeevau@hotmail.com)

Figure 1: site picture

WMO SPICE / Site objectives:

1. Assessing the inaccuracy of automatic precipitation gauges and developing a method for correcting measurements
2. Recommending automatic devices to measure solid precipitations, suitable to the Russian conditions
3. Finding out whether it is possible to use the current automatic precipitation gauges to calibrate the DMRL-S dual-polarized Doppler meteorological radar

Update the DFIR versus Bush gauge analysis of uncertainty using data from Valdai (past and future).

Site Description

The Valdai site participated the WMO Intercomparison for solid precipitation, 1987-1993. This site hosts the only bush gauge (The bush at Caribou Creek site, Canada, was developed after SPICE onset), which represents the primary field reference for the measurement of solid precipitation (R0). Additionally, the site has a DFIR with a manual Tretyakov collector and a Tretyakov shield. This has been operational between 1970 - 1976 and 1988 – 2010.
The site has six Tretyakov precipitation gauges (orifice area: 200 cm2) with standard shields, three bush–sheltered Tretyakov gauges, and one Nipher precipitation gauge (500 cm2).
The Valdai Control System (VCS) is located in a bushy site and consists of three standard Tretyakov precipitation gauges set up in thick bushes, in a centre of a round glade 10 m in diameter and in a centre of a glade of the same size within a wooden fence
The station is situated on a flat lakeside meadow 30 m from the Valdai Lake. The bush platform is situated in an osier thicket, cut at a height of 2 m from the ground. The operative part of the platform is placed 50 m from the nearest bush edge. The distance between the two station platforms is 150 m.

The landscape is characterized by moraine hills. The southern taiga features mainly forests of deciduous fir tree and tracts of pine. Regional forest cover: 75%; scrubland: 20%. In the vicinity of the precipitation station, bush consists mainly of crack willow (salix fragilis). Ground slope: 0°

SECTION 2: Site pictures and layout

Site pictures

Figure 2. R0 Control System

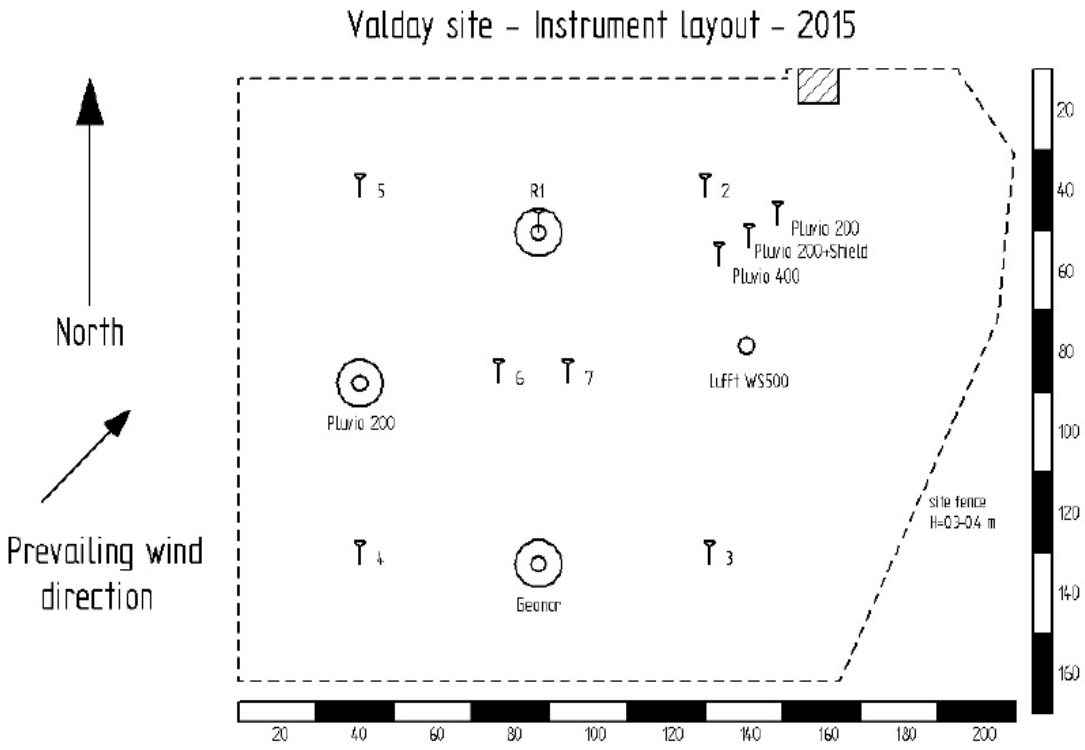
Figure 3. DFIR

Figure 4. Tryetyakov gauge

Figure 5. OTT Pluvio2 Under Test

Overview of **Russia** SPICE site **Valday**

Site Layout



Layout

Figure 6. Site Layout of **Valday**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Valday** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer	Project
Tretyakov gauge T0	**Manual**	**R0**	**Tretyakov**	**1 / day**	**N**	**2.0**	**Precipitation**	**--**	**SITE**
Tretyakov gauge T21	**Manual**	**R0**	**Tretyakov**	**1 / day**	**N**	**2.0**	**Precipitation**	**--**	**SITE**
Tretyakov gauge T8	**Manual**	**R0**	**Tretyakov**	**1 / day**	**N**	**2.0**	**Precipitation**	**--**	**SITE**
Observer	Manual	anc.	R0	1 / day			Precipitation type	--	SITE
Tretyakov gauge	**Manual**	**R1**	**Tretyakov + DFIR**	**1 / day**	**N**	**3.0**	**Precipitation**	**--**	
Observer	Manual	anc.	R1	1 / day			Precipitation type	--	SITE
Pluvio2 200 RH	**WG**	**R2**	**DFIR**	**5 min**	**N**	**3.0**	**Precipitation**	**OTT**	**SPICE**
Pluvio2 200mm	**WG**	**R3**	**Single Alter**	**5 min**	**N**	**2.0**	**Precipitation**	**OTT**	**SPICE**
Pluvio2 200mm	**WG**	**R3**	**Unshielded**	**5 min**	**N**	**2.0**	**Precipitation**	**OTT**	**SPICE**
Tretyakov gauge T2	Manual	UT	Tretyakov	1 / day	N	2.0	Precipitation	USSR Завод «Гидрометприбор», Tbilisi	SITE
Tretyakov gauge T3	Manual	UT	Tretyakov	1 / day	N	2.0	Precipitation	USSR Завод «Гидрометприбор», Tbilisi	SITE
Tretyakov gauge T4	Manual	UT	Tretyakov	1 / day	N	2.0	Precipitation	USSR Завод «Гидрометприбор», Tbilisi	SITE
Tretyakov gauge T5	Manual	UT	Tretyakov	1 / day	N	2.0	Precipitation	USSR Завод «Гидрометприбор», Tbilisi	SITE
Tretyakov gauge T6	Manual	UT	Tretyakov	1 / day	N	2.0	Precipitation	USSR Завод «Гидрометприбор», Tbilisi	SITE
Tretyakov gauge T7	Manual	UT	Tretyakov	1 / day	N	2.0	Precipitation	USSR Завод «Гидрометприбор», Tbilisi	SITE
Pluvio2 400mm	WG	UT	Single Alter	1 min	N	2.0	Precipitation	OTT	SITE
Geonor T200B	WG	UT	Double Alter	1 min	N	3.0	Precipitation	Geonor	SITE

WS500 compact AWS		Anc.	n/a	5 min	n/a	2.0	Wind speed and direction, air temp. and humidity, atm. Press.	Lufft	SITE

Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height (m)**	**Parameter**	**Manufacturer**	**Project**

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: 0 = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Volga River Hydro Meteorological Observatory, Russia

SECTION 1: Site Location

Country	**RUSSIA**
Site name	Volga River Hydro Meteorological Observatory
latitude	56°41'N
longitude	43°25'E
elevation (m)	100

Site Manager: Arkadi V. Koldaev (avk425@mail.ru)

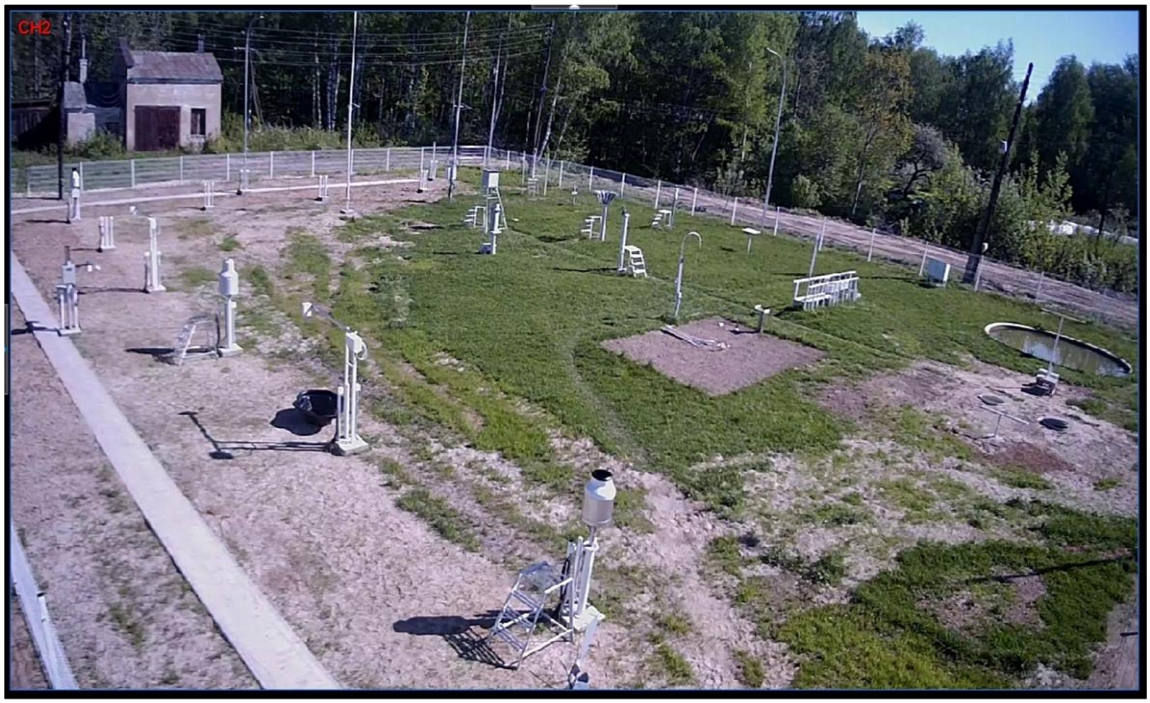


Figure 1: site picture

WMO SPICE / Site objectives:

Intercomparison of different types of sensor with different physical principles of measurement with the aim of improving the national observational net and automatic measurements of snow depth

Site Description

The observatory is located at the south-western edge of the Unzhensko-Vetluzhskaia plain, on the high and steep right bank of the Volga River (Gorky Reservoir) which is scarred by ravines and gullies. It is a wooded area of mixed forest, but the regional forest cover is only 15% due to the continual felling of trees. In the area where the site is located the vegetation consists of deciduous trees up to 20 m high and bushes, which surround the site on three sides.

Projects underway:

- Routine meteorological and hydrological measurements in duty mode (7 days a week 365 days a year)
- Intercomparisons of liquid precipitation sensors.
- Atmospheric electricity research (joint program with Russian Academy of Science)
- Summer practice for students of Nizhny Novgorod State University.

Personnel on site:

- 24 hours a day.
- Technicians with special meteorological skills can provide any type of maintenance

SECTION 2: Site pictures and layout

Site pictures

Figure 2. From North

Figure 3. From South

Figure 4. From East

Figure 5. From West

Overview of **Russia** SPICE site **Volga**

Site Layout

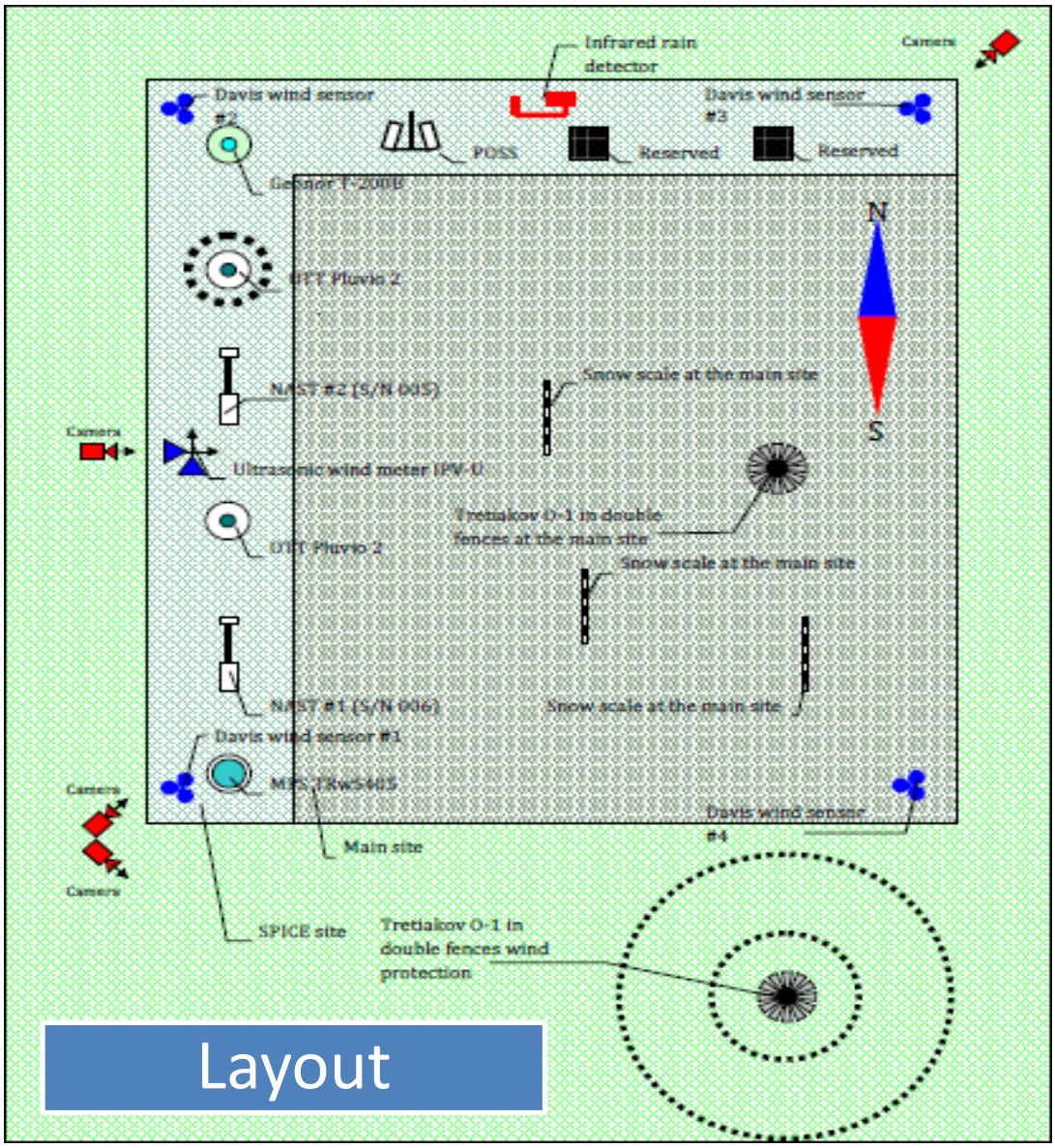


Figure 6. Site Layout of VOLGA

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Volga** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer	Project
Tretyakov gauge	M	R1	Tret+DFIR	2/day	N	3	Precipitation + SWE		SITE
Pluvio2 200mm #2	WG	R3	SA	1 min	N	2.0	Precipitation	OTT	SPICE
Pluvio2 200mm #1	WG	R3	Un	1 min	N	2.0	Precipitation	OTT	SPICE
Snow meas	M		n/a	??	n/a		Snow depth		SITE
Infrared Rain detector	OD	UT	n/a	??			Precipitation detector		SITE
POSS System	OD	Anc.		1 min	N	2.0	Precipitation Occurrence	Atmospheric Environment Service	SITE
T-200B	WG	UT	Un	1 min			Precipitation	Geonor	SITE
TRWS405	WG	UT	Un	1 min			Precipitation	MPS	SITE
Remote Snow depth sensor NAST (1)	Laser	UT		5 or 15 min?			Snow depth	Aqua Nubis	SITE
Remote Snow depth sensor NAST (2)	Laser	UT		5 or 15 min?			Snow depth	Aqua Nubis	SITE
Thermometer		Anc.		1 min		2.0	Air temperature, HI-Low Air temperature	Davis	SITE
Hygrometer		Anc.		1 min		2.0	Humidity	Davis	SITE
Pressure transducer		Anc.		1 min		2.0	Atmospheric pressure	Davis	SITE
Anemometer		Anc.		1 min		2.0	Wind speed, Hi Wind speed, Wind dir., Dominant Wind dir.	Davis	SITE
Anemometer #1 SW		Anc.		1 min		2.0	Wind speed, Hi Wind speed, Wind dir., Dominant Wind dir.	Davis	SITE
Anemometer		Anc.		1 min		2.0	Wind speed,	Davis	SITE

#2 NW						Hi Wind speed, Wind dir., Dominant Wind dir.		
Anemometer #3 NE		Anc.		1 min	2.0	Wind speed, Hi Wind speed, Wind dir., Dominant Wind dir.	Davis	SITE
Anemometer #3 SE		Anc.		1 min	2.0	Wind speed, Hi Wind speed, Wind dir., Dominant Wind dir.	Davis	SITE
Ultrasonic Anemometer IPV-U		Anc.		1 min	2.5	Wind speed, Hi Wind speed, Wind dir.	LOMO-Meteo	SITE
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height (m)**	**Parameter**	**Manufacturer**	**Project**

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: 0 = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Aramon-Formigal, Spain

SECTION 1: Site Location

Country	**SPAIN**
Site name	**Aramon-Formigal**
latitude	**42.76146**
longitude	**-0.39243**
elevation (m)	**1800**

Site Manager: Samuel Buisan (sbuisans@aemet.es)

Figure 1: site picture

WMO SPICE / Site objectives:

The SPICE objectives of primary interest for AEMET are:
1. Precipitation amount, over various time periods of time. Intercomparison between different rain gauges used at AEMET
2. Snow on the ground (snow depth) and linkages with previous objective.

The dense network of weather stations of AEMET do not use any shield for the rain gauges. No study has been performed in order to take into account the losses of precipitation during snowfalls. At this moment, the 75% of AEMET automatic stations use the THIES rain gauge (around 540 stations). The manned AEMET Spanish network of rain gauges uses the Hellman rain gauge (around 2500 stations). Since all these data is available to national, international scientific community for studies related with climate and meteorology, the results of the experiment could help to understand different processes related with solid precipitation.

Site Description

ARAMON - Formigal is the largest Spanish ski resort and it is part of the AEMET network of manned stations in the Pyrenees for more than 10 years. The site is located along the hillside in a flat place. No vegetation and no forested area on the surroundings. The Atlantic Ocean is located at a distance of 150 km. Mechanical lifts and buildings are at a distance of at least 100m

SECTION 2: Site pictures and layout

Site pictures

Figure 2. From South

Figure 3. From East

Figure 4. From North

Figure 5. From South-East

Overview of **SPAIN** SPICE site **ARAMON Formigal**

Site Layout

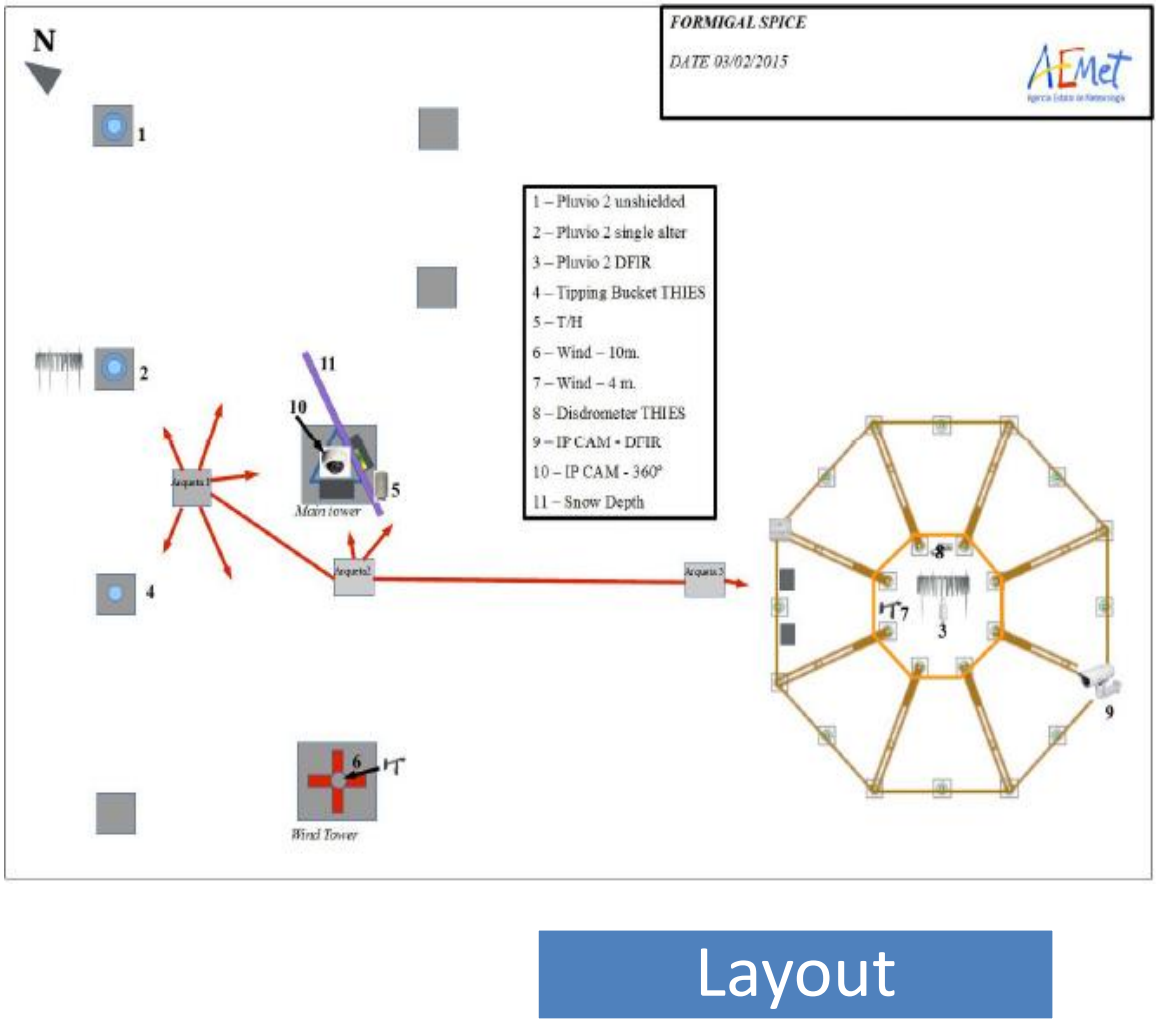


Figure 6. Site Layout of **ARAMON - Formigal**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **ARAMON Formigal** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project.

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer	Project
Pluvio2 200 cm^2, 1500 mm	**WG**	**R2**	**DFIR**	1 min	Y	3.5	**Precipitation**	**OTT**	**SITE**
LPM	Laser	Anc.	DFIR	1 min	Y	3.5	Precipitation detector (type and frequency)	Thies	SITE
Pluvio2 200 cm^2, 1500 mm	**WG**	**R3**	**A**	1 min	Y	3.5	**Precipitation**	**OTT**	**SITE**
Pluvio2 200 cm^2, 1500 mm	**WG**	**R3**	**Un**	1 min	Y	3.5	**Precipitation**	**OTT**	**SITE**
Ultrasonic Snow Level Sensor DCU-7210OTT	US	UT site	n/a				Snowdepth	APG Sensors	SITE
Raingauge 5.403235228	TB	UT site	Un	1 min	Y	1.5	Precipitation	Thies	SITE
SR-50A	US	UT site	n/a	1 min		4.2	Snowdepth	Campbell Sci.	SITE
Alpine 05103-45		Anc.	DFIR	1 min		4.0	Wind Speed and direction	YOUNG	SITE
1.1005.54.700		Anc.	n/a	1 min		4.0	Air Temperature and Relative Humidity	Thies	SITE
Alpine 05103-45		Anc.	n/a	1 min		10	Wind Speed and direction	YOUNG	SITE
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height (m)**	**Parameter**	**Manufacturer**	**Project**

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: 0 = Unshielded; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Weissfluhjoch, Switzerland

SECTION 1: Site Location

Country	**SWITZERLAND**
Site name	Weissfluhjoch
latitude	46°49’47.16’’ N
longitude	9°48’33.51’’E
elevation (m)	2537 m

Site Manager: Yves-Alain Roulet (yves-alain.roulet@meteoswiss.ch)

Figure 1: site picture

WMO SPICE / Site objectives:

The Swiss Institute for Snow and Avalanche Research (SLF) has a long history of projects, dealing with research in snow topics and specialized in snow measurements and analysis.
The collaboration between SLF and MeteoSwiss in SPICE offers an existing test site that has expandable possibilities with personnel on-site (Versuchsfeld Weissfluhjoch). Both institutes are very interested in the general objectives of SPICE. The test site was already equipped, and has been complemented with additional equipment to meet the SPICE requirements. MeteoSwiss is particularly interested in the recommendation on shielding configuration for the Pluvio2 weighing gauge, which is used on the Swiss operational measurement network. Also strong commitment of MeteoSwiss for WMO activities is a motivation to actively participate to SPICE, and therefore help other WMO members who could profit from the outcome of this intercomparison, in particular emerging countries moving from manual to automatic measurements. Outome and recommendations on the use of emerging technologies (e.g. disdrometers) are also of interest for MeteoSwiss, in the perspective of possible application for operational measurements.

Site Description

The site is located in the Eastern part of the Swiss Alps on a flat part of the mountain with no vegetation, except small amounts of grass. Various topographical obstacles (mountain ridge) to the East and West, about 500m from the site. The site is located above Davos, and is reachable by train and ski lift in winter, and by car in summer.

The Weissfluhjoch test site (S2) has been equipped with the basic infrastructure (cables, concrete pads, DFIR) in summer and autumn 2012.
The start of the experiment (partial commissioning) was performed on 8 February 2013. Since then, data from the gauges have been collected on a 10 min basis. The complete commissioning with all gauges in final configuration was performed on October 1st 2013, and data have been collected on a 1 min basis since then. Ancillary data from the SLF test field are also retrieved, as well as hourly webcam images pointing alternatively to every gauge.

The site provides extensive measurements for the assessment of both snowfall (additional equipment set-up by MeteoSwiss for the project) and snow on the ground (long term measurements performed by SLF), including automatic measurement and daily manual measurement (fresh snow, snow height, SWE).

SECTION 2: Site pictures and layout

Site pictures

Figure 2. From North

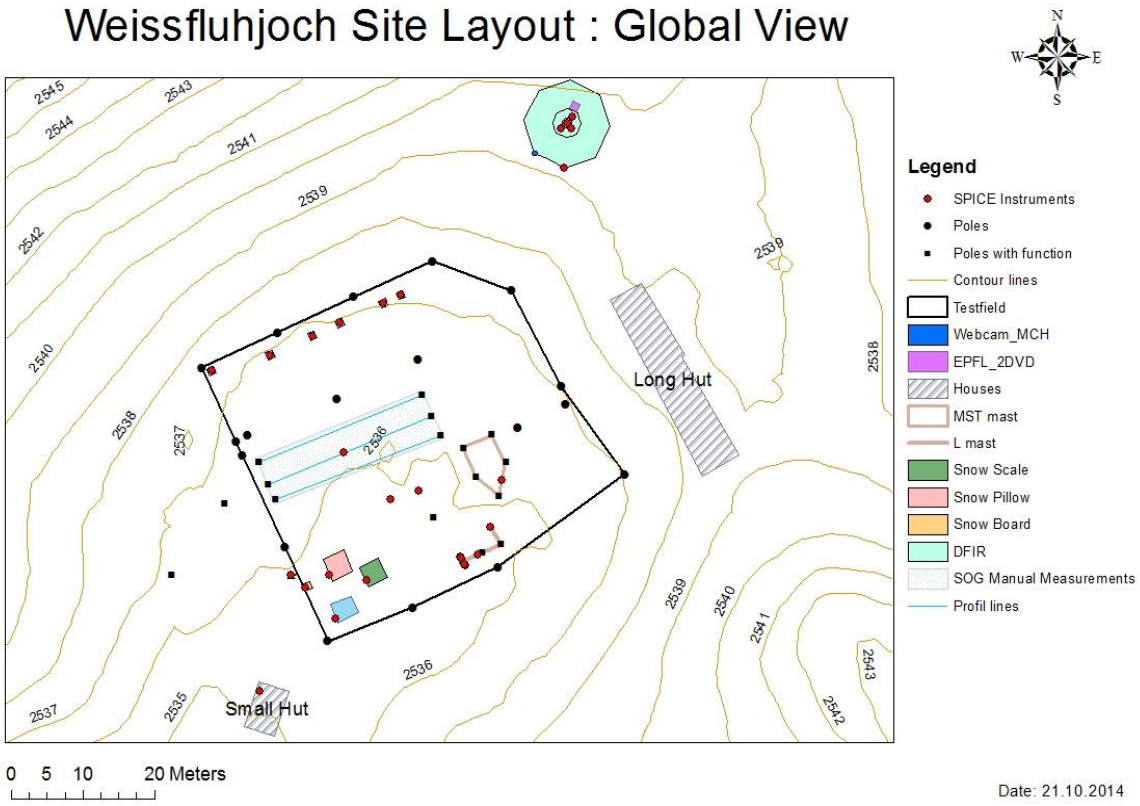
Figure 3. From East

Figure 4. From South

Figure 5. From West

Overview of **Switzerland** SPICE site **Weissfluhjoch**

Site Layout



Layout

Figure 6. General site Layout of **Weissfluhjoch**

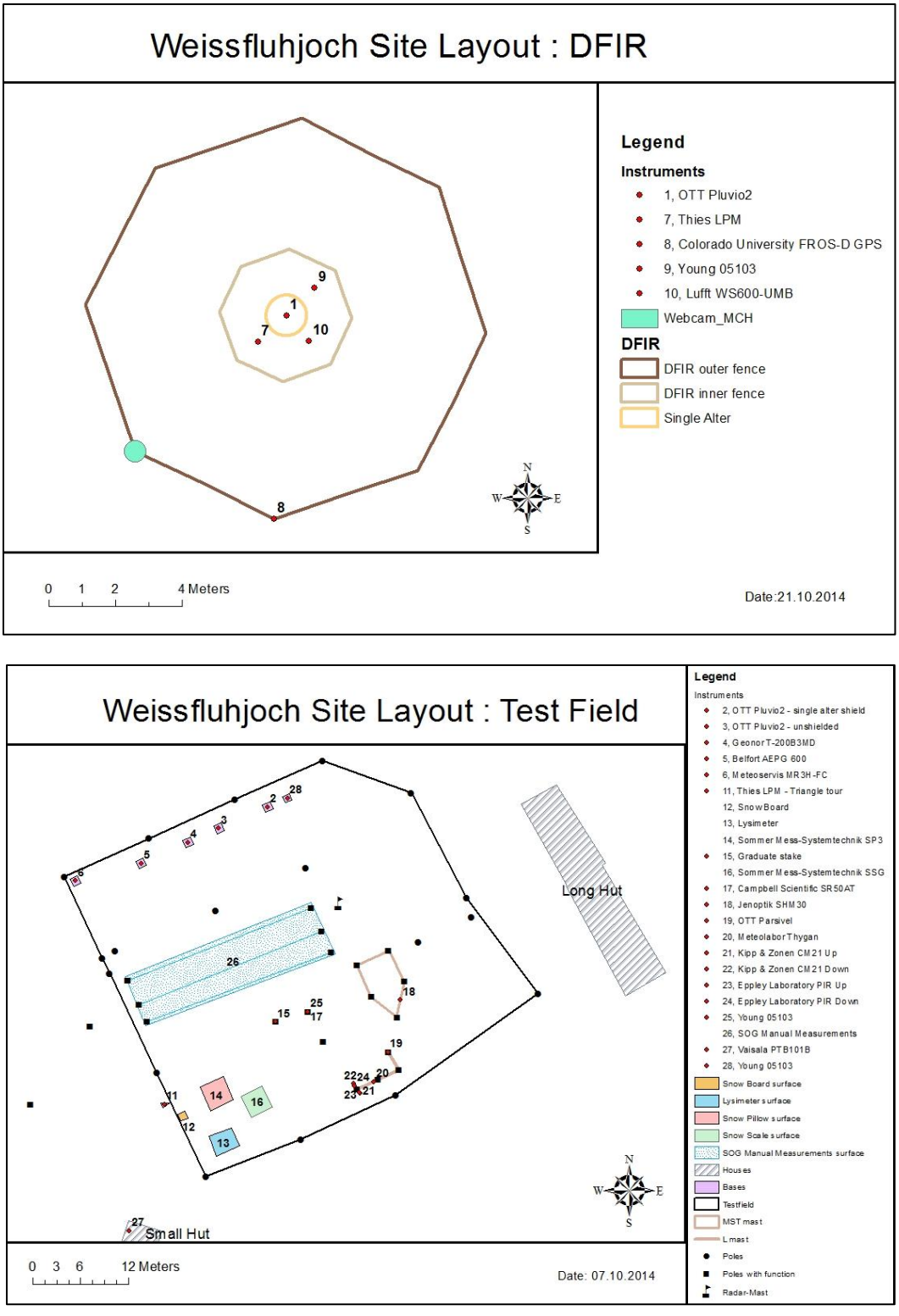


Figure 7. Detailed site Layouts of **Weissfluhjoch** with location of each sensor.

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Weissfluhjoch** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height	Parameter	Manufacturer	Project
Pluvio2 1500mm	**WG**	**R2**	**DFIR**	**1 min**	**Y**	**3.5**	**Precipitation**	**OTT**	**SPICE**
Pluvio2 1500mm	**WG**	**R3**	**SA**	**1 min**	**Y**	**3.5**	**Precipitation**	**OTT**	**SPICE**
Pluvio2 1500mm	**WG**	**R3**	**Un**	**1 min**	**Y**	**3.5**	**Precipitation**	**OTT**	**SPICE**
Snow on the ground	**M**	**RSoG**	**n/a**	**2/month**			**Snow depth**		**SITE**
LPM disdrometer, 5.4110.xx.x00	OD	Anc.	DFIR	1 min	Y	4.0	Precipitation type	Thies	SPICE
T-200B3MD 1500mm	WG	UT Manuf.	SA	1 min	Y	3.5	Precipitation	Geonor	SPICE
AEPG 600 200cm2, 600mm	WG	UT Manuf.	BDA	1 min	Y	3.5	Precipitation	Belfort	SPICE
MR3H-FC (ZAMG ver.)	TB	UT Manuf.	Un	1 min	Y	3.5	Precipitation	Meteoservis	SPICE
LPM disdrometer 5.4110.01.200 V2.50 STD	OD	UT Manuf.	Thies shield	1 min	Y	5.0	Precipitation, precipitation type	Thies	SPICE
SHM30 (012840-630-22)	L	UT	n/a	5 min	N	5.0	Snow depth	Jenoptik	SPICE
Lysimeter (pipe heated)	SMA	UT	n/a	10 min			Snow depth	SLF	SITE
SP 3		UT	n/a	1 min			Snow depth	Sommer Mess-Systemtechnik	SITE
SSG Snowscale	P	UT	n/a	30 min			Snow depth	Sommer Mess-Systemtechnik	SITE

SR50	US	UT	n/a	30 min	N	4.6	Snow depth	Campbell	SITE
WS600-UMB		Anc./UT	n/a	1 min	Y		Air temp.+ hum., Wind speed + dir., atm. Press., precipitation	Lufft	SITE
CM21		Anc.	n/a	2 min		5.0	In and out Shortwave radiation	Kipp&Zonen	SITE
Parsivel1	OD	Anc.	n/a	5 min	Y	6.0	Precipitation type	OTT	SITE
PIR Pyrgeometer		Anc.	n/a	2 min		5.0	In and out Longwave radiation	Eppley	SITE
Thygan		Anc.	n/a	10 min	Y	5.0	Air temperature and humidity	Meteolabor AG	SITE
PTB101B		Anc.	n/a	10 min			Atm. pressure	Vaisala	SITE
05103 Wind monitor		Anc.	DFIR	1 min		4.0	Wind speed and direction	Young	SITE
05103 Wind monitor		Anc.		30 min		5.5	Wind speed and direction	Young	SITE
05103 Wind monitor		Anc.		1 min		3.5	Wind speed and direction	Young	SITE
Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height**	**Parameter**	**Manufacturer**	**Project**

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3); **RSoG** = Ref. Snow on the ground

Shield: 0 = Un; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Site Description: Marshall, USA

SECTION 1: Site Location

Country	**U.S.A.**
Site name	Marshall
latitude	39.949 N
longitude	-105.195 W
elevation (m)	1742

Site Manager: Roy Rasmussen (rasmus@ucar.edu)

Figure 1: site picture

WMO SPICE / Site objectives:

Motivations:

1. Quantify solid precipitation errors,
2. Define standard measurement and analysis methods,
3. Develop transfer functions between different types of measurement systems

The Challenges of automatic snowfall rate measurements identified for the site are:

1. Wind under-catch: Gauge acting as obstacle to the flow, generating updrafts
2. Cap over of the orifice by snow accumulating on the gauge
3. Minimum detectable signal often large (to overcome noise)
4. Minimum detectable signal impacted by wind speed (higher the wind, the larger the minimum detectable signal)
5. Eliminating blowing snow false accumulations
6. High maintenance
7. Need to empty the bucket after snow fills up and refill bucket with glycol and oil.

Site Description

Proponent: National Oceanographic and Atmospheric Administration (NOAA)/Atmospheric Turbulence and Diffusion Division

Site Proposed: The NOAA/FAA/NCAR Winter Precipitation Testbed (Marshall Site), Boulder, CO

The Winter Precipitation Test Bed was initially established in 1991 at 40 km NW of Denver and 10 km SE of Boulder, at the base of the Rocky Mountains on a flat and level with semi-arid grasses less than 0.25 meters high, to address FAA needs for real-time snowfall rates in support of ground deicing.

The NOAA Climate Reference Network program started using the site in the late 90's to evaluate snow measuring instrumentation for climate purposes (Bruce Baker, lead).

The Test Bed has been used to investigate a number of important aspects of winter precipitation:

1. Under-catch of snow as a function of shield type and the development of transfer functions
2. Develop and test new wind shields
3. Evaluate the use of gauge/shield combinations for real-time and climate snow measurements
4. Develop and test new precipitation instruments (hotplate)
5. Real-time measurement of snow for aircraft ground deicing purposes
6. The use of visibility to measure snow intensity
7. Snow size distributions and terminal velocity
8. Radar- reflectivity snowfall relationships

Other programs running on site:

- GPS snow depth measurements sponsored by the University of Colorado
- Testing of various instruments by NCAR
- Ozone sonde launches once a month by NOAA

Projects underway: NOAA Climate Reference Network precipitation gauge testing.

SECTION 2: Site pictures and layout

Site pictures

Figure 2. Marshall site (Top is North)

Figure 3. Zoom of lower-right corner of Fig. 2 (Top is North)

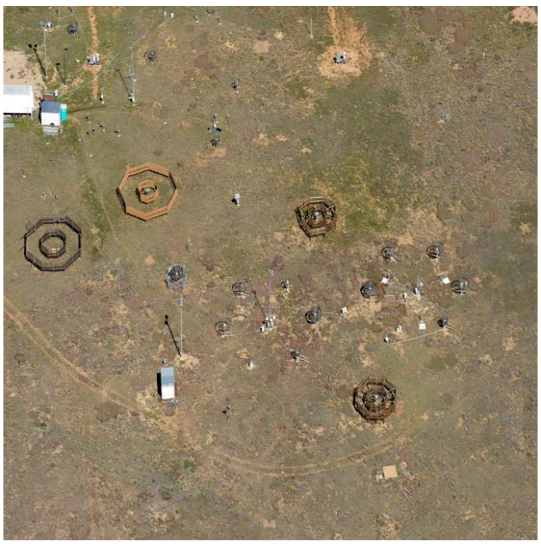


Figure 4. Pimary SPICE inst. Zoom of upper-left corner of Fig. 2 (Top is West)

Figure 5. View to the South

Overview of **U.S.A.** SPICE site **Marshall**

Site Layout

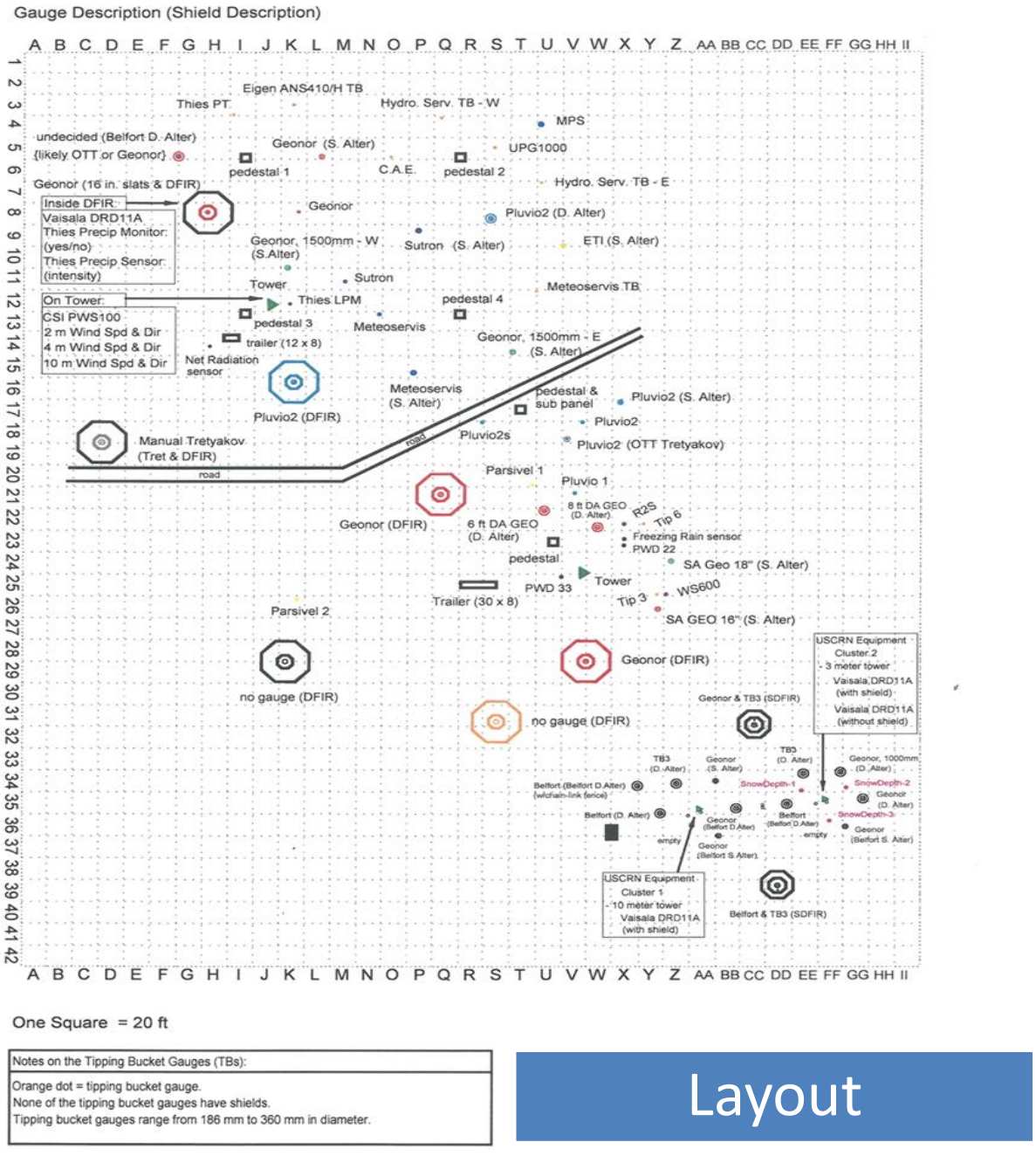


Figure 6. Site Layout of **Marshall**

SECTION 3: Site instrumentation list

Table 1. Instrumentation in **Marshall** with instrument name, instrument type, measured parameter, amount of instruments in the site, manufacturer, start date or working period and name of the project

Instrument name	Type	Status	Shield	Time res.	Heated (Y/N)	Height (m)	Parameter	Manufacturer	Project
Manual Bucket	**M**	**R1**	**DFIR**	2/day	N	3	**Bucket Accumulation**		**SITE**
T-200B3W 600mm	**WG**	**R2**	**Tall DFIR**	6s 1min	Y	3	**Precipitation**	**Geonor**	**SITE**
T-200B3W 600mm	**WG**	**R3**	**SA**	6s 1min	Y	2	**Precipitation**	**Geonor**	**SITE**
T-200B3W 600mm	**WG**	**R3**	**Un**	6s 1min	Y	2	**Precipitation**	**Geonor**	**SITE**
Pluvio2	**WG**	**R3**	**SA**	6s 1min	Y	2	**Precipitation**	**OTT**	**SITE**
Pluvio2	**WG**	**R3**	**Un**	6s 1min	Y	2	**Precipitation**	**OTT**	**SITE**
DRD11A	OD	Anc.	DFIR R2				Precipitation detector	Vaisala	SITE
Precipitation detector	OD	Anc.	DFIR R2				Precipitation detector	Thies	SITE NOT in Comm rep
TB200B3 1500mm SPICE-GEO1500-East	WG	UT	SA	6s 1min	Y	2.06	Precipitation	Geonor	SPICE
TB200B3 1500mm SPICE-GEO1500-West	WG	UT	SA	6s 1min	Y	2.06	Precipitation	Geonor	SPICE
MRW-500 SPICE-METEO-WG2	WG	UT	Un	6s 1min	Y	1.88	Precipitation	METEOSERVIS	SPICE
MRW-500 SPICE-METEO-WG1	WG	UT	Mod. Alter 33.5 cm slats	6s 1min	Y	1.92	Precipitation	METEOSERVIS	SPICE
Pluvio2 1500mm SPICE-OTT-Tret	WG	UT	Tretyakov	6s 1min	Y	1.5	Precipitation	OTT	SPICE
TRWS405 750mm SPICE-MPS-WG	WG	UT	Un	6s 1min	Y	1.85	Precipitation	MPS	SPICE
Laser Precipitation Monitor SPICE-	Laser	UT	Mod. A	6s 1min	Y	3.0	Precipitation and Precipitation	Thies	SPICE

Thies-LPM							type		
PWS 100 3m heated SPICE-Campbell-PWS	OD	UT	n/a	6s 1min	Y	3.0	Present weather	Campbell	SPICE
Precipitation transmitter SPICE-Thies_PT	TB	UT	Un	6s 1min	Y	1.37	Precipitation	Thies	SPICE
Meteorological particle sensor	?	UT	?					Droplet Measurement technologies	SPICE, not in comm rep.
ANS410H Precipitation transmitter, SPICE-Eigen_TB	TB	UT	Un	6s 1min	Y	1.19	Precipitation	Eigenbrodt	SPICE
TB PMB25R North, SPICE-CAE-TB1	TB	UT	Un	6s 1min	Y	1.9	Precipitation	CAE	SPICE
TBH-LP-TB3 /0.2 /P, SPICE-HYDRO-TB_East	TB	UT	Un	6s 1min	Y	1.97	Precipitation	Hydrological Services	SPICE
TBH-LP-TB3 /0.2 /P, SPICE-HYDRO-TB_West	TB	UT	Un	6s 1min	Y	1.9	Precipitation	Hydrological Services	SPICE
MR3H-FC SPICE-METEO-TB	TB	UT	Un	6s 1min	Y	1.27	Precipitation	Meteoservis	SPICE
UPG 1000 SPICE-UPG-TB	TB	UT	Un	6s 1min	Y	1.77	Precipitation	Environmental Measurement Systems	SPICE
TPG-001-1 SPICE-SUTRON-WG2	WG	UT	Un	6s 1min	Y	1.8	Precipitation	Sutron	SPICE
TPG-003-1 SPICE-SUTRON-WG1	WG	UT	Single Alter	6s 1min	Y	1.79	Precipitation	Sutron	SPICE
TB200B3 600mm North SPICE-GEO-DFIR-Modified	WG	UT	Mod. DFIR 4ft	6s 1min	Y	1.81	Precipitation	Geonor	SPICE
TB200B3 600mm South	WG	UT	Mod. DFIR	6s 1min	Y	1.81	Precipitation	Geonor	NCAR
Pluvio2 SPICE-OTT-DFIR-Modified	WG	UT	Mod. DFIR 4ft	6s 1min	Y	1.7	Precipitation	OTT	SITE
Parsivel2 SPICE-Parsivel2	Radar	UT	n/a	6s 1min	Y	2	Precipitation	OTT	SITE

T200B – 3W 600mm NCAR-GEO-DA8	WG	UT	Double Alter 8ft	6s 1min	Y	1.8	Precipitation	Geonor	NCAR
T200B – 3W 600mm NCAR-GEO-DA6	WG	UT	DA 6ft	6s 1min	Y	1.8	Precipitation	Geonor	NCAR
T200B – 3W 600mm NCAR-GEO-SA18	WG	UT	18” SA	6s 1min	Y	1.8	Precipitation	Geonor	NCAR
T200B – 3W 600mm NCAR-GEO-SA16	WG	UT	16” SA	6s 1min	Y	1.8	Precipitation	Geonor	NCAR
Pluvio1 1500mm NCAR-OTT-Pluvio1	WG	UT	ASOS DA (Tret+A)	6s 1min	Y	1.4	Precipitation	OTT	NCAR
WS600 NCAR_Lufft_WS600		UT	n/a	6s 1min	Y	2	Air temperature and humidity, Wind speed and direction, atm. Pressure, precipitation	Lufft	NCAR
R2S NCAR_Lufft_R2S	Doppler radar	UT	n/a	6s 1min	Y	2	Air temperature precipitation type, intensity, accumulation	Lufft	NCAR
TPS-3100 Hotplate NCAR_Hotplate	HP	UT	n/a	6s 1min	Y	2	Air temperature, wind speed, precipitation type, accumulation	Yankee Environmental Systems	NCAR
PWD22 NCAR-PWD22	OD	UT	n/a	6s 1min	Y	3	Present weather, Air temperature, wind speed, visibility, precipitation	Vaisala	NCAR
PWD33 CRN_PWD33	OD	UT	n/a	6s 1min	Y	3	Present weather	Vaisala	NCAR
ODM470 NCAR-EIGEN	OD	UT	n/a	6s 1min	Y	2	Present weather, Air	Eigenbrodt	NCAR

							temperature, wind speed, visibility, precipitation		
TB200B3 600mm North CRN_GEO_SDFIR	WG	UT	Small 2/3 DFIR	6s 1min	Y	1.9	precipitation	Geonor	SITE CRN
AEPG 600mm South CRN_BEL_SDFIR	WG	UT	Small 2/3 DFIR	6s 1min	Y	1.75	precipitation	Belfort	SITE CRN
TB200B3 600mm CRN_GEO_BDA	WG	UT	Belfort DA	6s 1min	Y	1.9	precipitation	Geonor	SITE CRN
CRN T200B3 1000mm North CRN_GEO_DA1	WG	UT	North DA	6s 1min	Y	1.9	precipitation	Geonor	SITE CRN
CRN T200B3 600mm South CRN_GEO_DA2	WG	UT	South DA	6s 1min	Y	1.9	precipitation	Geonor	SITE CRN
CRN T200B3 600mm CRN_GEO_SA	WG	UT	18” laths 4ft SA	6s 1min	Y	1.9	precipitation	Geonor	SITE CRN
CRN T200B3 600mm CRN_GEO_BSA18	WG	UT	18” laths 4ft Belfort SA	6s 1min	Y	1.7	precipitation	Geonor	SITE CRN
CRN T200B3 600mm CRN_GEO_BSA24	WG	UT	24” laths 4ft Belfort SA	6s 1min	Y	1.7	precipitation	Geonor	SITE CRN
CRN AEPG CRN_BEL_DA	WG	UT	6ft DA	6s 1min	Y	1.8	precipitation	Belfort	
CRN AEPG CRN_BEL_BDA	WG	UT	Belfort DA	6s 1min	Y	1.8	precipitation	Belfort	
CRN SR50A North CRN_SNDPTH1	US	UT	n/a	6s 1min	N	1.0	Snow Depth	Campbell	CRN
CRN SR50A East CRN_SNDPTH2	US	UT	n/a	6s 1min	N	1.0	Snow Depth	Campbell	CRN
CRN SR50A West CRN_SNDPTH3	US	UT	n/a	6s 1min	N	1.0	Snow Depth	Campbell	CRN
Pluvio2 Small 400mm SPICE-OTT-	WG	UT	Un	6s 1min	Y	1.79	Precipitation	OTT	SPICE

Pluvio2s-WG1									
ETI #21 (NOAH II)??	WG	UT	?	6s 1min			Precipitation	ETI Instr. Systems	NOT in Comm rep
CRN TB3 #1 CRN_TB3_DA_West ?	TB	UT	Doubl e Alter	6s 1min	Y	1.9	Precipitation	Hydrological Service	SITE CRN
CRN Tipping Bucket #2	TB		Un				Precipitation	Hydrological Service	SITE CRN NOT in Comm Rep.
CRN TB3 #3 CRN_TB3_DA_East?	TB	UT	Doubl e Alter	6s 1min	Y	1.9	Precipitation	Hydrological Service	SITE CRN
CRN Tipping Bucket #4	TB		Un				Precipitation	Hydrological Service	SITE CRN NOT in Comm Rep
CRN TB3 #5 CRN_HYDRO_SDFIR _North	TB	UT	North Small DFIR 2/3 mod. DFIR	6s 1min	Y	1.8	Precipitation	Hydrological Service	SITE CRN
CRN TB3 #6 CRN_HYDRO_SDFIR _South	TB	UT	South Small DFIR 2/3 mod. DFIR	6s 1min	Y	1.8	Precipitation	Hydrological Service	SITE CRN
Freezing rain sensor 0872F1 NCAR_FZRA		Anc.		6s 1min	Y	2	Accretion accumulation and rate	BF Goodrich	SITE NCAR
CNR1 Net Radiation Sensor SPICE-RAD		Anc.	n/a	6s 1min		1.5	Incoming&out going shortwave and longwave radiation	Kipp & Zonen	SPICE
05103 Wind monitor SPICE-WIND-10m		Anc.		1min		10.0	Wind Direction	Young	SPICE
CS500		Anc.	n/a	6s 1min		3.0	Air Temperature	Campbell	SITE

NCAR_TempRH						and Relative Humidity		NCAR
05103 Wind monitor SPICE-WIND-2m		Anc.		1min	2.0	Wind speed and direction	Young	SPICE
05103 Wind monitor SPICE-WIND-3.5m		Anc.		1min	3.5	Wind speed and direction	Young	SPICE
DRD 11A		Anc.				Precipitation Detector	Vaisala	SITE NCAR
PTB101B NCAR_Pressure		Anc.	n/a	6s 1min	1.0	Barometric Pressure	Vaisala	NCAR
Aerovane Indicator, Series 149 NCAR-WIND-10m		Anc.		1min	10.0	Wind speed and direction	Belfort	NCAR
Temperature Probe 1		Anc.			2.0	Air temperature		
Temperature Probe 2		Anc.			2.0	Air temperature		
Wind Monitor NCAR_Wind2m		Anc.		1min	2.0	Wind speed and direction	Young	NCAR
Wind Monitor NCAR_Wind3m		Anc.		6s 1min	3.0	Wind speed and direction	Young	NCAR
DRD_11A CRN_DRD1	OD	UT?	Manuf. shield	1min	2.0	Precipitation detector	Vaisala	SITE CRN
DRD_11A CRN_DRD2	OD	UT?	Manuf. shield	6s 1min	2.0	Precipitation detector	Vaisala	SITE CRN
DRD_11A CRN_DRD3 2m	OD	UT?	Manuf. shield	6s 1min	2.0	Precipitation detector	Vaisala	SITE CRN
Aspirated Temperature Probe 1 060A-2/062, 2144-L CRN_Temp1		Anc.		6s 1min	2.0	Air Temperature	MetOne	CRN
Aspirated Temperature Probe 2 060A-2/062, 2144-L CRN_Temp2		Anc.		6s 1min	2.0	Air Temperature	MetOne	CRN

Aspirated Temperature Probe 3 060A-2/062, 2144-L CRN_Temp3		Anc.		6s 1min		2.0	Air Temperature	MetOne	CRN
Propeller Wind Monitor 05103 CRN_Wind10_prop		Anc.		6s 1min		10.0	Wind speed and direction	Young	CRN
Sonic 85000 CRN_Wind10_sonic		Anc.		6s 1min		10.0	Wind speed and direction	Young	CRN
3 Cup CRN_Wind1.5m		Anc.		6s 1min		2 (1.5)	Wind speed	MetOne	CRN
Wetness sensor (CRN2)		Anc.	Shielded						Not in site comm rep.
Wetness sensor (CRN1)		Anc.	Shielded						Not in site comm rep.
Wetness sensor (CRN3)		Anc.	Shielded						Not in site comm rep.
Wetness sensor (CRN2) CRN4?		Anc.	Un						Not in site comm rep.

Instrument name	**Type**	**Status**	**Shield**	**Time res.**	**Heated (Y/N)**	**Height (m)**	**Parameter**	**Manufacturer**	**Project**

Type: **WG** = Weighing gauge; **TB** = Tipping bucket; **OD** = optical detector; **DS+Speed** = Droplets size and speed; **G** = Gamma ray SWE; **F** = Frequency; **HP** = Hot Plate; **L** = Laser; **M** = Manual; **R**= Radar; **SMA** = Snow Melt Analyser; **C** = Capacitive; **U** = Ultrasonic; **AC&P** = AC generator and Potentiometer; **P** = Pressure measurement

Status: **UT** = Under Test; **Anc**. = Ancillary, **Rx** = Reference (x = 0,0a,1,2,3)

Shield: Un = Uned; 1/2 DFIR = Half Double Fence Intercomparison Reference; 2/3 Mod DFIR = 2/3 Modified Double Fence Intercomparison Reference; A = Alter; BDA = Belfort Double Alter; DFIR = Double Fence Intercomparison Reference; OCT = Octagonal; SA = Single alter; SA mod = Single alter modified; TRET = Tretyakov; DA = Double Alter

Definition of the protocol for selecting and including instruments in the experiment (2012-2014) (re-drafting of ET meeting working doc no. 3.12)

1. Participation requirements

Participation in the multi-site intercomparison will be accepted based on the following main conditions:

a. In situ instruments (catching and non-catching types) and configurations that are currently being used in national networks or being considered for use in national networks will be considered.
b. Precipitation instruments with emerging technologies will be included based on the recommendations from the WMO Members.
c. Those Members that are SPICE's site providers may use their own instrumentation (or instrumentation already installed on place), and are required to meet the instrument requirements as outlined in the Site Questionnaire provided.
d. Participating instruments should be prepared to work in harsh winter conditions (e.g. heating, insulation, etc.) and to operate reliably, including if unattended.
e. Participating instruments should be installed following procedures specified by the manufacturer or the instrument provider. If an instrument is operationally used in combination with a windshield, then the appropriate shield has to be provided by the participant or properly installed by the site-providers (for instruments belonging to the site provider).
f. Participating automatic weighing gauges (AWG) that are installed as reference gauges (in DFIR configuration), should also be installed on the field in order to quantify the effect of wind losses in the catchment of solid precipitation.
g. Participating instruments should have a measurement resolution of 0.1 mm (or better) for precipitation amount and 6 mm/h or better over 1 minute time interval for precipitation intensity.
h. The reporting time interval of instruments should be 1 minute or better.
i. To standardize the data acquisition among intercomparison sites and facilitate the data synchronization, instruments with digital output (preferably serial) are preferred, to allow for the synchronization of their internal clock. Instruments with other types of outputs may be accepted only with an appropriate adaptor interface. The instruments with a pulse output do not require a converter unless the manufacturer usually provides the sensor with an interface.
j. Participating instruments should be calibrated against any suitable recognized standard before shipment and they should be supplied with appropriate calibration certificates and description of the methods used for calibration.
k. Participants should agree that their instruments could be tested regularly on site to identify possible drifts of the calibration or malfunctioning (metrological confirmation).
l. Preferences will be given to having available two identical instruments, however it is not a condition for participation.

2. Selection criteria

The number of instruments tested on each participating site will be limited by the available capacity on the intercomparison site. The ET/IOC will select instruments for participation based on the following criteria:

a. Those instruments and configurations belonging to SPICE site's providers which are already in place and that may used for on-going experiments linked to SPICE will be taken into consideration (if participation requirements are satisfied).

b. Instruments will be selected to cover a wide variety of measurement techniques and configurations.
c. Preference will be given to instruments that are widely used, operationally.
d. Preference will be given to new and emerging measuring techniques.
e. For those equipment tested in a recognized laboratory, results of the laboratory tests will be taken into consideration.

3. Selection procedure

In order to assist in the selection procedure, the ET/IOC prepared and distributed two Questionnaires (Q1 on Solid Precipitation Gauges and Q2 on Snow Depth Gauges), to enable the potential participants to provide detailed information on the proposed instruments and systems.

The Second Letter of Invitation issued by the WMO Secretariat included references to the Instrument Questionnaires and to the Site Questionnaire. The responses are expected by March 15, 2012.

The responses from the potential participants will be evaluated by the IOC and evaluated through the mechanisms agreed during the ET/IOC meeting in Geneva 5-7 October 2011 (Final Report, sec. 3.17) and a list of potential participating instruments will be proposed. The participation requirements (see sec. 2) will be verified. In case of the number of proposed instruments will exceed the SPICE sites capacity the selection criteria in sec. 2 will be applied accordingly.

4. Responsabilities of participants

a. Participants (instruments suppliers) should agree to supply transportation costs (importation and exportation) of their instruments and to provide support for the customs clearance.
b. Appropriate documentation including all detailed instructions and manuals needed for installation, operation, calibration, and routine maintenance have to be provided in advance in order to facilitate the intercomparison experiments. Participants should also provide the sites' managers with the following parameters as a minimum: measurement range, resolution, linearity, measurement uncertainty, threshold, dead time, delay time, time constant, internal calculation or update cycle, possible output cycles.
c. If the instruments are evaluating the measurements at an internal time cycle, commands or procedures for a synchronization of the internal cycle with an external clock have to be provided.
d. Participants should assist and provide the site managers with methods or any useful electronic equipment to synchronize their instruments output (or measuring/averaging/filtering interval) with respect to a standard time or to the data acquisition system of the experiment site.
e. The presence of participants is not required during the intercomparison, however assistance in operation should be provided in order to allow the test to be carried out properly and with minimum effort by the host country. Moreover, a Meeting of Participants, HMEI and sites' local staff could be organized to check that Participants' instruments were operated according to the recommended procedures and to examine the data acquisition system.

5. Inclusion of participating instruments in the experiment

The ET/IOC agreed to include in the intercomparison experiment 2012-2013 and 2013-2014 the participating instruments provided by National Hydro-Meteorological Services or National

Research Institute of the WMO Members, potentially in cooperation with an instrument manufacturer, or by manufacturers who are member of the Hydro-Meteorological Equipment Industry (HMEI).

Following the evaluation and the decision of the IOC on potential participants (see sec.3), the official list of Participants will be finalized and the Secretariat will request the Participants to deliver their instruments and related equipment.

Due to the multi-site configuration of SPICE and to reduce costs for shipment and custom clearance, the inclusion of participating instruments into the SPICE experiment sites should be carried out according to the following principles:

- inclusion of at least one instrument for each measurement technique in each SPICE experiment site;
- distribution of participating instruments according to the geographical position of the intercomparison sites with respect to instruments' providers.

Allocation of SPICE instruments proposed by Instrument Providers

Summary by instrument

Instrument Provider	Country	Manufacturer	Instrument Name	SPICE site(s) where allocated and number of instruments allocated
Weighing Gauges				
Belfort	USA	Belfort Instrument Company	AEPG 36000-1DDH	Weissfluhjoch (1)
Geonor AS	Norway	Geonor AS	T-200B3/T-200BM3, 1500 mm	Bratt's Lake (1) Caribou Creek (1) Marshall (2) Mueller Hut (1) Weissfluhjoch (1)
Meteoservis	Czech Rep	Meteoservis v.o.s.	MRW500	Bratt's Lake (2) Marshall (2)
MPS System	Slovakia	MPS System	Total Rain Weighing sensor TRwS 204	Haukeliseter (1) Marshall (1)
OTT Hydromet GmbH	Germany	OTT Hydromet GmbH	OTT Pluvio^2 200_RH	Marshall (1) Sodankylä (1)
NIMH Bulgaria	Bulgaria	Sutron (USA)	TPG-0001-1	Marshall (1)
Tipping Bucket Gauges				
Adolf Thies GmbH&Co KG	Germany	Adolf Thies GmbH&Co KG	Precipitation transmitter 5.4032.35.008	Marshall (1)
CAE S.p.A.	Italy	CAE S.p.A.	Heated PMB25R	CARE (1) Marshall (1)
Environment Meas. Limited	UK	Environment Meas. Limited	UPG1000	Marshall (1) Sodankylä (1)
Hydrological Services America	USA	Hydrlogical Services Pty. Ltd	Heated TBH/TBH-LP	CARE (1) Marshall (2)
Meteoservis	Czech Rep	Meteoservis v.o.s.	MR3H-FC AH-01 with heating	CARE (1) Marshall (1) Sodankylä (1)
ZAMG	Austria	Meteoservis (Czech Rep)	MR3H-FC ZAMG-version	CARE (1) Weissfluhjoch (1)
Eigenbrodt GmgH	Germany	Eigenbrodt GmbH	ANS410H	Marshall (1)

Optical Instruments				
Adolf Thies GmbH&Co KG	Germany	Adolf Thies GmbH&Co KG	LPM 5.4110.01.200	Marshall (1) Weissfluhjoch (1)
Campbell Scientific Ltd	UK	Campbell Scientific Ltd	PWS100	Haukeliseter (1) Marshall (1)
Droplet Measurement Technologies	USA	Droplet Measurement Technologies	Meteorological particle sensor	Marshall (1)
OTT Hydromet GmbH	Germany	OTT Hydromet GmbH	Parsivel^2	Sodankylä (1)
Vaisala	Finland	Vaisala	PWD53/PWD33	Sodankylä (1)
Vaisala	Finland	Vaisala	PWD52	Sodankylä (1)
Vaisala	Finland	Vaisala	FS11P	Sodankylä (1)
Evaporative Plates				
Yankee	USA	Yankee	TPS3100 Hotplate	Bratt's Lake (1) Sodankylä (1) Haukeliseter (1)
Snow Water Equivalent Instruments				
Campbell Sci. Canada	Canada	Campbell Scientific	CS725 Gamma Ray (previously known as GMON3)	Sodankylä (1)
Snow Depth Sensors				
Campbell Sci. Canada	Canada	Campbell Scientific	SR50ATH-316SS	Sodankylä (2) Hala Gasienicowa (2) Col de Porte (2)
ESW GmbH	Germany	ESW GmbH	SHM30/012840-642-22	Sodankylä (2)
E.T.G. srl	Italy	E.T.G srl	SENULSNIV	Sodankylä (2)
Felix Technology Inc.	Canada	Felix Technology Inc.	SL300	Sodankylä (2)
Sommer GmbH & Co KG	Austria	Sommer	USH-8	Sodankylä (2)
Jenoptik	Germany	Jenoptik	SHM30	Col de Porte (2)
University of Colorado	USA	University of Colorado	FROS-D GPS	Weissfluhjoch (1)

Summary by site

Site	Instruments
Marshall (USA)	• Geonor WG T-200B3/T-200BM3, 1500 mm (2) • MeteoServis Weighing Rain Gauge MRW500 (2) • MPS Systems Total Rain Weighing sensor, model TRwS 204 (1) • OTT Pluvio^2 model 200_RH (1) • Sutron TPG model TPG-0001-1, provided by NIMH Bulgaria (1) • Thies Precipitation transmitter 5.4032.35.008 (1) • CAE Tipping Bucket (heated) model PMB25R (1) • Environmental Meas. Limited UPG1000 (1) • Hydrological Services Heated TB TBH/TBH-LP (2) • Meteoservis MR3H-FC Rain Gauge AH-01 with heating (1) • Eigenbrodt ANS410H (1) • Thies LPM 5.4110.01.200 (1) • Campbell Scientific Present Weather Sensor PWS100 (1) • Droplet Meas. Technologies Meteorological particle sensor (1)
CARE (Canada)	• CAE TB heated PMB25R (1) • Hydrological Services Heated TB TBH/TBH-LP (1) • Meteoservis MR3H-FC Rain Gauge AH-01 with heating (1) • TB MR3H-FC ZAMG-version (1)
Bratt's Lake (Canada)	• Geonor WG T-200B3/T-200BM3, 1500 mm (1) • MeteoServis Weighing Rain Gauge MRW500 (2) • Yankee Hotplate TPS 3100 (1)
Caribou Creek (Canada)	• Geonor WG T-200B3/T-200BM3, 1500 mm: 1
Weissfluhjoch (Switzerland)	• Belfort All environment Precipitation Gauge 36000-1DDH (1) • Geonor WG T-200B3/T-200BM3, 1500 mm (1) • TB MR3H-FC ZAMG-version (1) • Thies LPM 5.4110.01.200 (1) • University of Colorado FROS-D GPS (1)
Haukeliseter (Norway)	• MPS Systems Total Rain Weighing sensor TRwS 204 (1) • Campbell Scientific Present Weather Sensor PWS100 (1) • Yankee Hotplate TPS3100 (1)
Hala Gasienicowa (Poland)	• Campbell Scientific Canada Snow depth sensors SR50ATH-316SS (2)
Mueller Hut (New Zealand)	• Geonor WG T-200B3/T-200BM3, 1500 mm (1)
Sodankylä (Finland)	• OTT Pluvio^2 200_RH (1) • Environmental Meas. Limited UPG1000 (1) • Meteoservis MR3H-FC Rain Gauge AH-01 with heating (1)

	• OTT Parsivel: 1 • Vaisala PWD53/PWD33 (1) • Vaisala PWD52 (1) • Vaisala FS11P (1) • CS725 Gamma Ray Snow Water Equivalent, previously known as GMON3 (1) • Campbell Scientific Canada Snow depth sensors SR50ATH-316SS (2) • ESW Snow Depth Sensor SHM30/012840-642-22 (2) • ETG Snow Depth Sensor SENULSNIV (2) • Felix Technology Snow Depth Sensor SL300 (2) • Sommer Snow Depth Sensor USH-8 (2)
Col de Porte (France)	• Campbell Scientific Canada Snow depth sensors SR50ATH-316SS (2) • Jenoptik SHM30 (2)

WMO-SPICE Instrument Performance Report
Belfort AEPG 600

1) Technical specifications (from manufacturer provided documentation)

Instrument model:	Belfort Universal All Environment Precipitation Gauge AEPG 600
Physical principle:	Weighing gauge (WG). The frequencies of three vibrating wire transducers change in response to the weight of accumulated precipitation, and are used to determine the accumulation amount in mm.
Bucket capacity:	600 mm
Collecting area:	200 cm^2
Operating temperature range:	-55 °C to 55 °C
Measurement uncertainty:	0.1% of full scale or 0.25 mm, whichever is greater (as defined by manufacturer)
Sensitivity:	0.025 mm (as defined by manufacturer)

a)

b)

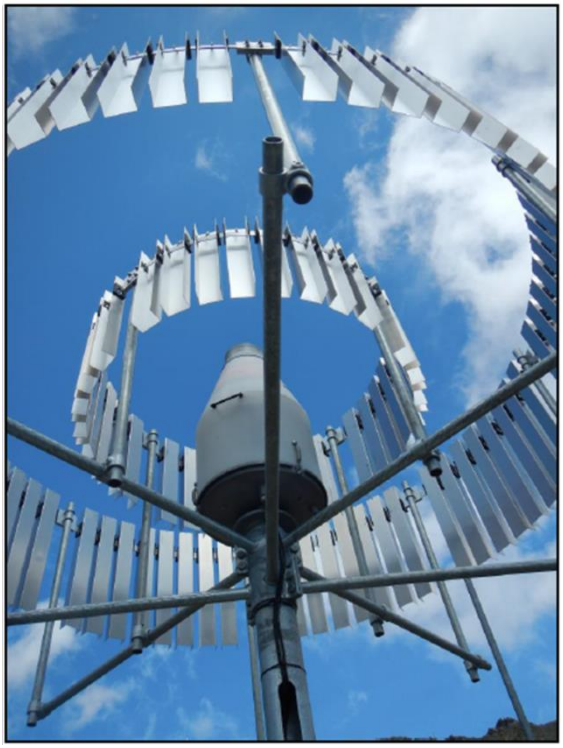


Figure 1: (a) Full-configuration and (b) zoomed-in (from beneath outer shield) views of Belfort AEPG 600 gauge in Belfort double-Alter shield at Weissfluhjoch.

2) Data output format

Gauge data output (according to instrument manual provided by manufacturer)	The AEPG 600 gauge measures the weight of a collection bucket interior to the gauge using vibrating wire technology. The frequency of each of three vibrating wires is measured by proprietary algorithms and processed by an on-board computer to calculate the average accumulated precipitation. This information is available in RS-232 format for transmission to a data collection device such as a computer or data logger. The output data message format is as follows: [Ambient temperature] [Orifice temperature] [Sensor 1 temperature] [Sensor 2 temperature] [Sensor 3 temperature] [Sensor 1 frequency] [Sensor 2 frequency] [Sensor 3 frequency] [Sensor 1 weight] [Sensor 2 weight] [Sensor 3 weight] [Total weight] [Rain amount] [Heater status] [Transducer 1, 2, 3 Status] The data acquisition interval is 1 minute.

3) SPICE test configuration

Number of transducers:	Three vibrating wire transducers, sampled independently
Shield:	Belfort double-Alter shield (BDA)
Test sites:	Weissfluhjoch (Switzerland)
Sensor provider(s):	The gauge evaluated was provided by the instrument manufacturer (Belfort Instrument)

A map of test site locations is provided in Figure 2.

3.1. Note on terms and acronyms used

Throughout this document, the following notations were used to identify the R2 reference (Ref) and senor under test (SUT) configurations:

Reference: 'DFIR' and 'DFAR' are used interchangeably for the R2 reference configurations. 'DFIR' refers to an automated gauge installed in a DFIR-fence, while 'DFAR' refers more explicitly to the Double-Fence Automated Reference configuration.

Sensor under test: The Belfort AEPG 600 gauge under test is installed in a Belfort double-Alter (BDA) shield, which is also denoted simply as a double-Alter (DA) shield.

Figure 2: Map of location of SPICE site where Belfort AEPG 600 gauge was tested.

A summary of the configuration of instruments as tested, the duration of tests and availability of data reflected in these results, and the ancillary measurements used, by site, is available in Tables 1, 2, and 3, respectively.

Table 1: Summary of gauge configuration and data output, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	Weissfluhjoch
Field configuration	Belfort double-Alter (BDA)
Height of installation (gauge rim)	3.5 m
Heating	SPICE algorithm
Antifreeze	Mixture of 3 L propylene glycol and 1 L water (Season 1) Mixture of 3.12 L propylene glycol and 4.56 L methanol (Season 2)
Oil	400 mL linseed oil (Season 1) 960 mL Isopar oil (Season 2)
Data output frequency	1 min
Data QC	SPICE QC methodology
Data temporal resolution	1 min
Processing interval for SPICE data analysis	30 min

Table 2: Data availability, by measurement season and site.

Measurement season	**Weissfluhjoch**
Season 1 (Oct. 2013 – Apr. 2014)	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓

Table 3: Summary of reference and ancillary measurements, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	Weissfluhjoch
R2 site reference	OTT Pluvio^2 Bucket RT (DFAR)
R2 precip detector	Thies LPM (DFAR)
Ancillary temp sensor	Meteolabor AG VTP6 Thygan (5 m)
Ancillary RH sensor	Meteolabor AG VTP6 Thygan (5 m)
Ancillary wind sensor	RM Young Wind Monitor 05103 (5.5 m)

4) Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6.1 of the WMO-SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

4.1.1.1. Characterization of performance in non-precipitating conditions

The assessment data are derived over 30 minute intervals during which the precipitation detector in the R2 reference configuration reports 0 minutes of precipitation. The accumulation over these intervals (accumulation in minute 30 – accumulation in minute 1), representing the variability of the gauge response due to wind, evaporation, temperature, etc., is recorded, along with the mean wind speed, and the change in temperature (temperature in minute 30 – temperature in minute 1).

4.1.1.2. Assessment of ability to detect and report accumulation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute assessment intervals:

- Ref 'Yes' : R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No' : R2 weighing gauge < 0.25 mm AND/OR precip detector recording < 18 min of precip;
- SUT 'Yes' : SUT accumulation ≥ 0.25 mm;
- SUT 'No' : SUT accumulation < 0.25 mm.

4.1.1.3. Assessment of ability to detect light precipitation

The data for this component of the assessment are derived in a similar manner as those in Section 4.1.1.2, but with different combinations of thresholds for the reference and SUT 'Yes' and 'No' conditions. These different threshold 'cases' have been selected to demonstrate the impact of the thresholds used in data derivation on the detection of light precipitation.

4.1.2. Skill score assessment

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores, as detailed in Section 3.6.1.3 of the WMO-SPICE Final Report.

For the assessments considered in this report, the ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

5) Environmental conditions

The environmental conditions at each site over the duration of the test period are expressed as probability density functions (PDFs) of mean air temperature, mean relative humidity, mean wind speed, vector mean wind direction, and precipitation rate for each component 30 minute assessment interval in Figure 3. The same parameters are also shown for all assessment intervals during which the site reference configuration detected precipitation (i.e. all Ref 'Yes' cases).

The precipitation percentage represents the number of minutes of precipitation during a 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration, expressed as a percentage. PDFs of precipitation percentage are also included in Figure 3.

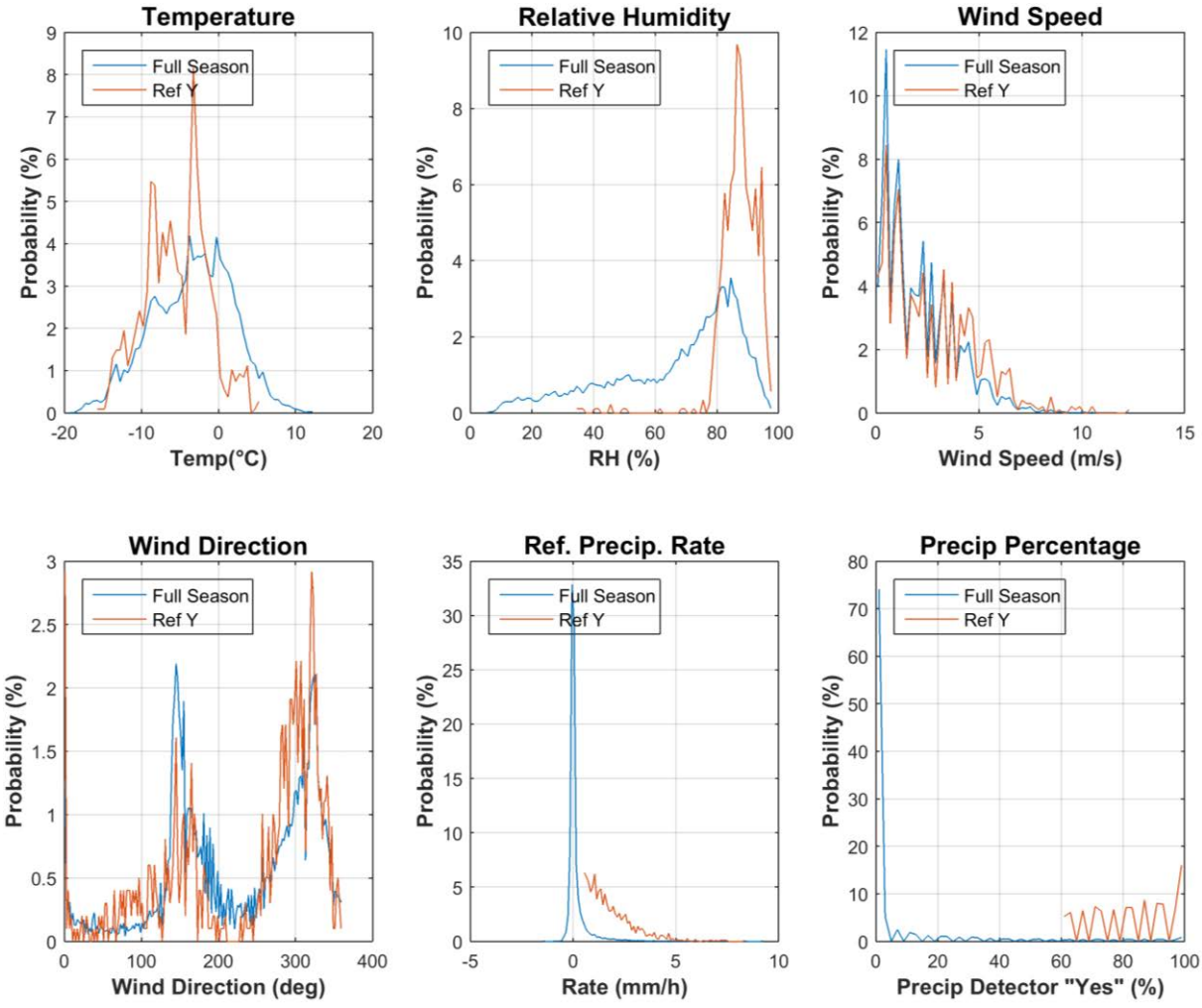


Figure 3: Summary of environmental conditions at Weissfluhjoch over the entire duration of formal tests (Full Season; blue PDFs) and during precipitation events reported by the site R2 reference (Ref Y; orange PDFs).

6) Evaluation of performance over the range of operating conditions

6.1. Characterization of SUT performance in non-precipitating conditions

The response of the SUT in the absence of precipitation was examined as defined in Section 4.1.1.1. The results are presented below, reflecting the distribution of the sensor response and its variability with wind and temperature, as measured during 30 minute assessment intervals.

6.1.1. Overall variability of SUT response

The overall variability of the SUT response in non-precipitating conditions is shown as a probability density function in Figure 4. The corresponding PDF for the reference configuration is provided for comparison.

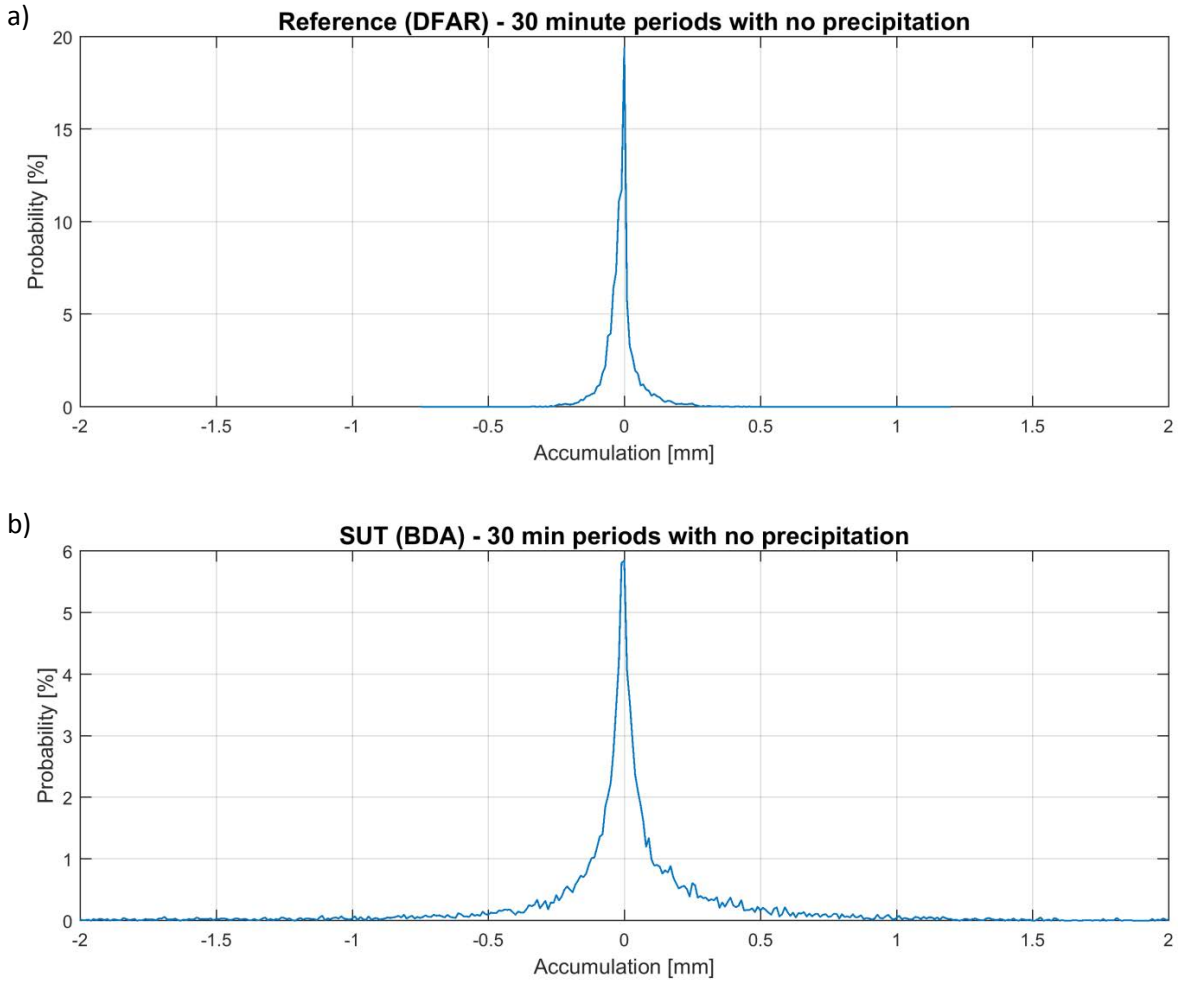


Figure 4: Probability density functions of the output signal (accumulation over 30 minute assessment intervals) in non-precipitating conditions for (a) the R2 reference configuration (DFAR) and (b) the Belfort AEPG 600 gauge in Belfort double-Alter shield at Weissfluhjoch.

The statistics of the output signal (accumulation over 30 minute assessment intervals) for the reference and SUT are provided in Table 4.

Table 4: Statistics of the R2 reference gauge and SUT output signal during non-precipitating conditions, as plotted in Figure 4.

Gauge	Average output signal (mm)	Standard deviation (mm)	Maximum output signal (mm)	Minimum output signal (mm)	Number of assessment intervals
Reference	-0.005	0.068	1.200	-0.750	11502
SUT	-0.002	0.607	7.872	-7.883	11502

6.1.2. Variability of SUT response as a function of temperature

The variability of the SUT response in the absence of precipitation for each assessment interval is plotted as function of the corresponding temperature difference in Figure 5. The temperature difference is defined as the difference in temperature between the end (minute 30) and beginning (minute 1) of the assessment interval. The corresponding plot for the reference configuration is provided for comparison.

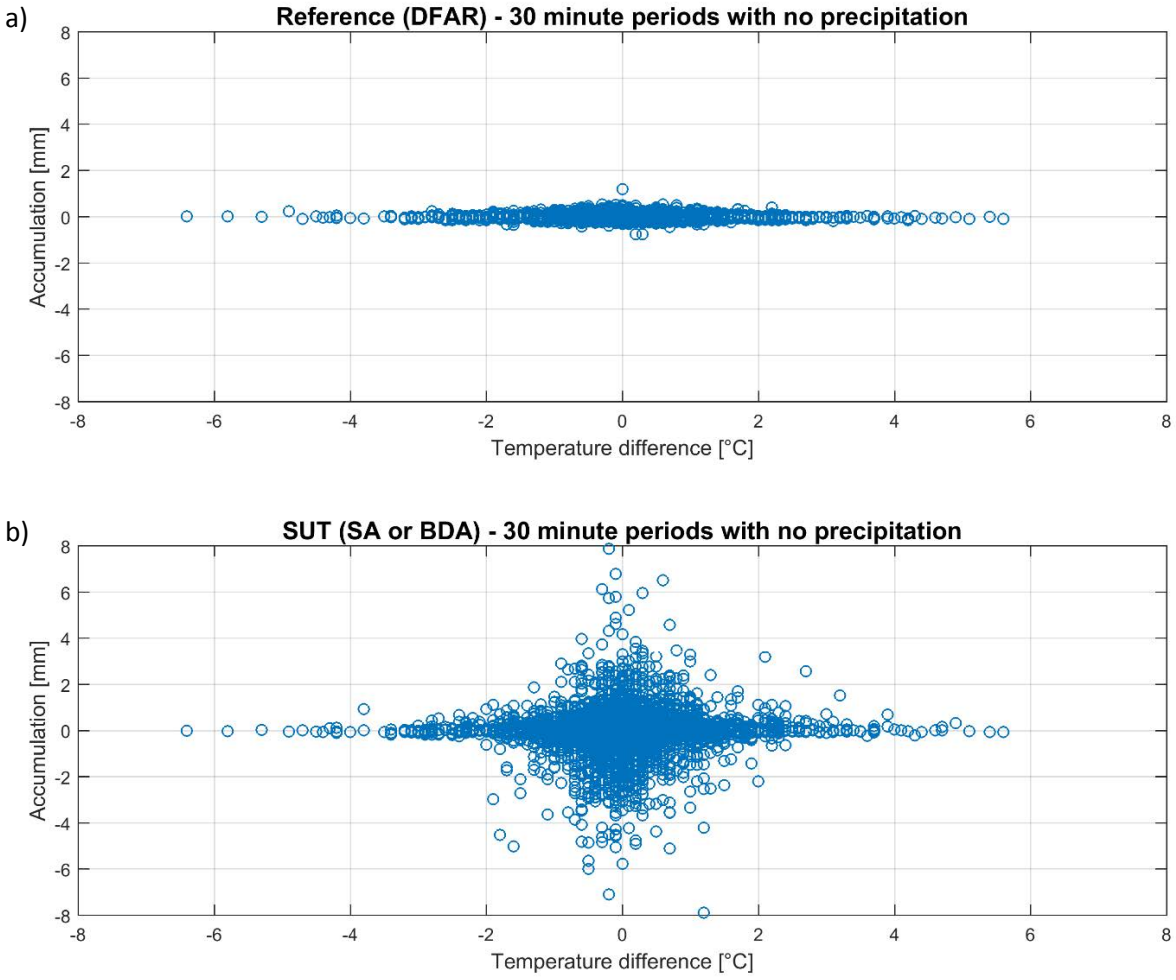


Figure 5: Variability of output signal (accumulation over each 30 minute assessment interval) as a function of the temperature difference over the interval in non-precipitating conditions for (a) the R2 reference configuration (DFAR) and (b) the Belfort AEPG 600 gauge in Belfort double-Alter shield at Weissfluhjoch.

6.1.3. Variability of SUT response as a function of wind speed

The variability of the SUT response in the absence of precipitation for each assessment interval is plotted as function of the mean wind speed in Figure 6. Here, the signal variability is represented as the standard deviation (STD) of the gauge accumulation output over each 30 minute interval. The corresponding plot for the reference configuration is provided for comparison.

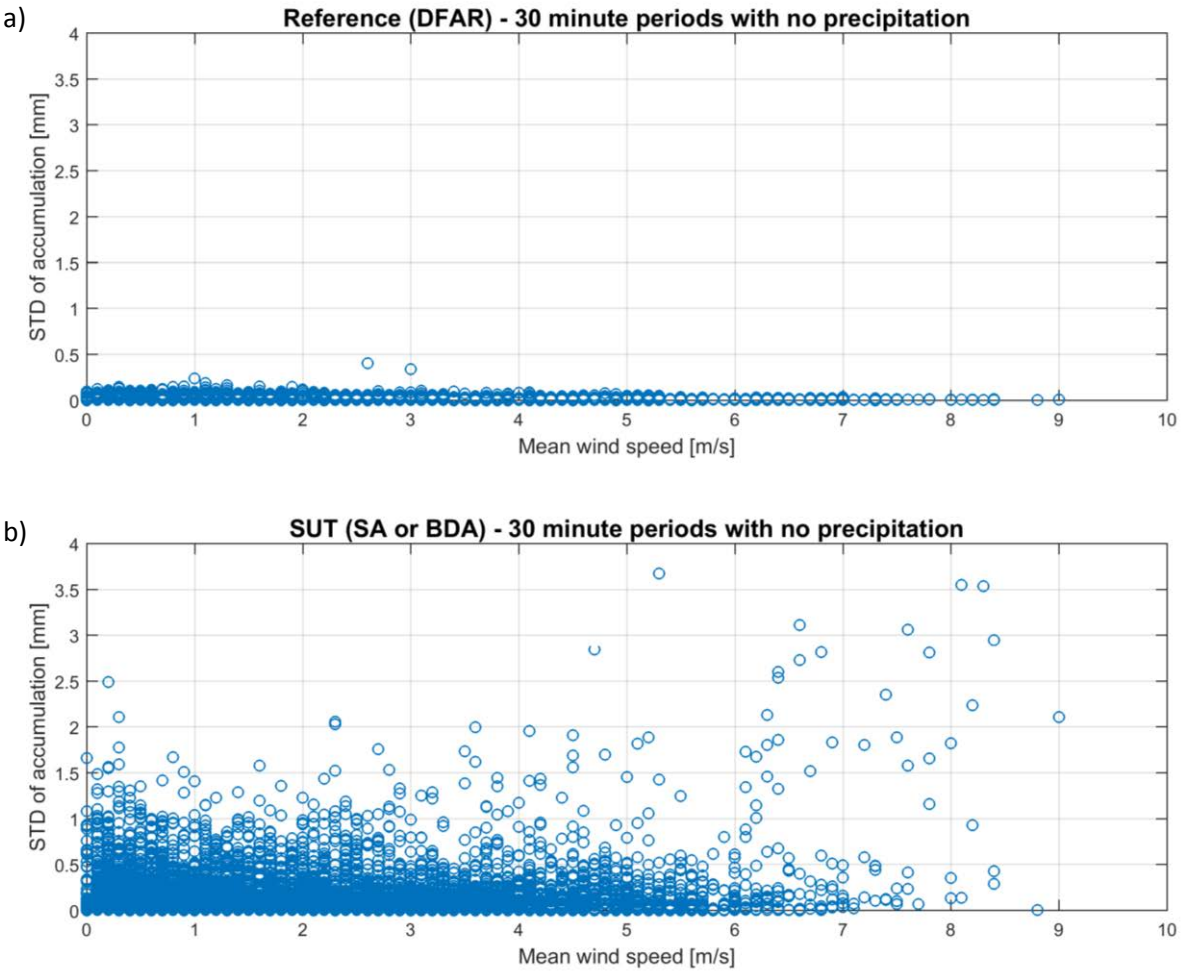


Figure 6: Variability of output signal (accumulation over each 30 minute assessment interval) as a function of mean wind speed in non-precipitating conditions for (a) the R2 reference configuration (DFAR) and (b) the Belfort AEPG 600 gauge in Belfort double-Alter shield at Weissfluhjoch.

6.2. Ability to detect and report precipitation

6.2.1. Skill score assessment

The overall ability of the SUT to detect and report the occurrence of precipitation relative to the site field reference R2 over 30 minute assessment intervals is expressed using selected skill scores (Section 4.1.2) and presented in Table 5. Scores are presented for 30 minute assessment intervals. The contingency results (Section 4.1.1) corresponding to these scores are presented in Table 6.

Table 5: Skill scores for Belfort AEPG 600 gauge at Weissfluhjoch during the formal test periods.

Probability of Detection, POD (%)	False Alarm Rate, FAR (%)	Bias, B (%)	Heidke Skill Score, HSS (%)
82.1	75.8	339	31.4

Table 6: Contingency table illustrating detection of precipitation by the Belfort AEPG 600 relative to the site reference at Weissfluhjoch, expressed as number of events over the entire test period.

Number of Events			
YY (hits)	YN (misses)	NY (false alarms)	NN (correct negatives)
885	193	2766	15131

6.3. Ability to report accumulated precipitation

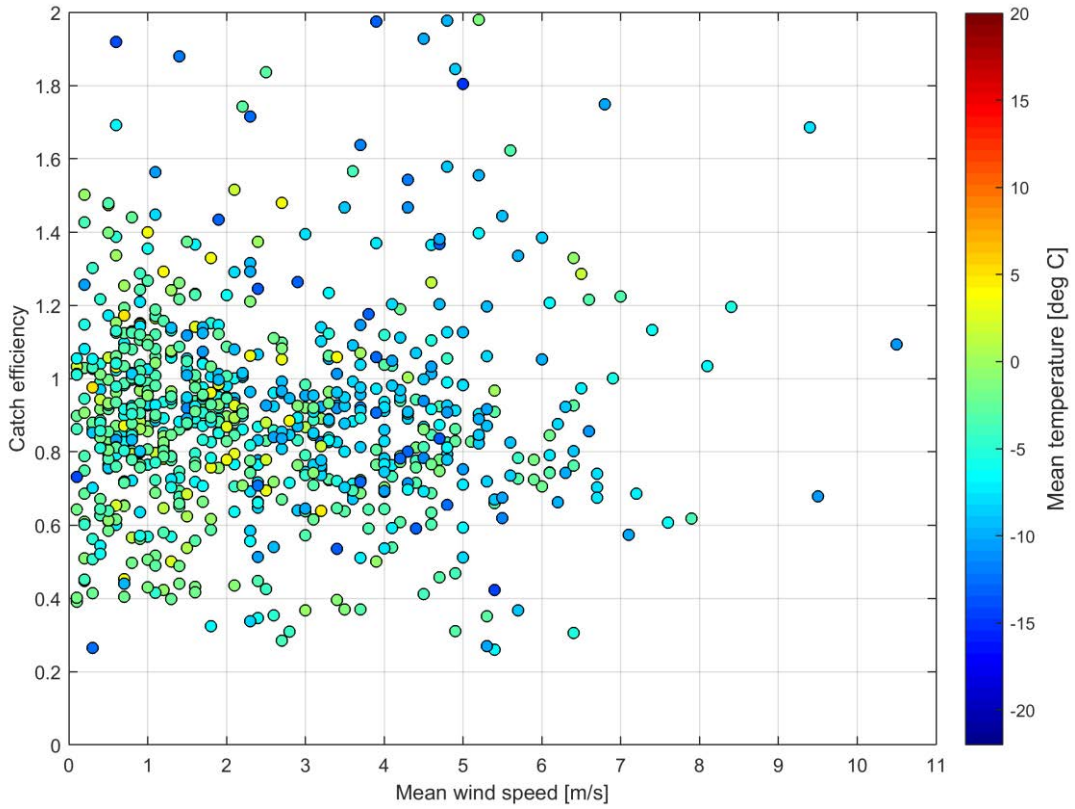
The SUT performance in terms of reporting accumulated precipitation is examined by comparing the amount reported by the sensor under test relative to the respective site reference during 30 minute assessment intervals. This is represented graphically using scatter and box and whisker plots of the catch efficiency as a function of mean wind speed at gauge height, as well as scatter plots of the amounts reported by the SUT versus the corresponding reference amounts (Figure 7). The SUT performance is also assessed in terms of the root mean square error, RMSE (Table 7).

Only assessment intervals during which the SUT and reference both reported precipitation (YY cases) are considered in this portion of the assessment. In the catch efficiency-wind speed scatter plots, the mean event temperature is indicated by colour, with the colour scale selected to be consistent across all sites with weighing gauges under test. In the box and whisker plots and accumulation-accumulation scatter plots, the predominant precipitation type is indicated by colour, as determined from the reported temperature (Section 4.1.4).

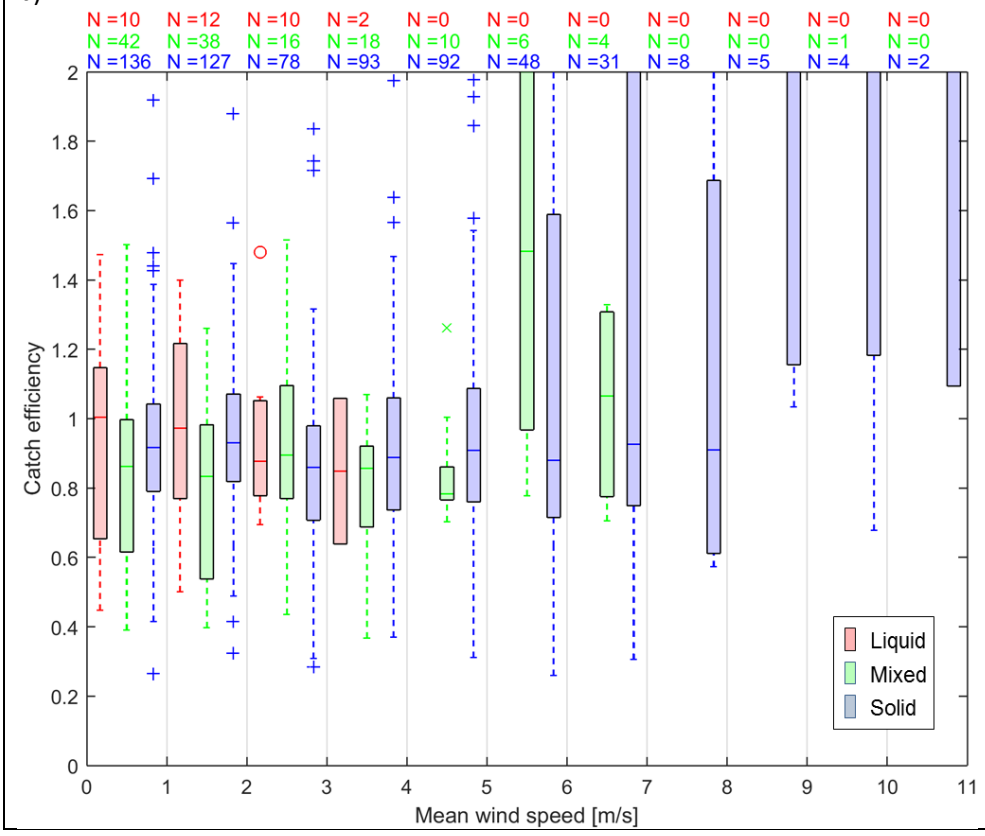
The catch efficiency plots in Figures 7a and 7b are limited to CE values ≤ 2 for clarity. For this particular dataset, approximately 7% of assessment intervals had CE values > 2, which impacted the resulting RMSE and overall CE values in Tables 7 and 8, respectively. To demonstrate the influence of these values, results are presented separately for all assessment intervals, and for those with CE values ≤ 2.

Figure 7: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the Belfort AEPG 600 gauge in Belfort double-Alter shield at Weissfluhjoch.

a)



b)



c)

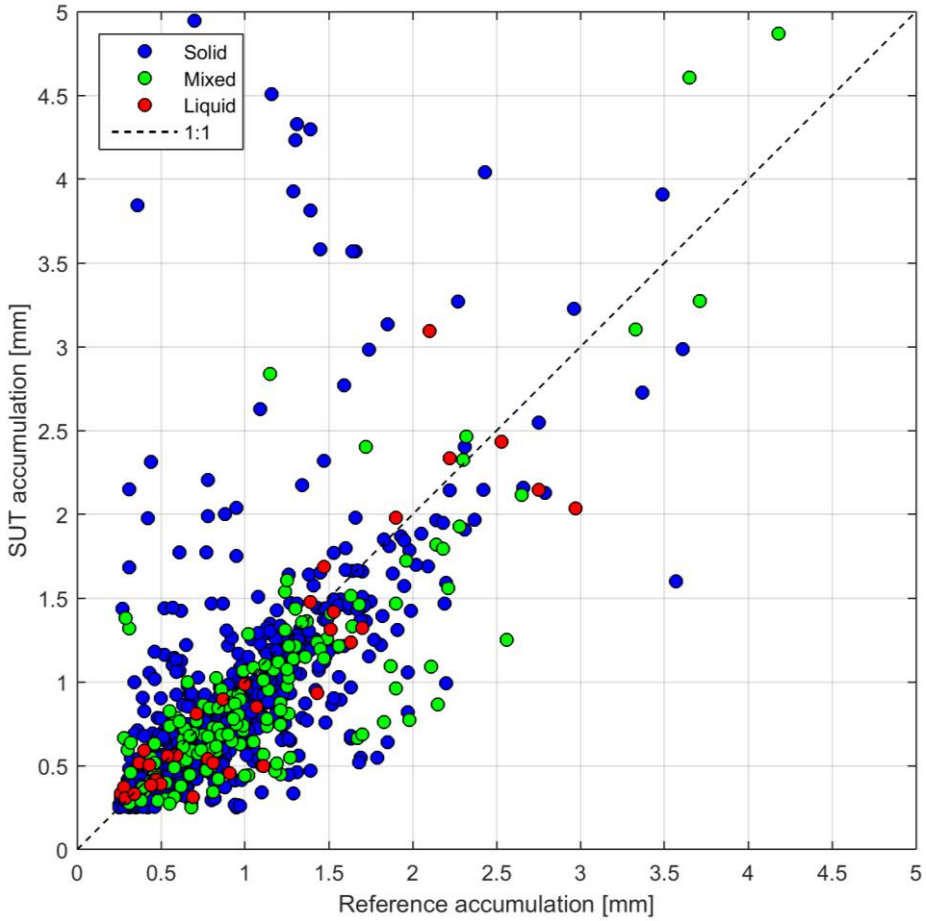

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5

Table 7: RMSE values for test configuration(s) by precipitation type for YY cases over the entire test period at Weissfluhjoch.

	RMSE (mm/30 min)			
	Liquid	Mixed	Solid	All precip types
All events	0.340	0.423	1.180	1.066
Events with CE ≤ 2	0.340	0.386	0.313	0.328

The overall catch ratio calculated using all 30 minute YY cases, over the entire test period, is provided in Table 8. To demonstrate the influence of the SUT accumulation threshold on the results, the overall catch ratio is also provided for all 30 minute YY cases determined using a lower SUT threshold of 0.1 mm/30 minutes. Note that these values reflect only the YY cases, and do not include the amounts corresponding to the cases when the SUT and the reference do not agree on the occurrence of precipitation.

Table 8: Overall catch ratio for test configuration(s) determined from YY cases over the entire test period at Weissfluhjoch, using two different SUT accumulation thresholds.

SUT accumulation threshold (mm/30 min)	Overall catch ratio	Overall catch ratio (events with CE ≤ 2)
0.25	1.10	0.88
0.1	1.10	0.86

6.4. Ability to detect light precipitation events

The impact of the threshold selection for data processing relative to the detection of light precipitation was examined using four different combinations of reference and SUT accumulation thresholds (four 'cases' in Table 9). Contingency results, probabilities of detection (POD), and false alarm rates (FAR) are presented for each case in Table 10. A quantitative comparison of the amounts reported in each case is beyond the scope of this assessment.

Table 9: Reference and SUT thresholds in each case for light precipitation detection assessment.

Case	Reference threshold (mm/30 min)	SUT threshold (mm/30 min)
1	0.25	0.25
2	0.1	0.1
3	0.25	No threshold
4	0.25	0

Table 10: Contingency results, probability of detection, and false alarm rate for each case in light precipitation detection assessment.

Case	Number of events				Skill score (%)	
	YY	YN	NY	NN	POD	FAR
1	885	193	2766	15131	82.1	75.8
2	1089	278	4909	12699	79.7	81.8
3	1078	0	17897	0	100	94.3
4	987	91	10224	7673	91.6	91.2

6.5. Assessment of events when the reference and the SUT do not agree on the occurrence of precipitation

Assessment intervals during which the site reference and SUT do not agree on the occurrence of precipitation – namely, the YN and NY cases (Section 4.1.1) – are characterized using histograms in Figures 8 and 9, respectively. The histograms include accumulated precipitation reported by the reference and SUT (0.25 mm/30 min thresholds for both), precipitation intensity as reported by the reference, and corresponding site conditions.

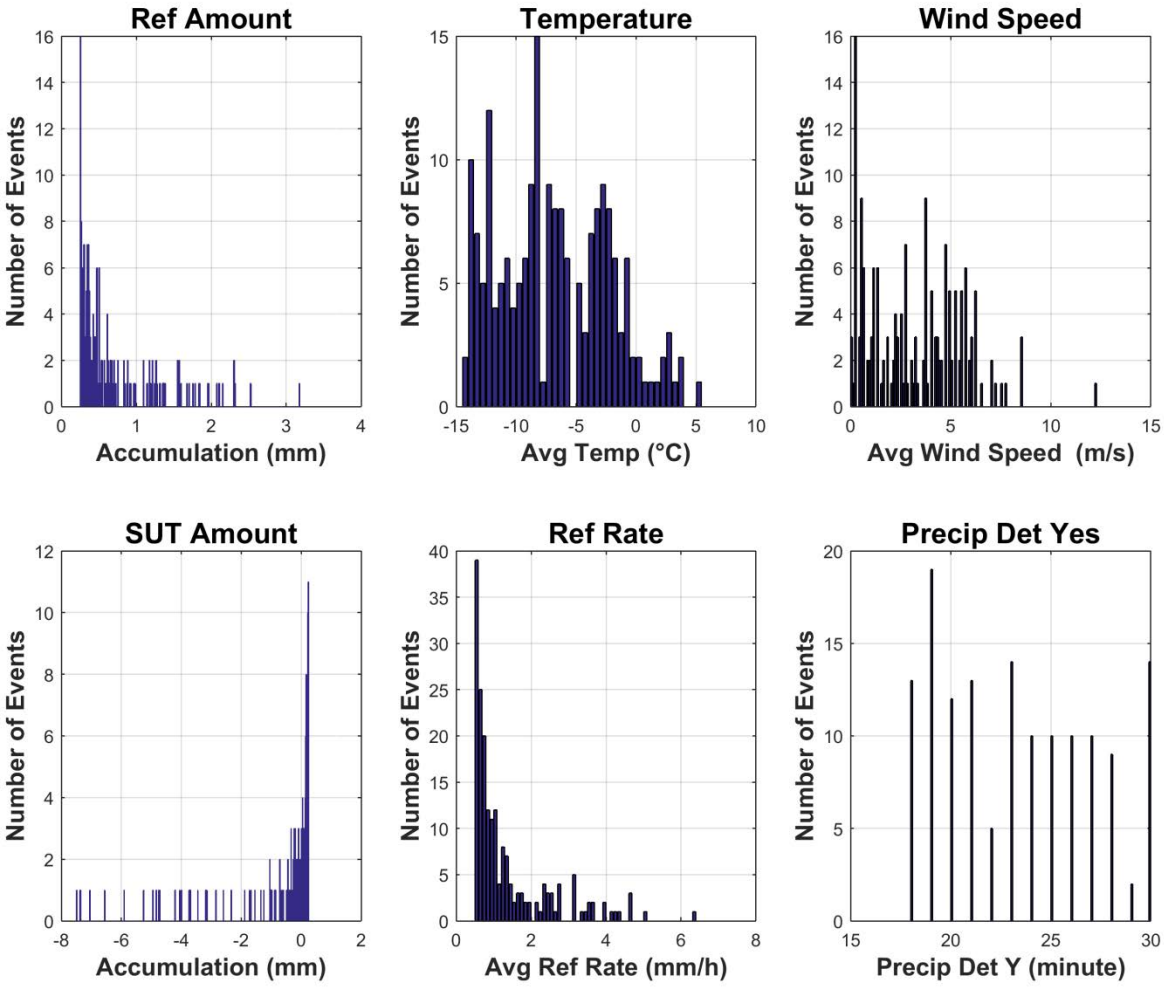


Figure 8: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for all YN events (193 total) over the test period.

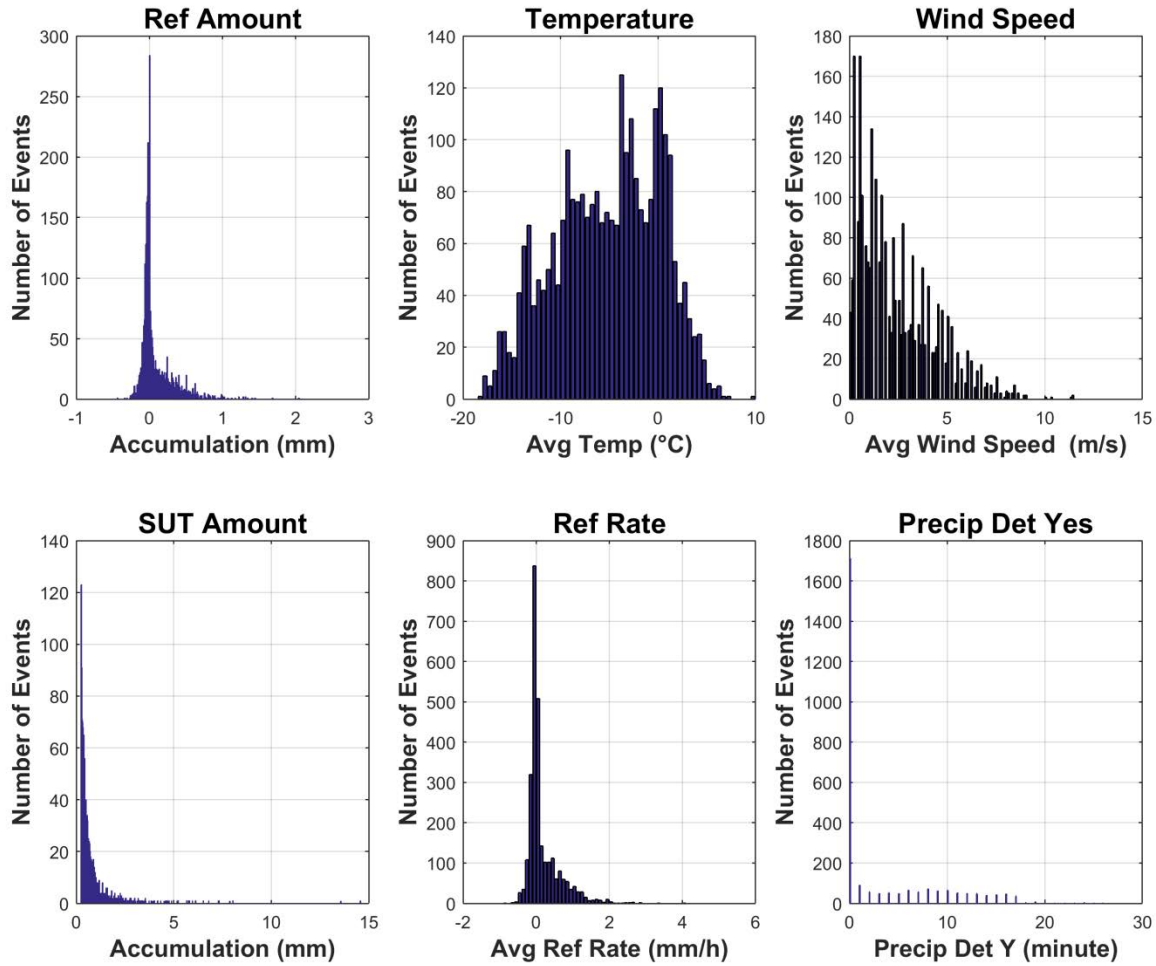


Figure 9: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for all NY events (2766 total) over the test period.

7) Interpretation of Results

The results indicate an instrument with significant uncertainty over the conditions tested; details are provided below. As the assessment results are based on only one test gauge, at one site, they cannot necessarily be applied to other gauges and/or sites. Additional testing of Belfort AEPG 600 gauges at sites in different climate regimes is recommended.

7.1. Operating conditions

The full range of conditions under which the test gauge was operated is illustrated in Figure 3. The conditions relevant to gauge operation are as follows:

- Temperatures between -20 °C and 13 °C;
- Precipitation intensity within 10 mm/hr.

The reported values fall within the manufacturer's specifications of -55 °C to 55 °C for temperature and 304 cm/hr for precipitation intensity.

7.2. Performance over the range of operating conditions

7.2.1. Non-precipitating conditions

The probability distribution of accumulation reports from the Belfort AEPG 600 test gauge during 30 min non-precipitating periods is broader than that of the reference configuration (Figure 4), indicating a higher magnitude of noise in the test gauge relative to the reference. Comparing the standard deviations of accumulation reports in non-precipitating conditions, the value for the test gauge exceeds that of the reference by almost an order of magnitude (Table 4), with the largest values observed at the highest wind speeds (Figure 6). Based on these results, it appears that the Belfort AEPG 600 test gauge at Weissfluhjoch is subject to significant noise/variability relative to the reference.

The magnitude of gauge responses in the absence of precipitation can be used to identify a detection threshold that minimizes the detection of false precipitation while enhancing the detection of light precipitation. This threshold is considered to be three times the standard deviation of the average gauge response during 30 minute non-precipitating periods. Based on the present results for the test gauge at Weissfluhjoch (Table 4), this minimum detection threshold is determined to be 1.8 mm for Belfort double-Alter shielded gauges.

7.2.2. Precipitating conditions

7.2.2.1. Ability to detect and report precipitation

The test gauge shows a high probability of detection (POD) for precipitation events (as defined in this assessment) relative to the reference, with a value of 82.1%. However, the gauge also shows a high false alarm rate (FAR) of 75.8%, likely resulting from the greater magnitude of noise/variability in the test gauge output relative to the reference noted above. The high Bias (B) value of 339% indicates

that the gauge over-reports the number of precipitation events relative to the reference, which can also likely be attributed to the variability in the test gauge output. The Heidke Skill Score (HSS), which considers the total number of correct 'Yes' and 'No' reports relative to the reference, as well as the number of correct responses expected due to chance alone, is 31.4%.

7.2.2.2. Ability to report accumulated precipitation

The results in Figure 7 illustrate significant variability in Belfort AEPG 600 gauge reports relative to those of the reference configuration over 30 minute intervals. This variability confounds the assessment of wind speed influence on catch efficiency in solid precipitation conditions; no clear trends are visible in the box and whisker plot for snow events in Figure 7b.

The results in Figure 7c indicate a number of events for which the SUT accumulation significantly exceeds the reference accumulation, falling above the 1:1 line. These are primarily solid precipitation events, and make up approximately 7% of the total number of events. Accordingly, the RMSE for solid events in Table 7 is 1.18 mm/30 min, compared to lower values of 0.34 mm/30 min and 0.42 mm/30 min for liquid and mixed precipitation, respectively. The overall RMSE, for all precipitation types, is 1.07 mm/min.

Limiting the assessment to events with catch efficiencies ≤ 2, the RMSE for liquid events is the same (0.34 mm/30 min), the value for mixed precipitation decreases slightly to 0.386 mm, while that for solid precipitation decreases significantly, to 0.31 mm/30 min (Table 7). The overall RMSE decreases to 0.33 mm/30 min. A similar trend is observed for the overall CE, which decreases from 1.10 for all events, to 0.88 for events with CE ≤ 2. The events with CE > 2 may represent contributions from blowing snow, either from elsewhere on the site and/or accumulated on the shield/gauge.

7.2.2.3. Ability to detect light precipitation

The detection thresholds and results presented in Tables 9 and 10, respectively, indicate that decreasing or removing the SUT detection threshold (while maintaining the reference detection threshold at 0.25 mm) leads to increases in both the Probability of Detection and the False Alarm Rate. The corresponding Heidke Skill Score values (not shown) decrease, indicating reductions in detection skill for the SUT. Decreasing both the reference and SUT detection thresholds to 0.1 mm, the Probability of Detection decreases while the False Alarm Rate increases. The resulting Heidke Skill Score value of 20.2% is lower than the 31.4% value for the standard 0.25 mm thresholds, indicating less detection skill for the SUT when considering contributions from light precipitation events with accumulations between 0.1 and 0.25 mm.

7.2.2.4. Assessment of events when the reference and SUT do not agree on the occurrence of precipitation

The YN ('miss') cases, when the reference detects a precipitation event and the SUT does not, are characterized in Figure 8. The majority of these cases have reference accumulations just above the 0.25 mm threshold, while the SUT accumulation is just below the threshold. As the SUT is more susceptible to the influence of wind relative to the reference gauge in a DFIR fence, it is likely that for

events with reference accumulation just above the threshold, the wind effect would reduce the SUT accumulation below the detection threshold.

The NY ('false alarm') cases, when the SUT detects a precipitation event and the reference does not, are characterized in Figure 9. Many of these cases show reference accumulations of 0 mm and SUT accumulations just above the 0.25 mm thresholds, which are attributed to the influence of noise in the SUT output. As the standard deviation of accumulation reports in the absence of precipitation (as indicated by an independent precipitation detector) is approximately 0.6 mm – over twice the precipitation detection threshold of 0.25 mm – it is clear that there is potential for accumulation reports ≥ 0.25 mm in the absence of precipitation.

8) Maintenance

Gauge calibration: each site completed the gauge field calibration and verification as per manufacturer recommendations, at least once a year or following the emptying of the gauge. The calibration records have been stored by each site host.

9) Performance Considerations

9.1. Data processing and outputs

The data output from Belfort AEPG 600 test gauge shows significant noise relative to the reference configuration at Weissfluhjoch, as indicated by comparison of outputs in non-precipitating conditions (Section 6.1.). Further investigation is required to identify the source of the noise, but it may be related to the specific configuration of the double-shield (see Section 9.2). Further processing of gauge outputs using filters, such as those employed as part of the SPICE data quality control procedure, is recommended to mitigate the influence of noise and facilitate subsequent analyses and interpretation.

9.2. Gauge configuration and operation

The Belfort AEPG 600 was tested in a Belfort double-Alter shield at Weissfluhjoch. With no unshielded or single-shielded gauges of the same type available for comparison, it is difficult to assess any potential benefits of the double-shield for gauge performance at increasing wind speeds. As noted above, the specific configuration of the double-shield employed at Weissfluhjoch may impact gauge performance. The shield is mounted on the gauge post (Figure 1), and may be susceptible to wind-induced vibration that can affect the gauge measurements. This is one potential explanation for the high magnitude of noise in the test gauge relative to the reference configuration (Sections 6.1 and 7.2.1). It is recommended that the shield assembly be installed and mounted separately from the gauge post, if possible, to eliminate the potential for shield-related noise in the gauge output.

Another aspect of the shield configuration that warrants consideration is that the Belfort double-Alter shield slat configuration employs springs and rubber gaskets that may require maintenance. This is a more significant concern for remote or unattended sites, as gauge performance may be impacted by periods when the slats are not functioning properly, or have fallen off. No such issues were encountered for the test gauge at Weissfluhjoch during SPICE.

It is recommended that all field configurations be fully tested and validated prior to use, in order to ensure that physical and signal interferences are not degrading the gauge signal. This would include the confirmation of grounding tailored to the soil conditions, sturdiness of foundation and mounting, and signal conditioning.

9.3. Ancillary measurements and adjustments

While the present dataset and analysis show no apparent relationship between catch efficiency and wind speed (due to noise in the gauge output), the application of adjustment functions is generally recommended to account for reductions in gauge catch efficiency with increasing wind speed, when such functions are available. Ancillary measurements of wind speed (preferably at gauge orifice height) and air temperature are required for the application of adjustment functions. Ancillary measurements from a sensitive precipitation detector that is independent from the test gauge are also recommended to help distinguish precipitation events from false reports due to noise.

WMO-SPICE Instrument Performance Report
Geonor T-200B3, 600mm

1) Technical Specifications (from manufacturer provided documentation)

Instrument model:	Geonor T-200B3
Physical principle:	Weighing gauge (WG) based on catchment principle; weight measurement by vibrating wire transducers.
Capacity volume :	600 mm
Collecting area :	200 cm^2
Operating temperature range:	-40 °C to 60 °C
Measurement uncertainty	0.1% of full scale (as defined by manufacturer)
Sensitivity	0.05 mm (as defined by manufacturer)

a) b)

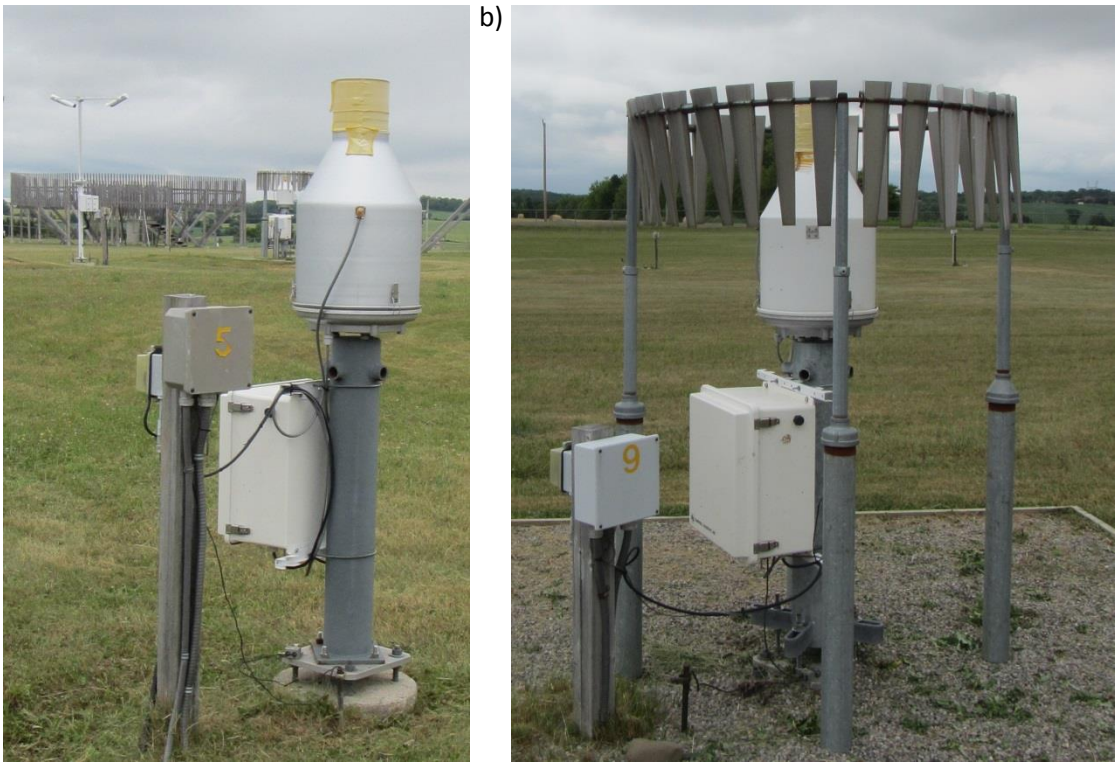


Figure 1: Geonor T-200B3 test gauges in (a) unshielded and (b) single-Alter shielded configurations at CARE site (Canada).

2) Data Output Format

Transducer output: Frequency (f); 0-5 V square wave

Data range: From ~1000 Hz (empty gauge) to ~3000 Hz (full gauge)

Gauge data output: Precipitation accumulation for each transducer, derived from the frequency output as outlined below. The arithmetic average of accumulation from the three transducers was used in subsequent data analysis.

The bucket content is weighed using a precision load cell with a high-tension vibrating wire (VW) transducer. Under load, the wire vibration frequency is proportional to the weight detected (and the associated precipitation amount, P), based on a quadratic relationship.

$P = A (f – f_0) + B (f – f_0)^2$

Where:

P = precipitation (in cm)
f = frequency reading (Hz)
A, B = calibration constants, available from the Calibration Certificates
$\mathbf{f_0}$ = frequency with empty bucket at calibration (Hz), available from the Calibration Certificates

The signal from a Geonor gauge transducer is amplified into a measurable quantity sampled with an external data logger, using a user defined strategy (duration, frequency) and logger specific functions. The sampling of the transducer signal is, generally, not continuous.

All sites testing Geonor T-200B3 gauges used Campbell Scientific CR3000 data loggers, and applied one of the two dedicated programming functions available:

1. Period average (PeriodAvg) command, which calculates the sensor's average frequency over a user-specified number of cycles within a defined interval. For example, the CARE SPICE site used a sampling strategy in which 250 cycles every sampled every 6 seconds, for each transducer.

2. Pulse count (PulseCount) command, which calculates the sensor's frequency by counting the number of pulses over a specified time period, which is then converted into a frequency.

3) SPICE test configuration

Number of transducers:	Three vibrating wire transducers for each test gauge, sampled independently using data loggers
Shield(s):	Unshielded (UN) and single-Alter Shield (SA)
Test Site(s):	CARE, Bratt's Lake, Caribou Creek (Canada); Marshall (USA);
Sensor Provider(s):	All gauges evaluated were provided by the host organizations; they were also part of the SPICE R3 reference configuration.

A map of test site locations is provided in Figure 2.

3.1. Note on terms and acronyms used

Throughout this document, the following notations were used to identify the R2 reference (Ref) and senor under test (SUT) configurations:

Reference: 'DFIR' and 'DFAR' are used interchangeably for the R2 reference configurations. 'DFIR' refers to an automated gauge installed in a DFIR-fence, while 'DFAR' refers more explicitly to the Double-Fence Automated Reference configuration.

Sensors under test: The Geonor T-200B3 test gauge with 600 mm capacity is shortened to 'Geo600' in some plot labels. 'SA' denotes single-Alter shielded test configurations and 'UN' denotes unshielded test configurations.

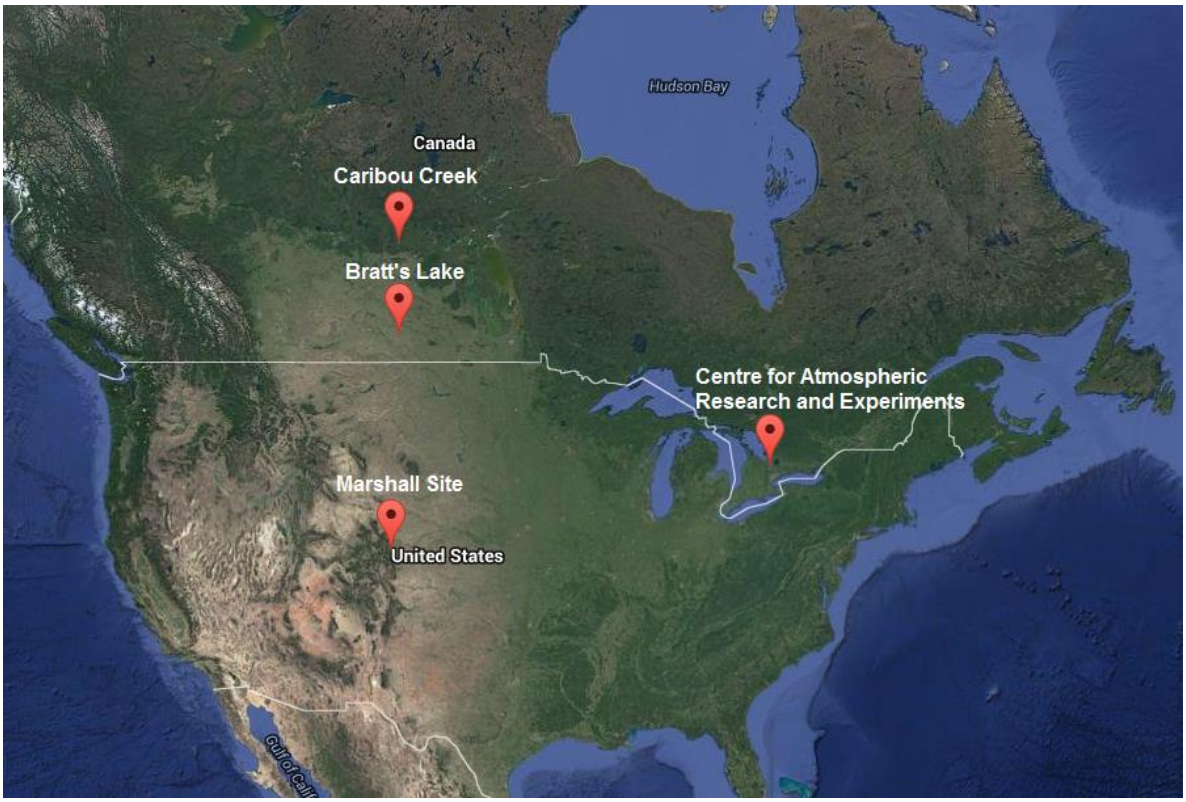


Figure 2: Map of SPICE sites where Geonor T-200B3 gauges were tested.

A summary of the configuration of instruments as tested, the duration of tests and availability of data reflected in these results, and the ancillary measurements used, by site, is available in Tables 1, 2, and 3, respectively.

Table 1: Summary of gauge configurations and data output, by site. Details and photos on individual site configurations are available in the respective site commissioning protocols.

	Bratt's Lake	CARE	Caribou Creek	Marshall
Field configuration (shield)	UN, SA	UN, SA	UN, SA	UN, SA
Height of installation (gauge rim)	2.2 m	2 m	2.2 m	2 m
Heating	SPICE algorithm	SPICE algorithm	No heating	SPICE algorithm
Antifreeze	Mixture of 60% methanol, 40% propylene glycol			
Oil	VoltEsso35	Bayoil (Season 1) Isopar (Season 2)	VoltEsso35	Automatic Transmission Fluid (ATF)
Data output frequency	1 min	6 sec	1 min	6 sec
Data QC	SPICE QC methodology			
Data temporal resolution	1 min			
Processing interval for SPICE data analysis	30 min			

Table 2: Data availability, by measurement season and site.

Measurement season	Bratt's Lake	CARE	Caribou Creek	Marshall
Season 1 (Oct. 2013 – Apr. 2014)	✓	✓	X	✓
Season 2 (Oct. 2014 –Apr. 2015)	✓	✓	✓	✓

Table 3: Summary of reference and ancillary measurements, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	Bratt's Lake	CARE	Caribou Creek	Marshall
R2 Site Reference	Geonor T-200B3 600 mm (DFAR)	Geonor T-200B3 600 mm (DFAR)	Geonor T-200B3 600 mm (DFAR)	Geonor T-200B3 600 mm (DFAR)
R2 Precip Detector	Thies Precip Sensor (DFAR)	Thies LPM (DFAR)	Thies Precip Sensor (DFAR)	Thies LPM (Site*)
Ancillary Temp Sensor	Vaisala HMP45 (2m)	Vaisala HMP155 (SS+, 1.5 m)	Vaisala HMP45 (2m)	MetOne, 060A-2/062, 2144-L (2 m)
Ancillary RH Sensor	Vaisala HMP45 (2m)	Vaisala HMP155 (SS+, 1.5 m)	Vaisala HMP45 (2m)	Campbell Scientific CS500 (2m)
Ancillary Wind Sensor	RM Young Wind Monitor 05103 (2m)	Vaisala NWS 425 (2m)	MetOne 13A (2m)	RM Young Wind Monitor 05103 (2 m)

*A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR shield; however, in cases where a more sensitive detector is available outside of the DFIR shield, or there are issues with the detector within the DFIR shield, a precipitation detector elsewhere on the site can be employed.

+ SS denotes that the sensor is installed inside a ventilated Stevenson Screen.

4) Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6.1 of the WMO-SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

4.1.1.1. Characterization of performance in non-precipitating conditions

The assessment data are derived over 30 minute intervals during which the precipitation detector in the R2 reference configuration reports 0 minutes of precipitation. The accumulation over these intervals (accumulation in minute 30 – accumulation in minute 1), representing the variability of the gauge response due to wind, evaporation, temperature, etc., is recorded, along with the mean wind speed, and the change in temperature (temperature in minute 30 – temperature in minute 1).

4.1.1.2. Assessment of ability to detect and report accumulation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute assessment intervals:

- Ref 'Yes' : R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No' : R2 weighing gauge < 0.25 mm AND/OR precip detector recording < 18 min of precip;
- SUT 'Yes' : SUT accumulation ≥ 0.25 mm;
- SUT 'No' : SUT accumulation < 0.25 mm.

4.1.1.3. Assessment of ability to detect light precipitation

The data for this component of the assessment are derived in a similar manner as those in Section 4.1.1.2, but with different combinations of thresholds for the reference and/or SUT 'Yes' and 'No' conditions. These different threshold 'cases' have been selected to demonstrate the impact of the thresholds used in data derivation on the detection of light precipitation.

4.1.2. Skill score assessment

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores, as detailed in Section 3.6.1.3 of the WMO-SPICE Final Report.

For the assessments considered in this report, the ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

5) Environmental conditions

The environmental conditions at each site over the duration of the test period are expressed as probability density functions (PDFs) of mean air temperature, mean relative humidity, mean wind speed, vector mean wind direction, and precipitation rate for each component 30 minute assessment interval in Figure 3. The same parameters are also shown for all assessment intervals during which the site reference configuration detected precipitation (i.e. all Ref 'Yes' cases) in Figure 4.

The precipitation percentage represents the number of minutes of precipitation during a 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration, expressed as a percentage. PDFs of precipitation percentage are also included in Figures 3 and 4.

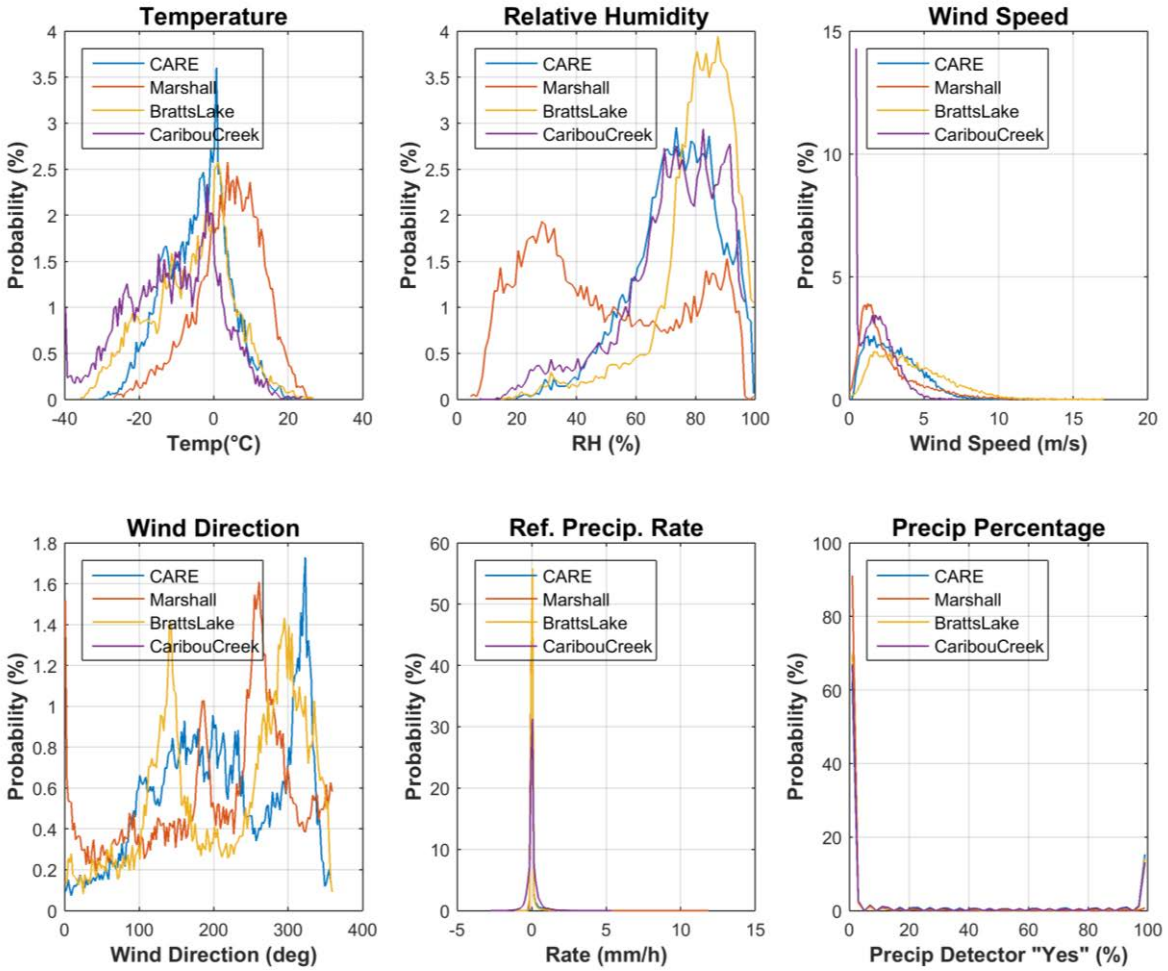


Figure 3: Summary of environmental conditions at sites with Geonor T-200B3 test gauges over the entire duration of formal tests, as per Table 2.

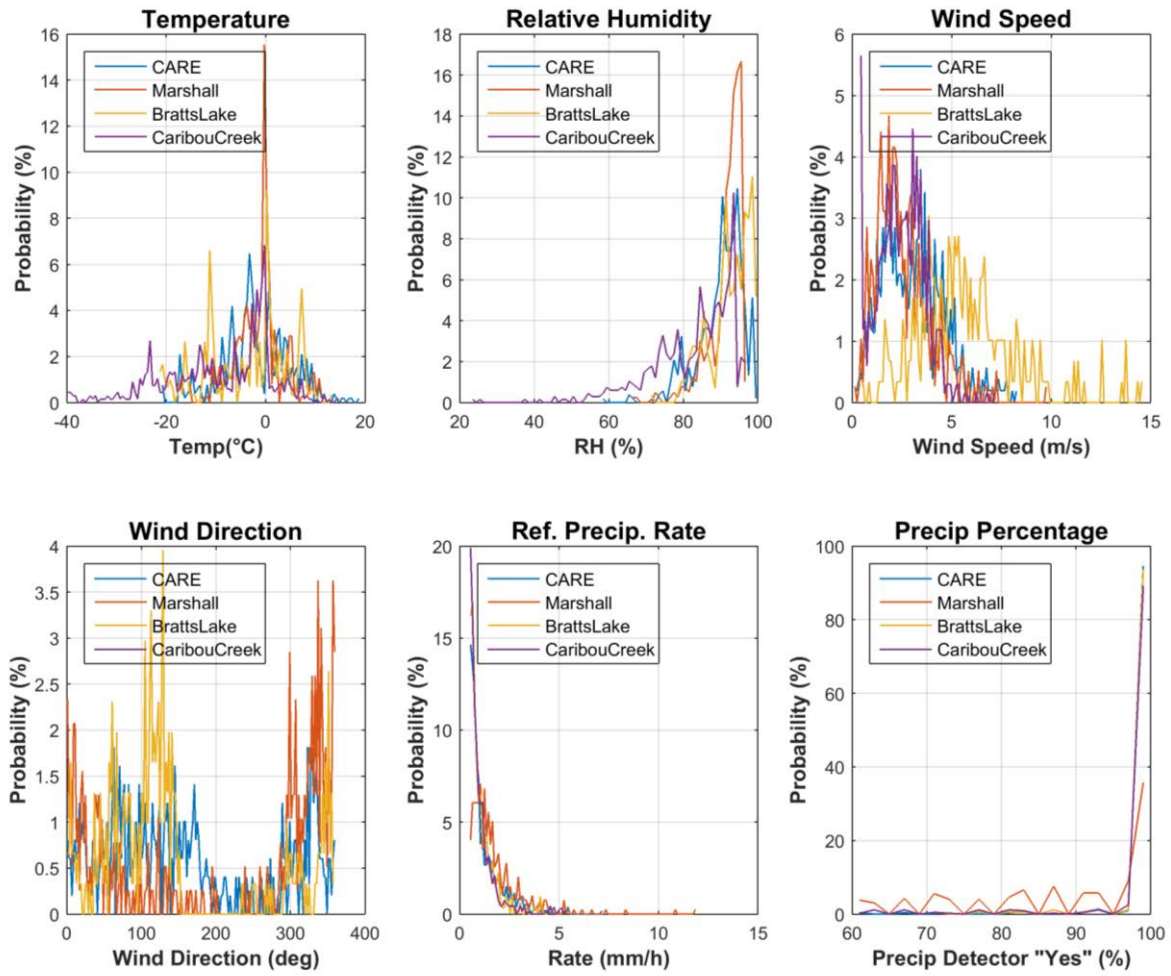


Figure 4: Summary of environmental conditions at sites with Geonor T-200B3 test gauges during precipitation events (as defined by the R2 reference), over the duration of formal tests.

6) Evaluation of the ability to perform over the range of operating conditions

6.1. Characterization of SUT performance in non-precipitating conditions

The response of the SUT in the absence of precipitation was examined as defined in Section 4.1.1.1. The results are presented below, reflecting the distribution of the sensor response and its variability with wind and temperature, as measured during 30 minute assessment intervals.

6.1.1. Overall variability of SUT response

The overall variability of the SUT response in non-precipitating conditions is shown as a probability density function for each test configuration in Figure 5. The corresponding PDF for the reference configuration at each test site is provided for comparison.

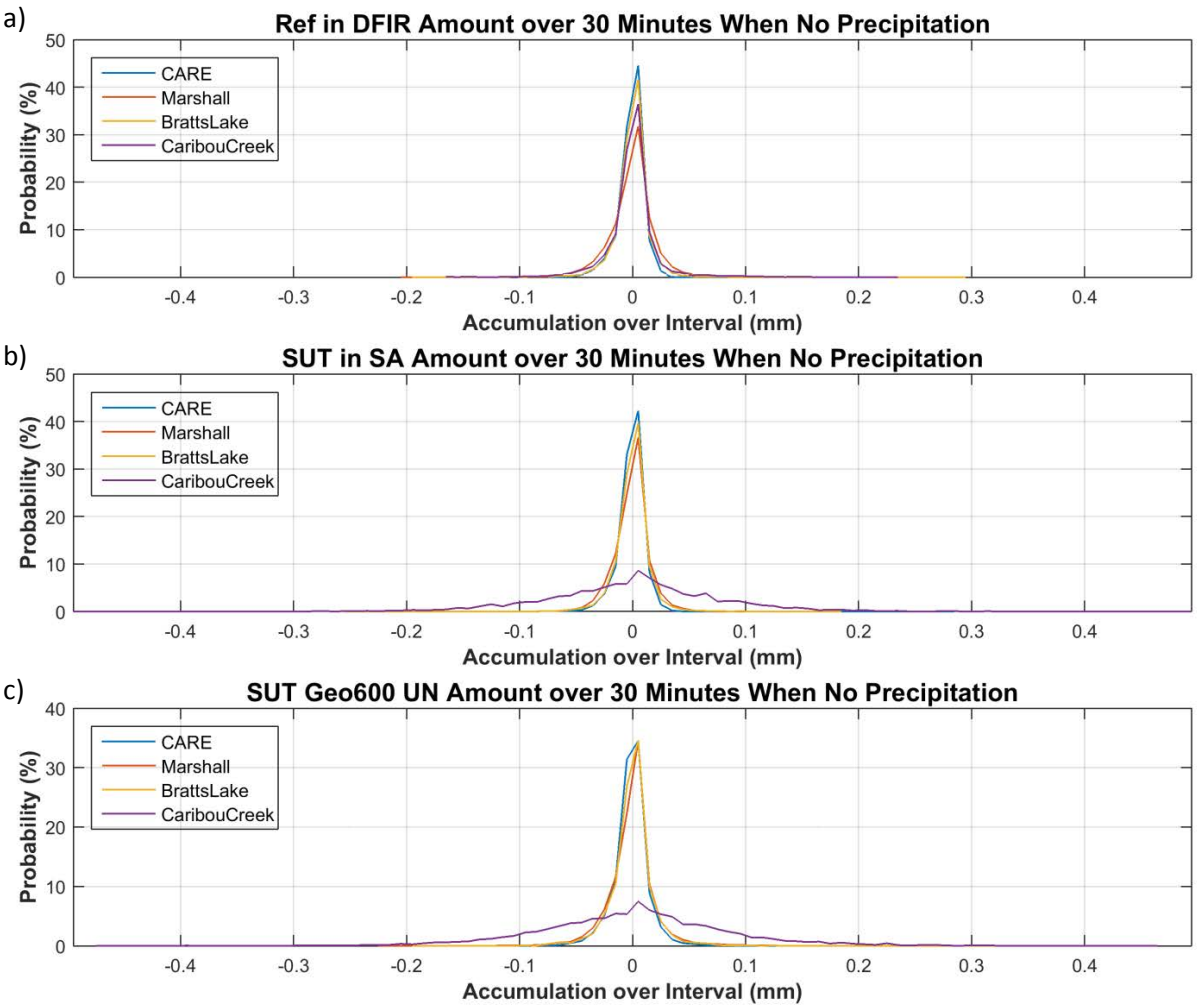


Figure 5: Probability density functions of the output signal (accumulation over 30 minute assessment intervals) in non-precipitating conditions for (a) the R2 reference configurations (DFAR) and Geonor T-200B3 test gauges in (b) single-Alter shielded and (c) unshielded configurations.

The statistics of the output signal (accumulation over 30 minute assessment intervals) for the reference and SUT at each site are provided in Table 4.

Table 4: Statistics of the R2 reference gauge (DFAR) and SUT output signal during non-precipitating conditions at sites testing Geonor T-200B3 gauges, as plotted in Figure 5.

Site	Configuration	Average output signal (mm)	Standard deviation (mm)	Maximum output signal (mm)	Minimum output signal (mm)	Number of assessment intervals
Bratt's Lake	DFAR	-0.000	0.016	0.293	-0.194	7033
	SA	-0.000	0.015	0.190	-0.259	7033
	UN	-0.001	0.019	0.321	-0.197	6740
CARE	DFAR	-0.001	0.011	0.204	-0.071	9449
	SA	-0.001	0.011	0.282	-0.070	9449
	UN	-0.001	0.015	0.199	-0.220	9448
Caribou Creek	DFAR	-0.000	0.024	0.233	-0.168	4897
	SA	-0.000	0.079	0.498	-0.492	4897
	UN	0.000	0.085	0.463	-0.473	4897
Marshall	DFAR	-0.001	0.022	0.196	-0.207	18176
	SA	-0.001	0.015	0.141	-0.143	18176
	UN	-0.001	0.021	0.186	-0.223	18222

6.1.2. Variability of SUT response as a function of temperature

The variability of the SUT response for each test configuration in the absence of precipitation is plotted as function of the temperature difference over each assessment interval in Figure 6. The temperature difference is defined as the difference in temperature between the end (minute 30) and beginning (minute 1) of the assessment interval. The corresponding plots for the reference configurations are provided for comparison.

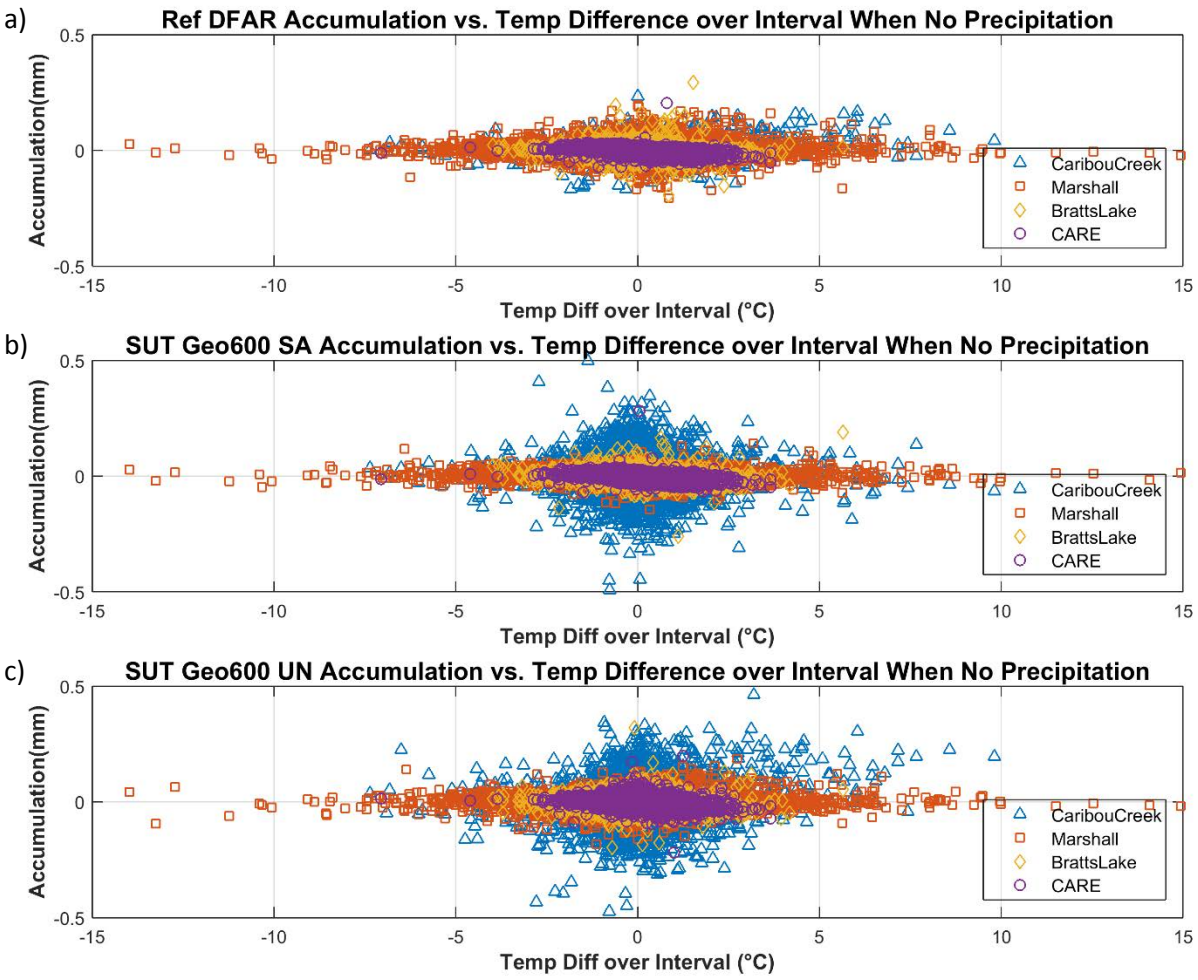


Figure 6: Variability of output signal (accumulation over each 30 minute assessment interval) as a function of the temperature difference over the interval in non-precipitating conditions for (a) R2 reference configurations (DFAR) and Geonor T-200B3 test gauges in (b) single-Alter shielded and (c) unshielded configurations.

6.1.3. Variability of SUT response as a function of wind speed

The variability of the SUT response for each test configuration in the absence of precipitation is plotted as function of the mean wind speed for each assessment interval in Figure 7. Here, the signal variability is represented as the standard deviation (STD) of the gauge accumulation output over each 30 minute interval. The corresponding plots for the reference configurations are provided for comparison.

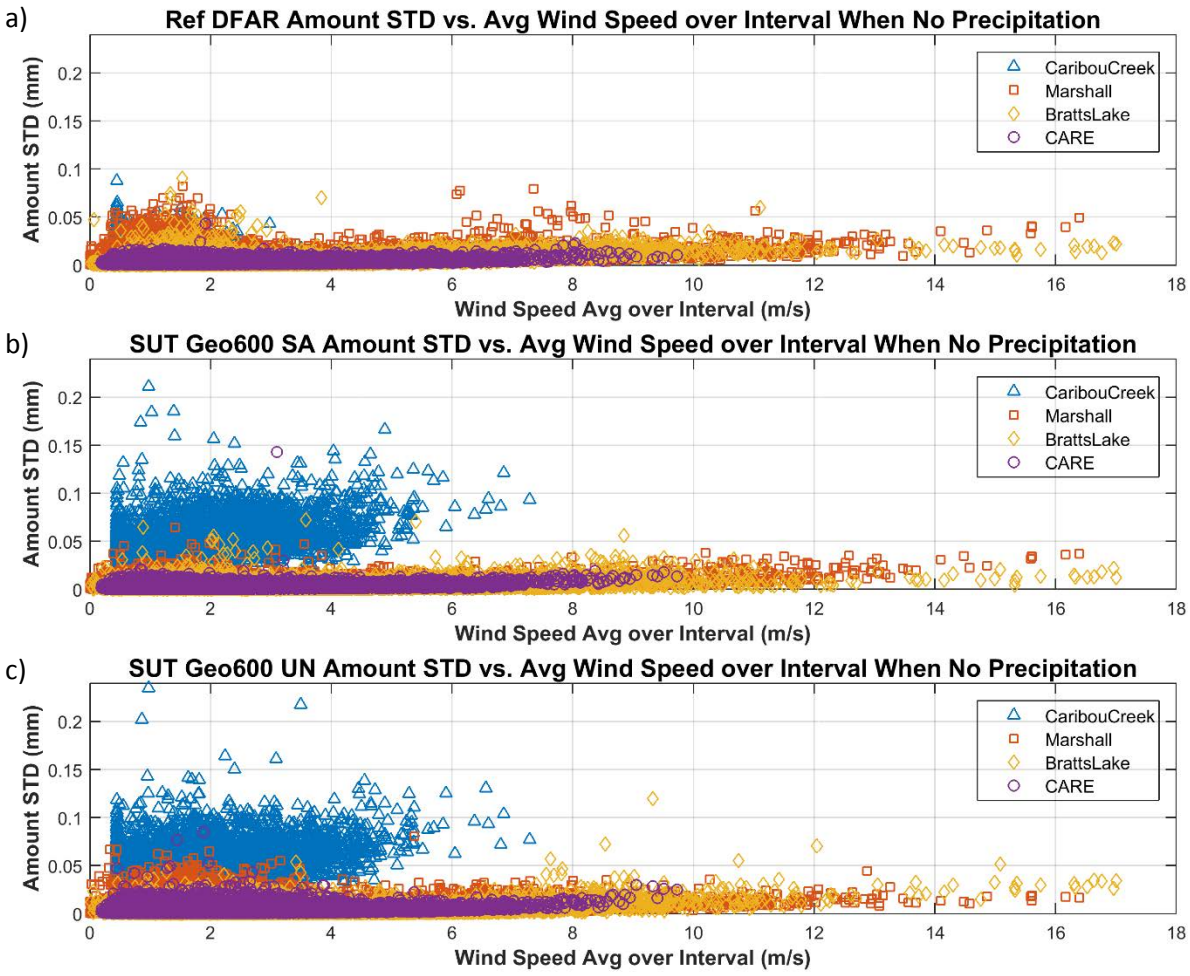


Figure 7: Variability of output signal (accumulation over each 30 minute assessment interval) as a function of mean wind speed in non-precipitating conditions for (a) R2 reference configurations (DFAR) and Geonor T-200B3 test gauges in (b) single-Alter shielded and (c) unshielded configurations.

6.2. Ability to detect and report precipitation

6.2.1. Skill score assessment

The overall ability of the SUT to detect and report the occurrence of precipitation relative to the site field reference R2 over 30 minute assessment intervals is expressed using selected skill scores (Section 4.1.2) and presented in Figure 8. The contingency results (Section 4.1.1) corresponding to these scores are presented in Table 5.

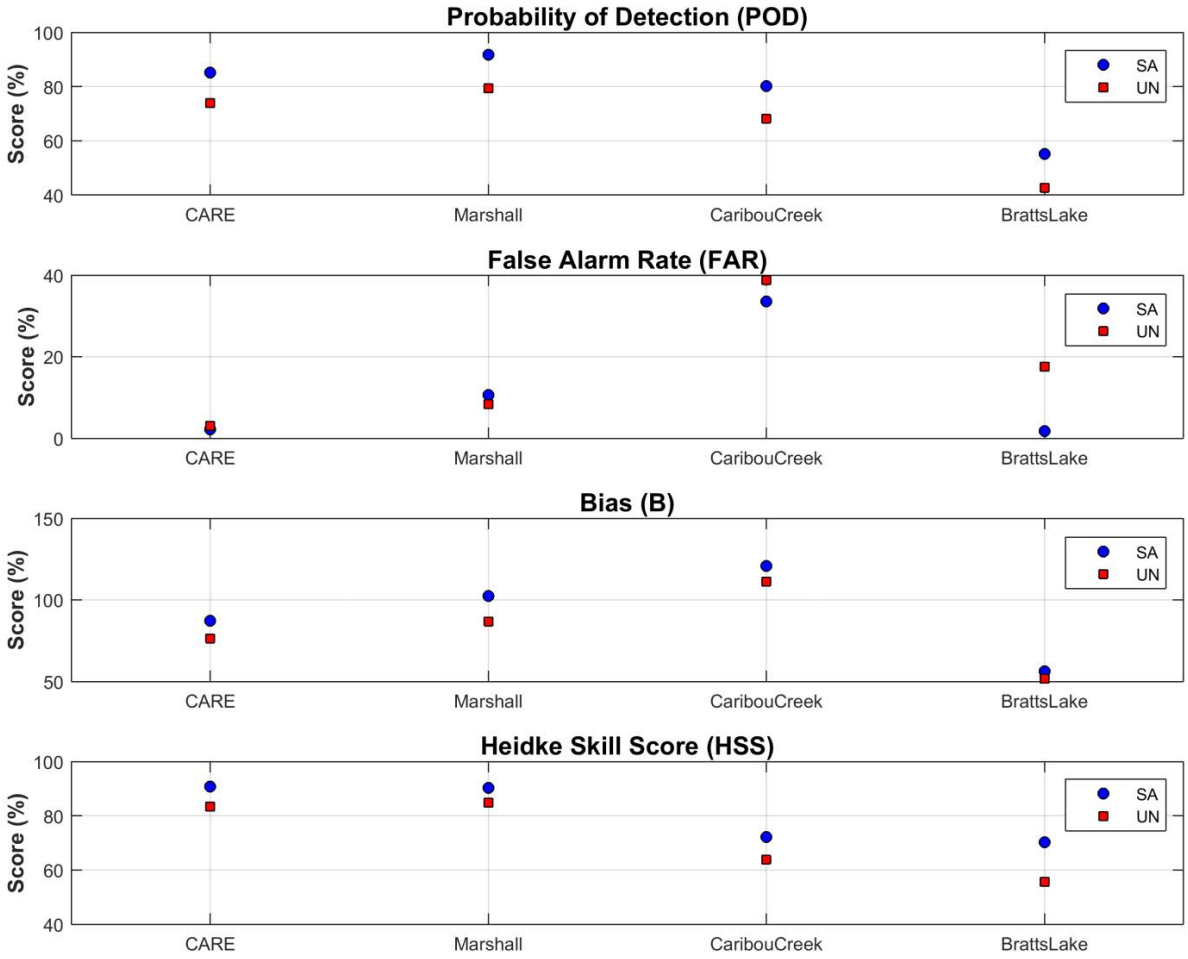


Figure 8: Skill scores for Geonor T-200B3 gauges over the duration of formal tests.

Table 5: Contingency table illustrating detection of precipitation by Geonor T-200B3 gauges under test relative to the corresponding site reference configurations, expressed as the number of events over the duration of formal tests.

Site	Shield	Number of Events			
		YY (hits)	YN (misses)	NY (false alarms)	NN (correct negatives)
Bratt's Lake	SA	170	138	3	16154
	UN	127	171	27	15211
CARE	SA	448	78	10	14623
	UN	389	137	12	14620
Caribou Creek	SA	93	23	47	6689
	UN	79	37	50	6686
Marshall	SA	363	33	43	19523
	UN	317	82	29	19585

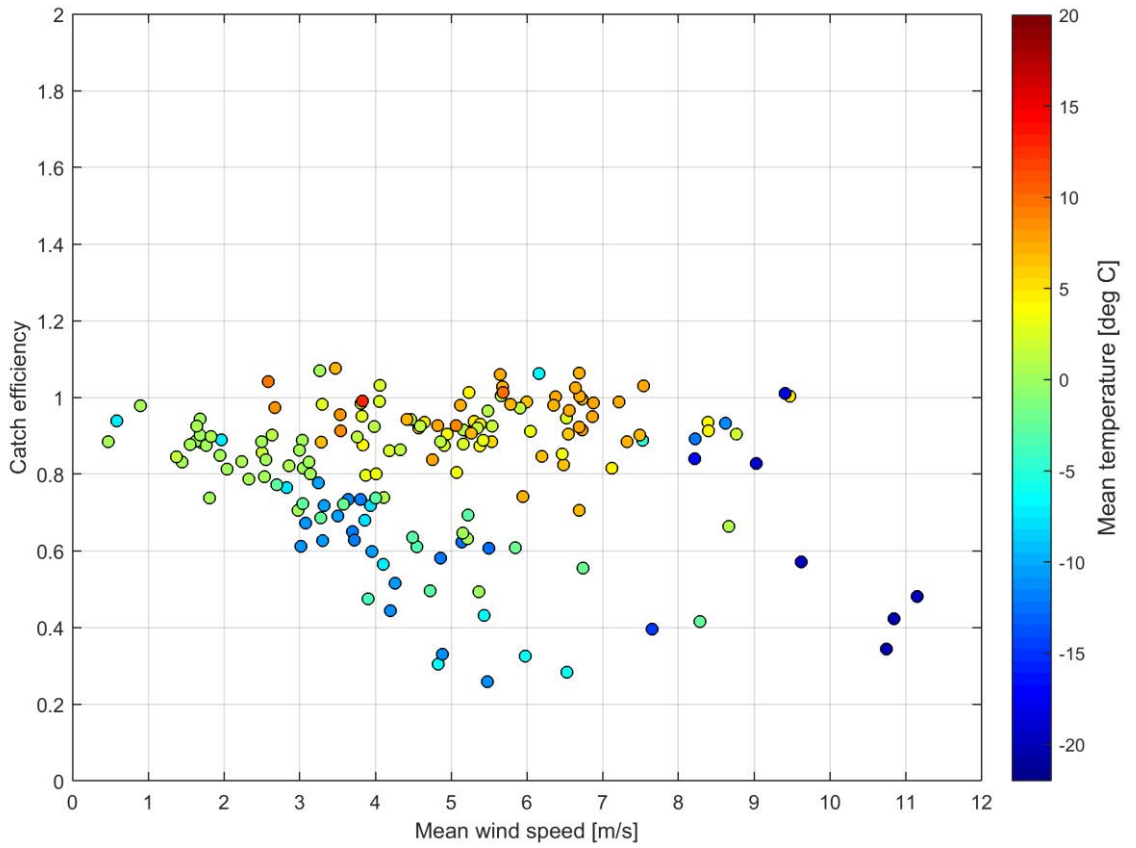
6.3. Ability to report accumulated precipitation

The SUT performance in terms of reporting accumulated precipitation is examined by comparing the amount reported by the sensor under test relative to the respective site reference during 30 minute assessment intervals. This is represented graphically using scatter and box and whisker plots of the catch efficiency as a function of mean wind speed at gauge height, as well as scatter plots of the amounts reported by the SUT versus the corresponding reference amounts (Figures 9 to 16). The SUT performance is also assessed in terms of the root mean square error, RMSE (Figure 17).

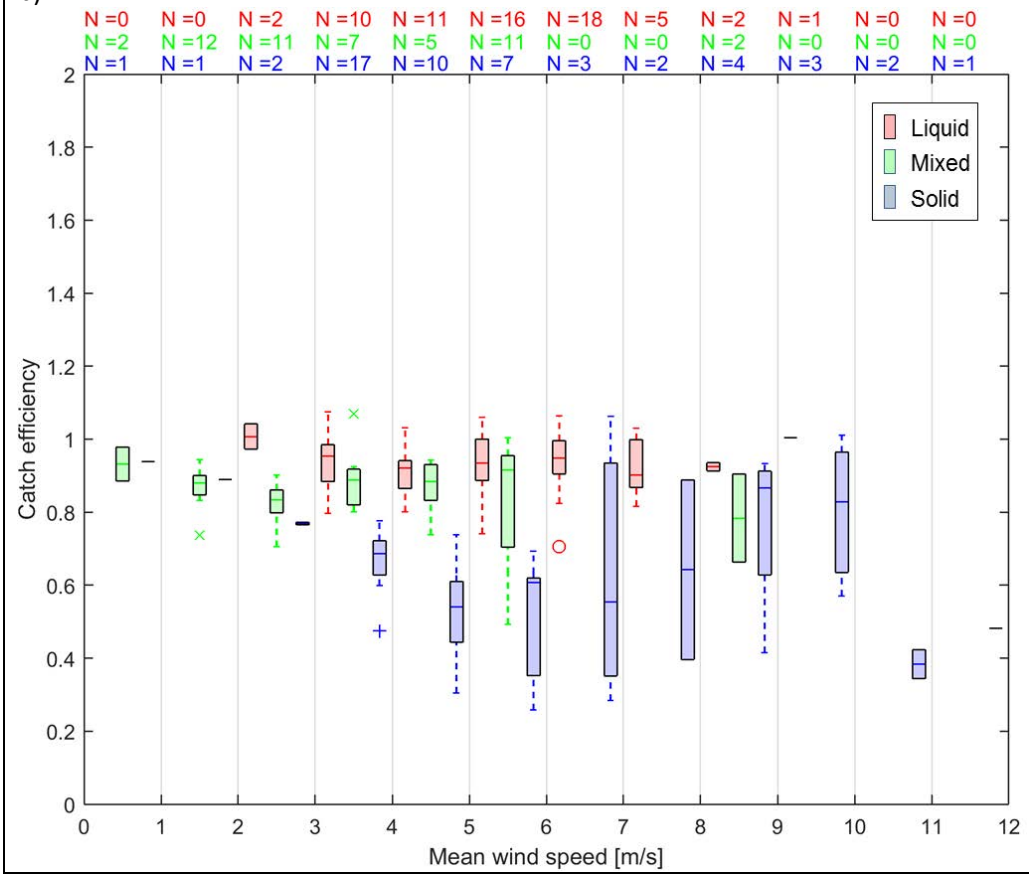
Only assessment intervals during which the SUT and reference both reported precipitation (YY cases) are considered in this portion of the assessment. In the catch efficiency-wind speed scatter plots, the mean event temperature is indicated by colour, with the colour scale selected to be consistent across all sites with weighing gauges under test. In the box and whisker plots and accumulation-accumulation scatter plots, the predominant precipitation type is indicated by colour, as determined from the reported temperature (Section 4.1.4).

Figure 9: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the SA-shielded Geonor T-200B3 gauge under test at Bratt's Lake.

a)



b)



c)

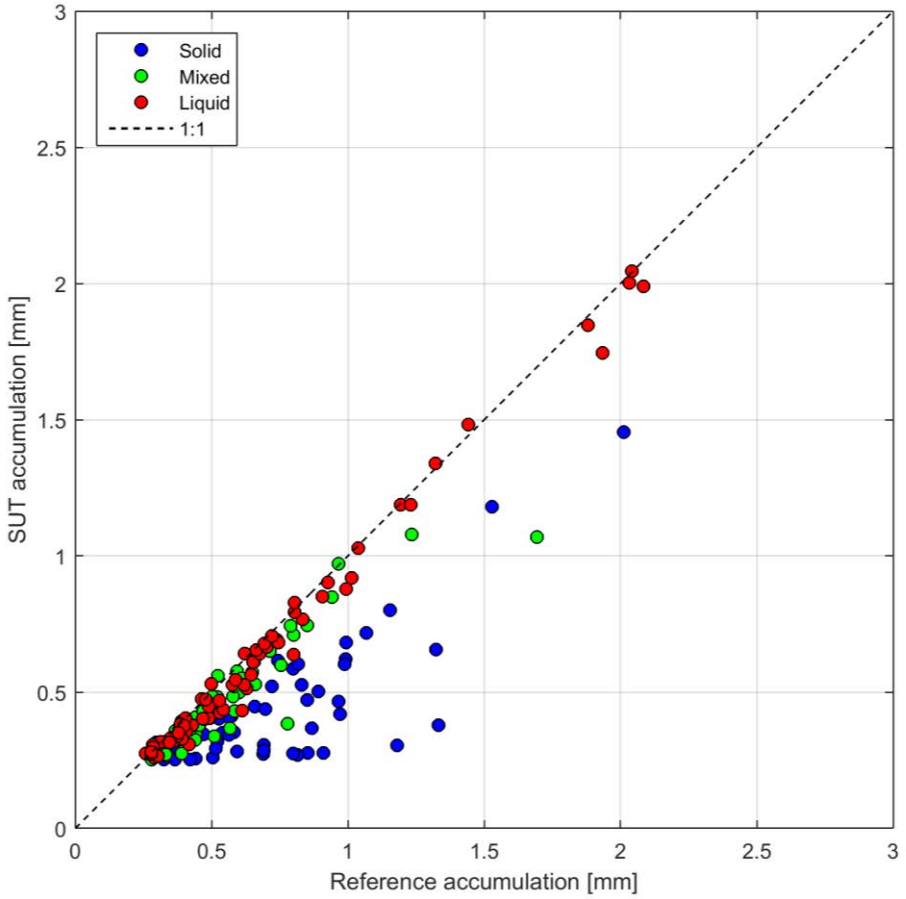
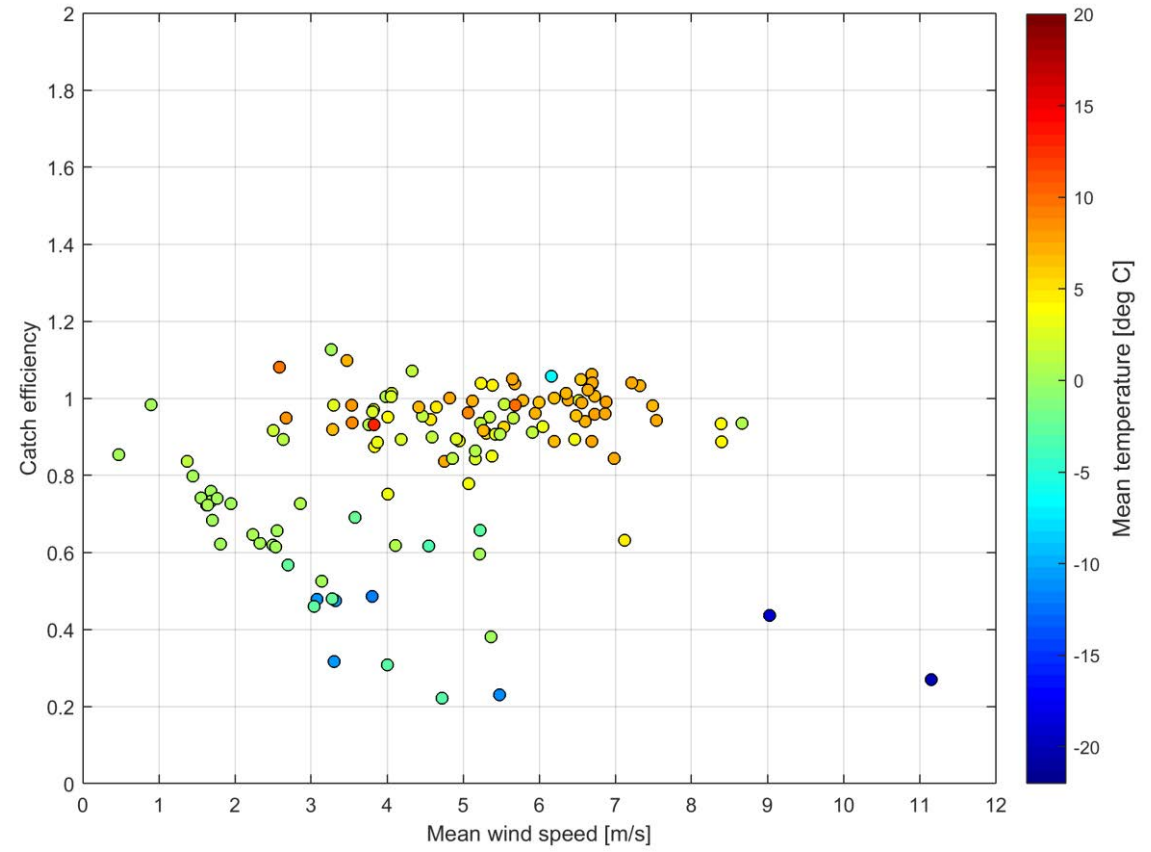
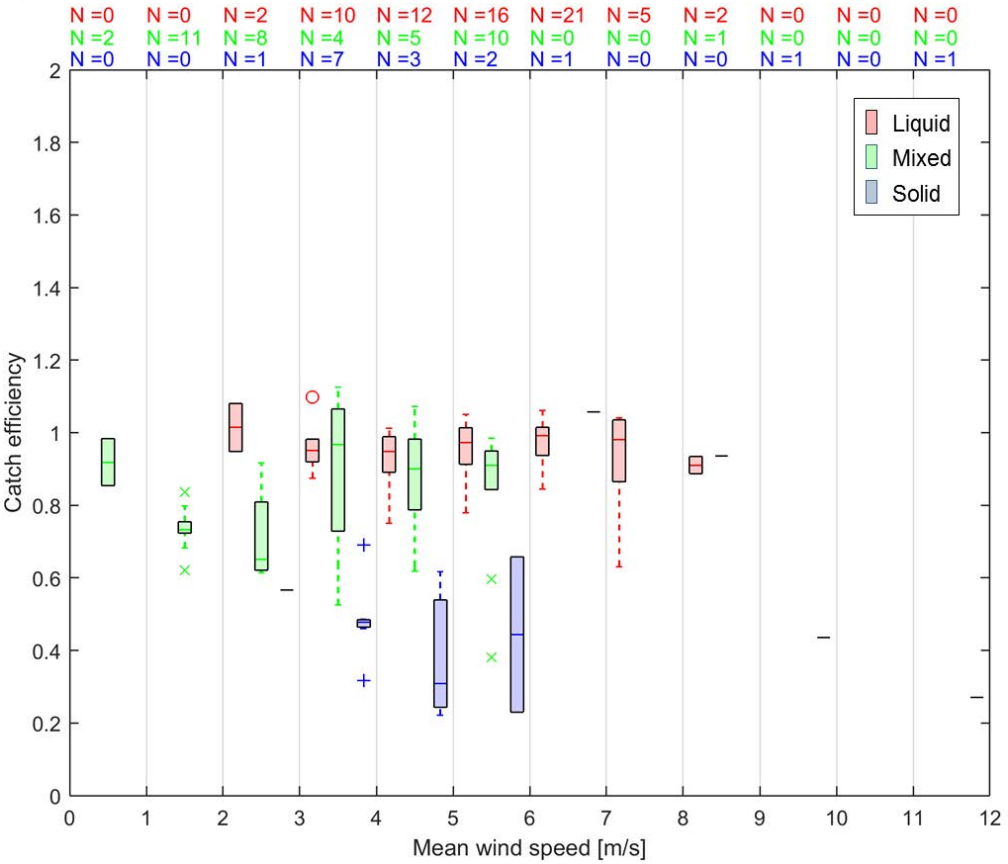

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
0.5
1
1.5
2
2.5
3

Figure 10: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded Geonor T-200B3 gauge under test at Bratt's Lake.

a)



b)



c)

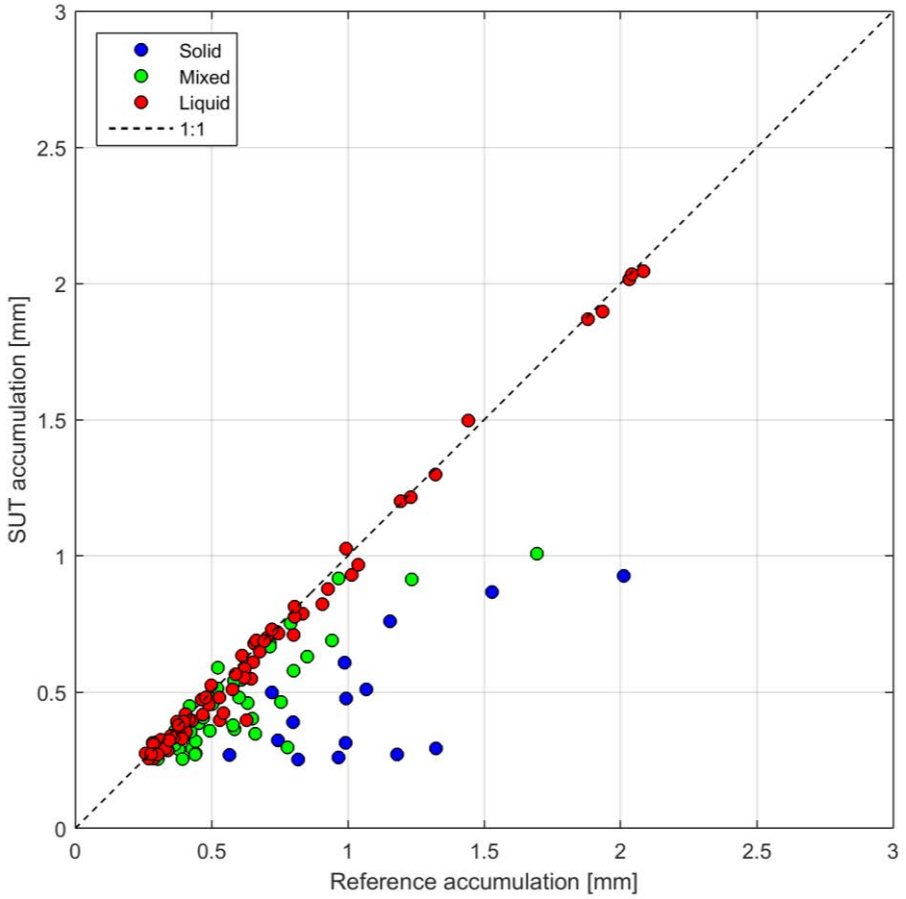
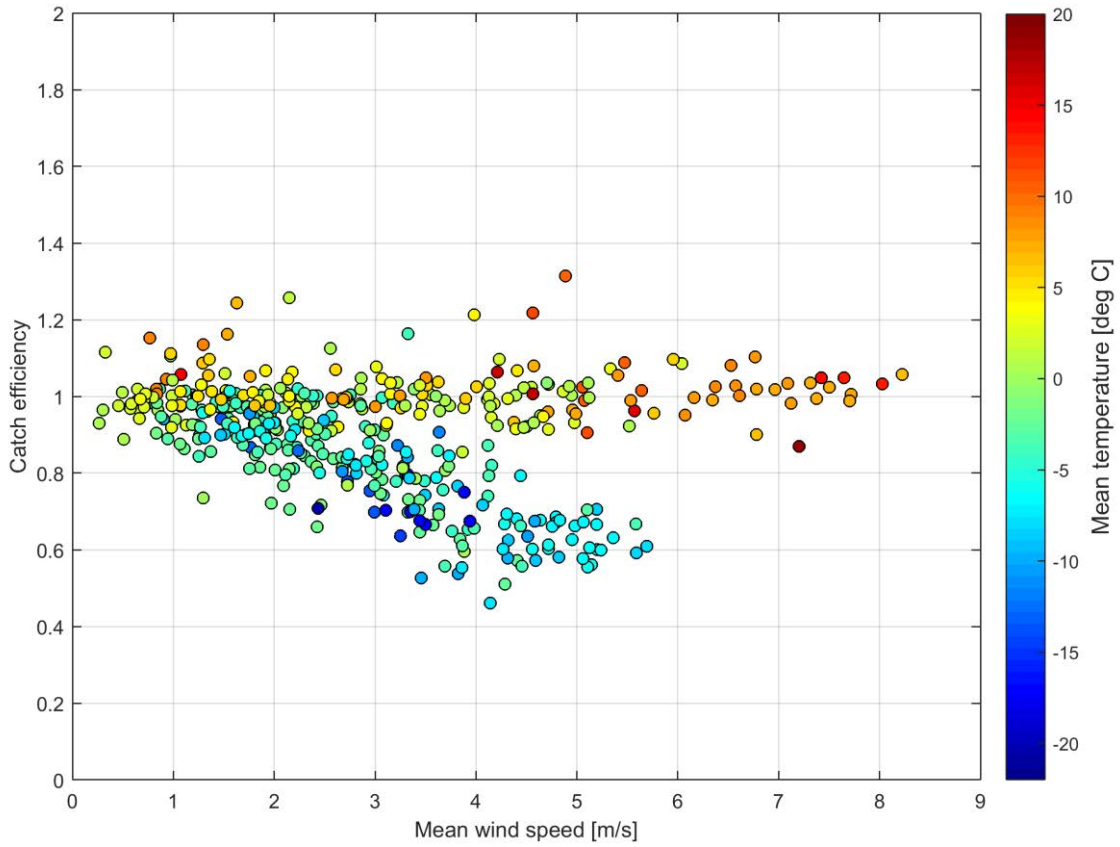
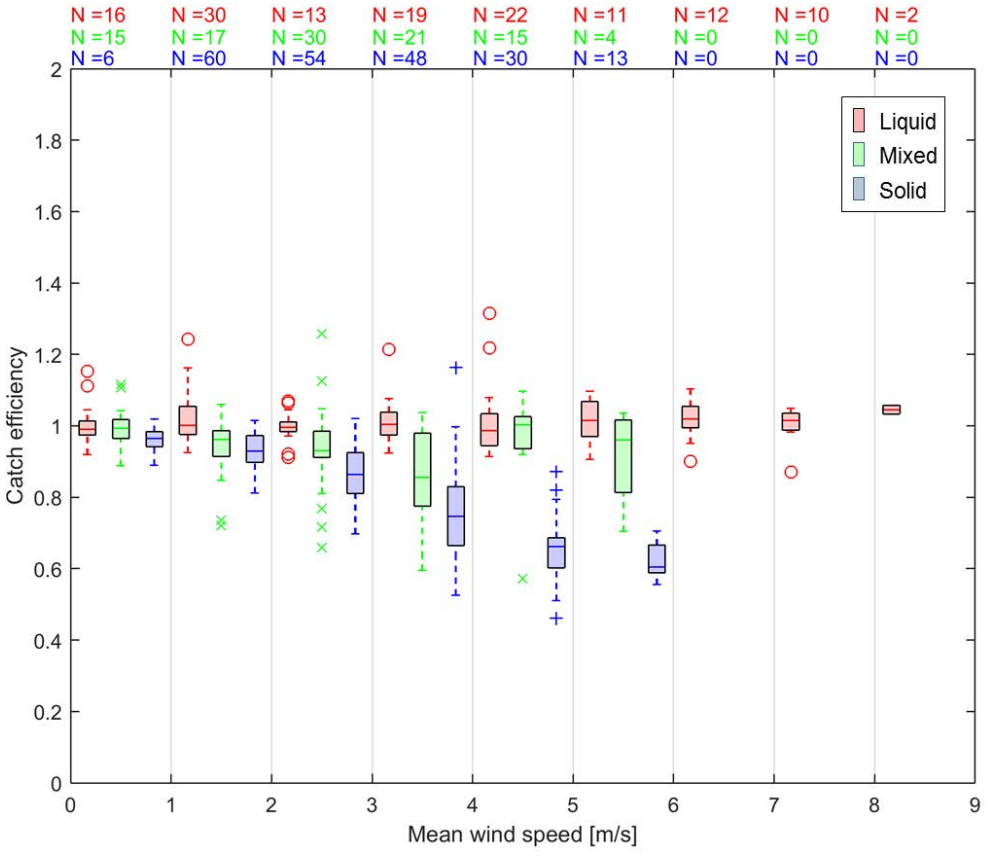

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
0.5
1
1.5
2
2.5
3

Figure 11: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the SA-shielded Geonor T-200B3 gauge under test at CARE.

a)



b)



c)

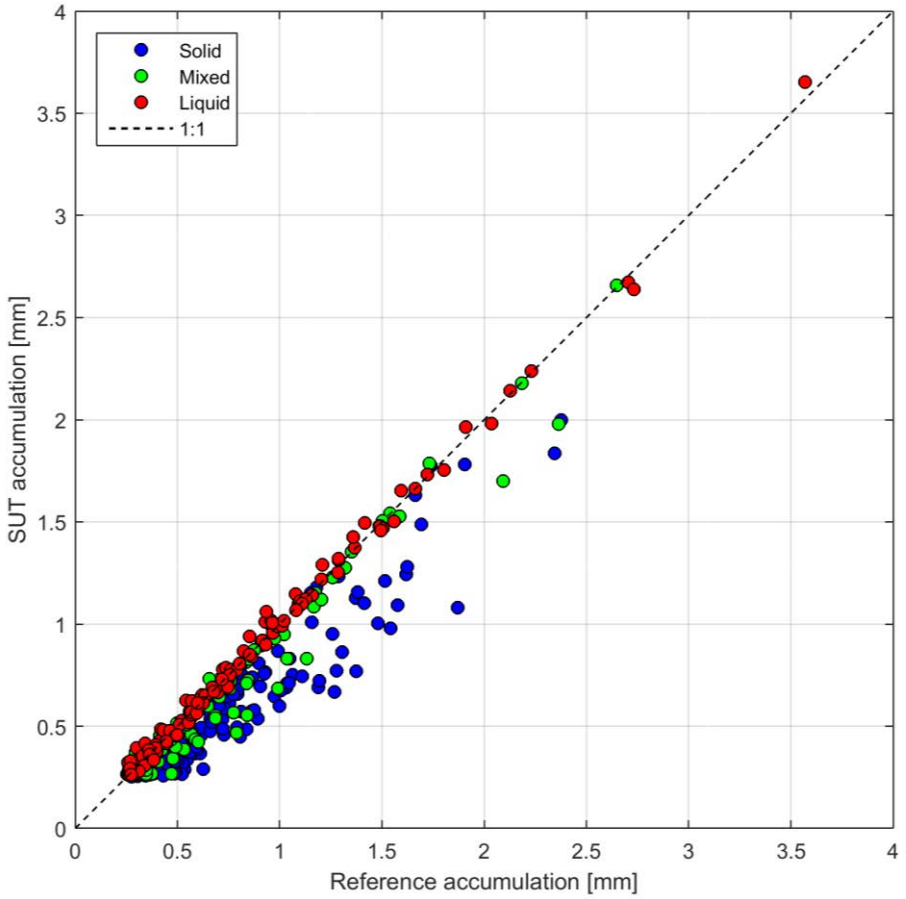
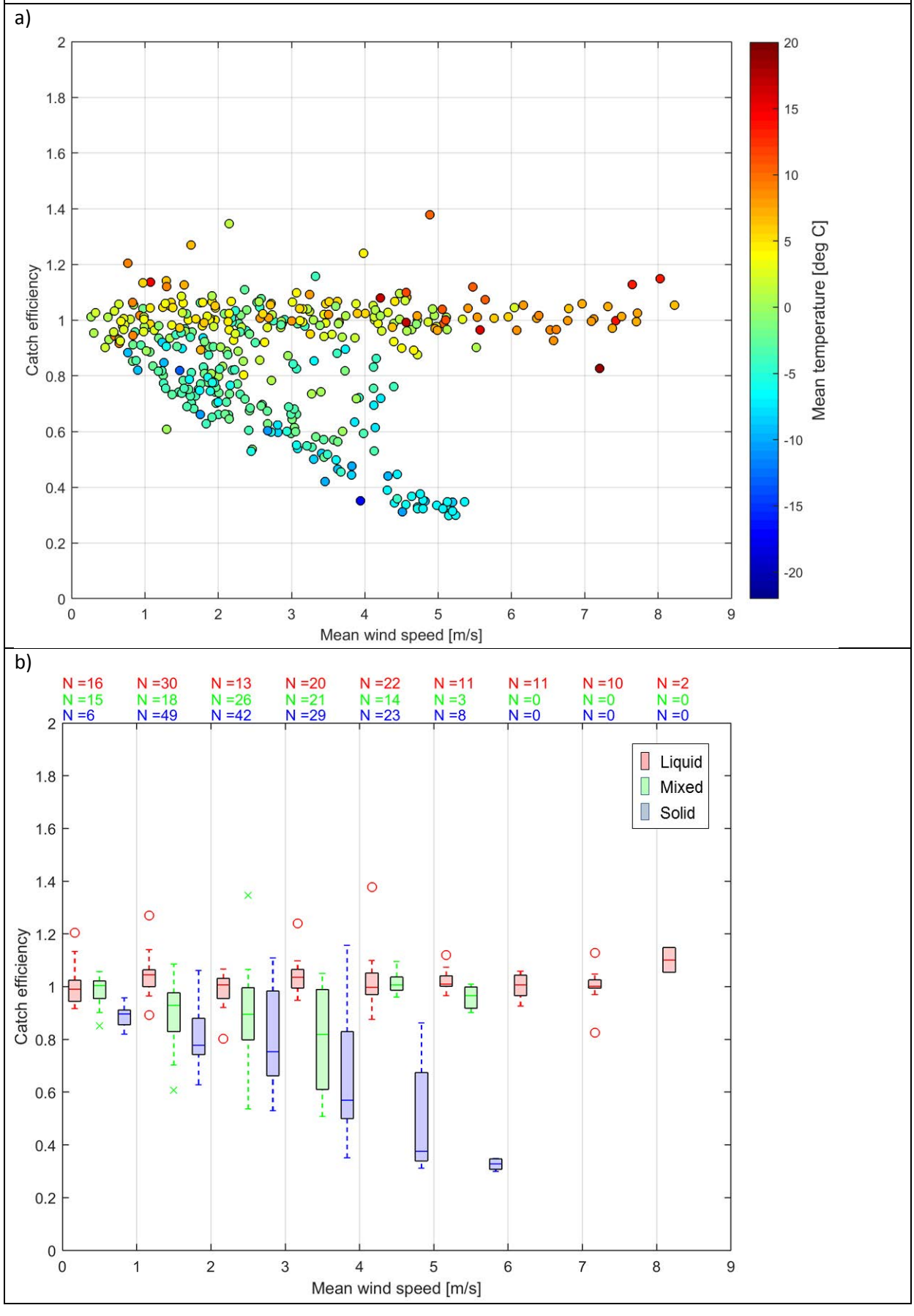

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
0.5
1
1.5
2
2.5
3
3.5
4

Figure 12: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded Geonor T-200B3 gauge under test at CARE.



c)

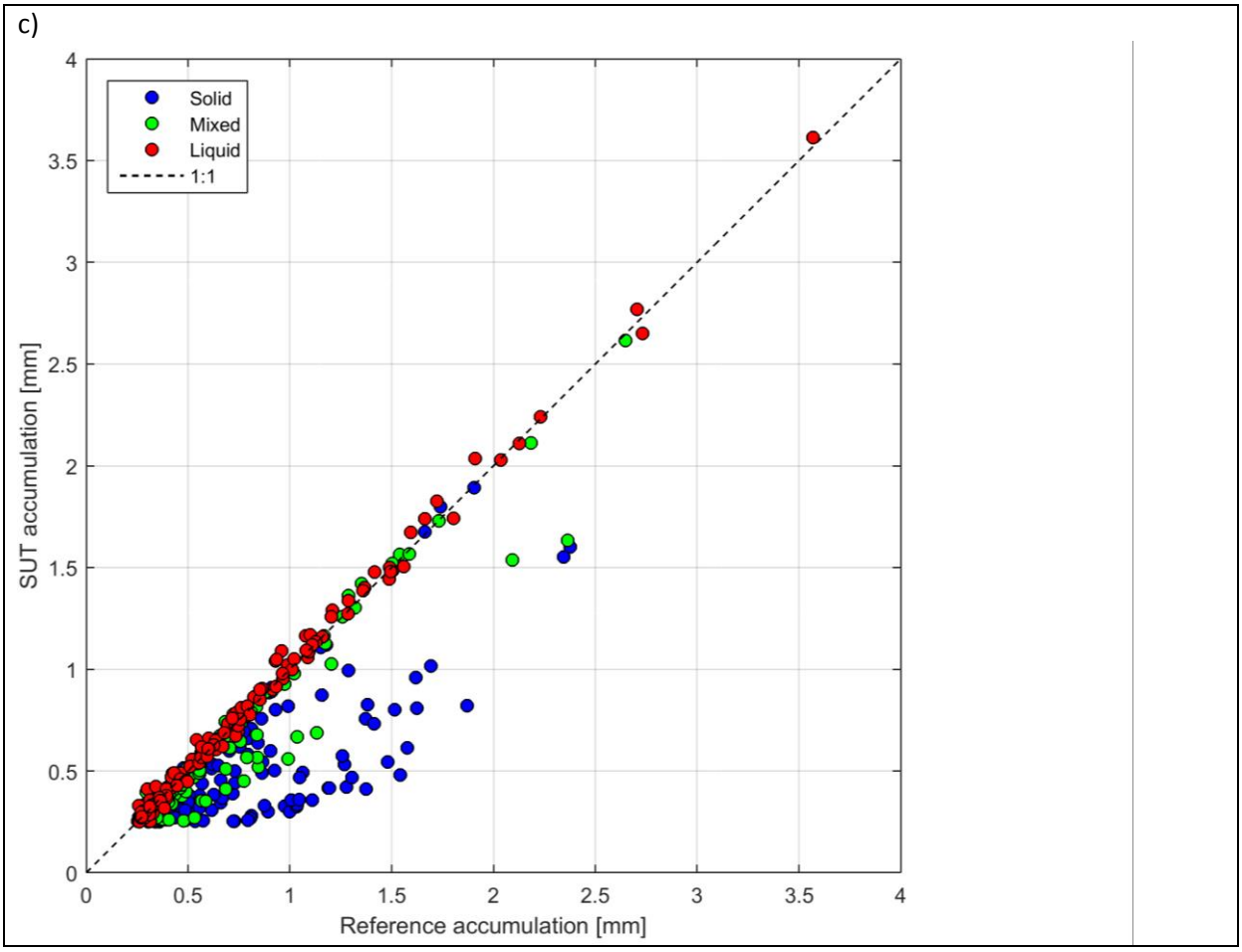
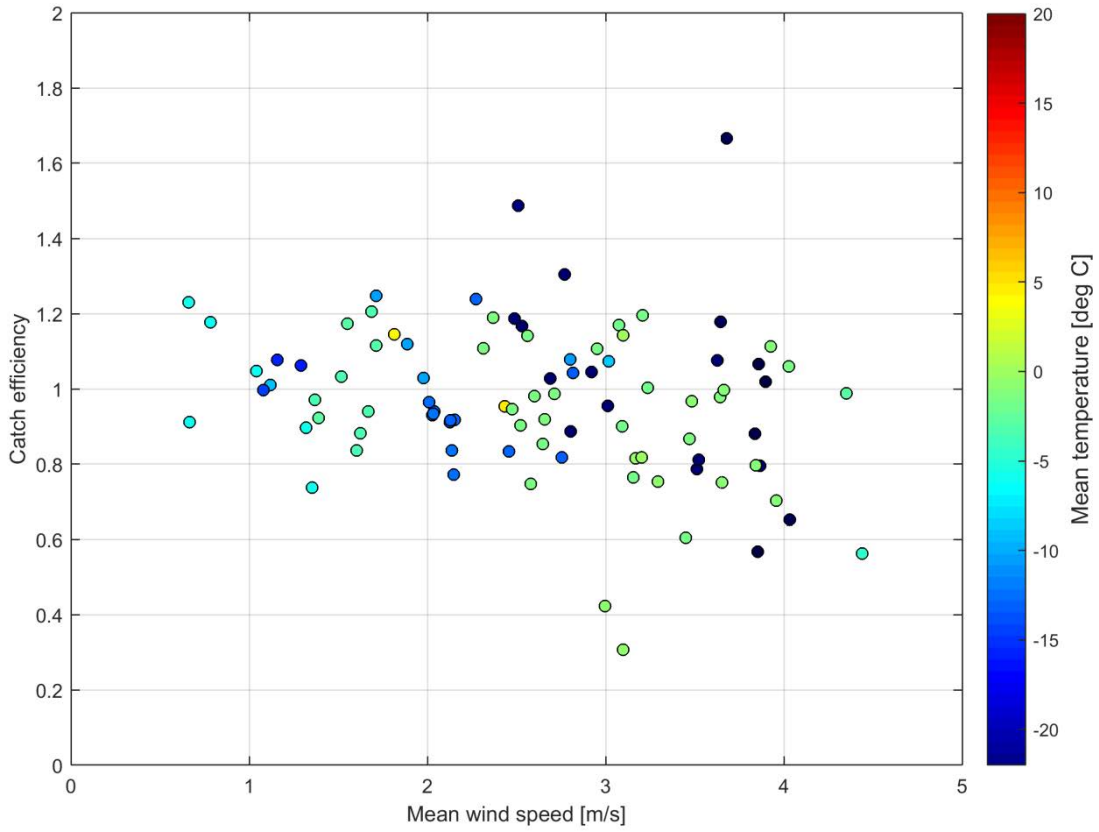
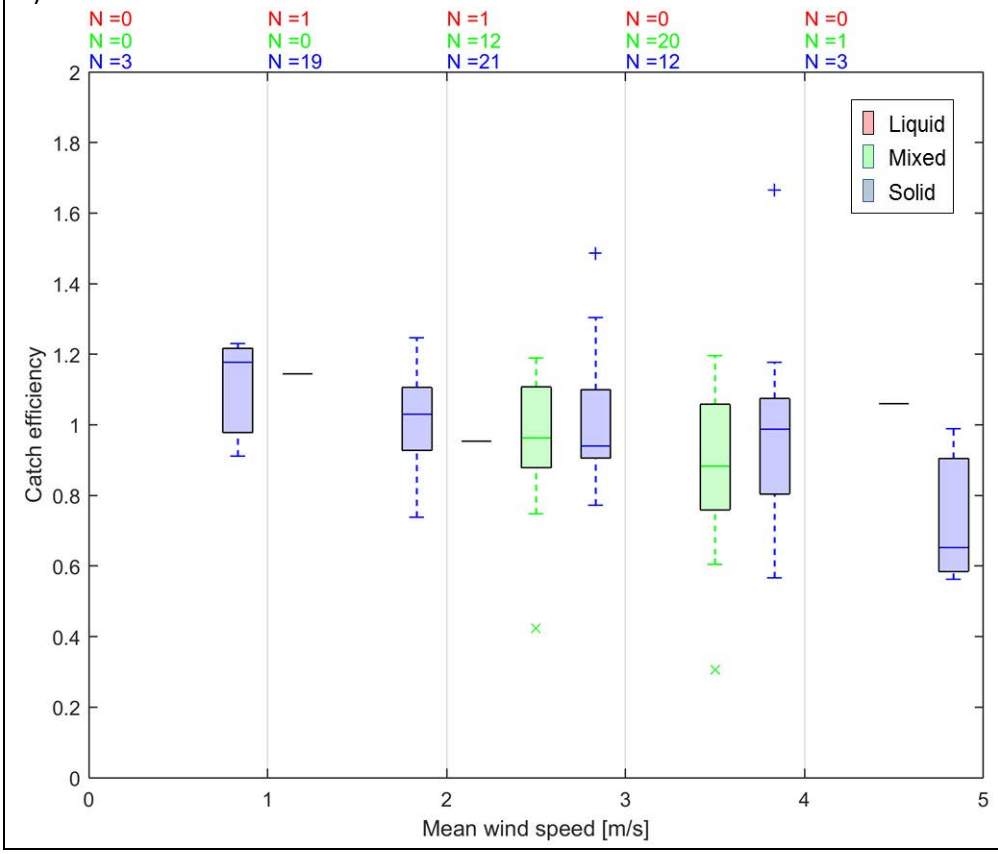

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
0.5
1
1.5
2
2.5
3
3.5
4

Figure 13: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the SA-shielded Geonor T-200B3 gauge under test at Caribou Creek.

a)



b)



c)

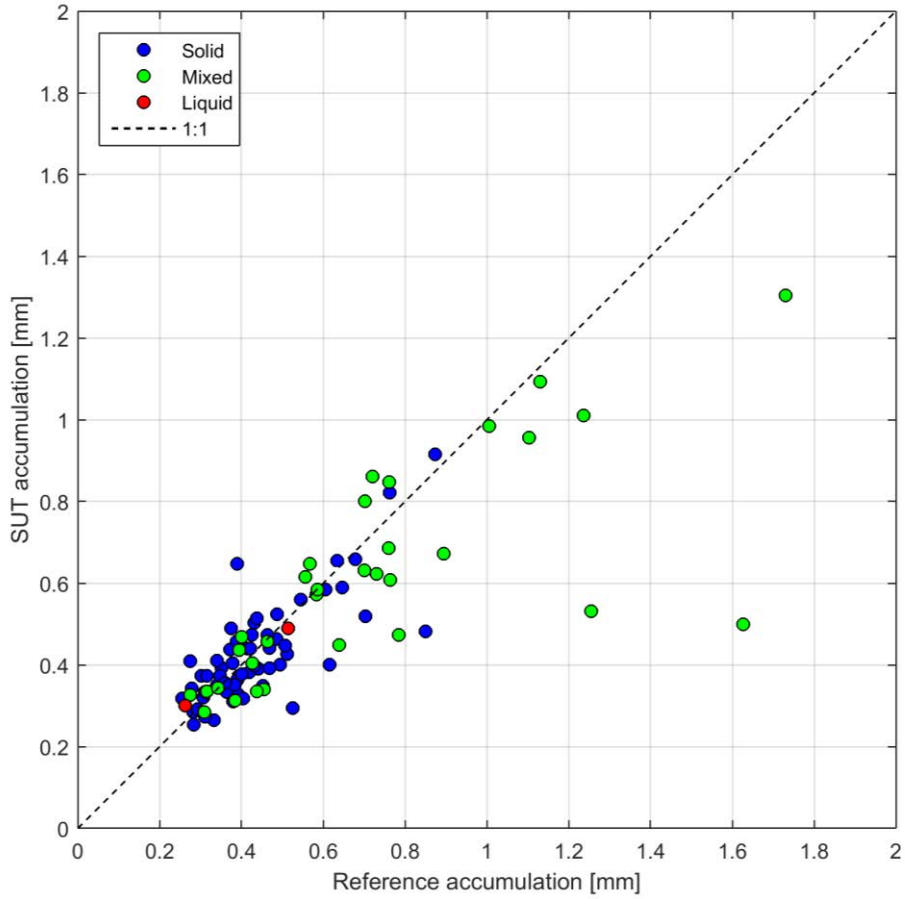
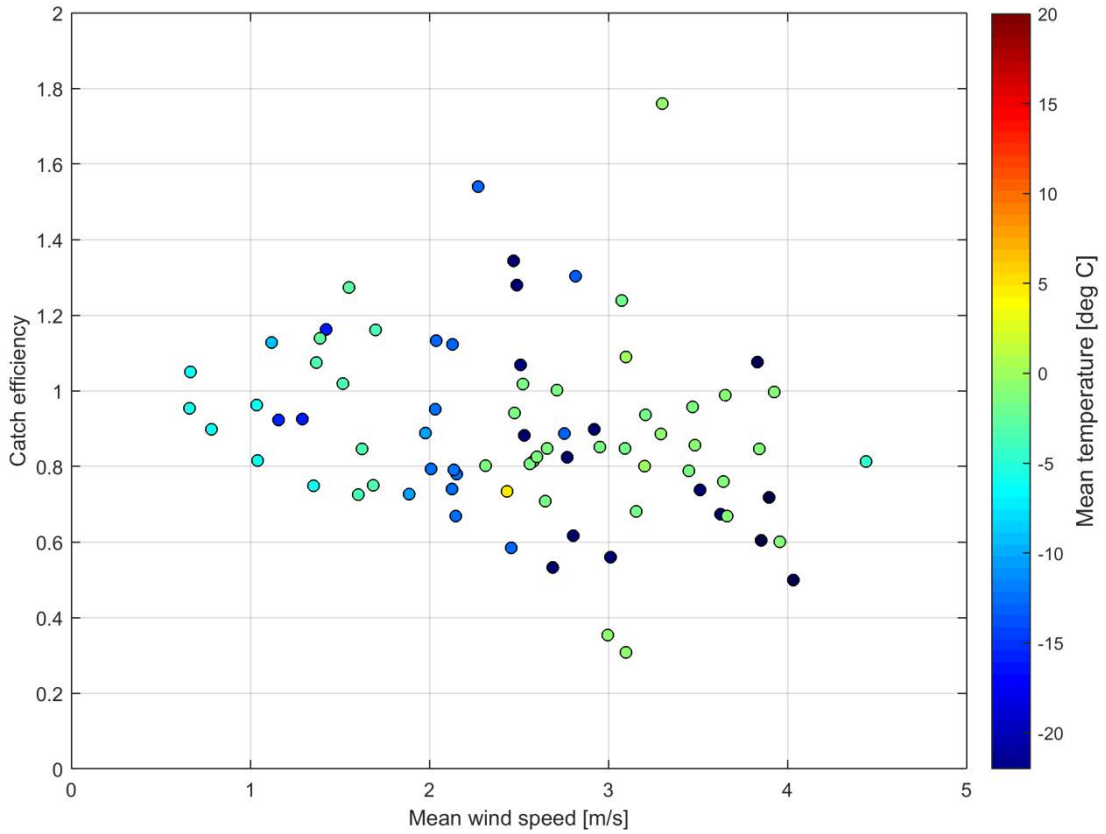
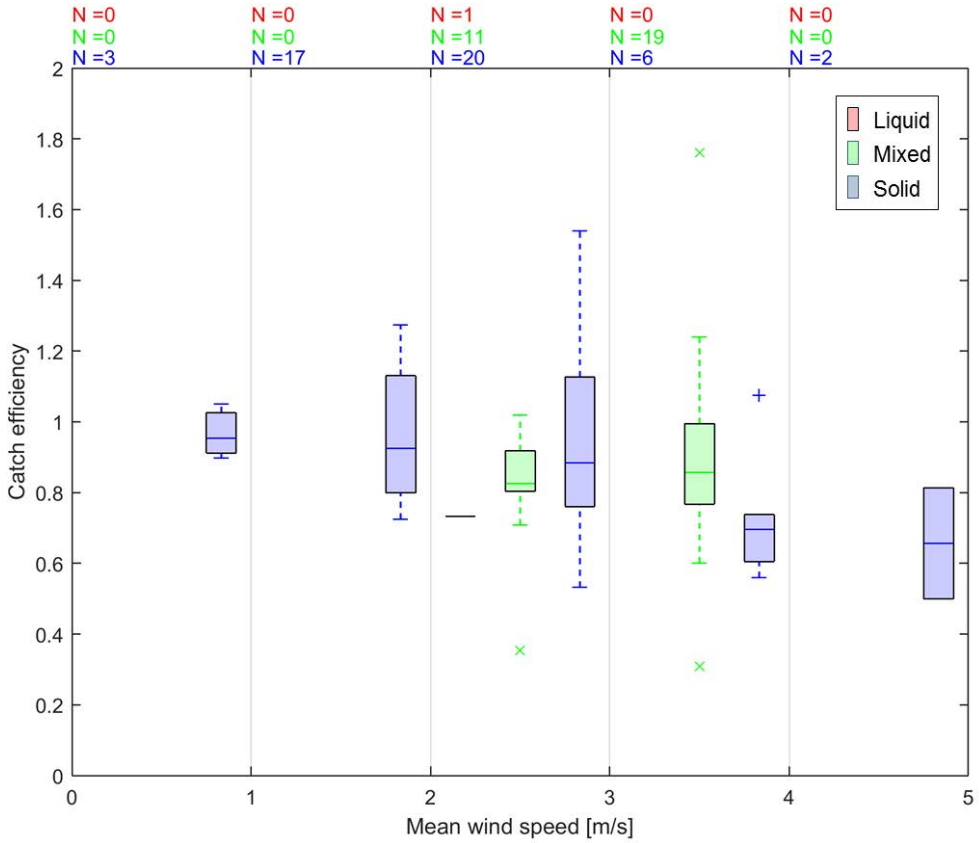

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

Figure 14: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded Geonor T-200B3 gauge under test at Caribou Creek.

a)



b)



c)

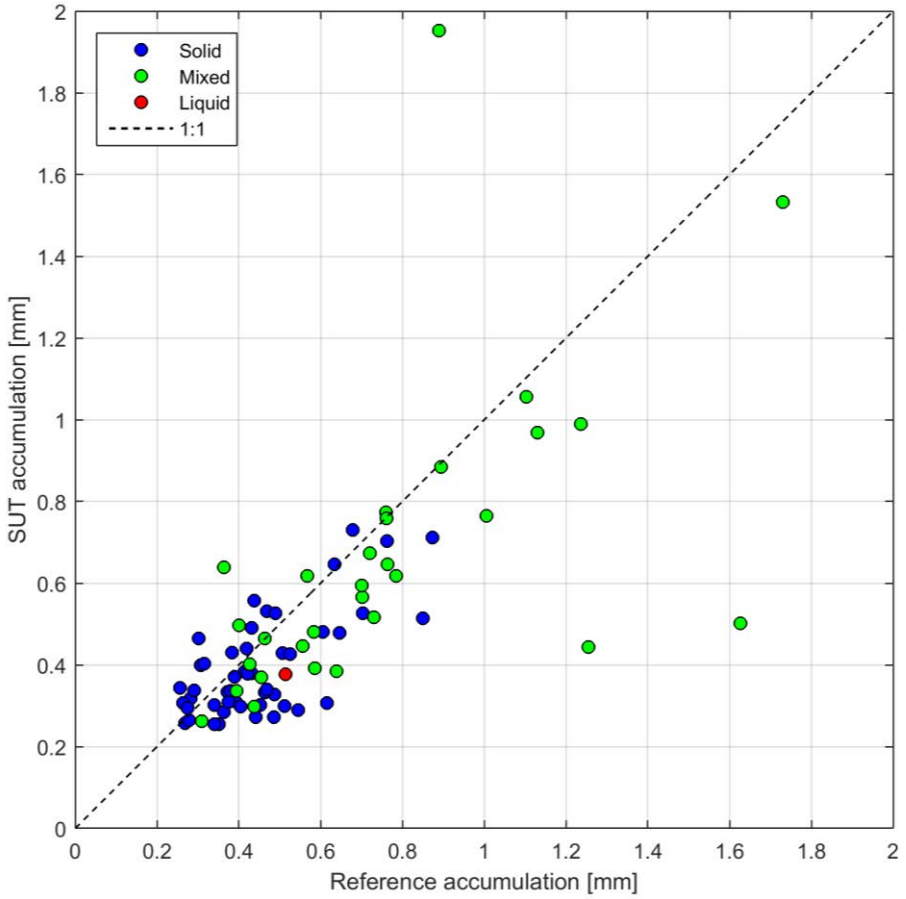
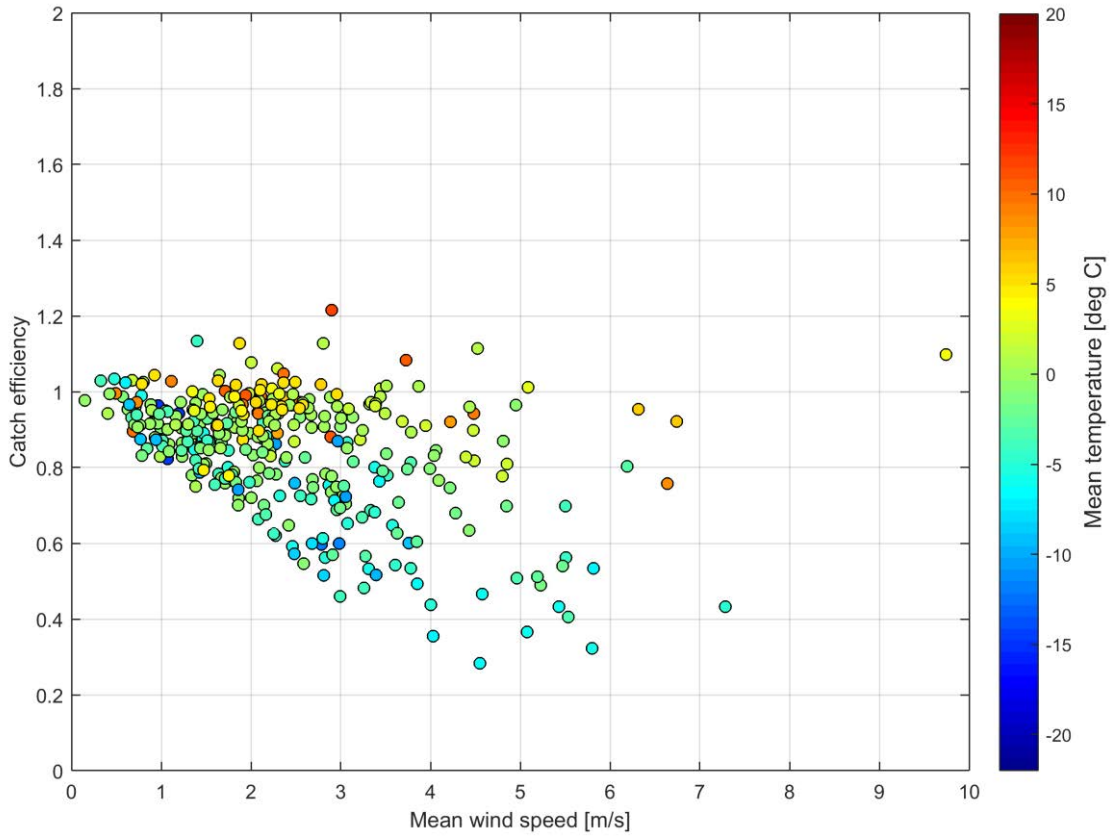
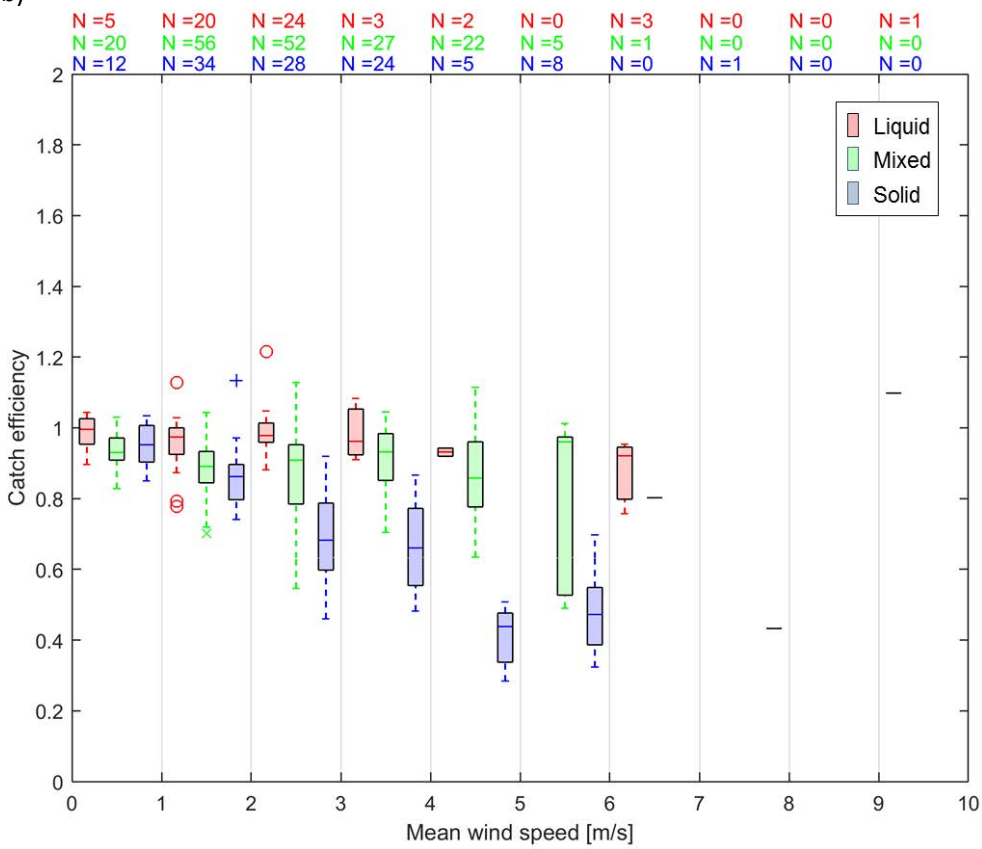

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

Figure 15: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the SA-shielded Geonor T-200B3 gauge under test at Marshall.

a)



b)



c)

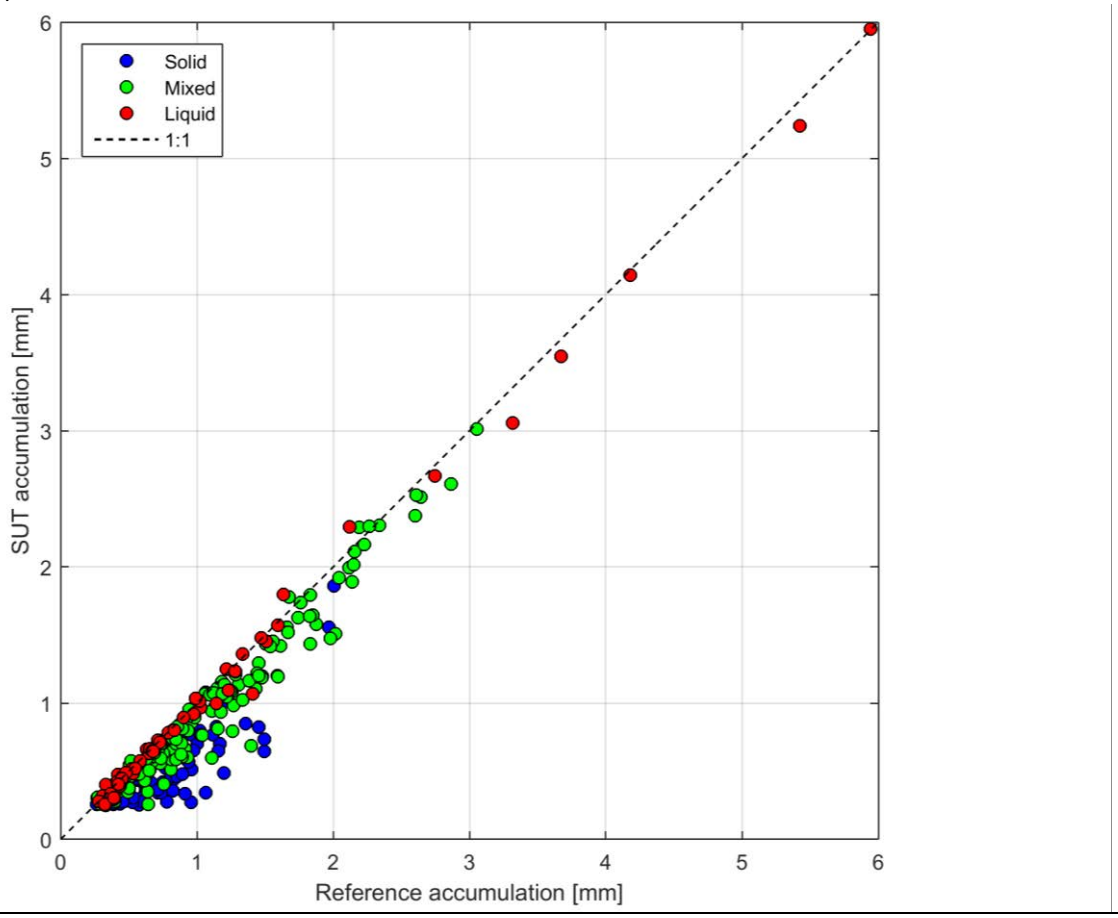
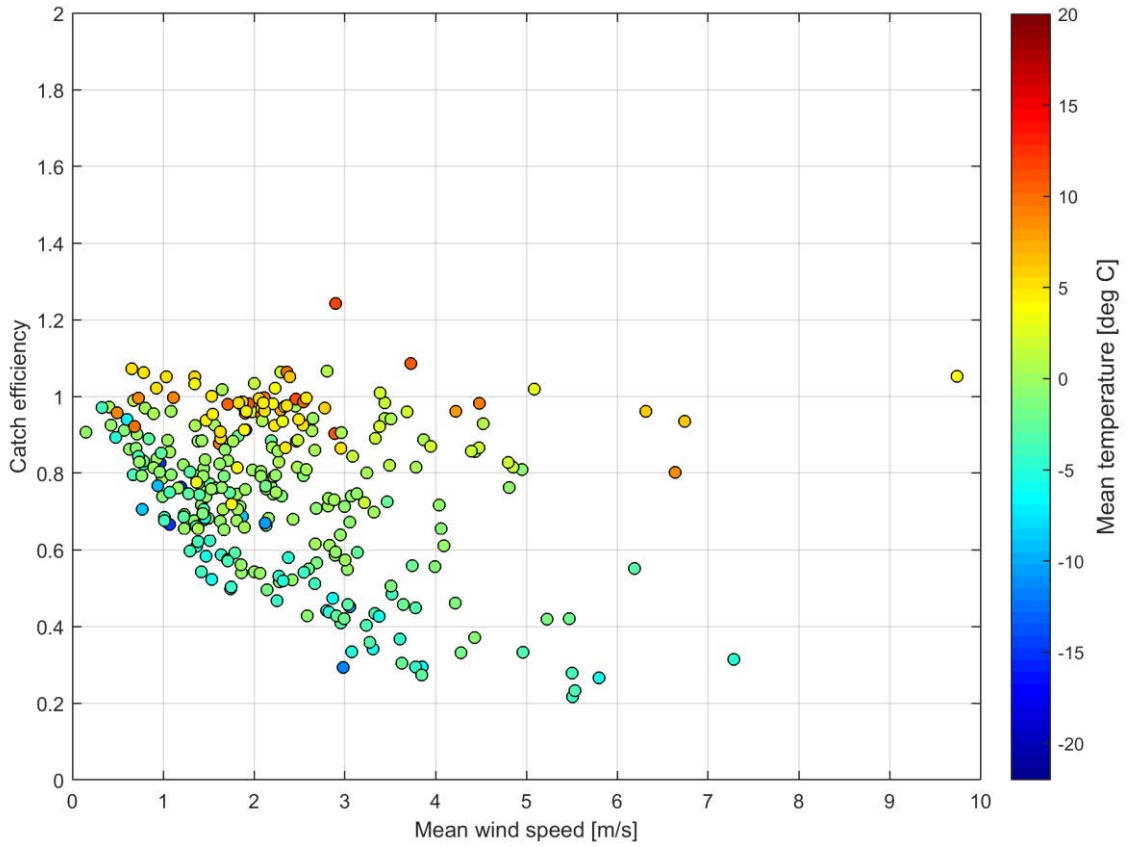
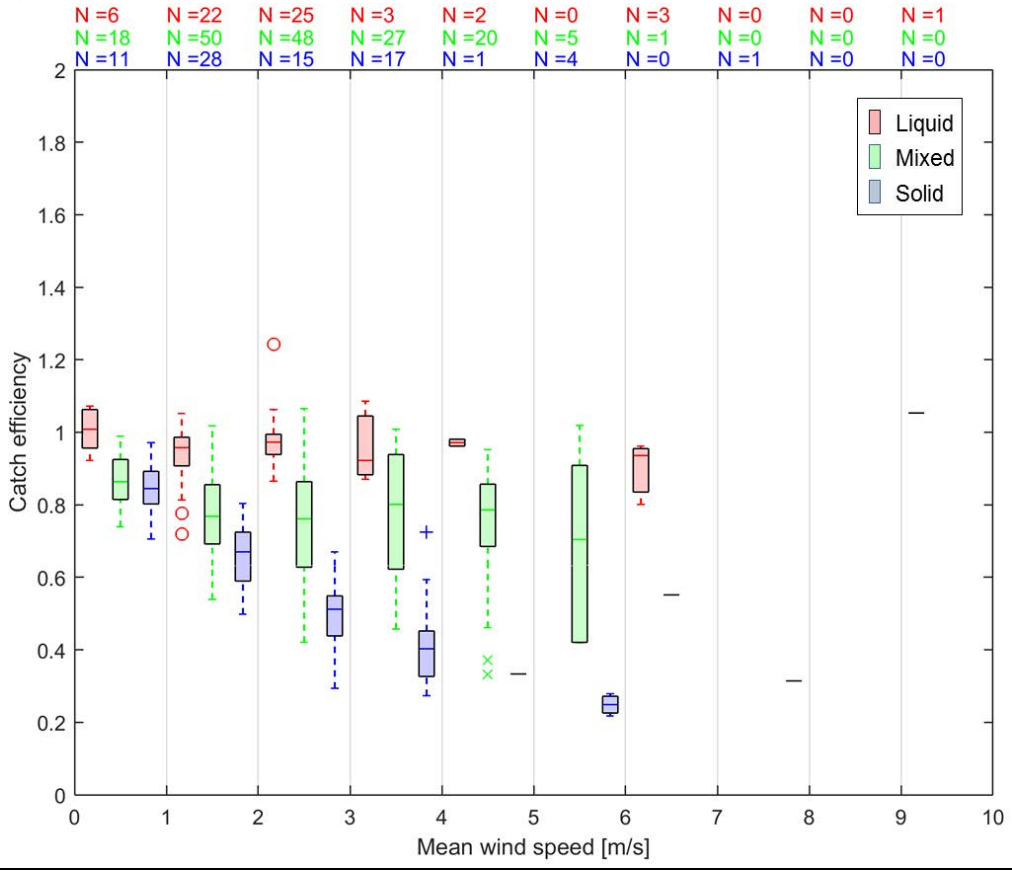

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
1
2
3
4
5
6

Figure 16: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded Geonor T-200B3 gauge under test at Marshall.

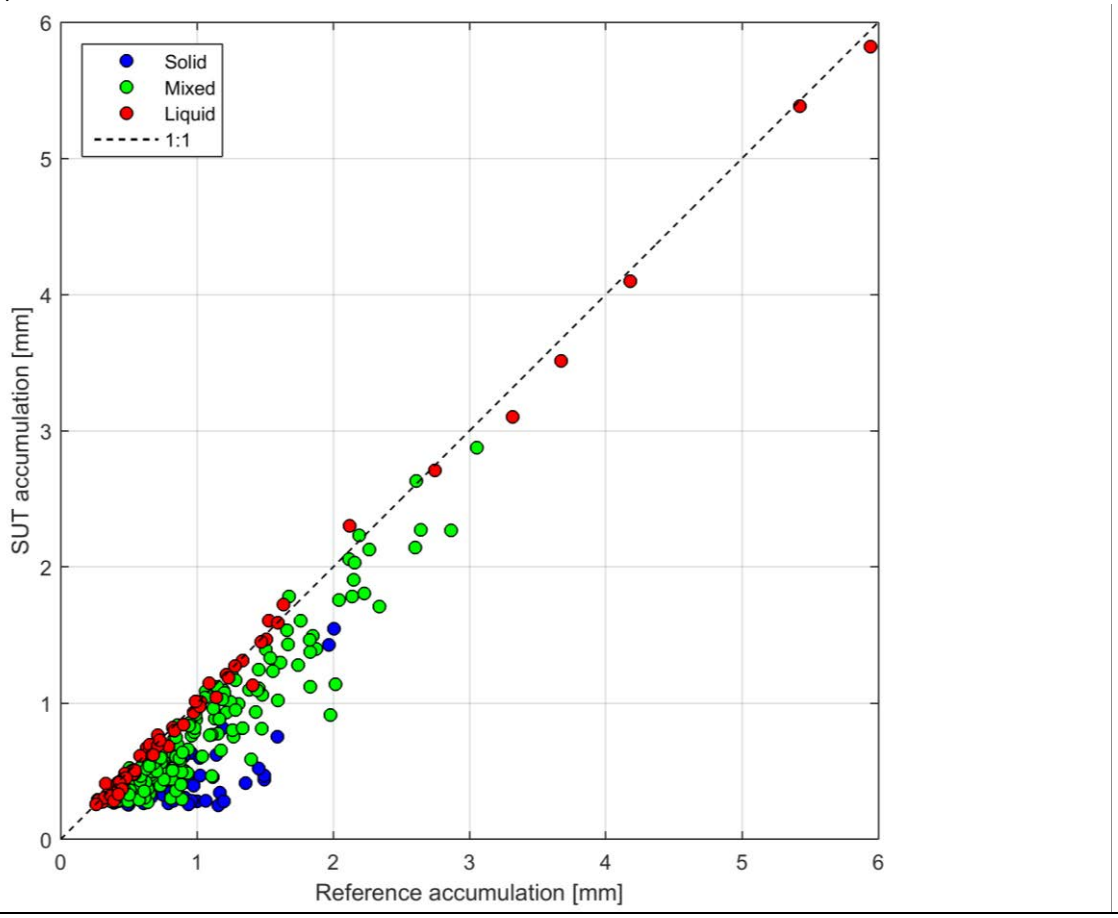
a)



b)



c)


Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
1
2
3
4
5
6

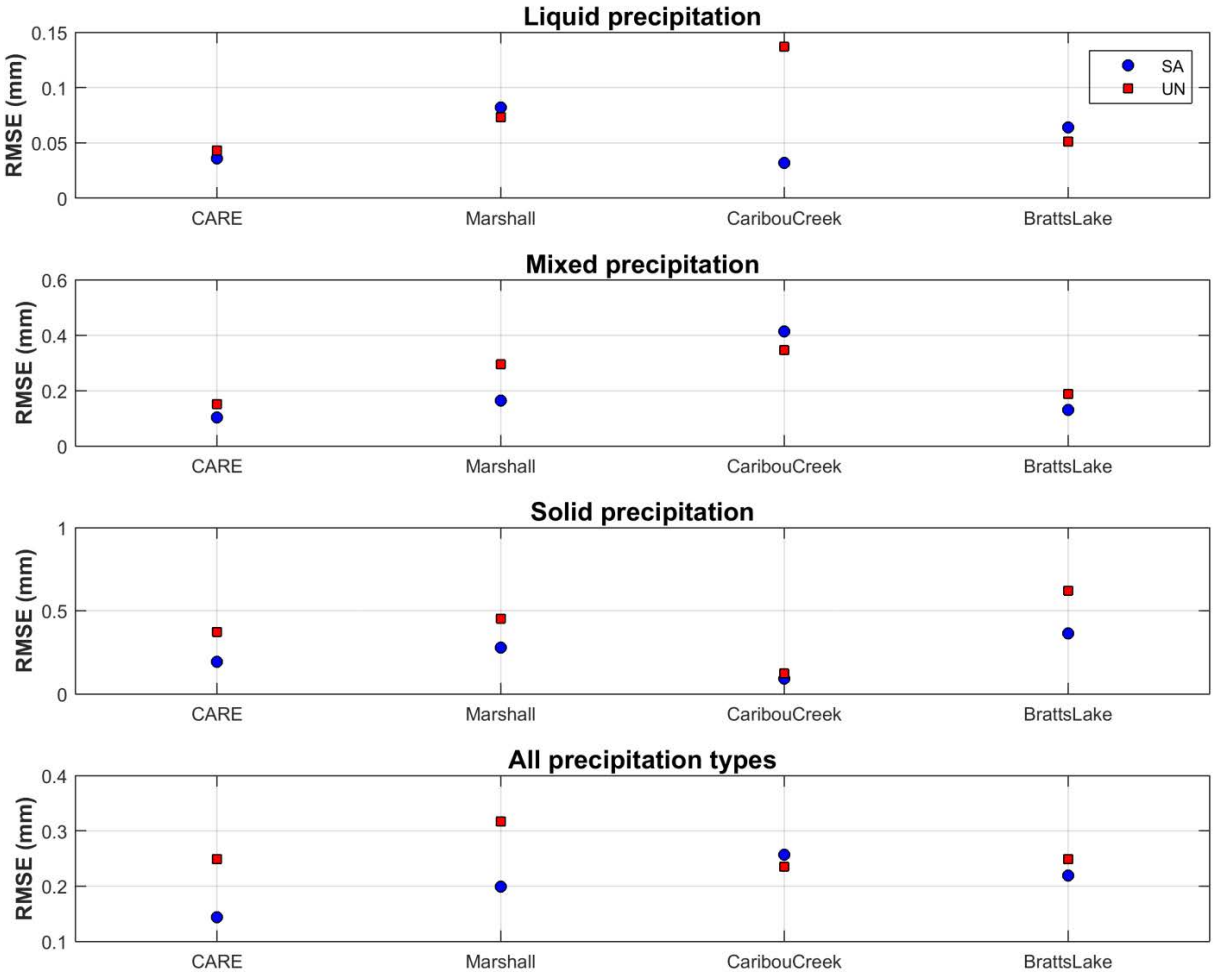


Figure 2: RMSE values calculated for test configurations by site and by precipitation type for YY cases over the duration of the test period.

Table 6: RMSE values calculated for all Geonor T-200B3 gauges assessed, represented by site and by precipitation type, as reflected in Figure 17.

Site	Configuration	RMSE (mm/30 min)			
		Liquid	Mixed	Solid	All precip types
Bratt's Lake	SA	0.064	0.130	0.364	0.219
	UN	0.051	0.188	0.620	0.249
CARE	SA	0.036	0.103	0.194	0.144
	UN	0.043	0.152	0.372	0.249
Caribou Creek	SA	0.032	0.414	0.092	0.257
	UN	0.137	0.347	0.124	0.235
Marshall	SA	0.082	0.164	0.279	0.199
	UN	0.073	0.295	0.453	0.317

The overall catch ratio calculated using all 30 minute YY cases, over the entire test period, is provided in Table 7. To demonstrate the influence of the SUT accumulation threshold on the results, the overall catch ratio is also provided for all 30 minute YY cases determined using a lower SUT threshold of 0.1 mm/30 minutes. Note that these values reflect only the YY cases, and do not include the amounts corresponding to the cases when the SUT and the reference do not agree on the occurrence of precipitation.

Table 7: Overall catch ratios for test configurations, determined from YY cases over the entire test period at each site, using two different SUT accumulation thresholds.

Site	Configuration	SUT accumulation threshold (mm/30 min)	Overall catch ratio
Bratt's Lake	SA	0.25	0.80
		0.1	0.73
	UN	0.25	0.82
		0.1	0.70
CARE	SA	0.25	0.90
		0.1	0.88
	UN	0.25	0.85
		0.1	0.81
Caribou Creek	SA	0.25	0.95
		0.1	0.96
	UN	0.25	0.87
		0.1	0.85
Marshall	SA	0.25	0.86
		0.1	0.85
	UN	0.25	0.77
		0.1	0.74

6.4. Ability to detect light precipitation events

The impact of the threshold selection for data processing relative to the detection of light precipitation was examined using four different combinations of reference and SUT accumulation thresholds (four 'cases' in Table 8) for the single-Alter-shielded Geonor T-200B3 gauge under test at the CARE site. Contingency results, probabilities of detection (POD), and false alarm rates (FAR) are presented for each case in Table 9. A quantitative comparison of the amounts reported in each case is beyond the scope of this assessment.

Table 8: Reference and SUT thresholds in each case for light precipitation detection assessment.

Case	Reference threshold (mm/30 min)	SUT threshold (mm/30 min)
1	0.25	0.25
2	0.1	0.1
3	0.25	No threshold
4	0.25	0

Table 9: Contingency results, Probability of Detection (POD), and False Alarm Rate (FAR) for each case in light precipitation detection assessment.

Case	Number of events				Skill score (%)	
	YY	YN	NY	NN	POD	FAR
1	448	78	10	14623	85.2%	2.18%
2	837	139	29	14154	85.8%	3.35%
3	526	0	14633	0	100%	96.5%
4	526	0	8501	6132	100%	94.2%

6.5. Assessment of events when the reference and the SUT do not agree on the occurrence of precipitation

Assessment intervals during which the site reference and SUT do not agree on the occurrence of precipitation – namely, the YN and NY cases (Section 4.1.1) – are characterized for the SA-shielded test gauge at CARE and the unshielded test gauge at Marshall using histograms in Figures 18 and 19, respectively. The histograms include accumulated precipitation reported by the reference and SUT (0.25 mm/30 min threshold for both), precipitation intensity as reported by the reference, and corresponding site conditions.

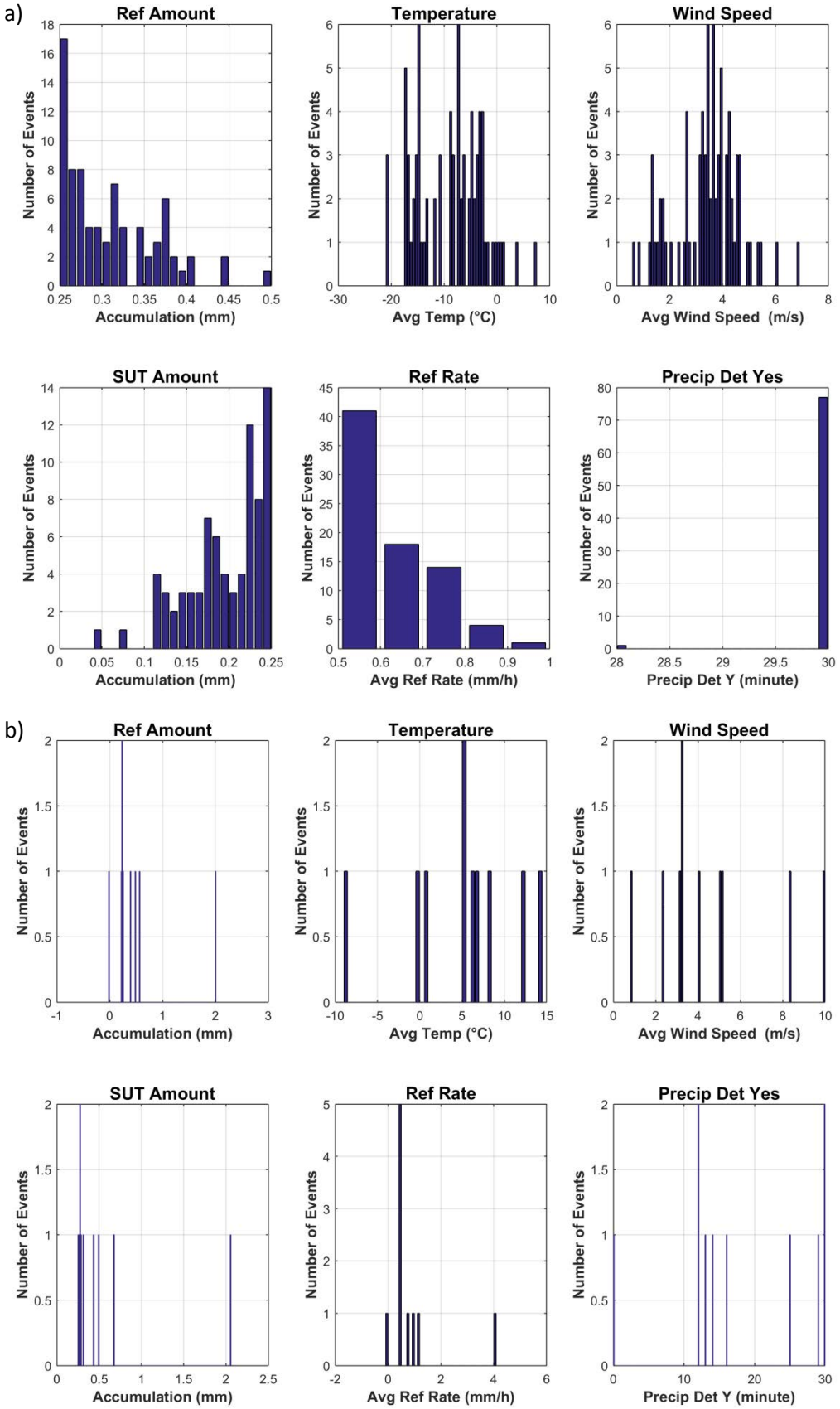


Figure 18: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for (a)YN events (78 total) and (b) NY events (10 total) for the SA-shielded Geonor T-200B3 gauge at CARE over the test period.

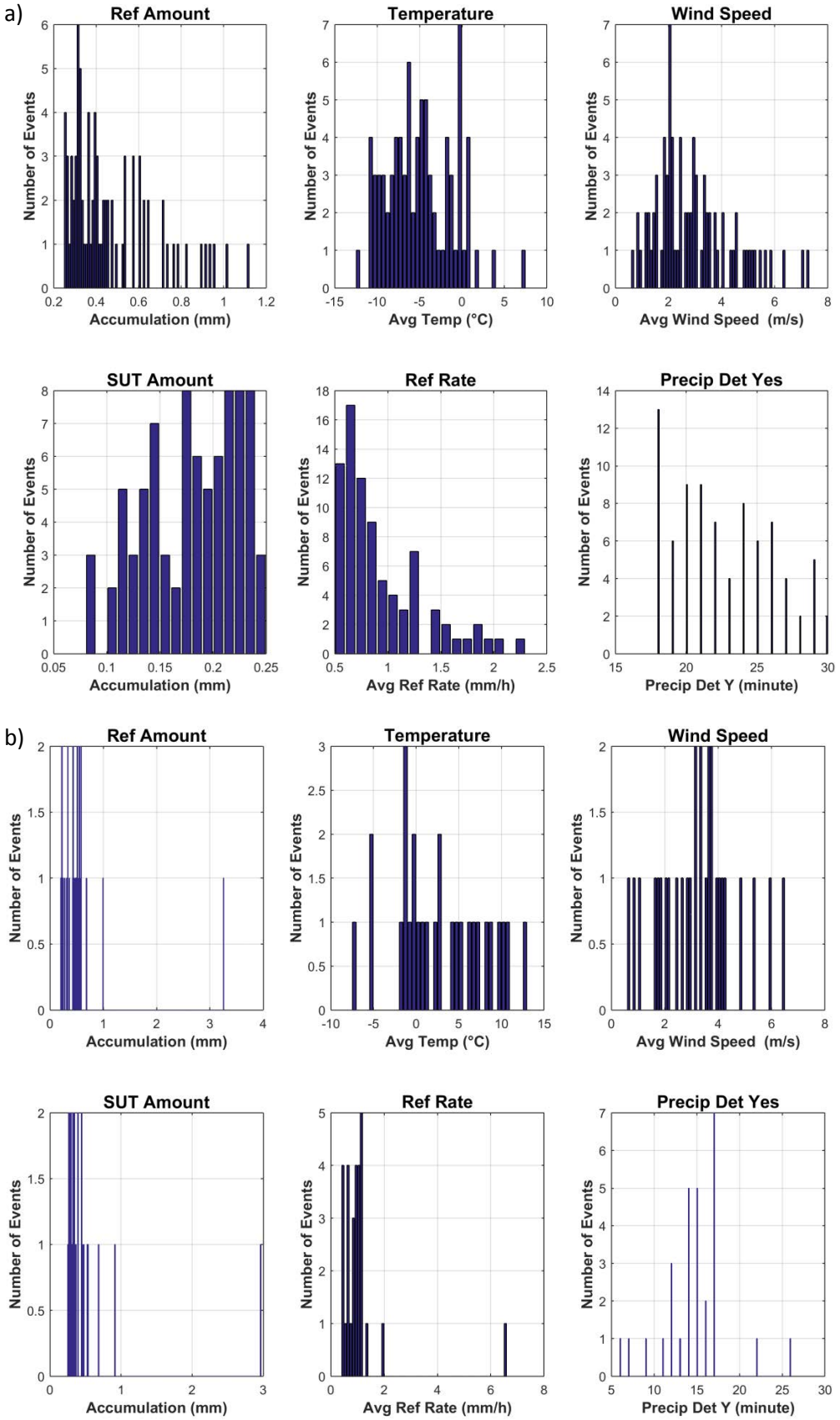


Figure 19: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for (a)YN events (82 total) and (b) NY events (29 total) for the unshielded Geonor T-200B3 gauge at Marshall over the test period.

7) Interpretation of Results

7.1. Operating conditions

The full ranges of conditions under which the test gauges at each site were operated are illustrated in Figure 3. The conditions relevant to gauge operation are as follows:

- Temperatures between approximately -40 °C and 26 °C;
- Precipitation intensity within approximately 12 mm/hr.

These conditions fall within the manufacturer's specified operating conditions of -40 °C to 60 °C for temperature.

The conditions at each site during precipitation events are shown in Figure 4. Of particular note are the mean wind speeds during precipitation events at Bratt's Lake, which extend to approximately 15 m/s; the mean wind speeds during precipitation at all other sites fall within 10 m/s, and are generally within about 7 to 8 m/s.

7.2. Performance over the range of operating conditions

7.2.1. Non-precipitating conditions

The results presented in Figures 5 to 7 and Table 4 show the variability in the reference and test gauge responses in the absence of precipitation from site to site. The PDFs of responses for the test gauges at Bratt's Lake, CARE, and Marshall are comparable to those for the corresponding reference configurations in Figure 5. The standard deviations of test gauge responses in Table 4 are of similar magnitude as those for the corresponding reference configurations, but vary by site and configuration from 0.011 mm for the single-Alter shielded gauge at CARE to 0.021 for the unshielded gauge at Marshall.

The PDFs of gauge responses in non-precipitating conditions for the test gauges at Caribou Creek in Figure 5 are markedly broader than those of the reference configuration at the same site (Figure 5a) and of those of the gauges in the same configuration at other sites (Figures 5b and 5c). This enhanced variability is also evident in the standard deviations of responses from these test gauges in Table 4, which are over three times as large as those for the reference configuration. This is attributed to a ground noise issue experienced at the site, and underscores the critical influence of test gauge configuration on gauge performance.

The variability of reference and test gauge responses in the absence of precipitation appears to be greatest during 30 minute periods with small or no temperature difference (Figure 6) and for mean wind speeds within about 3 m/s (Figure 7). With the exception of the test gauges at Caribou Creek, the magnitude of reports is typically within the accumulation detection for the detection of precipitation (0.25 mm for 30 minute periods); however, the maximum outputs in Table 4 indicate that reports for test gauges at the other sites can still exceed the detection threshold, resulting in the potential for false reports.

The magnitude of gauge responses in the absence of precipitation can be used to identify a detection threshold that minimizes the detection of false precipitation while enhancing the detection of light precipitation. This threshold is considered to be three times the standard deviation of the average gauge response during 30 minute non-precipitating periods. Based on the present results for test gauges at Bratt's Lake, CARE, and Marshall (Table 4), this minimum detection threshold is determined to be 0.04 mm for single-Alter shielded gauges and 0.06 mm for unshielded gauges. The Caribou Creek results are not included in the determination of threshold values due to the ground noise issue noted above.

7.2.2. Precipitating conditions

7.2.2.1. Ability to detect and report precipitation

The skill scores for Geonor T-200B3 test gauges in Figure 8 indicate that the single-Alter shielded configurations have higher Probability of Detection, higher Bias, and higher Heidke Skill Scores relative to the unshielded configurations at a given site. The false alarm rates are lower for the shielded configurations relative to the shielded configurations at all sites except Marshall, for which the values are comparable. The results for the test gauges at CARE and Marshall are similar for shielded (unshielded) test configurations: POD values above 85% (75%), FAR below 10% (10%), B between about 90-100% (75-90%), and HSS above 90% (80%).

The results for the test gauges at Caribou Creek show much higher False Alarm Rates, between about 30% and 40%, on account of noise related to the configuration. This results in more events being detected than the reference configuration (Bias > 100%) and less overall detection skill in terms of the Heidke Skill Score (HSS < 75%). The Probability of Detection and Bias values are lowest for the test gauges at Bratt's Lake, which can likely be attributed to the influence of higher mean wind speeds (Figures 3 and 4). This results in fewer events being detected relative to the reference, particularly for the unshielded configuration, which is more susceptible to wind-induced undercatch. The resulting Heidke Skill scores are 70% for the shielded configuration and 56% for the unshielded configuration.

7.2.2.2. Ability to report accumulated precipitation

The results presented in Figures 9 to 16, which are based on 30 minute events during which the reference and test gauge both detect precipitation (YY cases), illustrate the influence of wind speed and precipitation type on gauge catch efficiency. The discussion below will focus on snow events; the number of rain events during winter is limited, and the results for mixed events are variable due to the variability in the size and density of precipitation within the mixed regime, as well as the potential for transitions between phases. For solid precipitation events, the catch efficiency decreases with increasing wind speed; this decrease is more rapid for unshielded gauges relative to shielded gauges. For example, the median catch efficiencies for the shielded gauges at Bratt's Lake, CARE, and Marshall remain above approximately 0.4 for mean wind speeds between 5 and 6 m/s (Figures 9b, 11b, and 15b), while the median catch efficiencies are approximately 0.2 for the unshielded gauges at the same sites over the same mean wind speed range (Figures 10b, 12b, and 16b).The increase in median catch efficiency observed in the data for the shielded test gauge at Bratt's Lake at mean wind speeds greater than 6 m/s (Figure 9b) is attributed to accumulation

reports during blowing snow events. These events were corroborated by observer reports from a nearby airport.

The catch efficiency data for the test gauges at Caribou Creek (Figures 13a and 13b, 14a and 14b) are impacted by ground noise in the gauge output, which is believed to mitigate the expected decrease in median catch efficiency with increasing mean wind speed. As such, it is difficult to weigh the significance of the trends observed, in which the median catch efficiencies of both test configurations remain above 0.6 for mean wind speeds between 4 and 5 m/s. The difference in site conditions at Caribou Creek, which is in a southern Boreal climate, relative to those at the other sites, which are all in Continental climates, may also play a role.

Root mean square error values were computed from all 30 minute events during which each test configuration and the corresponding reference configuration both detected precipitation. Values are shown in Figure 17 and Table 6, and can be considered to represent the absolute uncertainty of each test configuration relative to the reference configuration in liquid, mixed, and solid precipitation, and in all precipitation types (note that the relative proportions of events of each phase differ by site). The RMSE for single-Alter shielded gauges in all precipitation types ranges from 0.14 mm/30 min for the test gauge at CARE to 0.22 mm/30 min for the test gauge at Bratt's Lake. The RMSE for the unshielded configurations in all precipitation types ranges from 0.25 mm/30 min for the test gauges at Bratt's Lake and CARE, to 0.32 mm/30 min for the test gauge at Marshall. The values for the shielded (0.26 mm/30 min) and unshielded (0.24 mm/30 min) gauges at Caribou Creek are similar to those for the continental sites, but may be impacted by gauge noise.

The overall catch ratio – computed from the total reference and SUT accumulation from all YY cases over the duration of formal tests – is provided for each test configuration in Table 8. For single-Alter shielded test configurations, the overall catch ratio ranges from 0.80 (Bratt's Lake; highest mean wind speeds) to 0.90 for the test gauge at CARE. For unshielded test configurations, the overall catch ratio ranges from 0.77 (Marshall) to 0.85 (CARE). The overall catch efficiencies for the shielded and unshielded configurations at Caribou Creek were higher, with values of 0.95 and 0.87, respectively; again, this is attributed to noise in the gauge output related to the configuration.

Decreasing the SUT accumulation threshold for precipitation events from 0.25 mm/30 min to 0.1 mm/min does not impact the overall catch ratio significantly, with the exception of the test gauges at Bratt's Lake, which are subject to the highest mean wind speeds. The reduction in the detection threshold decreases the overall catch efficiency of the shielded test configuration by approximately 9%, and that of the unshielded test configuration by almost 15%.

7.2.2.3. Ability to detect light precipitation

The detection thresholds and results presented in Tables 8 and 9, respectively, indicate that decreasing or removing the SUT detection threshold (while maintaining the reference detection threshold at 0.25 mm) leads to increases in both the Probability of Detection and False Alarm Rate to close to 100%, indicating negligible detection skill (e.g. Heidke Skill Score of close to 0%). Decreasing both the reference and SUT detection thresholds to 0.1 mm increases the POD by less than 1% and increases the FAR by just over 1%. These results suggest that reducing both detection thresholds

does not improve significantly the detection of light precipitation events relative to the 0.25 mm thresholds.

7.2.3.4. Assessment of events when the reference and SUT do not agree on the occurrence of precipitation

The YN ('miss') cases, when the reference detects a precipitation event and the SUT does not, and NY ('false alarm') cases, when the SUT detects a precipitation event and the reference does not, are characterized for selected test configurations in Figures 18 and 19. The majority of the YN cases (Figures 18a and 19a) have reference accumulations just above the 0.25 mm threshold and SUT accumulations below the threshold, likely resulting from enhanced wind effects for the single-Alter shielded and unshielded SUT configurations relative to the reference gauge in the DFIR-fence. The false alarm cases for the shielded test gauge at CARE (Figure 18b) are made up of cases for which either the reference accumulation threshold (0.25 mm) or precipitation occurrence threshold (18 minutes) was not met or exceeded, while those for the unshielded test gauge at Marshall (Figure 19b) are predominantly cases in which the precipitation occurrence threshold was not met or exceeded. These results illustrate the influence of the detection criteria on the assessment results.

8) Maintenance

Gauge calibration: each site completed the gauge field calibration and verification as per manufacturer recommendations, at least once a year or following the emptying of the gauge. The calibration records have been stored by each site host.

9) Performance Considerations

9.1. Data acquisition and processing

The Geonor T-200B3 must be operated in conjunction with a data logger. Additional processing is required to translate the gauge output into meteorologically-representative data. This requires an understanding of gauge performance, and generally, a well-educated user. The manufacturer is encouraged to propose an algorithm for deriving baseline datasets from the gauge output to allow for more consistency in the data derived by different users.

The SPICE data quality control and event selection procedures have been applied to filter accumulation data from Geonor T-200B3 gauges and derive precipitation amounts over specified intervals. Such procedures are recommended to reduce the variability in gauge reports and establish analysis-ready data products. The specific threshold employed for distinguishing between precipitation and signal noise is an important consideration; the present results indicate similar performance for accumulation thresholds of 0.25 mm and 0.1 mm.

9.2. Gauge configuration

The configuration and installation of Geonor T-200B3 gauges at a given site can impact significantly gauge performance. The single-Alter shielded test configurations show higher overall catch efficiencies, lower RMSE values, and more gradual decreases in median catch efficiency for solid precipitation with increasing wind speed relative to the unshielded test configurations. Shielded

configurations are therefore recommended, where possible. All field configurations must be fully tested and validated prior to use, in order to ensure that physical and signal interferences are not degrading the gauge signal. This would include the confirmation of grounding tailored to the soil conditions, sturdiness of foundation and mounting, and signal conditioning.

When heating is applied, testing of the heating circuits must be performed, including testing their impact to the sensor signal. Given the intermittence of the supply of power to the heating elements, the risk of interference needs to be minimized. It is recommended that the Geonor T-200B3 gauges are fitted with heating circuits integrated and validated by the manufacturer, in order to provide a consistent and proven solution. In its absence, a user without experience in configuring the heating circuits could implement circuits that would negatively affect the quality of the sensor output.

9.3. Ancillary measurements and adjustments

Ancillary measurements from a sensitive precipitation detector that is independent from the test gauge are recommended to help distinguish precipitation events from false reports due to noise. The application of adjustment functions is strongly recommended to account for the reduction in gauge catch efficiency as the wind speed increases. Additional ancillary measurements of wind speed (ideally at gauge height) and air temperature are required for the application of transfer functions.

WMO-SPICE Instrument Performance Report
Geonor T-200B3MD, 1500mm

1) Technical specifications (from manufacturer provided documentation)

Instrument model:	Geonor T-200B3MD (3: three transducers; MD: 1500mm capacity)
Physical principle:	Weighing gauge (WG) based on catchment principle; weight measurement by vibrating wire transducers
Capacity:	1500 mm
Collecting area :	200 cm^2
Operating temperature range:	-40 °C to 60 °C
Measurement uncertainty:	0.1% of full scale (as defined by manufacturer)
Sensitivity:	0.1 mm (as defined by manufacturer)

Figure 1: Geonor T-200B3MD test gauges in single-Alter shields at (a) Marshall (USA), (b) Weissfluhjoch (Switzerland), (c) Caribou Creek (Canada), and (d) Bratt's Lake (Canada), and (e) in Belfort double-Alter shield at CARE (Canada).

2) Data output format

Transducer output: Frequency (f); 0-5 V square wave

Data range: From ~1000 Hz (empty gauge) to ~3000 Hz (full gauge)

Gauge data output: Precipitation accumulation for each transducer, derived from the frequency output as outlined below. The arithmetic average of accumulation from the three transducers was used in subsequent data analysis.

The bucket content is weighed using a precision load cell with a high-tension vibrating wire (VW) transducer. Under load, the wire vibration frequency is proportional to the weight detected (and the associated precipitation amount, P), based on a quadratic relationship.

$P = A (f - f_0) + B (f - f_0)^2$

Where:

P = precipitation (in cm)
f = frequency reading (Hz)
A, B = calibration constants, available from the Calibration Certificates
$\mathbf{f_0}$ = frequency with empty bucket at calibration (Hz), available from the Calibration Certificates

The signal from a Geonor gauge transducer is amplified into a measurable quantity sampled with an external data logger, using a user defined strategy (duration, frequency) and logger specific functions. The sampling of the transducer signal is, generally, not continuous.

All sites testing Geonor T-200B3MD gauges used Campbell Scientific CR3000 data loggers, and applied one of the two dedicated programming functions available:

1. Period average (PeriodAvg) command, which calculates the sensor's average frequency over a user-specified number of cycles within a defined interval. For example, the CARE SPICE site used a sampling strategy in which 250 cycles were sampled every 6 seconds, for each transducer.

2. Pulse count (PulseCount) command, which calculates the sensor's frequency by counting the number of pulses over a specified time period, which is then converted into a frequency.

3) SPICE test configuration

Number of transducers:	Three vibrating wire transducers for each test gauge, sampled independently using data loggers
Shield(s):	Single-Alter shield (SA), Belfort double-Alter shield (BDA)
Test site(s):	CARE, Bratt's Lake, Caribou Creek (Canada); Marshall (USA); Weissfluhjoch (Switzerland)
Sensor provider(s):	The gauges tested at Bratt's Lake, Caribou Creek, Marshall and Weissfluhjoch were provided by the manufacturer of the instrument (Geonor AS, Norway). Environment and Climate Change Canada provided the gauge tested at the CARE site.

A map of test site locations is provided in Figure 2.

3.1. Note on terms and acronyms used

Throughout this document, the following notations were used to identify the R2 reference (Ref) and senor under test (SUT) configurations:

Reference: 'DFIR' and 'DFAR' are used interchangeably for the R2 reference configurations. 'DFIR' refers to an automated gauge installed in a DFIR-fence, while 'DFAR' refers more explicitly to the Double-Fence Automated Reference configuration. R2 reference configurations employing Geonor T-200B3 (600 mm) gauges are denoted as 'Geo600,' while those employing Bucket RT data from OTT Pluvio^2 gauges are denoted as 'Plv2BktRT.'

Sensors under test: 'SA' denotes single-Alter shielded test configurations and 'BDA' denotes Belfort double-Alter shielded test configurations.

Figure 2: Map of SPICE sites where Geonor T-200B3MD gauges were tested.

A summary of the configuration of instruments as tested, the duration of tests and availability of data reflected in these results, and the ancillary measurements used, by site, is available in Tables 1, 2, and 3, respectively.

Table 1: Summary of gauge configurations and data output, by site. Details and photos on individual site configurations are available in the respective site commissioning protocols.

	Bratt's Lake	CARE	Caribou Creek	Marshall	Weissfluhjoch
Field configuration (shield)	SA	BDA	SA	SA	SA
Height of installation (gauge rim)	2.5 m	2.0 m	2.5 m	2.0 m	3.5 m
Heating	No heating	Site-specific heating[1]	No heating	No heating	Site-specific heating[2]
Antifreeze	Mixture of 60% methanol, 40% propylene glycol†				
Oil	VoltEsso35	Bayoil (Season 1) Isopar (Season 2)	VoltEsso35	Automatic Transmission Fluid (ATF)	Linseed oil (Season 1) Isopar (Season 2)
Data output frequency	1 min	6 sec	1 min	6 sec	6 sec
Data QC	SPICE QC methodology				
Data temporal resolution	1 min				
Processing interval for SPICE data analysis	30 min				

[1] Heating configuration similar to that employed in SPICE, but with heaters powered individually, rather than in parallel (see Section 4.2.2.2 of WMO-SPICE Final Report for details).

[2] Manufacturer-recommended approach in which the heater is installed on the outer surface at the top of the inlet and is turned on for 10 minutes of every hour.

† For the test gauge at Weissfluhjoch, the 60% methanol, 40% propylene glycol mixture was used during Season 2, only. A mixture of 75% propylene glycol, 25% water was used during Season 1.

Table 2: Data availability, by measurement season and site.

Measurement season	Bratt's Lake	CARE	Caribou Creek	Marshall	Weissfluhjoch
Season 1 (Oct. 2013 – Apr. 2014)	✓	✓	X	✓	✓
Season 2 (Oct. 2014 –Apr. 2015)	✓	✓	✓	✓	✓

Table 3: Summary of reference and ancillary measurements, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	Bratt's Lake	CARE	Caribou Creek	Marshall	Weissfluhjoch
R2 Site Reference	Geonor T-200B3 600 mm (DFAR)	Geonor T-200B3 600 mm (DFAR)	Geonor T-200B3 600 mm (DFAR)	Geonor T-200B3 600 mm (DFAR)	OTT Pluvio^2 Bucket RT (DFAR)
R2 Precip Detector	Thies Precip Sensor (DFAR)	Thies LPM (DFAR)	Thies Precip Sensor (DFAR)	Thies LPM (Site*)	Thies LPM (DFAR)
Ancillary Temp Sensor	Vaisala HMP45 (2m)	Vaisala HMP155 (SS$^+$, 1.5 m)	Vaisala HMP45 (2m)	MetOne, 060A-2/062, 2144-L (2 m)	Meteolabor AG VTP6 Thygan (5 m)
Ancillary RH Sensor	Vaisala HMP45 (2m)	Vaisala HMP155 (SS$^+$, 1.5 m)	Vaisala HMP45 (2m)	Campbell Scientific CS500 (2m)	Meteolabor AG VTP6 Thygan (5 m)
Ancillary Wind Sensor	RM Young Wind Monitor 05103 (2m)	Vaisala NWS 425 (2m)	MetOne 13A (2m)	RM Young Wind Monitor 05103 (2 m)	RM Young Wind Monitor 05103 (5.5 m)

*A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR shield; however, in cases where a more sensitive detector is available outside of the DFIR shield, or there are issues with the detector within the DFIR shield, a precipitation detector elsewhere on the site can be employed.

$^+$ SS denotes that the sensor is installed inside a ventilated Stevenson Screen.

4) Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6.1 of the WMO-SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

4.1.1.1. Characterization of performance in non-precipitating conditions

The assessment data are derived over 30 minute intervals during which the precipitation detector in the R2 reference configuration reports 0 minutes of precipitation. The accumulation over these intervals (accumulation in minute 30 – accumulation in minute 1), representing the variability of the gauge response due to wind, evaporation, temperature, etc., is recorded, along with the mean wind speed, and the change in temperature (temperature in minute 30 – temperature in minute 1).

4.1.1.2. Assessment of ability to detect and report accumulation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute assessment intervals:

- Ref 'Yes' : R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No' : R2 weighing gauge < 0.25 mm AND/OR precip detector recording < 18 min of precip;
- SUT 'Yes' : SUT accumulation ≥ 0.25 mm;
- SUT 'No' : SUT accumulation < 0.25 mm.

4.1.1.3. Assessment of ability to detect light precipitation

The data for this component of the assessment are derived in a similar manner as those in Section 4.1.1.2, but with different combinations of thresholds for the reference and/or SUT 'Yes' and 'No' conditions. These different threshold 'cases' have been selected to demonstrate the impact of the thresholds used in data derivation on the detection of light precipitation.

4.1.2. Skill score assessment

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores, as detailed in Section 3.6.1.3 of the WMO-SPICE Final Report.

For the assessments considered in this report, the ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

5) Environmental conditions

The environmental conditions at each site over the duration of the test period are expressed as probability density functions (PDFs) of mean air temperature, mean relative humidity, mean wind speed, vector mean wind direction, and precipitation rate for each component 30 minute assessment interval in Figure 3. The same parameters are also shown for all assessment intervals during which the site reference configuration detected precipitation (i.e. all Ref 'Yes' cases) in Figure 4.

The precipitation percentage represents the number of minutes of precipitation during a 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration, expressed as a percentage. PDFs of precipitation percentage are also included in Figures 3 and 4.

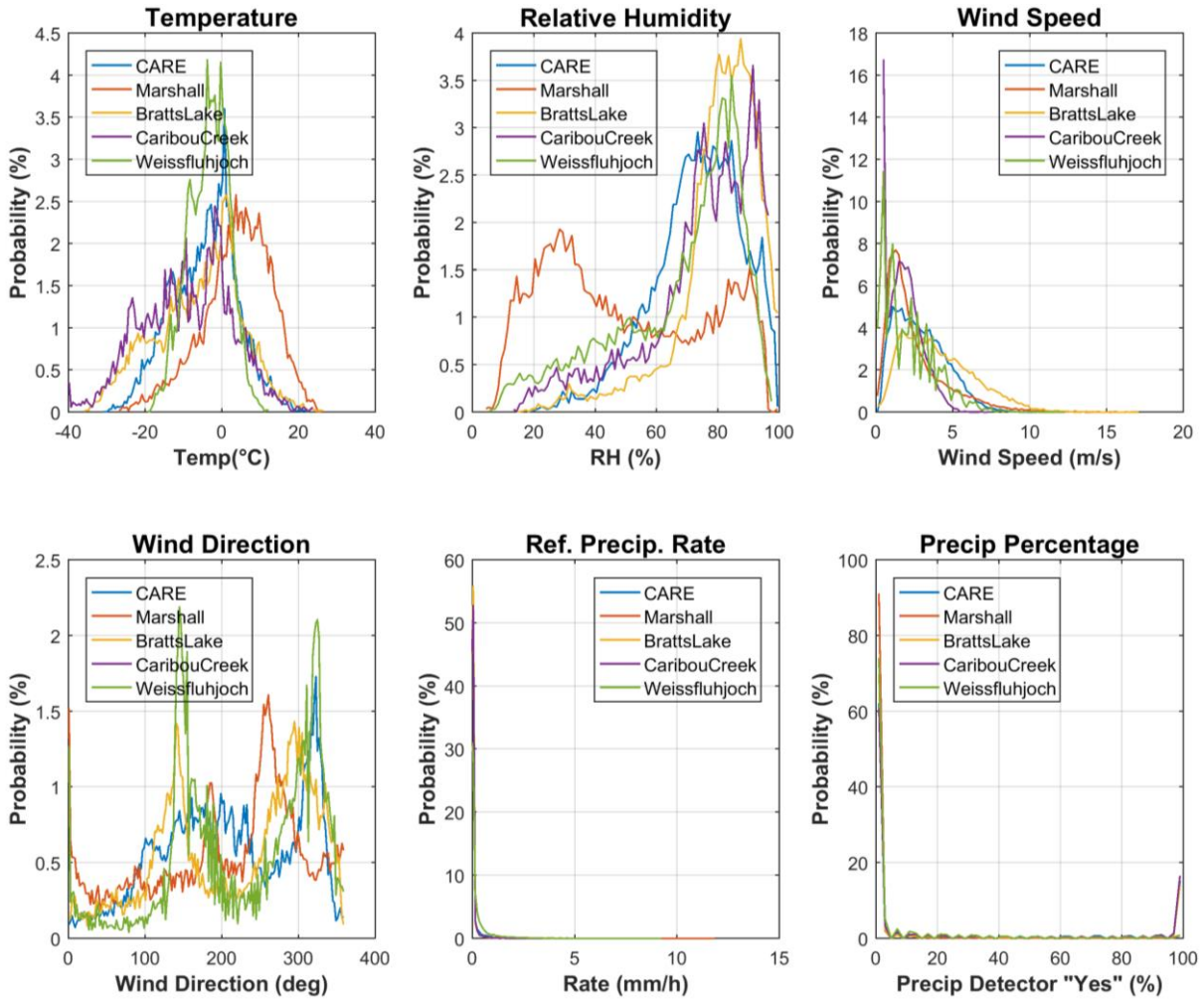


Figure 3: Summary of environmental conditions at sites with Geonor T-200B3MD test gauges over the entire duration of formal tests, as per Table 2.

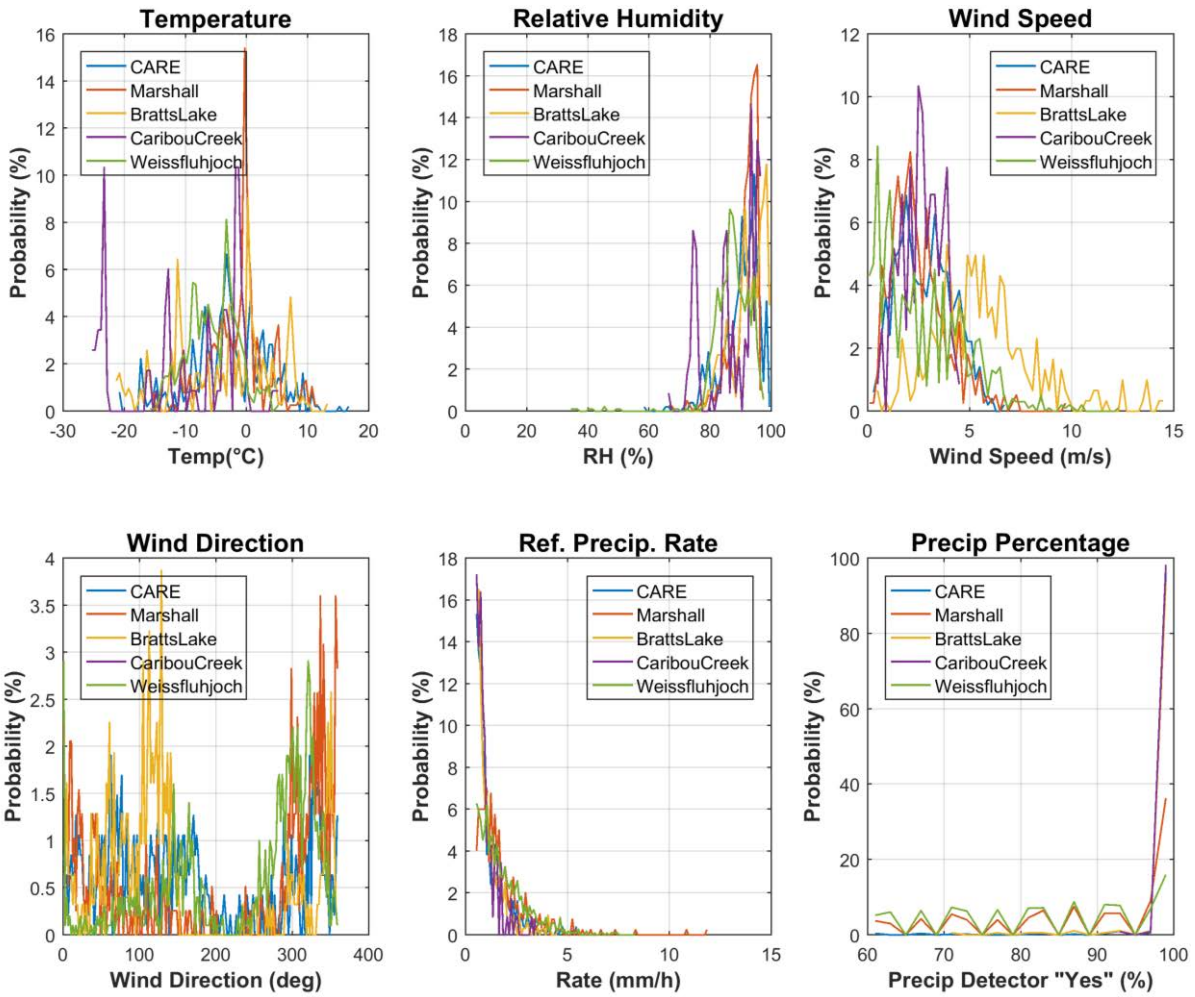


Figure 4: Summary of environmental conditions at sites with Geonor T-200B3MD test gauges during precipitation events (as defined by the R2 reference), over the duration of formal tests.

6) Evaluation of the ability to perform over the range of operating conditions

6.1. Characterization of SUT performance in non-precipitating conditions

The response of the SUT in the absence of precipitation was examined as defined in Section 4.1.1.1. The results are presented below, reflecting the distribution of the sensor response and its variability with wind and temperature, as measured during 30 minute assessment intervals.

6.1.1. Overall variability of SUT response

The overall variability of the SUT response in non-precipitating conditions is shown as a probability density function for each test configuration in Figure 5. The corresponding PDF for the reference configuration at each test site is provided for comparison.

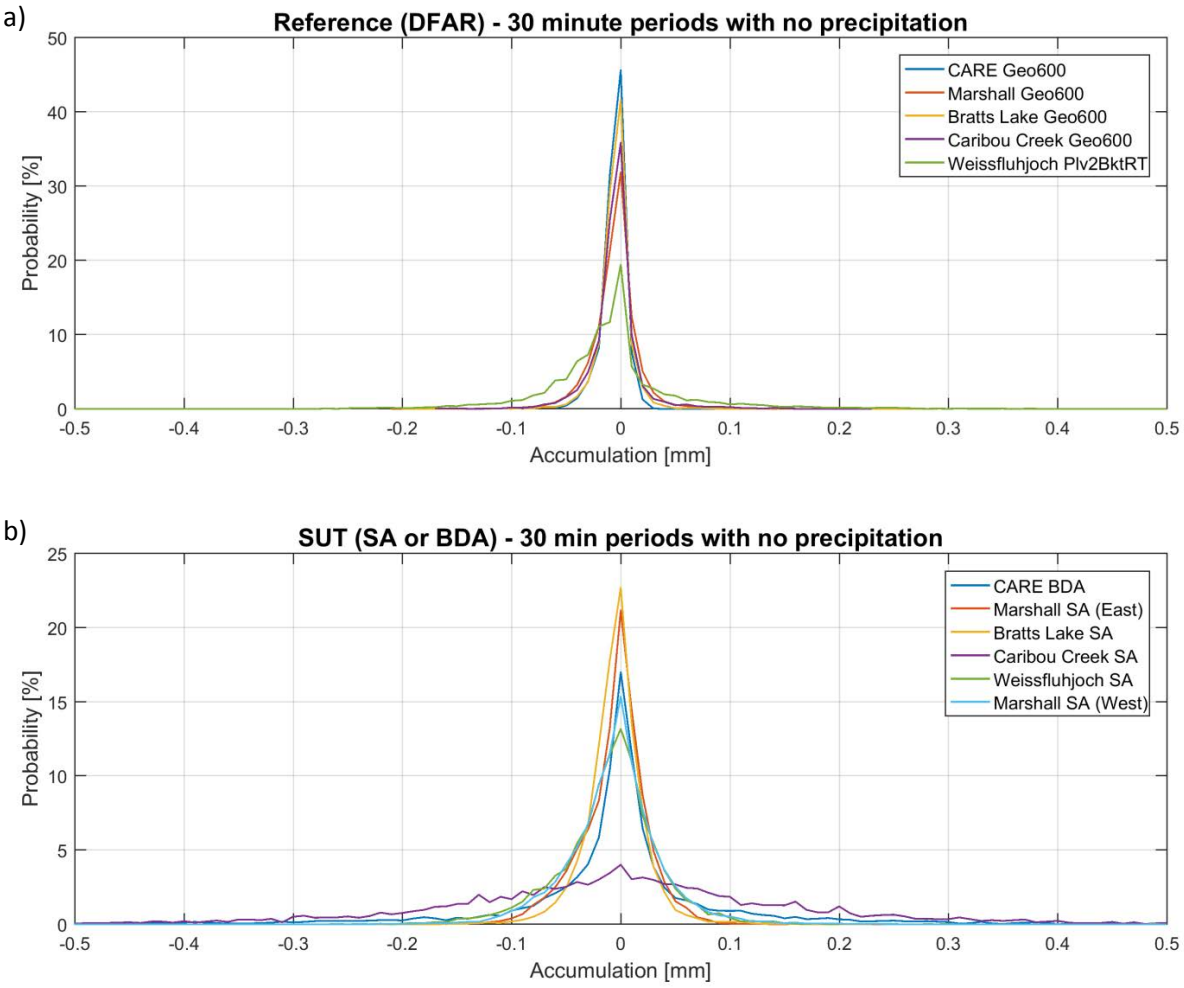


Figure 5: Probability density functions of the output signal (accumulation over 30 minute assessment intervals) in non-precipitating conditions for (a) the R2 reference configurations (DFAR) and (b) Geonor T-200B3MD test gauges in SA- and BDA-shielded configurations.

The statistics of the output signal (accumulation over 30 minute assessment intervals) for the reference and SUT at each site are provided in Table 4.

Table 4: Statistics of the R2 reference gauge (DFAR) and SUT output signal during non-precipitating conditions at sites testing Geonor T-200B3MD gauges, as plotted in Figure 5.

Site	Gauge (sensor)	Average output signal (mm)	Standard deviation (mm)	Maximum output signal (mm)	Minimum output signal (mm)	Number of assessment intervals
Bratt's Lake	DFAR	-0.000	0.016	0.293	-0.194	7226
	SUT	-0.000	0.028	0.235	-0.225	7226
CARE	DFAR	-0.001	0.011	0.204	-0.059	9187
	SUT	0.000	0.125	0.983	-1.370	9187
Caribou Creek	DFAR	-0.000	0.024	0.233	-0.168	4306
	SUT	-0.002	0.199	1.043	-1.843	4306
Marshall	DFAR	-0.001	0.022	0.196	-0.207	18123
	SUT (East)	-0.001	0.039	1.167	-0.415	18123
	SUT (West)	0.000	0.054	1.359	-1.501	17960
Weissfluhjoch	DFAR	-0.005	0.068	1.200	-0.750	11502
	SUT	-0.004	0.061	2.695	-0.870	11502

6.1.2. Variability of SUT response as a function of temperature

The variability of the SUT response for each test configuration in the absence of precipitation is plotted as function of the temperature difference over each assessment interval in Figure 6. The temperature difference is defined as the difference in temperature between the end (minute 30) and beginning (minute 1) of the assessment interval. The corresponding plots for the reference configurations are provided for comparison.

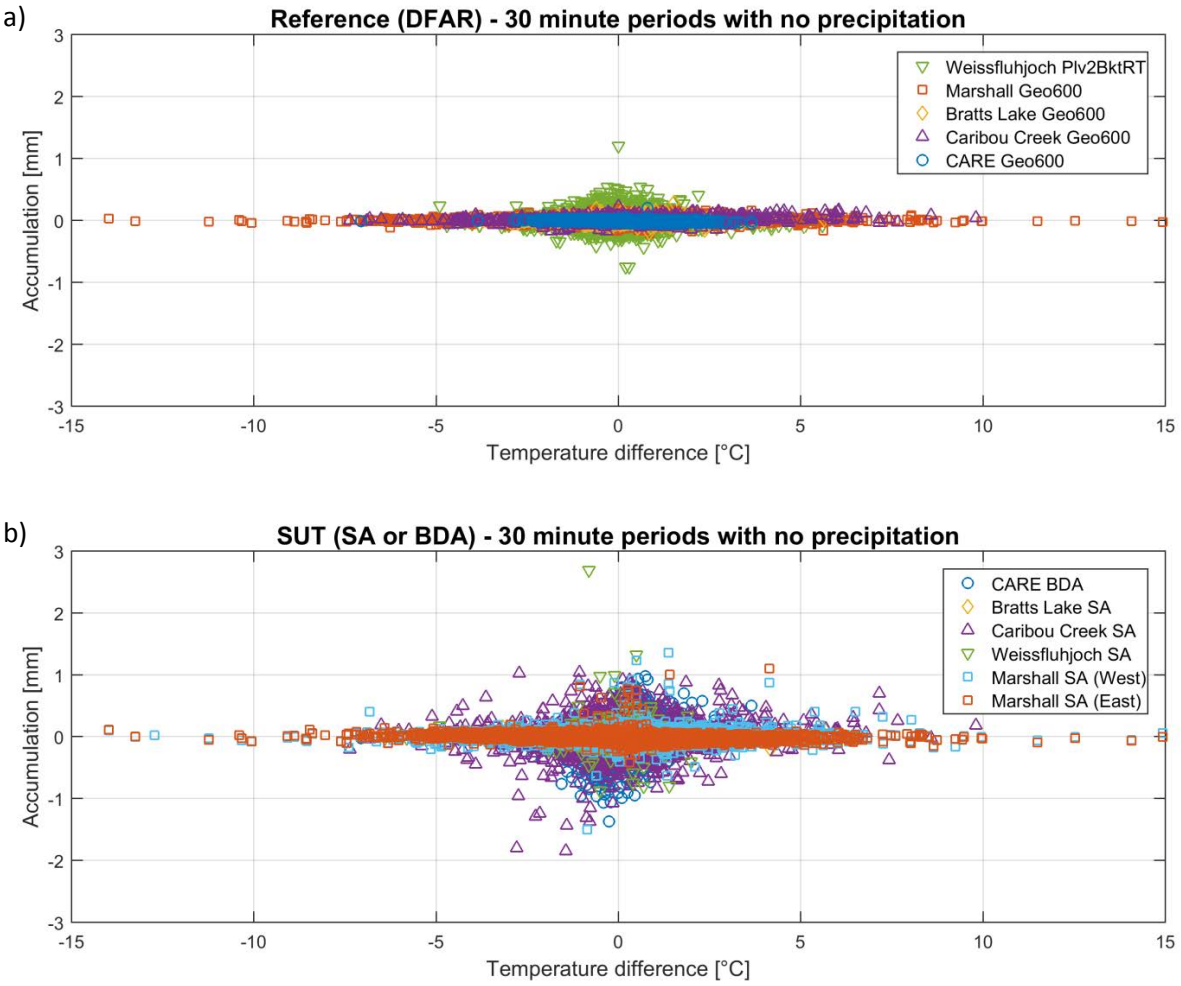


Figure 6: Variability of output signal (accumulation over each 30 minute assessment interval) as a function of the temperature difference over the interval in non-precipitating conditions for (a) the R2 reference configurations (DFAR) and (b) Geonor T-200B3MD test gauges in SA- and BDA-shielded configurations.

6.1.3. Variability of SUT response as a function of wind speed

The variability of the SUT response for each test configuration in the absence of precipitation is plotted as function of the mean wind speed for each assessment interval in Figure 7. Here, the signal variability is represented as the standard deviation (STD) of the gauge accumulation output over each 30 minute interval. The corresponding plots for the reference configurations are provided for comparison.

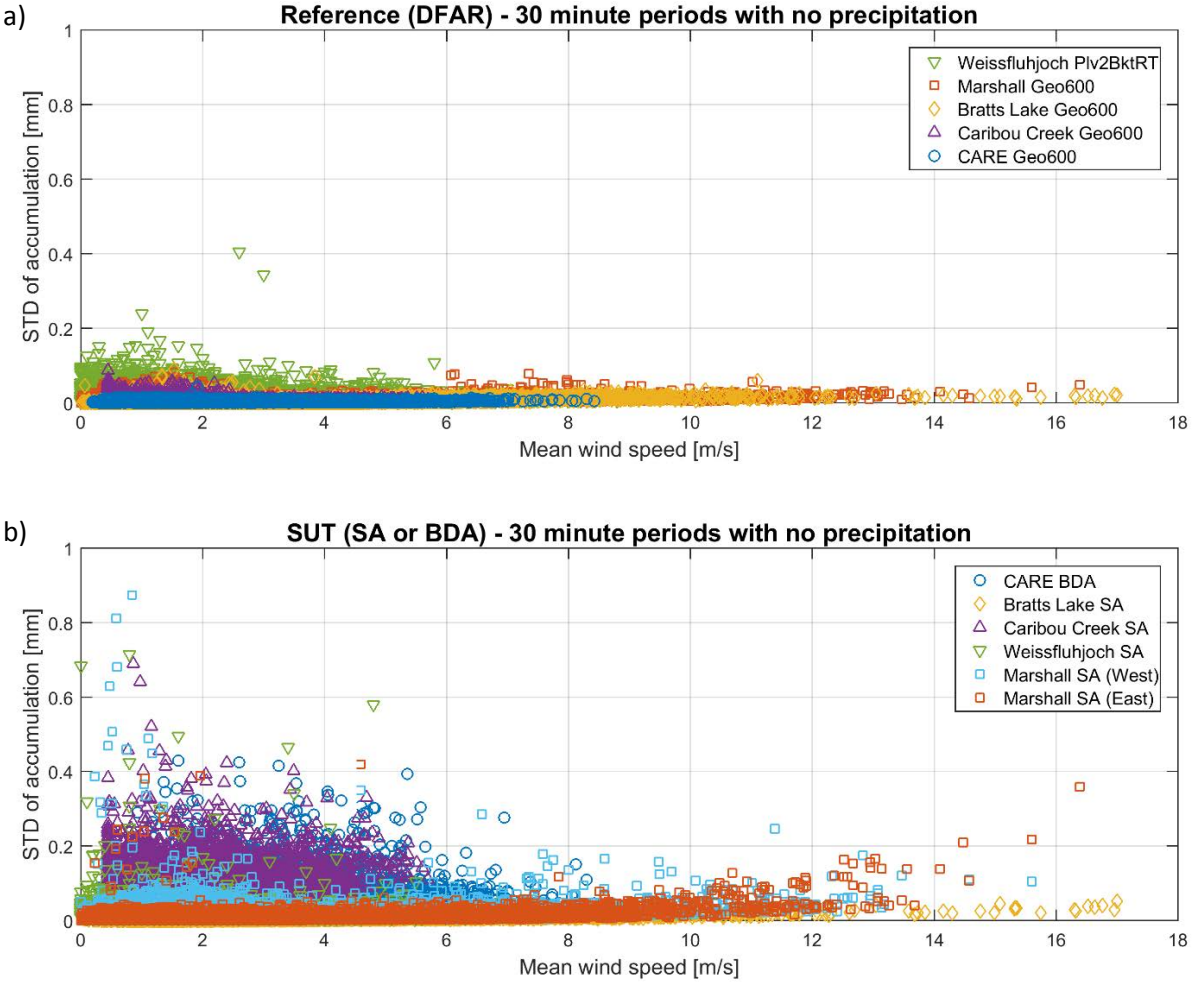


Figure 7: Variability of output signal (accumulation over each 30 minute assessment interval) as a function of mean wind speed in non-precipitating conditions for (a) the R2 reference configurations (DFAR) and (b) Geonor T-200B3MD test gauges in SA- and BDA-shielded configurations.

6.2. Ability to detect and report precipitation

6.2.1. Skill score assessment

The overall ability of the SUT to detect and report the occurrence of precipitation relative to the site field reference R2 over 30 minute assessment intervals is expressed using selected skill scores (Section 4.1.2) and presented in Figure 8. The contingency results (Section 4.1.1) corresponding to these scores are presented in Table 5.

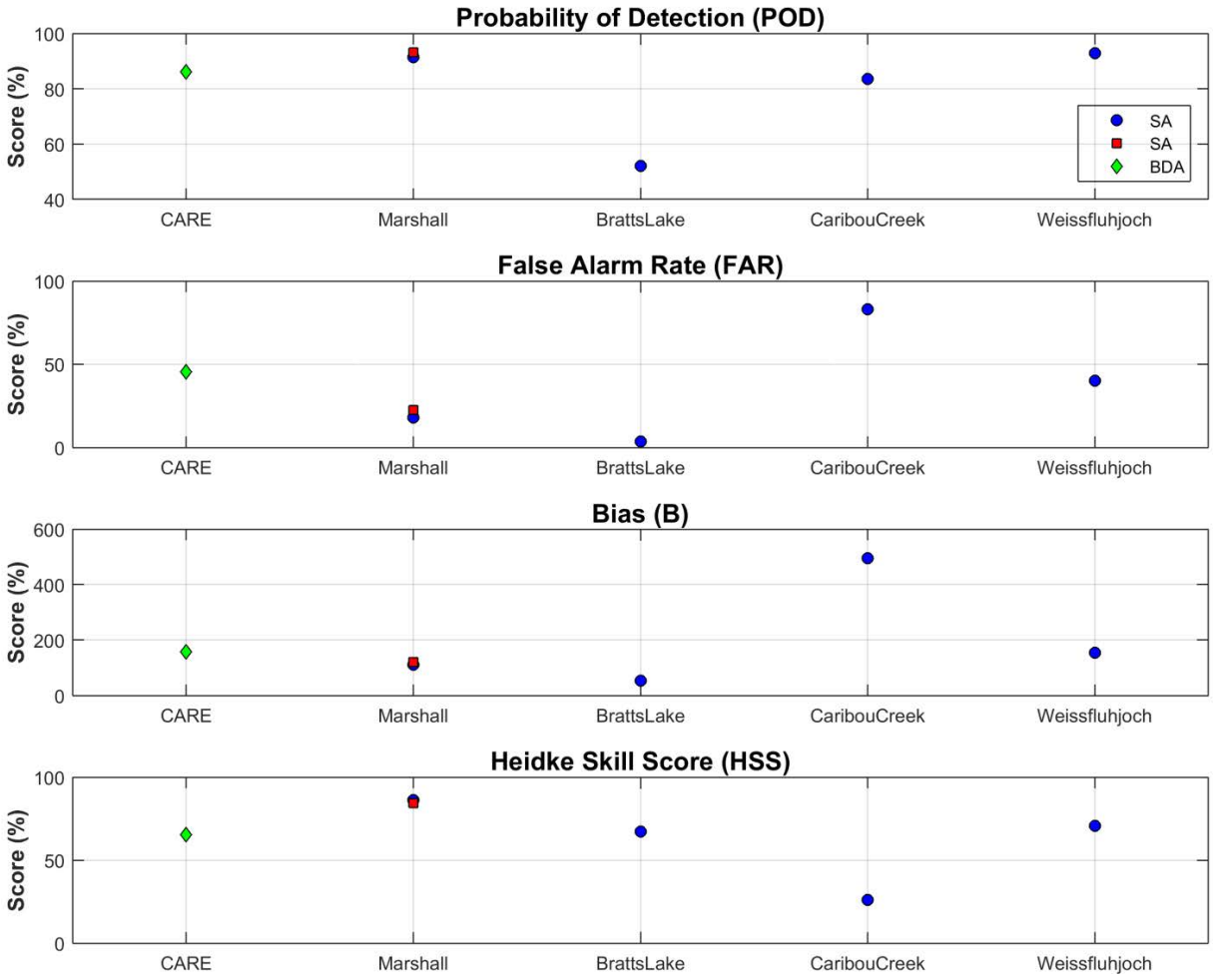


Figure 8: Skill scores for Geonor T-200B3MD gauges over the duration of formal tests.

Table 5: Contingency table illustrating detection of precipitation by Geonor T-200B3MD gauges under test relative to the corresponding site reference configurations, expressed as the number of events over the duration of formal tests.

Site	Number of Events			
	YY (hits)	YN (misses)	NY (false alarms)	NN (correct negatives)
CARE	426	69	355	13896
Bratt's Lake	164	151	6	16920
Caribou Creek	97	19	477	6260
Marshall (East)	365	34	80	19434
Marshall (West)	365	27	106	19179
Weissfluhjoch	1003	78	674	17222

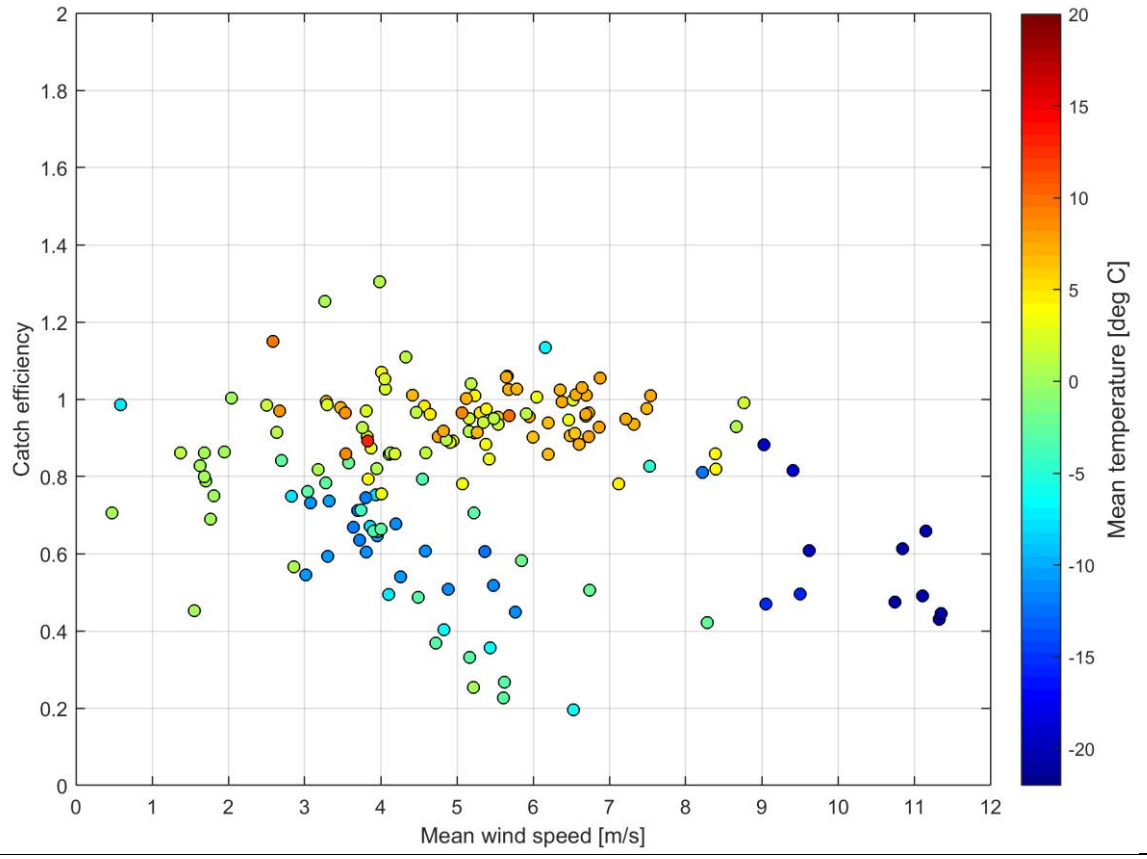
6.3. Ability to report accumulated precipitation

The SUT performance in terms of reporting accumulated precipitation is examined by comparing the amount reported by the sensor under test relative to the respective site reference during 30 minute assessment intervals. This is represented graphically using scatter and box and whisker plots of the catch efficiency as a function of mean wind speed at gauge height, as well as scatter plots of the amounts reported by the SUT versus the corresponding reference amounts (Figures 9 to 14). The SUT performance is also assessed in terms of the root mean square error, RMSE (Figure 15).

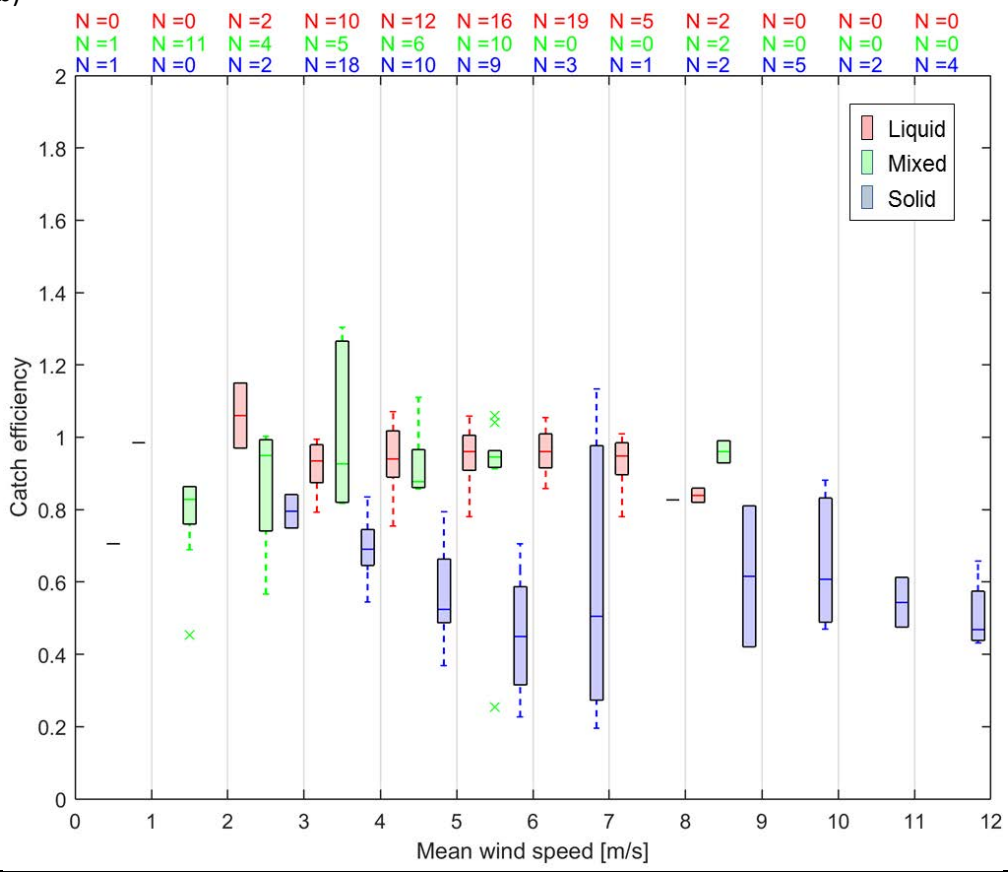
Only assessment intervals during which the SUT and reference both reported precipitation (YY cases) are considered in this portion of the assessment. In the catch efficiency-wind speed scatter plots, the mean event temperature is indicated by colour, with the colour scale selected to be consistent across all sites with weighing gauges under test. In the box and whisker plots and accumulation-accumulation scatter plots, the predominant precipitation type is indicated by colour, as determined from the reported temperature (Section 4.1.4).

Figure 9: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the SA-shielded Geonor T-200B3MD gauge under test at Bratt's Lake.

a)



b)



c)

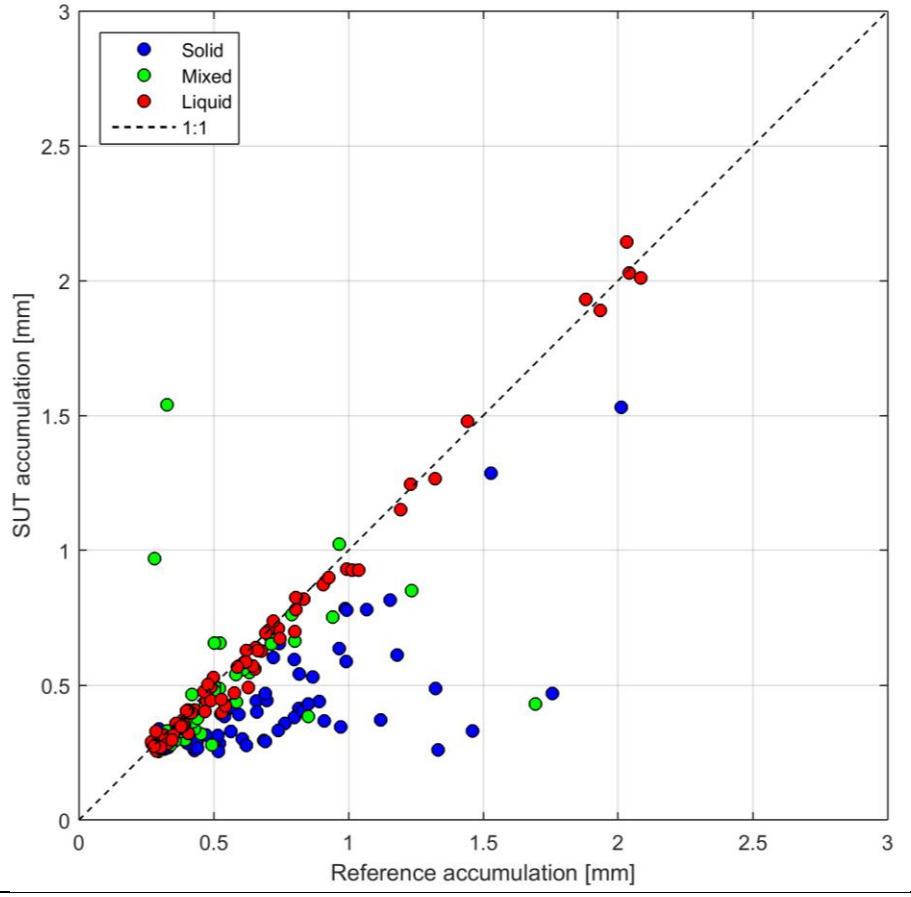
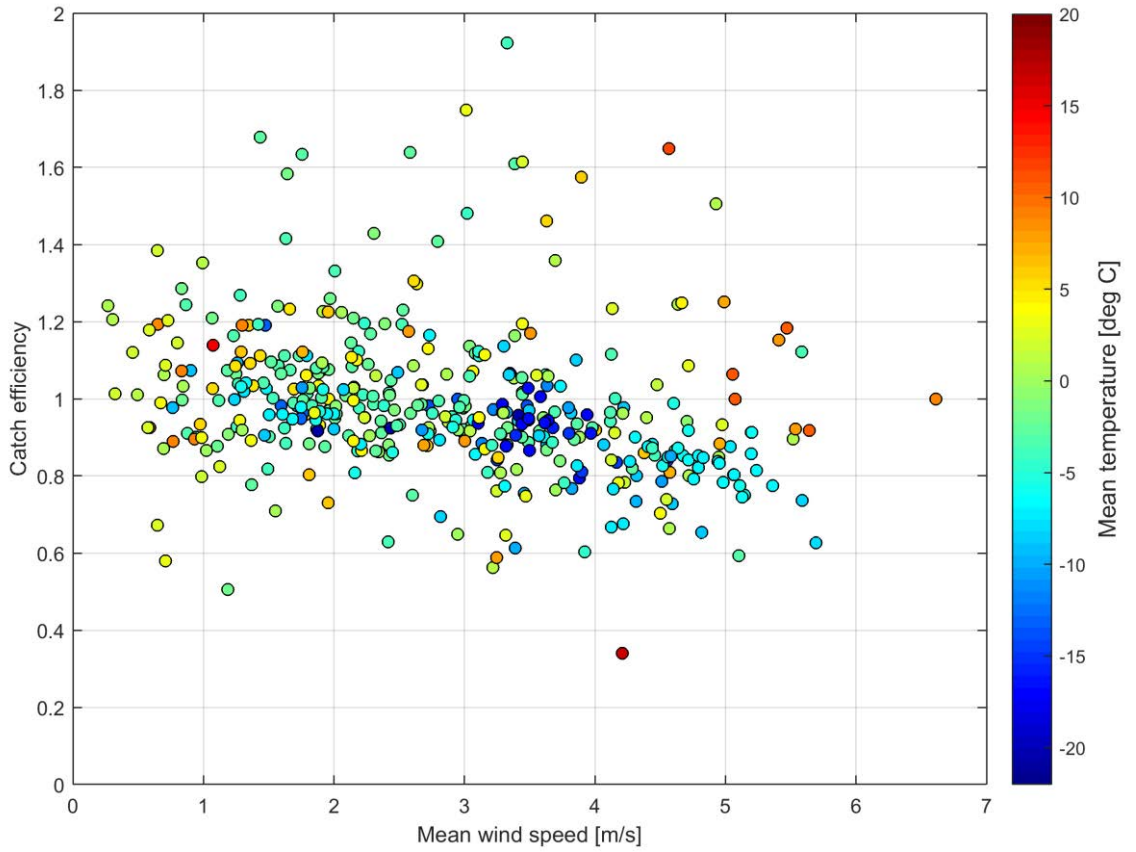
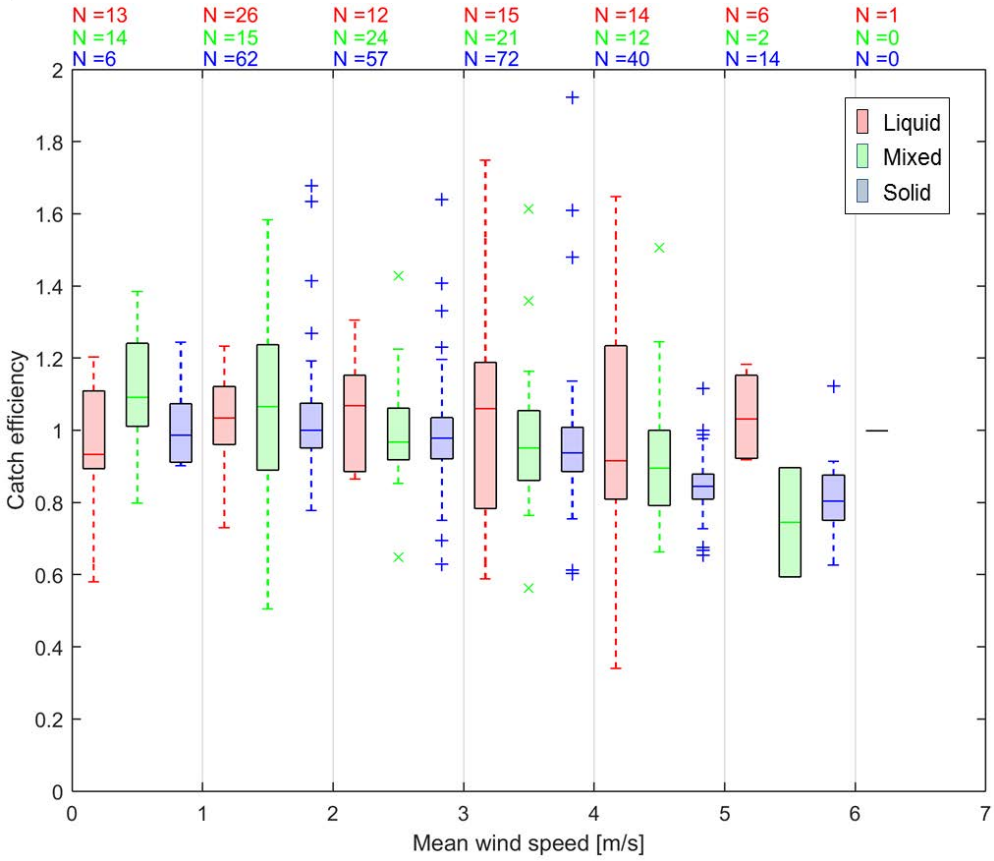

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
0.5
1
1.5
2
2.5
3

Figure 10: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the BDA-shielded Geonor T-200B3MD gauge under test at CARE.

a)



b)



c)

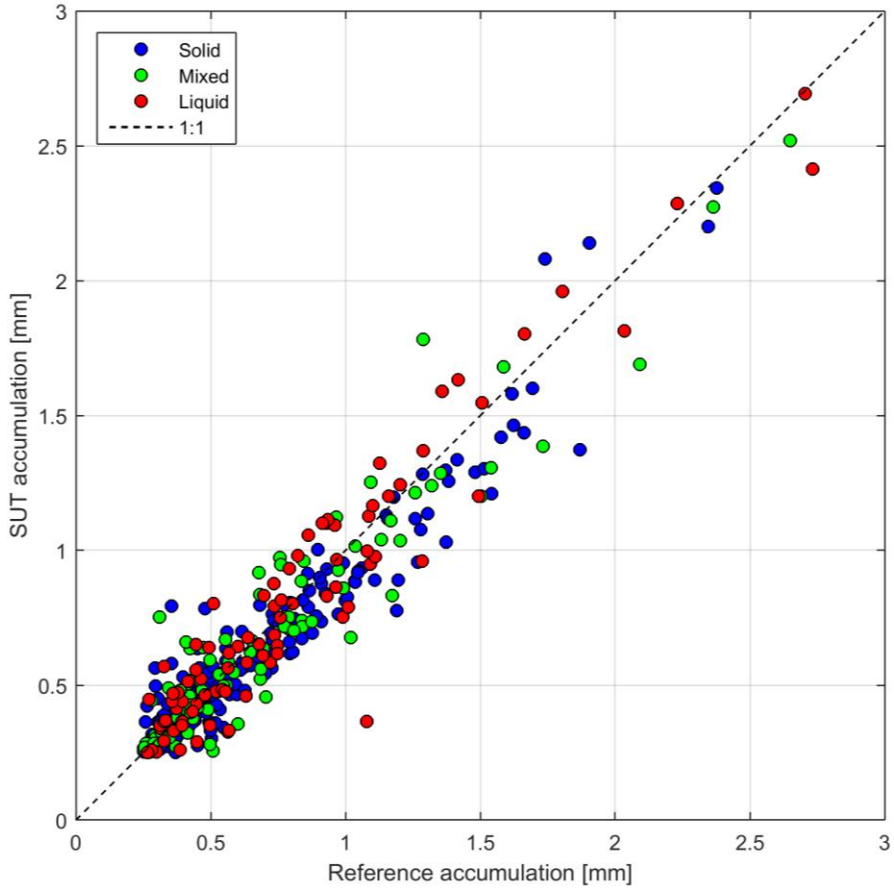
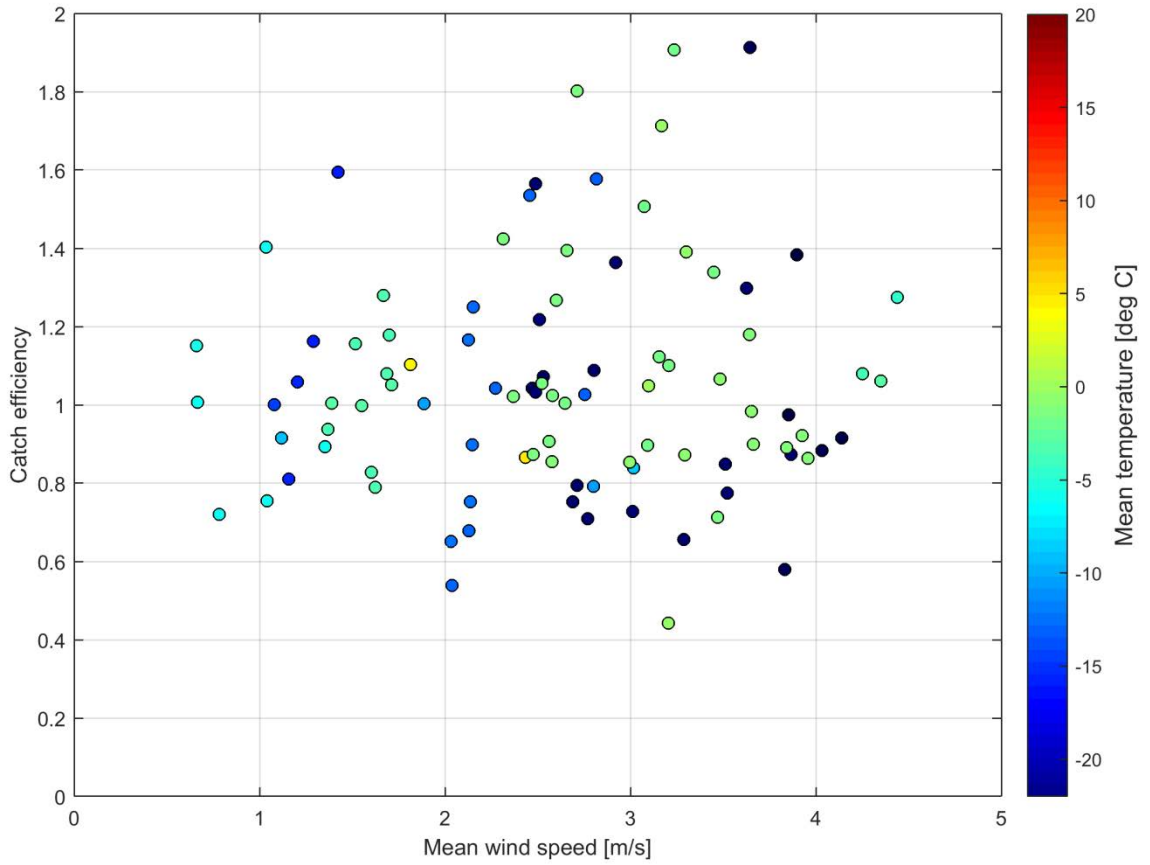
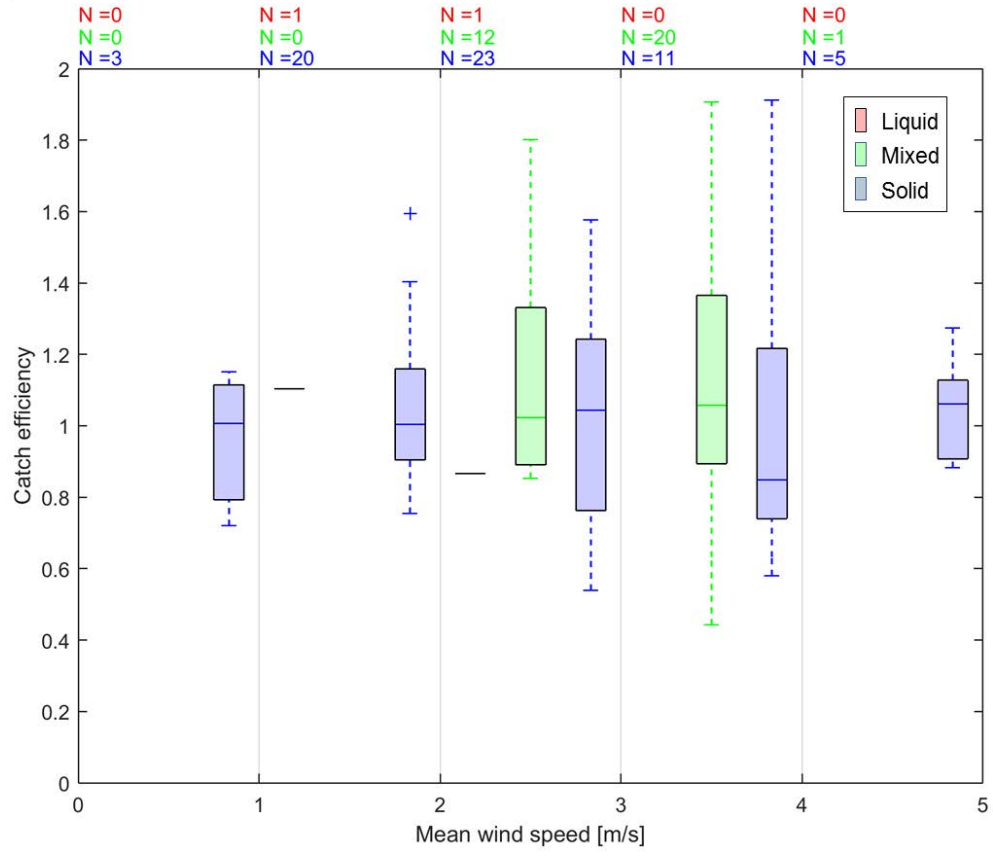

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
0.5
1
1.5
2
2.5
3

Figure 11: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the SA-shielded Geonor T-200B3MD gauge under test at Caribou Creek.

a)



b)



c)

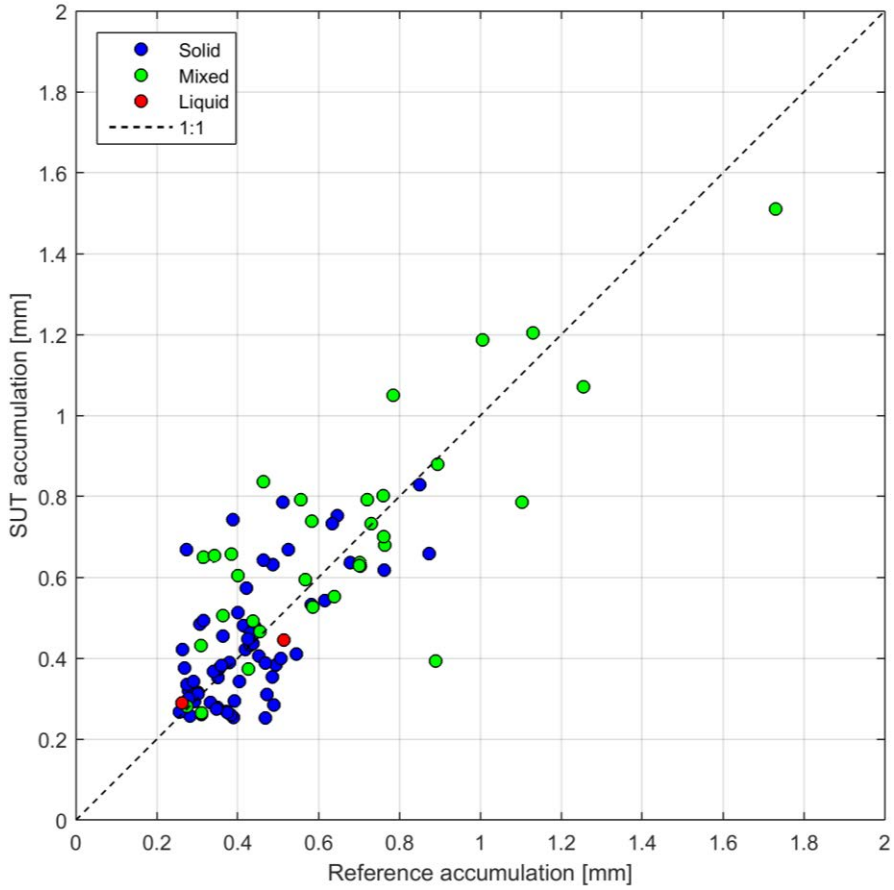
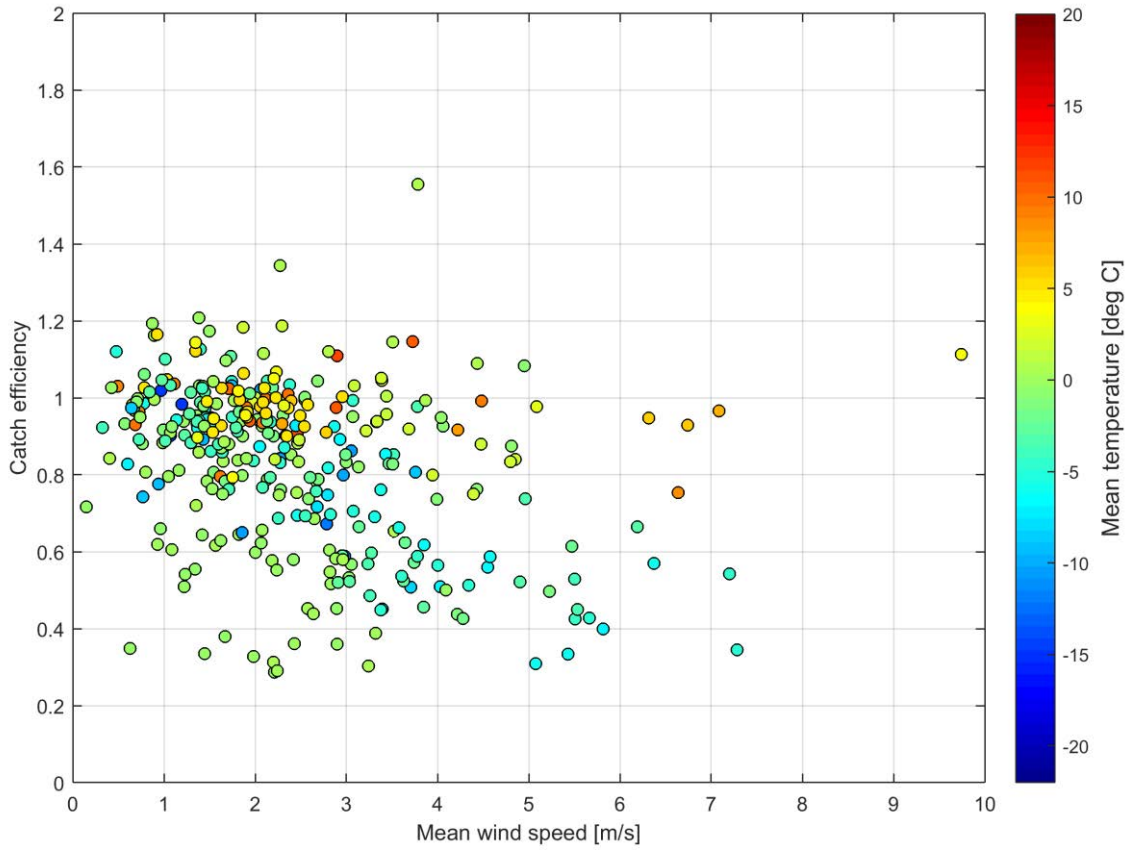
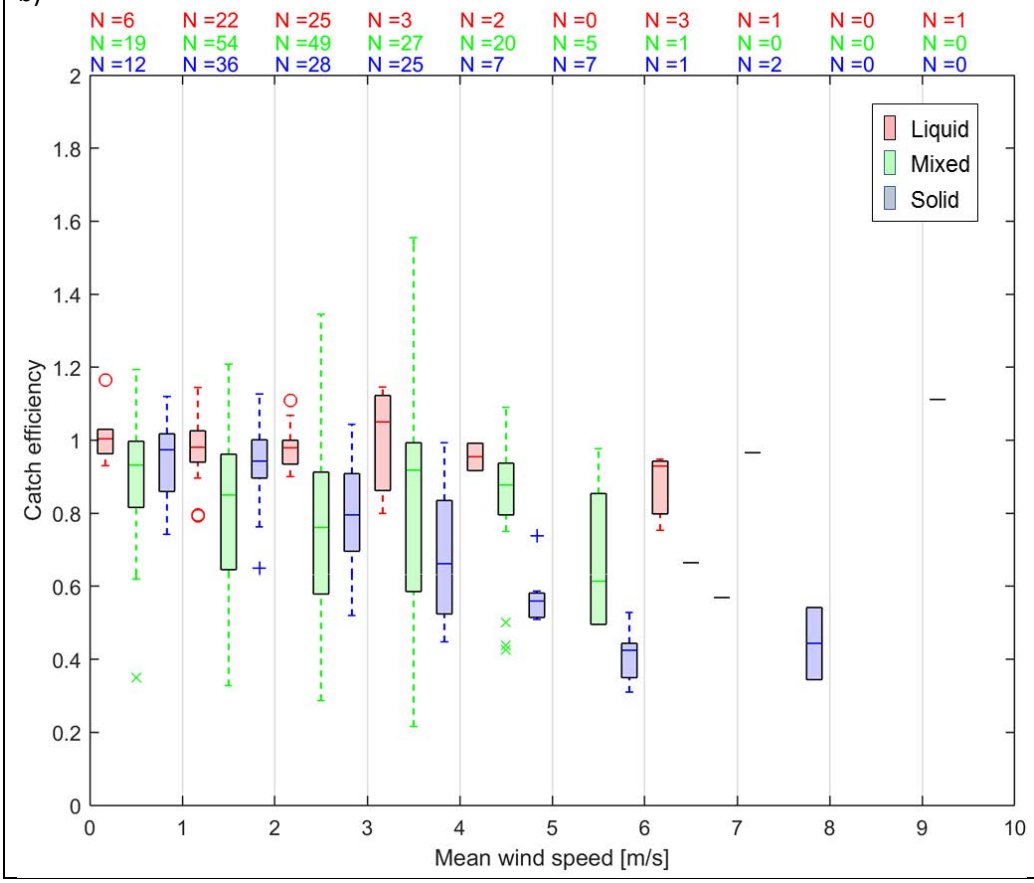

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

Figure 12: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the SA-shielded Geonor T-200B3MD gauge under test at Marshall (East sensor).

a)



b)



c)

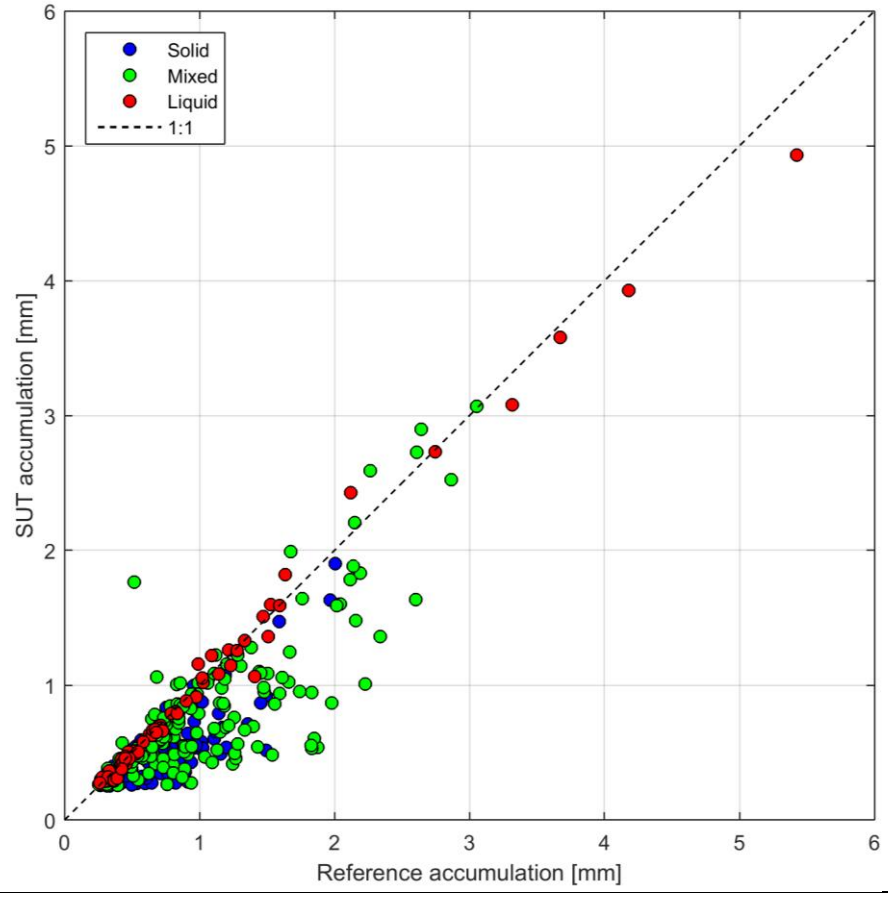
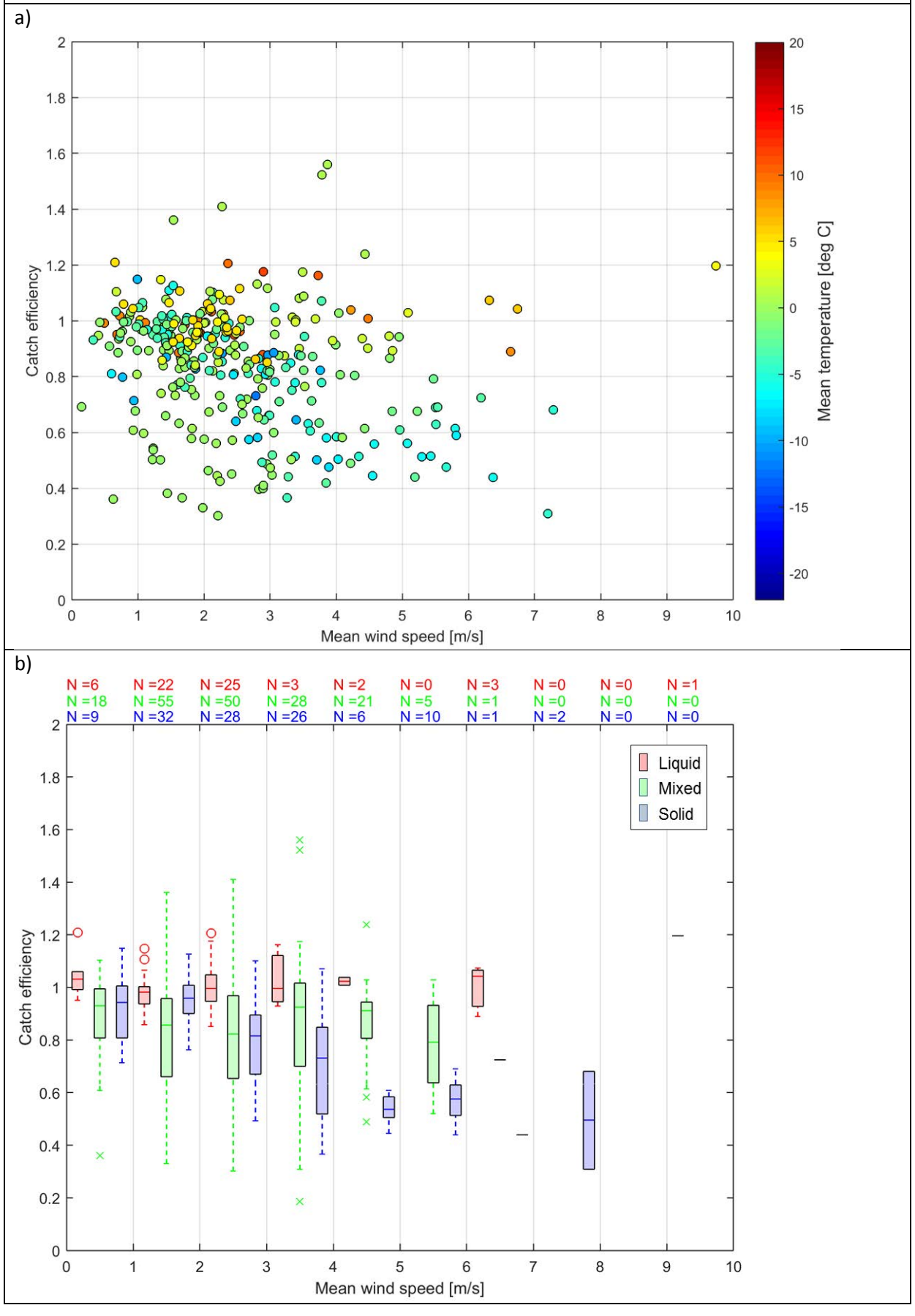

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
1
2
3
4
5
6

Figure 13: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the SA-shielded Geonor T-200B3MD gauge under test at Marshall (West sensor).



c)

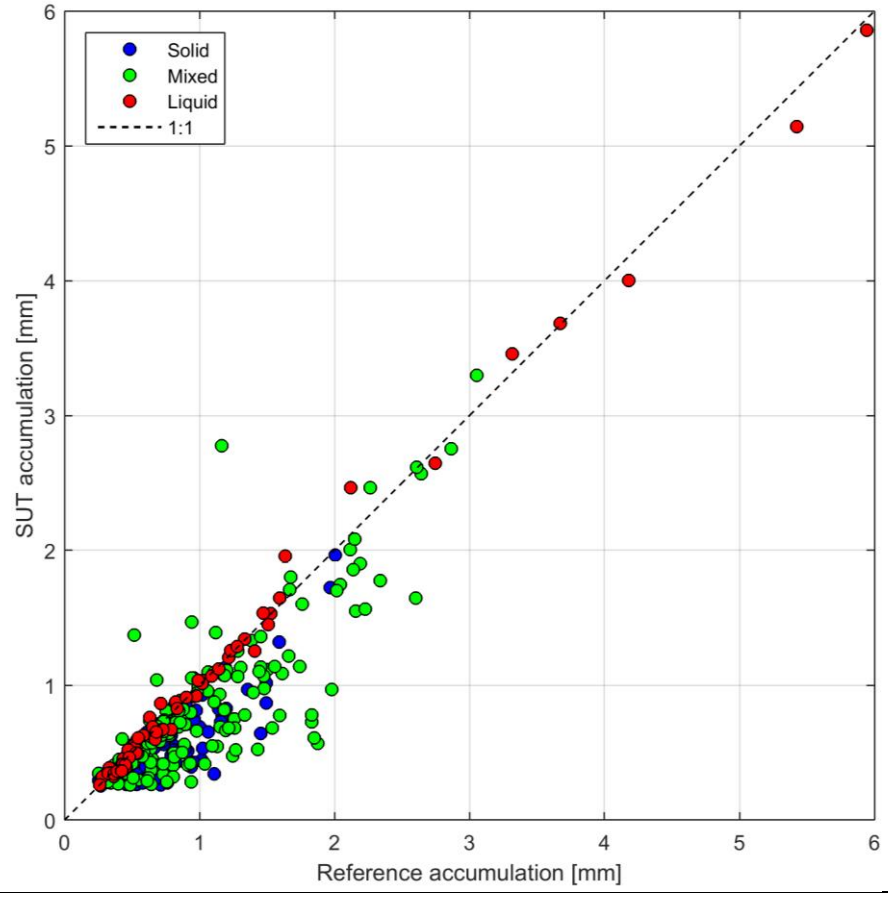
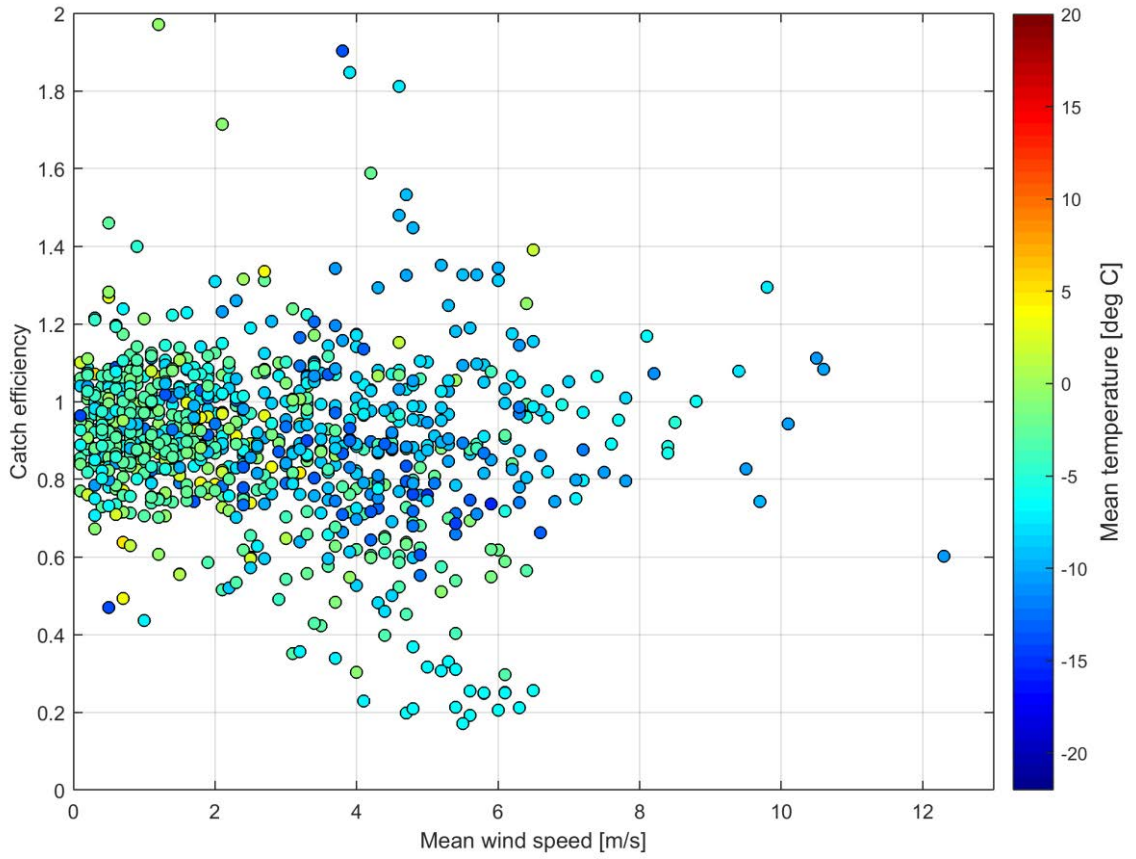
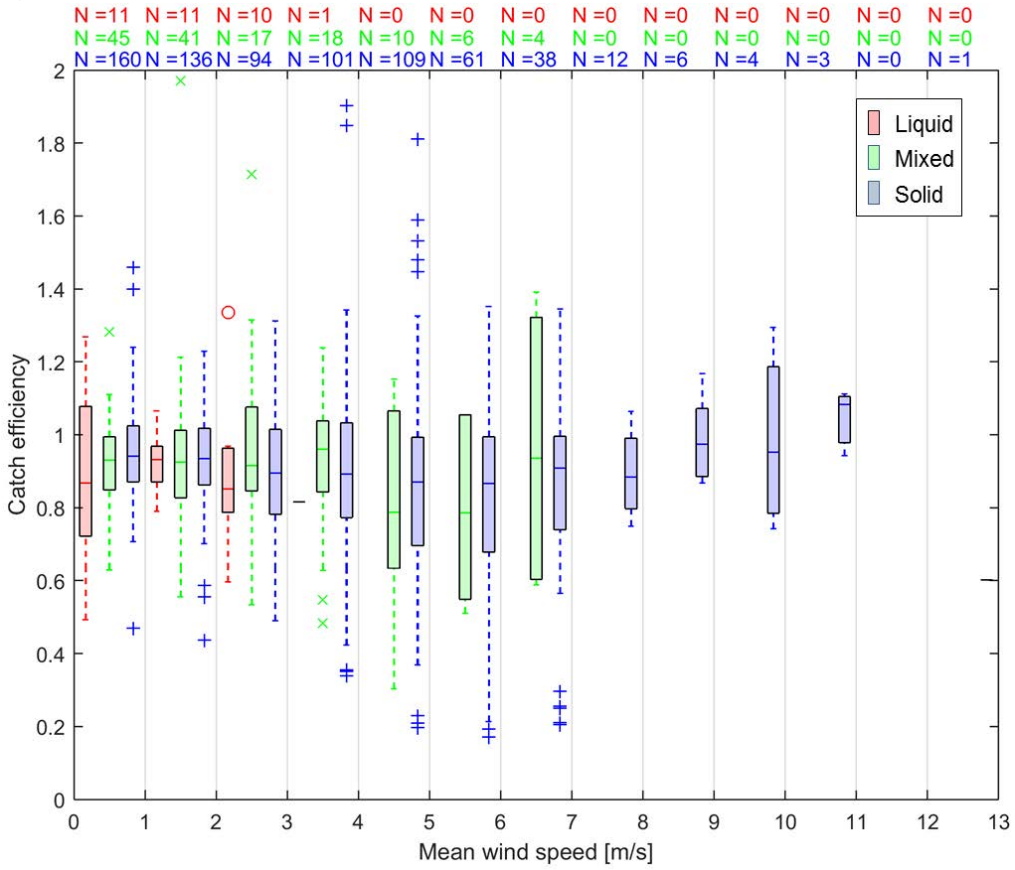

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
1
2
3
4
5
6

Figure 14: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the SA-shielded Geonor T-200B3MD gauge under test at Weissfluhjoch.

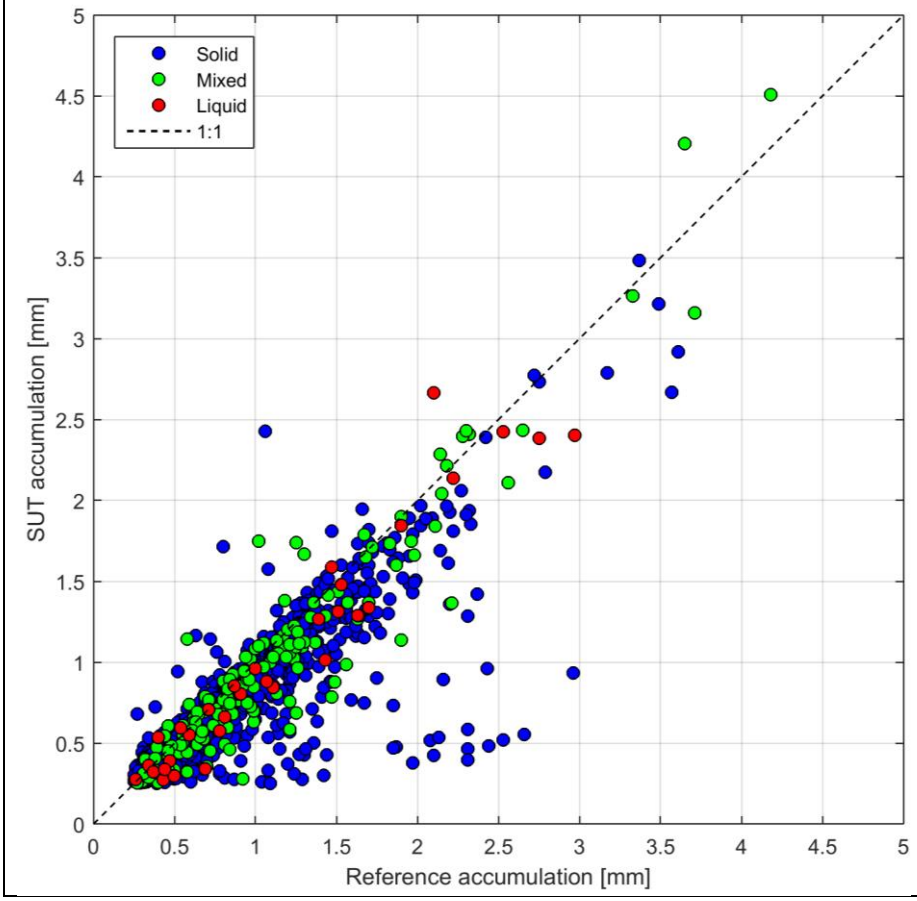
a)



b)



c)


Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5

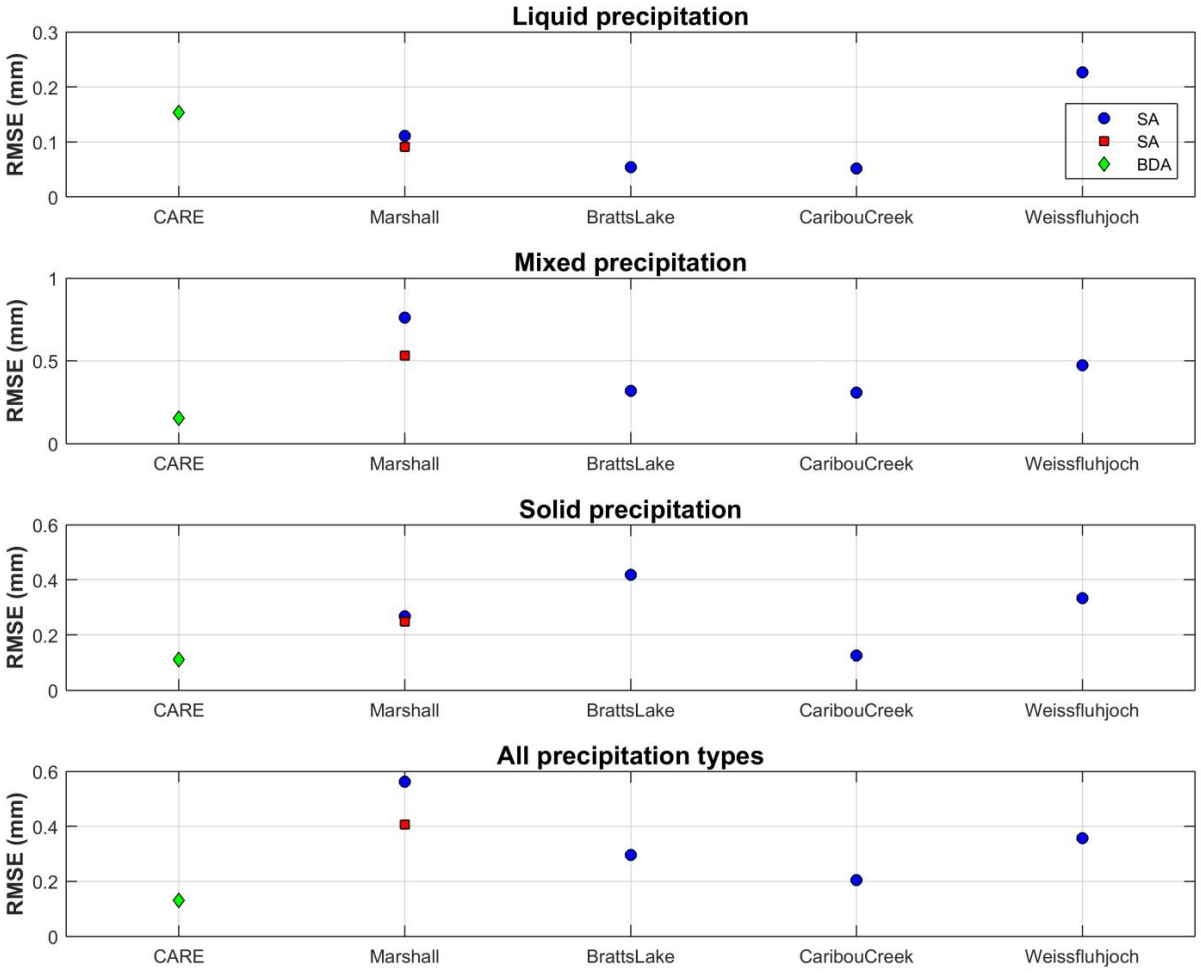


Figure 15: RMSE values calculated for test configurations by site and by precipitation type for YY cases over the duration of the test period.

Table 6: RMSE values calculated for test configurations by site and by precipitation type for YY cases over the duration of the test period, as presented in Figure 15.

Site	RMSE (mm/30 min)			
	Liquid	Mixed	Solid	All precip types
CARE	0.154	0.154	0.111	0.130
Bratt's Lake	0.054	0.319	0.419	0.296
Caribou Creek	0.052	0.308	0.125	0.205
Marshall (East)	0.111	0.762	0.268	0.563
Marshall (West)	0.091	0.531	0.249	0.407
Weissfluhjoch	0.227	0.474	0.334	0.357

The overall catch ratio calculated using all 30 minute YY cases, over the entire test period, is provided in Table 7. To demonstrate the influence of the SUT accumulation threshold on the results, the overall catch ratio is also provided for all 30 minute YY cases determined using a lower SUT threshold of 0.1 mm/30 minutes. Note that these values reflect only the YY cases, and do not include the amounts corresponding to the cases when the SUT and the reference do not agree on the occurrence of precipitation.

Table 7: Overall catch ratios for test configurations, determined from YY cases over the entire test period at each site, using two different SUT accumulation thresholds.

Site (sensor location)	Wind shield configuration	SUT accumulation threshold (mm/30 min)	Overall catch ratio
CARE	BDA	0.25	0.96
		0.1	0.94
Bratt's Lake	SA	0.25	0.80
		0.1	0.70
Caribou Creek	SA	0.25	1.06
		0.1	1.00
Marshall (East)	SA	0.25	0.83
		0.1	0.81
Marshall (West)	SA	0.25	0.85
		0.1	0.84
Weissfluhjoch	SA	0.25	0.87
		0.1	0.86

6.4. Ability to detect light precipitation events

The impact of the threshold selection for data processing relative to the detection of light precipitation was examined using four different combinations of reference and SUT accumulation thresholds (four 'cases' in Table 8) for a single-Alter-shielded Geonor T-200B3MD gauge under test at the Marshall site (East sensor). Contingency results, probabilities of detection (POD), and false alarm rates (FAR) are presented for each case in Table 9. A quantitative comparison of the amounts reported in each case is beyond the scope of this assessment.

Table 8: Reference and SUT thresholds in each case for light precipitation detection assessment.

Case	Reference threshold (mm/30 min)	SUT threshold (mm/30 min)
1	0.25	0.25
2	0.1	0.1
3	0.25	No threshold
4	0.25	0

Table 9: Contingency results, Probability of Detection (POD), and False Alarm Rate (FAR) for each case in light precipitation detection assessment.

Case	Number of events				Skill score (%)	
	YY	YN	NY	NN	POD	FAR
1	365	34	80	19434	91.5	18.0
2	436	3	419	19055	99.3	49.0
3	399	0	19514	0	100	98.0
4	399	0	11439	8075	100	96.6

6.5. Assessment of events when the reference and the SUT do not agree on the occurrence of precipitation

Assessment intervals during which the site reference and SUT do not agree on the occurrence of precipitation – namely, the YN and NY cases (Section 4.1.1) – are characterized for the BDA-shielded test gauge at CARE and the SA-shielded test gauges at Caribou Creek and Marshall (East sensor) using histograms in Figures 16, 17, and 18, respectively. The histograms include accumulated precipitation reported by the reference and SUT (0.25 mm/30 min threshold for both), precipitation intensity as reported by the reference, and corresponding site conditions.

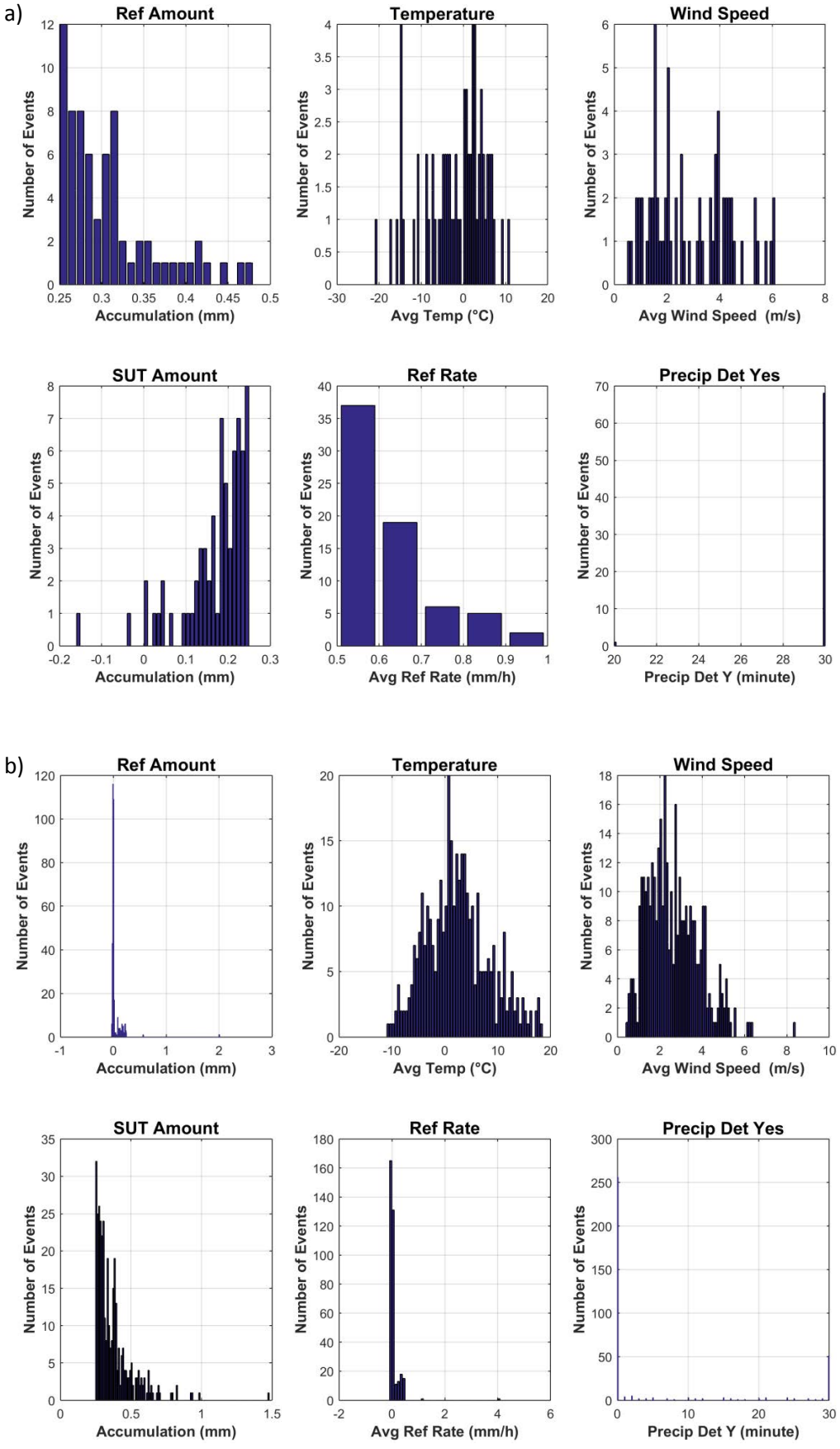


Figure 16: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for (a)YN events (69 total) and (b) NY events (355 total) for the BDA-shielded Geonor T-200B3MD gauge at CARE over the test period.

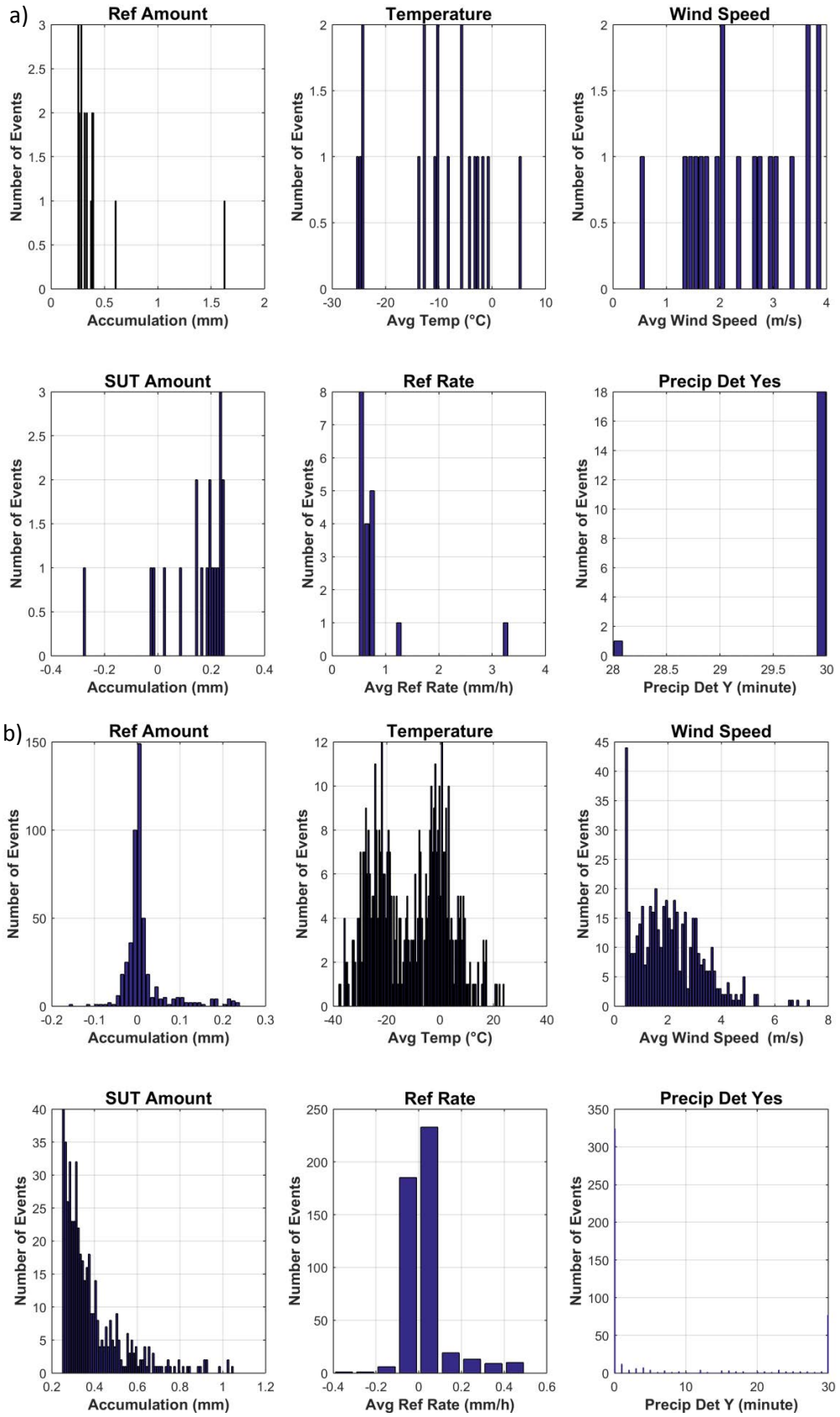


Figure 17: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for (a)YN events (19 total) and (b) NY events (477 total) for the SA-shielded Geonor T-200B3MD gauge at Caribou Creek over the test period.

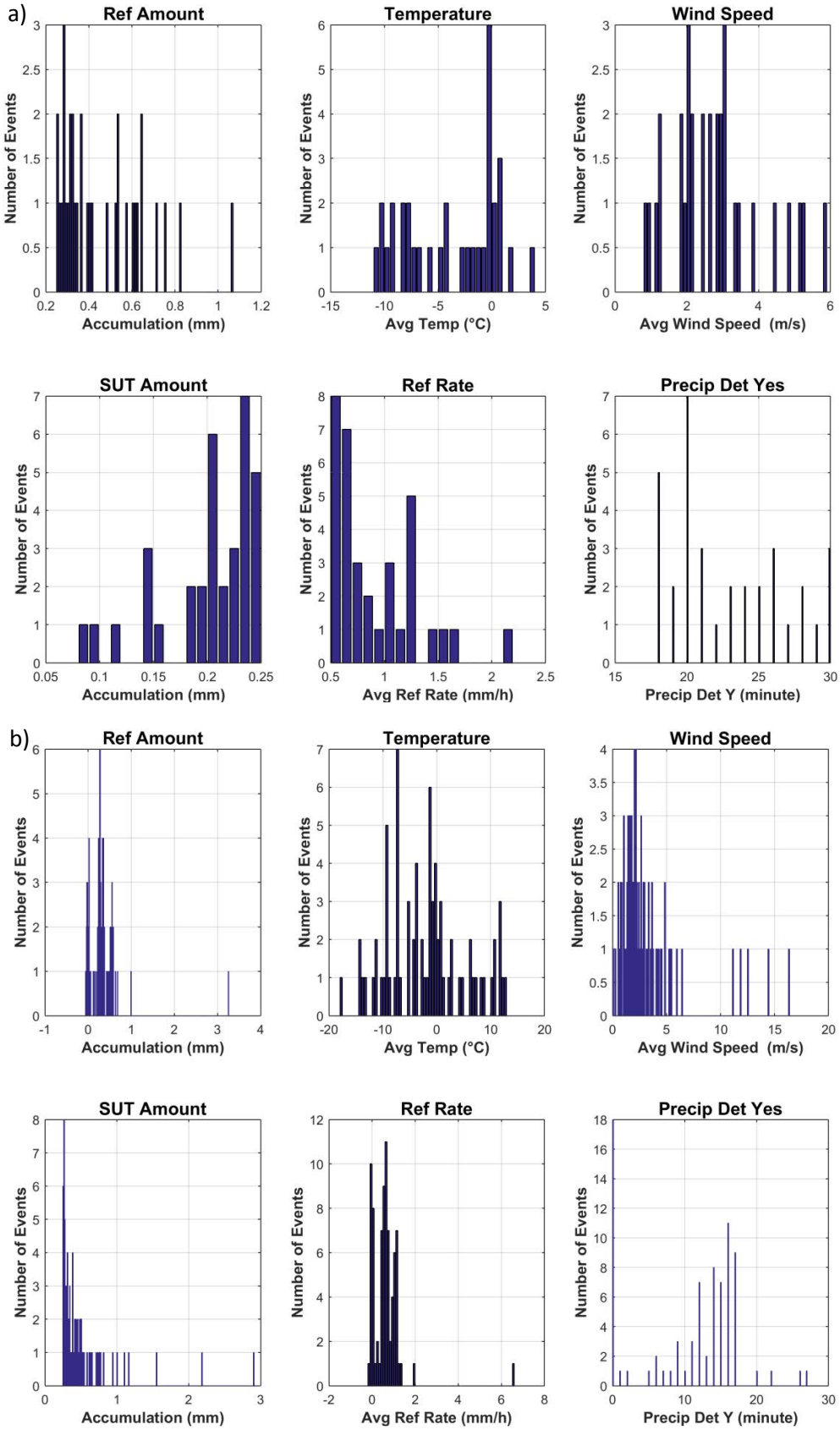


Figure 18: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for (a)YN events (34 total) and (b) NY events (80 total) for the SA-shielded Geonor T-200B3MD gauge at Marshall (East sensor) over the test period.

7) Interpretation of Results

7.1. Operating conditions

The full ranges of conditions under which the test gauges at each site were operated are illustrated in Figure 3. The conditions relevant to gauge operation are as follows:

- Temperatures between approximately -40 °C and 26 °C;
- Precipitation intensity within approximately 12 mm/hr.

These conditions fall within the manufacturer's specified operating conditions of -40 °C to 60 °C for temperature.

The conditions at each site during precipitation events are shown in Figure 4. Of particular note are the mean wind speeds during precipitation events at Bratt's Lake, which extend to approximately 15 m/s; the mean wind speeds during precipitation at all other sites fall within 10 m/s, and are generally within about 7 to 8 m/s.

7.2. Performance over the range of operating conditions

7.2.1. Non-precipitating conditions

The results presented in Figures 5 to 7 and Table 4 show variability in the test gauge responses in the absence of precipitation from site to site. All test gauges show enhanced variability relative to the corresponding reference configurations, which will impact their abilities to detect and report precipitation relative to the reference. The probability density function (PDF) for the test gauge at Caribou Creek in Figure 5 is broader relative to those for the other test gauges and sites. This is attributed to a ground noise issue experienced at the site, and underscores the critical influence of test gauge configuration on gauge performance. The variability of gauge responses in non-precipitating conditions do not show clear temperature or wind speed dependency (Figures 6 and 7). Test gauges at all sites show scatter in gauge responses for near-constant temperatures and at all wind speeds.

The standard deviation of gauge responses during 30 minute non-precipitating periods (Table 4) varied as follows: Bratt's Lake, 0.028 mm; CARE, 0.125 mm; Caribou Creek, 0.199 mm; Marshall, 0.039 mm and 0.054 mm for the East and West gauges, respectively; and Weissfluhjoch, 0.061 mm. Values as high as 2.70 mm were reported, indicating the potential for false alarms in terms of the detection of precipitation, as gauge reports in the absence of precipitation could exceed the detection threshold of 0.25 mm.

The magnitude of gauge responses in the absence of precipitation can be used to identify a detection threshold that minimizes the detection of false precipitation while enhancing the detection of light precipitation. This threshold is considered to be three times the standard deviation of the average gauge response during 30 minute non-precipitating periods. Based on the present test gauge results (excluding the test gauge at Caribou Creek, due to the noted ground noise issue), this minimum

detection threshold is determined to be 0.14 mm for the single-Alter shielded gauges and 0.38 mm for the Belfort double-Alter shielded gauge.

7.2.2. Precipitating conditions

7.2.2.1. Ability to detect and report precipitation

The skill scores for Geonor T-200B3MD test gauges (Figure 8) vary by site, with Probability of Detection values greater than 80% for the single-Alter shielded configurations Caribou Creek, Marshall, and Weissfluhjoch and the Belfort double-Alter shielded configuration at CARE. The test gauge at Bratt's Lake shows a notably lower POD (approximately 50%) and False Alarm Rate (approximately 4%), which may be related to the higher wind speeds experienced during precipitation events relative to the other sites (Figure 4). The test gauge at Caribou Creek shows a high FAR (83%) and Bias (495%), indicating a significant number of false reports relative to the reference; this is corroborated by the contingency results in Table 5, and attributed to the influence of the noted ground noise issue at the site. The resulting Heidke Skill Score for the test gauge at Caribou Creek is 26%, indicating less detection skill relative to the test gauges at other sites, which all exceed 65%.

The results for the test configuration at CARE, which employed a Belfort double-Alter shield, are similar to those observed for the test sites employing single-Alter shields (with the exception of Caribou Creek). This similarity is attributed to higher-than-expected noise in the gauge output noted by the site team, which are believed to have resulted from interference with the heater circuits. This is reflected in the high number of false alarm events detected for this test configuration (Table 5), and provides another example of the impact of gauge configuration on performance.

7.2.2.2. Ability to report accumulated precipitation

The results presented in Figures 9 to 14, which are based on 30 minute events during which the reference and test gauge both detect precipitation (YY cases), illustrate the influence of wind speed and precipitation type on gauge catch efficiency. The discussion below will focus on snow events; the number of rain events during winter is limited, and the results for mixed events are variable due to the variability in the size and density of precipitation within the mixed regime, as well as the potential for transitions between phases. The specific trends observed vary by site and by configuration. For the single-Alter shielded test gauges at Bratt's Lake and Marshall, the median catch efficiency for solid precipitation decreases with increasing mean wind speed, with values between 0.4 and 0.6 for speeds up to 6 m/s. The events at cold temperatures and mean wind speeds above 8 m/s at Bratt's Lake are believed to be blowing snow, based on observer reports from a nearby airport. The influence of a second shield is apparent in the data for the Belfort double-Alter shielded test gauge at CARE, which show a more gradual decrease with increasing mean wind speed, with median catch efficiencies of approximately 0.8 for mean wind speeds up to 6 m/s.

Wind speed trends for solid precipitation catch efficiency are less evident for the single-Alter shielded test gauges at Caribou Creek and Weissfluhjoch. At Caribou Creek, catch efficiency values greater than 1 are observed over the full range of wind speeds, which are attributed to noise resulting from the specific site configuration. At Weissfluhjoch, there are a large number of events

reported, and the median catch efficiency remains within about 0.8 to 1 for wind speeds up to 10 m/s. The reason for this consistency is unclear; however, it may be related to the specific flow field experienced by the test gauge within a site characterized by complex topography.

Root mean square error values were computed from all 30 minute events during which each test configuration and the corresponding reference configuration both detected precipitation. Values are shown in Figure 15 and Table 6, and can be considered to represent the absolute uncertainty of each test configuration relative to the reference configuration in liquid, mixed, and solid precipitation, and in all precipitation types (note that the relative proportions of events of each phase differ by site). The RMSE for single-Alter shielded gauges in all precipitation types ranges from 0.205 mm/30 min for the test gauge at Caribou Creek to 0.563 mm/30 min for the East test gauge at Marshall. The RMSE for the double-Alter shielded test gauge at CARE is 0.130 mm/30 min.

The overall catch ratio – computed from the total reference and SUT accumulation from all YY cases over the duration of formal tests – is provided for each test configuration in Table 7. For single-Alter shielded test configurations, the overall catch ratio ranges from 0.80 (Bratt's Lake) to 0.87 (Weissfluhjoch), with the exception of the test gauge at Caribou Creek, which is influenced by noise related to the configuration and has an overall catch efficiency of 1.06. The double-Alter shielded test gauge at CARE has an overall catch ratio of 0.96. Decreasing the SUT accumulation threshold for precipitation events from 0.25 mm/30 min to 0.1 mm/min does not impact the overall catch ratio significantly; the largest change observed is a decrease from 0.8 to 0.7 for the test gauge at Bratt's Lake.

7.2.2.3. Ability to detect light precipitation

The detection thresholds and results presented in Tables 8 and 9, respectively, indicate that decreasing or removing the SUT detection threshold (while maintaining the reference detection threshold at 0.25 mm) leads to increases in the Probability of Detection, but also increases the False Alarm Rate to close to 100%. Decreasing both the reference and SUT detection thresholds to 0.1 mm increases the POD, while increasing the FAR to almost 50%. These results suggest that gauge noise mitigates the ability of Geonor T-200B3MD gauges to detect light precipitation, as any reduction in the detection threshold intended to capture light precipitation is accompanied by a corresponding increase in false reports, as indicated by the false alarm rate.

7.2.3.4. Assessment of events when the reference and SUT do not agree on the occurrence of precipitation

The YN ('miss') cases, when the reference detects a precipitation event and the SUT does not, and NY ('false alarm') cases, when the SUT detects a precipitation event and the reference does not, are characterized for selected test configurations in Figures 16 to 18. The majority of the YN cases have reference accumulations just above the 0.25 mm threshold, and SUT accumulations just below the threshold. This difference is likely the result of enhanced wind effects for the SUT relative to the reference gauge in the DFIR-fence, which can reduce the SUT accumulation below the detection threshold. The number of false alarm events is comparatively higher for each of the test gauges considered, which is attributed to gauge noise increasing the accumulation above the detection event threshold, while the reference accumulation remains below the threshold.

8) Maintenance

Gauge calibration: each site completed the gauge field calibration and verification as per manufacturer recommendations, at least once a year or following the emptying of the gauge. The calibration records have been stored by each site host.

9) Performance Considerations

9.1. Data acquisition and processing

The Geonor T-200B3MD must be operated in conjunction with a data logger that converts the frequency signal into a precipitation amount. Additional processing is required to translate the gauge output into meteorologically-representative data. This requires an understanding of gauge performance, and generally, a well-educated user. The manufacturer is encouraged to propose an algorithm for deriving baseline datasets from the gauge output to allow for more consistency in the data derived by different users.

The SPICE data quality control and event selection procedures have been applied to filter accumulation data from Geonor T-200B3MD gauges and derive precipitation amounts over specified intervals. Such procedures are recommended to reduce the variability in gauge reports and establish analysis-ready data products. The specific threshold employed for distinguishing between precipitation and signal noise is an important consideration; as the threshold is decreased, there is higher likelihood of gauge noise producing false reports.

9.2. Gauge configuration

The configuration of Geonor T-200B3MD gauges at a given site can impact significantly the gauge performance. All field configurations must be fully tested and validated prior to use, in order to ensure that physical and signal interferences are not degrading the gauge signal. This would include the confirmation of grounding tailored to the soil conditions, sturdiness of foundation and mounting, and signal conditioning.

One aspect of the shield configuration that warrants consideration is that the Belfort double-Alter shield slat configuration employs springs and rubber gaskets that may require maintenance. This is a more significant concern for remote or unattended sites, as gauge performance may be impacted by periods when the slats are not functioning properly, or have fallen off. No such issues were encountered for the Belfort double-Alter shielded test gauge at CARE during SPICE.

When heating is applied, testing of the heating circuits must be performed, including testing their impact to the sensor signal. Given the intermittence of the supply of power to the heating elements, the risk of interference needs to be minimized. It is recommended that the Geonor T-200B3MD gauges are fitted with heating circuits integrated and validated by the manufacturer, in order to provide a consistent and proven solution. In its absence, a user without experience in configuring the heating circuits could implement circuits that would negatively affect the quality of the sensor output.

As unheated and heated gauges were not tested at the same site (with the same environmental conditions and general configuration), it is difficult to comment on the impact of heating on

measurements from Geonor T-200B3MD gauges, and to make any related recommendations. Users are encouraged to consult Section 4.2.2 of the WMO-SPICE Final Report for guidance and best practices related to heating.

9.3. Ancillary measurements and adjustments

The application of adjustment functions is strongly recommended to account for the reduction in gauge catch efficiency as the wind speed increases. Ancillary measurements of wind speed (preferably at gauge orifice height) and air temperature are required for the application of adjustment functions. Ancillary measurements from a sensitive precipitation detector that is independent from the test gauge are also recommended to help distinguish precipitation events from false reports due to noise.

WMO-SPICE Instrument Performance Report
MeteoServis MRW500

1) Technical Specifications (from manufacturer provided documentation)

Instrument model:	MeteoServis MRW500
Physical principle:	Weighing gauge (WG) based on catchment principle; weight measurement by strain gauges connected to internal control electronics.
Capacity:	1800 mm without using antifreeze liquid; 900 mm with antifreeze liquid.
Collecting area :	500 cm^2 (circular orifice)
Operating temperature range:	-35 °C to 60 °C
Measurement uncertainty:	0.2 mm for precipitation intensity (as defined by manufacturer)
Sensitivity (resolution):	0.1 mm (as defined by manufacturer)
Heater:	Embedded rim heater (continuous and shock heating), 350 W
Additional components:	The gauge consists of two containers, the catchment vessel and storage (accumulation) vessel. The fluid is transferred between them via pumps. Two rain detectors (for redundancy) are integrated; their data are used in the derivation of gauge data products.

Figure 1: (a) Unshielded and (b) single-Alter/Tret-shielded MRW500 gauges under test at Marshall; (c) unshielded and (d) single-Alter/Tret-shielded MRW500 gauges under test at Bratt's Lake.

2) Data Output Format

Gauge data output: Internal data processing, using a proprietary algorithm, is applied to the raw data from the strain gauge. This processing accounts for temperature effects and integrates precipitation detector data to generate the following outputs.

Instant value (Iv): the instantaneous measured weight of the catchment vessel, accounting for weight changes caused by pump operation. The delay in reporting the Iv after a weight change is up to 10 s.

Average value (Av): the average of the instantaneous values, usually over one minute.

Precipitation operating value (Ov): the output value of weight from the rain gauge, which shows the total accumulated precipitation from the start of collection. The reaction time is one minute. The value is modified by a gauge calibration factor. The value *Ov* should be used as the total precipitation caught by the rain gauge.

3) SPICE test configuration

Shield(s): A shield is provided by the manufacturer, and is specific to this gauge, with a configuration similar to a single-Alter (SA) shield, but with fixed slats similar to a Tretyakov (Tret) shield. The gauges tested were in both these SA/Tret shields and in unshielded (UN) configurations.

Test site(s): Marshall (USA); Bratt's Lake (Canada)

Sensor provider(s) All gauges tested were provided by Meteoservis v.o.s (Czech Republic)

A map of test site locations is provided in Figure 2.

3.1. Note on terms and acronyms used

Throughout this document, the following notations are used to identify the R2 reference (Ref) and senor under test (SUT) configurations:

Reference: 'DFIR' and 'DFAR' are used interchangeably for the R2 reference configurations. 'DFIR' refers to an automated gauge installed in a DFIR-fence, while 'DFAR' refers more explicitly to the Double-Fence Automated Reference configuration.

Sensors under test: 'SA/Tret' denotes single-Alter/Tretyakov shielded test configurations, and 'UN' denotes unshielded test configurations.

Figure 2: Map of SPICE sites where MeteoServis MRW500 gauges were tested.

A summary of the configuration of instruments as tested, the duration of tests and availability of data reflected in these results, and the ancillary measurements used, by site, is available in Tables 1, 2, and 3, respectively.

Table 1: Summary of gauge configurations and data output, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	Marshall	**Bratt's Lake**
Field configuration (shield)	Unshielded (UN) and single shield (SA/Tret)	
Height of installation (gauge rim)	192 cm	222 cm
Heating	Manufacturer recommended algorithm with rim heating 105 W	
Antifreeze	Manufacturer-provided antifreeze (Fridex EKO, based on propylene glycol), rated to -32 °C	Manufacturer-provided antifreeze (Fridex EKO, based on propylene glycol), rated to -32 °C (Season 1) Site-provided plumbing antifreeze (propylene-glycol based, non-toxic), rated to -50 °C (Season 2)†
Oil	Manufacturer-provided oil (LUKOSIOL M100, methyl-silicone oil)	
SUT data output frequency	1 min	1 min

Data QC	SPICE QC procedures
Data temporal resolution	1 min
Processing interval for SPICE data analysis	30 min

† Temperature conditions below the stated range of the manufacturer-provided antifreeze were experienced at Bratt's Lake during Season 1 (see Figure 3, below), an off-the-shelf solution suitable for temperatures as low as -50 °C was employed in Season 2.

Table 2: Data availability, by measurement season and site.

Measurement season	Marshall	Bratt's Lake
Season 1 (Oct. 2013 – Apr. 2014)	✓	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓	✓

Table 3: Summary of reference and ancillary measurements, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	Marshall	Bratt's Lake
R2 Site Reference	Geonor T-200B3 600 mm (DFAR)	Geonor T-200B3 600 mm (DFAR)
R2 Precip Detector	Thies LPM (Site*)	Thies Precip Sensor (DFAR)
Ancillary Temp Sensor	MetOne, 060A-2/062, 2144-L (2 m)	Vaisala HMP45 (2m)
Ancillary RH Sensor	Campbell Scientific CS500 (2m)	Vaisala HMP45 (2m)
Ancillary Wind Sensor	RM Young Wind Monitor 05103 (2 m)	RM Young Wind Monitor 05103 (2m)

*A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR shield; however, in cases where a more sensitive detector is available outside of the DFIR shield, or there are issues with the detector within the DFIR shield, a precipitation detector elsewhere on the site can be employed.

4) Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6.1 of the WMO-SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

4.1.1.1. Characterization of performance in non-precipitating conditions

The assessment data are derived over 30 minute intervals during which the precipitation detector in the R2 reference configuration reports 0 minutes of precipitation. The accumulation over these intervals (accumulation in minute 30 – accumulation in minute 1), representing the variability of the gauge response due to wind, evaporation, temperature, etc., is recorded, along with the mean wind speed, and the change in temperature (temperature in minute 30 – temperature in minute 1).

4.1.1.2. Assessment of ability to detect and report accumulation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute assessment intervals:

- Ref 'Yes' : R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No' : R2 weighing gauge < 0.25 mm AND/OR precip detector recording < 18 min of precip;
- SUT 'Yes' : SUT accumulation ≥ 0.25 mm;
- SUT 'No' : SUT accumulation < 0.25 mm.

4.1.1.3. Assessment of ability to detect light precipitation

The data for this component of the assessment are derived in a similar manner as those in Section 4.1.1.2, but with different combinations of thresholds for the reference and/or SUT 'Yes' and 'No' conditions. These different threshold 'cases' have been selected to demonstrate the impact of the thresholds used in data derivation on the detection of light precipitation.

4.1.2. Skill score assessment

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores, as detailed in Section 3.6.1.3 of the WMO-SPICE Final Report.

For the assessments considered in this report, the ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

5) Environmental conditions

The environmental conditions at each site over the duration of the test period are expressed as probability density functions (PDFs) of mean air temperature, mean relative humidity, mean wind speed, vector mean wind direction, and precipitation rate for each component 30 minute assessment interval in Figure 3. The same parameters are also shown for all assessment intervals during which the site reference configuration detected precipitation (i.e. all Ref 'Yes' cases) in Figure 4.

The precipitation percentage represents the number of minutes of precipitation during a 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration, expressed as a percentage. PDFs of precipitation percentage are also included in Figures 3 and 4.

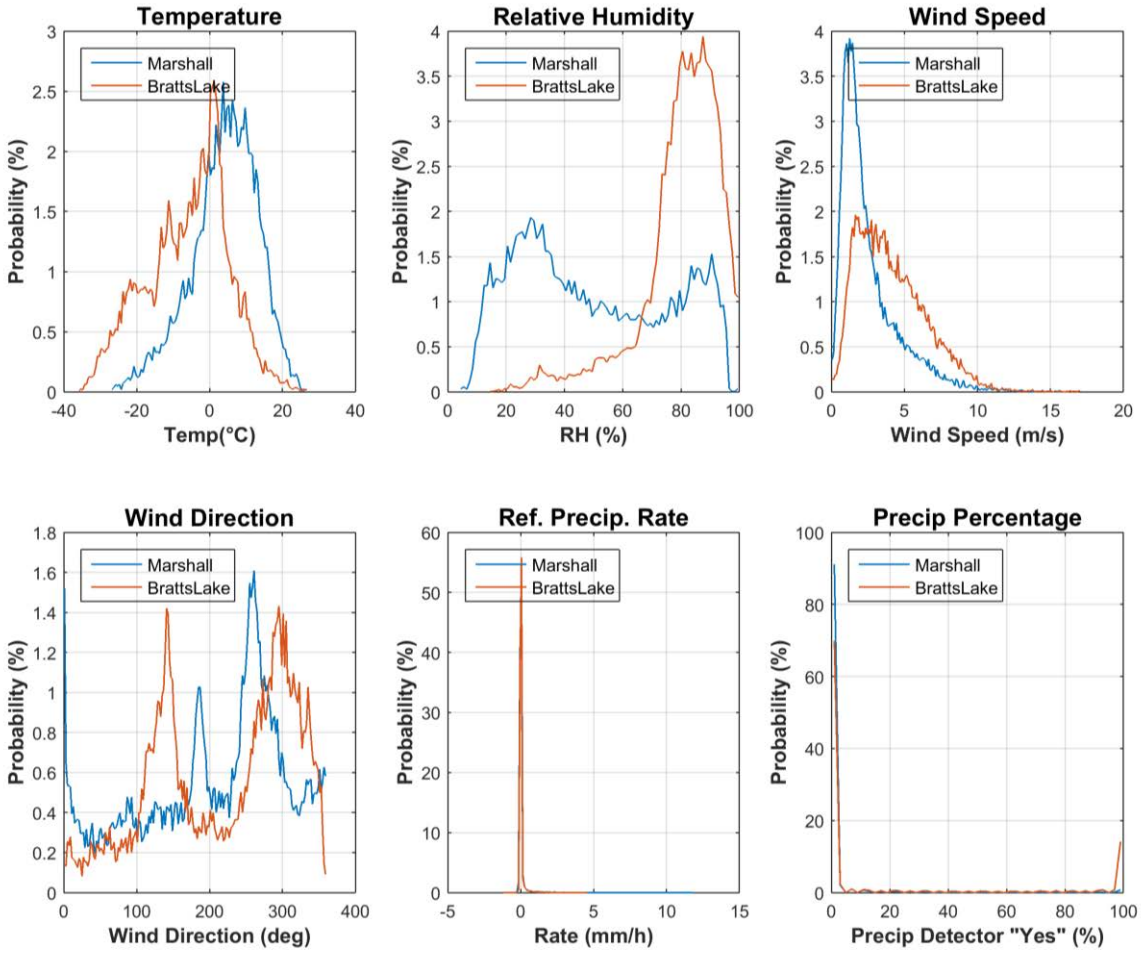


Figure 3: Summary of environmental conditions at Marshall and Bratt's Lake over the entire duration of formal tests, as per Table 2.

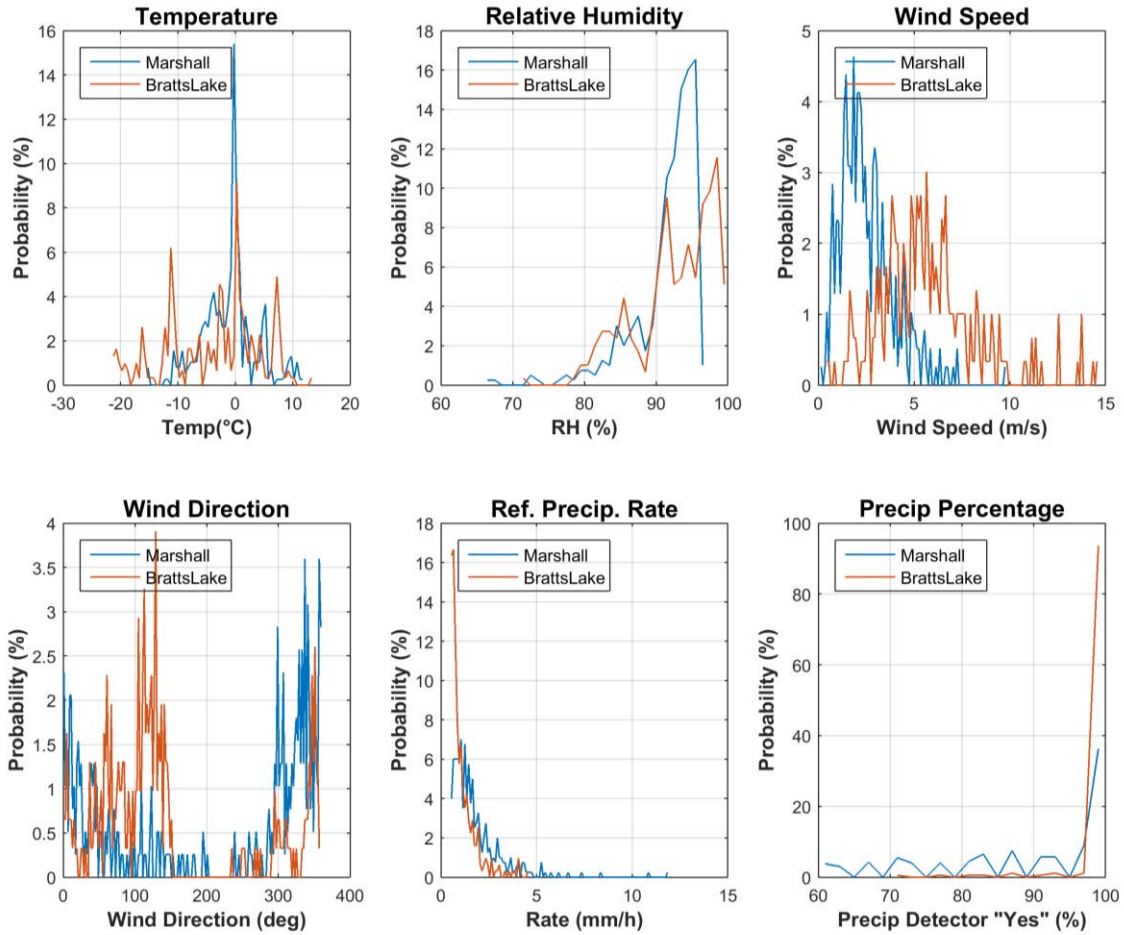


Figure 4: Summary of environmental conditions at Marshall and Bratt's Lake during precipitation events (as defined by the R2 reference), over the duration of formal tests.

6) Evaluation of the ability to perform over the range of operating conditions

6.1. Characterization of SUT performance in non-precipitating conditions

The response of the SUT in the absence of precipitation was examined as defined in Section 4.1.1.1. The results are presented below, reflecting the distribution of the sensor response and its variability with wind and temperature, as measured during 30 minute assessment intervals.

6.1.1. Overall variability of SUT response

The overall variability of the SUT response in non-precipitating conditions is shown as a probability density function for each test configuration in Figure 5. The corresponding PDF for the reference configuration is provided for comparison.

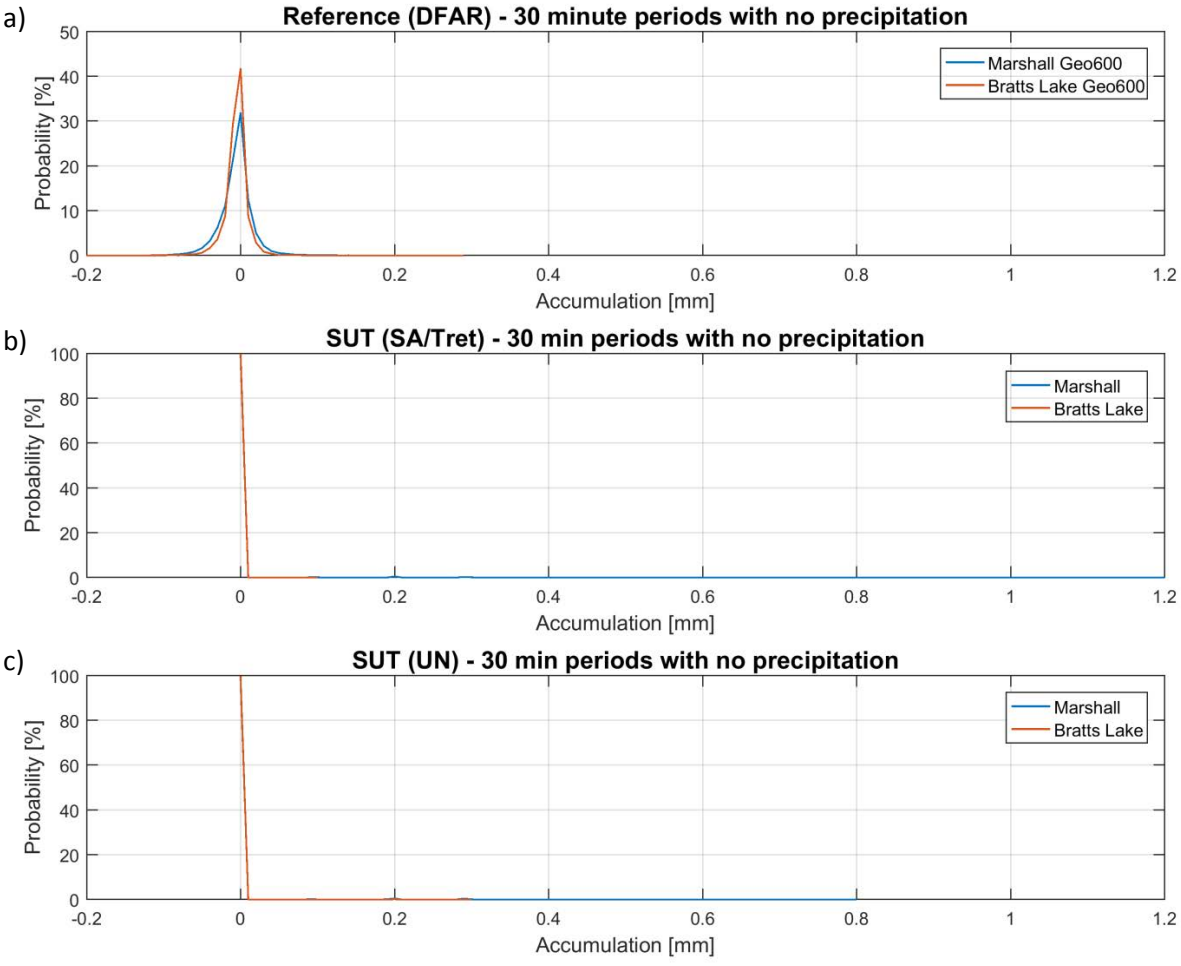


Figure 5: Probability density functions of the output signal (accumulation over 30 minute assessment intervals) in non-precipitating conditions for (a) the R2 reference configuration (DFAR), (b) MRW500 test gauges in SA/Tret configurations, and (c) MRW500 test gauges in unshielded configurations.

The statistics of the output signal (accumulation over 30 minute assessment intervals) for the reference and SUT at each site are provided in Table 4.

Table 4: Statistics of the R2 reference gauge and SUT output signal during non-precipitating conditions at Marshall and Bratt's Lake, as plotted in Figure 5.

Site	Gauge (shield)	Average output signal (mm)	Standard deviation (mm)	Maximum output signal (mm)	Minimum output signal (mm)	Number of assessment intervals
Marshall	DFAR	-0.001	0.022	0.196	-0.207	18186
	SUT (SA/Tret)	0.002	0.025	1.200	0.000	18186
	SUT (UN)	0.004	0.032	0.800	0.000	18175
Bratt's Lake	DFAR	-0.000	0.016	0.293	-0.194	7016
	SUT (SA/Tret)	0.000	0.002	0.100	0.000	7016
	SUT(UN)	0.000	0.011	0.300	0.000	5486

6.1.2. Variability of SUT response as a function of temperature

The variability of the SUT response for each test configuration in the absence of precipitation is plotted as function of the temperature difference over each assessment interval in Figure 6. The temperature difference is defined as the difference in temperature between the end (minute 30) and beginning (minute 1) of the assessment interval. The corresponding plots for the reference configurations are provided for comparison.

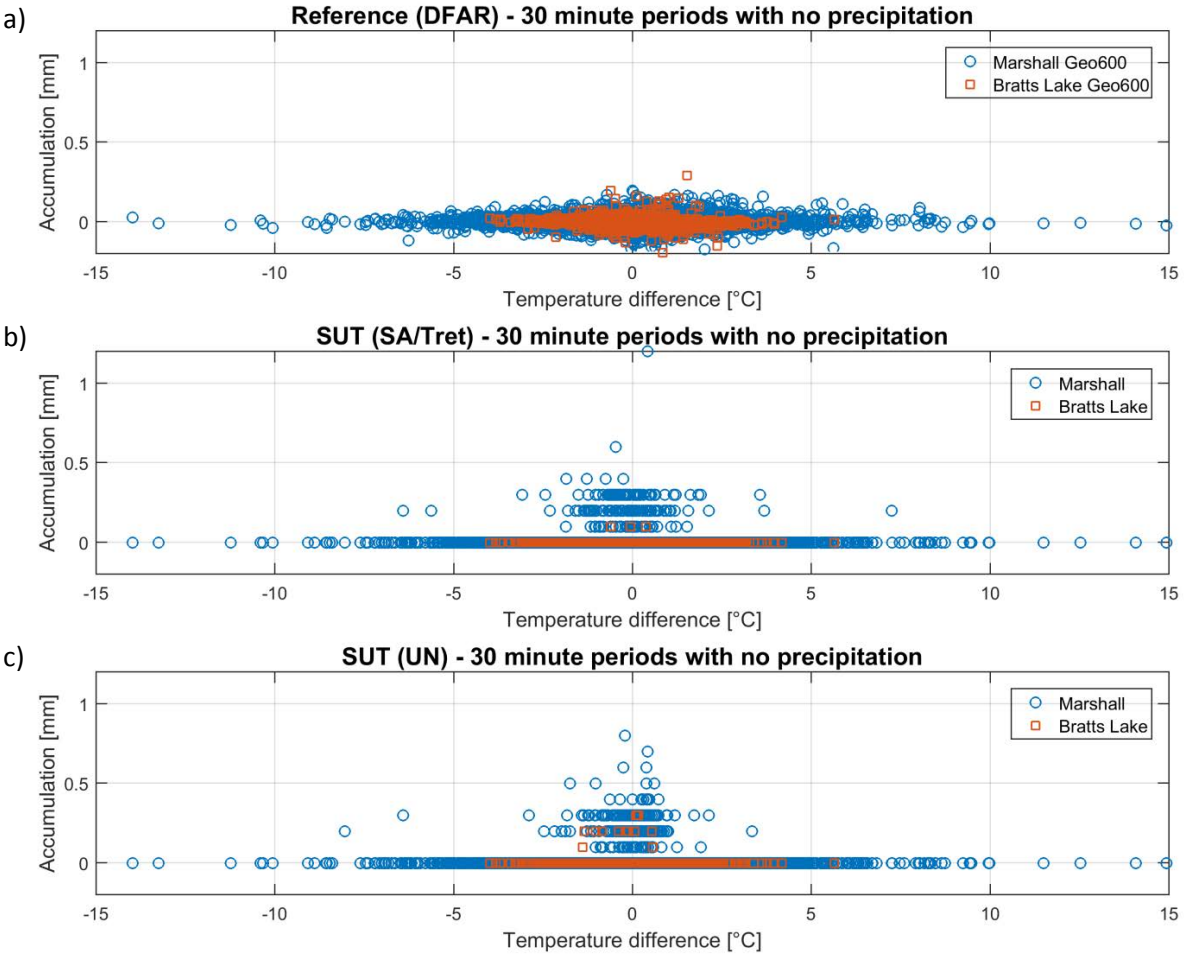


Figure 6: Variability of output signal (accumulation over each 30 minute assessment interval) as a function of the temperature difference over the interval in non-precipitating conditions for (a) the R2 reference configuration (DFAR), (b) MRW500 test gauges in SA/Tret configurations, and (c) MRW500 test gauges in unshielded configurations.

6.1.3. Variability of SUT response as a function of wind speed

The variability of the SUT response for each test configuration in the absence of precipitation is plotted as function of the mean wind speed for each assessment interval in Figure 7. Here, the signal variability is represented as the standard deviation (STD) of the gauge accumulation output over each 30 minute interval. The corresponding plots for the reference configurations are provided for comparison.

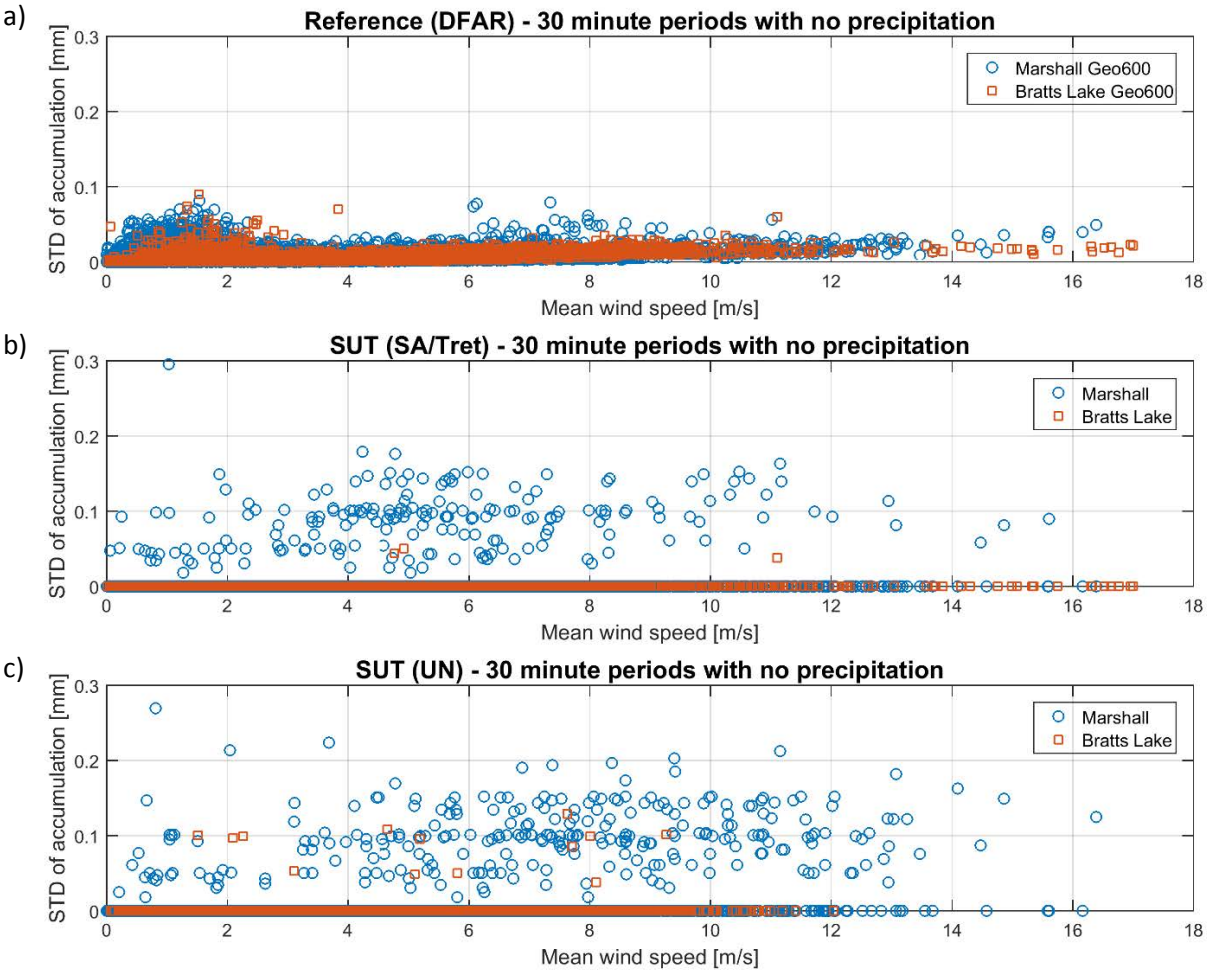


Figure 7: Variability of output signal (standard deviation of accumulation over each 30 minute assessment interval) as a function of mean wind speed in non-precipitating conditions for (a) the R2 reference configuration (DFAR), (b) MRW500 test gauges in SA/Tret configurations, and (c) MRW500 test gauges in unshielded configurations.

6.2. Ability to detect and report precipitation

6.2.1. Skill score assessment

The overall ability of the SUT to detect and report the occurrence of precipitation relative to the site field reference over 30 minute assessment intervals is expressed using selected skill scores (Section 4.1.2) and presented in Figure 8. The contingency results (Section 4.1.1) corresponding to these scores are presented in Table 5.

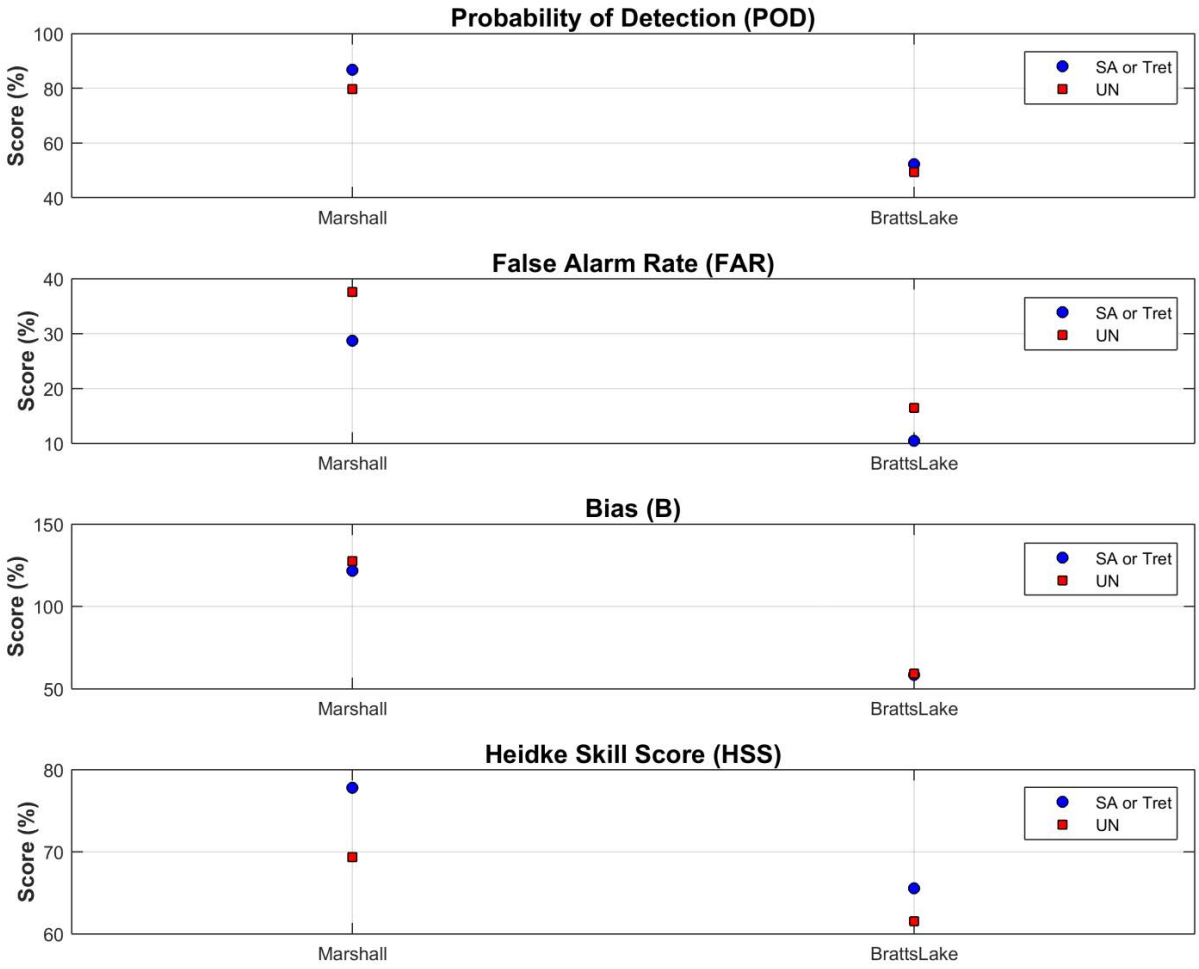


Figure 8: Skill scores for MRW500 test gauges over the duration of formal tests.

Table 5: Contingency table illustrating detection of precipitation by MRW500 gauges under test relative to the corresponding site reference configurations, expressed as number of events over the entire test period.

Site	Configuration	Number of Events			
		YY (hits)	YN (misses)	NY (false alarms)	NN (correct negatives)
Marshall	Unshielded	318	81	191	19375
	SA/Tret	346	53	139	19438
Bratt's Lake	Unshielded	127	130	25	12617
	SA/Tret	163	149	19	16307

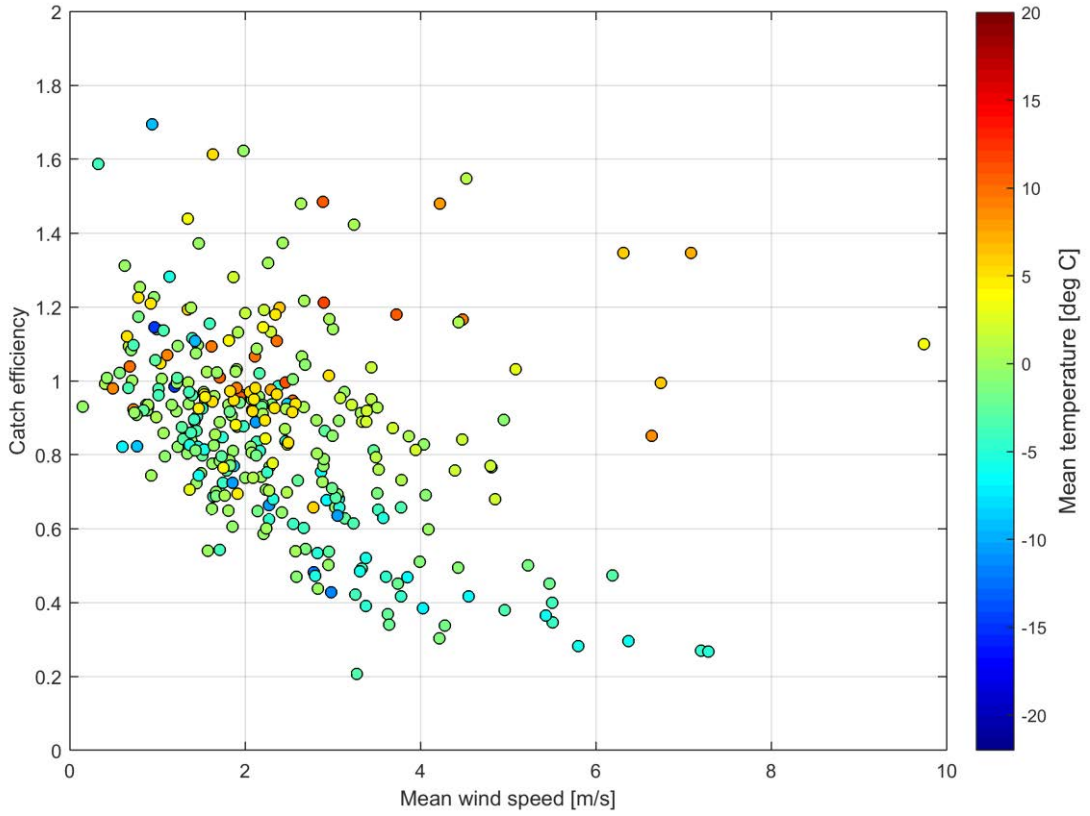
6.3. Ability to report accumulated precipitation

The SUT performance in terms of reporting accumulated precipitation is examined by comparing the amount reported by the sensor under test relative to the respective site reference during 30 minute assessment intervals. This is represented graphically using scatter and box and whisker plots of the catch efficiency as a function of mean wind speed at gauge height, as well as scatter plots of the amounts reported by the SUT versus the corresponding reference amounts (Figures 9 to 12). The SUT performance is also assessed in terms of the root mean square error, RMSE (Figure 13).

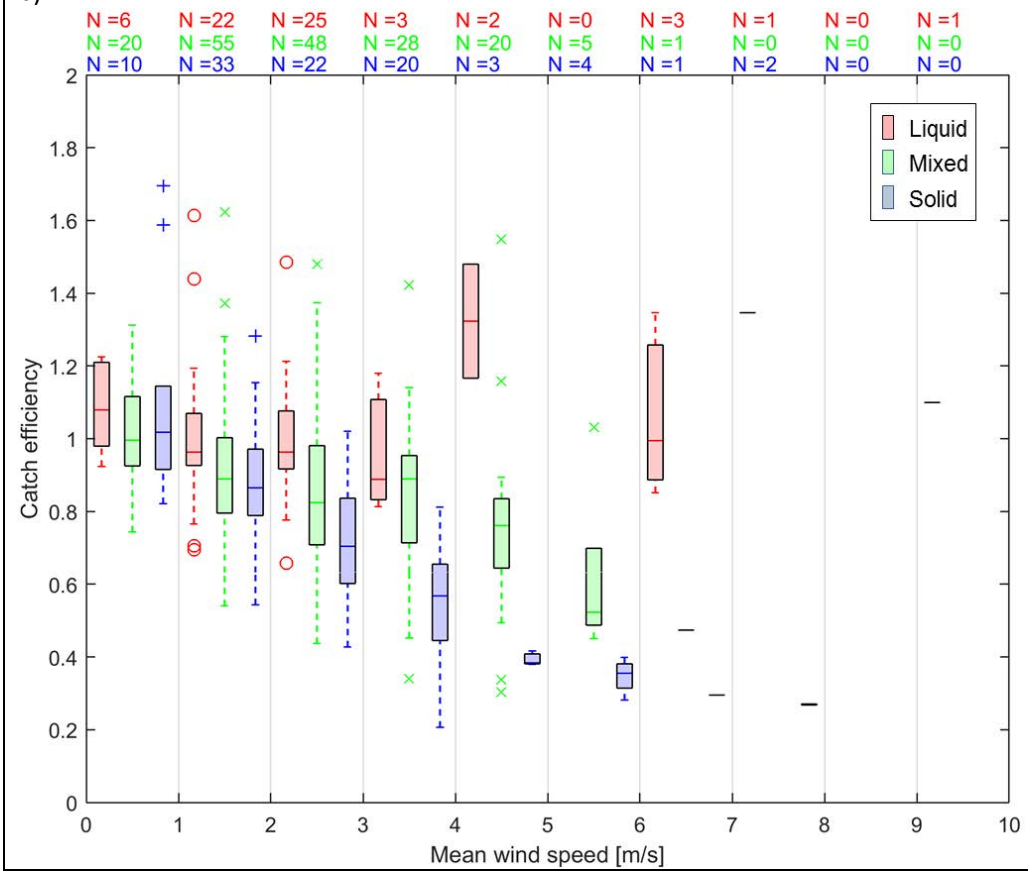
Only assessment intervals during which the SUT and reference both reported precipitation (YY cases) are considered in this portion of the assessment. In the catch efficiency-wind speed scatter plots, the mean event temperature is indicated by colour, with the colour scale selected to be consistent across all sites with weighing gauges under test. In the box and whisker plots and accumulation-accumulation scatter plots, the predominant precipitation type is indicated by colour, as determined from the reported temperature (Section 4.1.4).

Figure 9: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the SA/Tret-shielded MRW500 gauge under test at Marshall.

a)



b)



c)

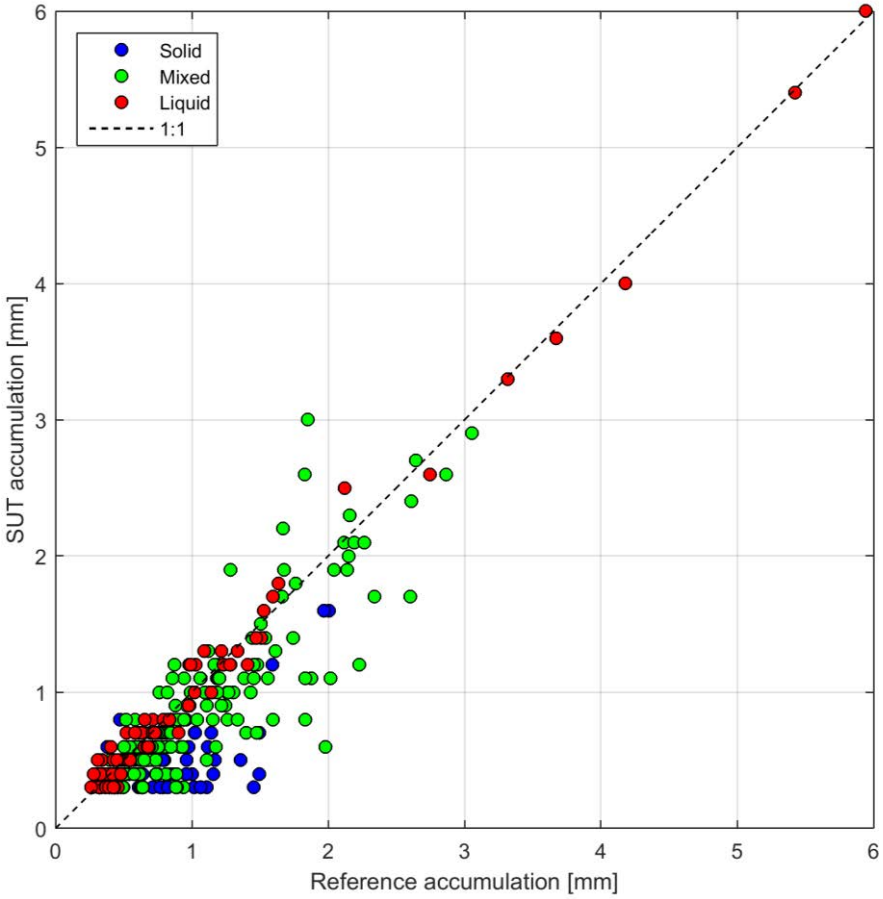
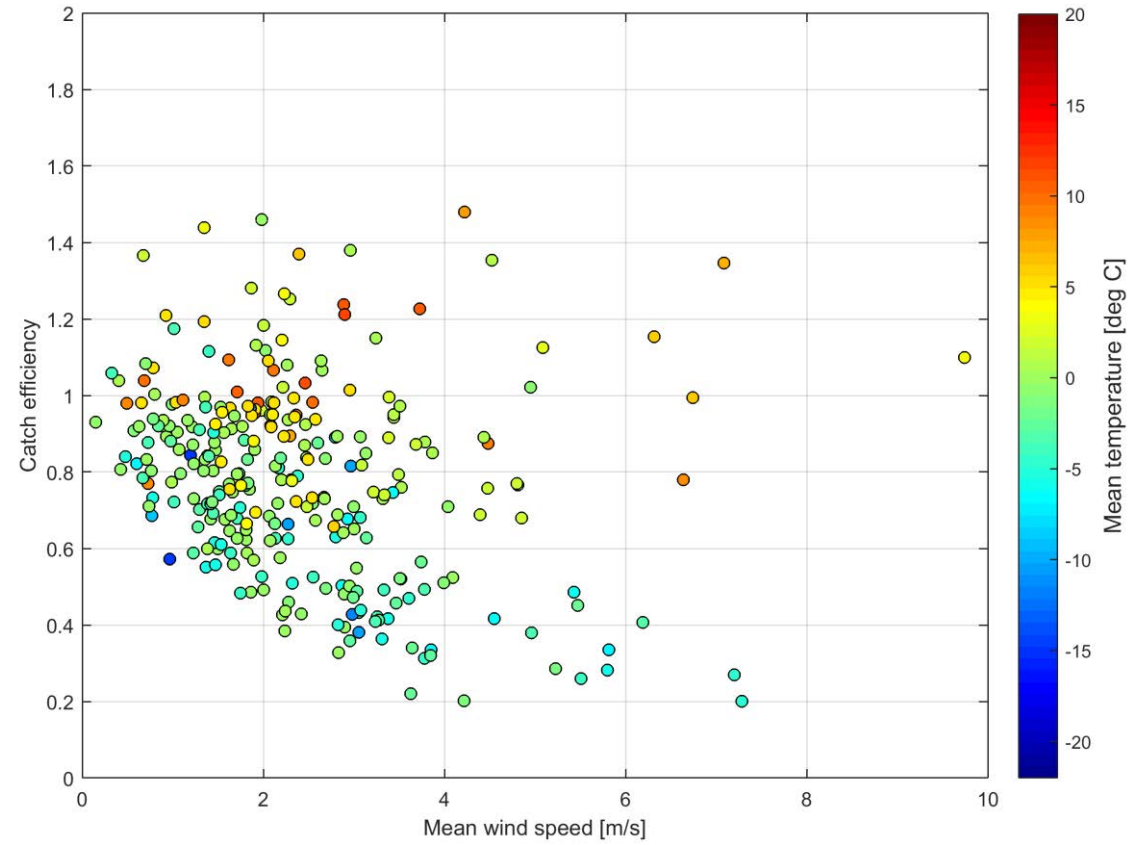
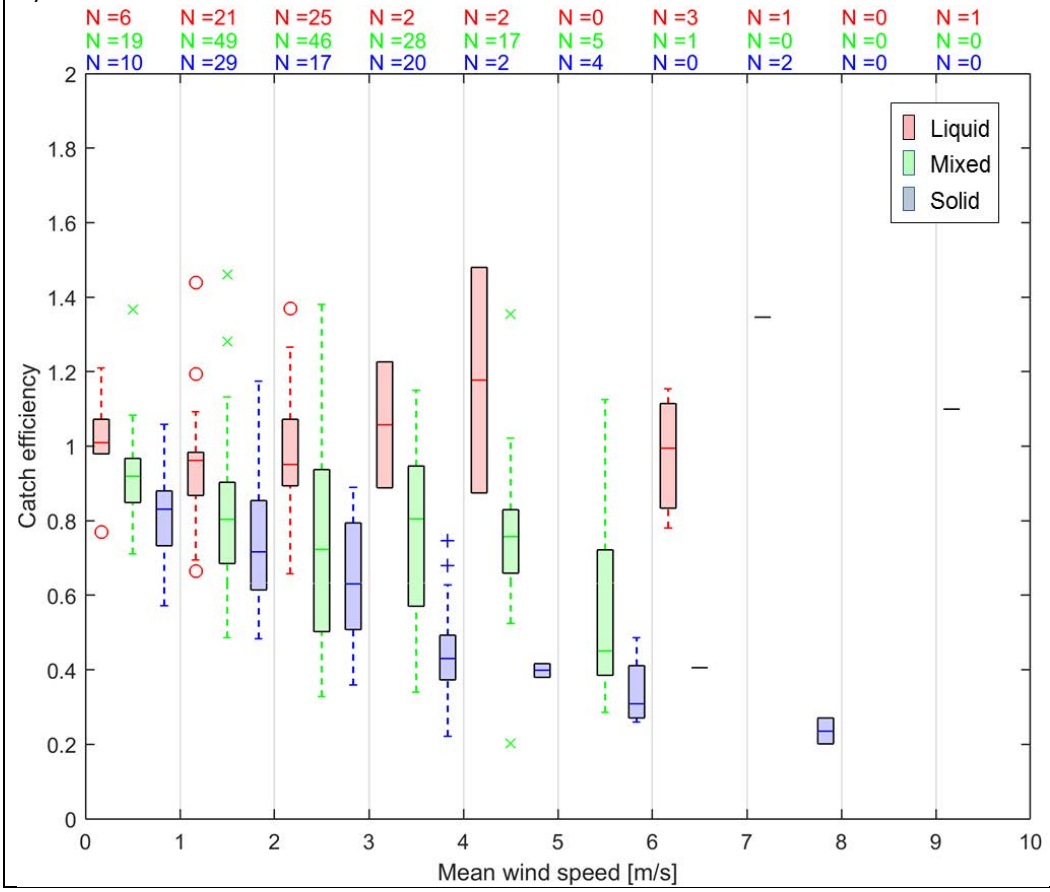

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
1
2
3
4
5
6

Figure 10: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded MRW500 gauge under test at Marshall.

a)



b)



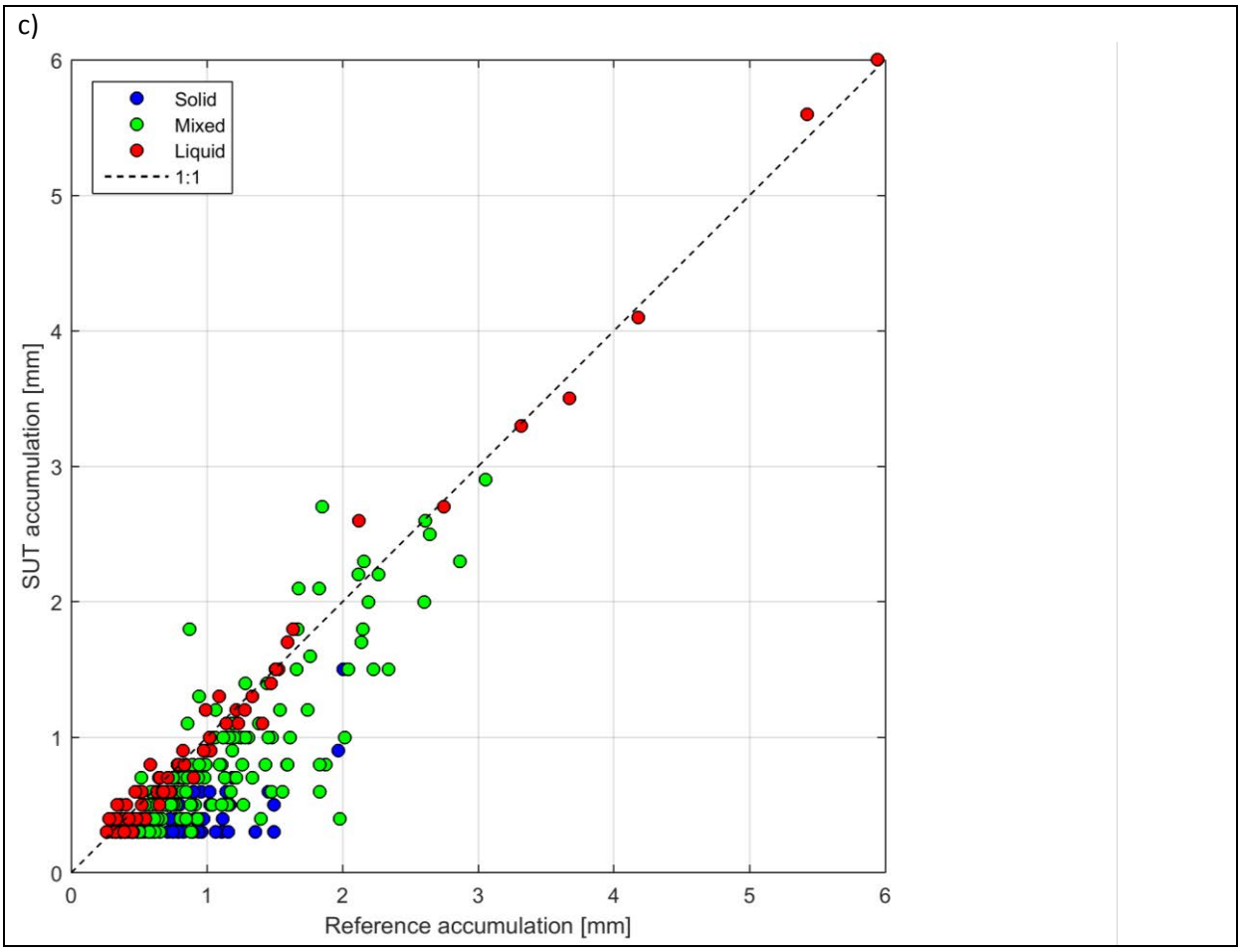
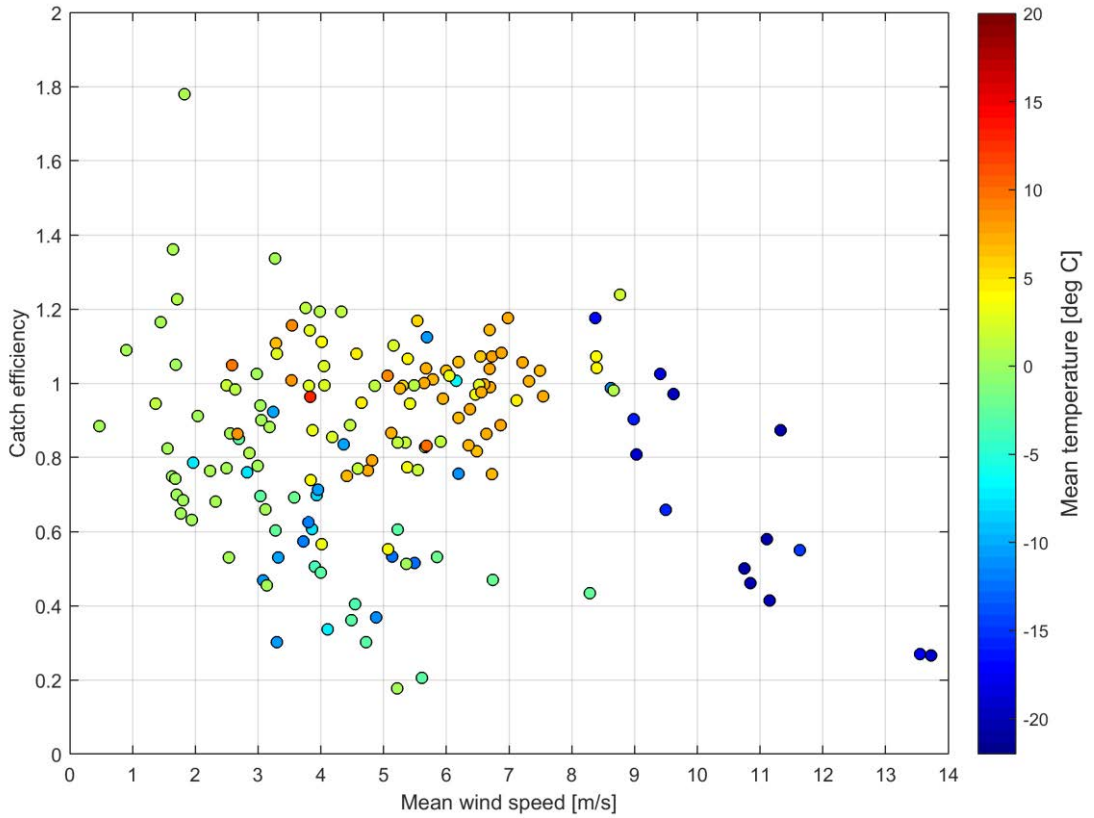
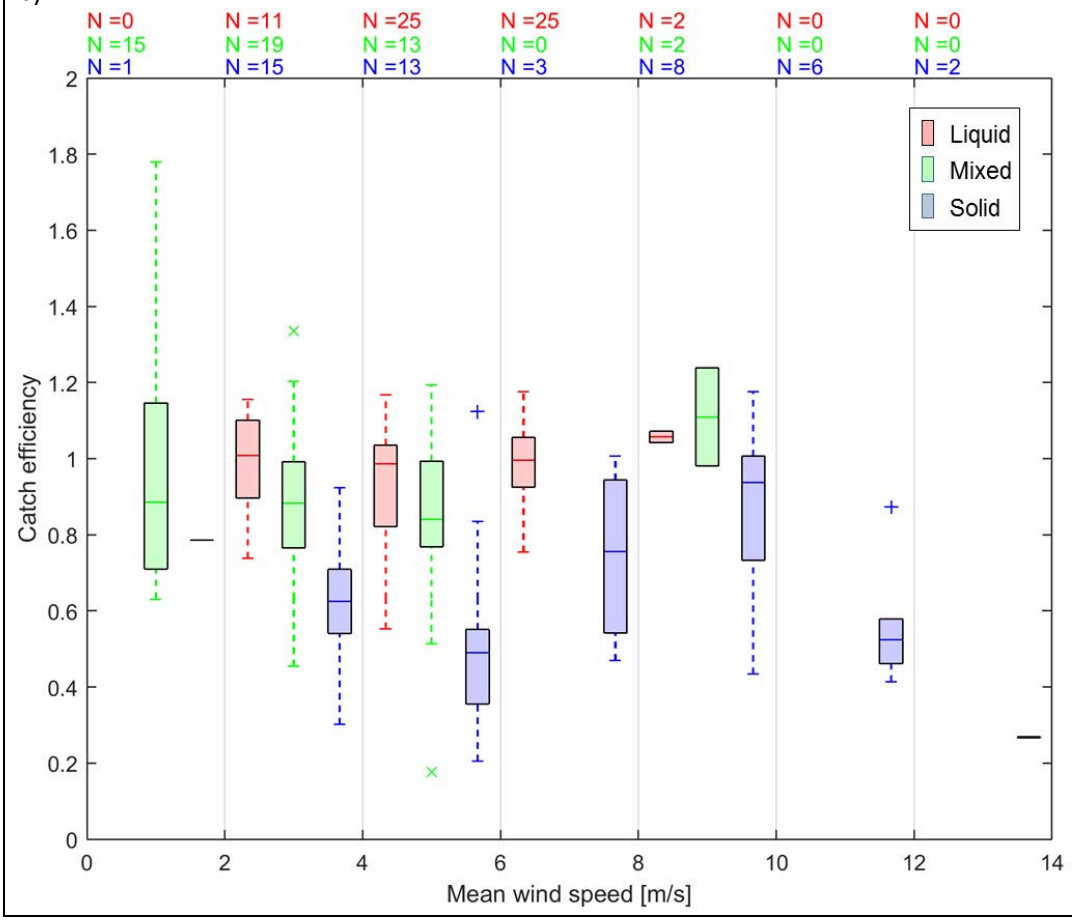

c)
Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
1
2
3
4
5
6

Figure 11: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the SA/Tret-shielded MRW500 gauge under test at Bratt's Lake.

a)



b)



c)

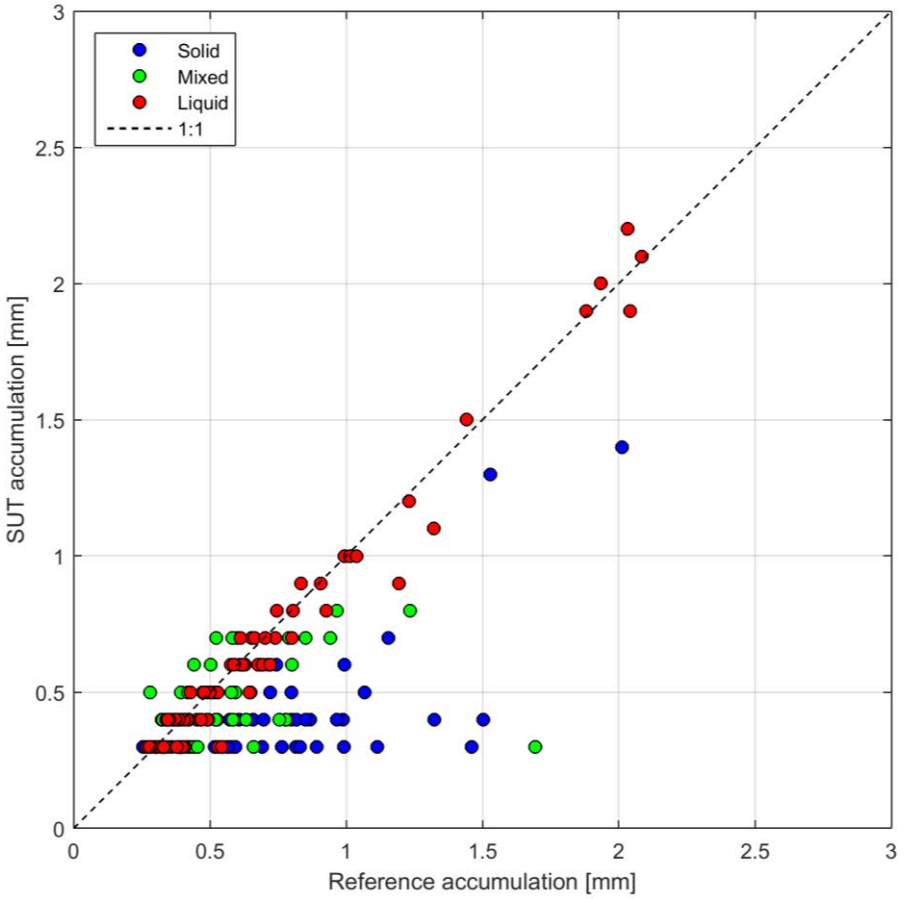
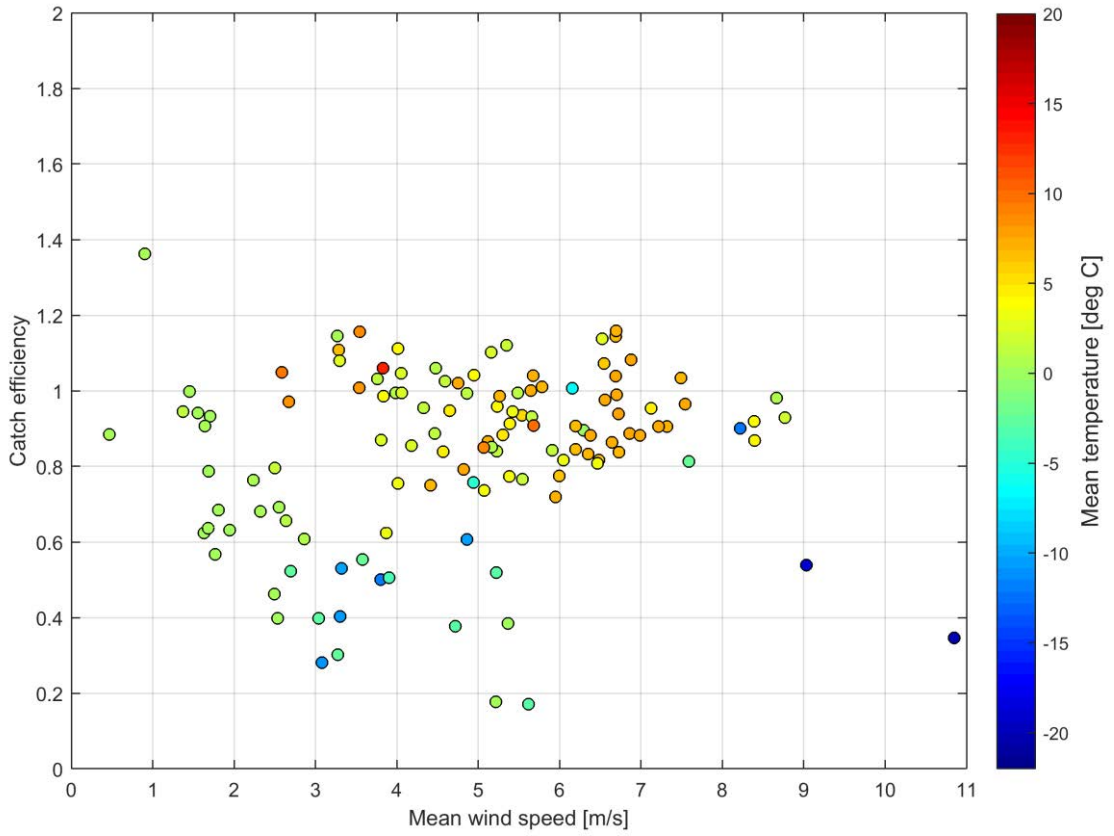
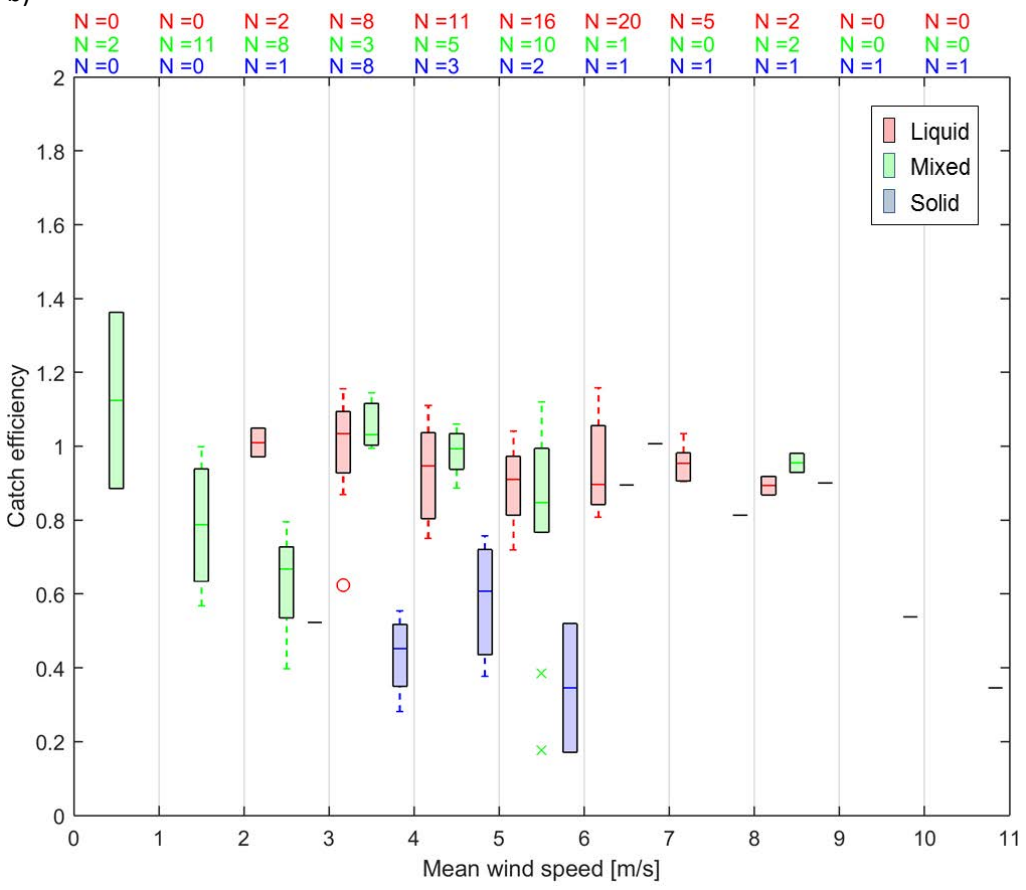

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
0.5
1
1.5
2
2.5
3

Figure 12: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded MRW500 gauge under test at Bratt's Lake.

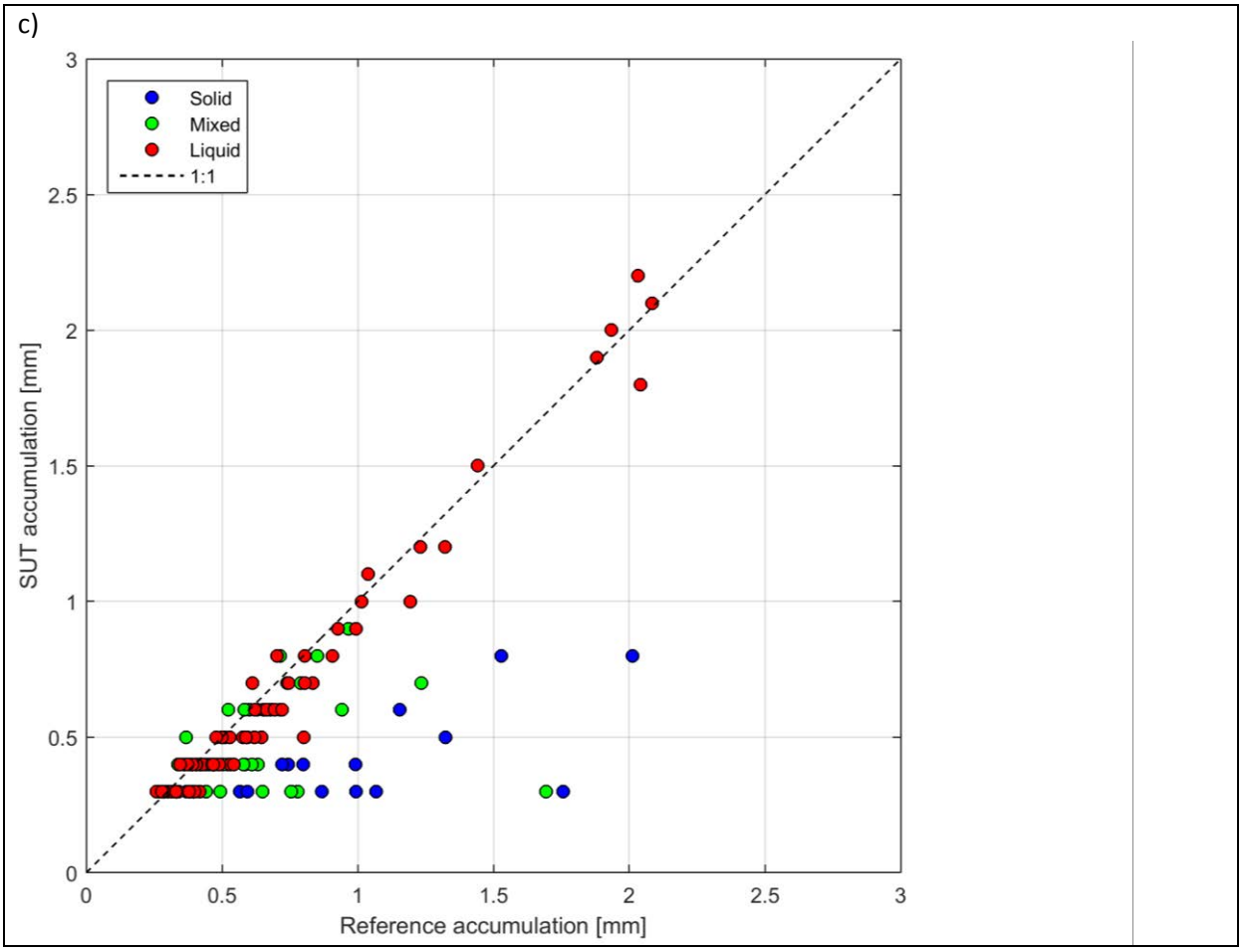
a)



b)



c)


Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
0.5
1
1.5
2
2.5
3

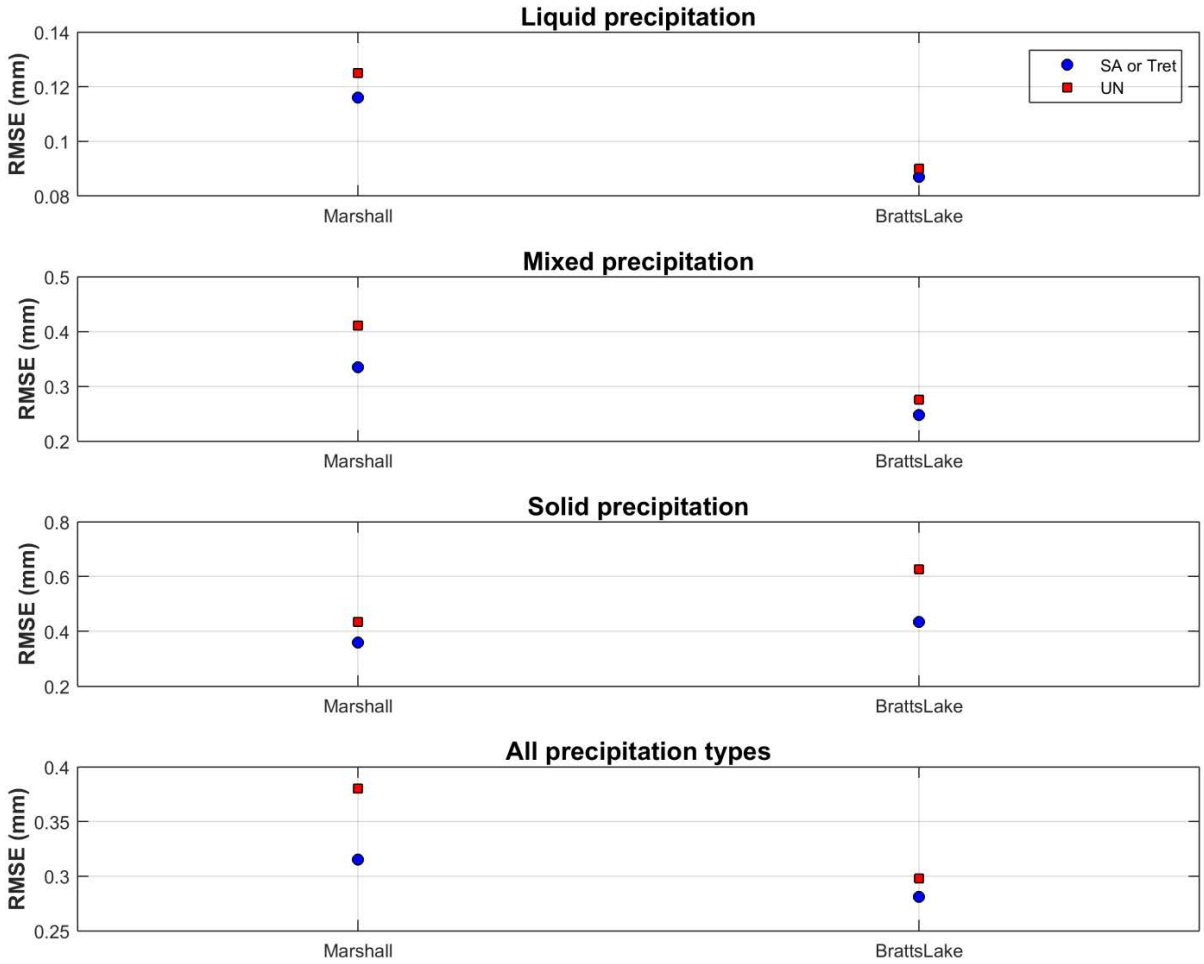


Figure 13: RMSE values calculated for test configuration(s) by precipitation type for YY cases over the entire test period.

Table 6: RMSE values calculated for test configuration(s) by site and by precipitation type, as presented in Figure 13.

Site	Configuration	RMSE (mm/30 min)			
		Liquid	Mixed	Solid	All precip types
Marshall	Unshielded	0.125	0.411	0.435	0.380
	SA/Tret	0.116	0.335	0.359	0.315
Bratt's Lake	Unshielded	0.090	0.276	0.626	0.298
	SA/Tret	0.087	0.248	0.435	0.281

The overall catch ratio calculated using all 30 minute YY cases, over the entire test period, is provided in Table 7. To demonstrate the influence of the SUT accumulation threshold on the results, the overall catch ratio is also provided for all 30 minute YY cases determined using a lower SUT threshold of 0.1 mm/30 minutes. Note that these values reflect only the YY cases, and do not include the amounts corresponding to the cases when the SUT and the reference do not agree on the occurrence of precipitation.

Table 7: Overall catch ratio for test configuration(s) determined from YY cases over the entire test period, using two different SUT accumulation thresholds.

Site	Configuration	SUT accumulation threshold (mm/30 min)	Overall catch ratio
Marshall	Unshielded	0.25	0.79
		0.1	0.75
	SA/Tret	0.25	0.85
		0.1	0.83
Bratt's Lake	Unshielded	0.25	0.80
		0.1	0.68
	SA/Tret	0.25	0.80
		0.1	0.68

6.4. Ability to detect light precipitation events

The impact of the threshold selection for data processing relative to the detection of light precipitation was examined using four different combinations of reference and SUT accumulation thresholds (four 'cases' in Table 8) for the SA/Tret-shielded MRW500 gauge under test at the Marshall site. Contingency results, probabilities of detection (POD), and false alarm rates (FAR) are presented for each case in Table 9. A quantitative comparison of the amounts reported in each case is beyond the scope of this assessment.

Table 8: Reference and SUT thresholds in each case for light precipitation detection assessment.

Case	Reference threshold (mm/30 min)	SUT threshold (mm/30 min)
1	0.25	0.25
2	0.1	0.1
3	0.25	No threshold
4	0.25	0

Table 9: Contingency results, probability of detection, and false alarm rate for each case in light precipitation detection assessment.

Case	Number of events				Skill score (%)	
	YY	YN	NY	NN	POD	FAR
1	346	53	139	19438	86.7	28.7
2	421	18	554	18983	95.9	56.8
3	399	0	19577	0	100	98
4	399	0	19577	0	100	98

6.5. Assessment of events when the reference and the SUT do not agree on the occurrence of precipitation

Assessment intervals during which the site reference and SUT do not agree on the occurrence of precipitation – namely, the YN and NY cases (Section 4.1.1) – are characterized for the test gauges in SA/Tret shields at Marshall and Bratt's Lake using histograms in Figures 14 and 15. The histograms include accumulated precipitation reported by the reference and SUT (0.25 mm/30 min threshold for both), precipitation intensity as reported by the reference, and corresponding site conditions.

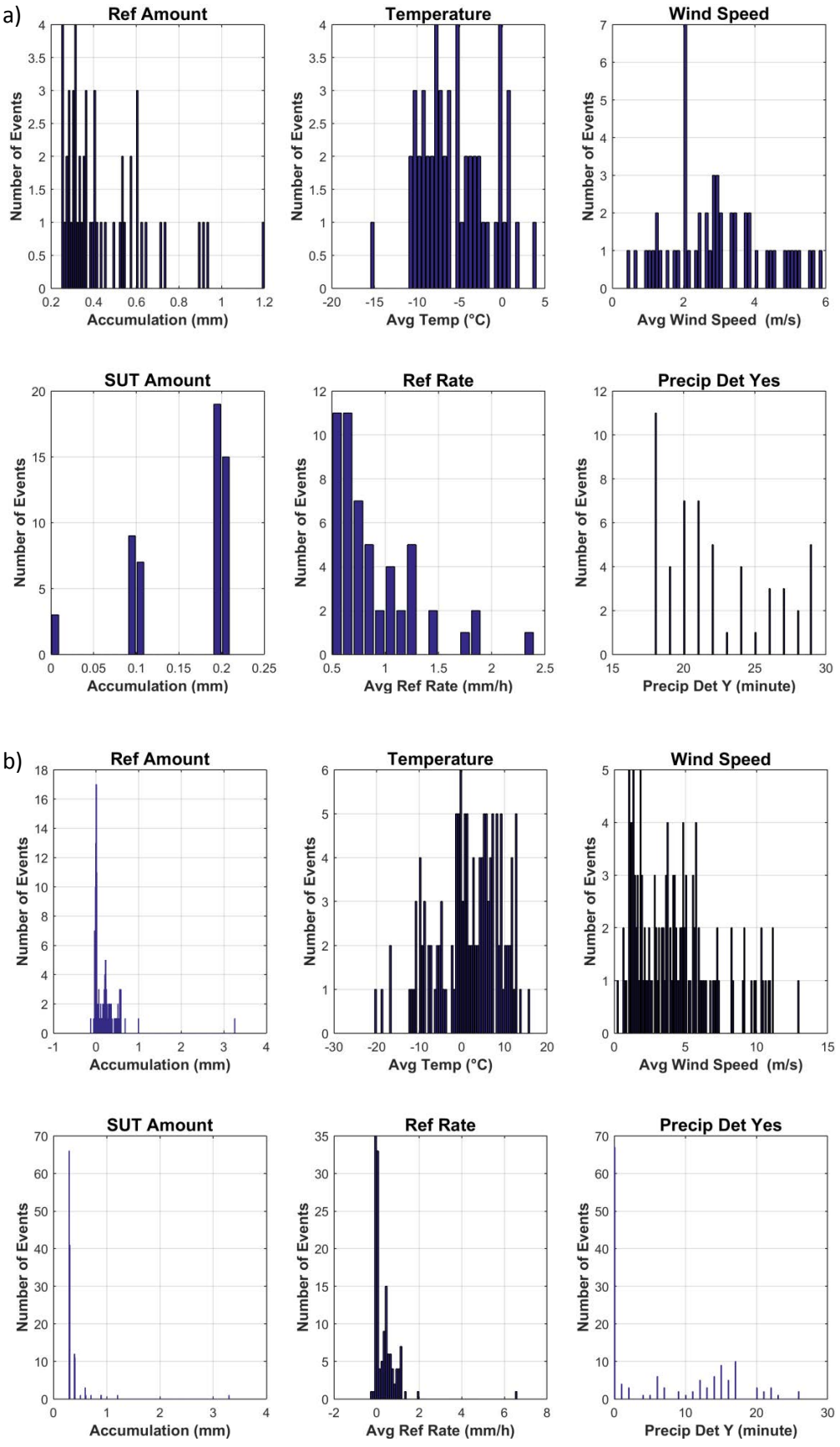


Figure 14: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for (a)YN events (53 total) and (b) NY events (139 total) for the SA/Tret-shielded MRW500 gauge at Marshall over the test period.

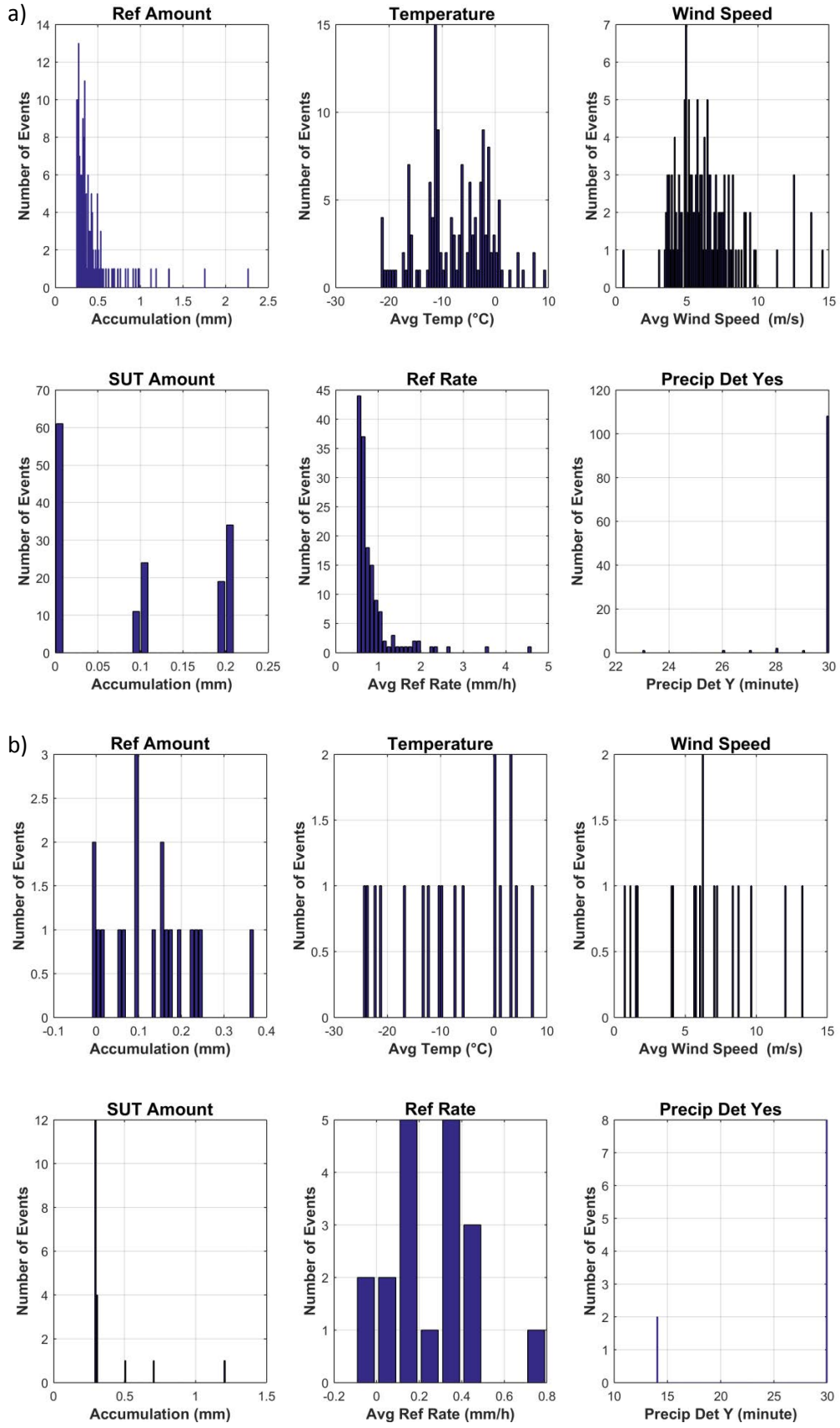


Figure 15: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for (a)YN events (149 total) and (b) NY events (19 total) for the SA/Tret-shielded MRW500 gauge at Bratt's Lake over the test period.

7) Interpretation of Results

7.1. Operating conditions

The full ranges of conditions under which the test gauges at each site were operated are illustrated in Figure 3. The conditions relevant to gauge operation are as follows:

- Temperatures between approximately -36 °C and 27 °C;
- Precipitation intensity within approximately 12 mm/hr.

These conditions fall largely within the manufacturer's specified operating conditions of -35 °C to 60 °C for temperature. While the temperatures at Bratt's Lake extend below the minimum recommended temperature (Figure 3), this applies to only 0.03% of the 30-minute events considered in the assessment. As the manufacturer-provided antifreeze is rated to -32 °C, a site-provided alternative (rated to -50 °C) was used for the second season at Bratt's Lake (see Table 1).

The conditions at each site during precipitation events are shown in Figure 4. Of particular note are the mean wind speeds during precipitation events at Bratt's Lake, which extend to approximately 15 m/s, with the maximum in the PDF (indicating the mean wind speeds with highest frequency of occurrence) at around 6 m/s. The mean wind speeds during precipitation events at Marshall are generally lower, with the maximum in the PDF at approximately 2 m/s.

7.2. Performance over the range of operating conditions

7.2.1. Non-precipitating conditions

The MRW500 data are processed using a manufacturer-developed algorithm, as outlined in Section 2. This processing results in narrow PDFs of accumulation over 30 minute periods without precipitation (Figure 5), indicating gauge reports of zero or very close to zero for both test configurations, at both sites. Indeed, the standard deviations of gauge outputs in the absence of precipitation for all test configurations are comparable to those of the corresponding reference configurations in Table 4.

Extreme values of 1.2 mm and 2.2 mm were reported by test configurations at Marshall and Bratt's Lake, respectively (Table 4), indicating the potential for false reporting by the algorithm in the absence of precipitation. The scatter plots in Figures 6 and 7 illustrate better the occurrence of non-zero reports; these reports are typically higher in magnitude for the test gauges at Bratt's Lake relative to those at Marshall, and show no clear trends with respect to increasing/decreasing temperature difference or increasing mean wind speed. It is difficult to comment further on the influence of data processing on non-zero reports, as the details of the algorithm are proprietary to the manufacturer.

The magnitude of gauge responses in the absence of precipitation can be used to identify a detection threshold that minimizes the detection of false precipitation while enhancing the detection of light precipitation. This threshold is considered to be three times the standard deviation of the average gauge response during 30 minute non-precipitating periods. Based on the present results for test

gauges at Bratt's Lake and Marshall (Table 4), this minimum detection threshold is determined to be 0.04 mm for single-Alter/Tretyakov-shielded gauges and 0.06 mm for unshielded gauges.

7.2.2. Precipitating conditions

7.2.2.1. Ability to detect and report precipitation

The skill scores and contingency results for the MRW500 test gauges (Figure 8 and Table 5, respectively) indicate that the ability to detect and report precipitation differs for test gauges in different configurations and at different sites. At both sites, the shielded configurations show higher POD, lower FAR, lower or comparable Bias, and higher HSS relative to the shielded configurations.

The specific conditions experienced during precipitation events at Bratt's Lake – notably, higher wind speeds – result in lower POD, FAR, B, and HSS for test gauges relative to the values for those at Marshall. The B values indicate that the test gauges at Marshall report a higher number of 30 min precipitation events than the reference (by about 25%), while those at Bratt's Lake report fewer events than reported by the reference (approximately 50%).

7.2.2.2. Ability to report accumulated precipitation

The results presented in Figures 9 to 12, which are based on 30 minute events during which the reference and test gauge both detect precipitation (YY cases), illustrate the influence of wind speed and precipitation type on gauge catch efficiency. The discussion below will focus on snow events; the number of rain events during winter is limited, and the results for mixed events are variable due to the variability in the size and density of precipitation within the mixed regime, as well as the potential for transitions between phases. The results for the test gauges at both sites illustrate the decrease in catch efficiency for solid precipitation with increasing mean wind speed. The median catch efficiencies for shielded and unshielded configurations are reduced to 0.5 – 0.6 and 0.4 – 0.5, respectively, for mean wind speeds between 3 m/s and 4 m/s. Hence, catch efficiency decreases more rapidly with increasing wind speed for unshielded MRW500 gauges relative to shielded MRW500 gauges at a given site. While the numbers of events reported at mean wind speeds greater than 4 m/s are limited, it appears that the median catch efficiencies for both configurations are reduced to the 0.2 to 0.3 range for wind speeds up to 8 m/s at Marshall, while the results for Caribou Creek are likely impacted by blowing snow at mean wind speeds greater than 6 m/s (corroborated by observer reports at a nearby airport).

Root mean square error values were computed from all 30 minute events during which each test configuration and the corresponding reference configuration both detected precipitation. The RMSE values presented in Figure 13 and Table 6 can be considered to represent the absolute uncertainty of each test configuration relative to the reference configuration at each site, in different precipitation conditions (liquid, mixed, solid, and all precipitation types). At both test sites, the RMSE values for shielded configurations are lower than those for unshielded configurations in each precipitation type classification, indicating that the former agree more closely with the reference reports. At Marshall (Bratt's Lake), the RMSE values for shielded and unshielded gauges are 0.32 mm/30 min and 0.38 mm/min (0.281 mm/min and 0.298 mm/min), respectively, in all precipitation conditions. Note that

these values will be impacted by differences in the relative proportions of events of each phase from site to site.

The overall catch ratio – computed from the total reference and SUT accumulation from all YY cases over the duration of formal tests – is provided for each test configuration in Table 7. The values for shielded and unshielded test configurations at both sites are similar, ranging from 0.79 to 0.85. Reducing the precipitation detection threshold for the test gauges from 0.25 mm/30 min to 0.1 mm/30 min decreases the overall catch ratio for all test configurations, with the largest decreases (approximately 15%) observed for the test gauges at Bratt's Lake.

7.2.2.3. Ability to detect light precipitation

The detection thresholds and results presented in Tables 8 and 9, respectively, indicate that decreasing the SUT detection threshold to zero or removing it entirely (while maintaining the reference detection threshold at 0.25 mm) increases both the Probability of Detection and False Alarm Rate to close to 100%; the corresponding Heidke Skill score values approach 0%, indicating no detection skill with these specific thresholds. Decreasing both the reference and SUT detection thresholds to 0.1 mm increases the POD for the test gauge considered (SA/Tret-shielded gauge at Marshall) from 87% to 96%, while increasing the FAR from 29% to 57%. These results suggest that lowering the thresholds enhances the detection of light precipitation, but also leads to the detection of more non-zero reports in the absence of precipitation.

7.2.3.4. Assessment of events when the reference and SUT do not agree on the occurrence of precipitation

The YN ('miss') cases, when the reference detects a precipitation event and the SUT does not, and NY ('false alarm') cases, when the SUT detects a precipitation event and the reference does not, are characterized for each test configurations in Figures 14 and 15. The 'miss' cases for test gauges at both sites (Figures 14a and 15a) are characterized by reference accumulations exceeding the detection threshold (0.25 mm) and SUT accumulations below the threshold; these differences most likely result from enhanced undercatch for single-shielded gauges relative to the reference in DFIR-fence. The SUT accumulation data indicate that most YN cases are just below the threshold for the test gauge at Marshall (0.2 mm), while most YN cases for the test gauge at Bratt's Lake correspond to zero accumulation reports; this difference is attributed to the higher mean wind speeds experienced at Bratt's Lake relative to Marshall (Figures 3 and 4).

The majority of 'false alarm' cases reported by the test gauge at Marshall are cases in which the reference accumulation threshold is met, but the precipitation occurrence threshold (18 minutes of precipitation for a 30 minute event) is not. In contrast, the majority of 'false alarm' cases reported by the test gauge at Bratt's Lake are cases in which the precipitation occurrence threshold is met, but the reference accumulation threshold is not. These results demonstrate that the assessment results depend not only on the instruments, but also the data derivation approach used in the assessment.

8) Maintenance

Gauge calibration: each site completed the gauge field calibration and verification as per manufacturer recommendations, at least once a year or following the emptying of the gauge. The calibration records have been stored by each site host.

9) Performance Considerations

9.1. Data processing and outputs

The MRW500 gauge data are processed using a proprietary algorithm, described in general terms in Section 2. The output data show noise of comparable magnitude to the reference configurations in the absence of precipitation, but also non-zero reports of up to approximately 2 mm. Users should be aware of the potential for false reports, and exercise caution when interpreting gauge data. Disclosure of the internal processing logic by the manufacturer would better enable users to understand the gauge output data, and how they change with modifications to the processing approach over time. The traceability of processing algorithms over time is an important consideration from a climate perspective.

Diagnostic parameters are available via the manufacturer's software. While there is a direct connection between a PC and the gauge, it doesn't appear that these parameters can be output routinely to a data logger. This capability, along with increased clarification on the meaning of each parameter, would increase the utility of gauge diagnostics for users.

9.2. Gauge configuration and operation

As tested, the Meteoservis MRW500 requires a 220 V power supply. The gauge has a large capacity and uses a wicking mechanism to enhance evaporation from the storage vessel, resulting in longer gauge service intervals (less frequent emptying relative to lower capacity gauges). A large volume of antifreeze (16 to 25 L, depending on site conditions) is required for extended operation in cold conditions (i.e. over a winter season). The complexity of the gauge, which uses pumps to keep the antifreeze solution mixed and the contents of the catchment vessel at appropriate levels, presents additional failure modes for the gauge that may necessitate service; however, no issues were experienced by the test gauges at either test site during the intercomparison.

Another concern with respect to the configuration is that pumps and tubes could be susceptible to freezing at low temperatures (i.e. below -25 °C). Freezing in the tubing and/or pumps is believed to have occurred for one of the test gauges at Bratt's Lake during a period in which the temperature decreased rapidly from above freezing to below -20 °C. It is hypothesized that condensation was formed during the decrease in temperature, which subsequently froze, impacting the output data (appearing as false precipitation reports).

Meteoservis MRW500 gauges were tested in both shielded and unshielded configurations. The shielded configurations at both test sites showed improved performance with respect to detecting precipitation and reporting accumulated precipitation relative to the unshielded configurations (each assessed relative to the reference configuration). From a performance standpoint, the shielded configuration is therefore recommended. The specific configuration of the shield provided by the manufacturer may not be suitable for use in heavy snowfall conditions (its proximity to the gauge

orifice may lead to snow capping), but no issues were experienced for either of the shielded configurations under test.

It is recommended that all field configurations be fully tested and validated prior to use, in order to ensure that physical and signal interferences are not degrading the gauge signal. This would include the confirmation of grounding tailored to the soil conditions, sturdiness of foundation and mounting, and signal conditioning.

9.3. Ancillary measurements and adjustments

The application of adjustment functions to measurements is strongly recommended to account for the reduction in gauge catch efficiency as the wind speed increases. Ancillary measurements of wind speed (preferably at gauge orifice height) and air temperature are required for the application of adjustment functions. In addition, ancillary measurements from a sensitive precipitation detector that is independent from the gauge are recommended to help distinguish precipitation events from false reports due to noise.

WMO-SPICE Instrument Performance Report
MPS TRwS405

1) Technical Specifications (from manufacturer provided documentation)

Instrument model:	MPS Total Rain weighing Sensor TRwS405
Physical principle:	Weighing gauge (WG) equipped with a strain gauge bridge to measure the weight of accumulated precipitation.
Capacity:	750 mm
Collecting area:	400 cm^2
Operating temperature range:	-40 °C to 70 °C
Measurement accuracy:	0.1% (as defined by manufacturer)
Resolution:	0.001 mm (as defined by manufacturer)

a)

b)

Figure 1: MPS TRwS405 gauges under test at (a) Marshall (USA) and (b) Haukeliseter (Norway).

2) Data Output Format

Gauge data output:

(according to instrument manual provided by manufacturer)

According to the instrument manual, the strain gauge measures the total weight of the bucket content and the on-board electronic module calculates the one minute sum of precipitation. The microprocessor-controlled electronics of the TRwS405 provide the following data:

- One minute rain intensity;
- Rain indication and duration;
- Total weight.

The manufacturer-designed algorithm, applied on-board, is designed to address the following (according to the instrument manual):

- Wind influence;
- Unreal jump transitions, particles (customer-selectable thresholds);
- Evaporation (the TRwS405 can also calculate evaporation as a secondary value);
- Temperature influence (using measurements from internal temperature sensor).

3) SPICE test configuration

Shield:	Unshielded
Test sites:	Marshall (USA), Haukeliseter (Norway)
Sensor provider(s):	All gauges evaluated were provided by the instrument manufacturer (MPS system s.r.o.)

A map of test site locations is provided in Figure 2.

3.1. Note on terms and acronyms used

Throughout this document, the following notations are used to identify the R2 reference (Ref) and senor under test (SUT) configurations:

Reference: 'DFIR' and 'DFAR' are used interchangeably for the R2 reference configurations. 'DFIR' refers to an automated gauge installed in a DFIR-fence, while 'DFAR' refers more explicitly to the Double-Fence Automated Reference configuration.

Sensors under test: 'UN' denotes unshielded test configurations.

Figure 2: Map of SPICE sites where MPS TRwS405 gauges were tested.

A summary of the configuration of instruments as tested, the duration of tests and availability of data reflected in these results, and the ancillary measurements used, by site, is available in Tables 1, 2, and 3, respectively.

Table 1: Summary of gauge configuration and data output, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	Marshall	**Haukeliseter**
Field configuration	Unshielded	Unshielded
Height of installation (gauge rim)	1.85 m	4.5 m
Heating	Rim heater; SPICE algorithm	
Antifreeze	Mixture of methanol (60%) and propylene glycol (40%)	Mixture of methanol (60%) and ethylene glycol (40%)
Oil	Automatic Transmission Fluid (ATF)	Statoil Hydraway 15LT
SUT data output frequency	1 minute	
Data QC	SPICE QC methodology	
Data temporal resolution	1 minute	
Processing interval for SPICE data analysis	30 minutes	

Table 2: Data availability, by measurement season and site.

Measurement season	Marshall	Haukeliseter
Season 1 (Oct. 2013 – Apr. 2014)	✓	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓	✓

Table 3: Summary of reference and ancillary measurements, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	Marshall	Haukeliseter
R2 Site Reference	Geonor T-200B3 600 mm (DFAR)	Geonor T-200B3M 1000 mm (DFAR)
R2 Precip Detector	Thies LPM (Site*)	Thies LPM X2 (Site*, 2013-2014) Thies LPM X5 (DFAR, 2014-2015)
Ancillary Temp Sensor	MetOne, model 060A-2/062, 2144-L (2 m)	PT100 (4.5 m)
Ancillary RH Sensor	Campbell Scientific CS500 (2m)	Campbell Scientific PWS100 (6 m)
Ancillary Wind Sensor	RM Young Wind Monitor 05103 (2 m)	Thies ultrasonic anemometer 3D (4.5 m; 2013-2014) RM Young Wind Monitor 05103 (4.5 m; 2014-2015)

*A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR shield; however, in cases where a more sensitive detector is available outside of the DFIR shield, or there are issues with the detector within the DFIR shield, a precipitation detector elsewhere on the site can be employed.

4) Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6.1 of the WMO-SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

4.1.1.1. Characterization of performance in non-precipitating conditions

The assessment data are derived over 30 minute intervals during which the precipitation detector in the R2 reference configuration reports 0 minutes of precipitation. The accumulation over these intervals (accumulation in minute 30 – accumulation in minute 1), representing the variability of the gauge response due to wind, evaporation, temperature, etc., is recorded, along with the mean wind speed, and the change in temperature (temperature in minute 30 – temperature in minute 1).

4.1.1.2. Assessment of ability to detect and report accumulation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute assessment intervals:

- Ref 'Yes' : R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No' : R2 weighing gauge < 0.25 mm AND/OR precip detector recording < 18 min of precip;
- SUT 'Yes' : SUT accumulation ≥ 0.25 mm;
- SUT 'No' : SUT accumulation < 0.25 mm.

4.1.1.3. Assessment of ability to detect light precipitation

The data for this component of the assessment are derived in a similar manner as those in Section 4.1.1.2, but with different combinations of thresholds for the reference and/or SUT 'Yes' and 'No' conditions. These different threshold 'cases' have been selected to demonstrate the impact of the thresholds used in data derivation on the detection of light precipitation.

4.1.2. Skill score assessment

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores, as detailed in Section 3.6.1.3 of the WMO-SPICE Final Report.

For the assessments considered in this report, the ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

5) Environmental conditions

The environmental conditions at each site over the duration of the test period are expressed as probability density functions (PDFs) of mean air temperature, mean relative humidity, mean wind speed, vector mean wind direction, and precipitation rate for each component 30 minute assessment interval in Figure 3. The same parameters are also shown for all assessment intervals during which the site reference configuration detected precipitation (i.e. all Ref 'Yes' cases) in Figure 4.

The precipitation percentage represents the number of minutes of precipitation during a 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration, expressed as a percentage. PDFs of precipitation percentage are also included in Figures 3 and 4.

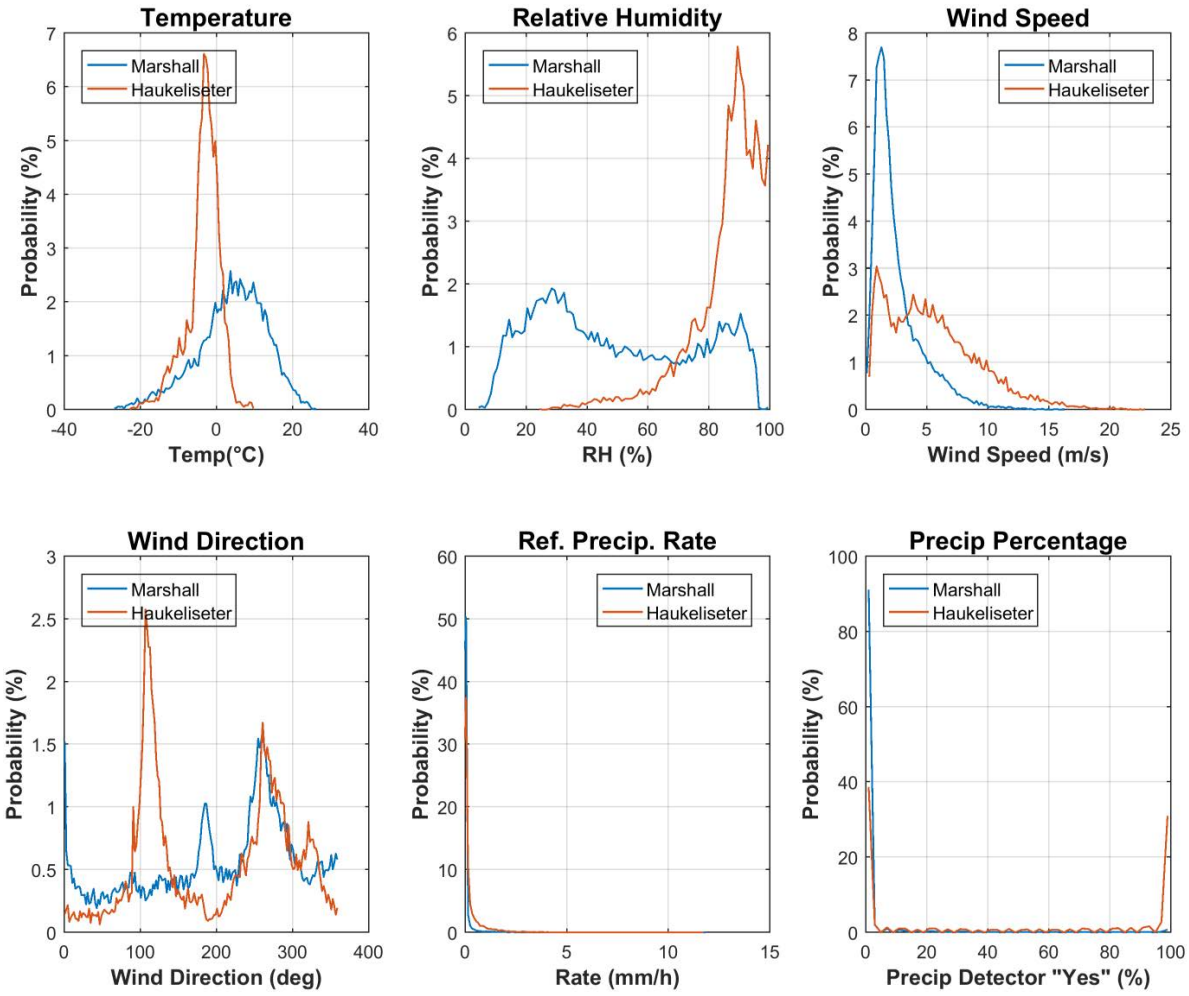


Figure 3: Summary of environmental conditions at Marshall and Haukeliseter over the entire duration of formal tests, as per Table 2.

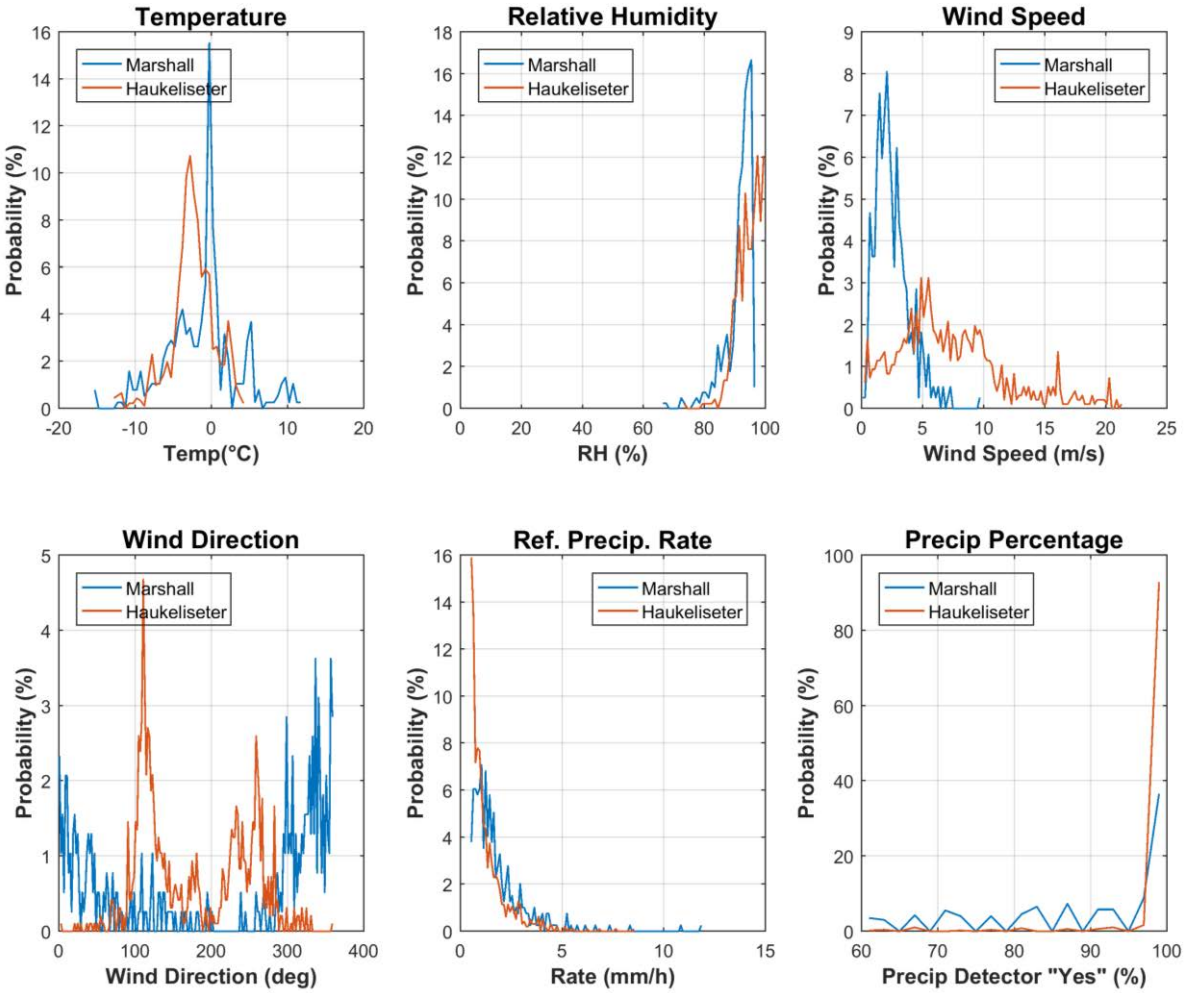


Figure 4: Summary of environmental conditions at Marshall and Haukeliseter during precipitation events (as defined by the R2 reference), over the duration of formal tests.

6) Evaluation of the ability to perform over the range of operating conditions

6.1. Characterization of SUT performance in non-precipitating conditions

The response of the SUT in the absence of precipitation was examined as defined in Section 4.1.1.1. The results are presented below, reflecting the distribution of the sensor response and its variability with wind and temperature, as measured during 30 minute assessment intervals.

6.1.1. Overall variability of SUT response

The overall variability of the SUT response in non-precipitating conditions is shown as a probability density function for each test configuration in Figure 5. The corresponding PDF for the reference configuration is provided for comparison.

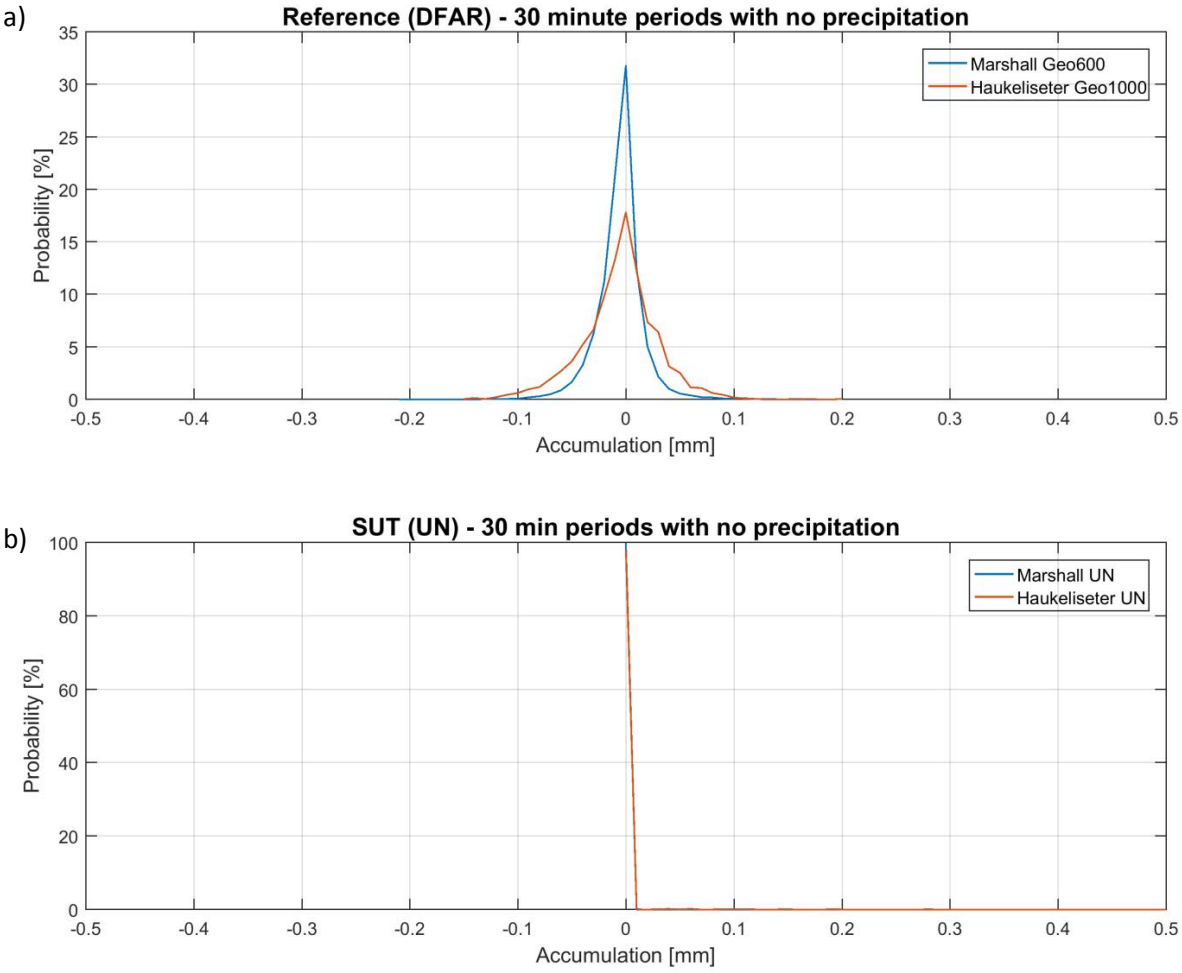


Figure 5: Probability density functions of the output signal (accumulation over 30 minute assessment intervals) in non-precipitating conditions for (a) the R2 reference configuration (DFAR) and (b) MPS TRwS405 test gauges in unshielded configurations at Marshall and Haukeliseter sites.

The statistics of the output signal (accumulation over 30 minute assessment intervals) for the reference and SUT at each site are provided in Table 4.

Table 4: Statistics of the R2 reference gauge and SUT output signal during non-precipitating conditions at Marshall and Haukeliseter, as plotted in Figure 5.

Site	Gauge	Average output signal (mm)	Standard deviation (mm)	Maximum output signal (mm)	Minimum output signal (mm)	Number of assessment intervals
Marshall	DFAR	0.0005	0.022	0.2	-0.2	18265
	SUT	0.0004	0.029	3.7	0	18265
Haukeliseter	DFAR	0.001	0.035	0.2	-0.14	2290
	SUT	0.005	0.059	1.5	0	2290

6.1.2. Variability of SUT response as a function of temperature

The variability of the SUT response for each test configuration in the absence of precipitation is plotted as function of the temperature difference over each assessment interval in Figure 6. The temperature difference is defined as the difference in temperature between the end (minute 30) and beginning (minute 1) of the assessment interval. The corresponding plots for the reference configurations are provided for comparison.

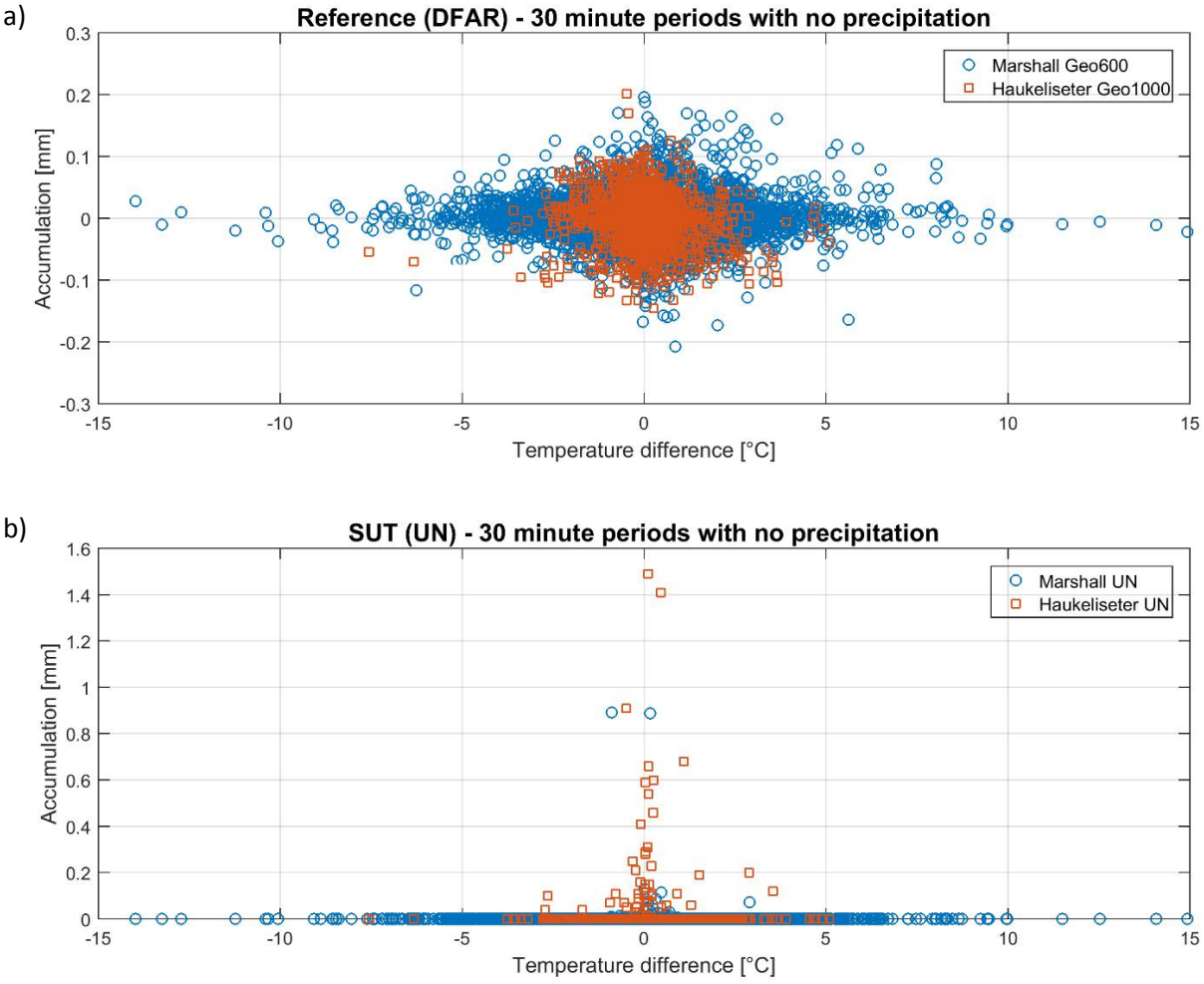
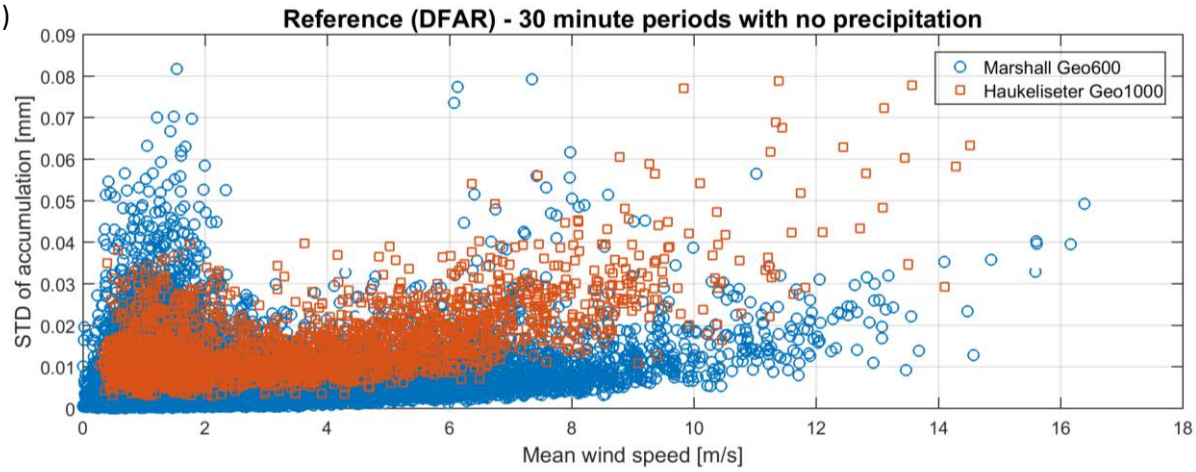


Figure 6: Variability of output signal (accumulation over each 30 minute assessment interval) as a function of the temperature difference over the interval in non-precipitating conditions for (a) the R2 reference configuration (DFAR) and (b) MPS TRwS405 test gauges in unshielded configurations at Marshall and Haukeliseter sites.

6.1.3. Variability of SUT response as a function of wind speed

The variability of the SUT response for each test configuration in the absence of precipitation is plotted as function of the mean wind speed for each assessment interval in Figure 7. Here, the signal variability is represented as the standard deviation (STD) of the gauge accumulation output over each 30 minute interval. The corresponding plots for the reference configurations are provided for comparison.

a)



b)

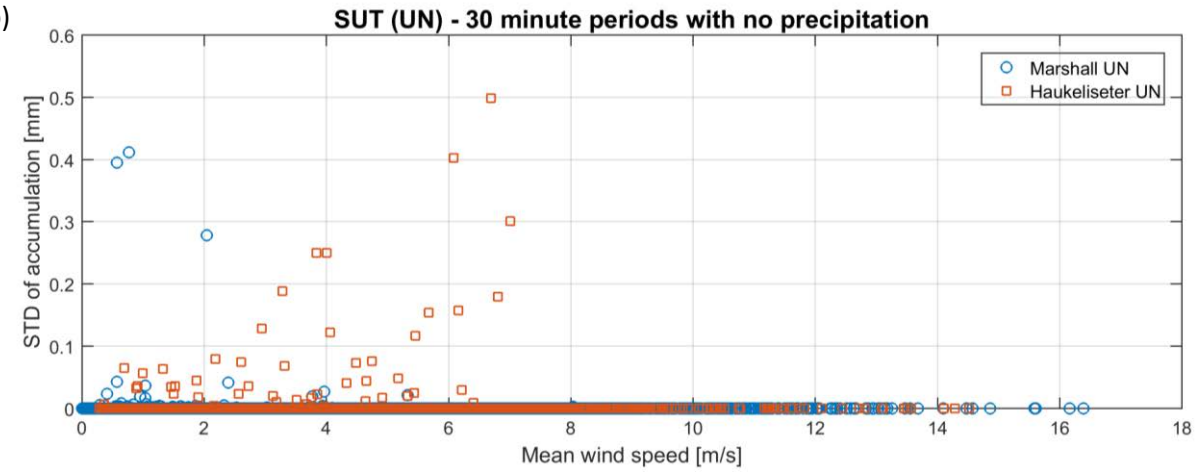


Figure 7: Variability of output signal (accumulation over each 30 minute assessment interval) as a function of mean wind speed in non-precipitating conditions for (a) the R2 reference configuration (DFAR) and (b) MPS TRwS405 test gauges in unshielded configurations at Marshall and Haukeliseter sites.

6.2. Ability to detect and report precipitation

6.2.1. Skill score assessment

The overall ability of the SUT to detect and report the occurrence of precipitation relative to the site field reference over 30 minute assessment intervals is expressed using selected skill scores (Section 4.1.2) and presented in Figure 8. The contingency results (Section 4.1.1) corresponding to these scores are presented in Table 5.

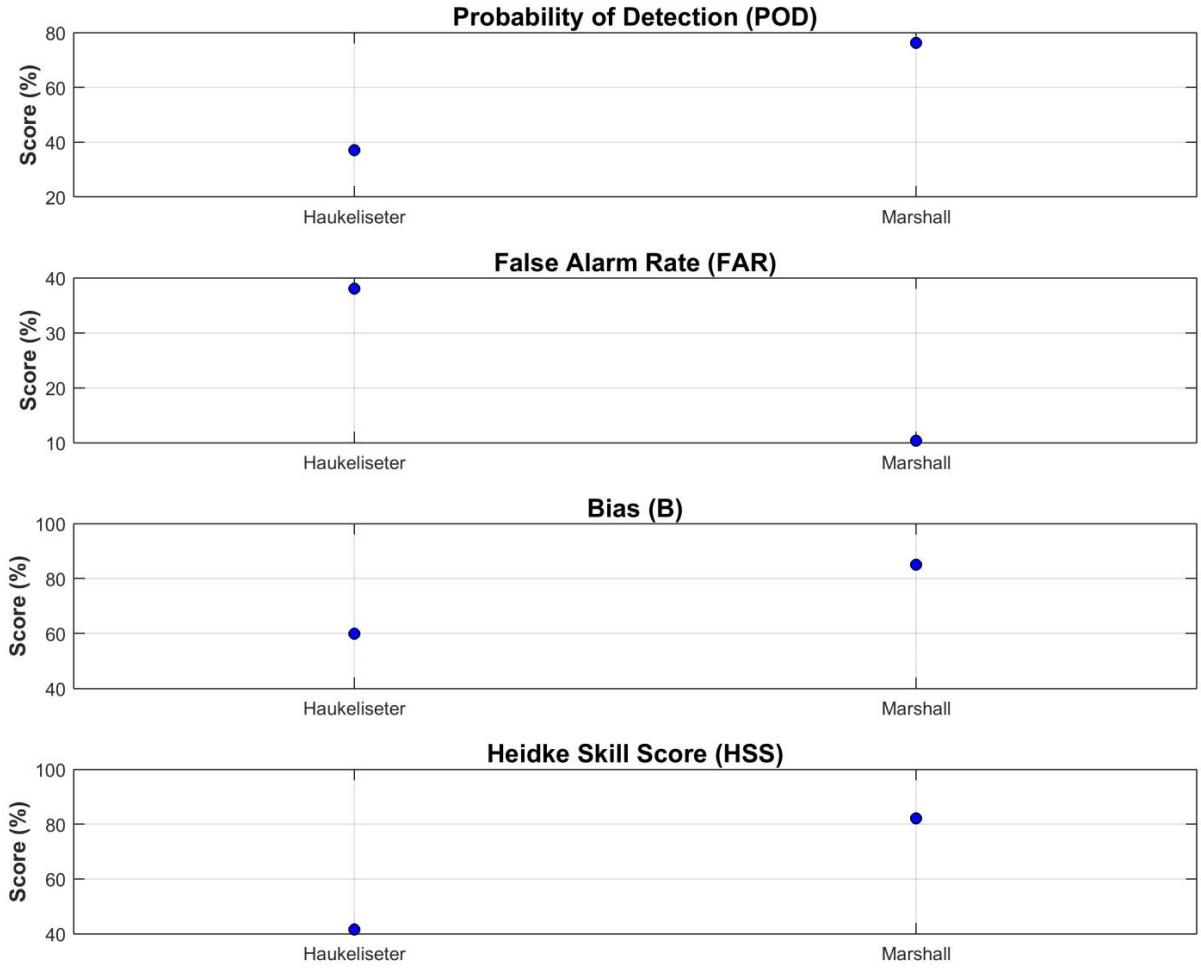


Figure 8: Skill scores for MPS TRwS405 gauges under test at Marshall and Haukeliseter over the duration of formal tests.

Table 5: Contingency table illustrating detection of precipitation by MPS TRwS405 gauges under test relative to the corresponding site reference configurations at Marshall and Haukeliseter, expressed as number of events over the entire test period.

Site	Number of Events			
	YY (hits)	YN (misses)	NY (false alarms)	NN (correct negatives)
Haukeliseter	357	605	219	7466
Marshall	302	94	35	19616

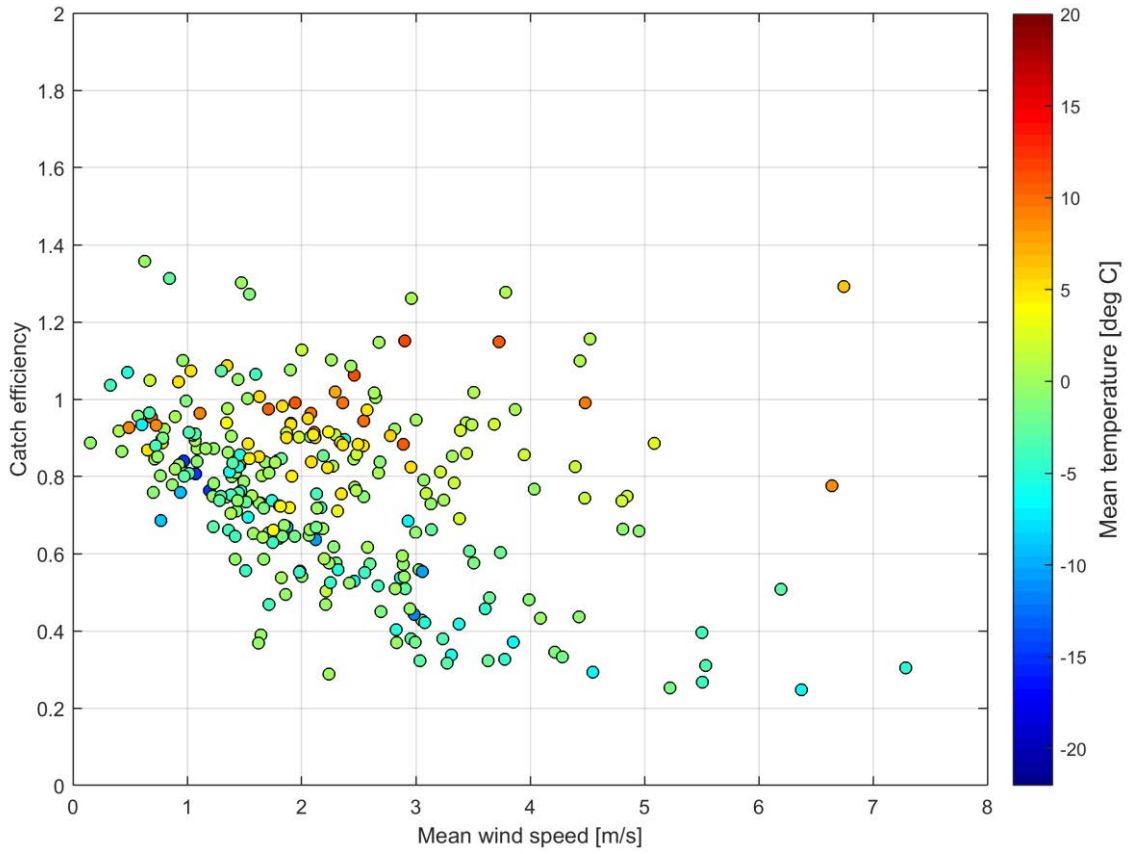
6.3. Ability to report accumulated precipitation

The SUT performance in terms of reporting accumulated precipitation is examined by comparing the amount reported by the sensor under test relative to the respective site reference during 30 minute assessment intervals. This is represented graphically using scatter and box and whisker plots of the catch efficiency as a function of mean wind speed at gauge height, as well as scatter plots of the amounts reported by the SUT versus the corresponding reference amounts (Figures 9 and 10). The SUT performance is also assessed in terms of the root mean square error, RMSE (Figure 11).

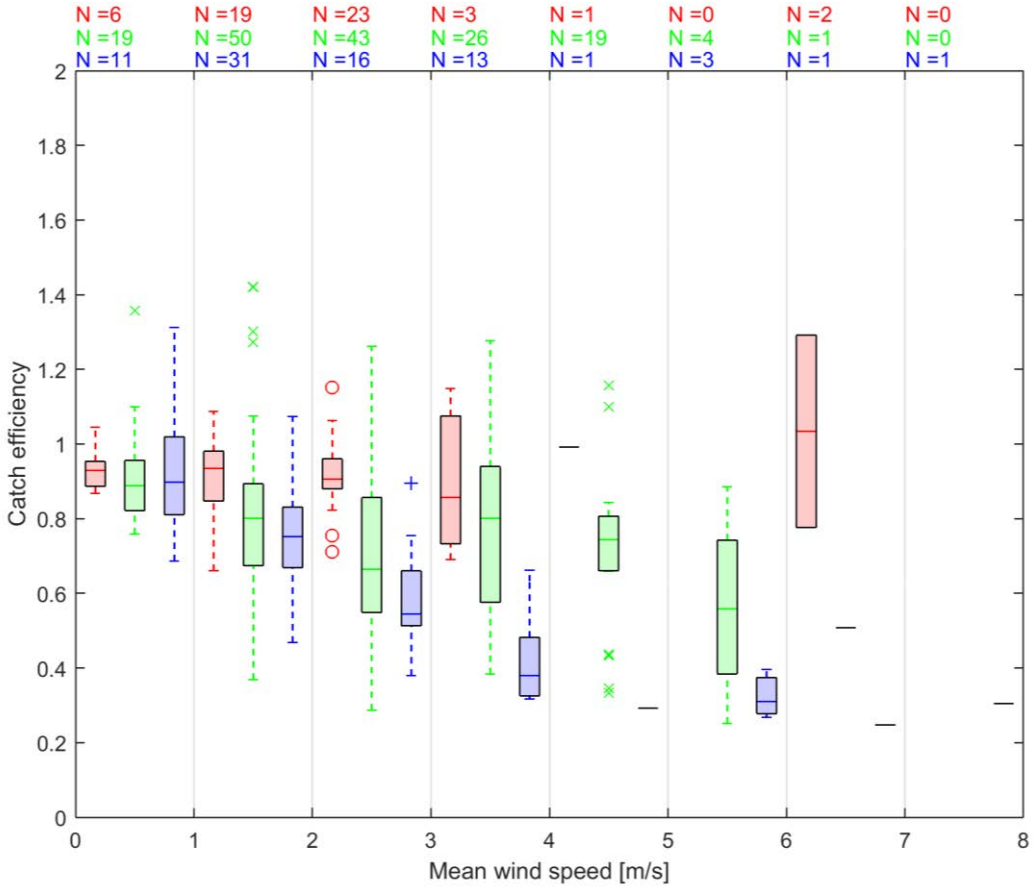
Only assessment intervals during which the SUT and reference both reported precipitation (YY cases) are considered in this portion of the assessment. In the catch efficiency-wind speed scatter plots, the mean event temperature is indicated by colour, with the colour scale selected to be consistent across all sites with weighing gauges under test. In the box and whisker plots and accumulation-accumulation scatter plots, the predominant precipitation type is indicated by colour, as determined from the reported temperature (Section 4.1.4).

Figure 9: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded MPS TRwS405 gauge under test at Marshall.

a)



b)


N =6
N =19
N =11
N =19
N =50
N =31
N =23
N =43
N =16
N =3
N =26
N =13
N =1
N =19
N =1
N =0
N =4
N =3
N =2
N =1
N =1
N =0
N =0
N =1
Catch efficiency
Mean wind speed [m/s]

c)

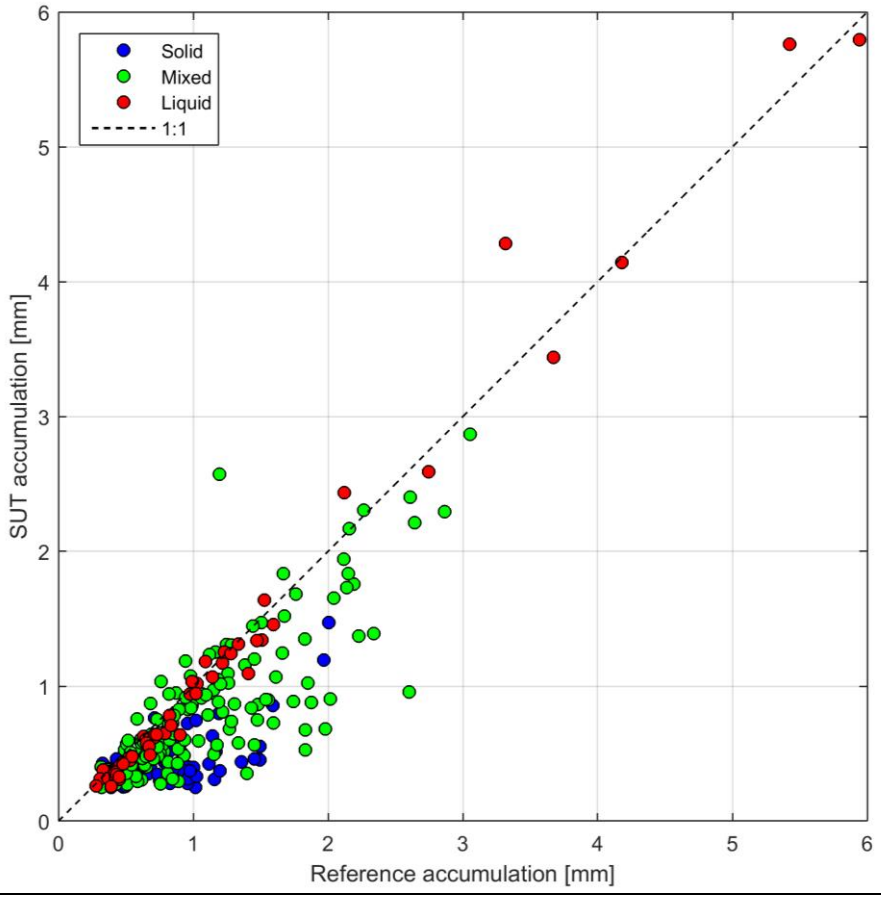
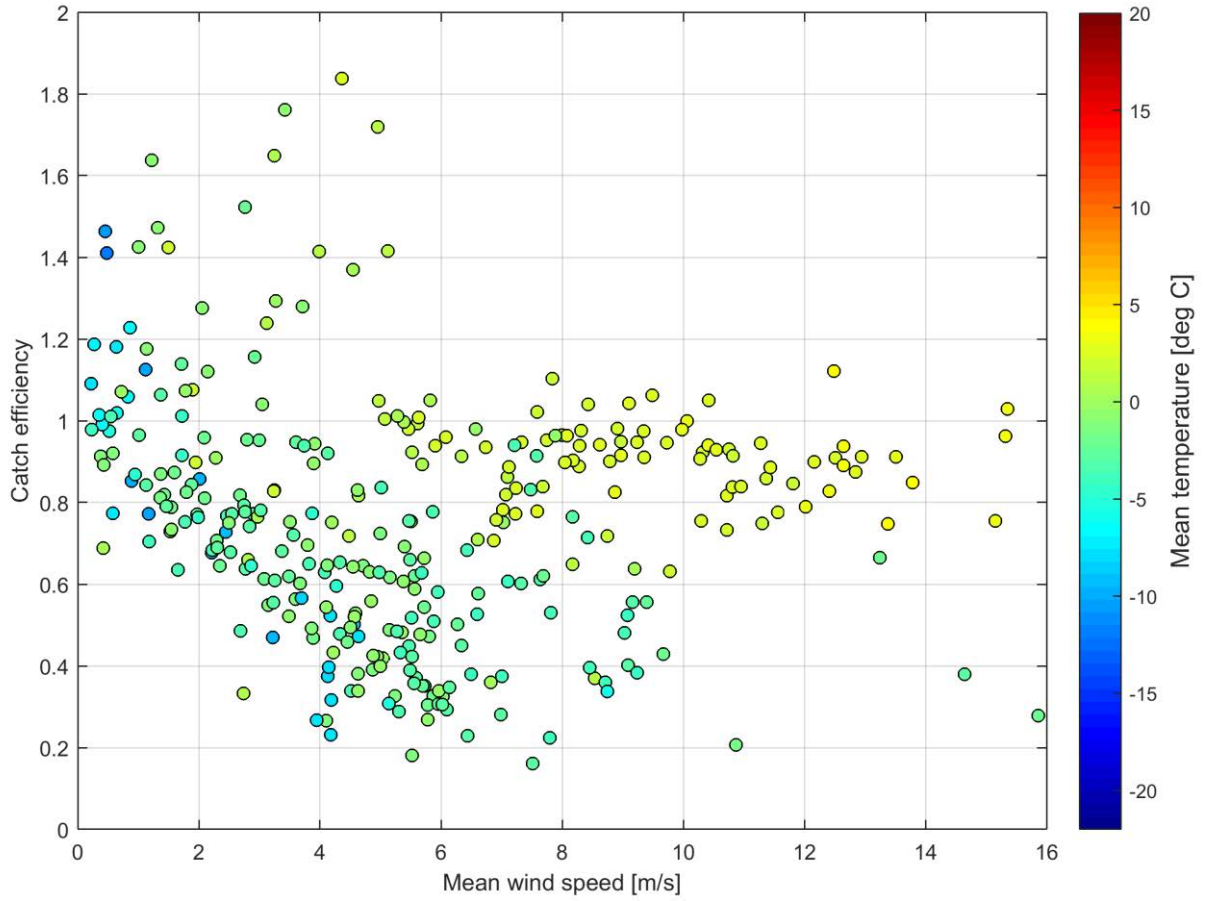
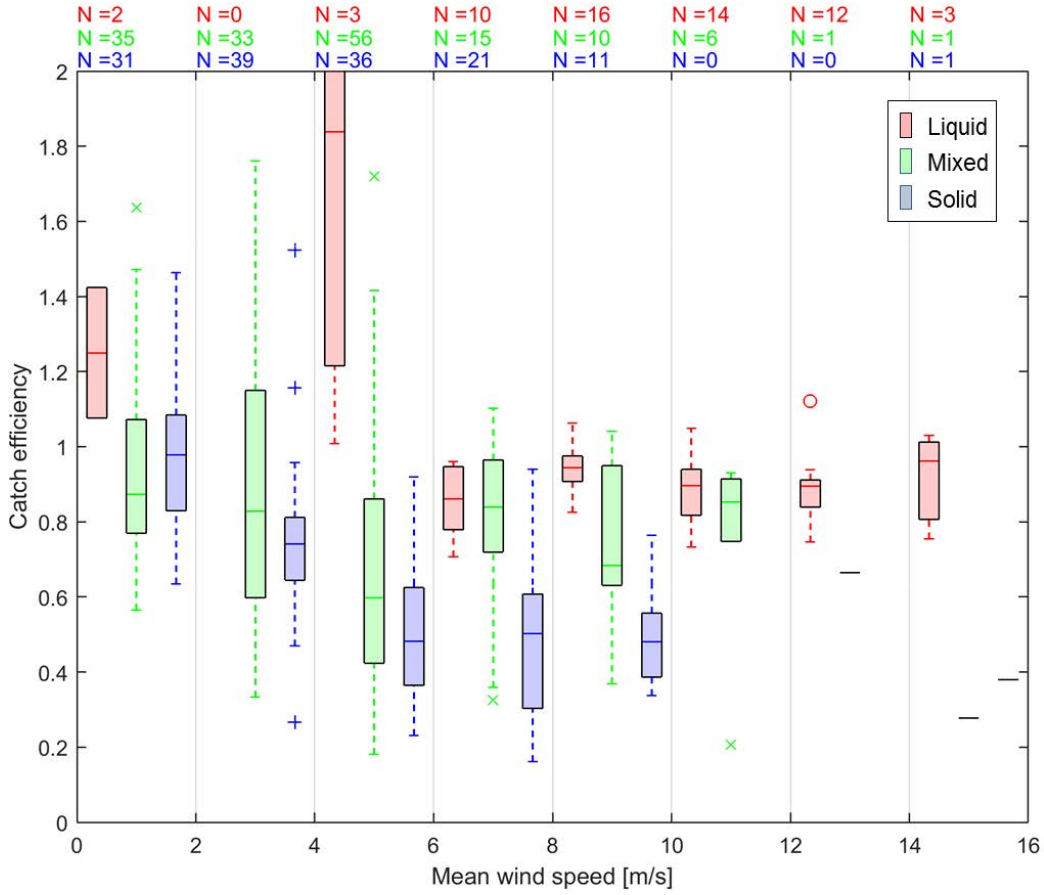

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]

Figure 10: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded MPS TRwS405 gauge under test at Haukeliseter.

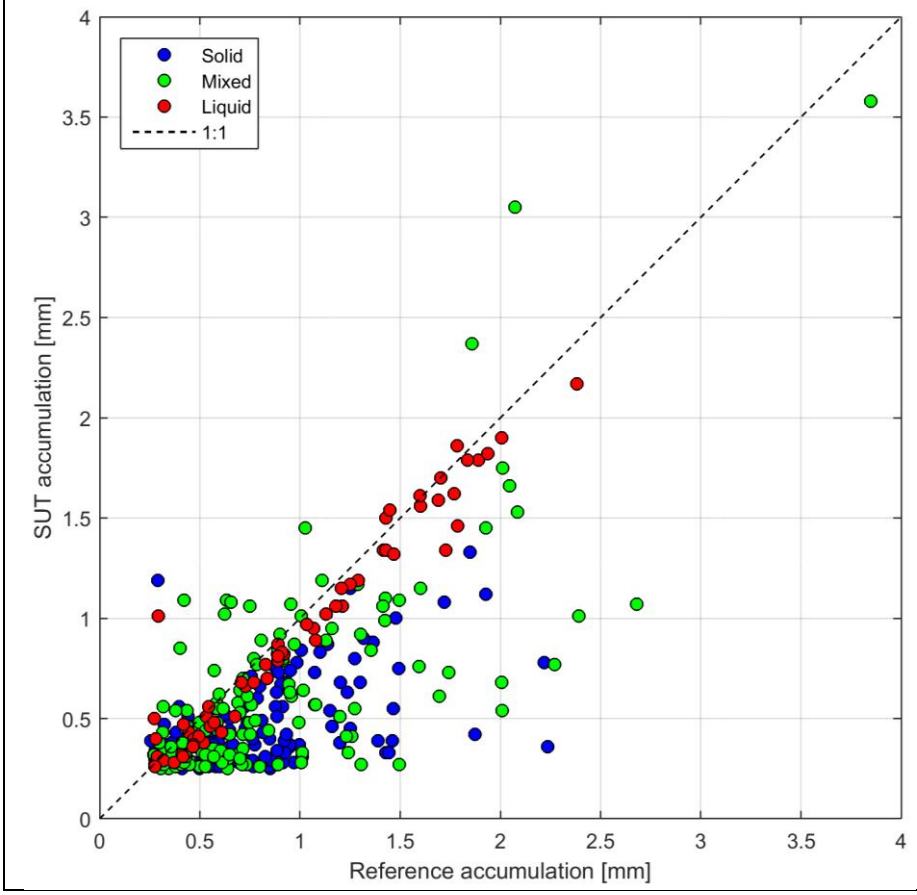
a)



b)


N =2
N =35
N =31
N =0
N =33
N =39
N =3
N =56
N =36
N =10
N =15
N =21
N =16
N =10
N =11
N =14
N =6
N =0
N =12
N =1
N =0
N =3
N =1
N =1
Liquid
Mixed
Solid
Catch efficiency
Mean wind speed [m/s]

c)


Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]

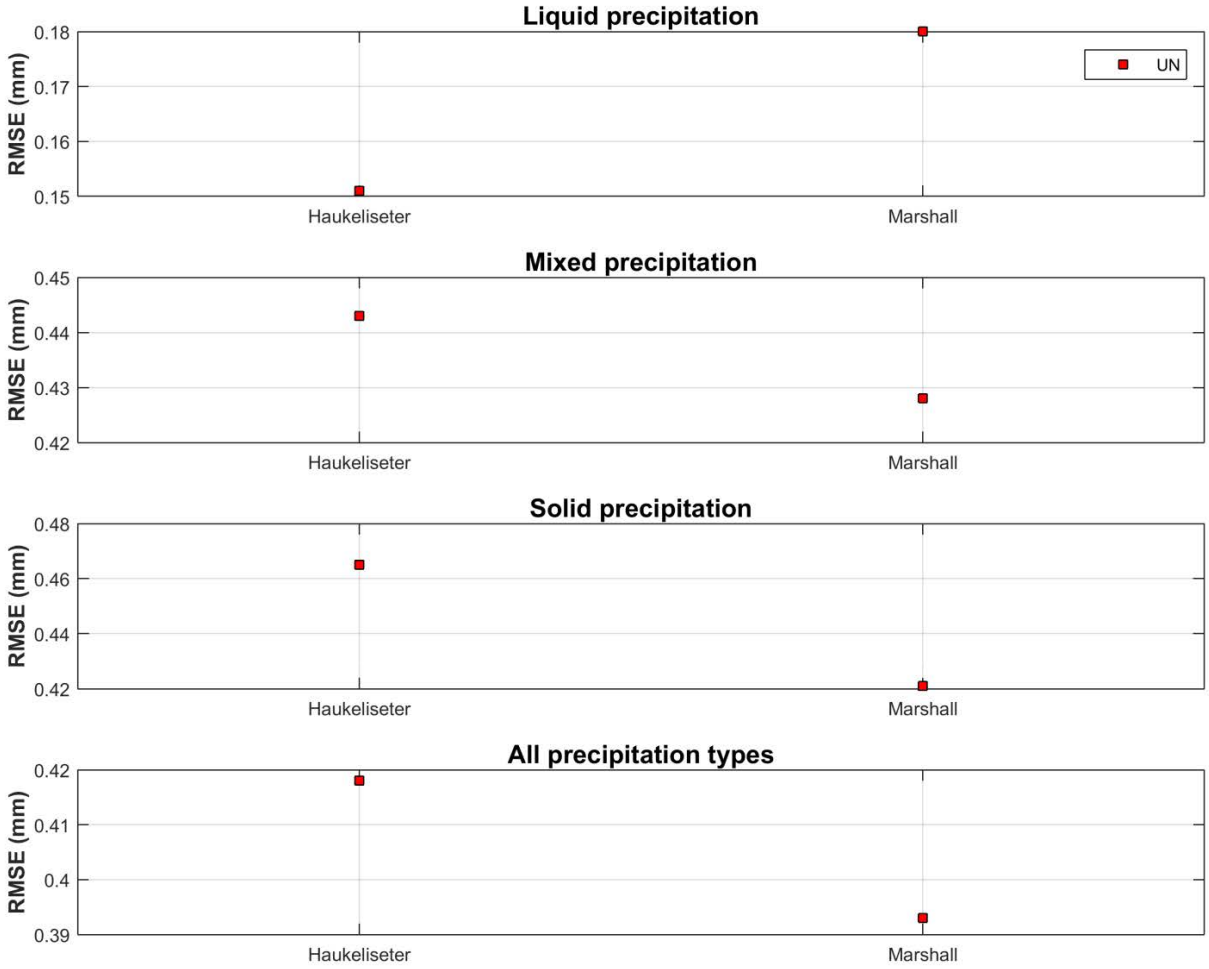


Figure 11: RMSE values calculated for test configuration(s) by precipitation type for YY cases over the entire test period at Marshall and Haukeliseter.

Table 6: RMSE values calculated for test configuration(s) by site and by precipitation type, as presented in Figure 11.

Site	RMSE (mm/30 min)			
	Liquid	Mixed	Solid	All precip types
Haukeliseter	0.151	0.443	0.465	0.418
Marshall	0.180	0.428	0.421	0.393

The overall catch ratio calculated using all 30 minute YY cases, over the entire test period, is provided in Table 7. To demonstrate the influence of the SUT accumulation threshold on the results, the overall catch ratio is also provided for all 30 minute YY cases determined using a lower SUT threshold of 0.1 mm/30 minutes. Note that these values reflect only the YY cases, and do not include the amounts corresponding to the cases when the SUT and the reference do not agree on the occurrence of precipitation.

Table 7: Overall catch ratio for test configuration(s) determined from YY cases over the entire test period at Marshall and Haukeliseter, using two different SUT accumulation thresholds.

Site	SUT accumulation threshold (mm/30 min)	Overall catch ratio
Haukeliseter	0.25	0.73
	0.1	0.68
Marshall	0.25	0.77
	0.1	0.74

6.4. Ability to detect light precipitation events

The impact of the threshold selection for data processing relative to the detection of light precipitation was examined using four different combinations of reference and SUT accumulation thresholds (four 'cases' in Table 8) for the unshielded MPS TRwS405 gauge under test at the Marshall site. Contingency results, probabilities of detection (POD), and false alarm rates (FAR) are presented for each case in Table 9. A quantitative comparison of the amounts reported in each case is beyond the scope of this assessment.

Table 8: Reference and SUT thresholds in each case for light precipitation detection assessment.

Case	Reference threshold (mm/30 min)	SUT threshold (mm/30 min)
1	0.25	0.25
2	0.1	0.1
3	0.25	No threshold
4	0.25	0

Table 9: Contingency results, probability of detection, and false alarm rate for each case in light precipitation detection assessment.

Case	Number of events				Skill score (%)	
	YY	YN	NY	NN	POD	FAR
1	302	94	35	19616	76.3	10.4
2	395	41	154	19457	90.6	28.1
3	396	0	19651	0	100	98.0
4	396	0	19651	0	100	98.0

6.5. Assessment of events when the reference and the SUT do not agree on the occurrence of precipitation

Assessment intervals during which the site reference and SUT do not agree on the occurrence of precipitation – namely, the YN and NY cases (Section 4.1.1) – are characterized for each test gauge using histograms in Figures 12 and 13. The histograms include accumulated precipitation reported by the reference and SUT (0.25 mm/30 min threshold for both), precipitation intensity as reported by the reference, and corresponding site conditions.

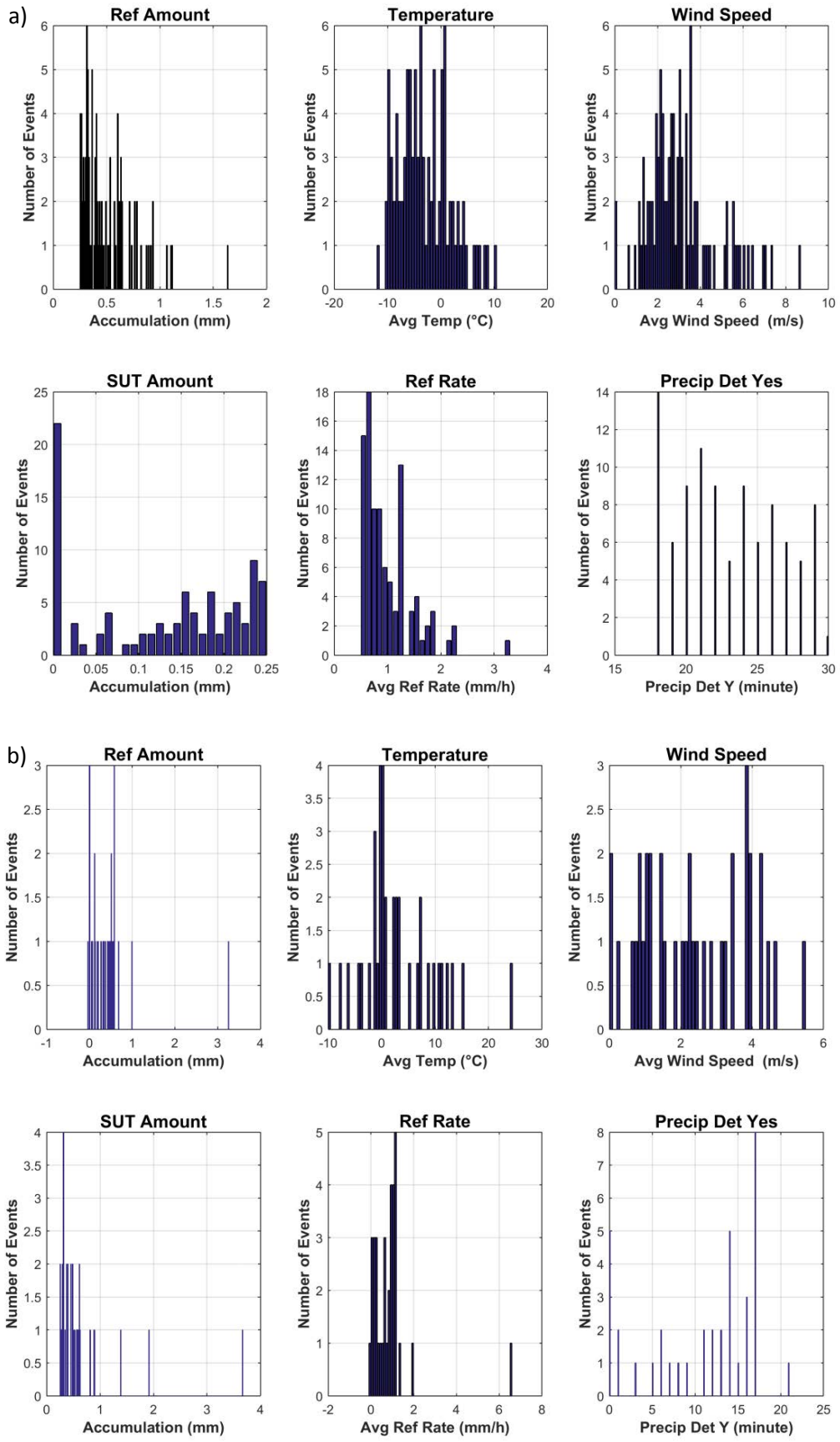


Figure 12: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for (a)YN events (94 total) and (b) NY events (35 total) for the unshielded MPS TRwS405 at Marshall over the test period.

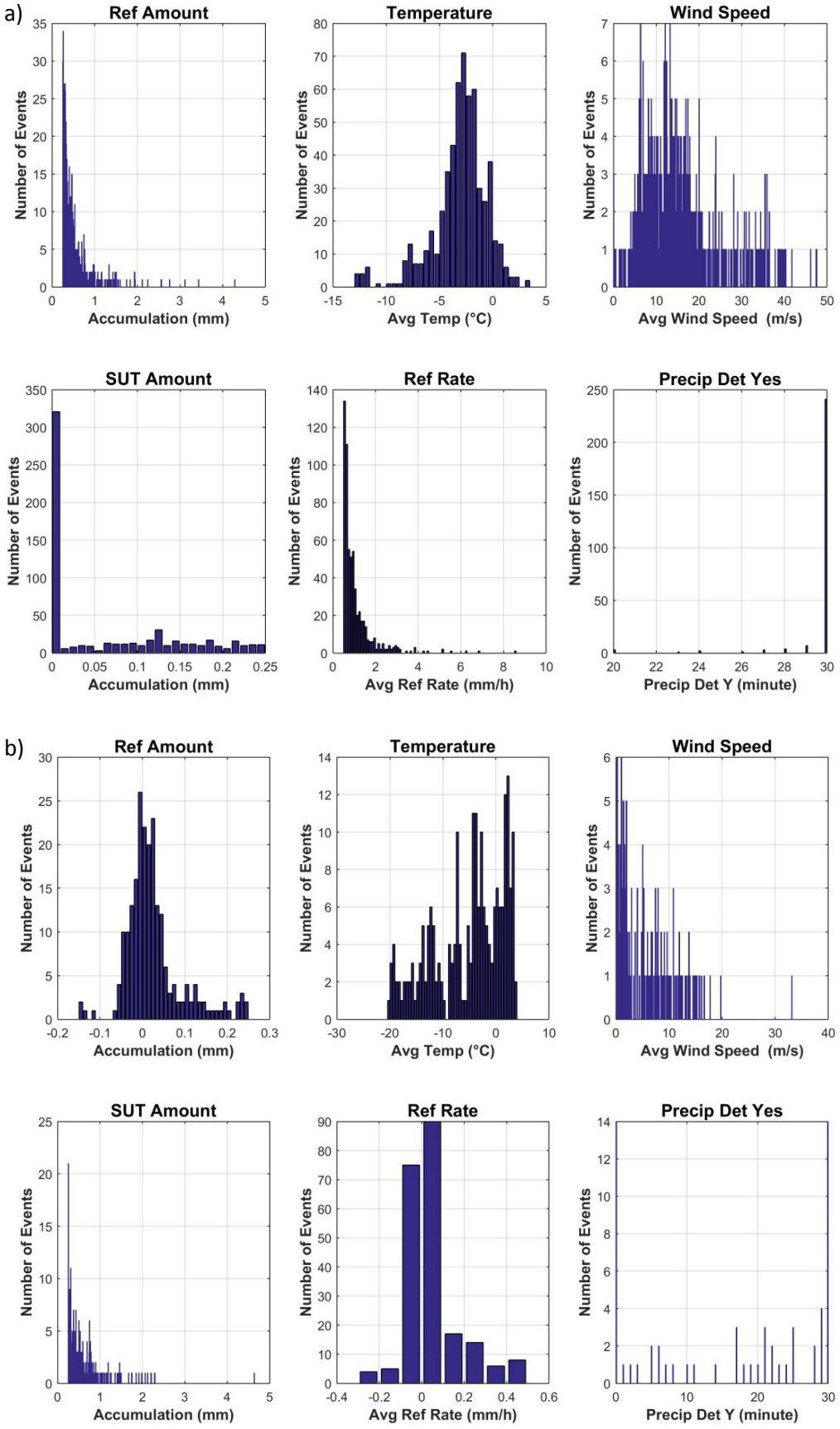


Figure 13: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for (a)YN events (605 total) and (b) NY events (219 total) for the unshielded MPS TRwS405 at Haukeliseter over the test period.

7) Interpretation of Results

7.1. Operating conditions

The full ranges of conditions under which the test gauges at each site were operated are illustrated in Figure 3. The conditions relevant to gauge operation are as follows:

- Temperatures between approximately -27 °C and 26 °C;
- Precipitation intensity within approximately 12 mm/hr.

These conditions fall within the manufacturer's specified operating conditions of -40 °C to 70 °C for temperature and 120 mm/min for maximum precipitation intensity.

The conditions at each site during precipitation events are shown in Figure 4. Of particular note are the mean wind speeds during precipitation events at Haukeliseter, which extend to approximately 20 m/s. The mean wind speeds during precipitation events at Marshall are generally within about 7 to 8 m/s.

7.2. Performance over the range of operating conditions

7.2.1. Non-precipitating conditions

The MPS TRwS405 data are processed using an on-board, manufacturer-developed algorithm (outlined in Section 2), which reduces the noise in gauge reports in the absence of precipitation. This is evident in Figure 5, which shows that the PDFs of gauge reports during 30 minute periods without precipitation are dominated by 0 mm reports. The standard deviations of test gauge outputs in non-precipitating conditions at Haukeliseter and Marshall are 0.06 mm and 0.03 mm, respectively (Table 4). While the PDFs of SUT reports in non-precipitating conditions indicate that reports of, or close to, 0 mm have the highest probability, there are a number of reports of higher magnitude, which appear to be more pervasive in the data from the test gauge at Haukeliseter (Figures 6 and 7). The maximum reported values are 1.5 mm and 3.7 mm for the test gauges at Haukeliseter and Marshall, respectively (Table 4), both of which exceed greatly the detection threshold for precipitation (0.25 mm) used in this assessment.

The magnitude of gauge responses in the absence of precipitation can be used to identify a detection threshold that minimizes the detection of false precipitation while enhancing the detection of light precipitation. This threshold is considered to be three times the standard deviation of the average gauge response during 30 minute non-precipitating periods. Based on the present test gauge results, this minimum detection threshold for unshielded MPS TRwS405 gauges is determined to be 0.13 mm.

7.2.2. Precipitating conditions

7.2.2.1. Ability to detect and report precipitation

The skill scores and contingency results for the MPS TRwS405 test gauges (Figure 8 and Table 5, respectively) indicate that the ability to detect and report precipitation was notably lower for the gauge at Haukeliseter relative to that at Marshall. Inspection of the contingency results for Haukeliseter (Table 5) shows that there was a greater number of 30 minute events during which the reference and SUT did not agree on the presence of precipitation (824 total YN and NY cases) than during which they agreed on the presence of precipitation (357 YY cases). This resulted in a low POD (~ 40%), high FAR (~40%), low B (~ 60%), and low HSS (~40%). In contrast, the test gauge at Marshall agreed with the reference on the detection of precipitation to a greater extent (302 YY cases compared to 129 total YN and NY cases in Table 5), resulting in a higher POD (~80%), lower FAR (~10%), B closer to 100% (~85%; indicates better agreement between the reference and SUT), and higher HSS (~ 80%). These differences likely result from differences in the configurations of test gauges, relative siting of test gauges and reference configurations, and specific conditions experienced (notably, the magnitude of wind speeds during precipitation events) at each site.

7.2.2.2. Ability to report accumulated precipitation

The results presented in Figures 9 and 10, which are based on 30 minute events during which the reference and test gauge both detect precipitation (YY cases), illustrate the influence of wind speed and precipitation type on gauge catch efficiency. The discussion below will focus on snow events; the number of rain events during winter is limited, and the results for mixed events are variable due to the variability in the size and density of precipitation within the mixed regime, as well as the potential for transitions between phases. For solid precipitation events, the median catch efficiency for each test gauge decreases with increasing mean wind speed. The decrease is more gradual for the test gauge at Haukeliseter, remaining above 0.4 for mean wind speeds up to 10 m/s, while that for the test gauge at Marshall falls below 0.4 for mean wind speeds between 3 m/s and 4 m/s.

Root mean square error values were computed from all 30 minute events during which each test configuration and the corresponding reference configuration both detected precipitation. Values are shown in Figure 11 and Table 6, and can be considered to represent the absolute uncertainty of each test configuration relative to the reference configuration in liquid, mixed, and solid precipitation, and in all precipitation types (note that the relative proportions of events of each phase differ by site). The RMSE values for the unshielded test gauges at both sites in all precipitation types are similar: 0.418 mm/30 min and 0.393 mm/30 min for the gauges at Haukeliseter and Marshall, respectively. The largest phase-specific RMSE values are observed for solid precipitation – 0.465 mm/30 min for the gauge at Haukeliseter and 0.421 mm/30 min for the gauge at Marshall – owing to the larger wind-induced undercatch for unshielded gauges relative to the reference configurations in DFIR-shields.

The overall catch ratio – computed from the total reference and SUT accumulation from all YY cases over the duration of formal tests – is provided for each test configuration in Table 7. The value for the test gauge at Haukeliseter (0.73) is lower than that for the test gauge at Marshall (0.77), and can likely be explained, at least in part, by the higher wind speeds during precipitation events at

Haukeliseter (Figure 4). Lowering the precipitation detection threshold for the test gauges from 0.25 mm/30 min to 0.1 mm/30 min decreases the overall catch ratio by about 4% to 7%.

7.2.2.3. Ability to detect light precipitation

The detection thresholds and results presented in Tables 8 and 9, respectively, indicate that decreasing or removing the SUT detection threshold (while maintaining the reference detection threshold at 0.25 mm) increases both the Probability of Detection and False Alarm Rate to close to 100%; the corresponding Heidke Skill score values (not shown) approach 0%, indicating minimal detection skill with these specific thresholds. Decreasing both the reference and SUT detection thresholds to 0.1 mm increases the POD from 76% to 91%, while increasing the FAR from 10% to 28%. These results suggest that lowering the threshold enhances the detection of light precipitation, but also leads to the detection of more non-zero reports in the absence of precipitation.

7.2.3.4. Assessment of events when the reference and SUT do not agree on the occurrence of precipitation

The YN ('miss') cases, when the reference detects a precipitation event and the SUT does not, and NY ('false alarm') cases, when the SUT detects a precipitation event and the reference does not, are characterized for each test configurations in Figures 12 and 13. The YN cases for both test gauges (Figure 12a and 13a) indicate a number of events with accumulation between 0 and 0.25 mm; as many of the events detected by the reference configurations had accumulations just above the 0.25 mm threshold, wind effects could reduce the magnitude of accumulation reports from the unshielded test configurations below the threshold. Both test configurations also reported a number of zero accumulation events; while some of these may result from wind effects, the large numbers of zero accumulation reports in Figures 12a and 13a suggest that there may be cases in which the gauge processing algorithm erroneously filtered out light precipitation. The majority of false alarm cases reported by the test gauge at Marshall were cases in which the precipitation detection threshold of 18 Y reports/30 min was not met (Figure 12b), while those reported at Haukeliseter were mostly cases in which the reference accumulation threshold of 0.25 mm/30 min was not met (Figure 13b).

8) Maintenance

Gauge calibration: each site completed the gauge field calibration and verification as per manufacturer recommendations, at least once a year or following the emptying of the gauge. The calibration records have been stored by each site host.

9) Performance Considerations

9.1. Data processing

The MPS TRwS405 uses on-board processing to generate precipitation reports. As a number of cases were identified in which (1) the gauge reported accumulation when an independent precipitation detector reported no precipitation was occurring and (2) the gauge reported zero accumulation when the reference configuration detected a precipitation event, it is recommended that users be aware of the potential for false reports, and exercise caution when interpreting gauge data.

Disclosure of the internal processing logic by the manufacturer would better enable users to understand the gauge output data, and how they change with modifications to the processing approach over time. The traceability of processing algorithms over time is an important consideration from a climate perspective.

9.2. Gauge configuration

MPS TRwS405 gauges were tested only in unshielded configurations; it is therefore difficult to make specific recommendations with respect to shielding. Based on experience with other test gauges, shielding of MPS TRwS405 gauges is recommended to mitigate the effects of wind-induced undercatch. Wind shields should be mounted separately from the gauge post to eliminate the potential for wind-induced vibration of the shield assembly to impact gauge measurements.

It is recommended that all field configurations be fully tested and validated prior to use, in order to ensure that physical and signal interferences are not degrading the gauge signal. This would include the confirmation of grounding tailored to the soil conditions, sturdiness of foundation and mounting, and signal conditioning.

9.3. Ancillary measurements and adjustments

The application of adjustment functions to measurements is strongly recommended to account for reductions in catch efficiency as the wind speed increases. Ancillary measurements of wind speed (preferably at gauge orifice height) and air temperature are required for the application of adjustment functions. In addition, ancillary measurements from a sensitive precipitation detector that is independent from the gauge are recommended to help distinguish precipitation events from false reports due to noise.

WMO-SPICE Instrument Performance Report
OTT Pluvio²

1) Technical Specifications (from manufacturer provided documentation)

Instrument model:	OTT Pluvio^2
Physical principle:	Weighing gauge (WG) based on catchment principle. High-precision stainless steel load cell used to weigh the bucket contents. An integrated temperature sensor compensates for temperature changes in the balance system.
Capacity:	1500 mm
Collecting area:	200 cm^2 (circular orifice)
Operating temperature range:	-40 °C to 60 °C
Measurement uncertainty:	Relative accuracy of ± 5% for Accumulated NRT output, ± 0.2% for Bucket RT output; absolute accuracy of ± 0.1 mm (all values as defined by manufacturer)
Resolution:	0.01 mm (for accumulation, as defined by manufacturer)
Heater:	Integrated orifice rim heating. Supply voltage for heating between 20 V and 28 V DC (typically 24 V DC). Maximum capacity heating capacity at low temperatures is approximately 53 W.

Figure 1: OTT Pluvio^2 test gauges at (a) Marshall (unshielded), (b) CARE (unshielded and single-Alter shielded), (c) Formigal (single-Alter shielded) and (d) Formigal (unshielded).

2) Data Output Format

Gauge data output*: (*As outlined in manual provided by manufacturer)	Every 6 seconds, the OTT Pluvio^2 calculates the bucket content (Bucket RT) using multiple raw values. These 6-second values for the precipitation intensity are added using a proprietary algorithm to derive the accumulated precipitation amount in various forms (see below). Internal data processing is performed using a proprietary algorithm, to account for temperature and wind effects. The difference between the current bucket content and the previous one gives the precipitation intensity in mm/min or mm/h. Depending on the specific algorithm applied, the data are available in real time (RT: within one minute of the measurement) or near real time (NRT: with a 5 minute delay). If very fine precipitation is involved (< 0.1 mm/min), the output is delayed by up to 65 minutes. The following data products are available in the Pluvio^2 output message: ▪ Intensity RT (fixed output interval: 1 minute; units: mm/hr) ▪ Accumulated RT/NRT (since the last measurement sample; units: mm) ▪ Accumulated NRT (since the last measurement sample; units: mm) ▪ Accumulated total NRT (since the last reset; units: mm) ▪ Bucket RT (current unfiltered bucket content; units: mm) ▪ Bucket NRT (current filtered bucket content; units: mm) ▪ Load cell temperature (units: °C) ▪ Status (since the last measurement sample)
Firmware version (as tested)	V1.30.0 NOTE: Since the start of the SPICE formal tests, OTT Hydromet has updated the firmware; however, the tests continued with the initial firmware version, to ensure consistency in data. Some of the findings in this report may no longer be representative to the firmware version used at the time of reading this report, but remain important from a historical data perspective.

3) SPICE test configuration

Shield configuration(s):	Unshielded (UN), single-Alter shield (SA), and Tretyakov shield (Tret)
Test site(s):	CARE (Canada); Formigal (Spain); Haukeliseter (Norway); Marshall (USA); Sodankylä (Finland); and Weissfluhjoch (Switzerland)
Instrument provider(s):	OTT Hydromet provided two Pluvio^2 instruments, which were tested at Marshall (Tret) and at Sodankylä (UN). Host organizations: all six test sites tested one pair of their own Pluvio^2 gauges (one SA, one UN); these were initially configured as part of the SPICE R3 reference, except at Haukeliseter, where only one Pluvio^2 (SA) was tested.
Heating configuration(s):	All Pluvio^2 gauges tested employed the integrated orifice rim heating

A map of test site locations is provided in Figure 2.

3.1. Note on terms and acronyms used

Throughout this document, the following notations were used to identify the R2 reference (Ref) and senor under test (SUT) configurations:

Reference: 'DFIR' and 'DFAR' are used interchangeably for the R2 reference configurations. 'DFIR' refers to an automated gauge installed in a DFIR-fence, while 'DFAR' refers more explicitly to the Double-Fence Automated Reference configuration. Different sites employed different gauges in the R2 reference configurations, which are differentiated in plot legends as follows: 'Geo600' for 600 mL capacity Geonor T-200B3 gauges; 'Geo1000' for 1000 mL capacity Geonor T-200B3M gauges; and 'Plv2BktRT' for the Bucket RT output from 1500 mL capacity OTT Pluvio^2 gauges.

Sensors under test: 'SA' denotes single-Alter shielded test configurations, 'Tret' denotes Tretyakov-shielded test configurations, and 'UN' denotes unshielded test configurations. 'UT' is used to denote sensors 'Under Test' that have been submitted by the manufacturer.

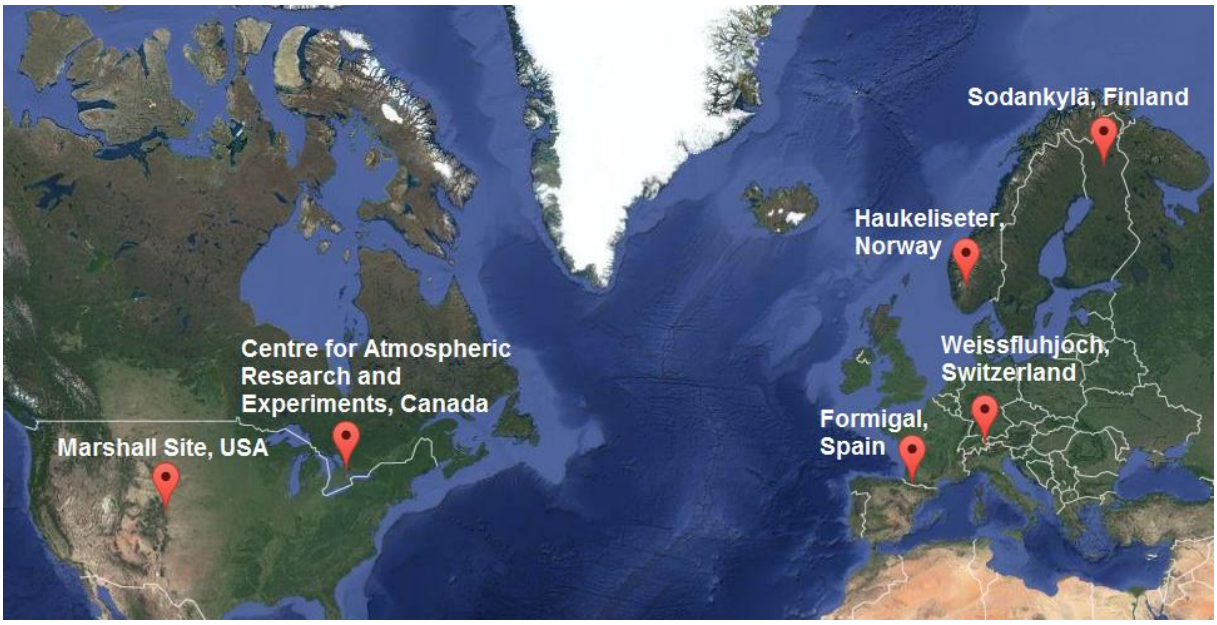


Figure 2: Map of SPICE sites where OTT Pluvio^2 gauges were tested.

A summary of the configuration of instruments as tested, the duration of tests and availability of data reflected in these results, and the ancillary measurements used, by site, is available in Tables 1, 2, and 3, respectively.

Table 1: Summary of gauge configurations and data output, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	CARE	Formigal	Haukeliseter	Marshall	Sodankylä	Weissfluhjoch
Field configuration (shield)	SA, UN	SA, UN	SA	SA, UN, Tret	SA, UN	SA, UN
Height of Installation (gauge rim)	2.0 m	3.5 m	4.5 m	2.0 m	1.5 m	3.5 m
Heating	Manufacturer recommended algorithm with integrated rim heating					
Antifreeze	60% methanol, 40% propylene glycol	Ethylene glycol	60% methanol, 40% ethylene glycol	60% methanol, 40% propylene glycol	Meltium (Potassium formate, 49-51% solution)	60% methanol, 40% propylene glycol†
Oil	Bayoil (Season 1) Isopar (Season 2)	5W30 motor oil	Hydraulic oil, Statoil Hydraway 15LT	Automatic Transmission Fluid (ATF)	N/A	Linseed oil (Season 1) Isopar (Season 2)
Data output frequency	6 sec	1 min	1 min	6 sec	6 sec	1 min
Data QC	SPICE QC methodology					
Data temporal resolution	1 minute					
Processing interval for data analysis	30 minutes					

† For the test gauge at Weissfluhjoch, the 60% methanol, 40% propylene glycol mixture was used during Season 2, only. A mixture of 75% propylene glycol, 25% water was used during Season 1.

Table 2: Data availability, by measurement season and site.

Measurement season	CARE	Formigal	Haukeliseter	Marshall	Sodankylä	Weissfluhjoch
Season 1 (Oct. 2013 – Apr. 2014)	✓	X	✓	✓	✓	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓	✓	✓	✓	✓	✓

Table 3: Summary of reference and ancillary measurements, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	CARE	Formigal	Haukeliseter	Marshall	Sodankylä	Weissfluhjoch
R2 Site Reference	Geonor T-200B3 600 mm (DFAR)	OTT Pluvio^2 Bucket RT (DFAR)	Geonor T-200M3 1000 mm (DFAR)	Geonor T-200B3 600 mm (DFAR)	OTT Pluvio^2 Bucket RT (DFAR)	OTT Pluvio^2 Bucket RT (DFAR)
R2 Precip Detector	Thies LPM (DFAR)	Thies LPM (DFAR)	Thies LPM X2 (Site*, 2013-2014), Thies LPM X5 (DFAR, 2014-2015)	Thies LPM (Site*)	OTT Parsivel2 (Site*, 2013-2014), OTT Parsivel2 (DFAR, 2014-2015)	Thies LPM (DFAR)
Ancillary Temp Sensor	Vaisala HMP155 (SS$^+$, 1.5 m)	Thies Thermo Hygrometer (4 m)	PT100 (4.5 m)	MetOne, 060A-2/062, 2144-L (2 m)	Vaisala HMP155 (2 m)	Meteolabor AG VTP6 Thygan (5 m)
Ancillary RH Sensor	Vaisala HMP155 (SS$^+$, 1.5 m)	Thies Thermo Hygrometer (4 m)	Campbell Scientific PWS100 (6 m)	Campbell Scientific CS500 (2m)	Vaisala HMP155 (2 m)	Meteolabor AG VTP6 Thygan (5 m)
Ancillary Wind Sensor	Vaisala NWS425 (2 m)	RM Young Alpine 05103-45 (10 m)	Thies ultrasonic anemometer 3D (4.5 m, 2013-2014) RM Young Wind Monitor 05103 (4.5 m, 2014-2015)	RM Young Wind Monitor 05103 (2 m)	Thies acoustic 2D wind sensor (3.5 m)	RM Young Wind Monitor 05103 (5.5 m)

*A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR-fence; however, in cases where a more sensitive detector is available outside of the DFIR-fence, or there are issues with the detector within the DFIR-fence, a precipitation detector elsewhere on the site can be employed.

$^+$ SS denotes that the sensor is installed inside a ventilated Stevenson Screen.

3.2. Pluvio^2 data analysis strategy

As noted in Section 2, the Pluvio^2 weighing gauge outputs several data products of varying complexity, resulting from the application of the manufacturer's proprietary algorithm. As the Pluvio^2 was used as part of the SPICE field reference system, the baseline data product selected for the reference system was the Bucket RT dataset, which is the closest to the physical measurement, and least likely to be impacted as the internal processing algorithm is modified, over time. As the internal algorithm is updated, the derived products will evolve, and their traceability in time could be achieved through reference to the Bucket RT data.

For this reason, the approach for deriving the results in this report is as follows:

- The reference dataset was derived from the Bucket RT data; this is consistent for all sites at which the SPICE R2 field reference system employed a Pluvio^2 gauge;
- The assessment of the Pluvio^2 gauges provided to SPICE by the manufacturer, which were tested on the Marshall and Sodankylä sites, was conducted using both the Bucket RT and Accumulated NRT data;
- For all Pluvio^2 instruments provided by the host organizations, the results are reported only for the Bucket RT data;
- For comparison, results for the Accumulated NRT dataset from one of the instruments provided by the host organizations, a Pluvio^2 instrument in a single-Alter shield at the CARE site, are included in this report. This additional analysis was performed to link the results of test gauges provided by the manufacturer and those provided by the host organizations.

Given the large number of Pluvio^2 instruments tested during SPICE, the results in this report are organized by site, and by configuration. Where both the Bucket RT and Accumulated NRT data were assessed, the results are presented together, by gauge.

Note: In the figures and tables that follow, the Bucket RT data are also referred to as 'BktRT' and the Accumulated NRT data are also referred to as 'Accum NRT' and 'AccNRT.'

4) Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6.1 of the WMO-SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

4.1.1.1. Characterization of performance in non-precipitating conditions

The assessment data are derived over 30 minute intervals during which the precipitation detector in the R2 reference configuration reports 0 minutes of precipitation. The accumulation over these intervals (accumulation in minute 30 – accumulation in minute 1), representing the variability of the gauge response due to wind, evaporation, temperature, etc., is recorded, along with the mean wind speed, and the change in temperature (temperature in minute 30 – temperature in minute 1).

4.1.1.2. Assessment of ability to detect and report accumulation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute assessment intervals:

- Ref 'Yes' : R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No' : R2 weighing gauge < 0.25 mm AND/OR precip detector recording < 18 min of precip;
- SUT 'Yes' : SUT accumulation ≥ 0.25 mm;
- SUT 'No' : SUT accumulation < 0.25 mm.

4.1.1.3. Assessment of ability to detect light precipitation

The data for this component of the assessment are derived in a similar manner as those in Section 4.1.1.2, but with different combinations of thresholds for the reference and/or SUT 'Yes' and 'No' conditions. These different threshold 'cases' have been selected to demonstrate the impact of the thresholds used in data derivation on the detection of light precipitation.

4.1.2. Skill score assessment

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores, as detailed in Section 3.6.1.3 of the WMO-SPICE Final Report.

For the assessments considered in this report, the ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

5) Environmental conditions

The environmental conditions at each site over the duration of the test period are expressed as probability density functions (PDFs) of mean air temperature, mean relative humidity, mean wind speed, vector mean wind direction, and precipitation rate for each component 30 minute assessment interval in Figure 3. The same parameters are also shown for all assessment intervals during which the site reference configuration detected precipitation (i.e. all Ref 'Yes' cases) in Figure 4.

The precipitation percentage represents the number of minutes of precipitation during a 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration, expressed as a percentage. PDFs of precipitation percentage are also included in Figures 3 and 4.

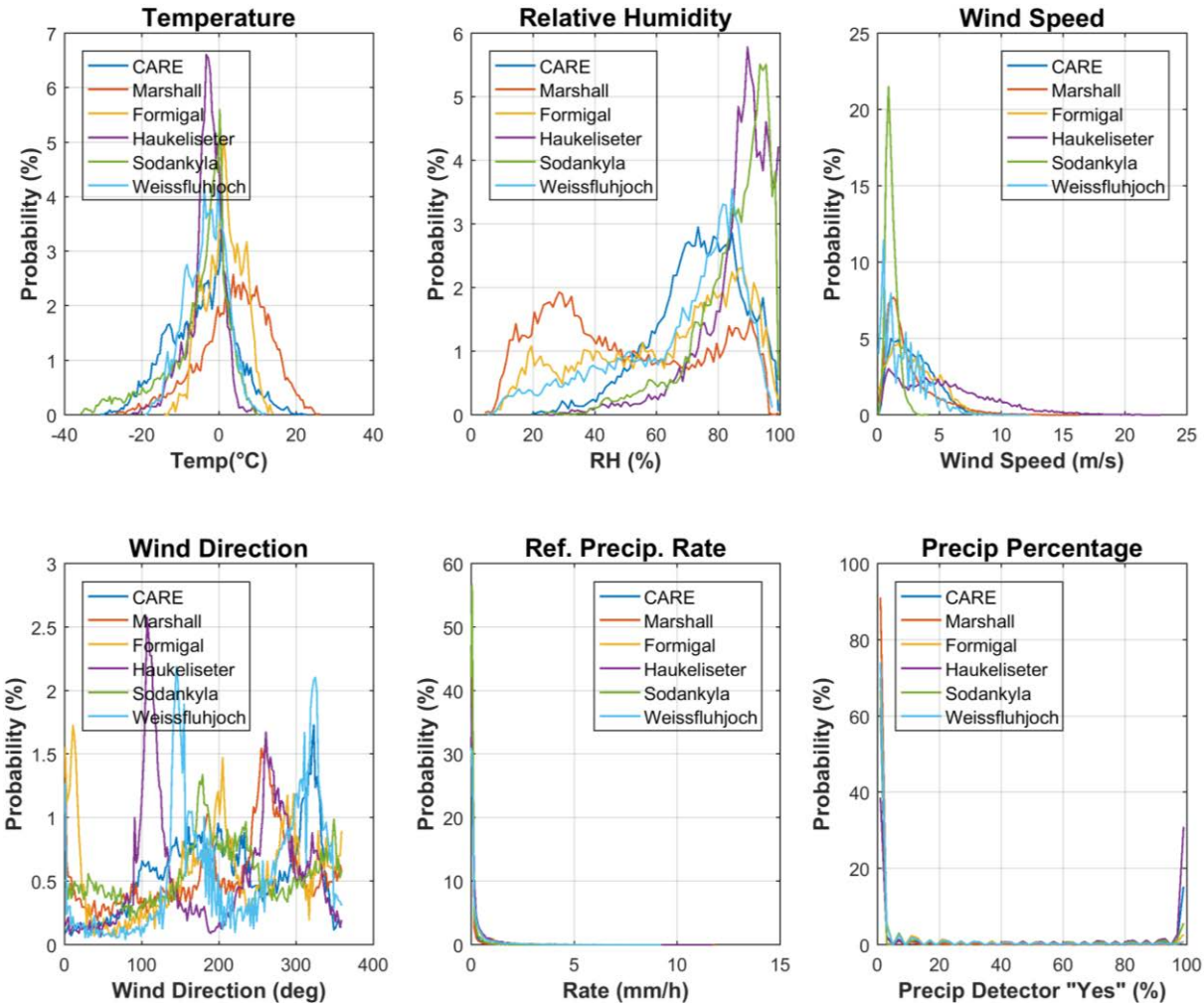


Figure 3: Summary of environmental conditions at sites with OTT Pluvio^2 test gauges over the entire duration of formal tests, as per Table 2.

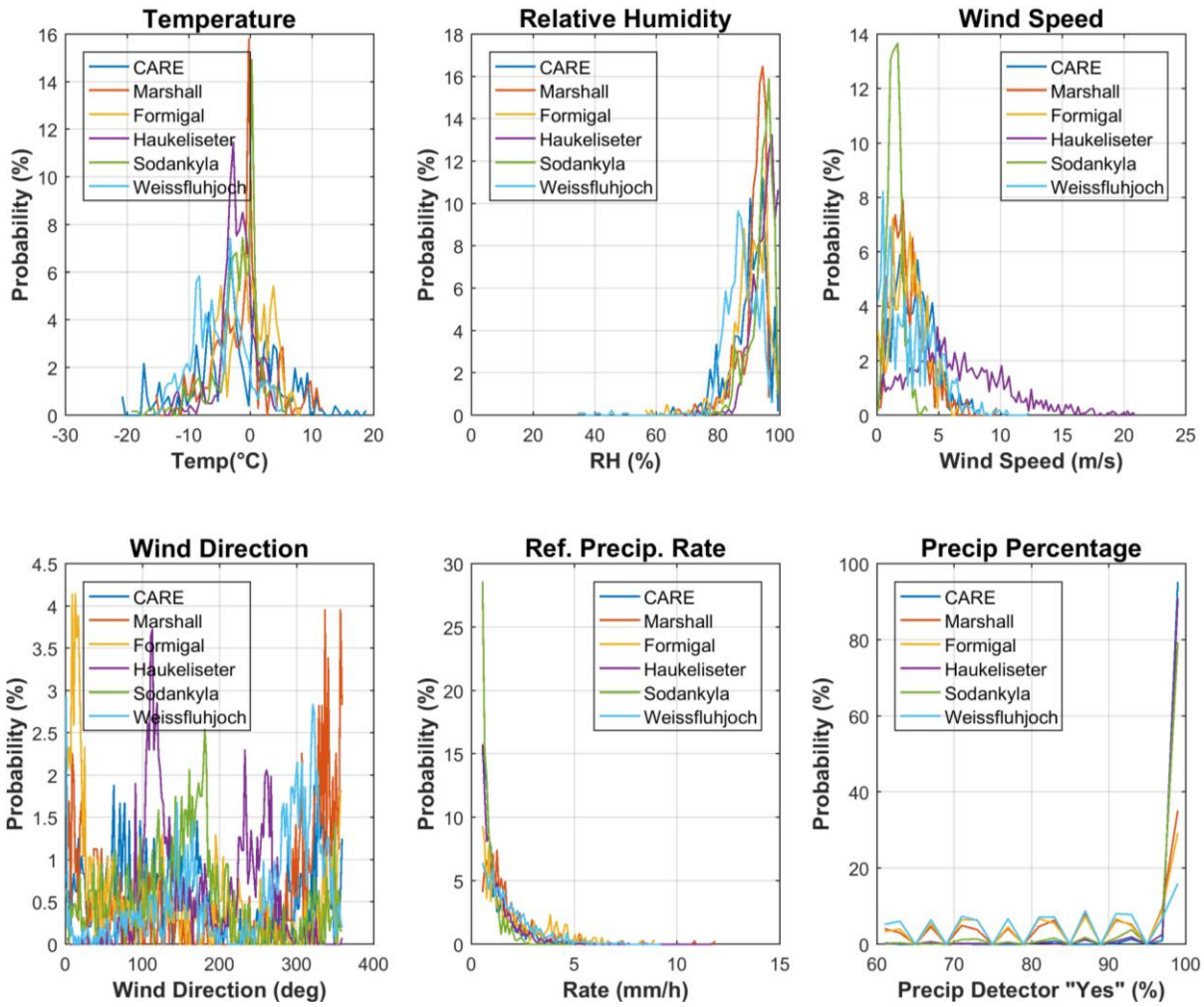


Figure 4: Summary of environmental conditions at sites with OTT Pluvio^2 test gauges during precipitation events (as defined by R2 reference), over the duration of formal tests.

6) Evaluation of the ability to perform over the range of operating conditions

6.1. Characterization of SUT performance in non-precipitating conditions

The response of the SUT in the absence of precipitation was examined as defined in Section 4.1.1.1. The results are presented below, reflecting the distribution of the sensor response and its variability with wind and temperature, as measured during 30 minute assessment intervals.

6.1.1. Overall variability of SUT response

The overall variability of the Bucket RT and Accumulated NRT outputs in non-precipitating conditions are shown as probability density functions for applicable test configurations in Figures 5 and 6, respectively. The corresponding PDFs for the reference configuration at each test site are provided for comparison.

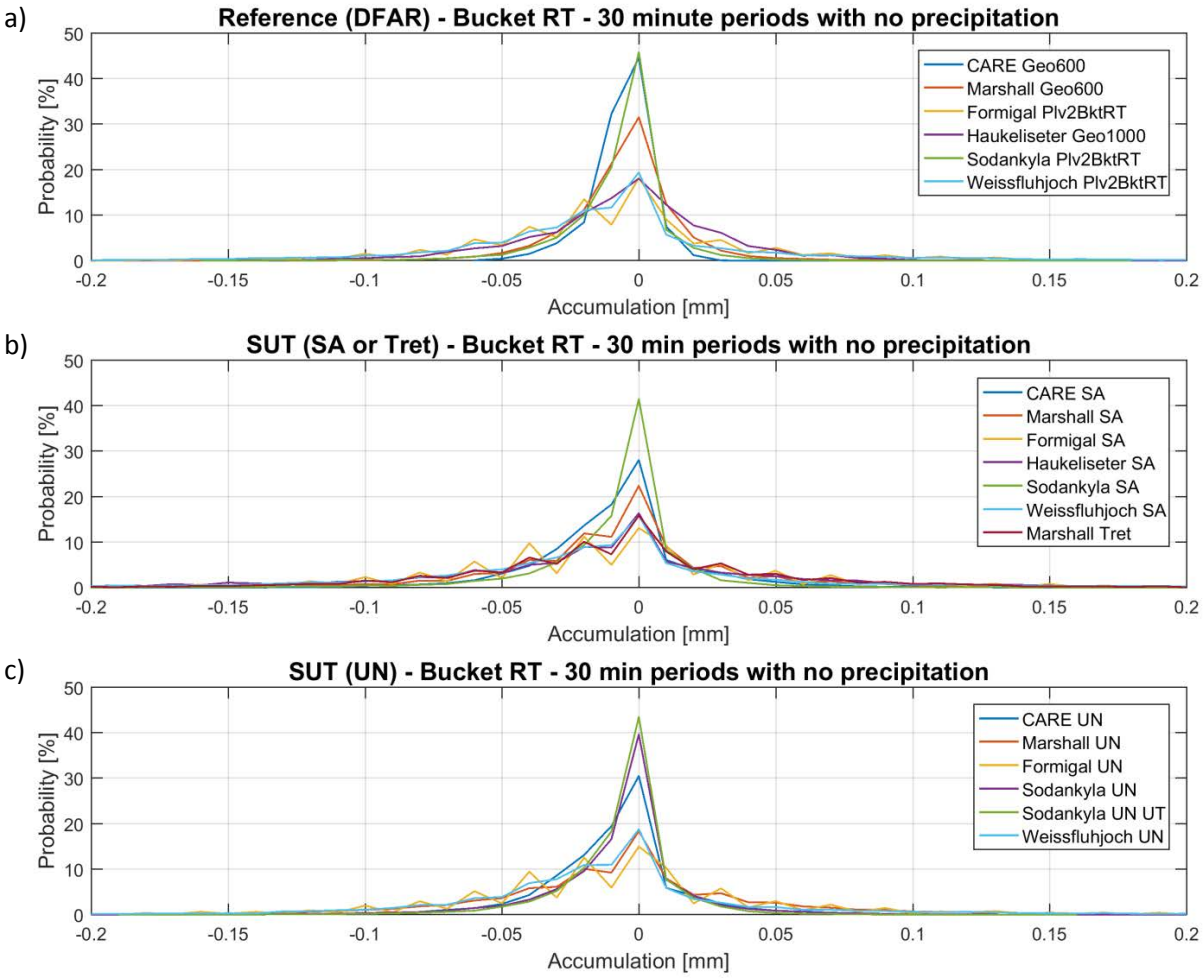


Figure 5: Probability density functions of the Bucket RT output (accumulation over 30 minute assessment intervals) in non-precipitating conditions for (a) the R2 reference configurations (DFAR), (b) Pluvio^2 test gauges in SA- and Tret-shielded configurations, and (c) Pluvio^2 test gauges in unshielded configurations.

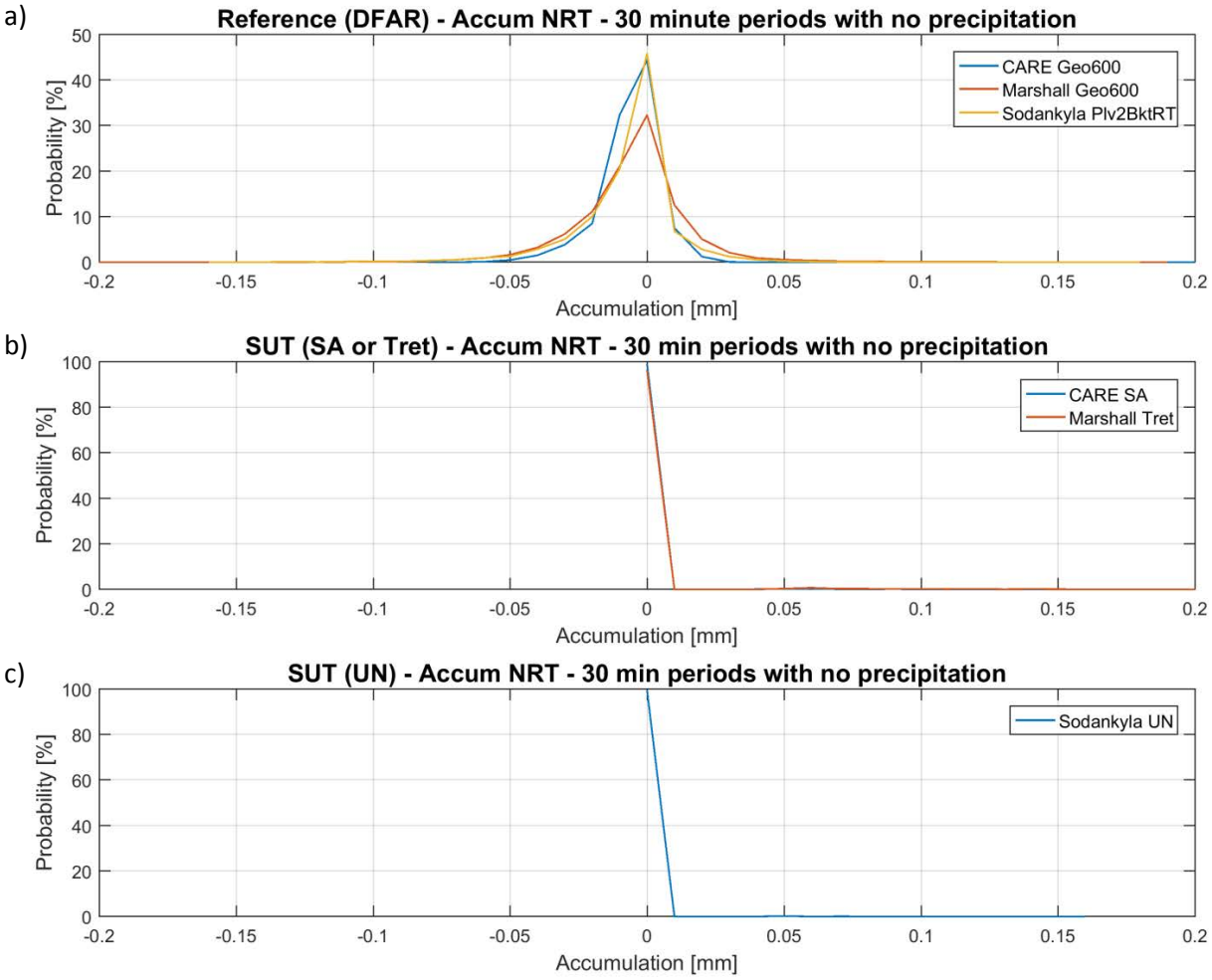


Figure 6: Probability density functions of the Accumulated NRT output (accumulation over 30 minute assessment intervals) in non-precipitating conditions for (a) the R2 reference configurations (DFAR), (b) Pluvio^2 test gauges in SA- and Tret-shielded configurations, and (c) Pluvio^2 test gauges in unshielded configurations.

The statistics of the output signal (accumulation over 30 minute assessment intervals) for the reference and SUT(s) at each site are provided in Table 4.

Table 4: Statistics of the R2 reference gauge (DFAR) and SUT output signal during non-precipitating conditions at sites testing OTT Pluvio^2 gauges, as plotted in Figures 5 and 6. Values for Accumulated NRT output data are indicated in red text.

Site	Config.	Output data	Avg output signal (mm)	Standard deviation (mm)	Maximum output signal (mm)	Minimum output signal (mm)	Number of assessment intervals
CARE	DFAR	N/A	-0.001	0.011	0.204	-0.059	9371
	SA	BucketRT	-0.005	0.030	0.220	-0.260	9371
	SA	AccumNRT	0.001	0.008	0.250	0.000	9482
	UN	BucketRT	-0.005	0.030	0.610	-0.330	9493
Formigal	DFAR	BucketRT	-0.002	0.057	0.380	-0.270	3380
	SA	BucketRT	-0.003	0.078	0.600	-0.410	3380
	UN	BucketRT	-0.002	0.070	0.630	-0.380	3373
Haukeliseter	DFAR	N/A	0.000	0.034	0.202	-0.145	2773
	SA	BucketRT	-0.005	0.087	0.700	-0.480	2773
Marshall	DFAR	N/A	-0.001	0.022	0.196	-0.207	17919
	SA	BucketRT	-0.002	0.047	0.400	-0.290	17919
	UN	BucketRT	-0.001	0.067	0.900	-0.860	16189
	Tret*	BucketRT	-0.001	0.067	0.640	-0.590	16226
	Tret*	AccumNRT	0.004	0.030	0.660	0.000	16242
Sodankylä	DFAR	BucketRT	-0.004	0.020	0.190	-0.160	12015
	SA	BucketRT	-0.004	0.046	1.100	-1.960	12015
	UN	BucketRT	-0.003	0.038	1.210	-0.970	11980
	UN UT*	BucketRT	-0.004	0.022	0.160	-0.190	12021
	UN UT*	AccumNRT	0.000	0.006	0.160	0.000	12015
Weissfluhjoch	DFAR	BucketRT	-0.005	0.068	1.200	-0.750	11508
	SA	BucketRT	-0.007	0.104	0.910	-0.670	11508
	UN	BucketRT	-0.006	0.069	0.660	-0.620	11511

* Test gauge provided by manufacturer.

6.1.2. Variability of SUT response as a function of temperature

The variability of the Bucket RT and Accumulated NRT output for applicable test configurations in the absence of precipitation is plotted as function of the temperature difference over each assessment interval in Figures 7 and 8, respectively. The temperature difference is defined as the difference in temperature between the end (minute 30) and beginning (minute 1) of the assessment interval. The corresponding plots for the reference configurations are provided for comparison.

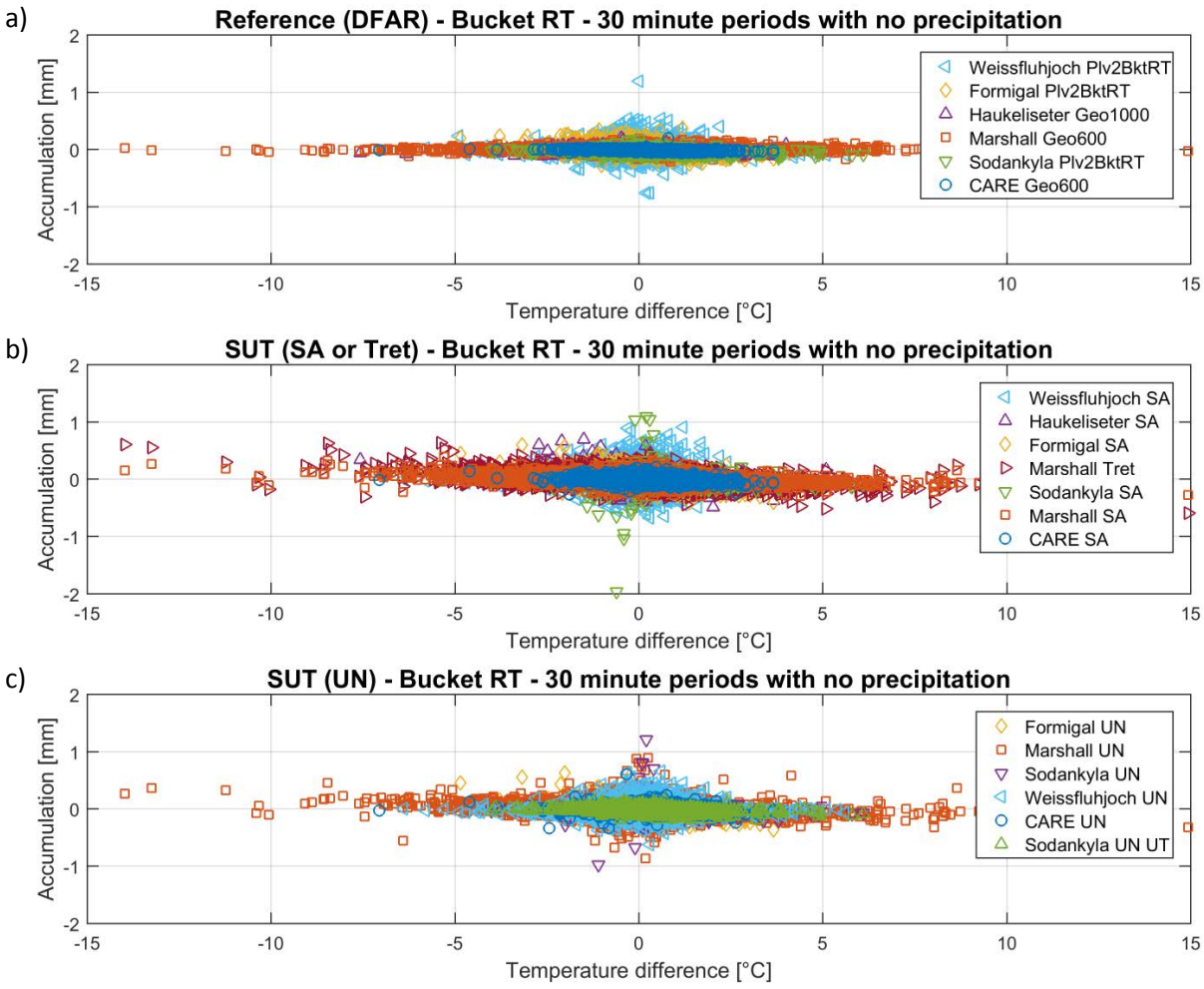


Figure 7: Variability of Bucket RT output (accumulation over each 30 minute assessment interval) as a function of the temperature difference over the interval in non-precipitating conditions for (a) the R2 reference configurations (DFAR), (b) Pluvio^2 test gauges in SA- and Tret-shielded configurations, and (c) Pluvio^2 test gauges in unshielded configurations.

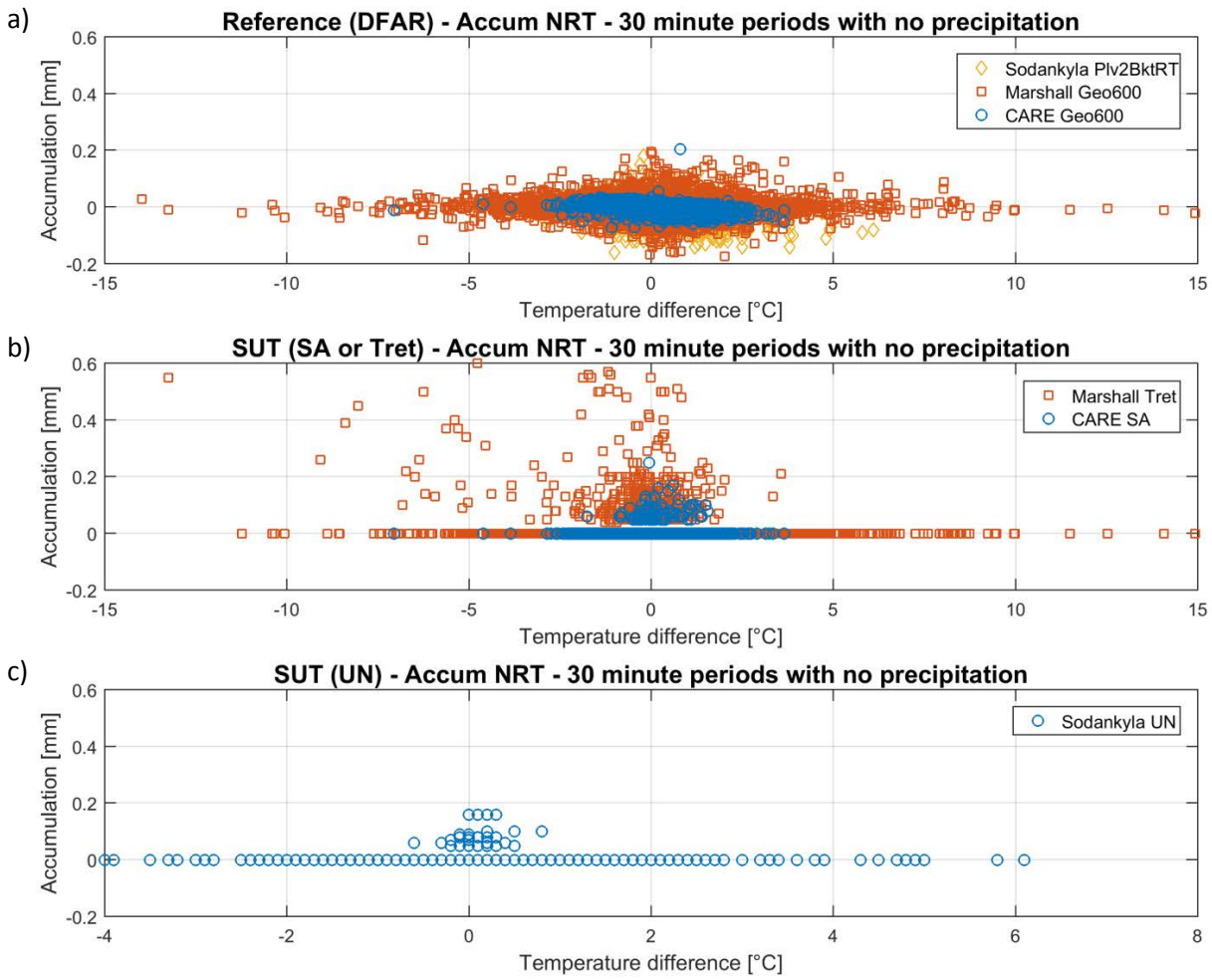


Figure 8: Variability of gauge output (accumulation over each 30 minute assessment interval) as a function of the temperature difference over the interval in non-precipitating conditions for (a) the R2 reference configurations (DFAR), and for Accumulated NRT output from (b) Pluvio^2 test gauges in SA- and Tret-shielded configurations, and (c) Pluvio^2 test gauges in unshielded configurations. Note that the temperature axis scale for the unshielded Pluvio^2 gauge in (c) is truncated relative to those in (a) and (b) for clarity.

6.1.3. Variability of SUT response as a function of wind speed

The variability of the Bucket RT and Accumulated NRT output for applicable test configurations in the absence of precipitation is plotted as function of the mean wind speed for each assessment interval in Figures 9 and 10, respectively. Here, the signal variability is represented as the standard deviation (STD) of the gauge accumulation output over each 30 minute interval. The corresponding plots for the reference configurations are provided for comparison.

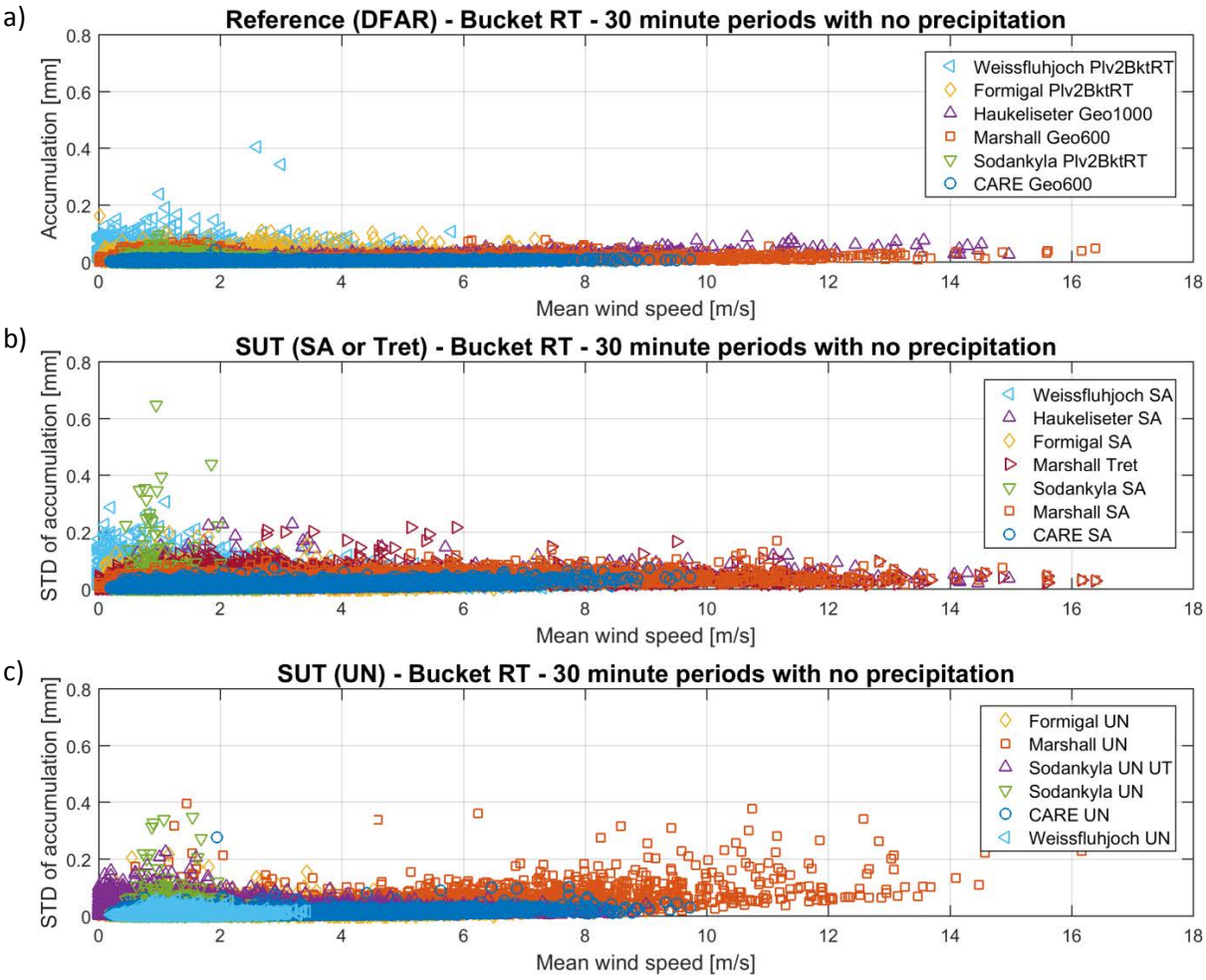


Figure 9: Variability of Bucket RT output (standard deviation of accumulation over each 30 minute assessment interval) as a function of mean wind speed in non-precipitating conditions for (a) the R2 reference configurations (DFAR), (b) Pluvio^2 test gauges in SA- and Tret-shielded configurations, and (c) Pluvio^2 test gauges in unshielded configurations.

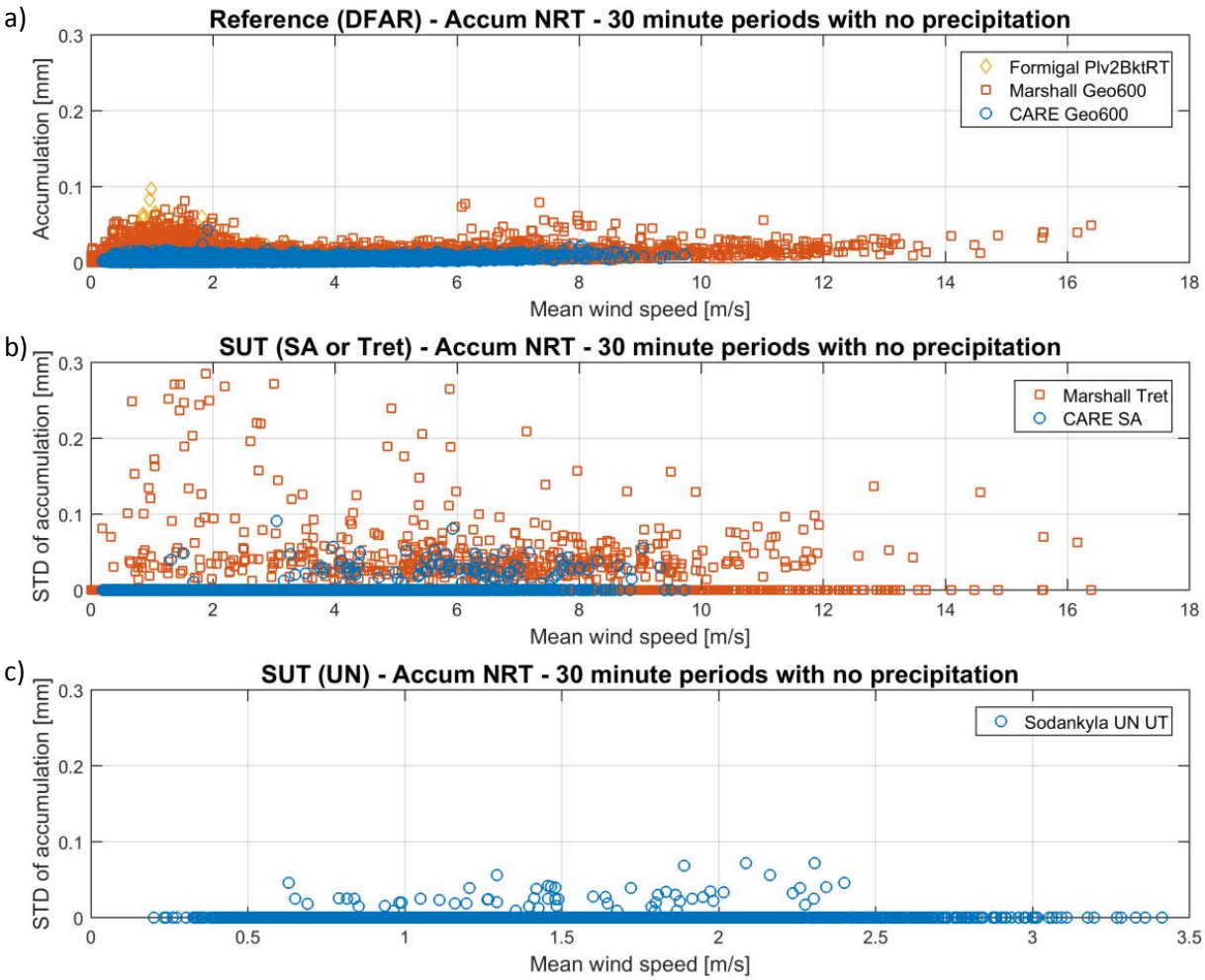


Figure 10: Variability of Accumulated NRT output (standard deviation of accumulation over each 30 minute assessment interval) as a function of mean wind speed in non-precipitating conditions for (a) the R2 reference configurations (DFAR), (b) Pluvio^2 test gauges in SA- and Tret-shielded configurations, and (c) Pluvio^2 test gauges in unshielded configurations. Note that the mean wind speed axis in (c) is truncated relative to those in (a) and (b) for clarity.

6.2. Ability to detect and report precipitation

6.2.1. Skill score assessment

The overall ability of the SUT to detect and report the occurrence of precipitation relative to the site field reference R2 over 30 minute assessment intervals is expressed using selected skill scores (Section 4.1.2) and presented in Figures 11 and 12 for Bucket RT and Accumulated NRT outputs, respectively. The contingency results (Section 4.1.1) corresponding to these scores are presented in Table 5.

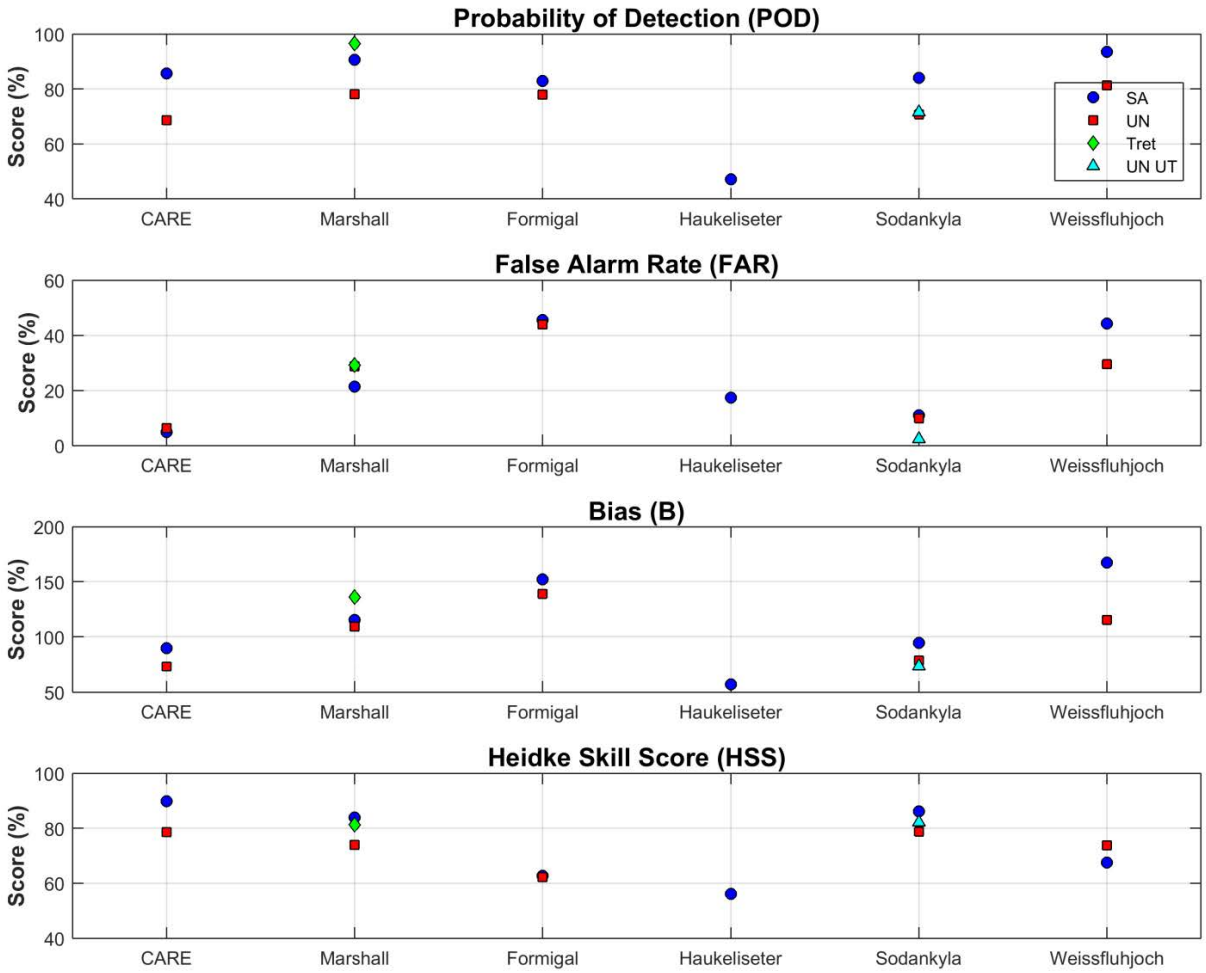


Figure 11: Skill scores for Bucket RT output from Pluvio^2 gauges over the duration of formal tests.

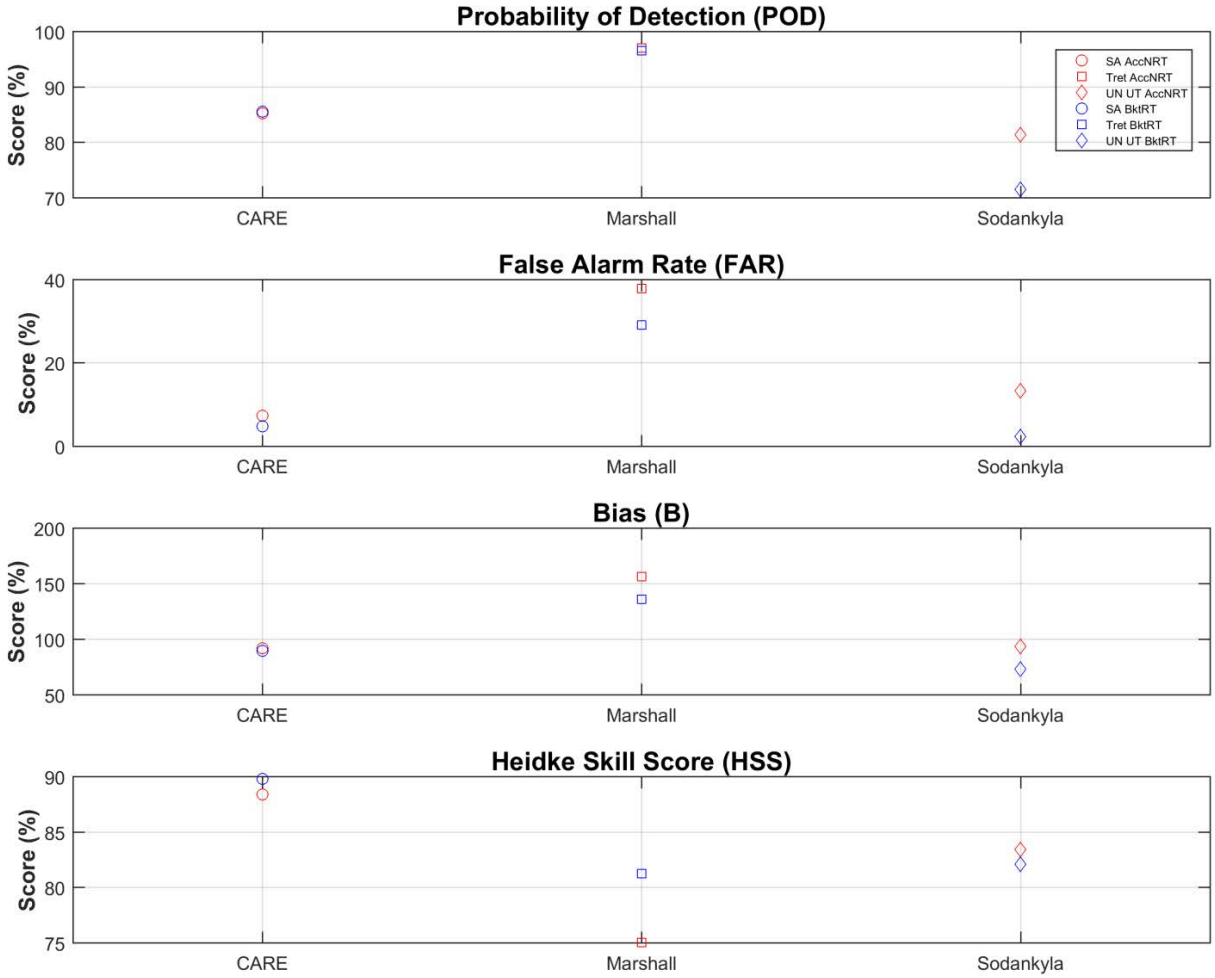


Figure 12: Skill scores for Accumulated RT (red markers) and Bucket RT (blue markers) outputs from Pluvio^2 gauges under test that were provided by the gauge manufacturer or site host, over the duration of formal tests.

Table 5: Contingency table illustrating detection of precipitation by Pluvio^2 gauges under test relative to the corresponding site reference configurations, expressed as the number of events over the duration of formal tests. Results for Bucket RT and Accumulated NRT data are indicated by black and red text, respectively.

			Number of Events			
Site	Configuration	Output data	YY (hits)	YN (misses)	NY (false alarms)	NN (correct negatives)
CARE	SA	Bucket RT	435	73	22	14428
	SA	Accum NRT	440	76	35	14540
	UN	Bucket RT	353	161	24	14566
Formigal	SA	Bucket RT	320	66	267	4934
	UN	Bucket RT	301	85	235	4960
Haukeliseter	SA	Bucket RT	708	794	149	10674
Marshall	SA	Bucket RT	330	34	90	19183
	UN	Bucket RT	295	83	119	17362
	Tret	Bucket RT	365	13	150	17368
	Tret	Accum NRT	366	11	224	17310
Sodankylä	SA	Bucket RT	529	100	65	19453
	UN	Bucket RT	446	184	45	19426
	UN UT	Bucket RT	451	179	11	19510
	UN UT	Accum NRT	512	117	79	19434
Weissfluhjoch	SA	Bucket RT	1119	78	886	18210
	UN	Bucket RT	972	224	407	18699

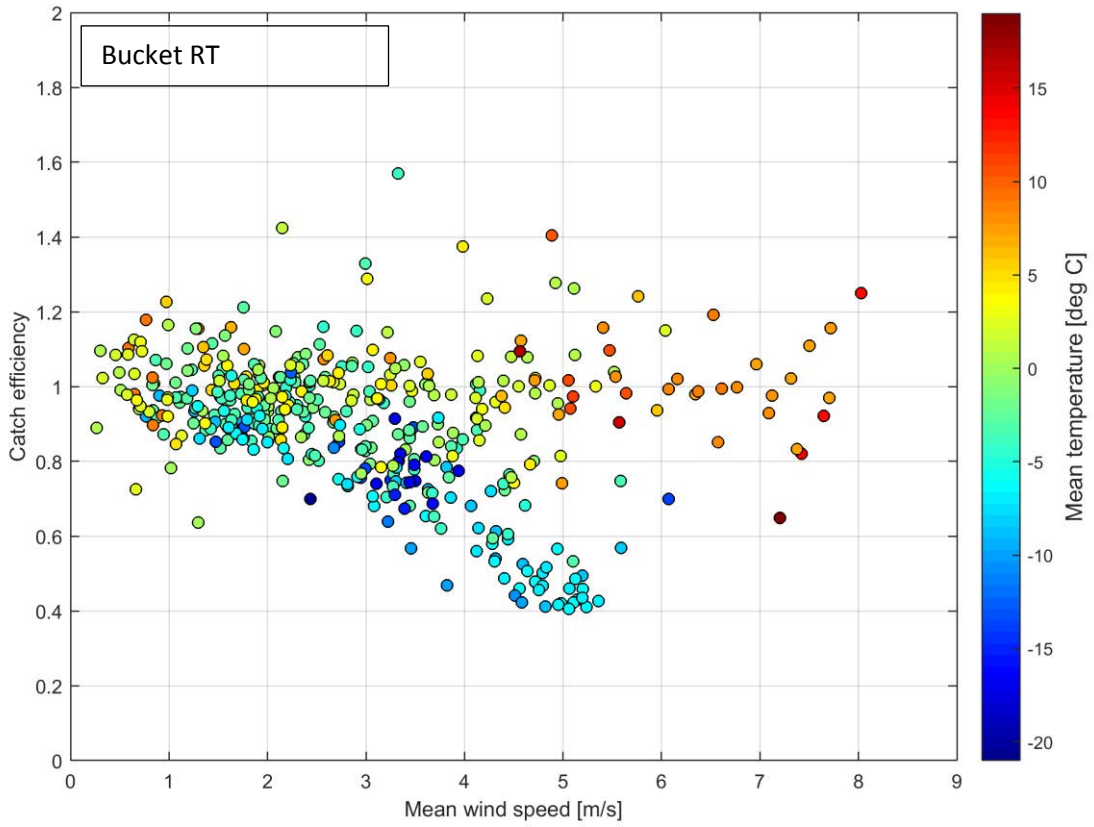
6.3. Ability to report accumulated precipitation

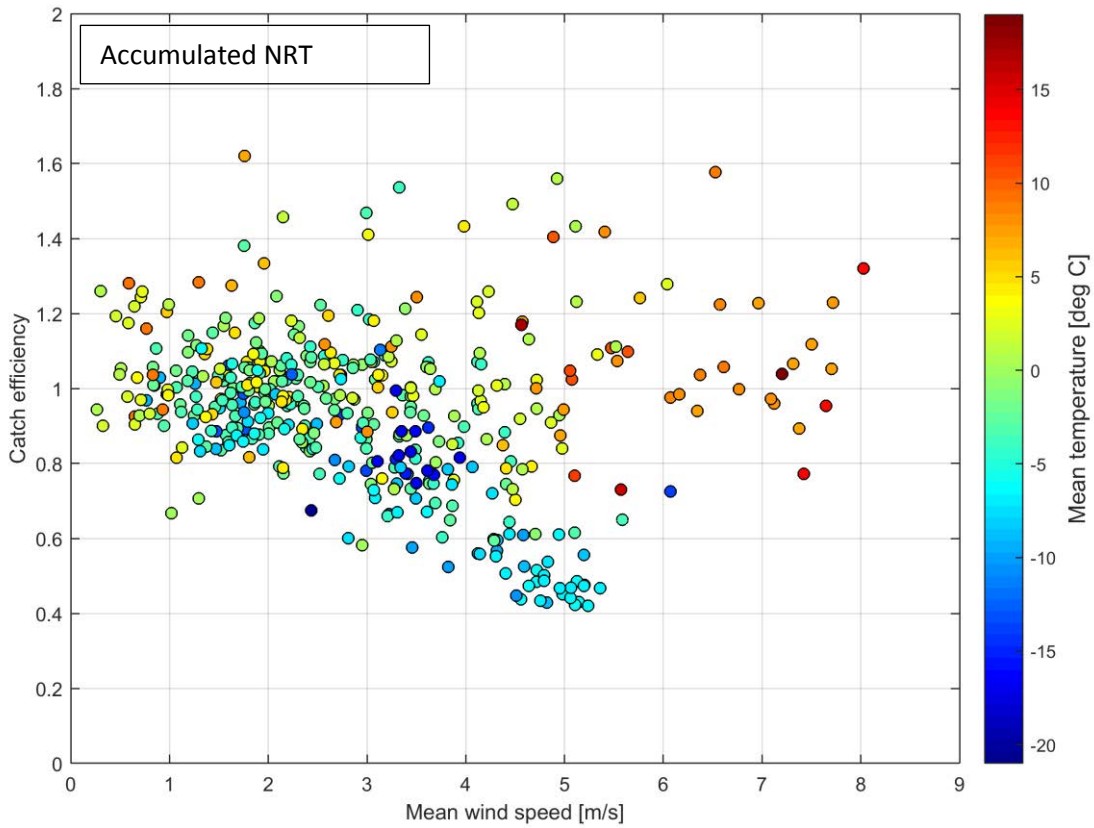
The SUT performance in terms of reporting accumulated precipitation is examined by comparing the amount reported by the sensor under test relative to the respective site reference during 30 minute assessment intervals. This is represented graphically using scatter and box and whisker plots of the catch efficiency as a function of mean wind speed at gauge height, as well as scatter plots of the amounts reported by the SUT versus the corresponding reference amounts (Figures 12 to 24). The SUT performance is also assessed in terms of the root mean square error, RMSE (Figure 25). Results are presented for Bucket RT data for all test gauges, and for both Bucket RT and Accumulated NRT data for gauges under test from Instrument Providers.

Only assessment intervals during which the SUT and reference both reported precipitation (YY cases) are considered in this portion of the assessment. In the catch efficiency-wind speed scatter plots, the mean event temperature is indicated by colour, with the colour scale selected to be consistent across all sites with weighing gauges under test. In the box and whisker plots and accumulation-accumulation scatter plots, the predominant precipitation type is indicated by colour, as determined from the reported temperature (Section 4.1.4).

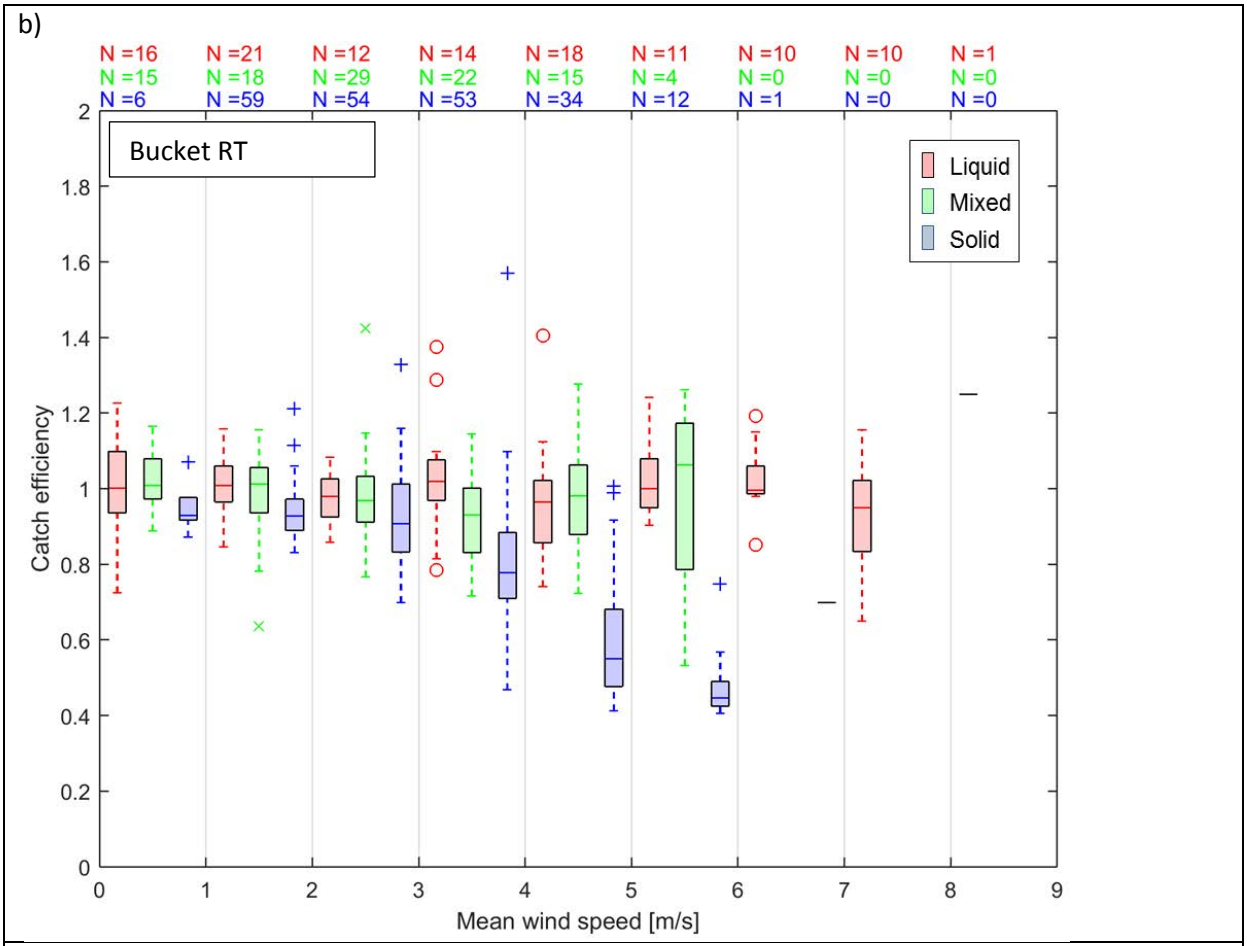
Figure 12: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the SA-shielded OTT Pluvio^2 gauge under test at CARE (Bucket RT and Accumulated NRT outputs).

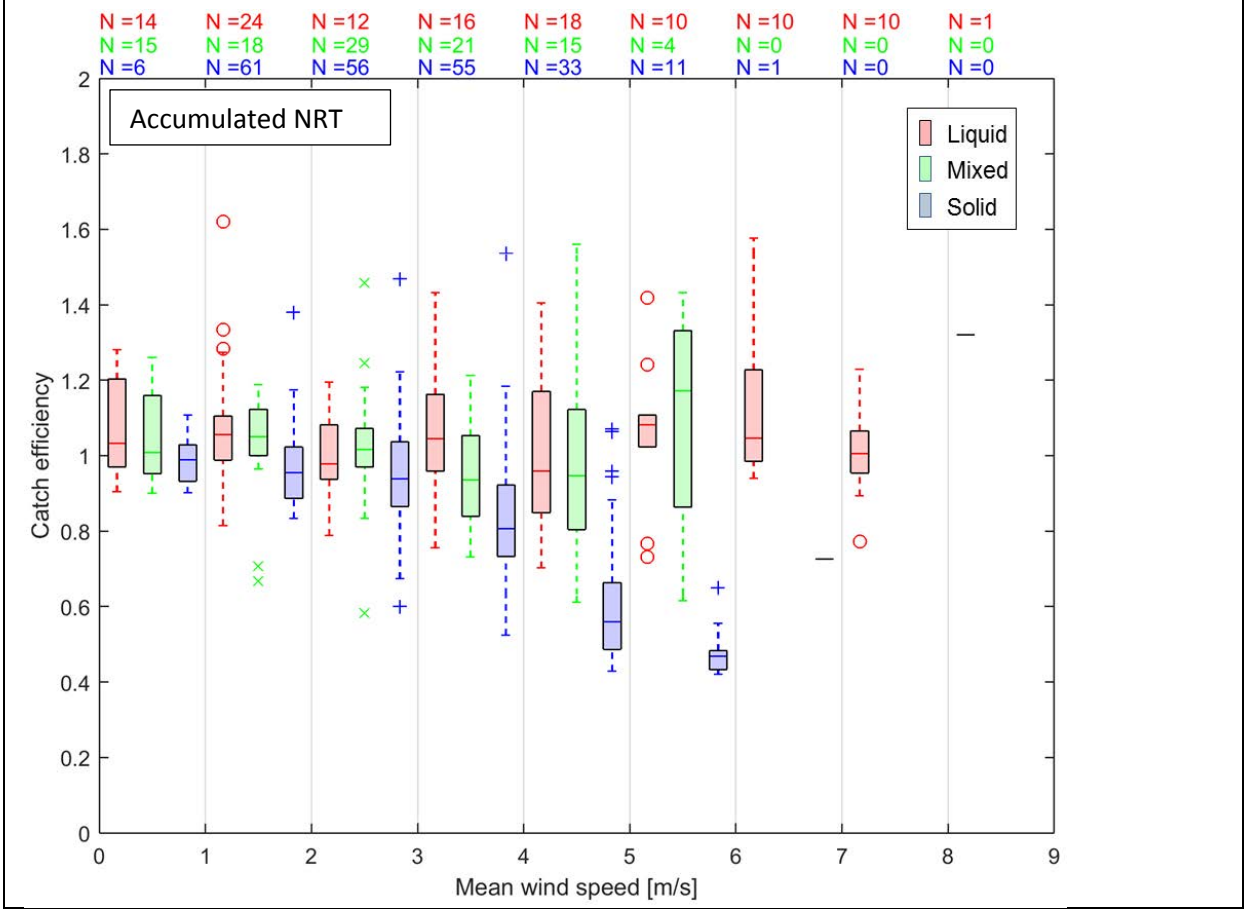
a)



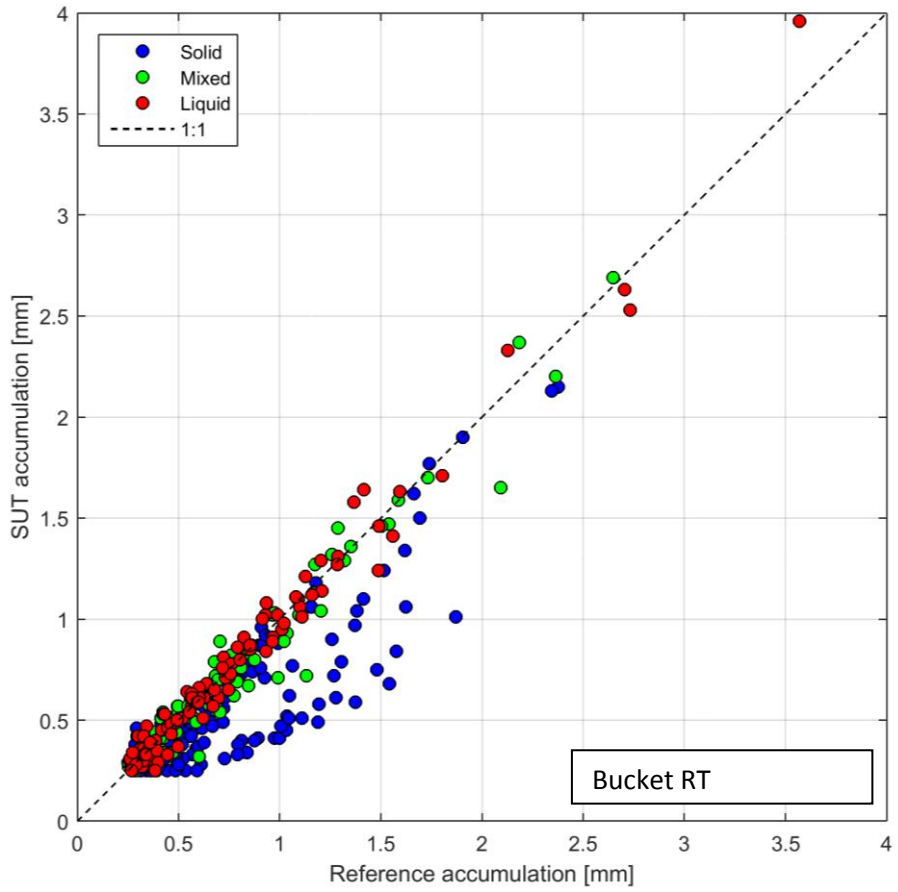


b)


N =16 N =21 N =12 N =14 N =18 N =11 N =10 N =10 N =1
N =15 N =18 N =29 N =22 N =15 N =4 N =0 N =0 N =0
N =6 N =59 N =54 N =53 N =34 N =12 N =1 N =0 N =0
Bucket RT
Liquid
Mixed
Solid
Catch efficiency
Mean wind speed [m/s]


N =14 N =24 N =12 N =16 N =18 N =10 N =10 N =10 N =1
N =15 N =18 N =29 N =21 N =15 N =4 N =0 N =0 N =0
N =6 N =61 N =56 N =55 N =33 N =11 N =1 N =0 N =0
Accumulated NRT
Liquid
Mixed
Solid
Catch efficiency
Mean wind speed [m/s]

c)


Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Bucket RT

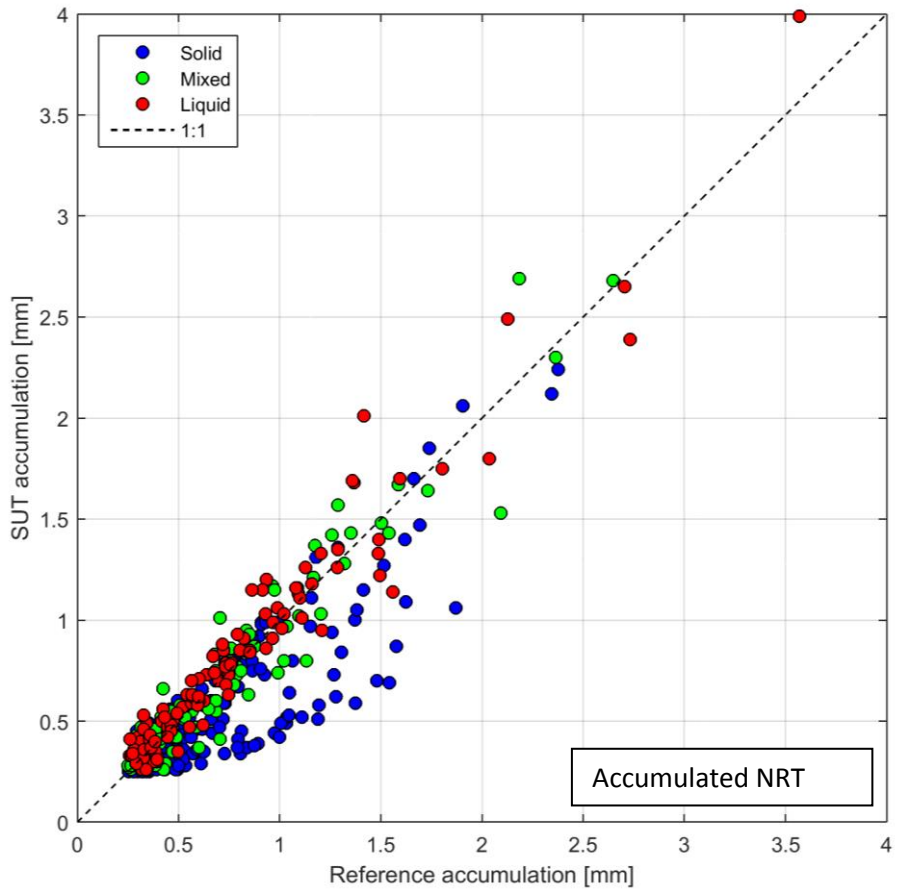
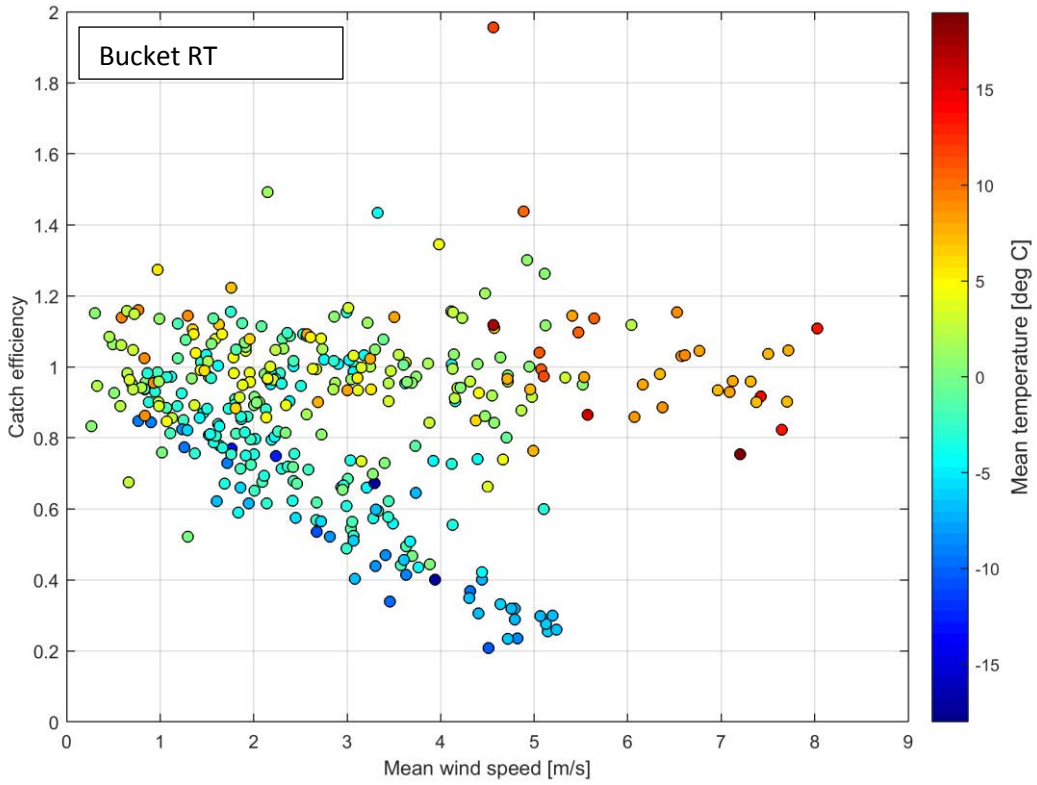
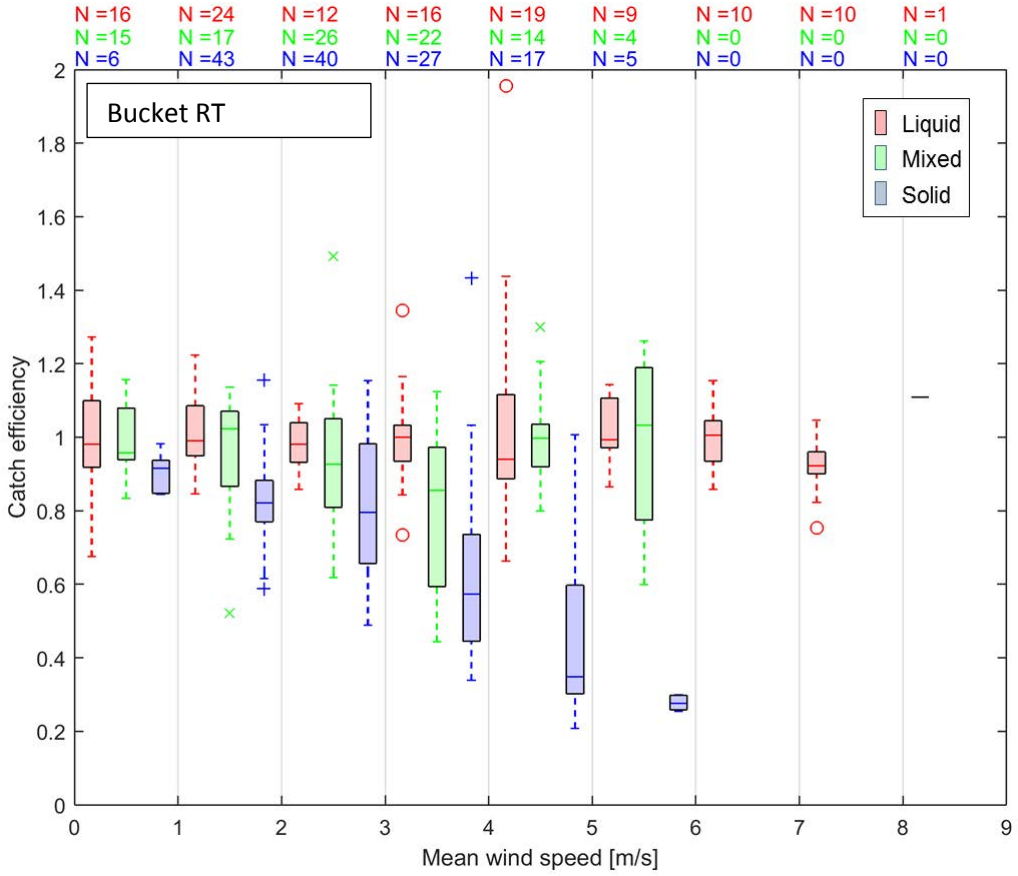

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Accumulated NRT

Figure 13: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded OTT Pluvio^2 gauge under test at CARE (Bucket RT output).

a)



b)



c)

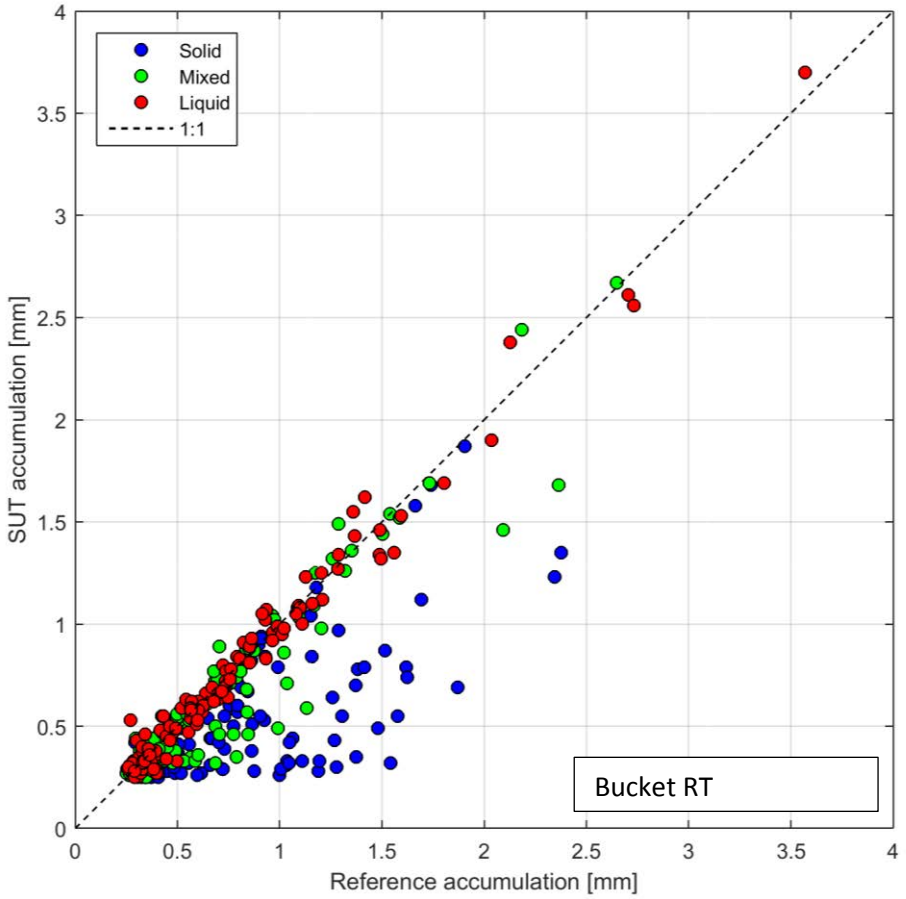
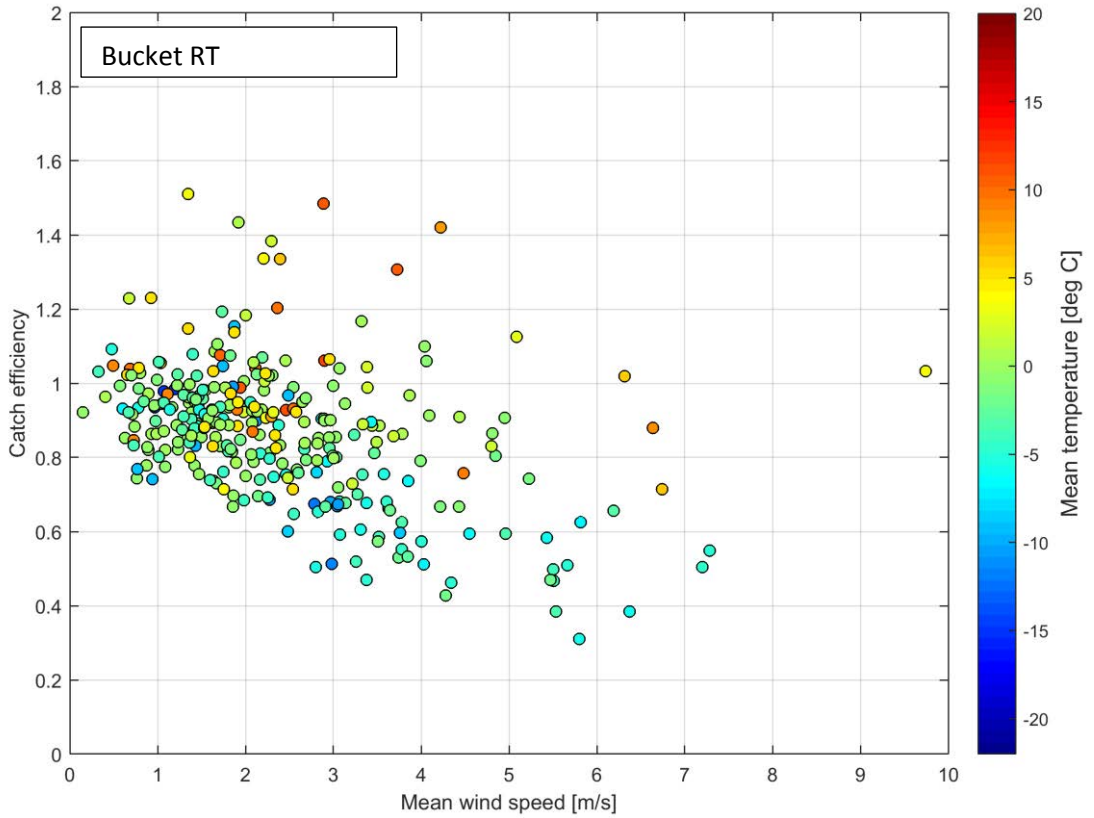
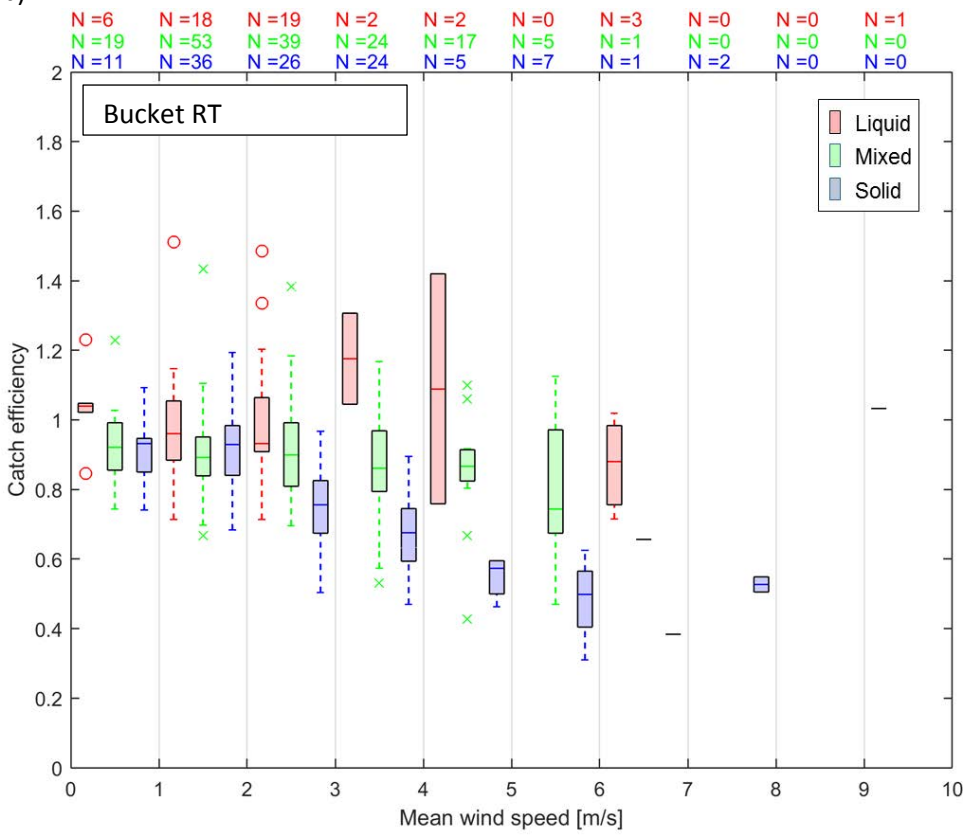

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Bucket RT
0
0.5
1
1.5
2
2.5
3
3.5
4

Figure 14: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the single-Alter shielded OTT Pluvio^2 gauge under test at Marshall (Bucket RT output).

a)



b)



c)

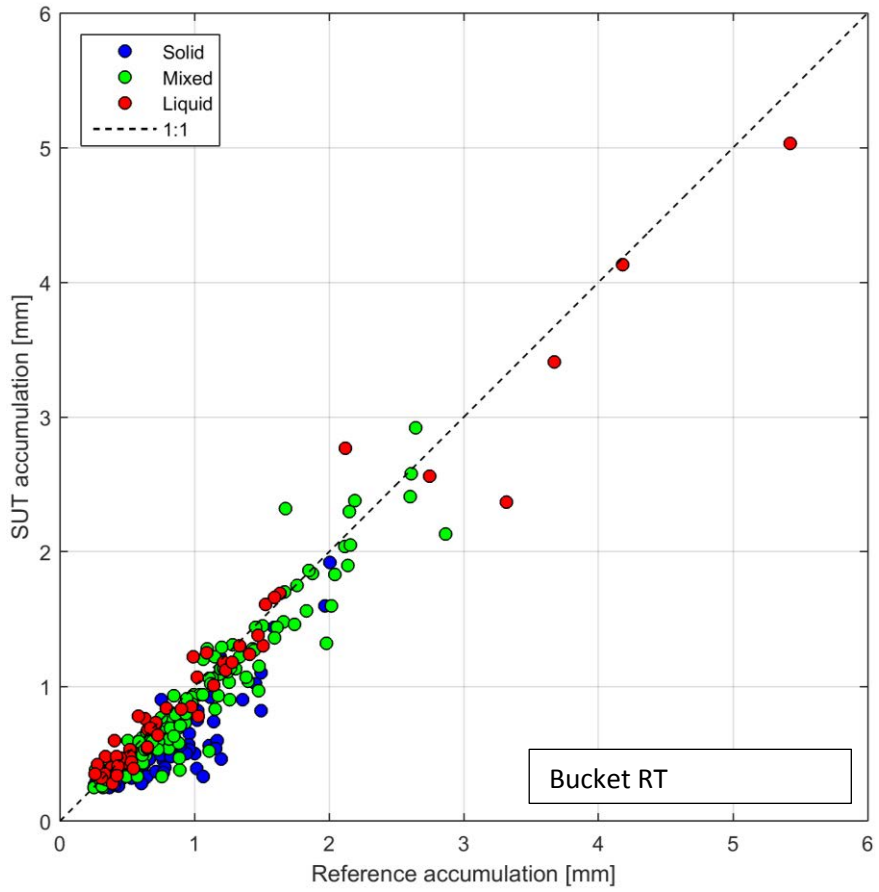
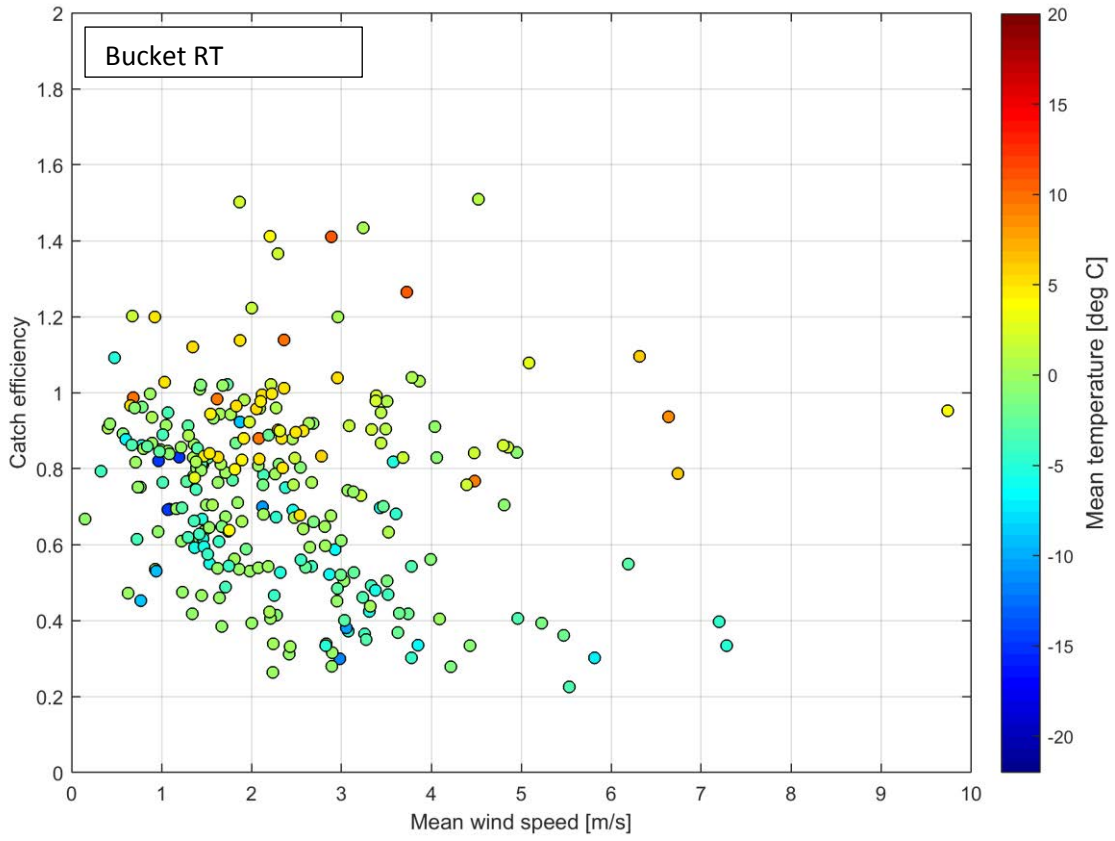
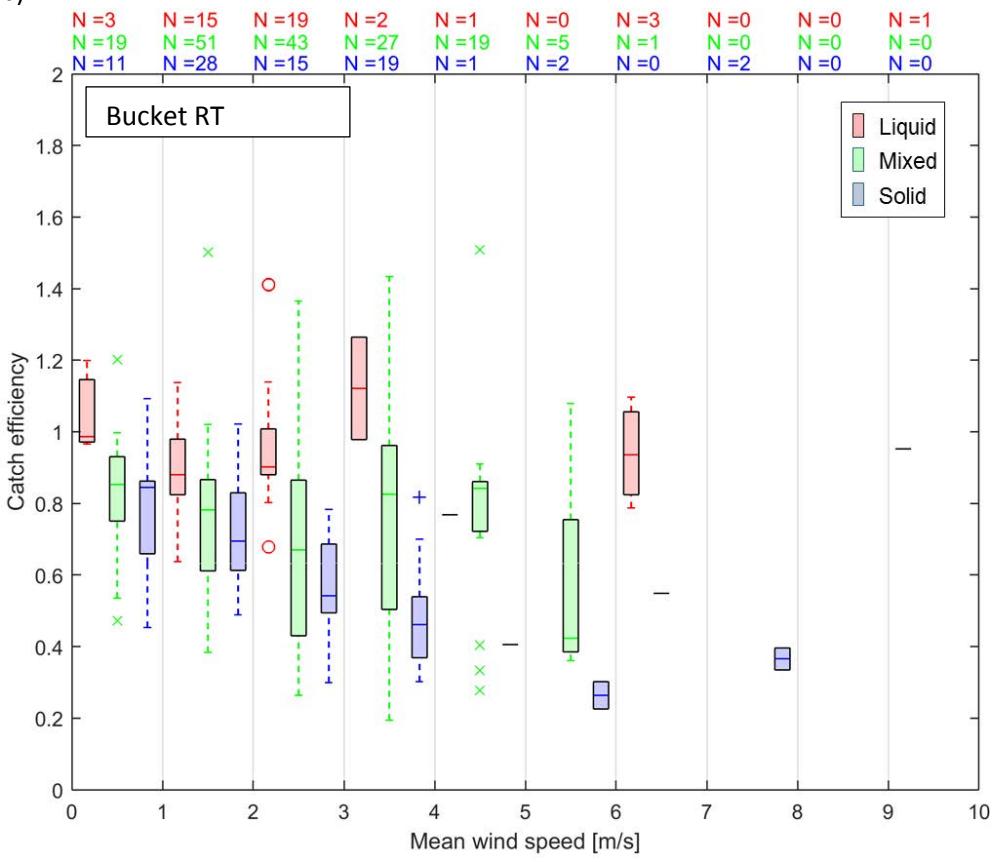

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Bucket RT
0
1
2
3
4
5
6

Figure 15: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded OTT Pluvio^2 gauge under test at Marshall (Bucket RT output).

a)



b)



c)

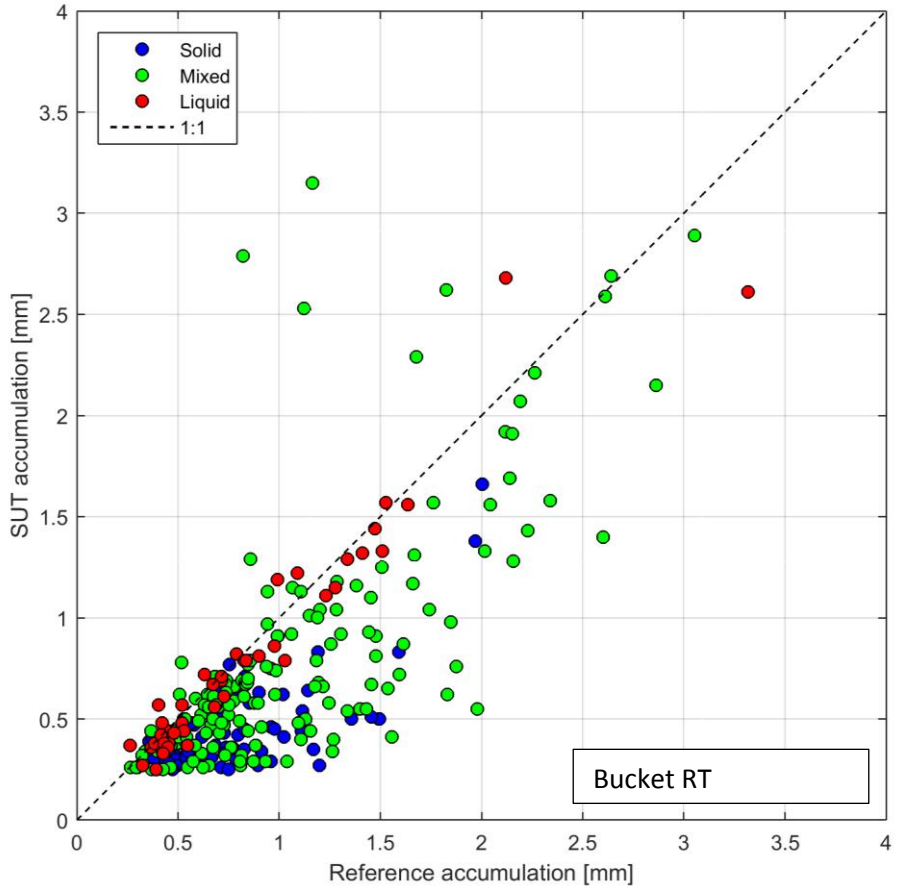
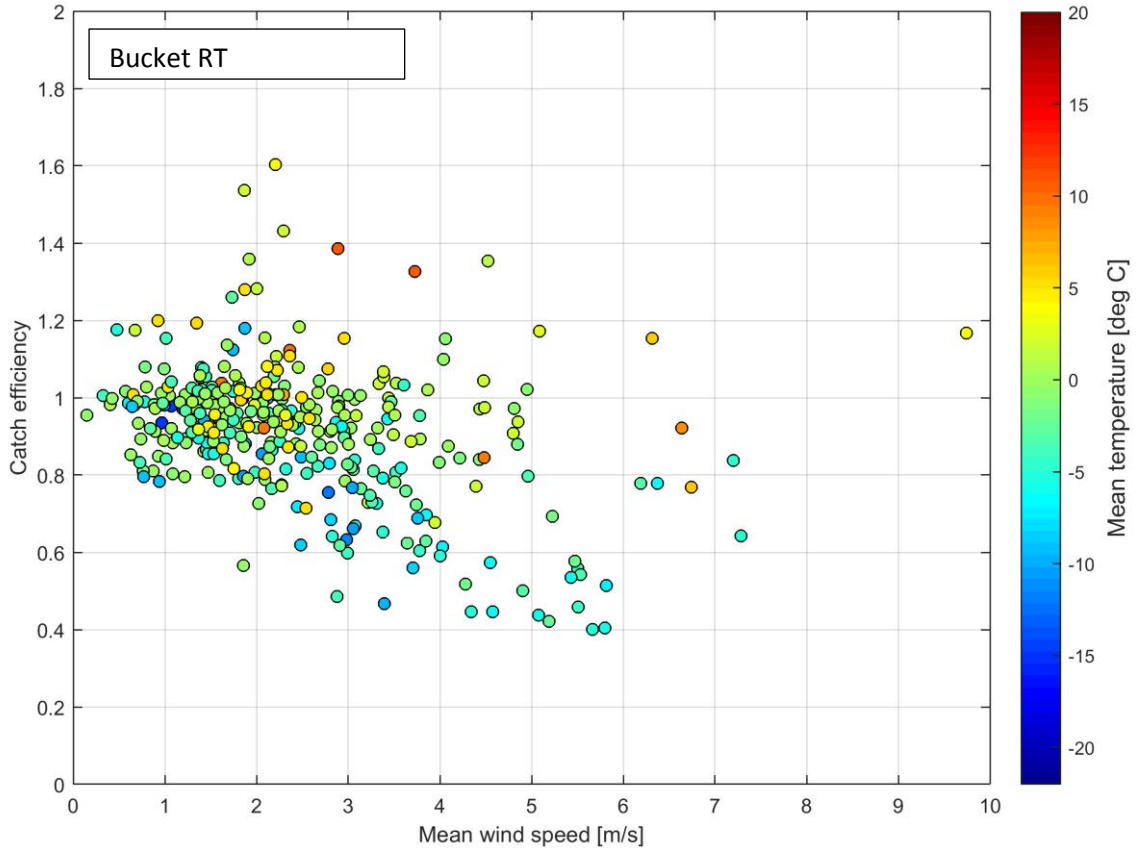
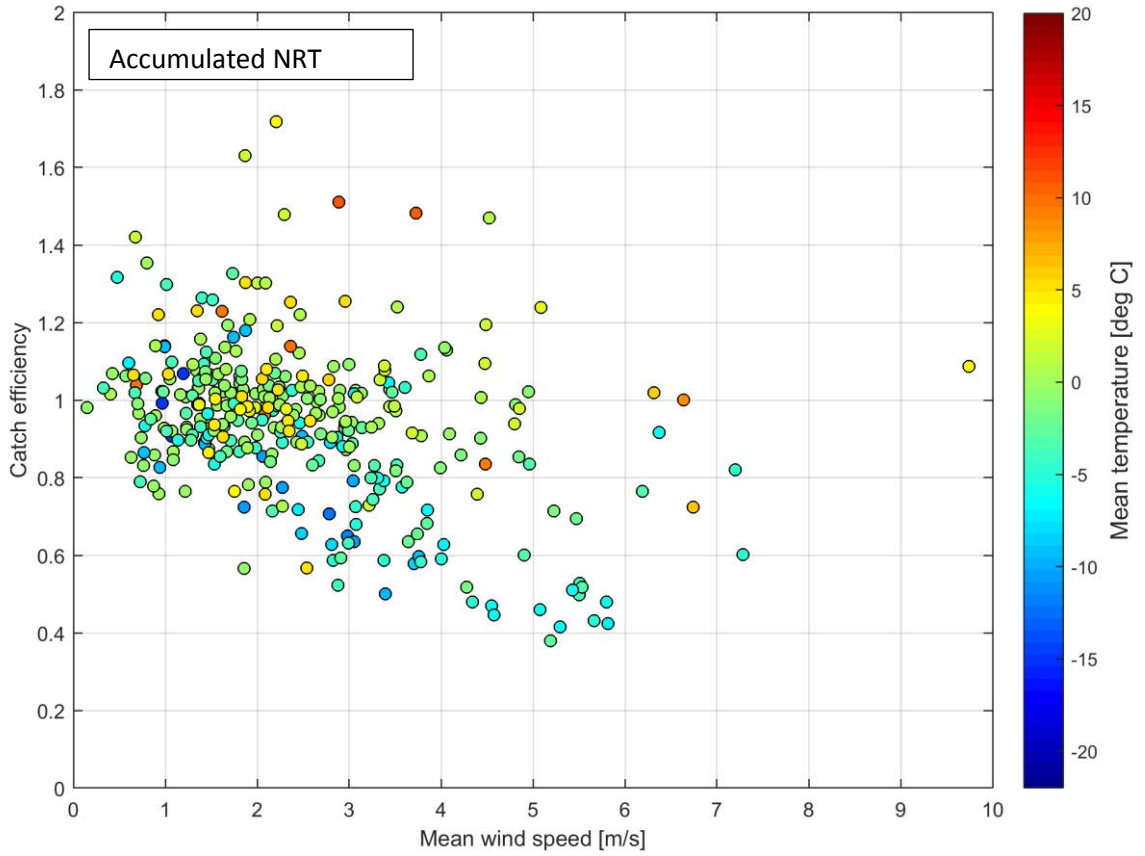

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Bucket RT
0
0.5
1
1.5
2
2.5
3
3.5
4

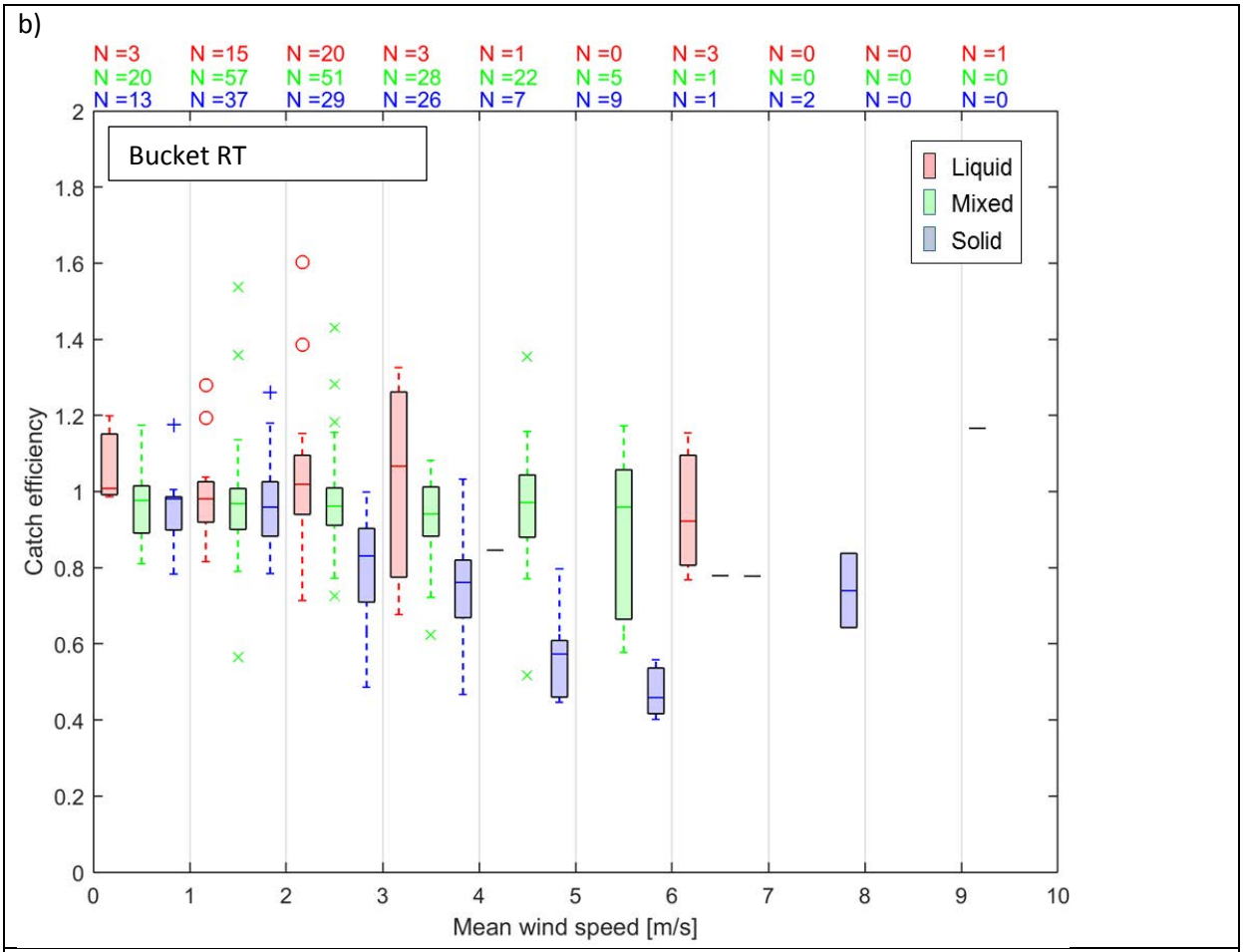
Figure 16: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the Tretyakov-shielded OTT Pluvio^2 gauge under test at Marshall provided by the gauge manufacturer (Bucket RT and Accumulated NRT outputs).

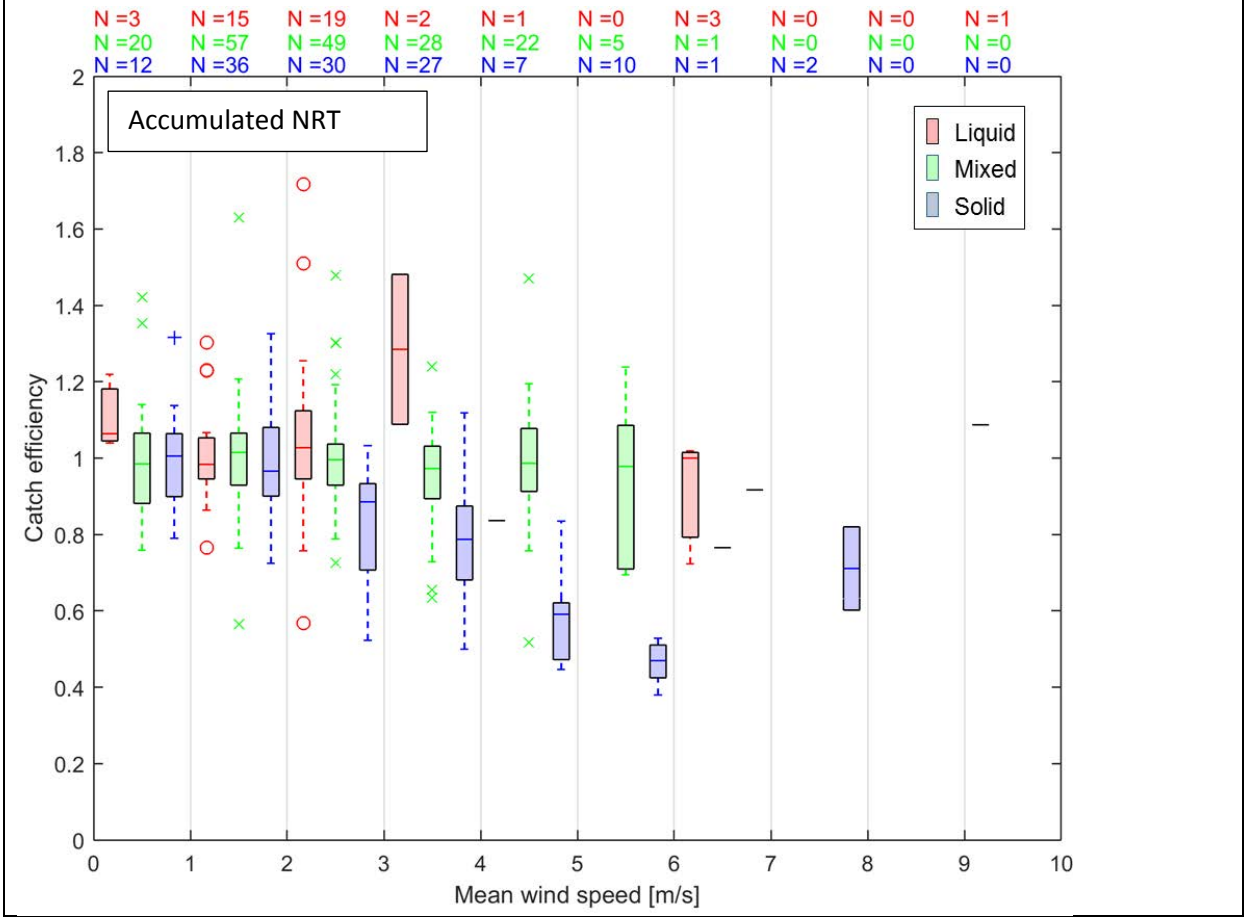
a)



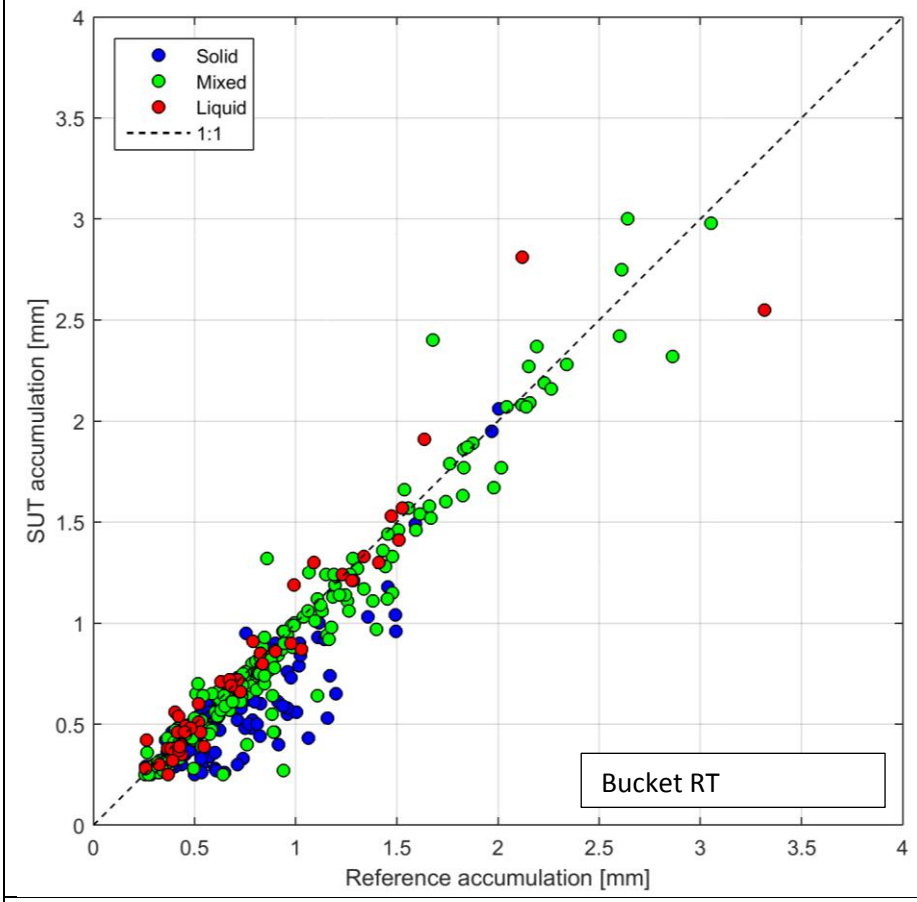


b)


N =3 N =15 N =20 N =3 N =1 N =0 N =3 N =0 N =0 N =1
N =20 N =57 N =51 N =28 N =22 N =5 N =1 N =0 N =0 N =0
N =13 N =37 N =29 N =26 N =7 N =9 N =1 N =2 N =0 N =0
Bucket RT
Liquid
Mixed
Solid
Catch efficiency
Mean wind speed [m/s]


N =3 N =15 N =19 N =2 N =1 N =0 N =3 N =0 N =0 N =1
N =20 N =57 N =49 N =28 N =22 N =5 N =1 N =0 N =0 N =0
N =12 N =36 N =30 N =27 N =7 N =10 N =1 N =2 N =0 N =0
Accumulated NRT
Liquid
Mixed
Solid
Catch efficiency
Mean wind speed [m/s]

c)


Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Bucket RT

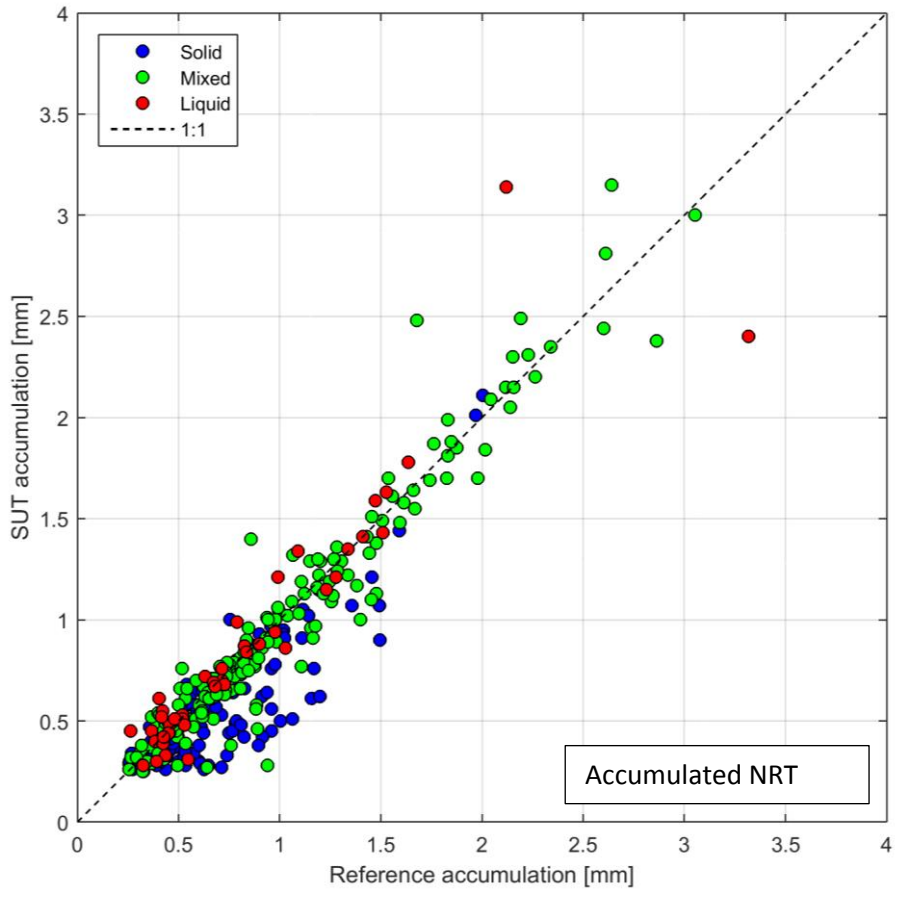
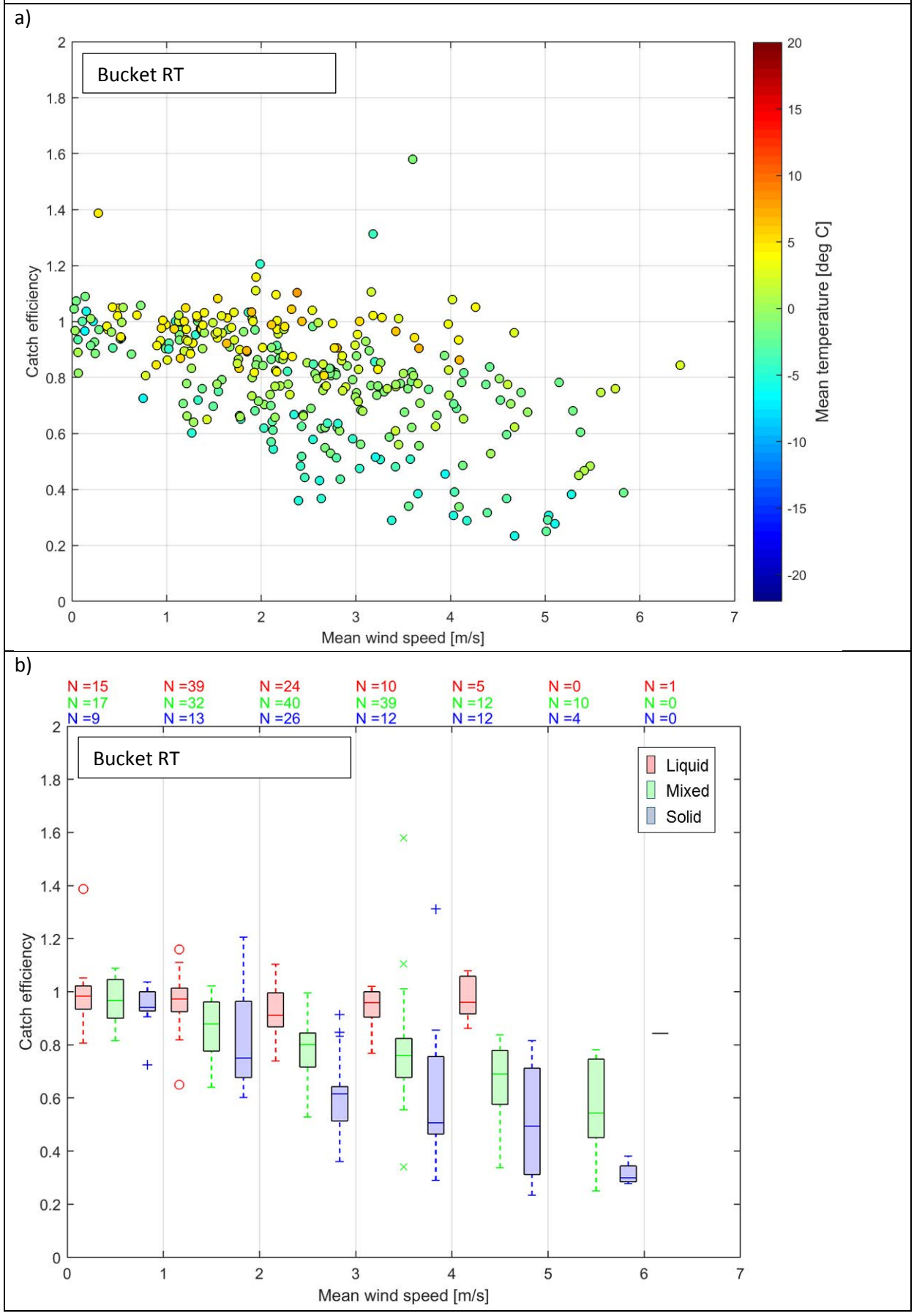

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Accumulated NRT

Figure 17: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the single-Alter shielded OTT Pluvio^2 gauge under test at Formigal (Bucket RT output).



c)

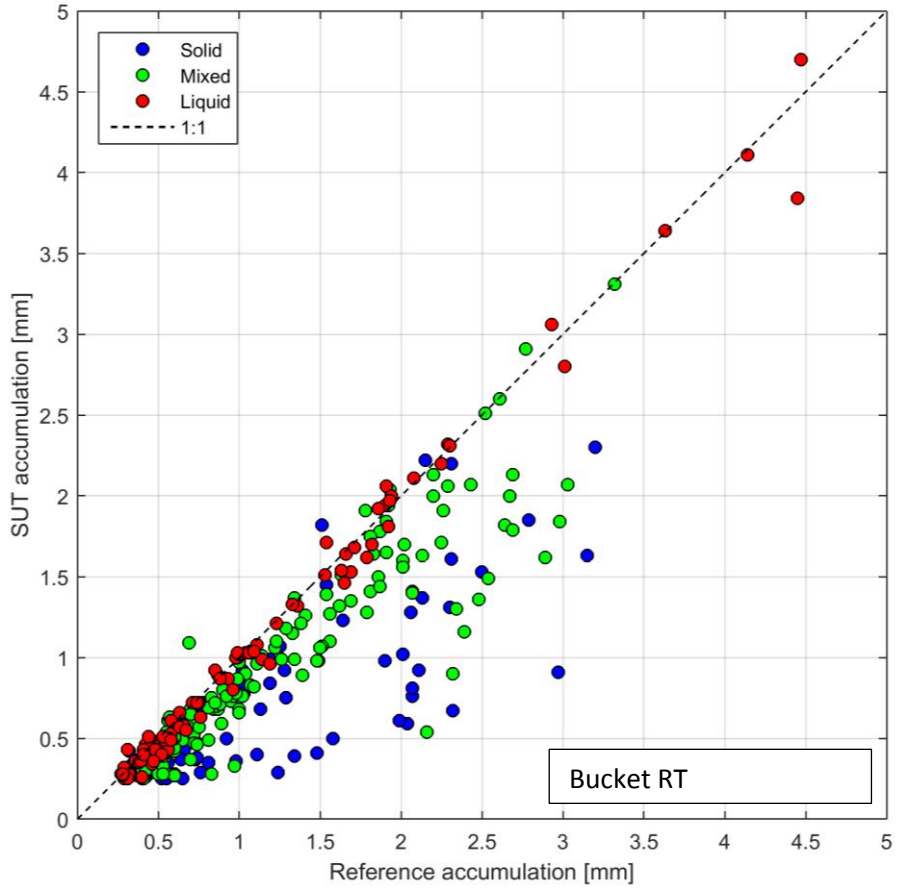
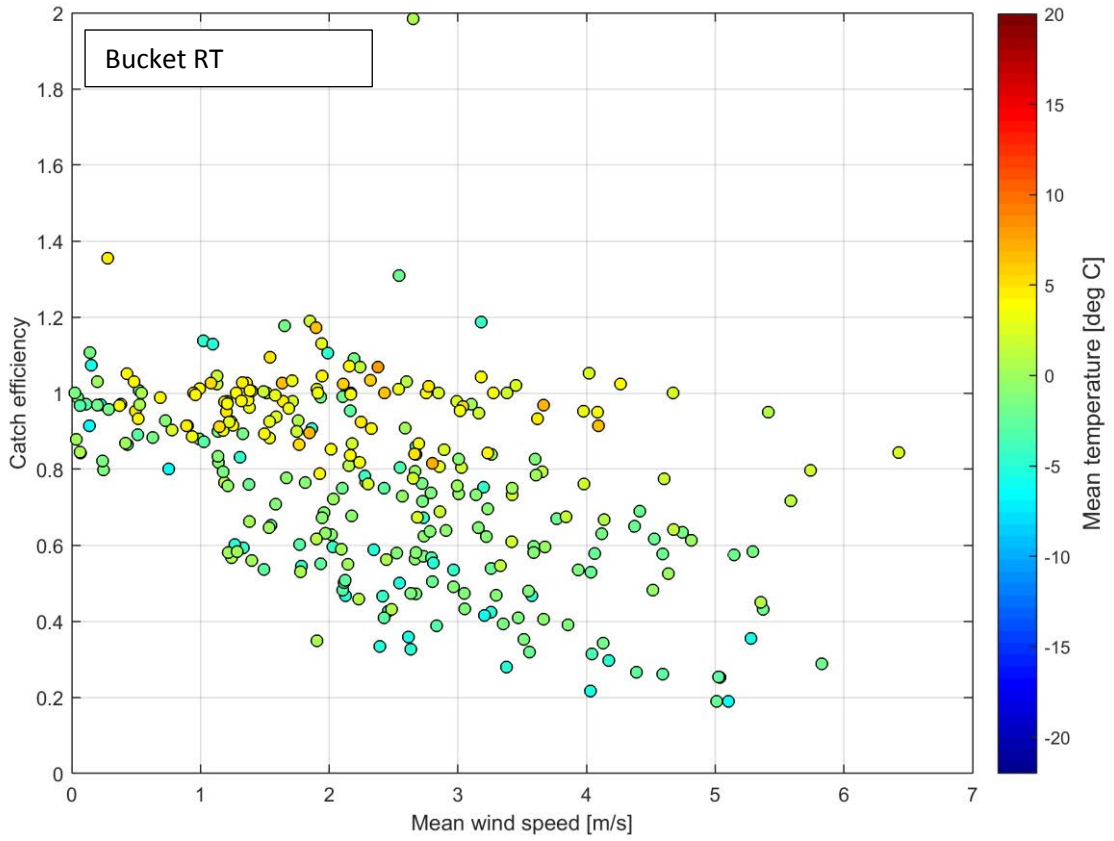
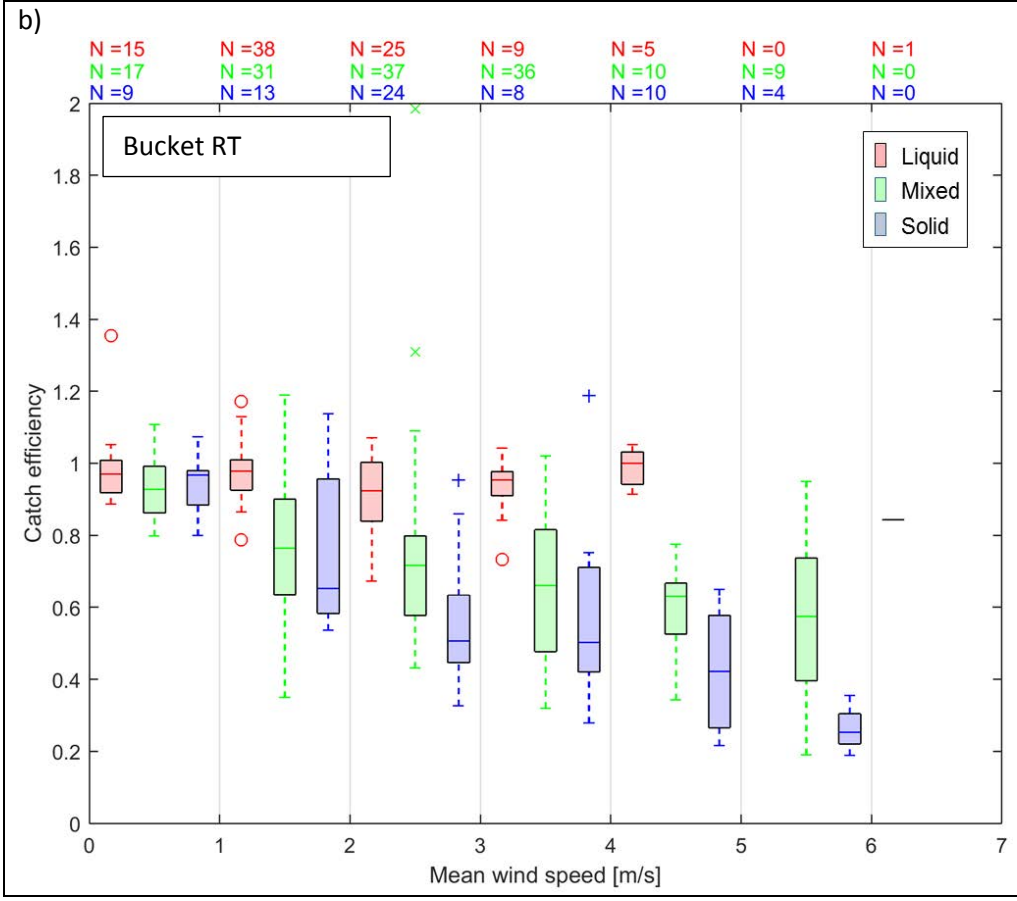

Solid
Mixed
Liquid
1:1
Bucket RT
SUT accumulation [mm]
Reference accumulation [mm]
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5

Figure 18: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded OTT Pluvio^2 gauge under test at Formigal (Bucket RT output).

a)



b)



c)

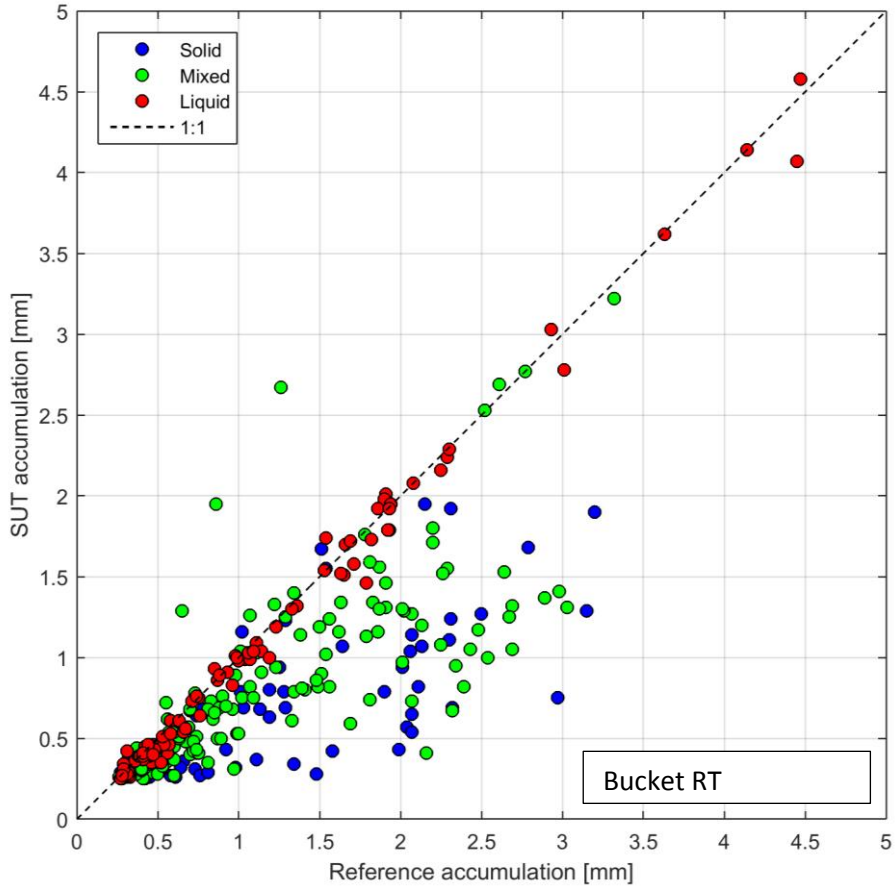
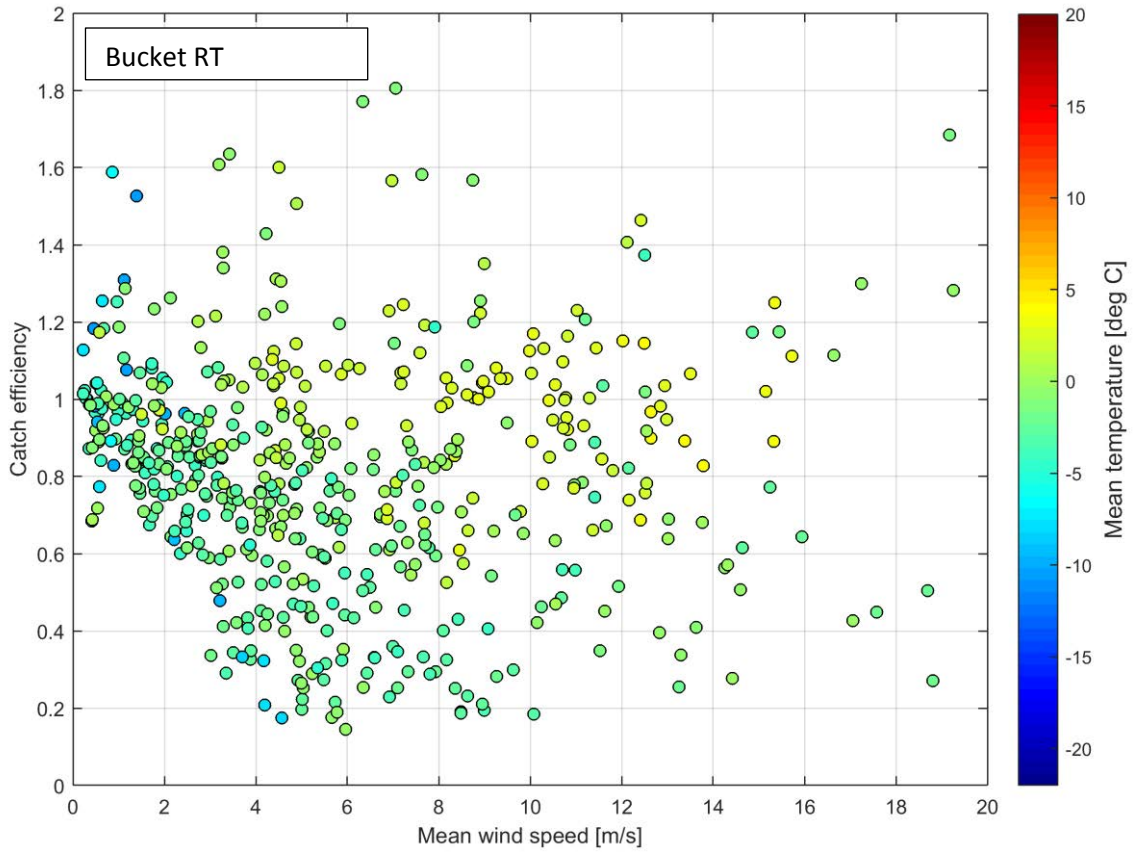
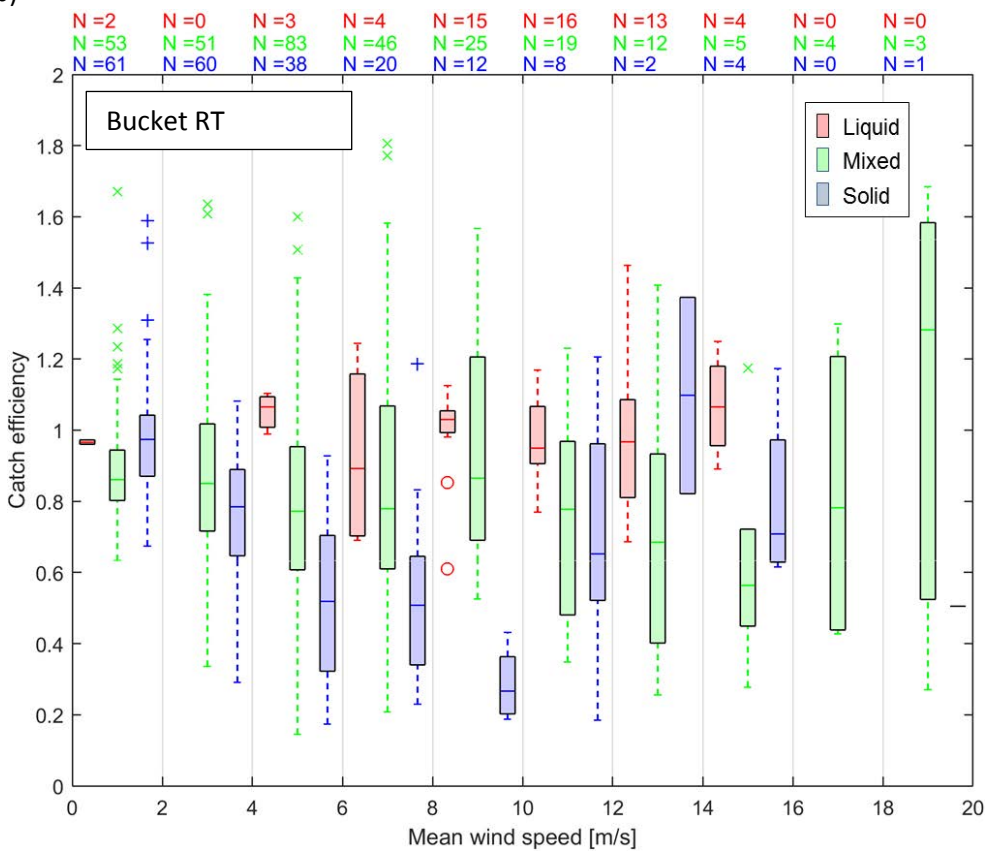

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Bucket RT
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5

Figure 19: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the single-Alter shielded OTT Pluvio^2 gauge under test at Haukeliseter (Bucket RT output).

a)



b)



c)

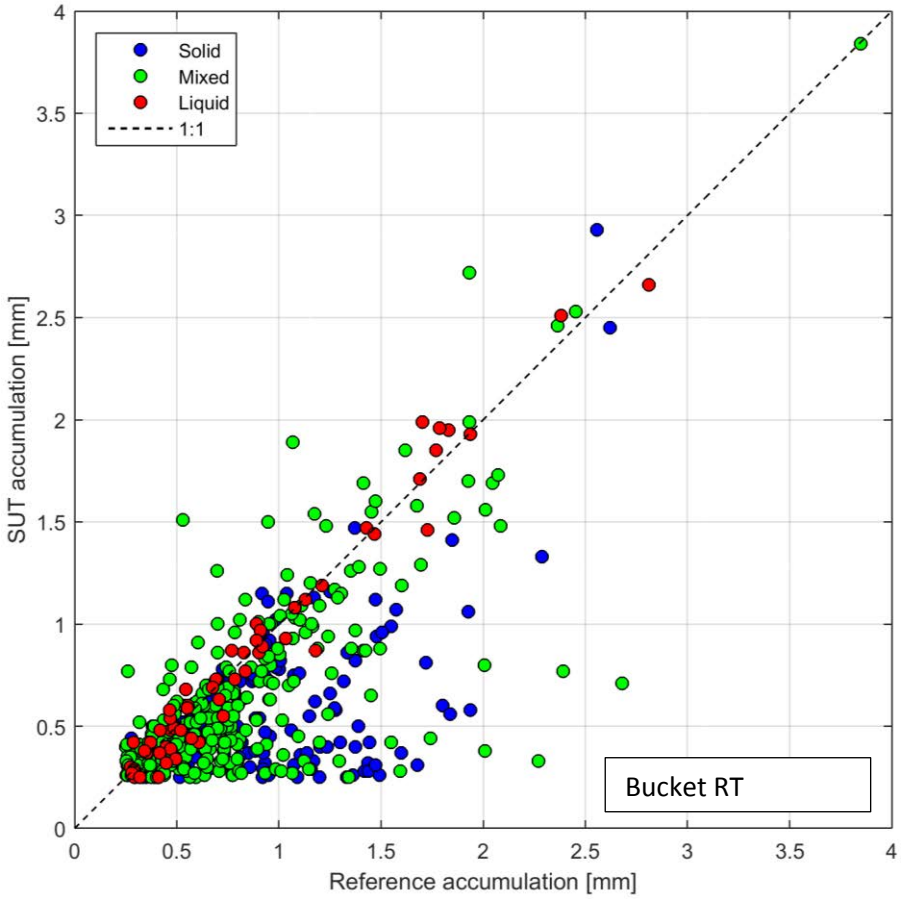
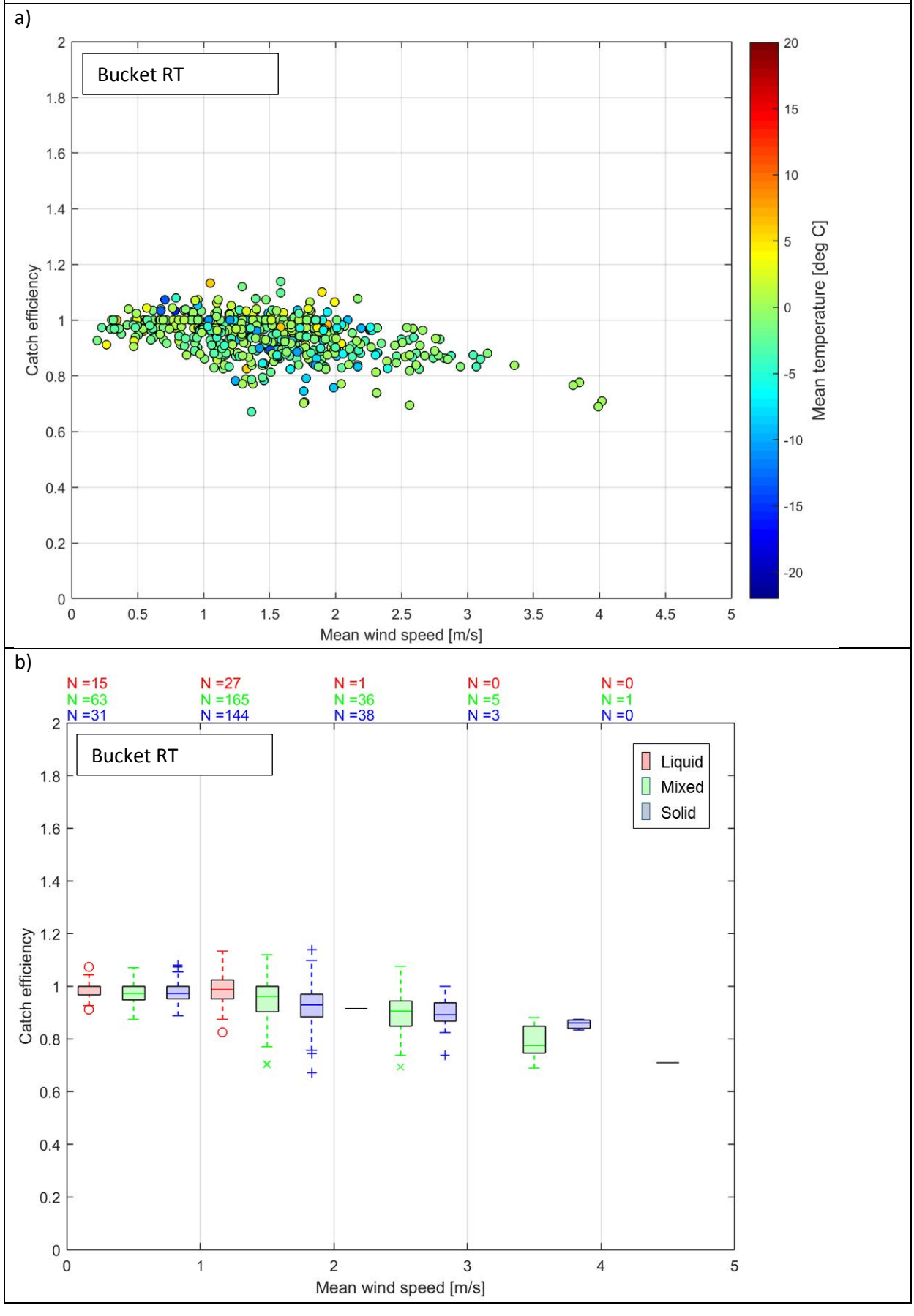

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Bucket RT
0
0.5
1
1.5
2
2.5
3
3.5
4

Figure 20: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the single-Alter shielded OTT Pluvio^2 gauge under test at Sodankylä (Bucket RT output).



c)

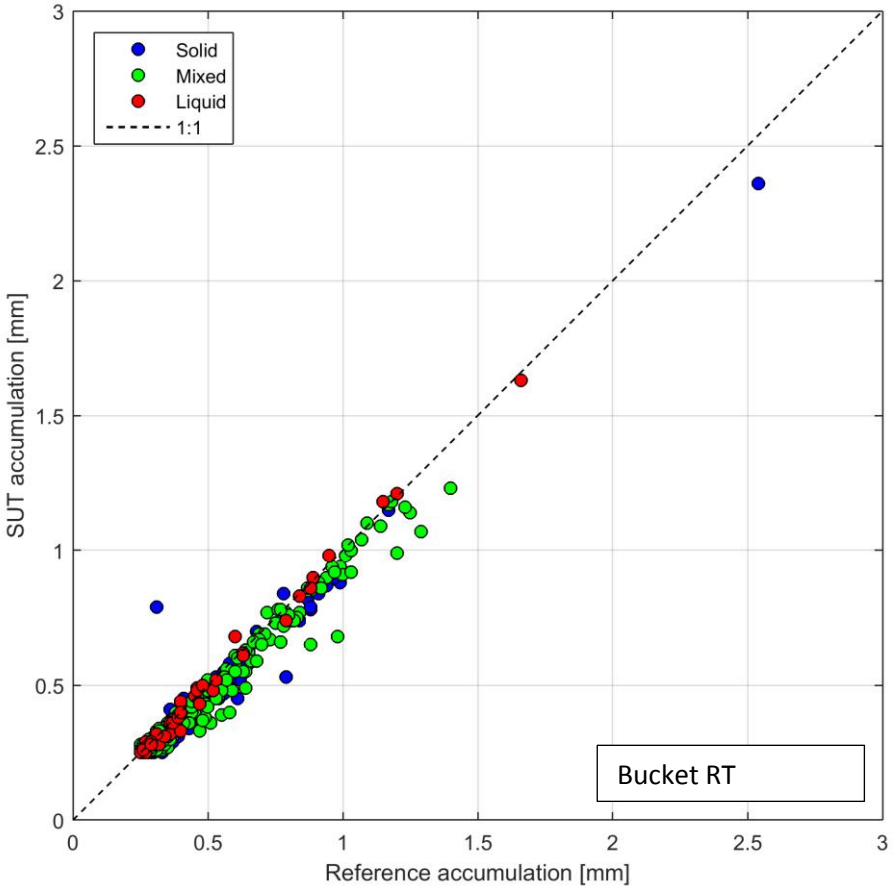
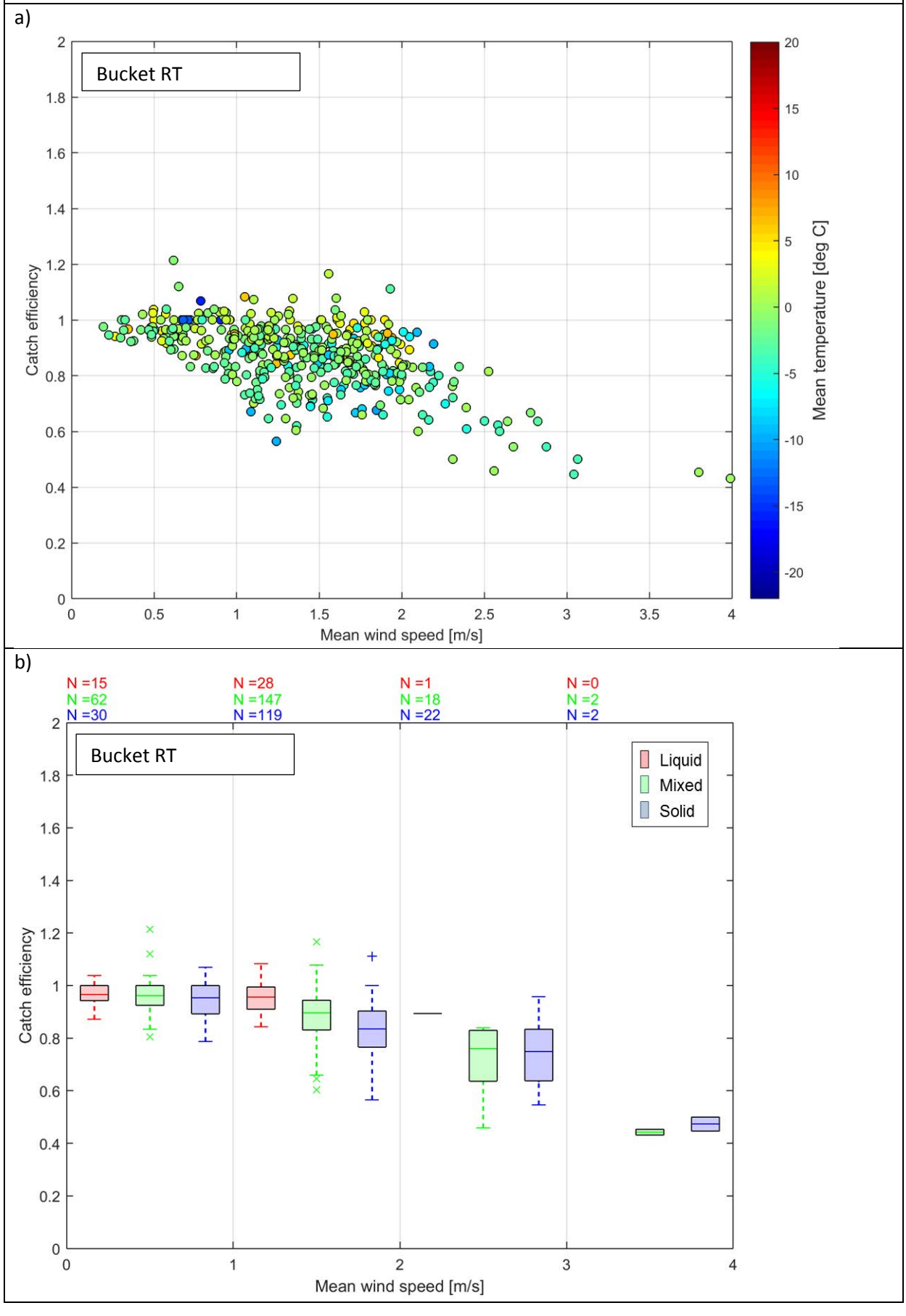

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Bucket RT
0
0.5
1
1.5
2
2.5
3

Figure 21: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded OTT Pluvio^2 gauge under test at Sodankylä (Bucket RT output).



c)

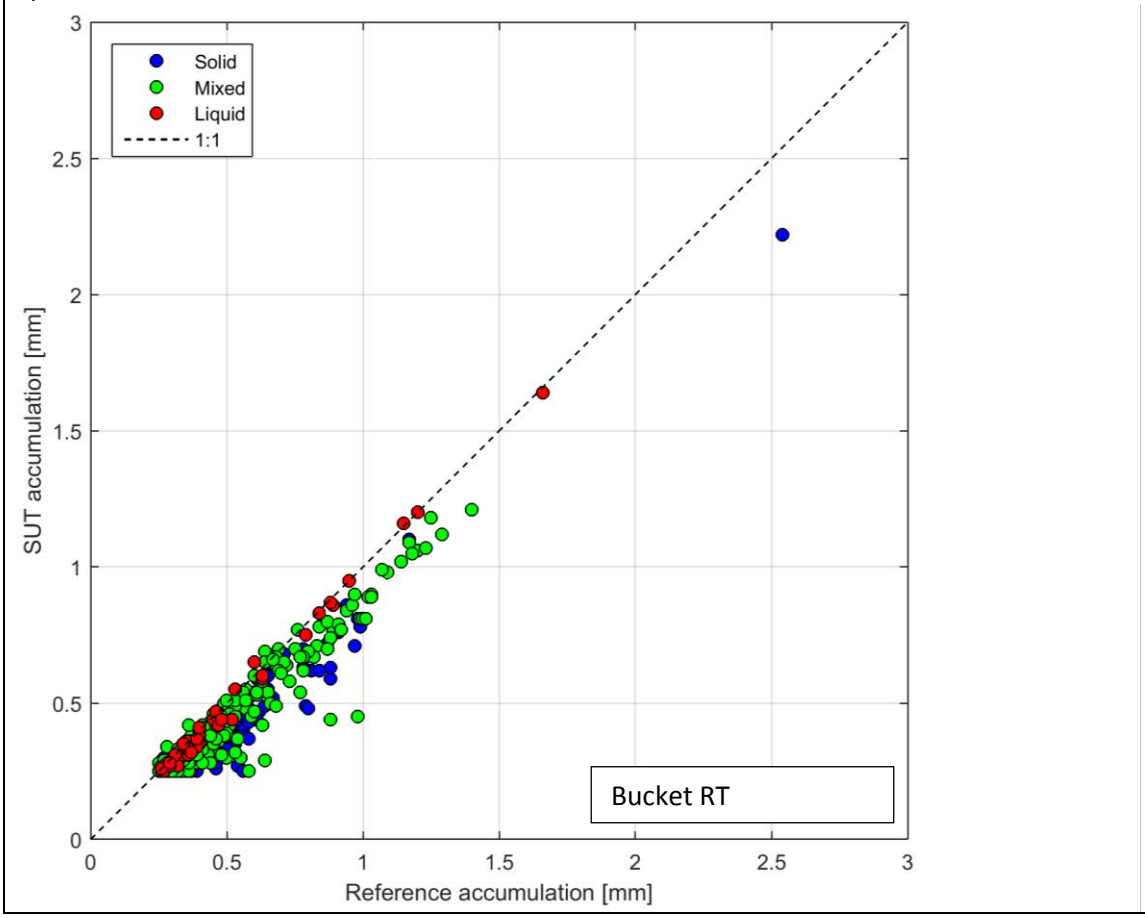
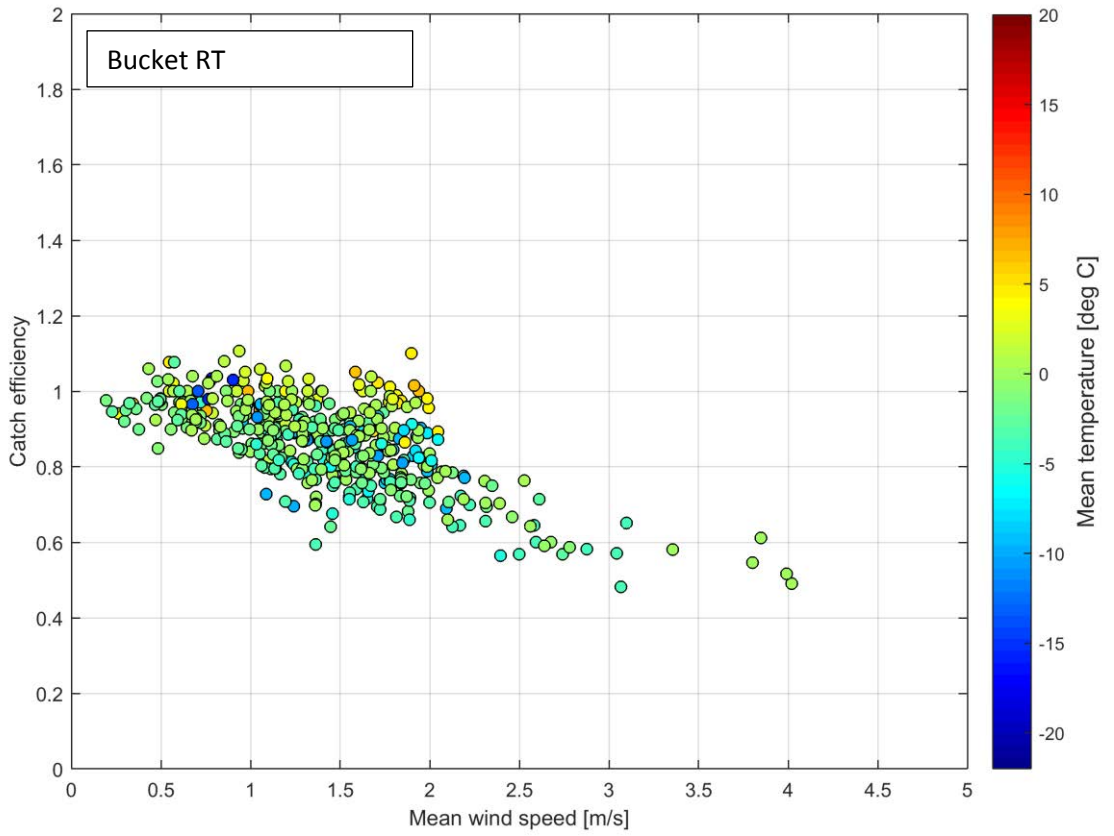
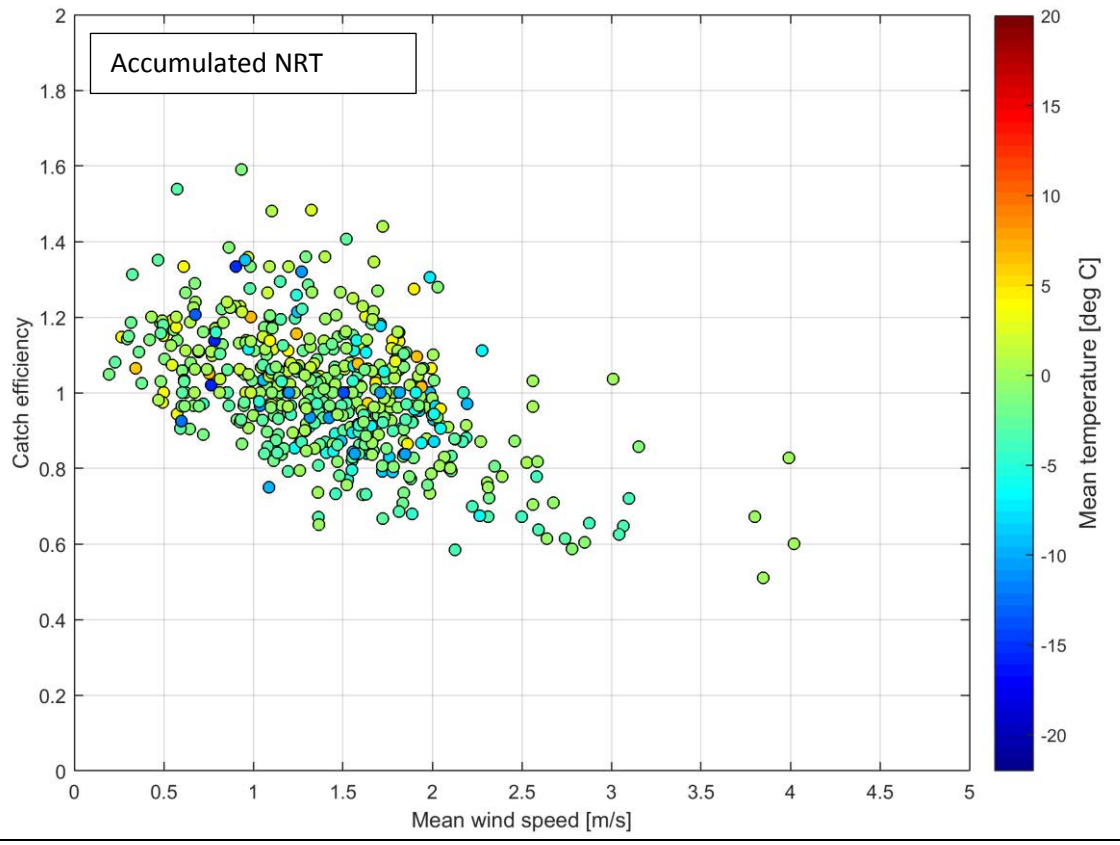

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Bucket RT
0
0.5
1
1.5
2
2.5
3

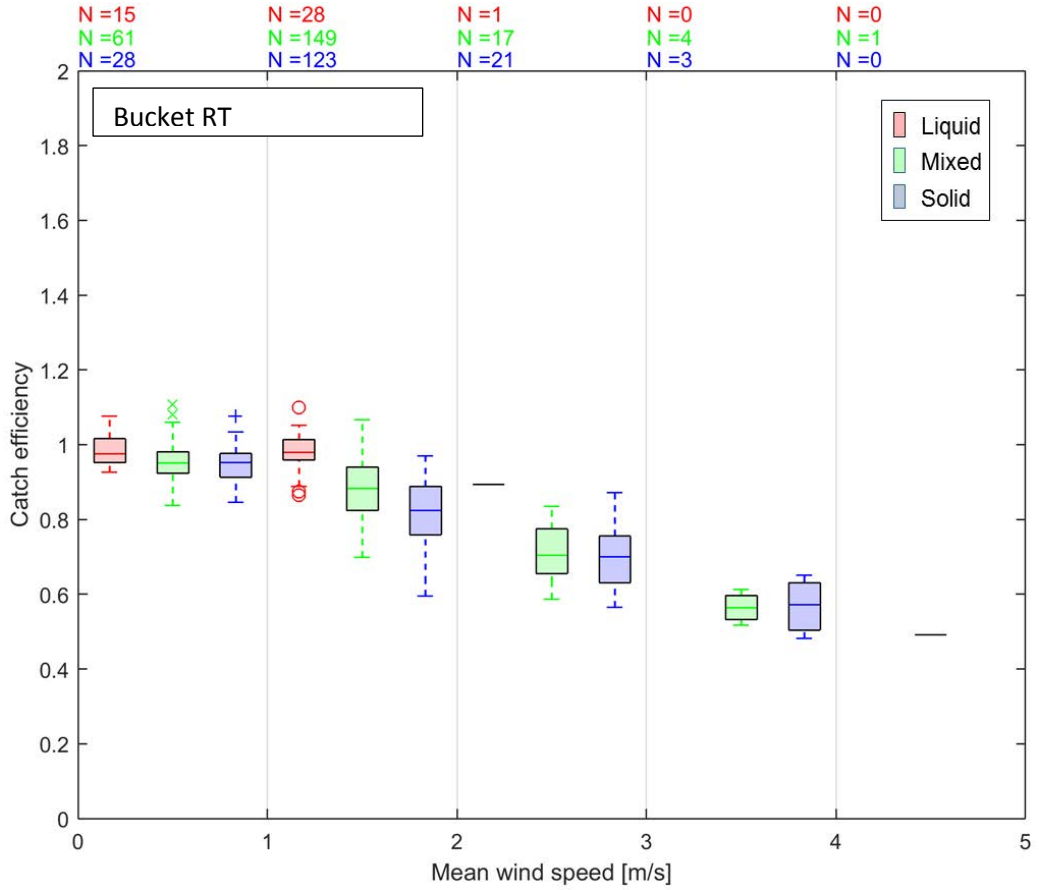
Figure 22: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded OTT Pluvio^2 gauge under test at Sodankylä provided by the gauge manufacturer (Bucket RT and Accumulated NRT outputs).

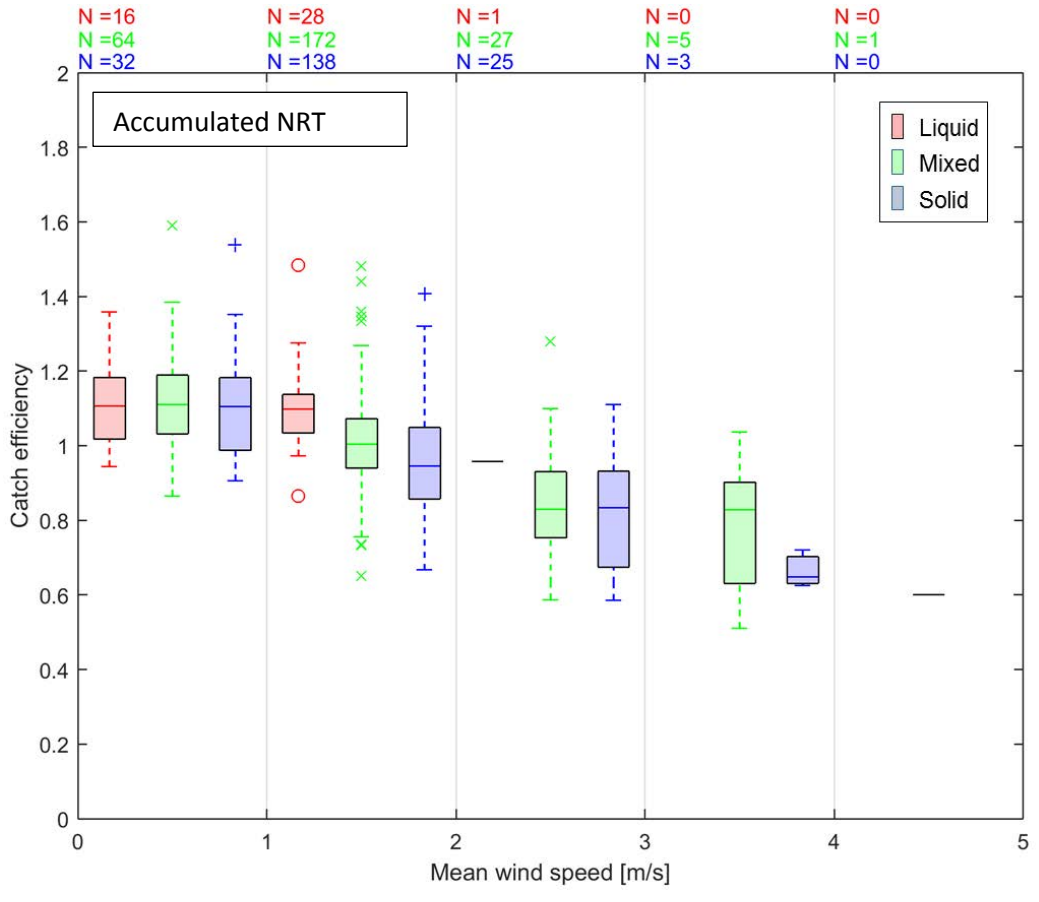
a)



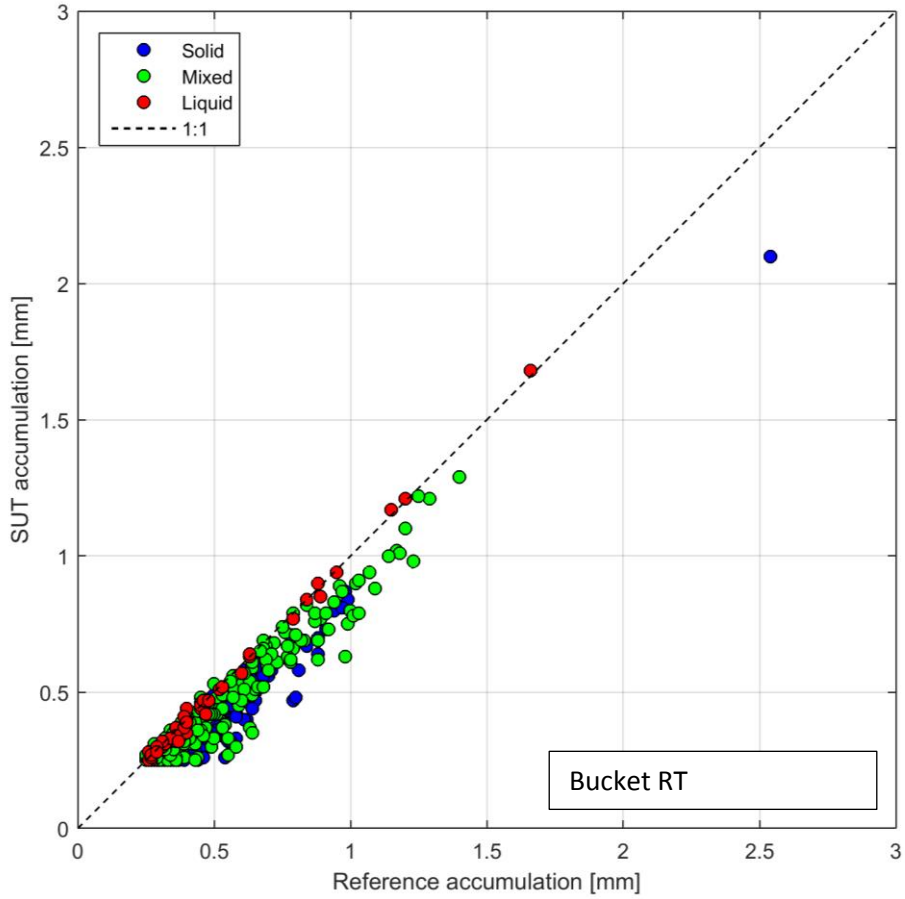


b)


N =15
N =61
N =28
N =28
N =149
N =123
N =1
N =17
N =21
N =0
N =4
N =3
N =0
N =1
N =0
Bucket RT
Liquid
Mixed
Solid
Catch efficiency
Mean wind speed [m/s]


N =16
N =64
N =32
N =28
N =172
N =138
N =1
N =27
N =25
N =0
N =5
N =3
N =0
N =1
N =0
Accumulated NRT
Liquid
Mixed
Solid
Catch efficiency
Mean wind speed [m/s]

c)


Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Bucket RT

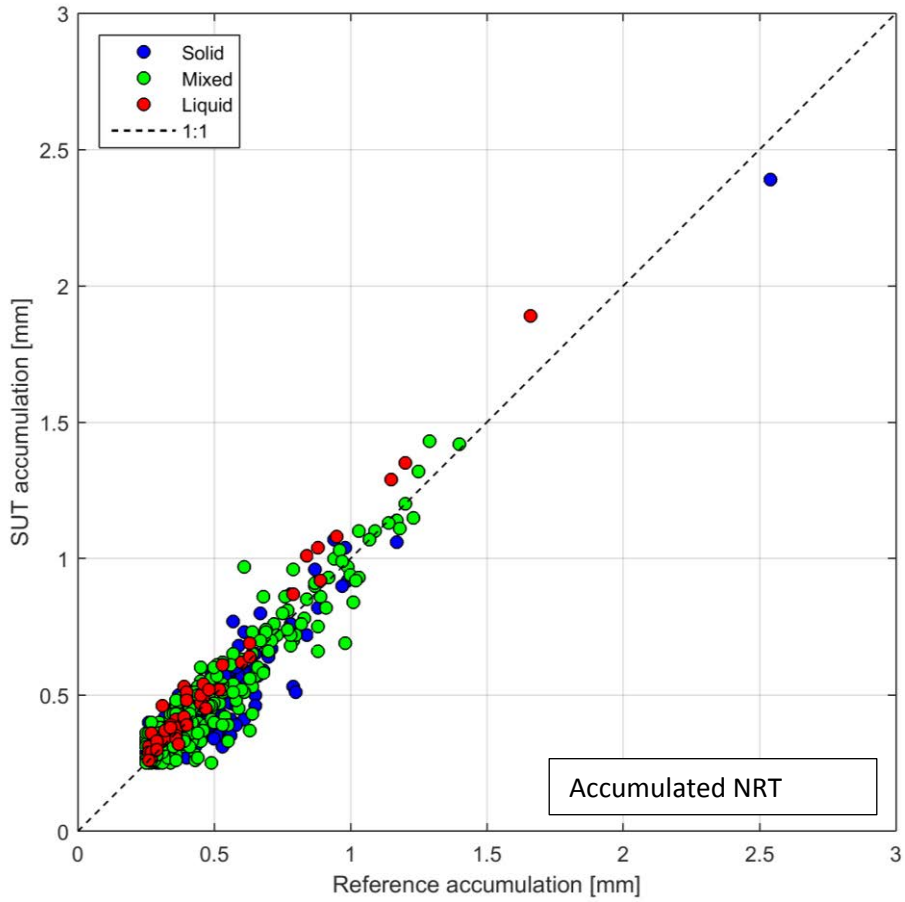
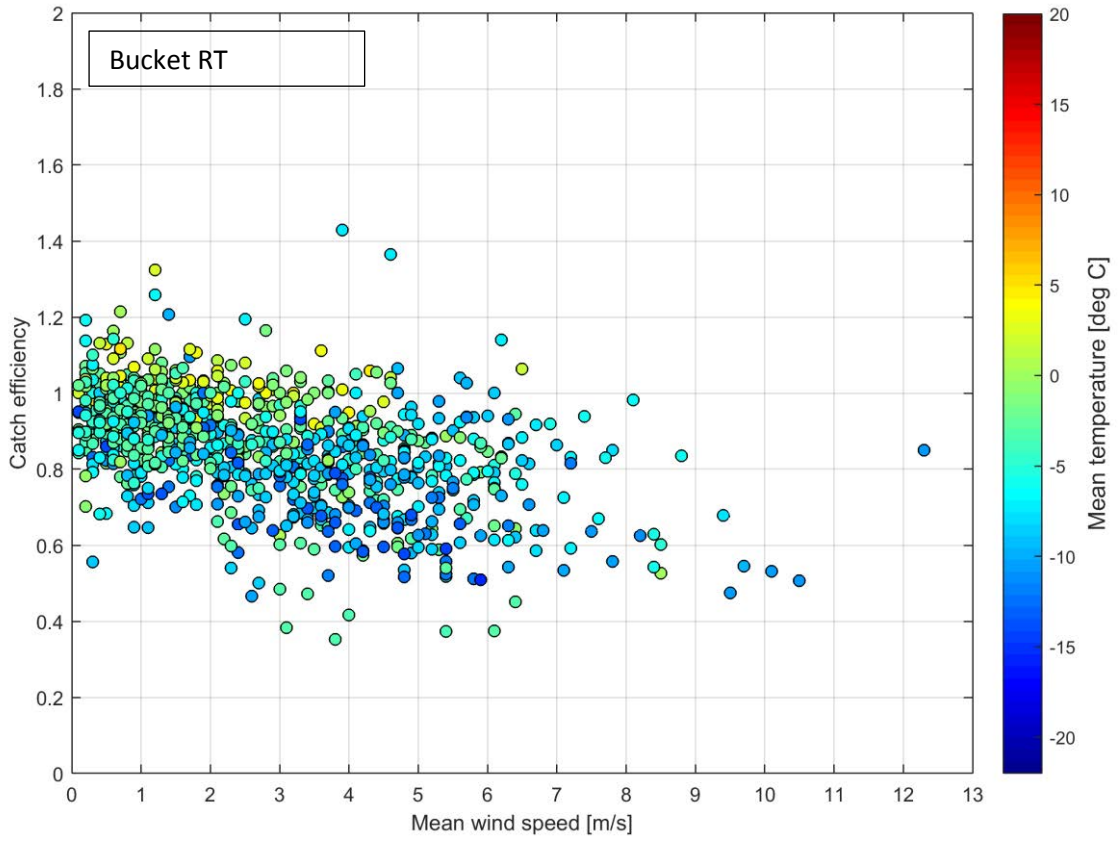
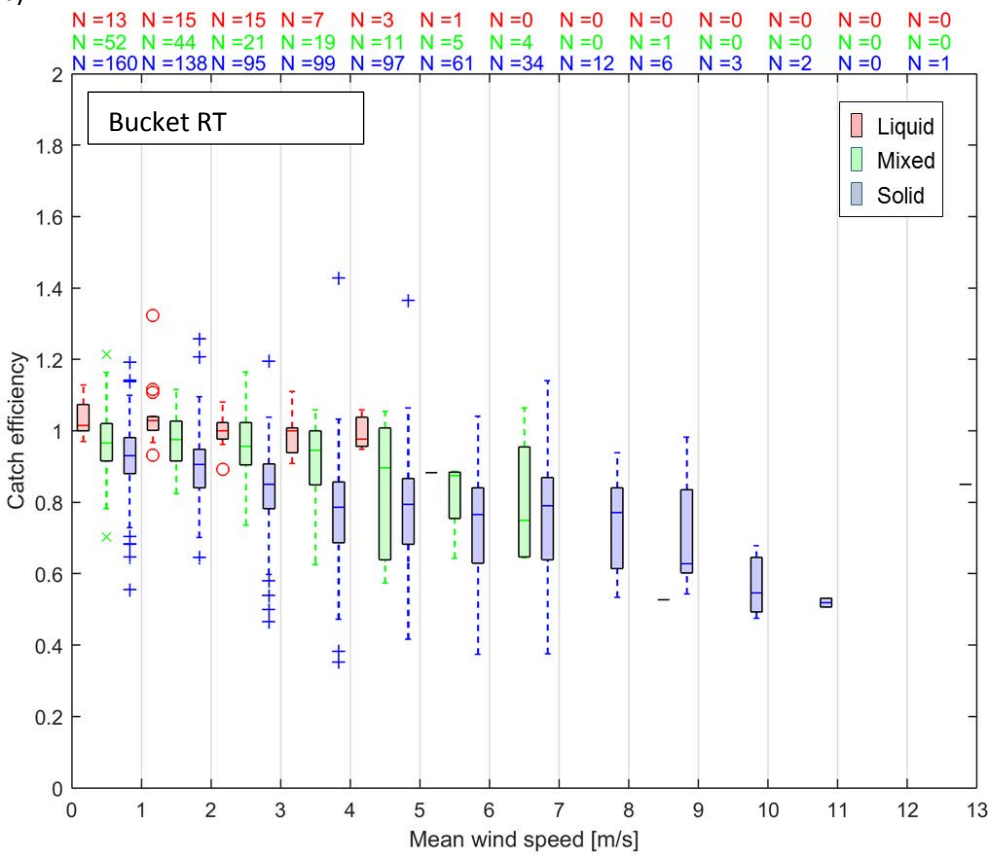

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Accumulated NRT

Figure 23: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the single-Alter shielded OTT Pluvio^2 gauge under test at Weissfluhjoch (Bucket RT output).

a)



b)



c)

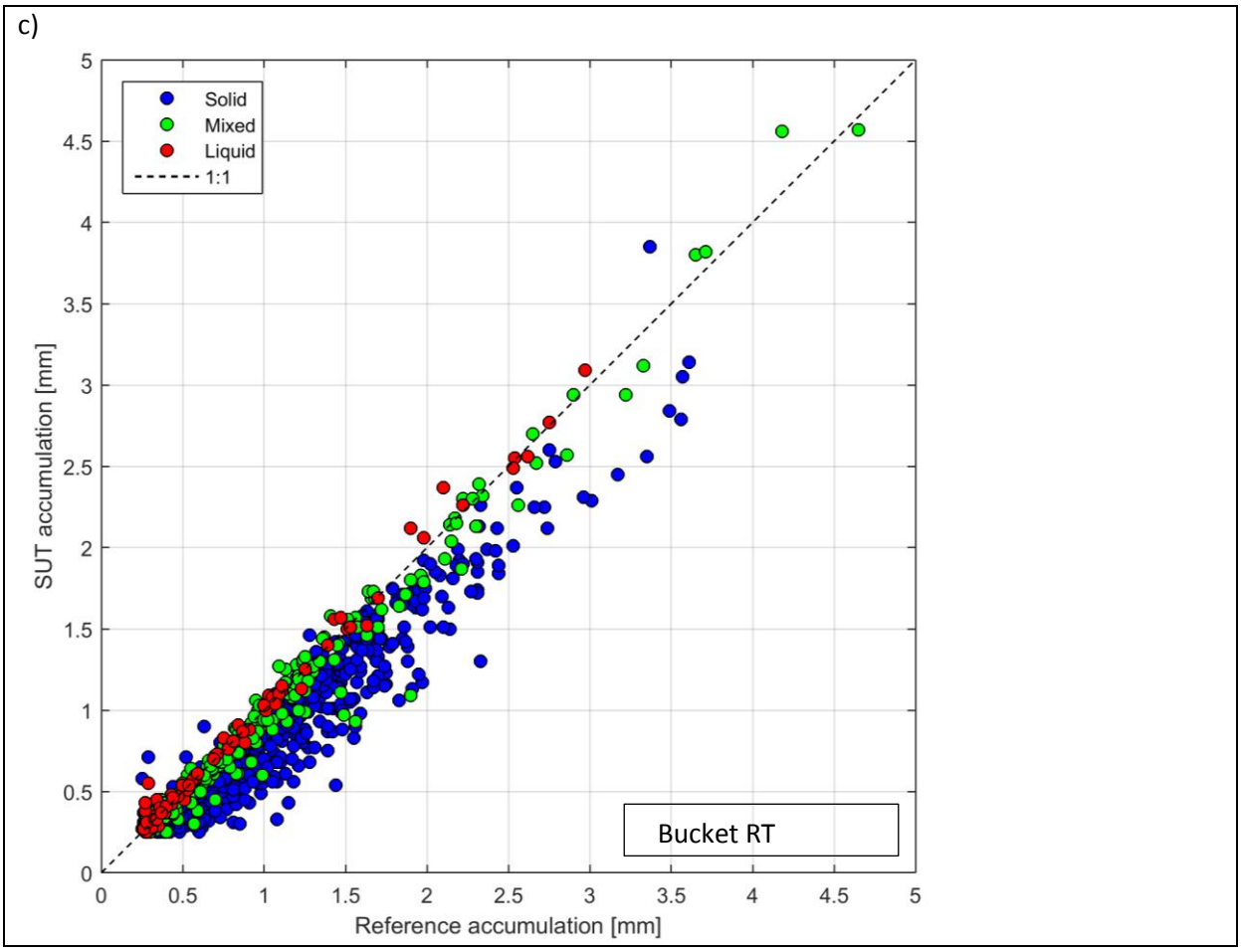
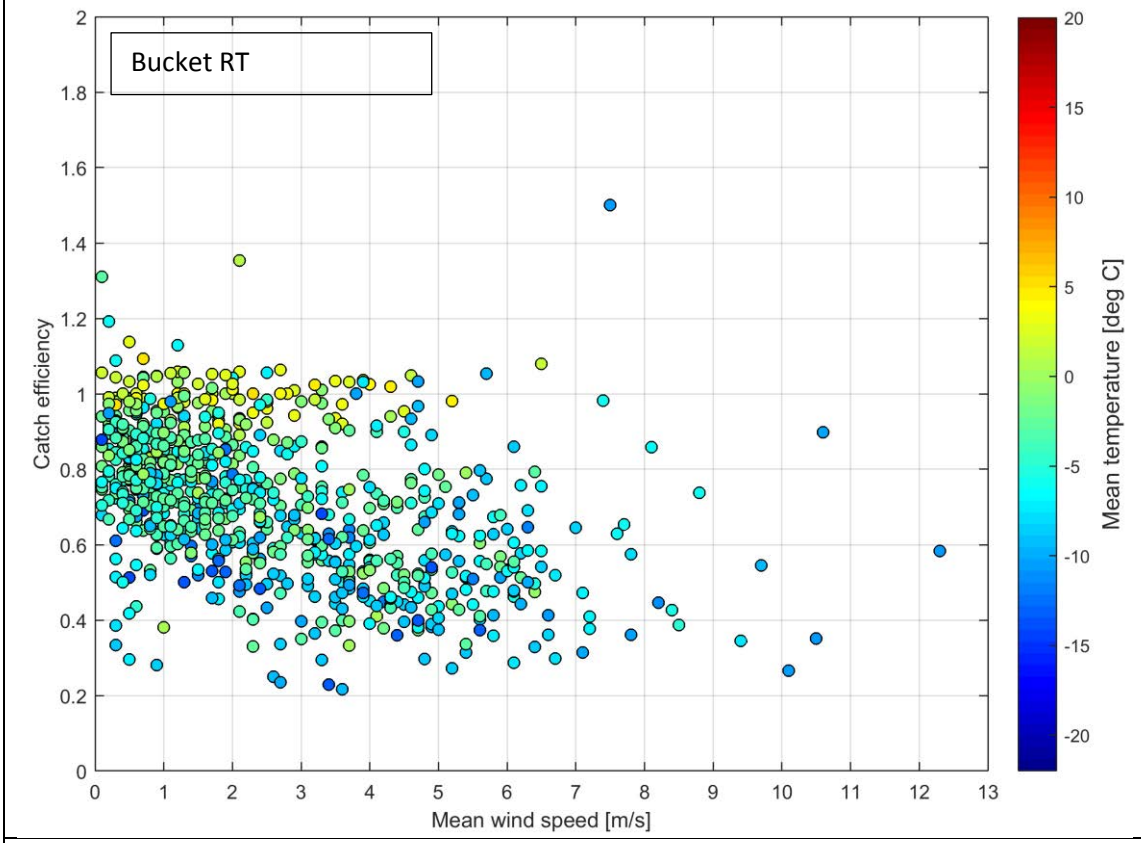
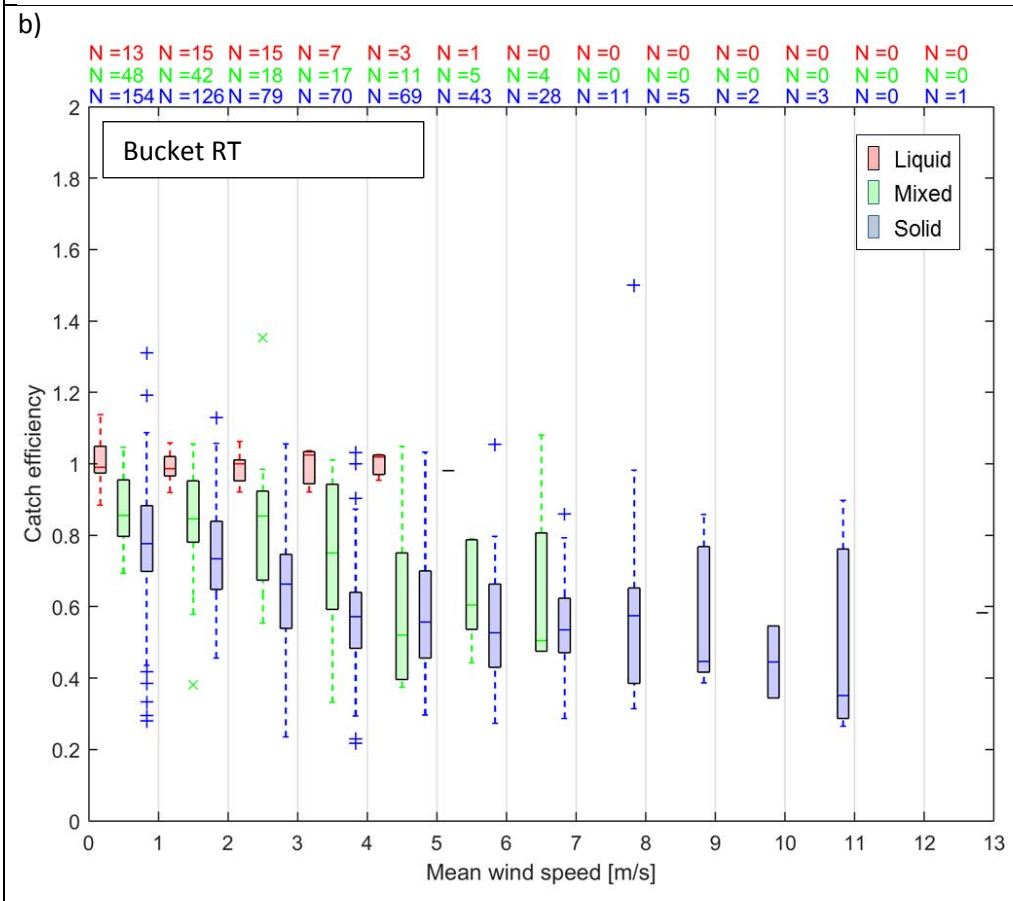

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Bucket RT
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5

Figure 24: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded OTT Pluvio^2 gauge under test at Weissfluhjoch (Bucket RT output).

a)



b)



c)

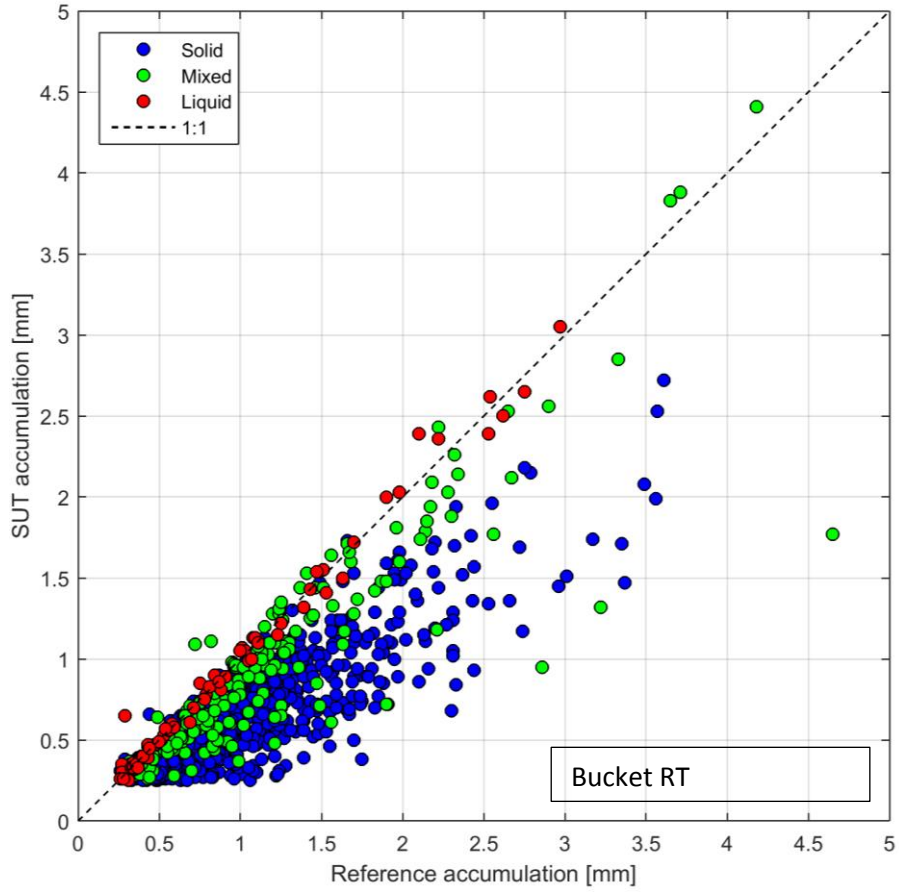
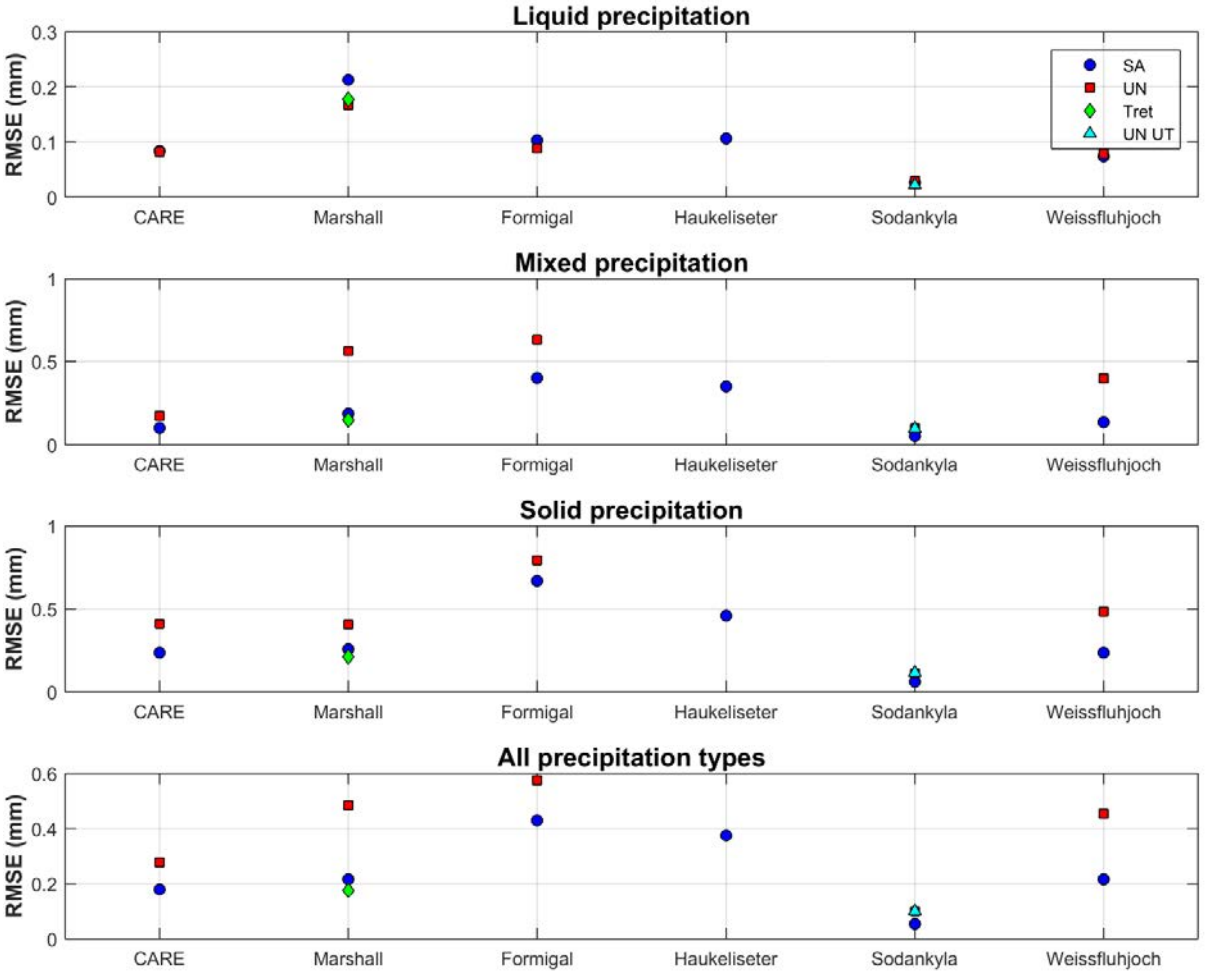
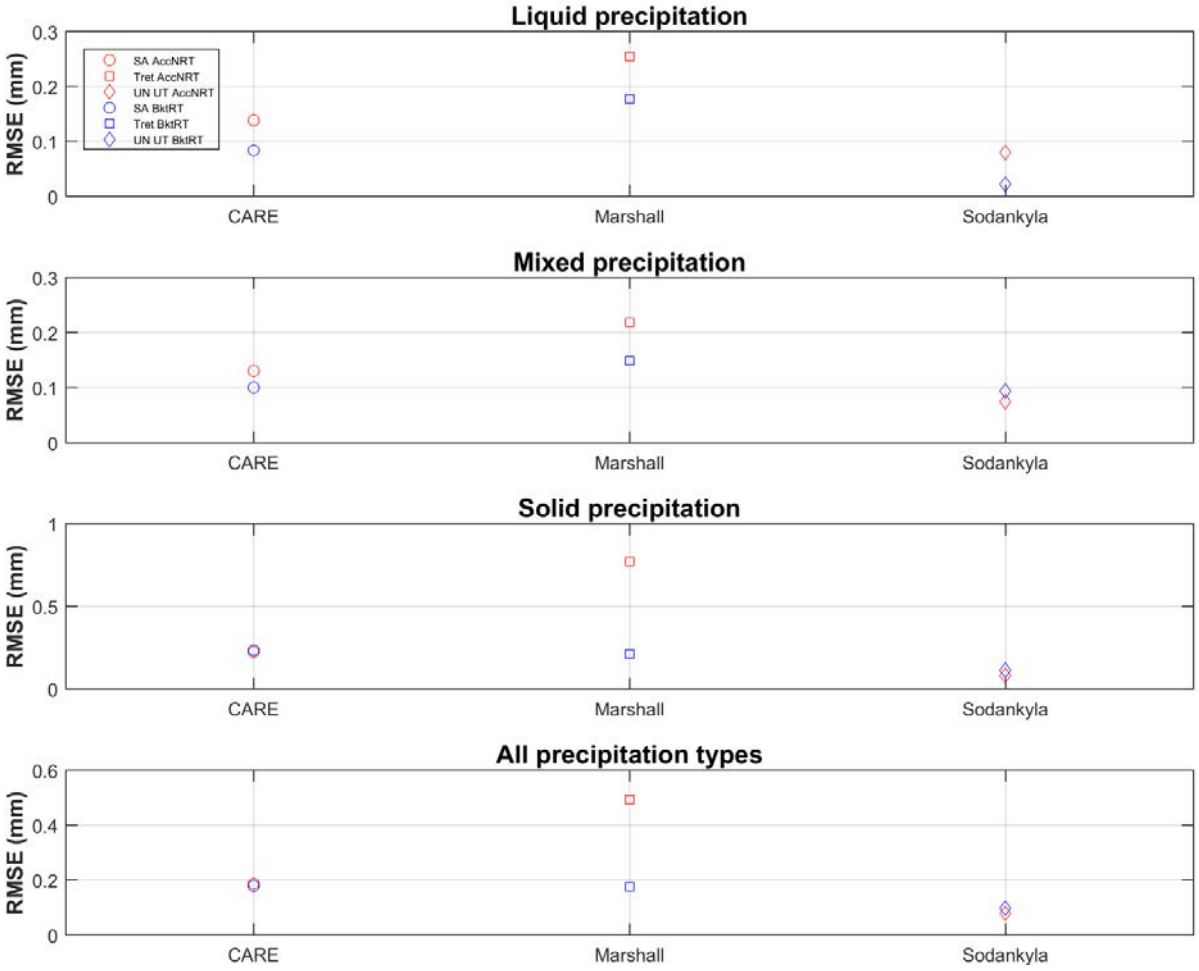

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
Bucket RT
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5

Figure 25: RMSE values over the duration of the test period for (a) Bucket RT data and (b) Accumulated NRT data. Results are presented by site, by configuration, and by precipitation type.

a)



b)


Liquid precipitation
Mixed precipitation
Solid precipitation
All precipitation types
RMSE (mm)
SA AccNRT
Tret AccNRT
UN UT AccNRT
SA BktRT
Tret BktRT
UN UT BktRT
CARE
Marshall
Sodankyla

The RMSE results are tabulated in Table 6 for Bucket RT (black text) and Accumulated NRT (red text) outputs.

Table 6: RMSE values calculated for test configurations by site and by precipitation type for YY cases over the duration of the test period, as presented in Figure 25. Results are presented separately for Bucket RT (black text) and Accumulated NRT (red text) outputs.

Site	Shield	Dataset	RMSE (mm/30 min)			
			Liquid	Mixed	Solid	All precip types
CARE	SA	Bucket RT	0.084	0.100	0.235	0.179
	SA	Accum NRT	0.138	0.131	0.226	0.187
	UN	Bucket RT	0.082	0.173	0.409	0.276
Formigal	SA	Bucket RT	0.103	0.401	0.667	0.429
	UN	Bucket RT	0.089	0.632	0.793	0.575
Haukeliseter	SA	Bucket RT	0.106	0.352	0.459	0.375
Marshall	SA	Bucket RT	0.213	0.184	0.256	0.216
	UN	Bucket RT	0.167	0.564	0.407	0.484
	Tret	Bucket RT	0.178	0.149	0.211	0.176
	Tret	Accum NRT	0.255	0.219	0.771	0.493
Sodankylä	SA	Bucket RT	0.027	0.053	0.060	0.054
	UN	Bucket RT	0.030	0.098	0.112	0.099
	UN UT	Bucket RT	0.022	0.094	0.116	0.099
	UN UT	Accum NRT	0.080	0.075	0.084	0.079
Weissfluhjoch	SA	Bucket RT	0.074	0.134	0.237	0.215
	UN	Bucket RT	0.079	0.398	0.485	0.453

The overall catch ratios calculated using all 30 minute YY cases, over the entire test period, are provided in Table 7. To demonstrate the influence of the SUT accumulation threshold on the results, the overall catch ratios are also provided for all 30 minute YY cases determined using a lower SUT threshold of 0.1 mm/30 minutes. Note that these values reflect only the YY cases, and do not include the amounts corresponding to the cases when the SUT and the reference do not agree on the occurrence of precipitation.

Table 7: Overall catch ratios for test configurations, determined from YY cases over the entire test period at each site, using two different SUT accumulation thresholds. Results are presented for Bucket RT (black text) and Accumulated NRT (red text) outputs.

			Overall catch efficiency	
Site	Shield	Dataset	0.25 mm/30 min SUT threshold	0.1 mm/30 min SUT threshold
CARE	SA	Bucket RT	0.890	0.873
	SA	Accum NRT	0.924	0.906
	UN	Bucket RT	0.838	0.780
Formigal	SA	Bucket RT	0.793	0.778
	UN	Bucket RT	0.739	0.712
Haukeliseter	SA	Bucket RT	0.771	0.670
Marshall	SA	Bucket RT	0.873	0.852
	UN	Bucket RT	0.747	0.705
	Tret	Bucket RT	0.916	0.910
	Tret	Accum NRT	1.072	1.067
Sodankylä	SA	Bucket RT	0.939	0.926
	UN	Bucket RT	0.866	0.835
	UN UT	Bucket RT	0.861	0.839
	UN UT	Accum NRT	0.989	0.958
Weissfluhjoch	SA	Bucket RT	0.863	0.856
	UN	Bucket RT	0.701	0.679

6.4. Ability to detect light precipitation events

The impact of the threshold selection for data processing relative to the detection of light precipitation was examined using four different combinations of reference and SUT accumulation thresholds (four 'cases' in Table 8) for the single-Alter shielded Pluvio^2 gauge under test at the CARE site. Only the Bucket RT output data are considered in this assessment. Contingency results, probabilities of detection (POD), and false alarm rates (FAR) are presented for each case in Table 9. A quantitative comparison of the amounts reported in each case is beyond the scope of this assessment.

Table 8: Reference and SUT thresholds in each case for light precipitation detection assessment.

Case	Reference threshold (mm/30 min)	SUT threshold (mm/30 min)
1	0.25	0.25
2	0.1	0.1
3	0.25	No threshold
4	0.25	0

Table 9: Contingency results, Probability of Detection (POD), and False Alarm Rate (FAR) for each case in light precipitation detection assessment.

Case	Number of events				Skill score (%)	
	YY	YN	NY	NN	POD	FAR
1	435	73	22	14428	85.6	4.8
2	792	166	141	13859	82.7	15.1
3	508	0	14450	0	100	96.6
4	508	0	7594	6856	100	93.7

6.5. Assessment of events when the reference and the SUT do not agree on the occurrence of precipitation

Assessment intervals during which the site reference and SUT do not agree on the occurrence of precipitation – namely, the YN and NY cases (Section 4.1.1) – are characterized for the SA-shielded test gauges at Haukeliseter, Formigal, and Sodankylä (Bucket RT output only) using histograms in Figures 26, 27, and 28, respectively. The histograms include accumulated precipitation reported by the reference and SUT (0.25 mm/30 min threshold for both), precipitation intensity as reported by the reference, and corresponding site conditions.

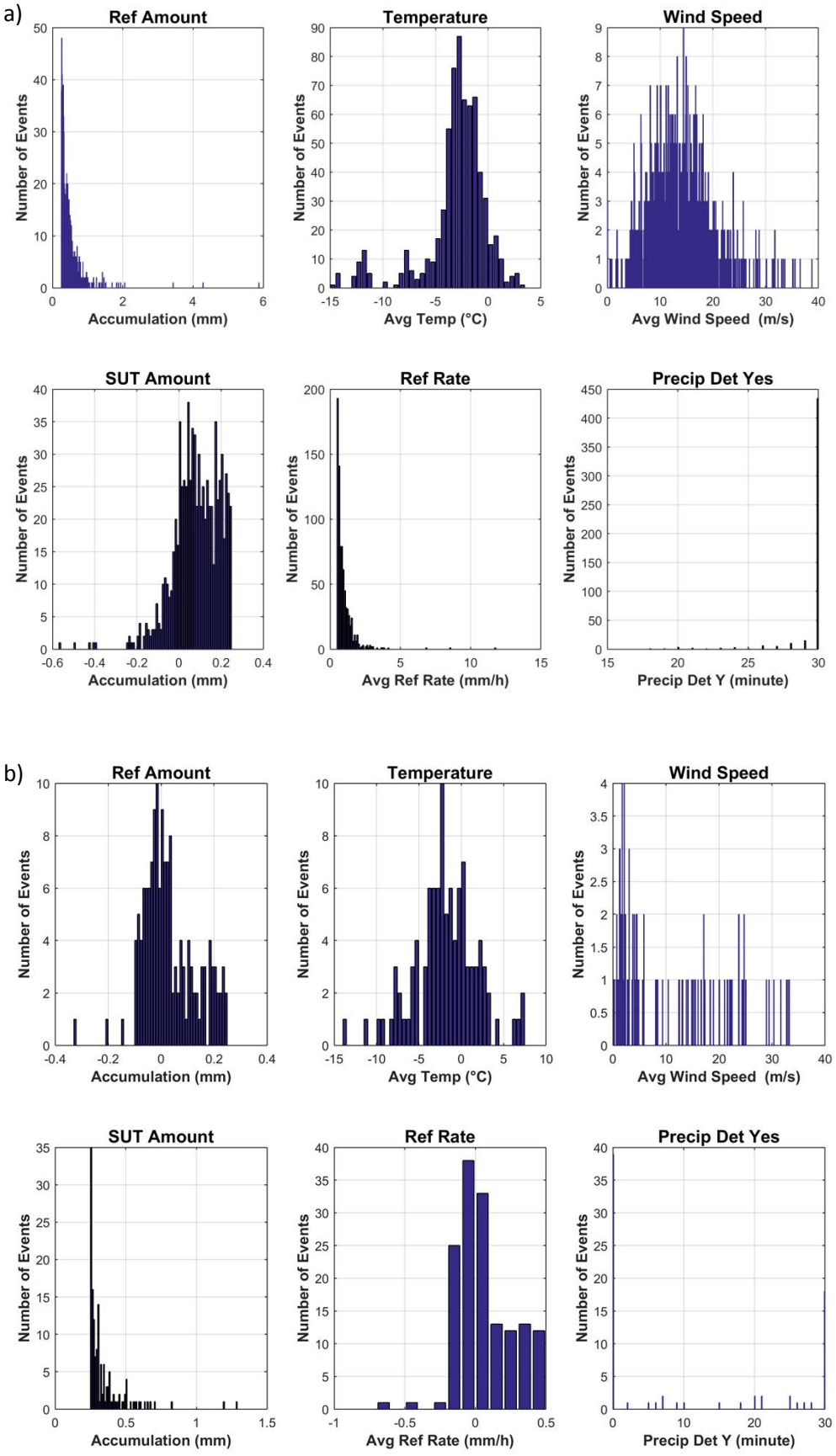


Figure 26: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for (a)YN events (794 total) and (b) NY events (149 total) for the SA-shielded Pluvio^2 test gauge at Haukeliseter (Bucket RT output).

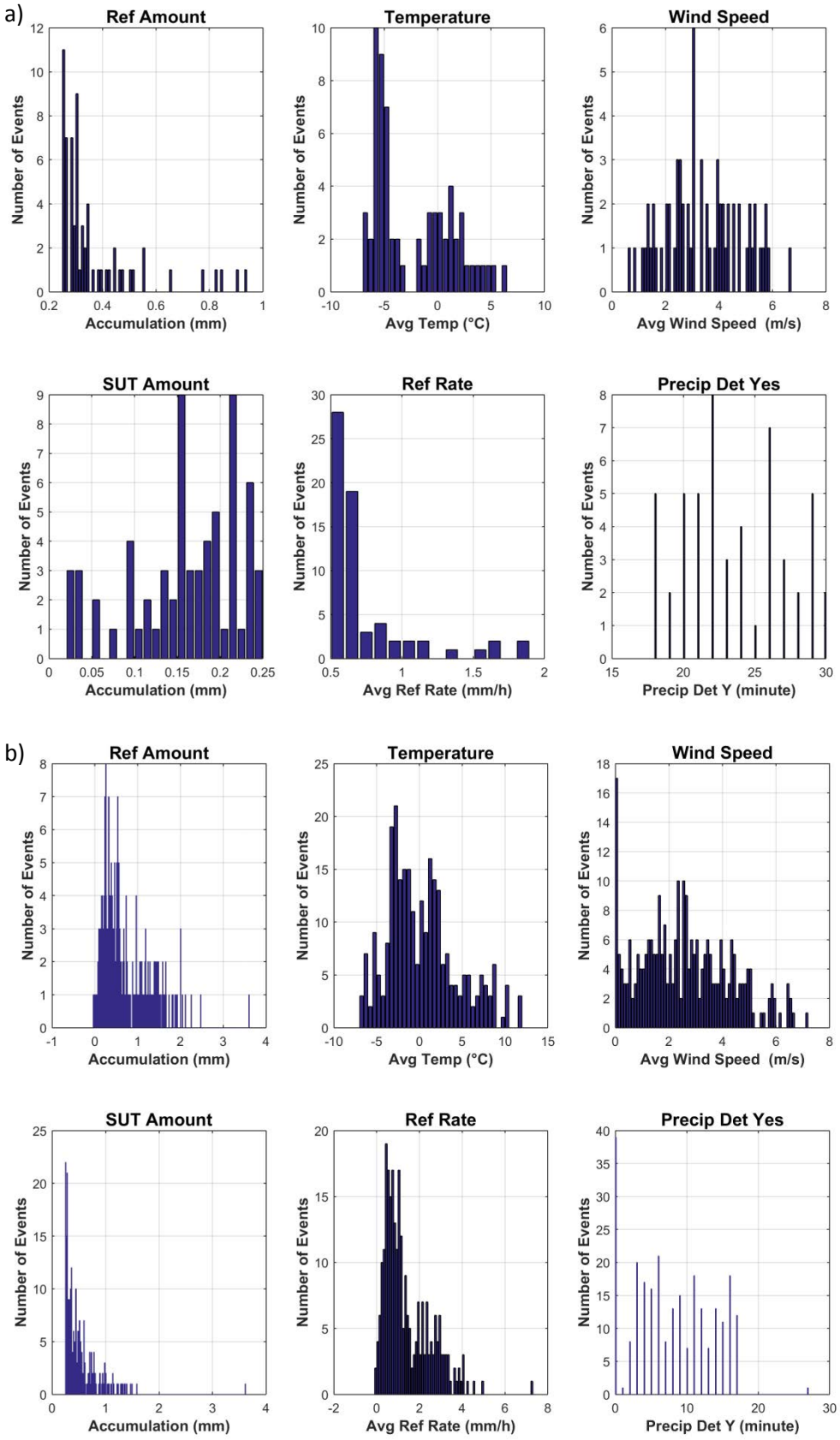


Figure 27: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for (a)YN events (66 total) and (b) NY events (267 total) for the SA-shielded Pluvio^2 test gauge at Formigal (Bucket RT output).

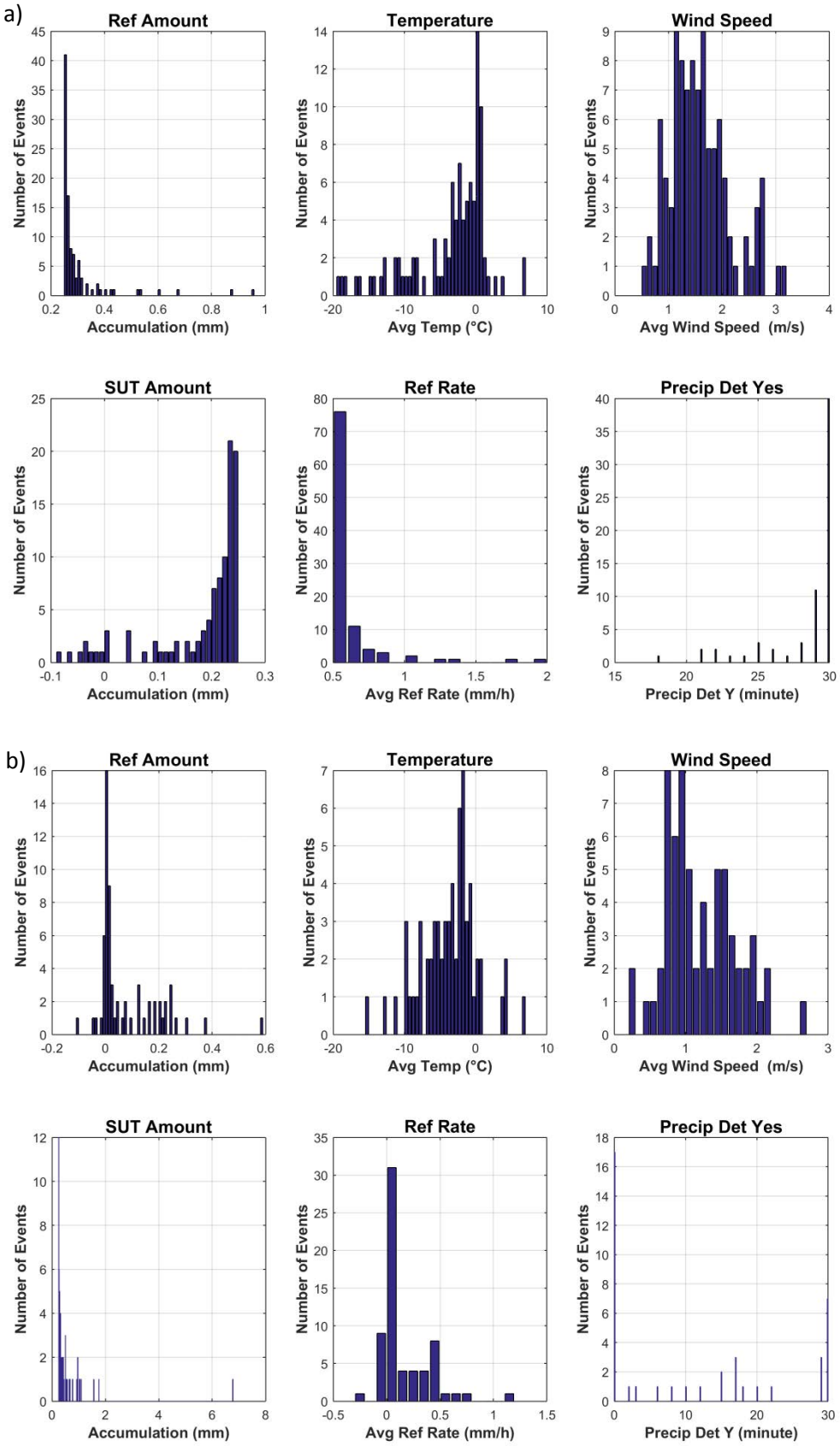


Figure 28: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for (a)YN events (100 total) and (b) NY events (65 total) for the SA-shielded Pluvio^2 test gauge at Sodankylä (Bucket RT output).

7) Interpretation of Results

7.1. Operating conditions

The full ranges of conditions under which the test gauges at each site were operated are illustrated in Figure 3. The conditions relevant to gauge operation are as follows:

- Temperatures between approximately -36 °C and 26 °C;
- Precipitation intensity within approximately 12 mm/hr.

These conditions fall within the manufacturer's specified operating conditions of -40 °C to 60 °C for temperature.

The conditions at each site during precipitation events are shown in Figure 4. Of particular note are the mean wind speeds during precipitation events at Haukeliseter, which extend to approximately 21 m/s; the mean wind speeds during precipitation at all other sites fall within 12 m/s.

7.2. Performance over the range of operating conditions

7.2.1. Non-precipitating conditions

The results presented in Figures 5 to 10 and Table 4 show the variability in the reference and test gauge responses in the absence of precipitation (30 minute intervals with zero precipitation reported by a sensitive and independent precipitation detector) from site to site. The PDFs of Bucket RT responses from test gauges in both unshielded and single-Alter configurations are generally broader than those of the corresponding reference configurations in Figure 5. This is reflected by the larger standard deviation values for these test gauge responses relative to those from the corresponding reference configurations in Table 4.

For shielded and unshielded test gauges at the same site, the PDFs of Bucket RT responses in non-precipitating conditions in Figure 5b and 5c are similar, as are the corresponding standard deviation values in Table 4. These results suggest that the gauge responses in the absence of precipitation are largely independent of the specific shield configuration employed. Greater variability (and larger standard deviation values) may be expected for gauges with shields mounted directly to the gauge post, which may be subject to wind-induced vibration of the shield/gauge assembly. This is one potential explanation for the higher standard deviations observed for SA-shielded gauges relative to unshielded gauges at Formigal, Sodankylä, and Weissfluhjoch.

The PDFs of Accumulated NRT responses from the test gauges submitted by the manufacturers or sites in Figure 6 are much narrower than those of the corresponding reference configurations, and are dominated by 0 mm reports. Accordingly, the standard deviation values for these configurations in Table 4 are smaller than those for the reference configurations at CARE and Sodankylä. The standard deviation of gauge reports in the absence of precipitation for the Accumulated NRT output from the Tretyakov-shielded gauge at Marshall is slightly higher than that of the corresponding reference configuration in Figure 6. This may result from a small number of reports of higher magnitude in the tail of the distribution for the test gauge, which are not apparent in the PDF.

The variability of Bucket RT responses from test gauges in the absence of precipitation appears to be greatest during 30 minute periods with small or no temperature difference (Figure 7), with an apparent, albeit very gradual, trend toward increasing accumulation reports with increasing negative temperature difference. This trend may be related to the static temperature compensation applied to the Bucket RT output by the gauge firmware. The variability of Bucket RT outputs with increasing mean wind speed in non-precipitating conditions varies from site to site (Figure 9), with some enhanced variability observed at low mean wind speeds (within about 6 m/s), and a notable increase in variability at mean wind speeds greater than about 6 m/s for the unshielded test gauge at Marshall, which could potentially result from blowing snow.

No clear trends are observed for the variability of Accumulated NRT reports in non-precipitating conditions as a function of mean wind speed (Figure 10). Considering the variability of the same reports as a function of the temperature difference over the interval in non-precipitating conditions (Figure 8), increased variability is observed for 30 minute periods with small or no temperature difference for all configurations tested. The Tretyakov-shielded configuration at Marshall also showing large variability for intervals over which the temperature decreased; the reason(s) for this is (are) unclear, and could be related to the specific gauge configuration, gauge siting, environmental condition, and/or the processing algorithm applied by the gauge firmware.

Evident upon inspection of Figures 7 to 10 and Table 4 is the potential for false precipitation reports in the absence of precipitation for all test configurations and sites. In the present assessment, an independent precipitation detector is used to corroborate the occurrence of precipitation, which is expected to eliminate many of the cases for which the 0.25 mm over 30 minutes threshold is exceeded; however, sites using only accumulation thresholds for the identification of precipitation events may be more susceptible to false reports, even if employing the processed, Accumulated NRT output.

The magnitude of gauge responses in the absence of precipitation can be used to identify a detection threshold that minimizes the detection of false precipitation while enhancing the detection of light precipitation. This threshold is considered to be three times the standard deviation of the average gauge response during 30 minute non-precipitating periods. Based on the present results, this minimum detection threshold is determined to be 0.15 mm for the Bucket RT output from unshielded gauges, and 0.20 mm for the Bucket RT output from shielded gauges. While these values are similar, the slightly higher threshold for shielded gauge may result from wind-induced vibration of configurations in which the shield assembly is mounted to the gauge post.

For the Accumulated NRT output, this detection threshold is 0.02 mm for unshielded gauges and 0.06 mm for shielded gauges. Hence, the application of the processing algorithm reduces the variability of test gauge reports in the absence of precipitation, lowering the detection thresholds relative to the Bucket RT output.

7.2.2. Precipitating conditions

7.2.2.1. Ability to detect and report precipitation

The skill scores for the Bucket RT output from OTT Pluvio^2 test gauges in Figure 11 vary by configuration, and by site. In general, single-Alter shielded and Tretyakov shielded configurations have higher Probability of Detection, higher Bias, and higher Heidke Skill Scores relative to the unshielded configurations at a given site. The False Alarm Rate is comparable for both shield configurations, or lower for the shielded configuration, for test gauges with the shield assembly mounted separately from the gauge post (CARE, Marshall). For test gauges with the shield assembly mounted to the gauge post, the False Alarm Rate is comparable for both configurations, or higher for the shielded configuration, likely on account of wind-induced vibration increasing accumulation reports above the detection threshold of 0.25 mm.

With the exception of the test gauge at Haukeliseter, the Probability of Detection for shielded gauges at all sites is greater than 80%, while that for all unshielded test gauges and sites is approximately 80%, or lower (Figure 11). The POD for the shielded test gauge at Haukeliseter is the lowest among the gauges tested, which is attributed to the higher mean wind speeds reducing gauge catch. Wind speed effects at Haukeliseter mitigate the number of false alarms (low FAR), but also bias the gauge reports low relative to the reference (low Bias), resulting in the lowest overall Heidke Skill Scores.

The effects of environmental conditions are demonstrated further by the higher relative False Alarm Rates for the test gauges at Formigal (Alpine climate with Maritime influence) and Weissfluhjoch (Apline climate with complex topography) in Figure 11. These high FAR values result from the large numbers of NY cases relative to YY cases for the test gauges at these sites as indicated in Table 5. These higher relative FAR values result in lower relative Heidke Skill Scores compared to the test gauges at CARE and Marshall (both in Continental climates) and at Sodankylä (Northern Boreal climate, sheltered site).

For test gauges submitted by the gauge manufacturer (Tretyakov-shielded gauge at Marshall, unshielded gauge at Sodankylä) or site host organization (single-Alter shielded gauge at CARE), the skill scores determined using both the Bucket RT and Accumulated NRT outputs are compared in Figure 12. The results vary from site, on account of differences in the specific configuration, siting, and environmental conditions. The skill scores are comparable for the test gauge at CARE, with a higher False Alarm Rate for the Accumulated NRT data relative to the Bucket RT data, likely resulting from the processing approach applied to the NRT data. This could be 'phantom accumulation' resulting from changes in temperature over the assessment interval, or an impact of the consideration of light accumulation in the NRT data, which is not considered in the Bucket RT data; most likely, the observed trend results from a combination of these factors. The higher FAR determined for the Accumulated NRT data results in a higher Bias and lower Heidke Skill score relative to the Bucket RT output.

A similar trend is observed for the Tretyakov-shielded test gauge at Marshall in Figure 12, but with greater differences in the skill scores determined for each gauge output. Higher False Alarm Rates are observed for this gauge relative to the SA-shielded gauge at CARE, possibly related to the fixed shield slats of the Tretyakov shield and their proximity to the gauge orifice, which may increase the

propensity for snow to accumulate on the slats and blow into the orifice. This could explain how the Heidke Skill scores for both outputs are lower relative to those for the test gauge at CARE, despite higher Probabilities of Detection; however, site-to-site differences in siting, installation, and conditions also likely play a role. The higher FAR for the Accumulated NRT output relative to the Bucket RT output for the test gauge at Marshall is likely related to the processing approach, as considered above. This results in the lower Heidke Skill score for the Accumulated NRT output relative to that for the Bucket RT output.

For the unshielded test gauge at Sodankylä, all skill scores are higher for the Accumulated NRT output relative to the Bucket RT output. The precipitation conditions at Sodankylä (Northern Boreal climate) are characterized by lower event accumulations relative to CARE and Marshall (both in Continental climates), which may be well-suited for the Pluvio^2 processing algorithm, which tracks the accumulation contribution from light precipitation and adds it to the processed output when it exceeds a defined threshold (0.2 mm within an hour).

7.2.2.2. Ability to report accumulated precipitation

The results presented in Figures 12 to 24, which are based on 30 minute events during which the reference and test gauge both detect precipitation (YY cases), illustrate the influence of wind speed and precipitation type on gauge catch efficiency. The discussion below will focus on snow events; the number of rain events during winter is limited, and the results for mixed events are variable due to the variability in the size and density of precipitation within the mixed regime, as well as the potential for transitions between phases. For solid precipitation events, the catch efficiency decreases with increasing wind speed; this decrease is more rapid for unshielded gauges relative to shielded gauges. For example, for the test gauges at CARE and Marshall, the median catch efficiencies for mean wind speeds between 5 m/s and 6 m/s are between 0.4 and 0.5 for single-Alter and Tretyakov shielded gauges, and between 0.2 and 0.3 for unshielded gauges (see Figures 12 and 13 for test gauges at CARE, and Figures 14 to 16 for test gauges at Marshall). While the same general trend – in which the median catch efficiency is higher within a given mean wind speed bin for a shielded gauge, relative to an unshielded gauge – applies, the specific relationships vary among single-Alter shielded gauges and among unshielded gauges at the remaining sites (Formigal, Haukeliseter, Sodankylä, Weissfluhjoch). These results illustrate how results for the same test gauge, in the same shield configuration, are impacted by the specific gauge configuration, siting, and environmental conditions at a given site.

For test gauges submitted by the manufacturer or site hosts, results can be compared for both Bucket RT and Accumulated NRT outputs, the former being closer to the raw gauge output, and the latter being a more processed output (see Section 2). For the single-Alter shielded and Tretyakov shielded test gauges at CARE and Marshall, respectively, the median catch efficiencies for solid precipitation considering the Bucket RT and Accumulated NRT outputs agree within 0.1 in each 1 m/s mean wind speed bin, with the NRT output consistently reporting the higher median value (see Figure 12b for the test gauge at CARE, and Figure 16b for the test gauge at Marshall). A similar trend is observed for the unshielded test gauge at Sodankylä in Figure 22b, but with a greater difference between the median catch efficiencies in each mean wind speed bin, with values for Accumulated NRT outputs exceeding those for Bucket RT outputs by 0.2 to 0.3. This greater difference at Sodankylä is attributed to the contribution of light precipitation, which occurs frequently at

Sodankylä, and may be better captured using the Accumulated NRT processing algorithm. It should be noted, however, that the greater number of catch efficiency values > 1 for the Accumulated NRT output in Figure 22 indicate the potential for the algorithm to over-report the accumulation over 30 minutes relative to the site reference configuration.

Root mean square error values were computed from all 30 minute events during which each test configuration and the corresponding reference configuration both detected precipitation. Values are shown in Figure 25 and Table 6, and can be considered to represent the absolute uncertainty of each test configuration relative to the reference configuration in liquid, mixed, and solid precipitation, and in all precipitation types (note that the relative proportions of events of each phase differ by site). For the Bucket RT output (Figure 25a), the RMSE for shielded gauges in all precipitation types ranges from approximately 0.05 mm/30 min for the gauge at Sodankylä (sheltered, low wind site with characteristically light precipitation), to approximately 0.2 mm/30 min for the gauges at CARE, Marshall, and Weissfluhjoch, to about 0.4 mm/30 min for the gauges at Formigal and Haukeliseter. The RMSE values for the Bucket RT output for unshielded gauges exceeded those for the shielded gauges at all sites where both configurations were tested, with the differences varying by site from about 0.05 mm/30 min for the gauges at Sodankylä to about 0.3 mm/30 min for the gauges at Marshall.

The RMSE values for Bucket RT and Accumulated NRT outputs are provided for test gauges submitted by the gauge manufacturer and site hosts in Figure 25b. The RMSE values for both outputs in all precipitation types are comparable for the single-Alter shielded test gauge at CARE (about 0.2 mm/30 min) and the unshielded test gauge submitted by the gauge manufacturer at Sodankylä (about 0.1 mm/30 min). This suggests that on a per-event basis (30 minute intervals in the present assessment), the absolute uncertainty of test gauges is not impacted significantly by the specific data output considered. For the Tretyakov-shielded test gauge at Marshall, however, the RMSE value for the Accumulated NRT output is about 0.5 mm/30 min, compared to approximately 0.2 mm/30 min for the Bucket RT output. This difference may be related to the presence of the Tretyakov shield, which has fixed slats and is in close proximity to the gauge orifice. Snow may accumulate on the slats and blow into the orifice, which may be interpreted as light precipitation by the processing algorithm.

The overall catch ratio – computed from the total reference and SUT accumulation from all YY cases over the duration of formal tests – is provided for each test configuration and gauge output in Table 7. For the Bucket RT output from shielded test configurations, the overall catch ratio ranges from 0.77 for the single-Alter shielded gauge at Haukeliseter (highest mean wind speeds) to 0.94 for the single-Alter shielded gauge at Sodankylä (lowest mean wind speeds). For the Bucket RT output from unshielded test configurations, the overall catch ratio ranges from 0.70 for the gauge at Weissfluhjoch to 0.87 for the gauge at Sodankylä. At all sites testing shielded and unshielded configurations, the overall catch efficiency is greater for the shielded configuration.

The overall catch efficiencies for Accumulated NRT outputs exceed those for Bucket RT outputs for all configurations submitted by site hosts or the gauge manufacturer (Table 7), with values ranging from 0.92 (SA-shielded test gauge at CARE) to 1.07 (Tretyakov-shielded test gauge at Marshall). The closer proximity of overall catch efficiency values to 1 for the Accumulated NRT outputs indicates improved agreement with the reference configuration at each site relative to the Bucket RT outputs; however,

given the potential for the NRT output to over-report accumulation relative to the reference configuration (see, for example, the event catch efficiencies > 1 in Figures 16 and 22), these data should be used with caution. Further work is required to distinguish and quantify the contributions from false reports in the Accumulated NRT output.

Decreasing the SUT accumulation threshold for precipitation events from 0.25 mm/30 min to 0.1 mm/min decreases the overall catch efficiency for all test configurations, sites, and data outputs (Table 7). These results demonstrate that lowering the detection threshold for the test configurations does not improve the overall agreement with the reference configuration in terms of the total event accumulation over the test period. Sensitivity of the results to the specific threshold selected are apparent; however, with the exception of the configuration at Haukeliseter (which is subject to markedly higher wind speeds than test configurations at the other sites), the differences in the overall catch efficiency using the lower threshold (0.1 mm/30 min) are within 6% of the values using the standard (0.25 mm/30 min) threshold.

7.2.2.3. Ability to detect light precipitation

The detection thresholds and results presented in Tables 8 and 9, respectively, indicate that decreasing or removing the SUT detection threshold (while maintaining the reference detection threshold at 0.25 mm) leads to increases in both the Probability of Detection and False Alarm Rate to close to 100%, indicating negligible detection skill (e.g. Heidke Skill Score of close to 0%). Decreasing both the reference and SUT detection thresholds to 0.1 mm decreases the POD by about 3% and increases the FAR by over 200%. These results suggest that reducing both detection thresholds does not improve the detection of light precipitation events relative to the 0.25 mm thresholds; rather, it increases significantly the number of false alarm (NY) events.

7.2.3.4. Assessment of events when the reference and SUT do not agree on the occurrence of precipitation

The YN ('miss') cases, when the reference detects a precipitation event and the SUT does not, and NY ('false alarm') cases, when the SUT detects a precipitation event and the reference does not, are characterized for selected test configurations in Figures 26, 27, and 28. The majority of the YN cases (Figures 26a, 27a, and 28a) have reference accumulations just above the 0.25 mm threshold and SUT accumulations below the threshold, likely resulting from enhanced wind effects (and lower gauge catch) for the single-Alter shielded SUT configurations relative to the reference gauge in the DFIR-fence. The results for the false alarm (NY) cases in Figures 26b, 27b, and 28b illustrate the influence of the detection criteria on the assessment results. For example, the NY cases for the test gauge at Formigal in Figure 27b are all events during which the precipitation detector condition (≥ 18 minutes of precipitation occurrence) was not met, while the NY cases for the test gauges at Haukeliseter (Figure 26b) and Sodankylä (Figure 28b) are predominated by events during which both the 0.25 mm reference accumulation threshold and precipitation detector condition are not met.

8) Maintenance

Gauge calibration: each site completed the gauge field calibration and verification as per manufacturer recommendations, at least once a year or following the emptying of the gauge. The calibration records have been stored by each site host.

9) Performance Considerations

9.1. Data processing

The Pluvio^2 output includes data products of various complexities, resulting from the application of gauge-specific algorithms. Some of these products (e.g. Accumulated NRT) are user ready. While summary descriptions of these products are provided by the manufacturer, the instrument manual should clearly identify and describe each product and the related processing approach to better equip users with the information required for data interpretation. Disclosure of the processing logic used in the algorithms could help to explain differences among the data products, and could lead to recommendations for how the data available could be used in a complementary way to improve the overall reliability of data in operations.

The proprietary nature of the algorithms also presents a challenge for the use of Pluvio^2 data for climate applications, where traceability over time is important. For example, the impact of any changes made to the algorithms over time on the gauge data will be difficult to assess. Further disclosure of the processing approach is recommended in this regard, as well as the addition of the firmware version to the standard output message to facilitate the tracking of changes over time.

For the specific firmware version tested, the potential for false precipitation reports was identified in both the Bucket RT and Accumulated NRT data outputs. Additional work is required to characterize the specific conditions leading to false reports, and to quantify their contributions to total precipitation accumulation over different time scales (e.g. seasonal).

The SPICE data quality control and event selection procedures have been applied to filter both Bucket RT and Accumulated NRT data from Pluvio^2 gauges and derive precipitation amounts over specified intervals. Such procedures are recommended to reduce the variability in gauge reports and establish analysis-ready data products. The specific threshold employed for distinguishing between precipitation and signal noise is an important consideration; the present results indicate similar performance in terms of overall catch efficiency for accumulation thresholds of 0.25 mm and 0.1 mm.

9.2. Gauge configuration

The configuration and installation of Pluvio^2 gauges at a given site can impact significantly gauge performance. The single-Alter shielded test configurations show higher overall catch efficiencies, lower RMSE values, and more gradual decreases in median catch efficiency for solid precipitation with increasing wind speed relative to the unshielded test configurations. Shielded configurations are therefore recommended, where possible. All field configurations must be fully tested and validated prior to use. Where possible, mounting the wind shield separate from the gauge post is recommended to avoid data impacts due to wind-induced vibration of the shield assembly.

The shape of the gauge housing appears to be prone to snow accumulation on the 'shoulder' when the wind speeds are low (Figure 29). Given the proximity of this accumulation to the gauge opening, there are risks to data from two perspectives: 1) snow collected on the shoulder could be blown into the gauge and result in false accumulation; and 2) excessive accumulation on the shoulders could lead to gauge capping and missed/erroneous reports. An additional aspect related to capping is that the advanced data products (e.g. Accumulated NRT) include filters that could remove a sudden jump in data caused by the capping bridge (partial or total) dropping inside the bucket. In such cases, the accumulation related to the snow cap would not be reported (in addition to any precipitation that was not collected due to the cap), leading to underestimation of accumulation totals. Addressing the issue of snow accumulation is strongly recommended, either through redesign of the outer shell or the use of heaters in the shoulder region.

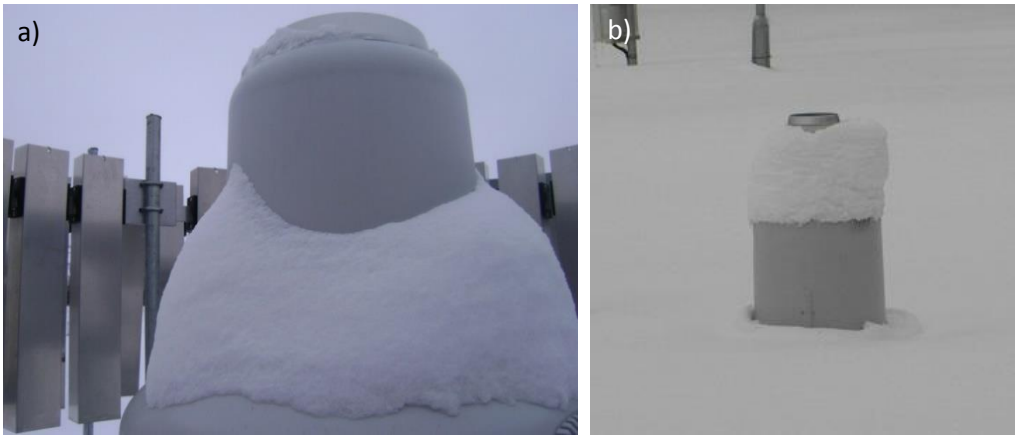


Figure 29: Snow accumulation on the shoulder of Pluvio^2 gauges at (a) CARE and (b) Sodankylä.

The Pluvio^2 derived data products (e.g. Accumulated NRT) account for the daily evaporation of the bucket content; this is not the case for the Bucket RT (and Bucket NRT) data products. In the absence of a film of oil in the bucket to prevent evaporation, the Bucket RT data continues to report the weight of the bucket content, which could decrease significantly depending on the external conditions. In the absence of a film of oil, the content of the bucket will evaporate, thereby reducing the need for emptying the bucket as frequently. This could be an advantage from an operational perspective, by lengthening service intervals, but the impact must be understood well. It is recommended that the instrument manual clearly identifies the advantages and the impact of the use of oil, to give the users sufficient information to decide how to best use the instrument.

9.3. Ancillary measurements and adjustments

Ancillary measurements from a sensitive precipitation detector that is independent from the test gauge are recommended to help distinguish precipitation events from false reports due to noise or gauge processing. The application of adjustment functions is strongly recommended to account for the reduction in gauge catch efficiency as the wind speed increases. Additional ancillary measurements of wind speed (ideally at gauge height) and air temperature are required for the application of transfer functions.

WMO-SPICE Instrument Performance Report
Sutron TPG

1) Technical Specifications (from manufacturer provided documentation)

Instrument model:	Sutron Total Precipitation Gauge TPG-0001-1
Physical principle :	Weighing gauge (WG) based on catchment principle; employs a load cell transducer for weight measurement
Capacity :	914 mm (36” standard)
Collecting area:	324 cm^2 (8” diameter aluminum inlet with knife edge)
Operating temperature range:	-40 °C to 60 °C
Measurement uncertainty:	0.6 mm, from -25°C to +60°C; 0.3 % of full scale, from -40°C to +60°C (as defined by manufacturer in 05/27/2010 datasheet)
Sensitivity:	0.025 mm (as defined by manufacturer)

a)

b)

Figure 1: Sutron TBG gauges under test in (a) unshielded and (b) single-Alter shielded configurations at Marshall site.

2) Data output format

Gauge data output: (according to instrument manual provided by manufacturer)	Precipitation in the gauge collection container is weighed by a load cell that generates a raw signal in millivolts. This raw signal is converted to a precipitation amount using the following equation: Precip (mm) = Signal (mV) x slope + offset + field calibration offset where the values of the slope, offset, and field calibration offset are generated during the calibration process. The user can select sample and average intervals, allowing the precipitation amount to be computed as an average of a number of samples.

3) SPICE test configuration

Number of transducers:	One (the standard model has a single load cell)
Shield:	Unshielded (UN) and single-Alter Shield (SA)
Test sites:	Marshall (USA)
Sensor provider(s):	All gauges evaluated were provided by the instrument manufacturer (Sutron)

A map showing the test site location is provided in Figure 2.

3.1. Note on terms and acronyms used

Throughout this document, the following notations are used to identify the R2 reference (Ref) and senor under test (SUT) configurations:

Reference: 'DFIR' and 'DFAR' are used interchangeably for the R2 reference configurations. 'DFIR' refers to an automated gauge installed in a DFIR-fence, while 'DFAR' refers more explicitly to the Double-Fence Automated Reference configuration.

Sensors under test: 'SA' denotes single-Alter shielded test configurations and 'UN' denotes unshielded test configurations. 'Tret' appears in some plot titles, and refers to Tretyakov-shielded test gauges (not applicable to Sutron TPG gauges under test).

Figure 2: Map of location of SPICE site where Sutron TPG gauges were tested.

A summary of the configuration of instruments as tested, the duration of tests and availability of data reflected in these results, and the ancillary measurements used, by site, is available in Tables 1, 2, and 3, respectively.

Table 1: Summary of gauge configuration and data output, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	Marshall
Field configuration	Unshielded (UN); single-Alter shield (SA)
Height of installation (gauge rim)	2 m
Heating	SPICE algorithm
Antifreeze	Mixture of 60% methanol and 40% propylene glycol
Oil	Automatic Transmission Fluid (ATF)
SUT data output frequency	6 sec
Data QC	SPICE QC methodology
Data temporal resolution	1 min
Processing interval for SPICE data analysis	30 min

Table 2: Data availability, by measurement season and site.

Measurement season	Marshall
Season 1 (Oct. 2013 – Apr. 2014)	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓

Table 3: Summary of reference and ancillary measurements, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	Marshall
R2 site reference	Geonor T-200B3 600 mm (DFAR)
R2 precip detector	Thies LPM (Site*)
Ancillary temp sensor	MetOne, model 060A-2/062, 2144-L
Ancillary RH sensor	Campbell Scientific CS500 (2m)
Ancillary wind sensor	RM Young Wind Monitor 05103 (2 m)

*A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR shield; however, in cases where a more sensitive detector is available outside of the DFIR shield, or there are issues with the detector within the DFIR shield, a precipitation detector elsewhere on the site can be employed.

4) Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6.1 of the WMO-SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

4.1.1.1. Characterization of performance in non-precipitating conditions

The assessment data are derived over 30 minute intervals during which the precipitation detector in the R2 reference configuration reports 0 minutes of precipitation. The accumulation over these intervals (accumulation in minute 30 – accumulation in minute 1), representing the variability of the gauge response due to wind, evaporation, temperature, etc., is recorded, along with the mean wind speed, and the change in temperature (temperature in minute 30 – temperature in minute 1).

4.1.1.2. Assessment of ability to detect and report accumulation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute assessment intervals:

- Ref 'Yes' : R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No' : R2 weighing gauge < 0.25 mm AND/OR precip detector recording < 18 min of precip;
- SUT 'Yes' : SUT accumulation ≥ 0.25 mm;
- SUT 'No' : SUT accumulation < 0.25 mm.

4.1.1.3. Assessment of ability to detect light precipitation

The data for this component of the assessment are derived in a similar manner as those in Section 4.1.1.2, but with different combinations of thresholds for the reference and/or SUT 'Yes' and 'No' conditions. These different threshold 'cases' have been selected to demonstrate the impact of the thresholds used in data derivation on the detection of light precipitation.

4.1.2. Skill score assessment

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores, as detailed in Section 3.6.1.3 of the WMO-SPICE Final Report.

For the assessments considered in this report, the ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

5) Environmental conditions

The environmental conditions at each site over the duration of the test period are expressed as probability density functions (PDFs) of mean air temperature, mean relative humidity, mean wind speed, vector mean wind direction, and precipitation rate for each component 30 minute assessment interval in Figure 3. The same parameters are also shown for all assessment intervals during which the site reference configuration detected precipitation (i.e. all Ref 'Yes' cases).

The precipitation percentage represents the number of minutes of precipitation during a 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration, expressed as a percentage. PDFs of precipitation percentage are also included in Figure 3.

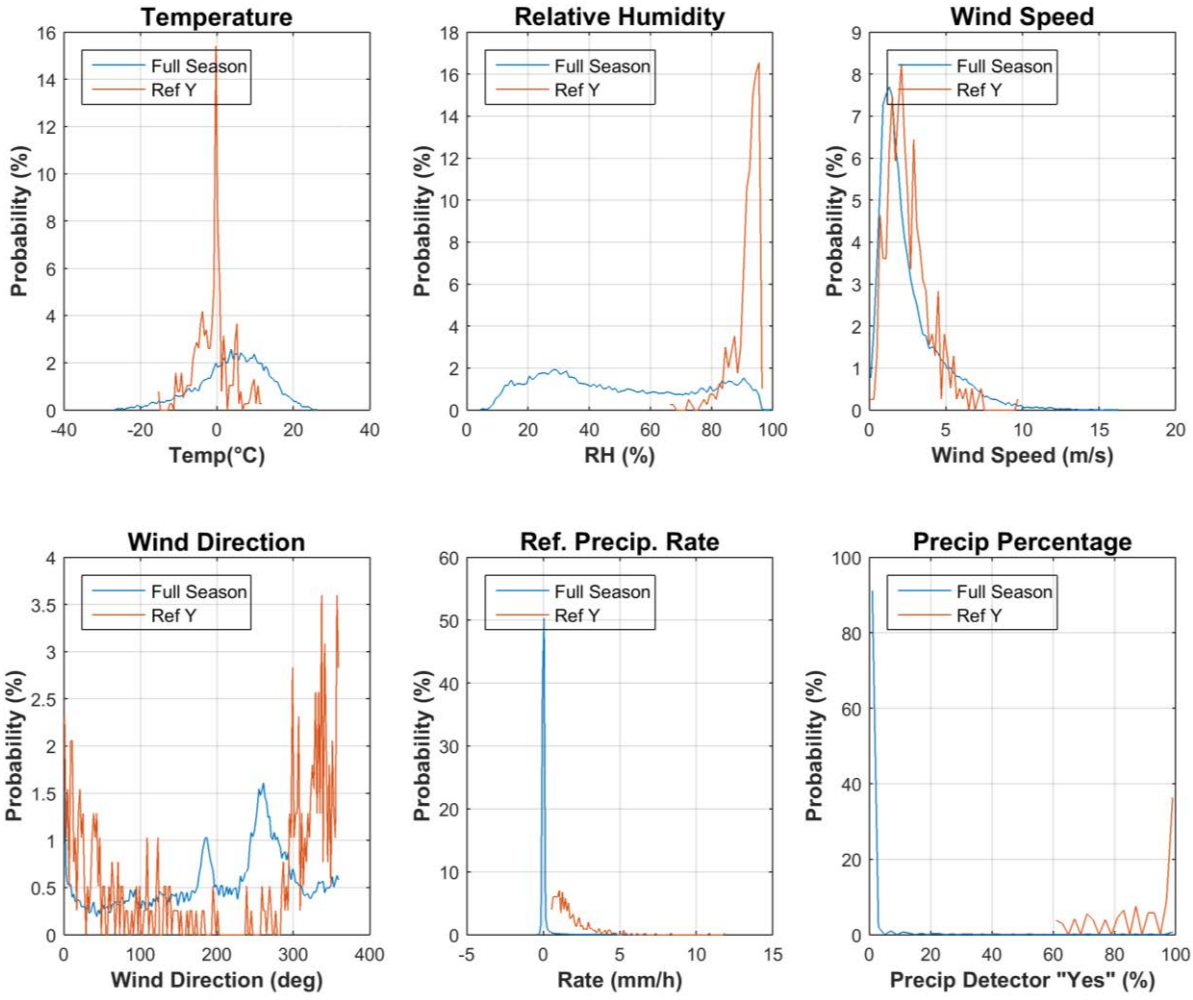


Figure 3: Summary of environmental conditions at Marshall over the entire duration of formal tests (Full Season; blue PDFs) and during precipitation events reported by the site R2 reference (Ref Y; orange PDFs).

6) Evaluation of the ability to perform over the range of operating conditions

6.1. Characterization of SUT performance in non-precipitating conditions

The response of the SUT in the absence of precipitation was examined as defined in Section 4.1.1.1. The results are presented below, reflecting the distribution of the sensor response and its variability with wind and temperature, as measured during 30 minute assessment intervals.

6.1.1. Overall variability of SUT response

The overall variability of the SUT response in non-precipitating conditions is shown as a probability density function for each test configuration in Figure 4. The corresponding PDF for the reference configuration is provided for comparison.

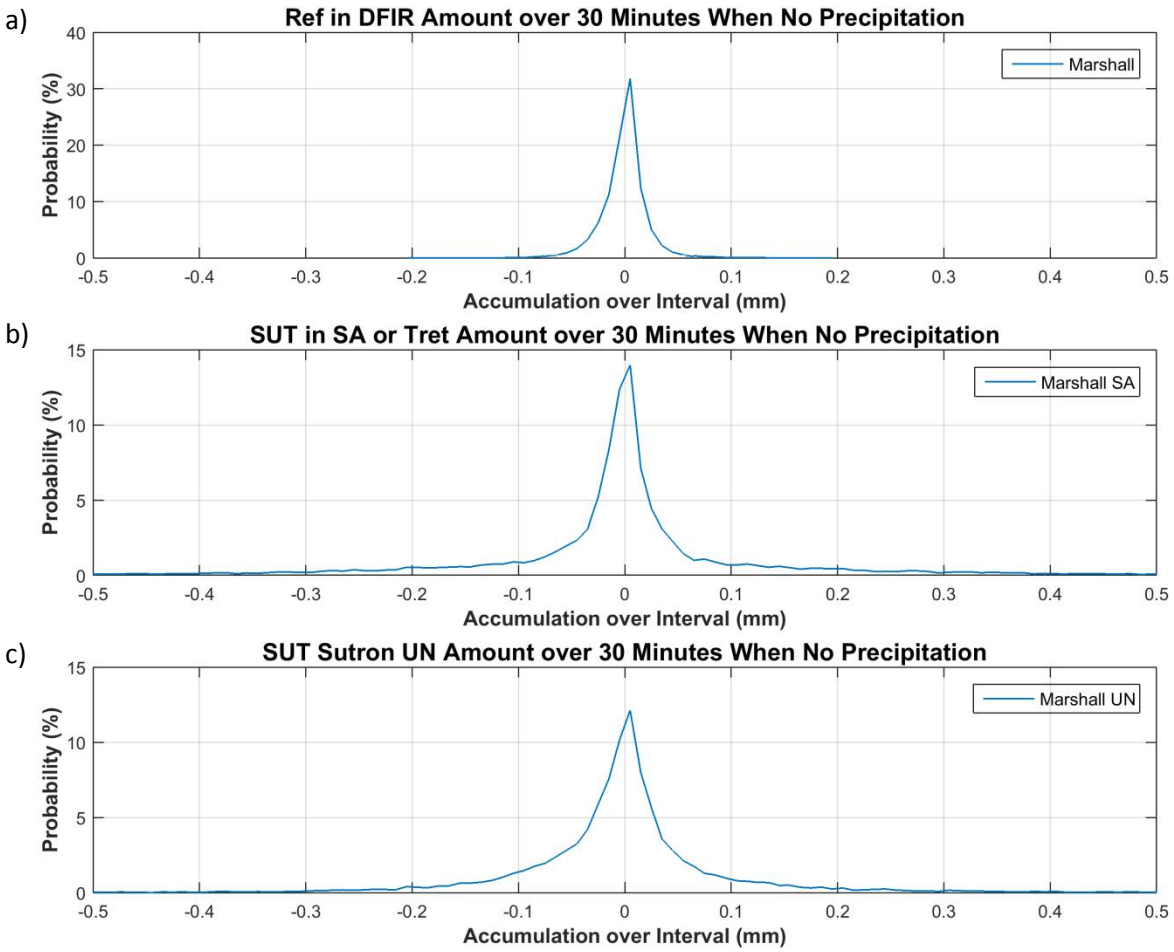


Figure 4: Probability density functions of the output signal (accumulation over 30 minute assessment intervals) in non-precipitating conditions for (a) the R2 reference configuration (DFAR) and Sutron TPG gauges in (b) single-Alter shielded and (c) unshielded configurations at Marshall.

The statistics of the output signal (accumulation over 30 minute assessment intervals) for the reference and SUT are provided in Table 4.

Table 4: Statistics of the R2 reference gauge and SUT output signal during non-precipitating conditions, as plotted in Figure 4.

Gauge	Average output signal (mm)	Standard deviation (mm)	Maximum output signal (mm)	Minimum output signal (mm)	Number of assessment intervals
Reference	-0.001	0.022	0.196	-0.207	18248
SUT (SA)	0.000	0.185	3.887	-1.882	18248
SUT (UN)	-0.001	0.135	4.043	-2.246	18256

6.1.2. Variability of SUT response as a function of temperature

The variability of the SUT response for each test configuration in the absence of precipitation is plotted as function of the temperature difference over each assessment interval in Figure 5. The temperature difference is defined as the difference in temperature between the end (minute 30) and beginning (minute 1) of the assessment interval. The corresponding plot for the reference configuration is provided for comparison.

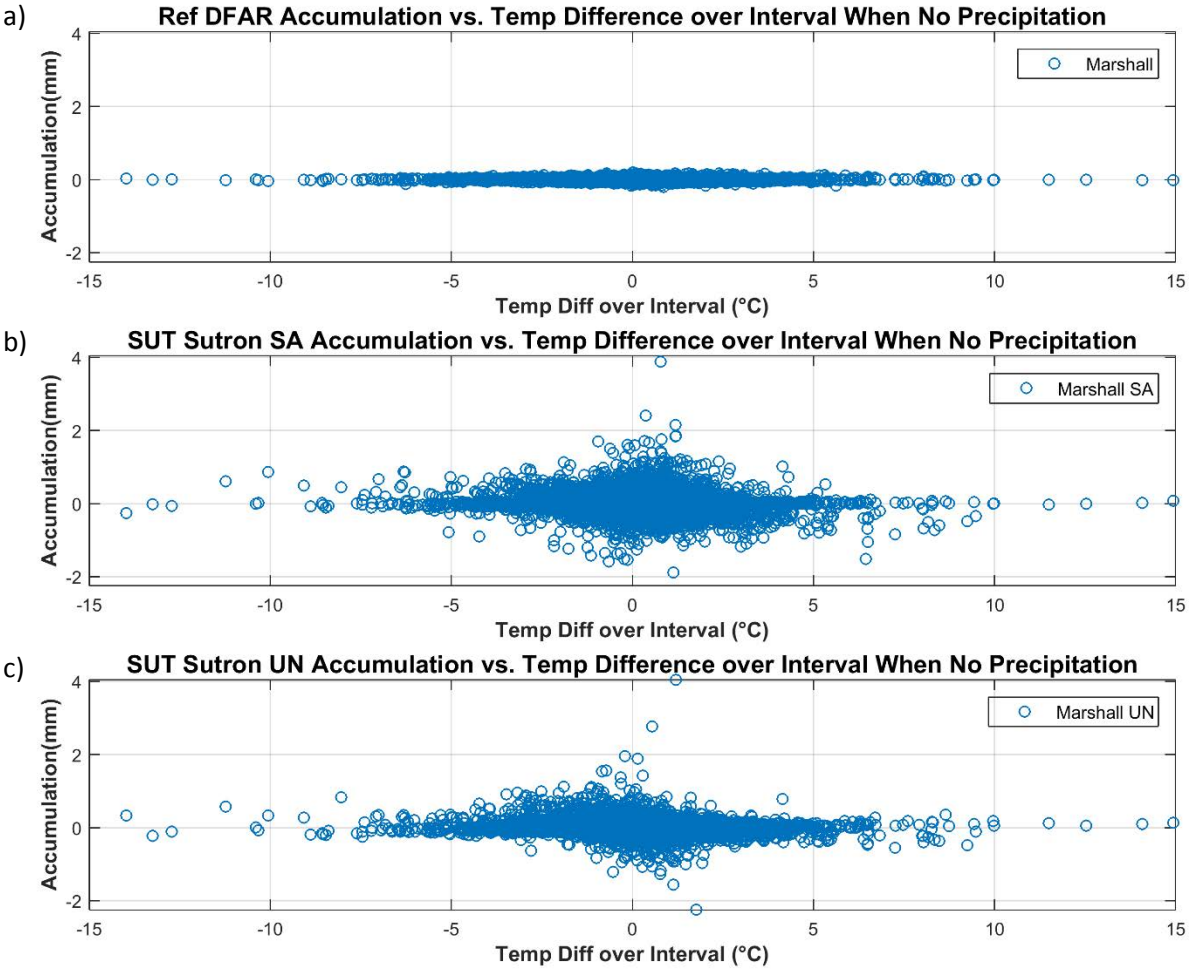


Figure 5: Variability of output signal (accumulation over each 30 minute assessment interval) as a function of the temperature difference over the interval in non-precipitating conditions for (a) the R2 reference configuration (DFAR) and Sutron TPG gauges in (b) single-Alter shielded and (c) unshielded configurations at Marshall.

6.1.3. Variability of SUT response as a function of wind speed

The variability of the SUT response for each test configuration in the absence of precipitation is plotted as function of the mean wind speed for each assessment interval in Figure 6. Here, the signal variability is represented as the standard deviation (STD) of the gauge accumulation output over each 30 minute interval. The corresponding plot for the reference configuration is provided for comparison.

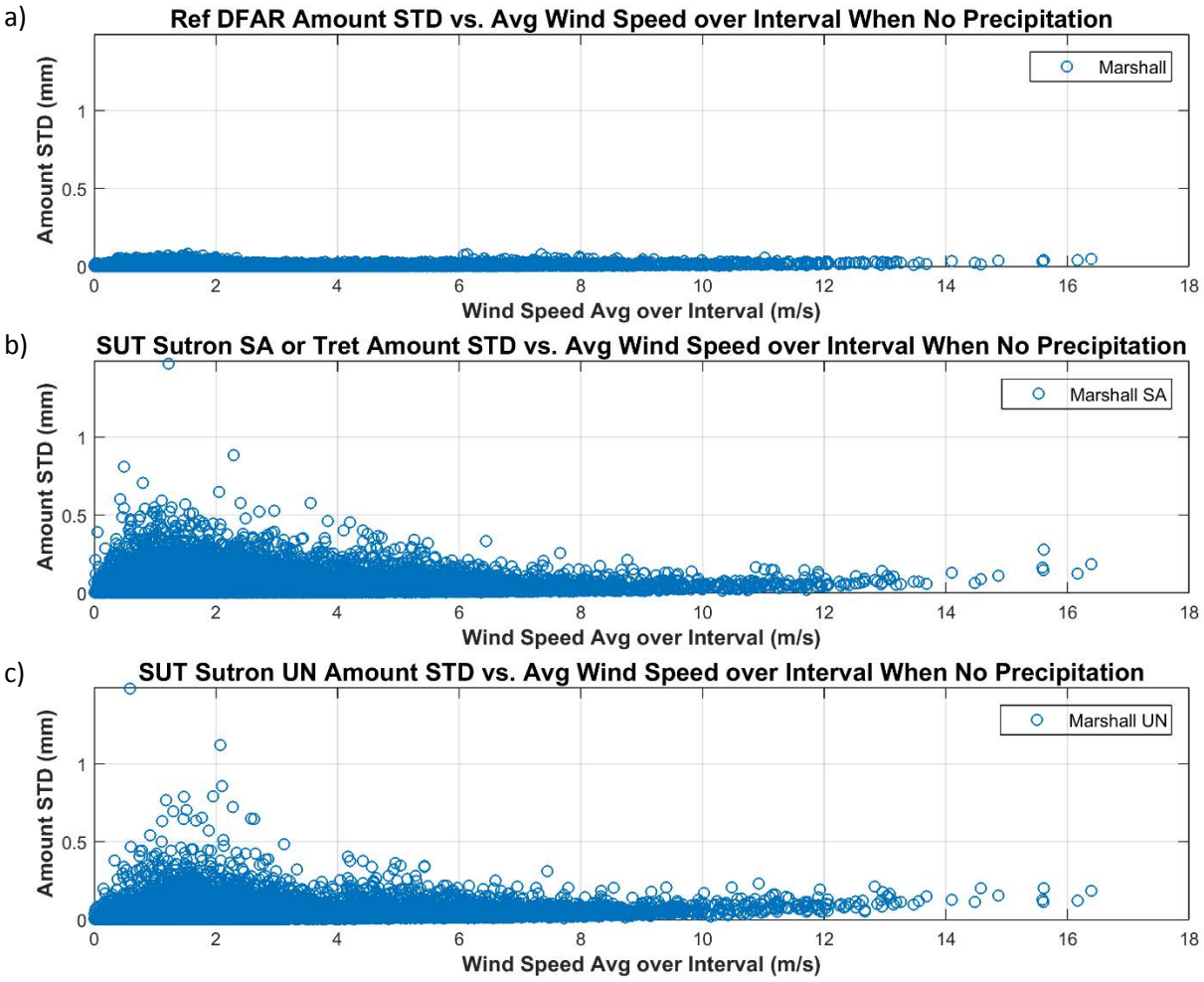


Figure 6: Variability of output signal (accumulation over each 30 minute assessment interval) as a function of mean wind speed in non-precipitating conditions for (a) the R2 reference configuration (DFAR) and Sutron TPG gauges in (b) single-Alter shielded and (c) unshielded configurations at Marshall.

6.2. Ability to detect and report precipitation

6.2.1. Skill score assessment

The overall ability of the SUT to detect and report the occurrence of precipitation relative to the site field reference over 30 minute assessment intervals is expressed using selected skill scores (Section 4.1.2) and presented in Table 5. Scores are presented for 30 minute assessment intervals. The contingency results (Section 4.1.1) corresponding to these scores are presented in Table 6.

Table 5: Skill scores for Sutron TPG gauges under test at Marshall during the formal test periods.

SUT configuration	Probability of Detection, POD (%)	False Alarm Rate, FAR (%)	Bias, B (%)	Heidke Skill Score, HSS (%)
Single-Alter	91.2	75.8	377	36.2
Unshielded	79.7	64.5	225	47.6

Table 6: Contingency table illustrating detection of precipitation by Sutron TPG gauges under test relative to the site reference at Marshall, expressed as number of events over the entire test period.

SUT configuration	Number of Events			
	YY (hits)	YN (misses)	NY (false alarms)	NN (correct negatives)
Single-Alter	364	35	1141	18498
Unshielded	318	81	579	19068

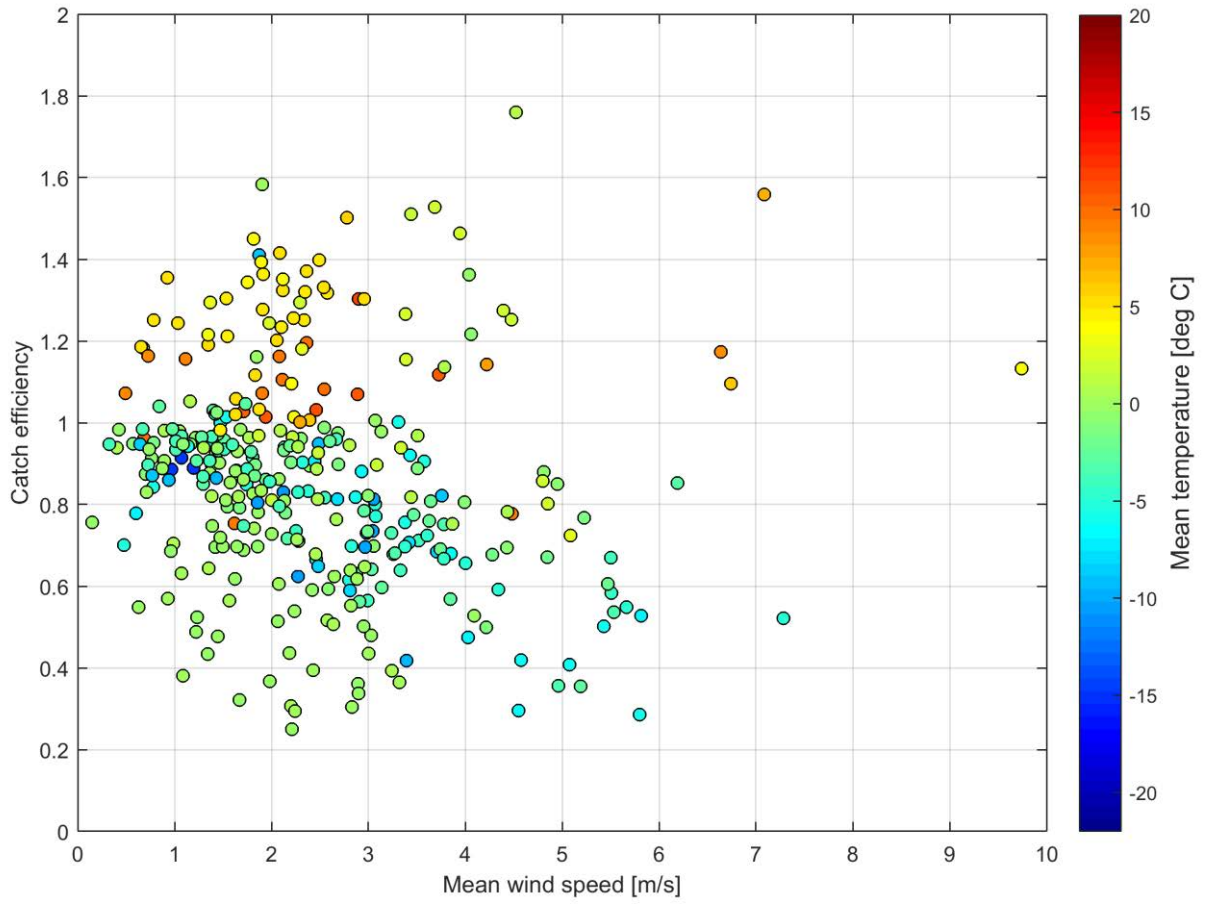
6.3. Ability to report accumulated precipitation

The SUT performance in terms of reporting accumulated precipitation is examined by comparing the amount reported by the sensor under test relative to the respective site reference during 30 minute assessment intervals. This is represented graphically using scatter and box and whisker plots of the catch efficiency as a function of mean wind speed at gauge height, as well as scatter plots of the amounts reported by the SUT versus the corresponding reference amounts (Figures 7 and 8). The SUT performance is also assessed in terms of the root mean square error, RMSE (Table 7).

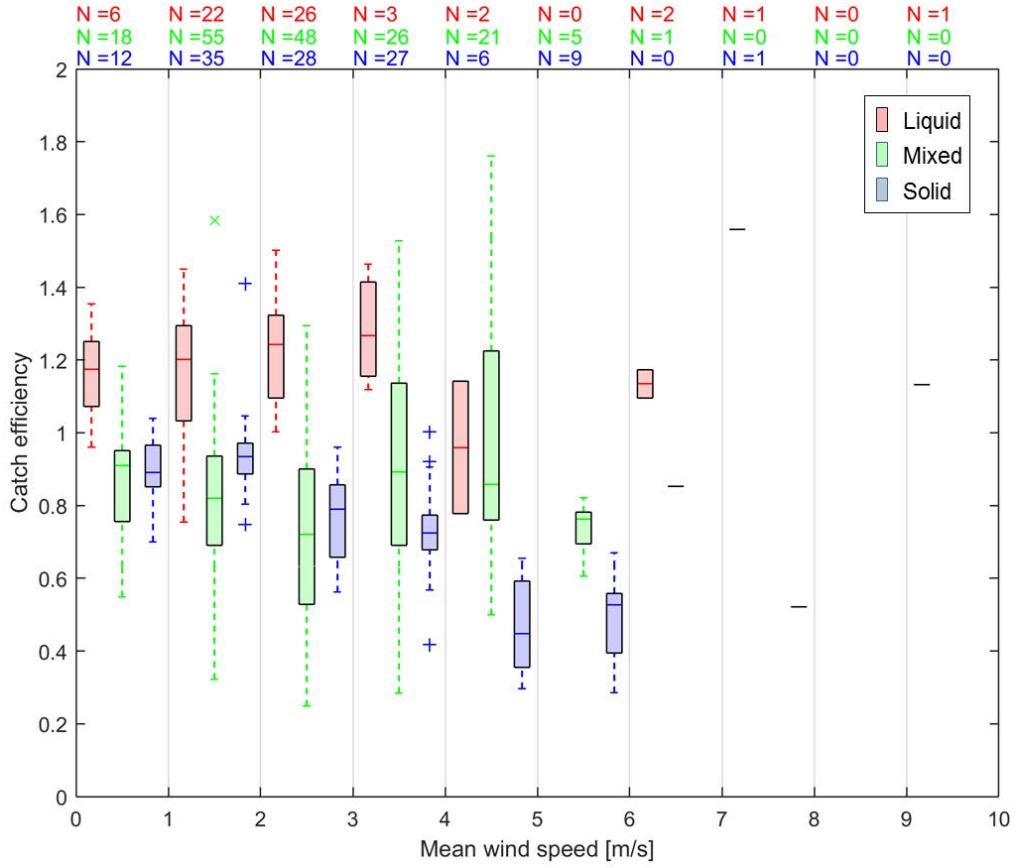
Only assessment intervals during which the SUT and reference both reported precipitation (YY cases) are considered in this portion of the assessment. In the catch efficiency-wind speed scatter plots, the mean event temperature is indicated by colour, with the colour scale selected to be consistent across all sites with weighing gauges under test. In the box and whisker plots and accumulation-accumulation scatter plots, the predominant precipitation type is indicated by colour, as determined from the reported temperature (Section 4.1.4).

Figure 7: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the Sutron TPG gauge in single-Alter shield at Marshall.

a)



b)


N =6 N =22 N =26 N =3 N =2 N =0 N =2 N =1 N =0 N =1
N =18 N =55 N =48 N =26 N =21 N =5 N =1 N =0 N =0 N =0
N =12 N =35 N =28 N =27 N =6 N =9 N =0 N =1 N =0 N =0
Liquid
Mixed
Solid
Catch efficiency
Mean wind speed [m/s]

c)

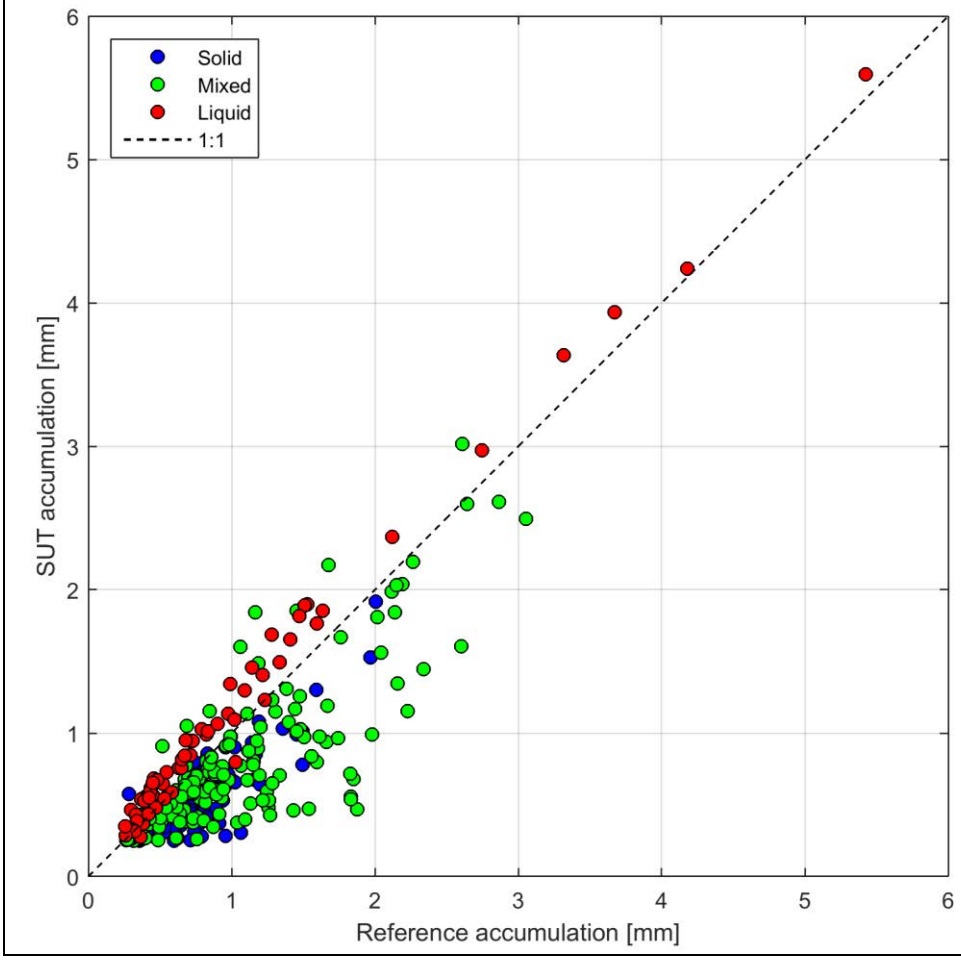
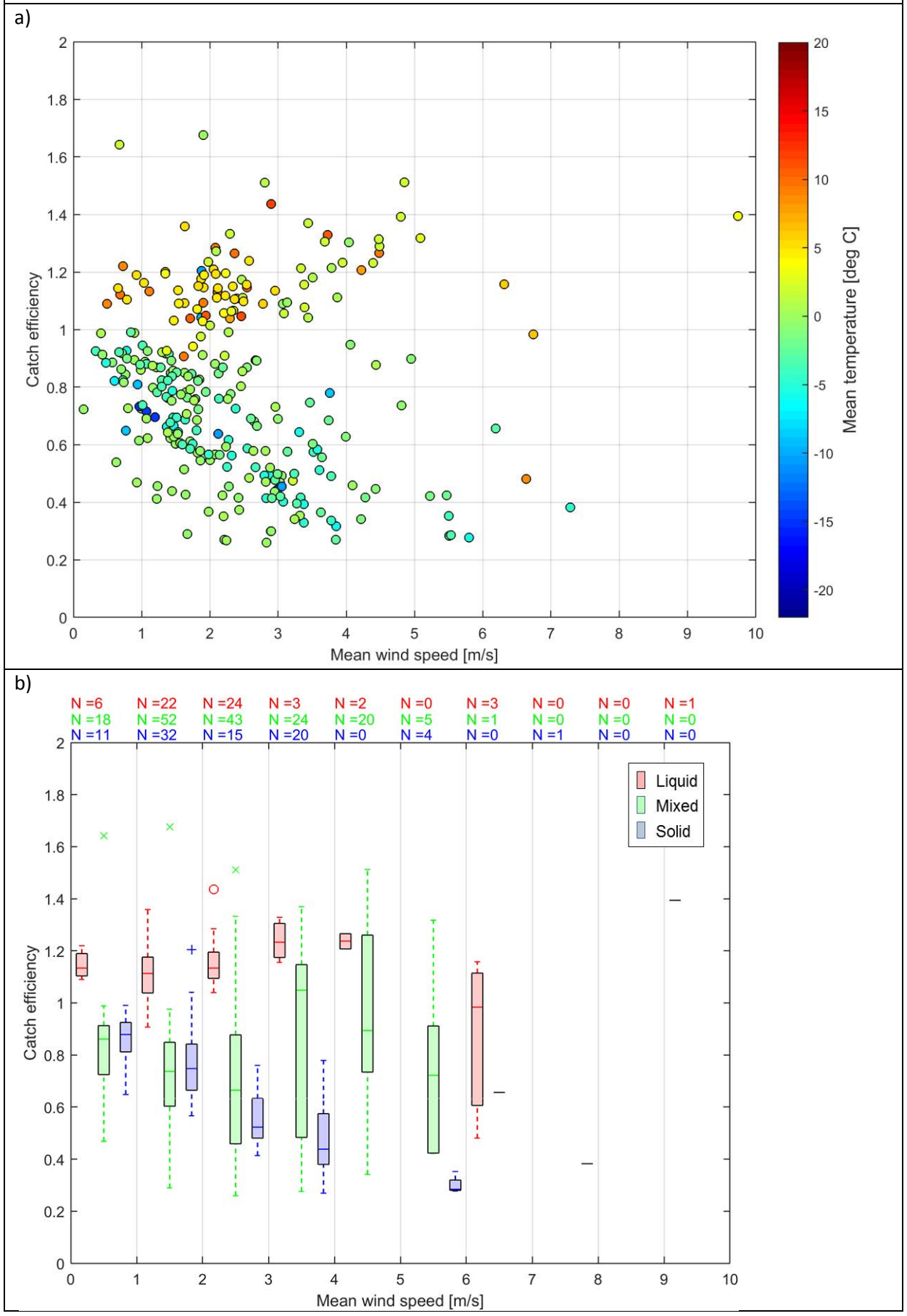

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]

Figure 8: (a) Catch ratio scatter plots, (b) catch ratio box and whisker plots, and (c) accumulation-accumulation scatter plots for the unshielded Sutron TPG gauge at Marshall.



c)

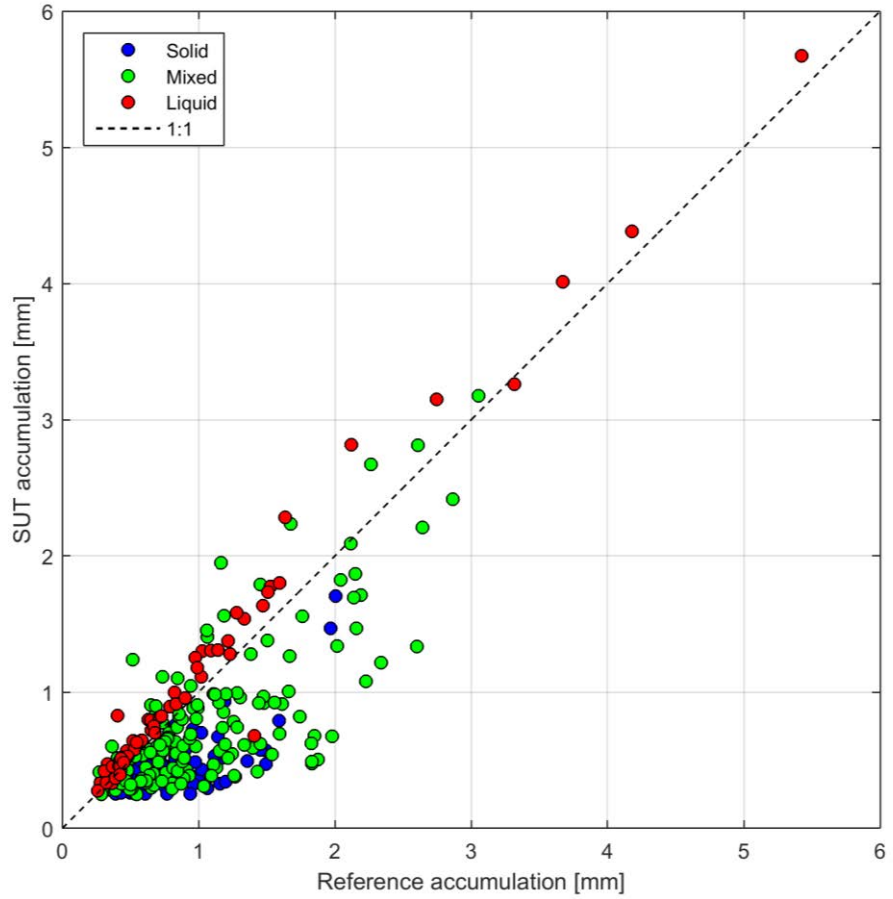

Solid
Mixed
Liquid
1:1
SUT accumulation [mm]
Reference accumulation [mm]
0
1
2
3
4
5
6

Table 7: RMSE values for test configuration(s) by precipitation type for YY cases over the entire test period at Marshall.

SUT configuration	RMSE (mm/30 min)			
	Liquid	Mixed	Solid	All precip types
Single-Alter	0.19	0.75	0.24	0.56
Unshielded	0.22	0.76	0.40	0.60

The overall catch ratio calculated using all 30 minute YY cases, over the entire test period, is provided in Table 8. To demonstrate the influence of the SUT accumulation threshold on the results, the overall catch ratio is also provided for all 30 minute YY cases determined using a lower SUT threshold of 0.1 mm/30 minutes. Note that these values reflect only the YY cases, and do not include the amounts corresponding to the cases when the SUT and the reference do not agree on the occurrence of precipitation.

Table 8: Overall catch ratio for test configuration(s) determined from YY cases over the entire test period at Marshall, using two different SUT accumulation thresholds.

SUT configuration	SUT accumulation threshold (mm/30 min)	Overall catch ratio
Single-Alter	0.25	0.87
	0.1	0.86
Unshielded	0.25	0.82
	0.1	0.78

6.4. Ability to detect light precipitation events

The impact of the threshold selection for data processing relative to the detection of light precipitation was examined using four different combinations of reference and SUT accumulation thresholds (four 'cases' in Table 9) for the Sutron TPG gauge in single-Alter shield. Contingency results, probabilities of detection (POD), and false alarm rates (FAR) are presented for each case in Table 10. A quantitative comparison of the amounts reported in each case is beyond the scope of this assessment.

Table 9: Reference and SUT thresholds in each case for light precipitation detection assessment.

Case	Reference threshold (mm/30 min)	SUT threshold (mm/30 min)
1	0.25	0.25
2	0.1	0.1
3	0.25	No threshold
4	0.25	0

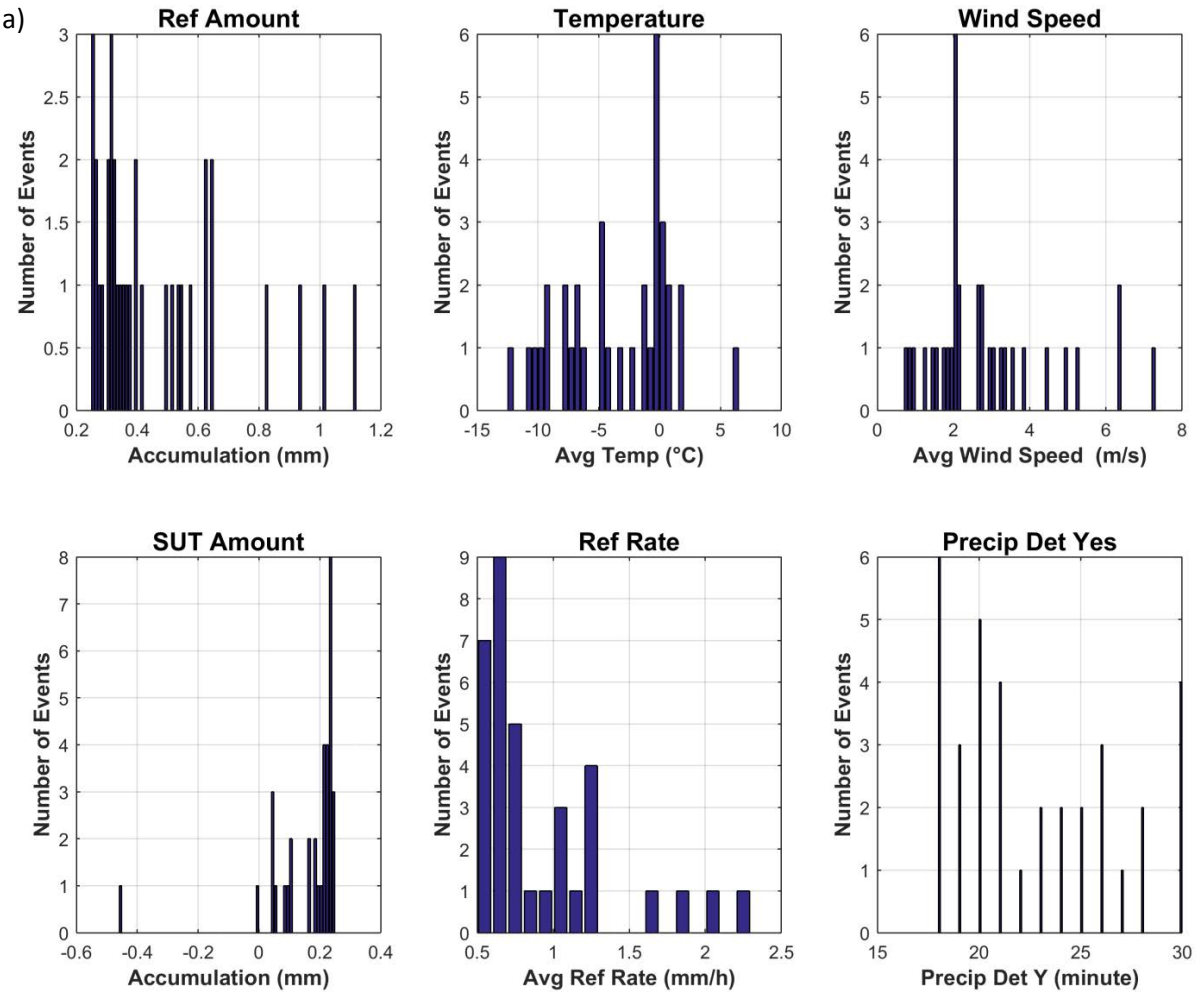
Table 10: Contingency results, probability of detection, and false alarm rate for each case in light precipitation detection assessment.

Case	Number of events				Skill score (%)	
	YY	YN	NY	NN	POD	FAR
1	364	35	1141	18498	91.2	75.8
2	428	11	2724	16875	97.5	86.4
3	399	0	19639	0	100	98
4	397	2	9996	9643	99.5	96.2

6.5. Assessment of events when the reference and the SUT do not agree on the occurrence of precipitation

Assessment intervals during which the site reference and SUT do not agree on the occurrence of precipitation – namely, the YN and NY cases (Section 4.1.1) – are characterized separately for each test gauge using histograms in Figures 9 and 10. The histograms include accumulated precipitation reported by the reference and SUT (0.25 mm/30 min threshold for both), precipitation intensity as reported by the reference, and corresponding site conditions.

a)



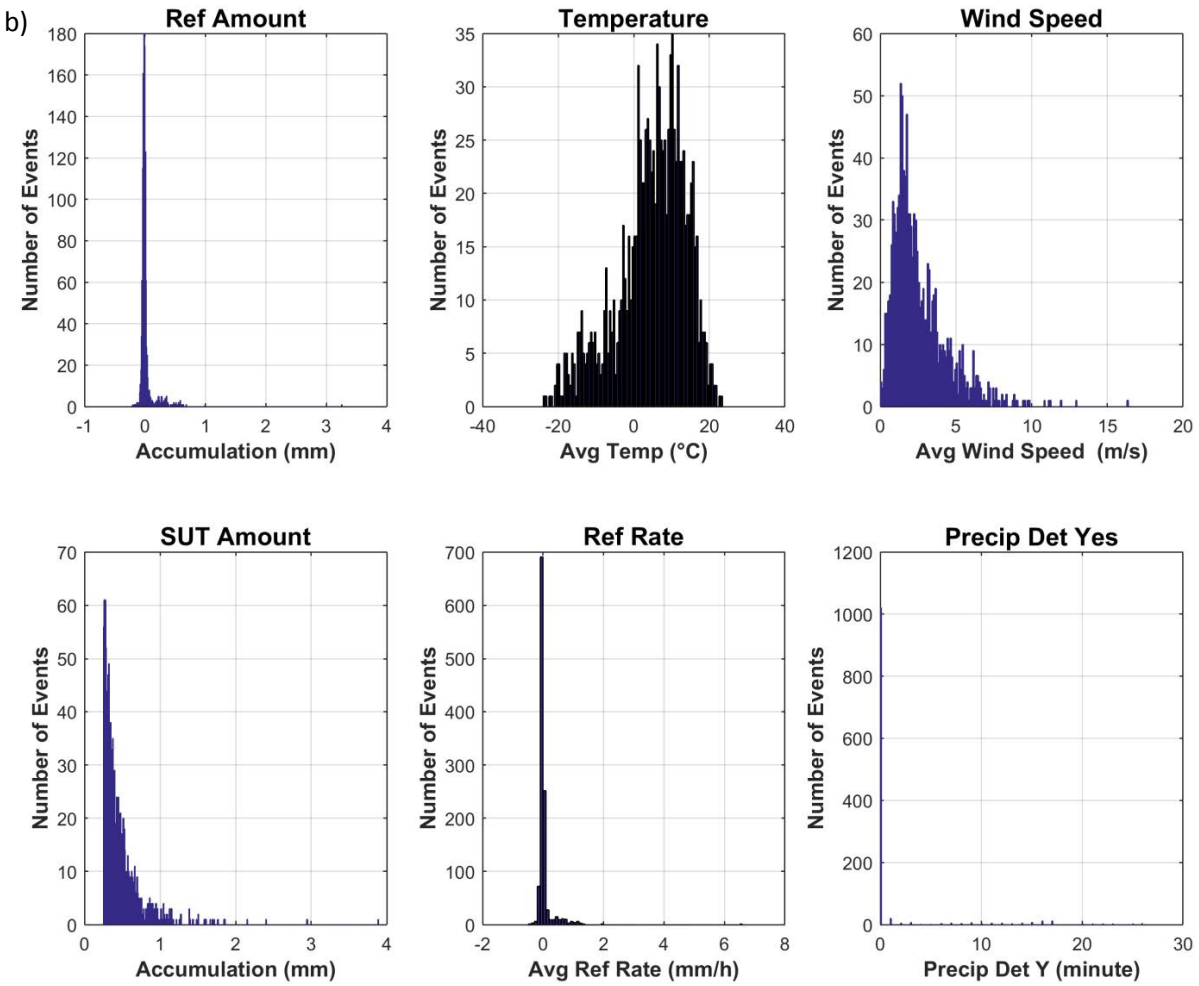
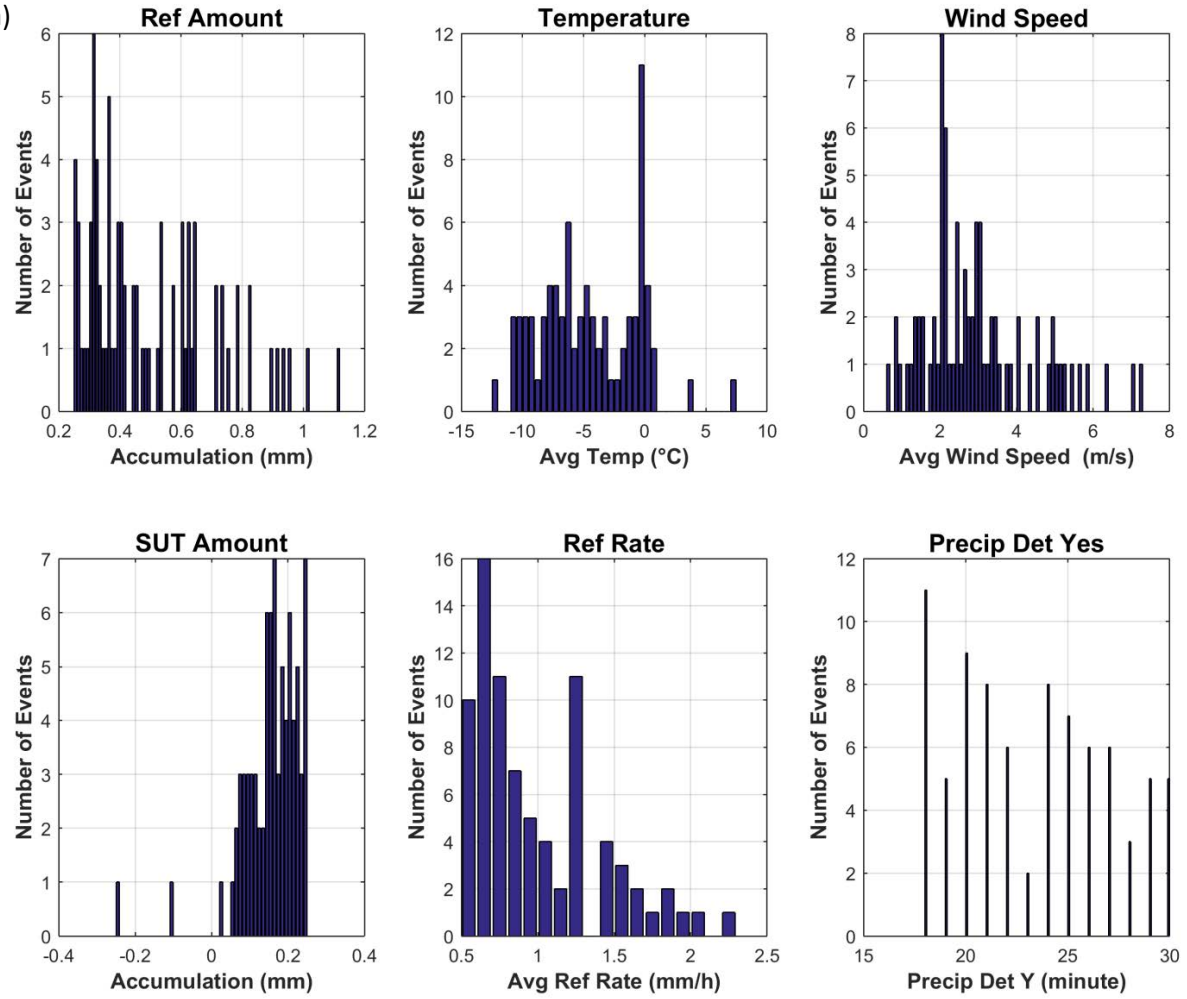


Figure 9: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for (a)YN events (35 total) and (b) NY events (1141 total) for the single-Alter shielded Sutron TPG over the test period.

a)


Ref Amount
Number of Events
Accumulation (mm)
Temperature
Number of Events
Avg Temp (°C)
Wind Speed
Number of Events
Avg Wind Speed (m/s)
SUT Amount
Number of Events
Accumulation (mm)
Ref Rate
Number of Events
Avg Ref Rate (mm/h)
Precip Det Yes
Number of Events
Precip Det Y (minute)

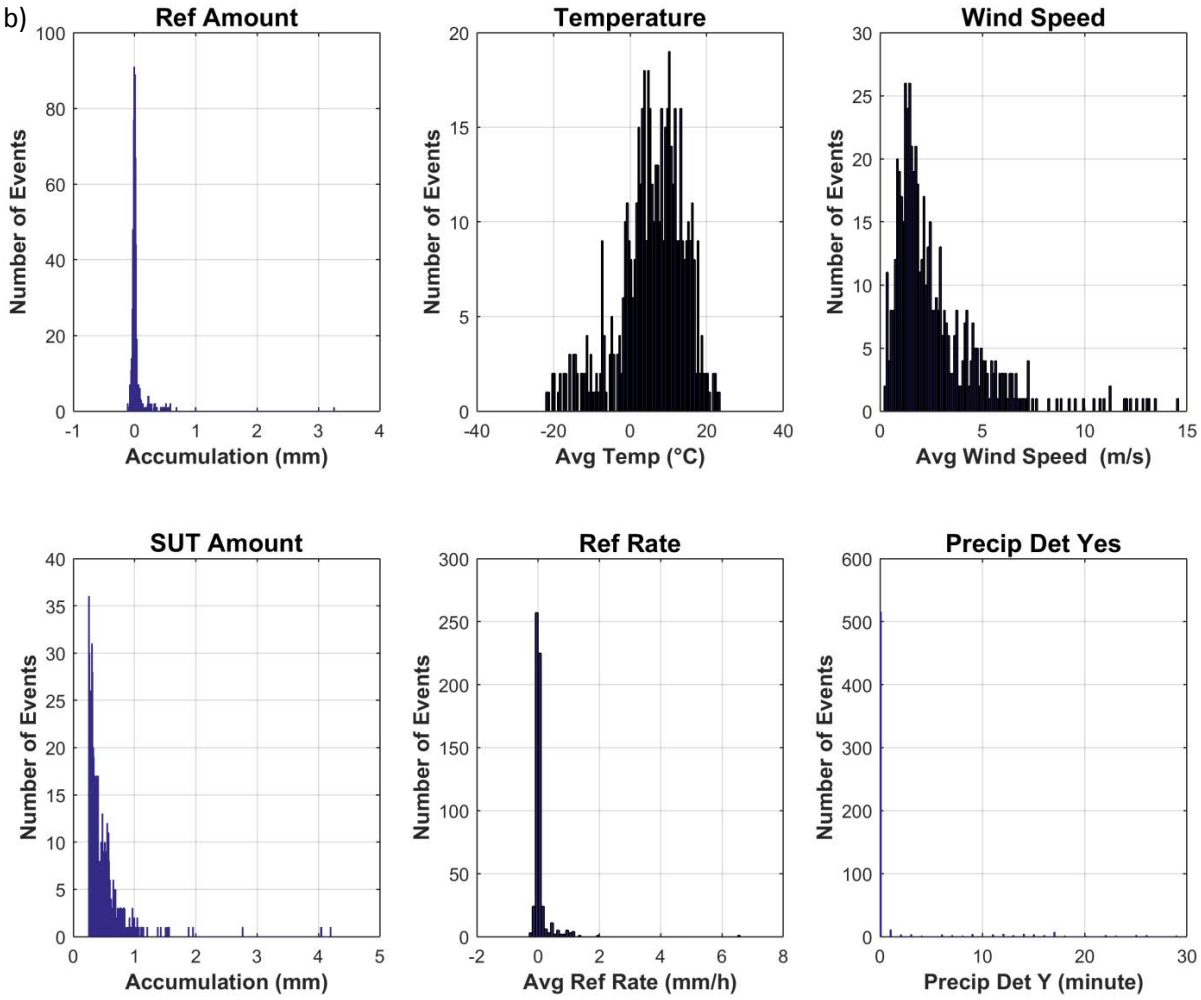


Figure 10: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, reference precipitation rate, and number of minutes with 'Yes' responses from the precipitation detector in the R2 reference configuration for (a)YN events (81 total) and (b) NY events (579 total) for the unshielded Sutron TPG over the test period.

7) Interpretation of Results

7.1. Operating conditions

The full range of conditions under which the test gauges at Marshall were operated is illustrated in Figure 3. The gauges were operated at temperatures between approximately -27 °C and 26 °C, which fall within the manufacturer's specified operating range of -40 °C to 60 °C. The conditions during precipitation events are also shown in Figure 3, and indicate that most events occurred at mean temperatures below freezing and mean wind speeds below 8 m/s.

7.2. Performance over the range of operating conditions

7.2.1. Non-precipitating conditions

Accumulation reports over 30 minute periods without precipitation (as indicated by an independent precipitation detector) are characterized for the reference and Sutron TPG gauges at the Marshall site in Figures 4 to 6 and Table 4. The PDFs for both test configurations are broader than that of the reference configuration (Figure 4), indicating greater variability in gauge reports from the Sutron TPG in the absence of precipitation relative to the reference. This enhanced variability, or noise, is also apparent in the standard deviation results in Table 4; the standard deviations of accumulation reports during non-precipitating periods are approximately 6 to 8 times greater for the unshielded and shielded Sutron TPG gauges, respectively, than for the reference configuration at Marshall. The variability in test gauge responses appears to be greatest for events characterized by small changes in temperature (Figure 5) and mean wind speeds below about 6 m/s (Figure 6).

The magnitude of gauge responses in the absence of precipitation can be used to identify a detection threshold that minimizes the detection of false precipitation while enhancing the detection of light precipitation. This threshold is considered to be three times the standard deviation of the average gauge response during 30 minute non-precipitating periods. Based on the present results for test gauges at Marshall (Table 4), this minimum detection threshold is determined to be 0.56 mm for single-Alter shielded gauges and 0.40 mm for unshielded gauges.

7.2.2. Precipitating conditions

7.2.2.1. Ability to detect and report precipitation

The ability of the Sutron TPG gauges under test to detect and report precipitation relative to the reference configuration was assessed over 30 minute periods. The skill score results in Table 5 indicate higher Probability of Detection values for the shielded Sutron TPG relative to the unshielded gauge (POD of 91% for SA, 80% for UN), but also a higher false alarm rate (FAR of 76% for SA, 64% for UN), resulting in a higher Bias (B of 377% for SA, 225% for UN) and lower Heidke Skill Score (HSS of 36% for SA, 48% for UN). The high False Alarm Rates for both test configurations result from the large numbers of NY events in Table 6, which exceed the combined number of YY and YN events for each test configuration. This is attributed to the noise in the test gauge outputs, which can cause accumulation reports to exceed the detection threshold in very light or non-precipitating conditions. The high Bias values indicate that both test configurations detect over twice the number of events

detected by the reference; while these results are dependent upon the specific thresholds selected, they illustrate that gauge noise can result in false reports and over-reporting of precipitation under the specific conditions tested.

7.2.2.2. Ability to report accumulated precipitation

The results presented in Figures 7 to 8, which are based on 30 minute events during which the reference and test gauge both detect precipitation (YY cases), illustrate the influence of wind speed and precipitation type on gauge catch efficiency. The discussion below will focus on snow events; the number of rain events during winter is limited, and the results for mixed events are variable due to the variability in the size and density of precipitation within the mixed regime, as well as the potential for transitions between phases. For snow events, the median catch efficiency decreases more rapidly with increasing wind speed for the unshielded test configuration (Figure 8b) relative to the shielded test configuration (Figure 7b). For the shielded configuration, the median catch efficiency remains above 0.7 for mean wind speeds between 3 and 4 m/s, and falls to within 0.4 to 0.6 for mean wind speeds between 4 and 6 m/s. For the unshielded configuration, the median catch efficiency for solid precipitation is between 0.4 and 0.5 for mean wind speeds between 3 and 4 m/s, and falls to approximately 0.3 for mean wind speeds between 5 and 6 m/s.

Root mean square error values were computed from all 30 minute events during which each test configuration and the reference configuration both detected precipitation. The computed values are shown in Table 7, and can be considered to represent the absolute uncertainty of each test configuration relative to the reference configuration in liquid, mixed, and solid precipitation, and in all precipitation types. The RMSE values for both configurations in all precipitation types are similar, with the shielded gauge having a slightly lower value of 0.56 mm/30 min relative to the 0.60 mm/30 min value for the unshielded gauge. The values in solid precipitation show greater discrepancy, with 0.24 mm/30 min and 0.40 mm/30 min values reported for the shielded and unshielded gauges, respectively. These differences are attributed to the single-Alter shield mitigating the effects of wind-induced undercatch, which leads to larger RMSE values.

The overall catch ratio – computed from the total reference and SUT accumulation from all YY cases over the duration of formal tests – is provided for each test configuration in Table 8. The overall catch ratio for the single-Alter shielded Sutron TPG is 0.87, compared to 0.82 for the unshielded gauge. As both gauges were installed at the same site, and subject to the same environmental conditions, these differences are attributed to the enhanced catch efficiency of the shielded gauge in solid precipitation conditions; this is corroborated by the lower RMSE for the shielded gauge in solid precipitation in Table 7. Decreasing the SUT accumulation threshold for precipitation events from 0.25 mm/30 min to 0.1 mm/min does not impact the overall catch ratio significantly. Given the magnitude of noise observed for both test configurations in non-precipitating conditions (see Section 6.1 and the discussion in Section 7.2.1), the 0.1 mm threshold would be within the expected noise level of the test gauges.

7.2.2.3. Ability to detect light precipitation

The detection thresholds and results presented in Tables 9 and 10, respectively, indicate that decreasing the SUT detection threshold to zero or removing it entirely (while maintaining the

reference detection threshold at 0.25 mm) increases both the Probability of Detection and False Alarm Rate to close to 100%. These values correspond to an instrument with low detection skill relative to the reference configuration (Heidke Skill Score close to zero). Decreasing both the reference and SUT detection thresholds to 0.1 mm increases the POD from 91% to 98%, while also increasing the FAR from 76% to almost 86%. These results suggest that gauge noise mitigates the ability of Sutron TPG gauges to detect light precipitation, as any reduction in the detection threshold intended to capture light precipitation is accompanied by a corresponding increase in false reports, as indicated by the false alarm rate.

7.2.3.4. Assessment of events when the reference and SUT do not agree on the occurrence of precipitation

The YN ('miss') cases, when the reference detects a precipitation event and the SUT does not, and NY ('false alarm') cases, when the SUT detects a precipitation event and the reference does not, are characterized for the shielded and unshielded test configurations in Figures 9 and 10, respectively.
The majority of the YN cases have reference accumulations just above the 0.25 mm threshold, and SUT accumulations just below the threshold. This difference is likely the result of enhanced wind effects for the test gauges relative to the reference gauge in the DFIR-fence, which can reduce the SUT accumulation below the detection threshold. Following this logic, the number of 'miss' cases should be greater for the unshielded configuration relative to the shielded configuration; indeed, there are 35 'miss' cases for the shielded gauge (Figure 9a) compared to 81 for the unshielded gauge (Figure 10a).

The number of 'false alarm' events is comparatively higher for each of the test gauges considered – 1141 for the shielded gauge (Figure 9b) and 579 for the unshielded gauge (Figure 10b). The majority of these events are characterized by reference accumulations of 0 mm, zero minutes of precipitation reported by the precipitation detector, and test gauge reports just above the detection threshold of 0.25 mm. Accordingly, the false alarm events are attributed primarily to the noise in test gauge reports observed in the absence of precipitation (Section 6.1 and discussion above in Section 7.2.1). The larger number of false alarm events for the shielded gauge relative to the unshielded gauge results from the higher magnitude of noise variability observed for the shielded gauge, as indicated by the standard deviations of accumulation reports in non-precipitating conditions in Table 4. Further investigation is required to determine why the magnitude of noise variability differed for the test configurations. For the shielded configuration, the shield was mounted separately from the gauge (Figure 1b), and is not expected to impact gauge performance (e.g. by vibration).

8) Maintenance

Gauge calibration: each site completed the gauge field calibration and verification as per manufacturer recommendations, at least once a year or following the emptying of the gauge. The calibration records have been stored by each site host.

9) Performance Considerations

9.1. Data processing

Filtering of Sutron TPG gauge data are recommended to mitigate the influence of noise. In the present analysis, the SPICE data quality control procedure was applied to filter the test data; however, the noise observed in the processed data from both test gauges relative to the reference in the absence of precipitation (Section 6.1) suggests that additional filtering may be required. The specific threshold employed for distinguishing between precipitation and signal noise is an important consideration; as the threshold is decreased, there is higher likelihood of gauge noise producing false reports.

9.2. Gauge configuration

Sutron TPG gauges were tested in single-Alter and unshielded configurations at the Marshall site. The shielded gauge shows higher overall catch efficiency and lower RMSE relative to the unshielded gauge, but also higher false alarm rates, attributed to noise observed in the absence of precipitation. These results illustrate that the specific configuration of Sutron TPG gauges at a given site can impact the gauge performance. All field configurations must be fully tested and validated prior to use, in order to ensure that physical and signal interferences are not degrading the gauge signal. This would include the confirmation of grounding tailored to the soil conditions, sturdiness of foundation and mounting, and signal conditioning.

9.3. Ancillary measurements and adjustments

The application of adjustment functions is strongly recommended to account for the reduction in gauge catch efficiency as the wind speed increases. Ancillary measurements of wind speed (preferably at gauge orifice height) and air temperature are required for the application of adjustment functions. Ancillary measurements from a sensitive precipitation detector that is independent from the test gauge are also recommended to help distinguish precipitation events from false reports due to noise.

WMO-SPICE Instrument Performance Report
CAE PMB25R

1) Technical specifications (from manufacturer provided documentation)

Instrument model:	CAE PMB25R
Physical principle:	Heated tipping bucket (TB) gauge. Tips are recorded when the bucket fills to capacity, triggering a magnetic reed switch.
Bucket capacity:	0.2 mm
Collecting area:	1000 cm^2
Heating configuration:	Funnel (150 W), ring (110 W), and bucket (40 W) heaters; total maximum heating power of 300 W. Maintains temperature of ring and funnel at 2 °C.
Operating temperature range:	-30 °C to 60 °C
Measurement range:	0 mm/hr to 300 mm/hr intensity
Measurement uncertainty:	3%
Sensitivity:	0.1 mm reporting resolution (following internal processing)

a)

b)

Figure 1: CAE PMB25R installations at (a) CARE and (b) Marshall.

2) Data output format

Gauge data output: As tested, gauge reports accumulated precipitation in real-time and/or precipitation intensity following internal data processing (intensity correction) with a fixed output delay of 7 minutes.

Analysis is based on time-adjusted, minutely reports of precipitation intensity in mm/hr. Accumulated precipitation is computed as the cumulative sum of minutely accumulations derived from these intensity values.

3) SPICE test configuration

Shield: Unshielded

Test sites: CARE (Canada); Marshall (USA);

Sensor provider(s): CAE S.p.A.

Figure 2: Map of SPICE sites where CAE PMB25R gauges were tested.

A summary of the configuration of instruments as tested, the duration of tests and availability of data reflected in these results, and the ancillary measurements used, by site, is available in Tables 1, 2, and 3, respectively.

Table 1: Summary of gauge configurations and data output, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	CARE	Marshall
Field configuration	Unshielded	Unshielded
Height of installation (gauge rim)	1.5 m	1.9 m
SUT data output frequency	1 min	1 min
Data QC	SPICE QC methodology	
Data temporal resolution	1 min	
Processing interval for SPICE data analysis	30 min, 60 min	

Table 2: Data availability, by measurement season and site.

Measurement season	CARE	Marshall
Season 1 (Oct. 2013 – Apr. 2014)	✓	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓	✓

Table 3: Summary of reference and ancillary measurements, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	CARE	Marshall
R2 site reference	Geonor T-200B3 600 mm (DFAR)	Geonor T-200B3 600 mm (DFAR)
R2 precip detector	Thies LPM (DFAR)	Thies LPM (Site*)
Ancillary temp sensor	Vaisala HMP155 (Stevenson screen)	MetOne, model 060A-2/062, 2144-L
Ancillary RH sensor	Vaisala HMP155 (Stevenson screen)	Campbell Scientific CS500
Ancillary wind sensor	Vaisala NWS 425 (2 m)	RM Young Wind Monitor 05103 (2 m)

*A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR shield; however, in cases where a more sensitive detector is available outside of the DFIR shield, or there are issues with the detector within the DFIR shield, a precipitation detector elsewhere on the site can be employed.

4) Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6.1 of the WMO-SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute assessment intervals:

- Ref 'Yes' : R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No' : R2 weighing gauge < 0.25 mm AND/OR precip detector recording < 18 min of precip;
- SUT 'Yes' : SUT accumulation > 0 mm;
- SUT 'No' : SUT accumulation = 0 mm.

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores.

4.1.2. Skill score assessment

The ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

The above scores are computed using the formulations provided in Section 3.6.1.3 of the WMO-SPICE Final Report.

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

4.2. Sensor-specific considerations: response delays

Tipping bucket gauges require that an amount of precipitation corresponding to the bucket capacity is accumulated before a tip is triggered and the gauge records precipitation. This can result in response delays relative to the reference configuration. Heated tipping bucket gauges are subject to further delays, as any solid precipitation in the funnel must be melted before reaching the bucket and potentially triggering a tip, and the heating itself can potentially evaporate incident precipitation. These response delays will impact the comparison with the reference. For this reason, the assessment of TB gauge performance relative to the reference configuration is also considered over 60 minute intervals, using the same conditions and thresholds outlined above in Section 4.1.1.

Response delays are quantified by determining the time elapsed between the onset of precipitation as determined by the reference configuration, and the first tip recorded by the TB. The assessment is based on periods with at least 30 minutes of precipitation, as identified by the reference configuration, followed by at least 180 minutes without precipitation. This extended period without precipitation is intended to allow additional time for the melting and recording of precipitation by heated TB gauges. Additional details are provided in Section 3.6.1.4.4 of the WMO-SPICE Final Report.

5) Environmental conditions

The environmental conditions at each site over the duration of the test period are expressed as probability density functions (PDFs) of mean air temperature, mean relative humidity, mean wind speed, vector mean wind direction, and precipitation rate for each component 30 minute assessment interval in Figure 3. Figure 4 presents the same parameters for all assessment intervals during which the site reference configuration detected precipitation (i.e. all Ref 'Yes' cases).

The precipitation percentage represents the number of minutes of precipitation during a 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration, expressed as a percentage. PFDs of precipitation percentage are also included in Figures 3 and 4.

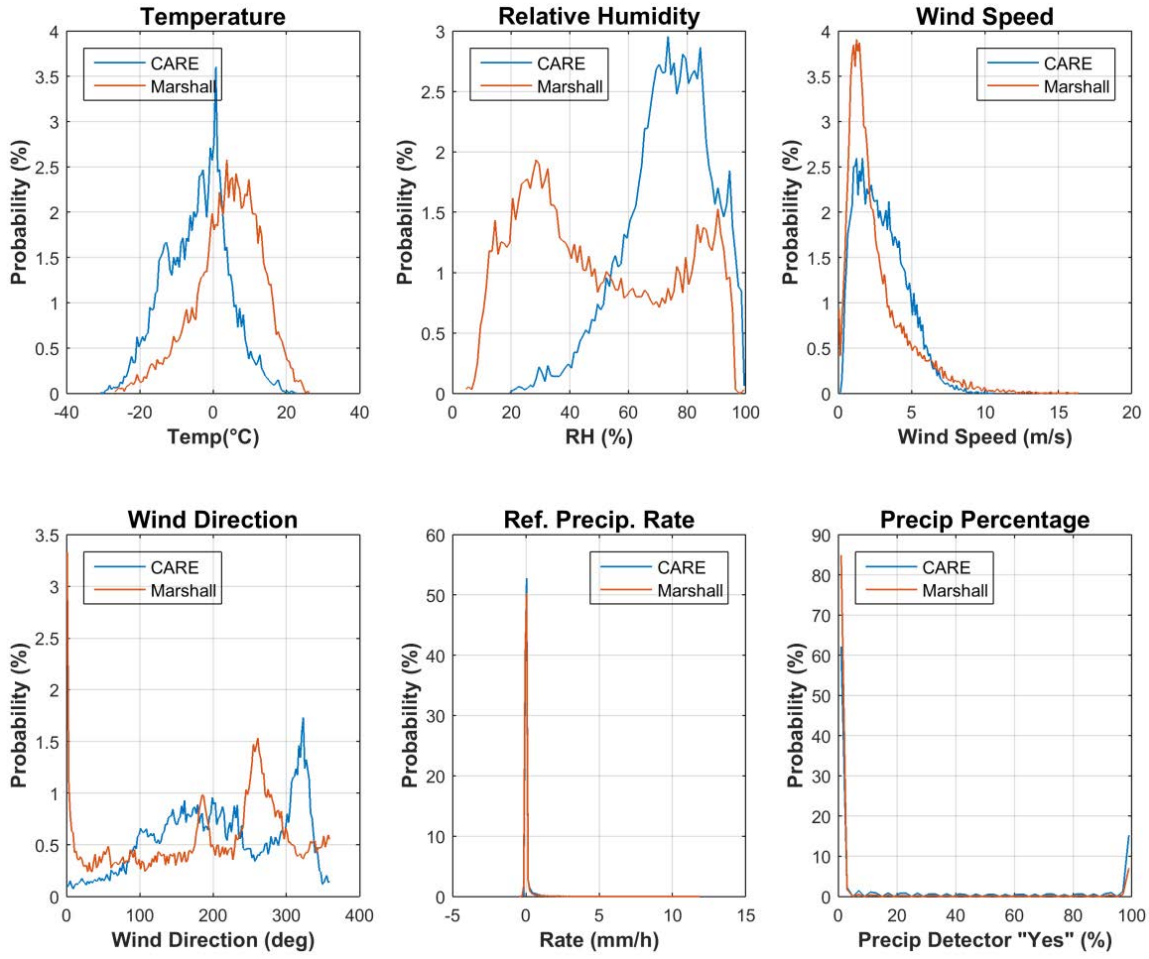


Figure 3: Summary of aggregated environmental conditions at the SPICE sites that operated CAE PMB25R gauges, over the entire duration of formal tests, as per Table 2.

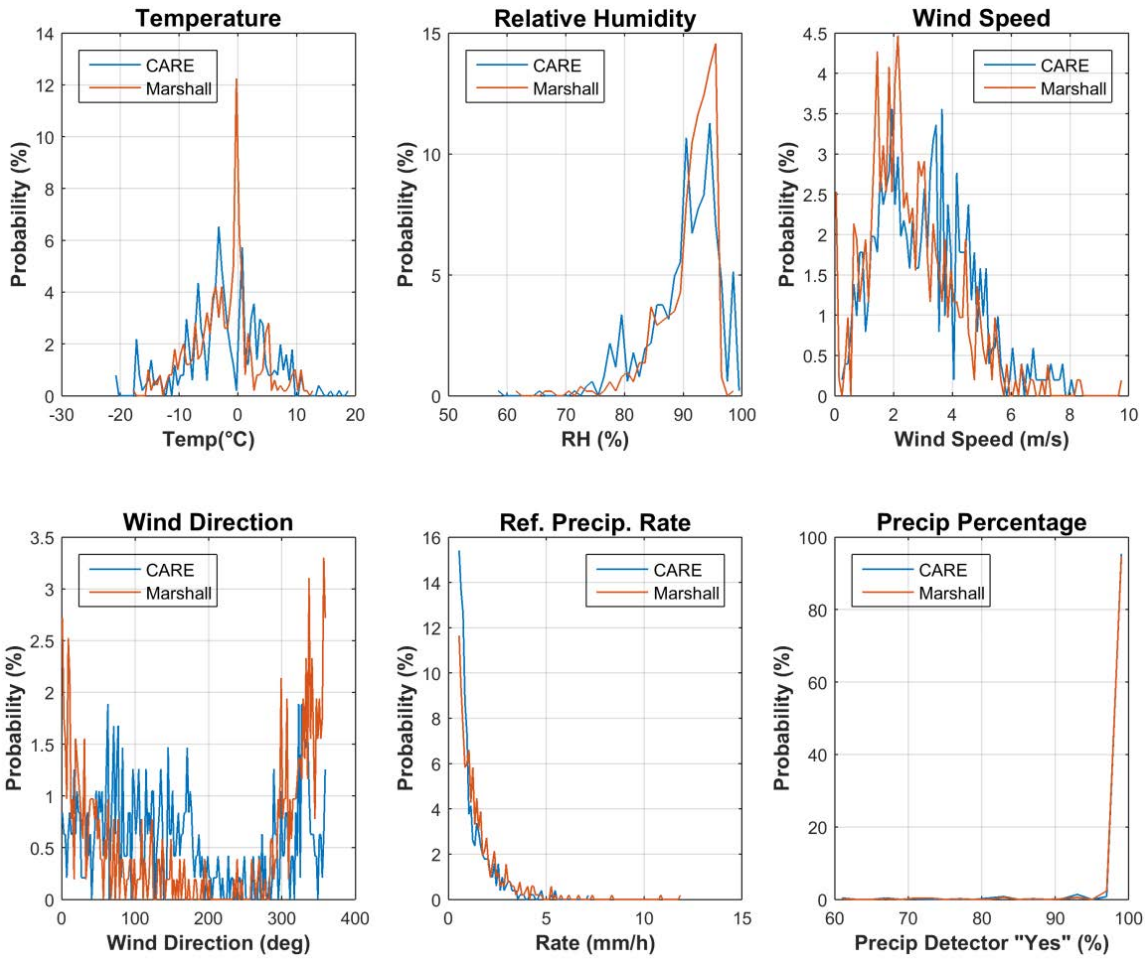


Figure 4: Summary of the aggregated environmental conditions at the SPICE sites that tested CAE PMB25R gauges during precipitation events, as reported by the site R2 reference, during the formal tests, as per Table 2.

6) Evaluation of performance over the range of operating conditions

6.1. Ability to detect and report precipitation

6.1.1. Skill score assessment

The overall ability of the SUT to detect and report the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores (Section 4.1.2) and presented in Figure 5. Scores are presented for both 30 minute and 60 minute assessment intervals. The contingency results (Section 4.1.1) corresponding to these scores are presented in Table 4.

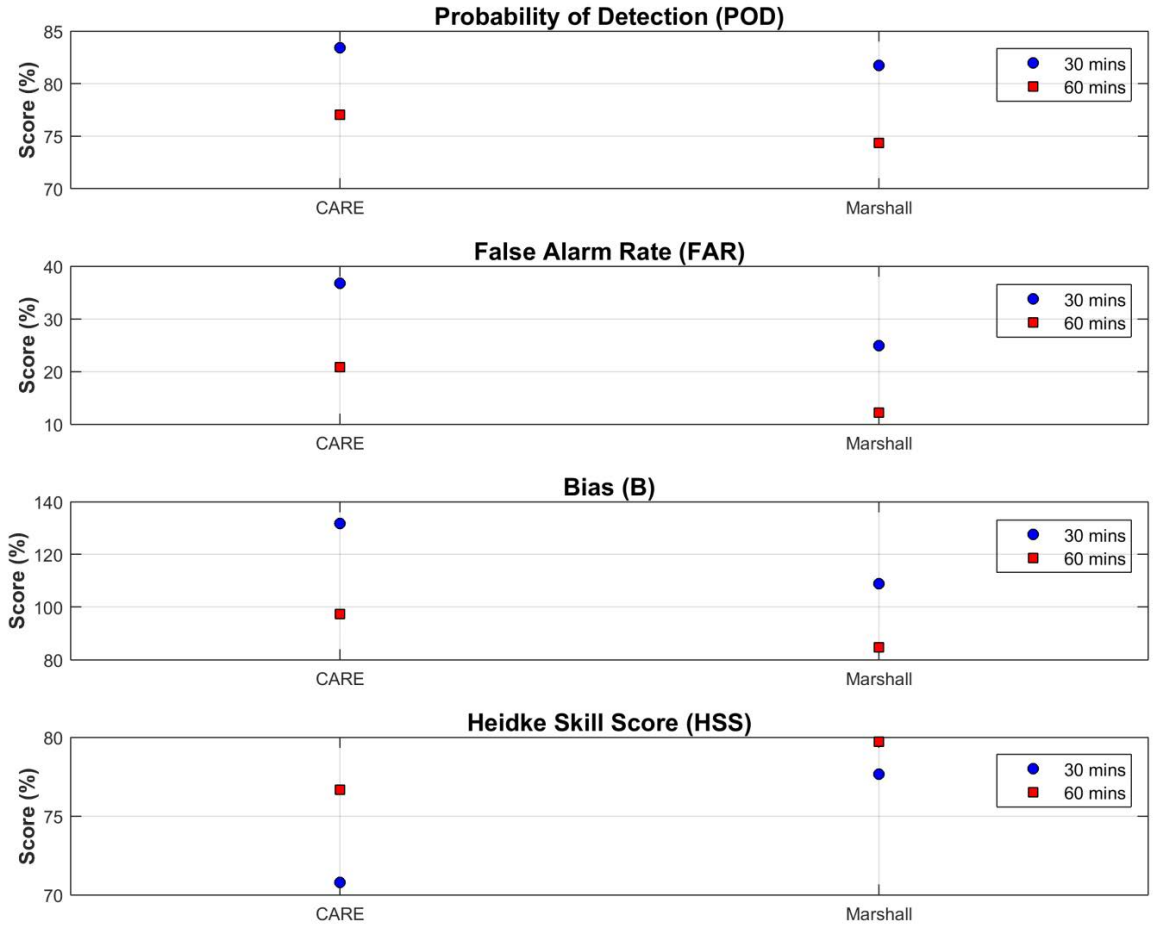


Figure 5: Skill scores for CAE PMB25R gauges during the formal test periods.

Table 4: Contingency table illustrating detection of precipitation by the CAE PMB25R relative to the specific site reference, expressed as number (and percentage) of events over the entire test period, by site.

Site	Time interval	Number of Events (% of Events)			
		YY, hits	YN, misses	NY, false alarms	NN, correct negatives
CARE	30 min	422 (2.9%)	84 (0.6%)	245 (1.7%)	13726 (94.8%)
	60 min	342 (4.7%)	102 (1.4 %)	90 (1.2%)	6689 (92.6%)
Marshall	30 min	421 (2.1%)	94 (0.5%)	140 (0.7%)	19395 (96.8%)
	60 min	316 (3.2%)	109 (1.1%)	44 (0.4%)	9543 (95.3%)

6.1.2. Characterization of response delays

Response delays were determined for the test gauges at each site using the approach outlined in Section 4.2 and compiled over the entire test period. The delays are represented as probability density functions in Figure 6. These PDFs provide an overall picture of response delays for each sensor in all precipitation types. The response delays are characterized further by separating the precipitation events by type/phase, and plotting the response delays as a function of the mean precipitation intensity observed by the reference gauge during the delay period (Figure 7).

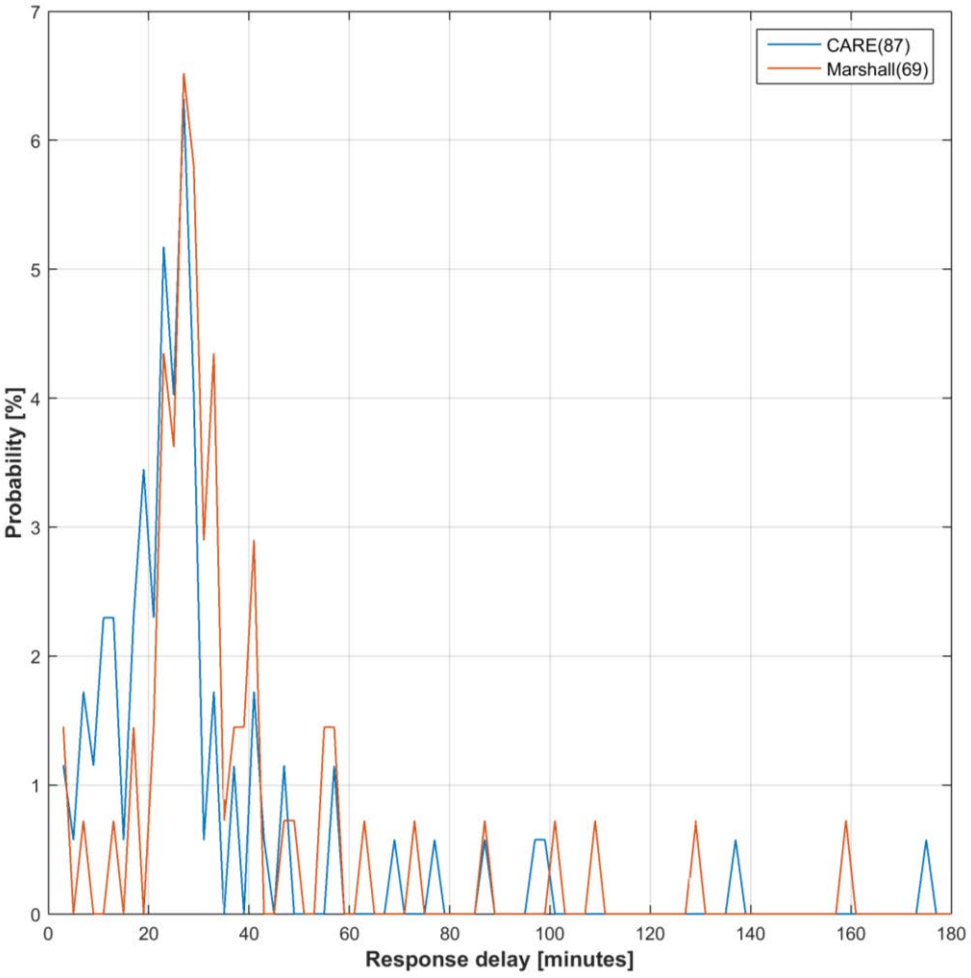


Figure 6: Response delays for heated CAE PMB25R gauges at CARE and Marshall relative to the R2 reference configuration at each site. The number of response assessment periods used to determine the delays for gauges at each site are indicated in parentheses in the legend.

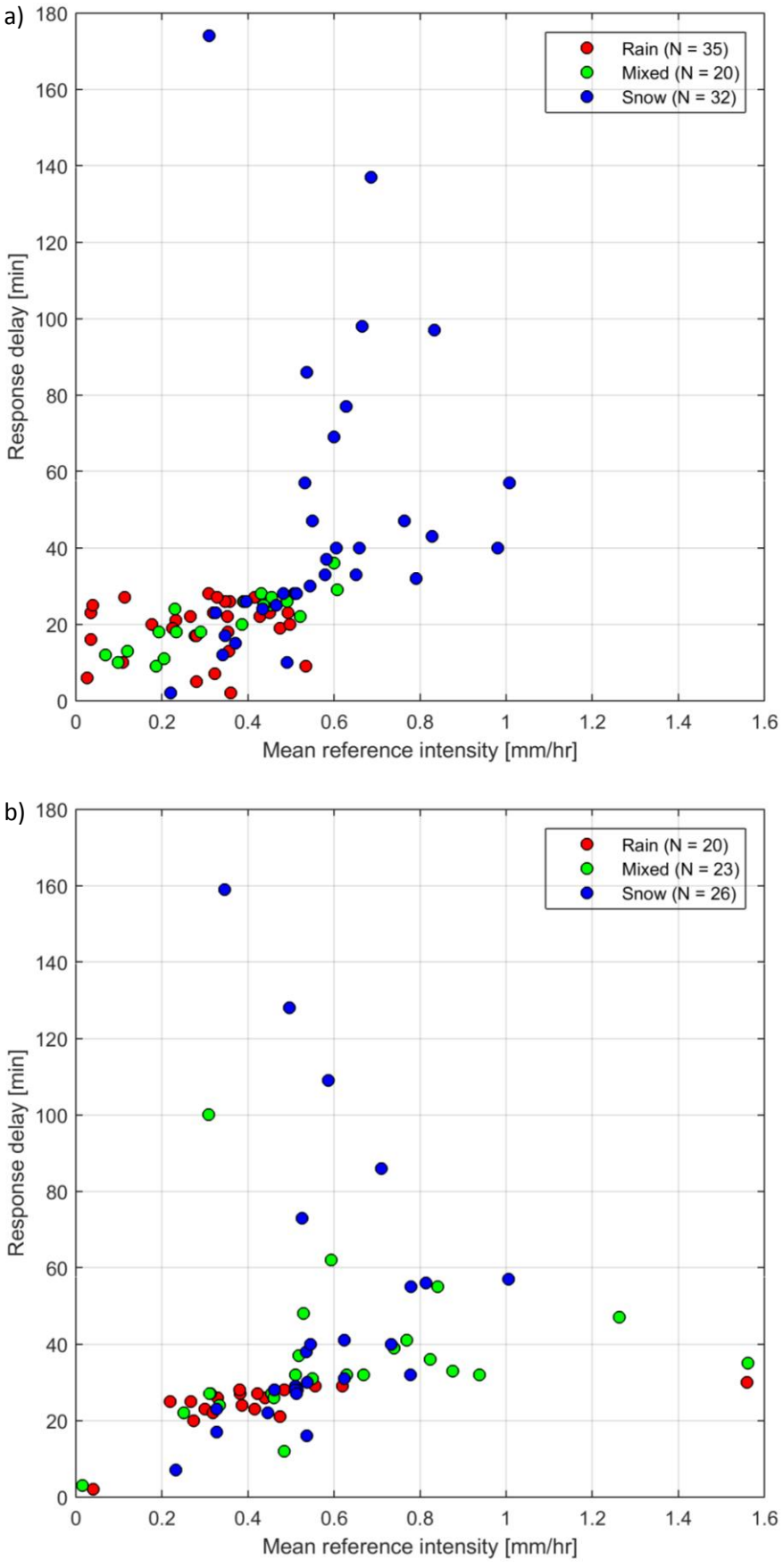


Figure 7: Response delays for CAE PMB25R gauges as a function of the mean precipitation intensity observed by the reference gauge during the delay period at (a) CARE and (b) Marshall. The predominant precipitation type for each event is determined from the maximum and minimum reported temperature during the delay period.

6.2. Ability to report accumulated precipitation

The SUT performance in terms of reporting accumulated precipitation is examined by comparing the amount reported by each sensor under test relative to the respective site reference during a given assessment interval. This is represented graphically using scatter and box plots of the catch efficiency as a function of mean wind speed at gauge height, as well as scatter plots of the amounts reported by the SUT versus the corresponding reference amounts (Figures 8 to 10). The SUT performance is also assessed in terms of the root mean square error, RMSE (Figure 11).

Only assessment intervals during which the SUT and reference both reported precipitation (YY cases) are considered in this portion of the assessment. To assess the influence of the assessment interval on SUT performance, analysis was conducted using both 30 and 60 minute intervals. The plots presented in Figures 8 to 10 include only the 30 minute results, in order to better constrain the wind speed and temperature data used in the assessment (which vary to a greater extent over 60 minute periods) and to increase the sample size of events.

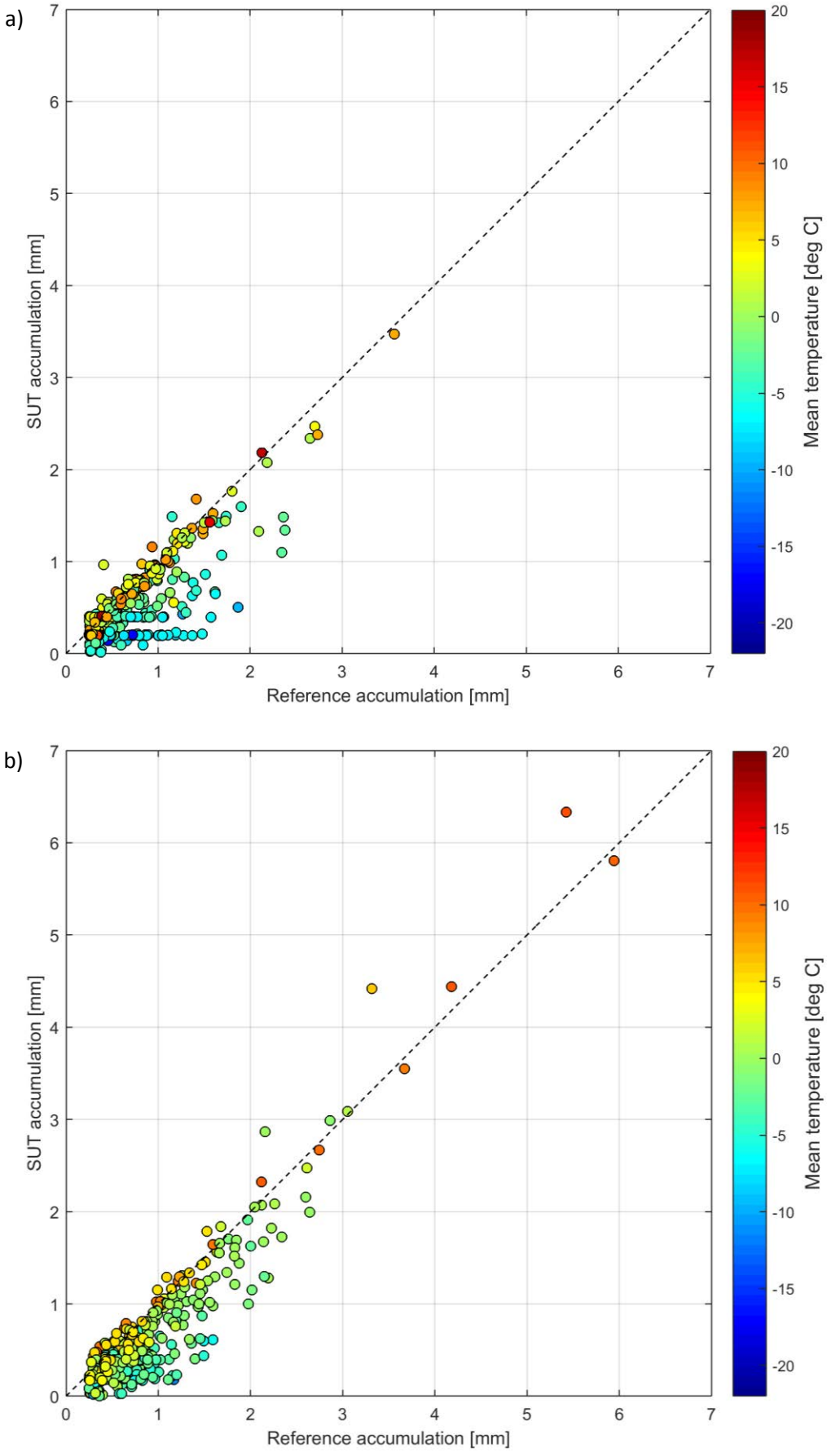


Figure 8: SUT accumulation vs. reference accumulation scatter plots for 30 minute precipitation events at (a) CARE and (b) Marshall. The mean event temperature is indicated by colour.

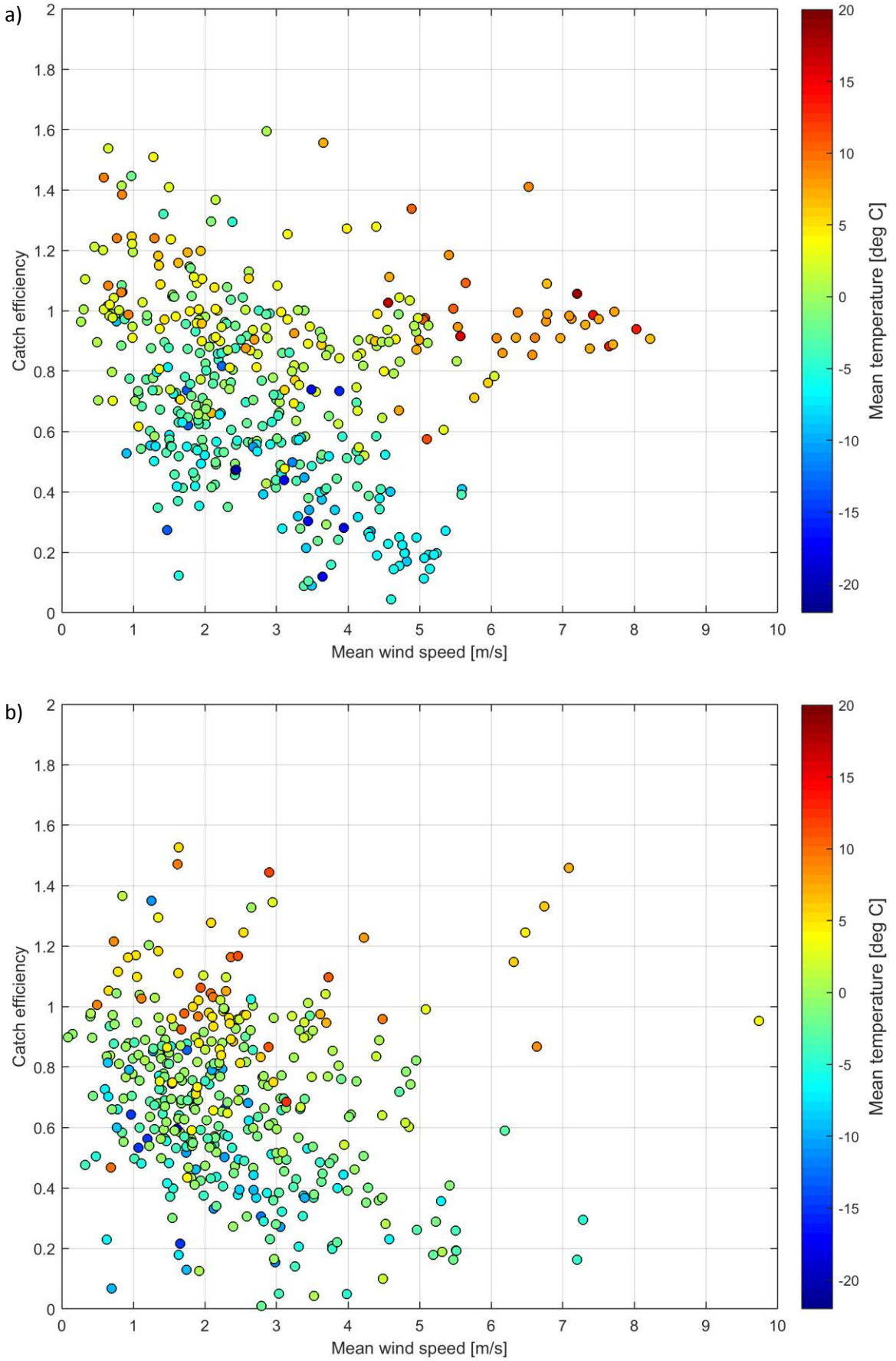


Figure 9: Catch ratio vs. mean wind speed scatter plots for 30 minute precipitation events at (a) CARE and (b) Marshall. The mean event temperature is indicated by colour.

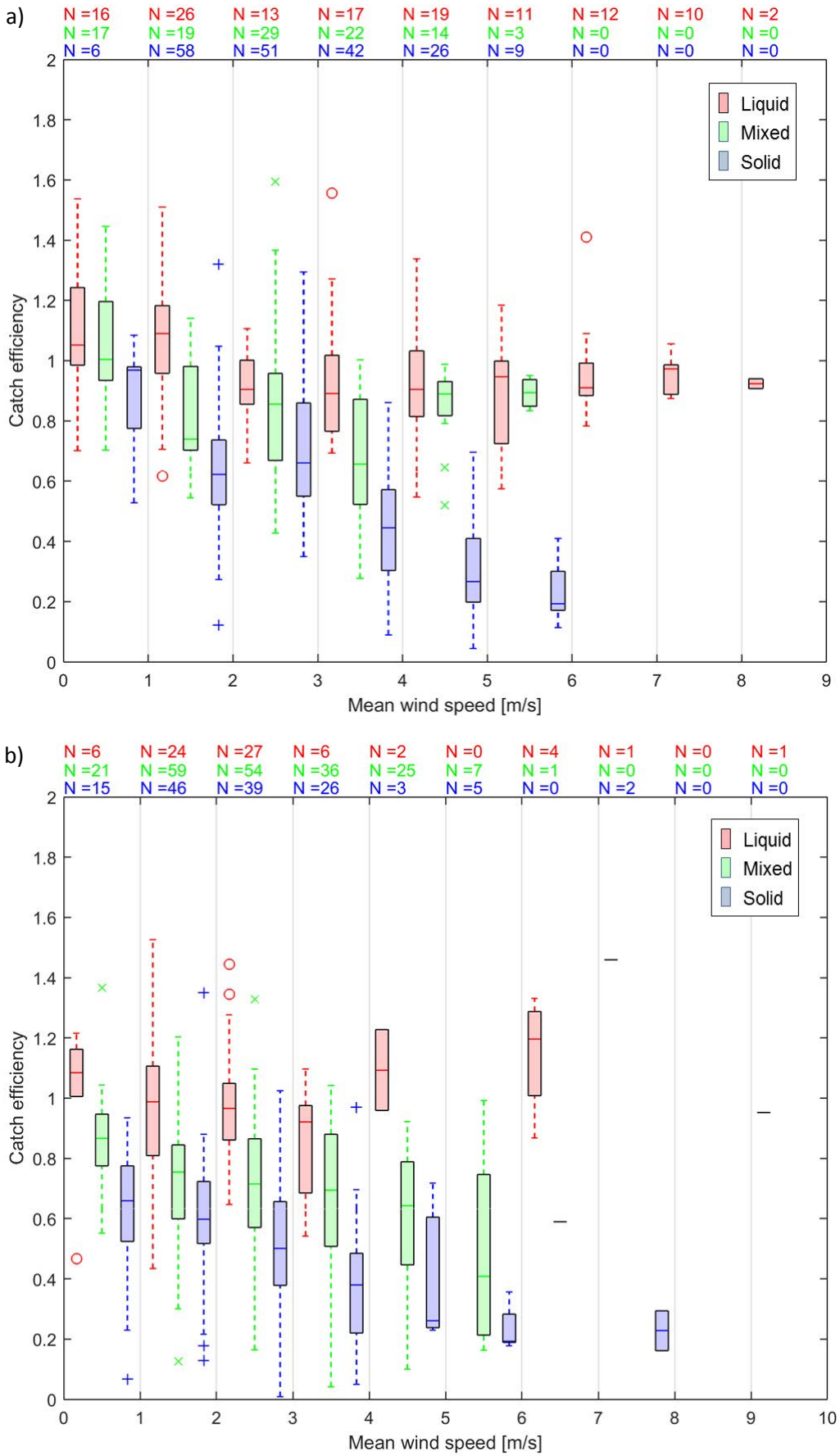


Figure 10: Catch ratio vs. mean wind speed box and whisker plots for 30 minute precipitation events at (a) CARE and (b) Marshall. The predominant precipitation type for each event is determined from the maximum and minimum reported temperature and indicated by colour.

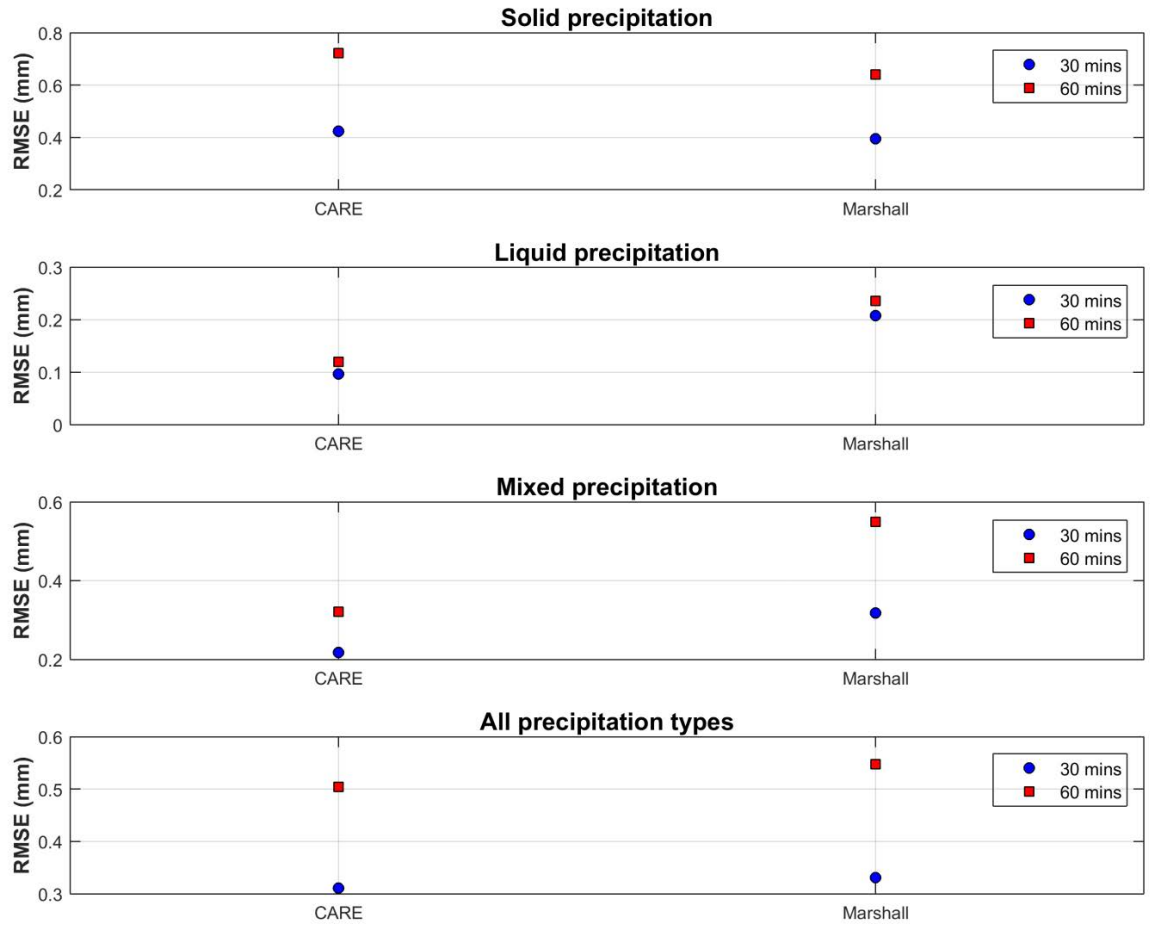


Figure 11: Graphical representation of the RMSE values, by site, and by precipitation type, for YY cases.

The total accumulation reported by each site reference and SUT, for all assessment intervals during which both detected and reported precipitation, are presented alongside the corresponding catch ratios in Table 5.

Table 5: SUT and reference accumulation and overall catch ratio, assessed for YY cases over the entire test period, by site.

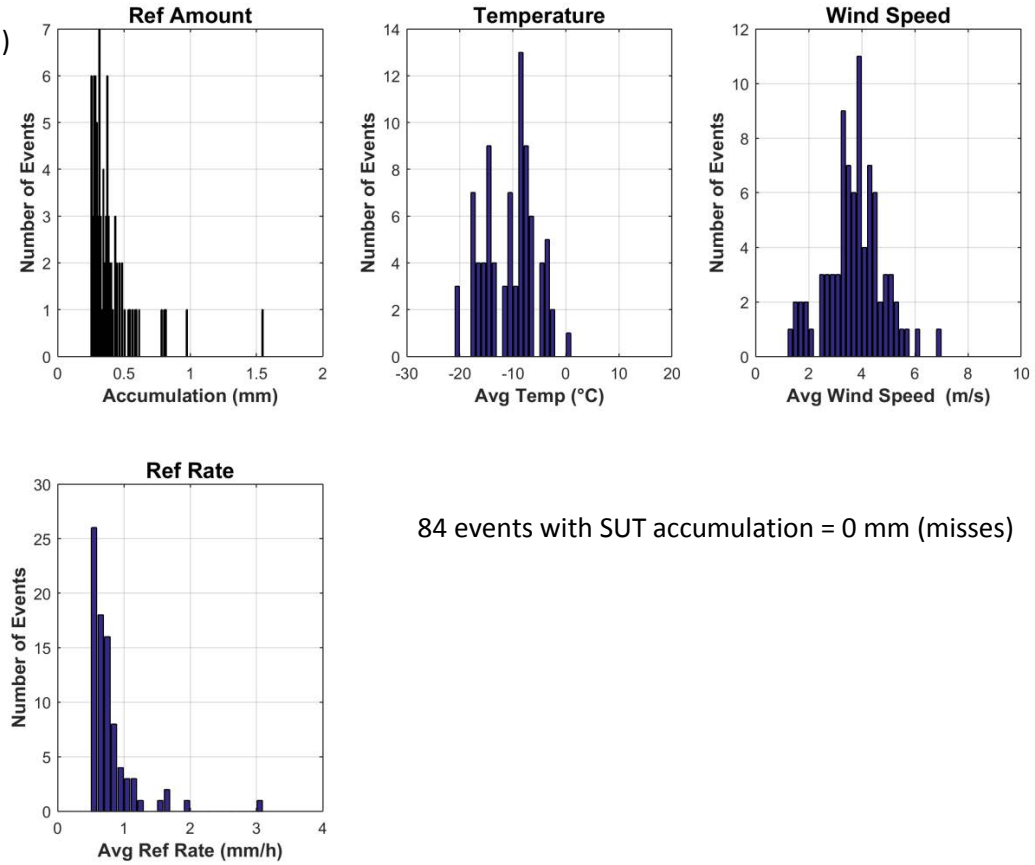
Site	Time interval	SUT accumulation	Reference accumulation	Overall catch ratio
CARE	30 min	209.4 mm	281.9 mm	0.74
	60 min	238.8 mm	330.2 mm	0.72
Marshall	30 min	266.6 mm	357.5 mm	0.75
	60 min	288.4 mm	404.8 mm	0.71

6.3. Assessment of events when the reference and the SUT do not agree on the occurrence of precipitation

Assessment intervals during which the site reference and SUT do not agree on the occurrence of precipitation – namely, the YN and NY cases (Section 4.1.1) – are characterized using histograms in Figures 12 and 13. The histograms include accumulated precipitation, precipitation intensity as reported by the reference, and corresponding site conditions.

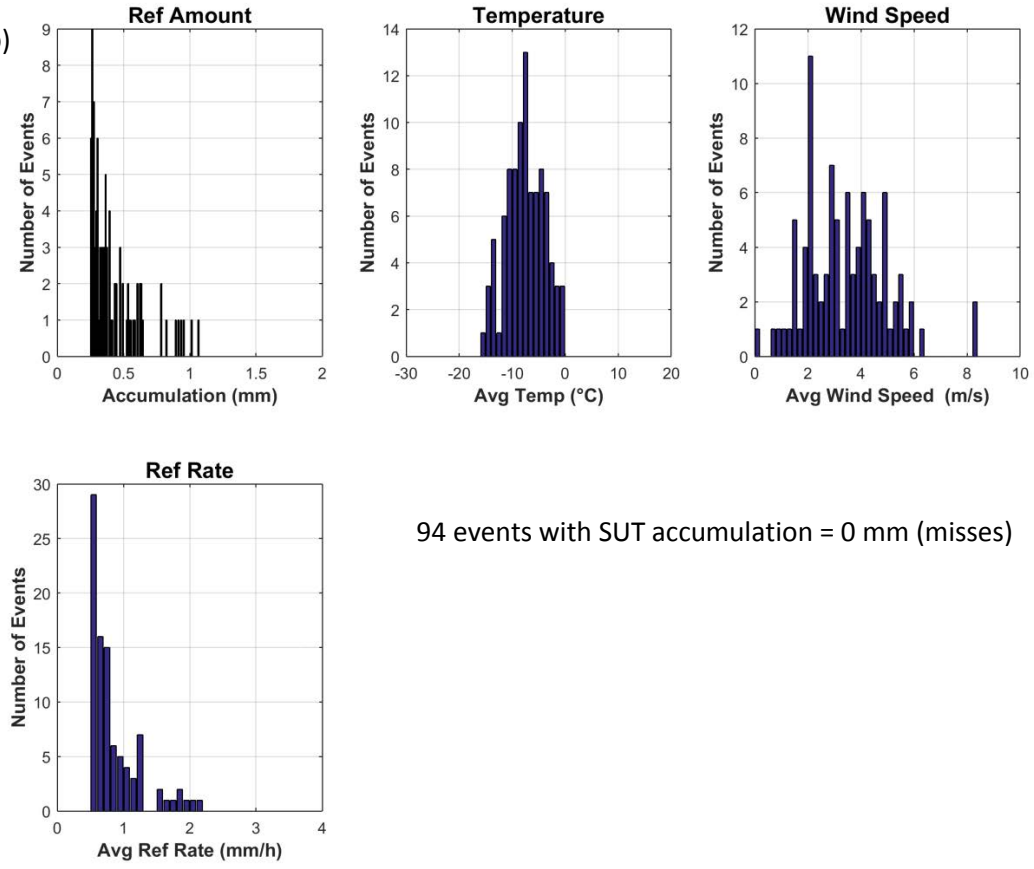
The total SUT and reference accumulations over the test period include contributions from YN and NY cases, which impact the overall catch efficiency. These contributions may be significant, given the response delays associated with heated tipping bucket gauges. Total accumulation and catch efficiency results presented in Table 5 (YY cases only) are expanded to include contributions from YN and NY cases in Table 6.

a)



84 events with SUT accumulation = 0 mm (misses)

b)



94 events with SUT accumulation = 0 mm (misses)

Figure 12: Histograms of reference accumulation, mean temperature, mean wind speed, and reference precipitation rate for YN cases at (a) CARE and (b) Marshall.

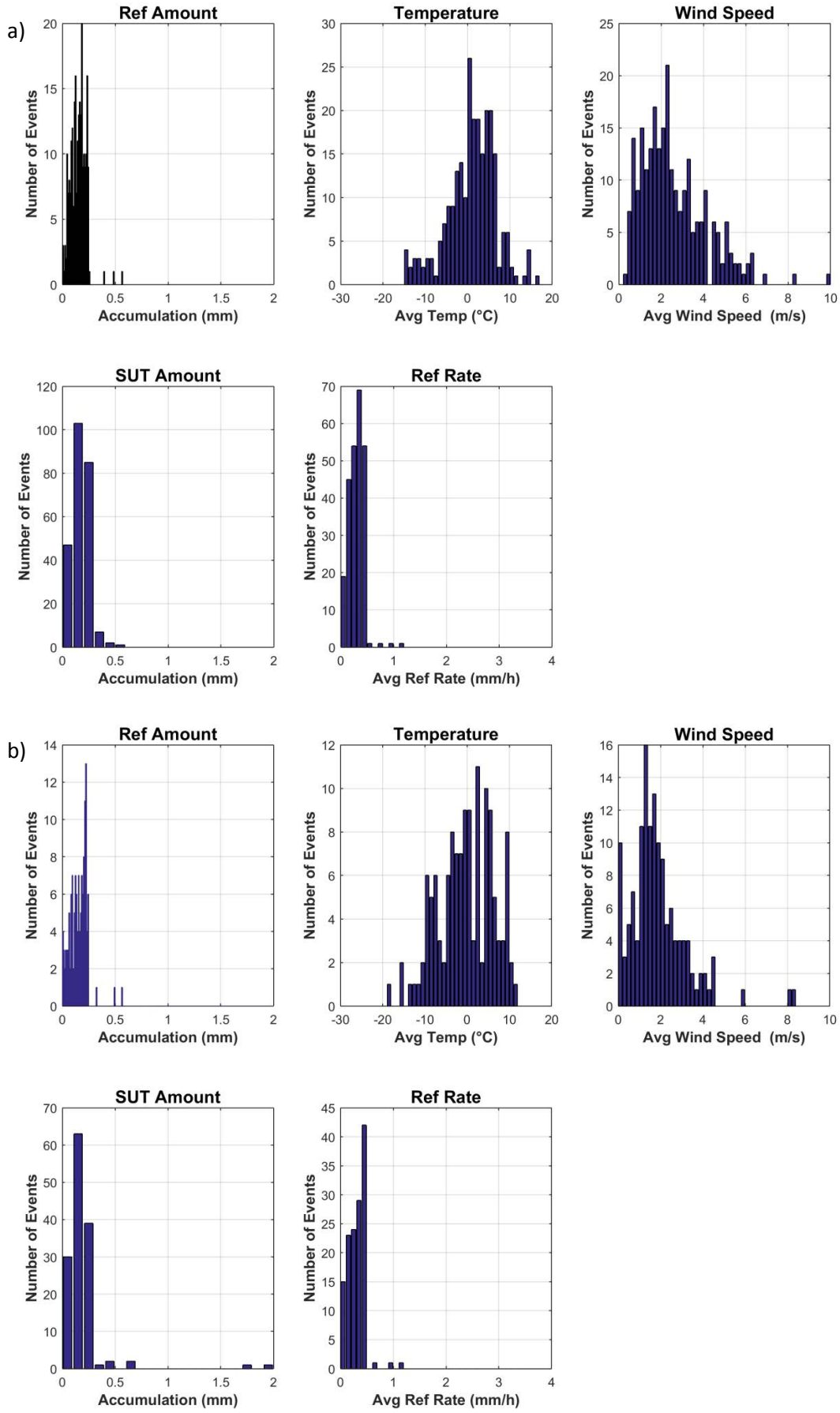


Figure 13: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, and reference precipitation rate for NY cases at (a) CARE and (b) Marshall.

Table 6: SUT and reference accumulation and overall catch ratio for each site/gauge, assessed for the entire test period, for all YY, YN, and NY cases.

Site	Time interval	SUT accumulation	Reference accumulation	Overall catch ratio
CARE	30 min	251.1 mm	351.9 mm	0.71
	60 min	254.2 mm	389.6 mm	0.65
Marshall	30 min	297.5 mm	421.7 mm	0.71
	60 min	302.6 mm	465.3 mm	0.65

7. Interpretation of Results

7.1. Operating conditions

The full range of conditions under which the test gauges were operated is illustrated in Figure 3. The conditions relevant to gauge operation are as follows:

- Temperatures between -31 °C and 27 °C;
- Precipitation intensity within 12 mm/hr.

The reported intensity values fall within the manufacturer's specifications. The reported temperatures extend below the minimum specified value of -30 °C; however, only two 30 minute periods were observed with temperatures < -30 °C over the duration of the experiment at both sites, and none during precipitation events (Figure 4), so the impact on the present analysis is not expected to be significant.

Both sites are located in the continental interior, but the conditions at Marshall are generally drier and less windy relative to those at CARE, with a larger proportion of events with RH < 80% and a smaller proportion of events with mean wind speeds > 3 m/s (Figure 3). In general, the distributions of temperature, relative humidity, temperature, and precipitation rate are similar for the precipitation events at CARE and Marshall (Figure 4).

7.2. Ability to detect and report precipitation

7.2.1. Skill score assessment

The performance of the sensors under test was considered in terms of skill scores, computed for both 30 and 60 minute assessment intervals (Figure 5). Using the shorter 30 minute assessment intervals, the Probability of Detection is higher for both gauges under test relative to the scores for the longer 60 min assessment intervals, but the False Alarm Rate (FAR) and Bias (B) is also higher for both gauges. As a result, the Heidke Skill Score (HSS) values are higher for both gauges when 60 minute assessment intervals are used. The comparison of skill scores for the test gauges at each site will be considered for only the 60 minute assessment intervals, as hourly reports are broadly used, operationally.

The probability of detecting precipitation was similar for the gauges under test, with POD values of approximately 77% and 74% for the gauges at CARE and Marshall, respectively. The relative differences between the False Alarm Rates and Biases for the test gauges at each site are attributed to the relatively higher number of false alarm events observed at CARE (Table 4): FAR of 21% for CARE and 12% for Marshall; Bias of 97% for CARE and 85% for Marshall. The resulting Heidke Skill Scores are similar for the gauges at each site: approximately 77% for CARE and 80% for Marshall.

7.2.2. Characterization of response delays

7.2.2.1. Response delay PDFs

Probability distribution functions (PDFs) of response delays between the onset of precipitation as observed by the reference configuration and the first response by the CAE PMB25R heated tipping bucket gauge are shown in Figure 6. For the gauges under test at both sites, response times within 60 minutes have the highest probability. Within this region of the PDFs, the distribution for CARE is shifted toward shorter times, while that for Marshall is shifter toward longer times. For both sites, the PDFs tail off gradually, with low probabilities of response times up to 180 minutes.

The subtle differences between the PDFs may result from differences in site or gauge configuration, or to differences in the conditions experienced at each site. The proportion of solid precipitation events at each site was similar (45% at CARE, 42% at Marshall), but CARE had more liquid events relative to mixed events (31% and 24%, respectively) compared to Marshall (19% and 39%, respectively), which could explain the shift in the PDF toward shorter response times.

For precipitation occurring within a given 30 or 60 minute interval (as indicated by the reference configuration), the response delay PDFs for the test gauges at both sites indicate the potential for the TB response to occur within a subsequent 30 or 60 minute interval. The delays with highest probability in the PDFs are at about 30 to 35 minutes, so it is more likely that the TB gauges will respond to precipitation detected by the reference within a 60 minute period than within a 30 minute period. This is reflected by the lower number of false alarm (NY) cases for the 60 minute intervals relative to 30 minute intervals in Table 4, and the lower FAR and Bias for these intervals in Figure 5.

Another important consideration is that even when both the SUT and reference detect precipitation within a given interval, the TB may actually be reporting precipitation collected during previous intervals, or may only report a fraction of the precipitation within that interval (reporting the rest during a subsequent interval). These ‘false YY cases’ will impact the assessment of TB gauge performance in terms of reporting accumulated precipitation (Section 7.3).

7.2.2.2. Influence of precipitation intensity, phase

The dependence of delay times on the mean intensity of precipitation preceding the TB response, as reported by the reference configuration, is illustrated in Figure 7. For the test gauge at CARE, response times are within 30 minutes for all precipitation types (liquid, mixed, solid) when the mean intensity of precipitation is within about 0.55 mm/hr. This suggests that for low intensities, the TB response is independent of precipitation phase, and that heating/melting of solid precipitation does not increase delay times significantly. For mean solid precipitation intensities exceeding ~ 0.55 mm/hr, however, the delay times are longer, extending to about 140 minutes, which can likely be attributed to heating/melting.

Similar trends are observed for the test gauge at Marshall, with delays within about 30 minutes for all precip types with intensities within about 0.5 mm/hr, and longer delays for solid precipitation at higher intensities. The Marshall data show mixed events with longer delays at higher intensities,

which can be attributed to events in which there was a solid precipitation component, or a transition from solid to mixed or liquid precipitation. The plots for both sites show events with longer delays (up to 175 minutes) for solid and mixed events with intensities below 0.5-0.55 mm/hr, which are attributed to events with variable intensity over a longer duration.

7.3. Ability to report accumulated precipitation

7.3.1. Performance when both SUT and reference detect precipitation

7.3.1.1. Temperature, phase, and wind speed influence

Focussing on the 30 minute events in which both the sensors under test and site reference configurations report precipitation (YY cases, or 'hits,' only), general stratification of the data by temperature is observed for CARE and Marshall in Figures 8 and 9. Colder temperatures are generally correlated with larger differences in accumulation between the reference and SUT (Figure 8), and hence lower catch efficiencies (Figure 9). The catch efficiencies for events at colder temperatures generally decrease with increasing wind speed (Figure 9), owing to the enhanced influence of wind speed on the collection of mixed and solid precipitation (relative to the collection of liquid precipitation) by the gauges under test.

Given the marked scatter of event data in Figure 9, the phase and wind speed influences on catch efficiency are more clearly illustrated by the box and whisker plots in Figure 10. For CARE, the median catch efficiency for solid precipitation decreases to about 0.6 for wind speeds between 2 and 3 m/s, falling to around 0.4 for winds up to 4 m/s, then 0.3 to 0.2 for wind speeds of 5 to 6 m/s. For Marshall, the decrease in catch efficiency with increasing wind speed is more abrupt, with values > 0.6 for wind speeds ≤ 1 m/s, values between 0.4 and 0.6 for wind speeds between 2 and 3 m/s, and between 0.2 and 0.4 for wind speeds > 3 m/s.

These general trends for solid precipitation characterize the maximum expected wind speed effects on catch efficiency the gauges at CARE and Marshall under the conditions tested. As observed in Figure 10, the trends for mixed and liquid precipitation fall within the above limits.

7.3.1.2. Root mean square error

The RMSE was calculated for all 30 and 60 minute events during which the reference and SUT both detected precipitation (YY cases), for each site. The overall RMSE calculated for each site (all precipitation types), and RMSE values calculated for each precipitation type, are shown in Figure 11. The duration of the assessment interval does not significantly impact the RMSE values observed for liquid precipitation events at both sites, which are impacted less by response delays and wind-induced undercatch than mixed and solid precipitation events. For the latter events, which are impacted to a greater extent by response delays and wind-induced undercatch, the RMSE values are larger for 60 minute assessment intervals for the test gauges at both sites. Given the broad use of hourly reports in meteorological operations, further discussion of RMSE values will be considered for 60 minute assessment intervals only.

The overall RMSE was within 0.6 mm for the test gauges at both CARE and Marshall. Hence, for the specific site and gauge configurations tested, over the range of conditions tested, the absolute uncertainty of the CAE PMB25R relative to a DFAR configuration can be considered to be within 0.6 mm for 60 minute intervals. The largest phase-specific RMSE values were determined for solid precipitation (within 0.8 mm over 60 minutes), owing to the larger wind-induced differences in reference and SUT accumulation relative to mixed (RMSE within 0.6 mm over 60 minutes) and liquid (RMSE within 0.3 mm over 60 minutes) precipitation.

7.3.1.3. Overall catch efficiency

The total accumulated precipitation recorded by the reference and sensor under test at each site was compiled for all 30 and 60 minute events during which both detected precipitation (YY cases) and used to calculate the overall catch ratio. The results are provided in Table 5. Focussing on the more operationally-relevant 60 minute values, the overall catch ratios for CARE and Marshall are 0.72 and 0.71, respectively, indicating similar overall performance by the gauges under test at both sites.

7.3.2. Characterization of conditions when the SUT and reference do not agree on the occurrence of precipitation

The conditions during 30 minute events in which the reference detected precipitation, but the SUT did not (YN cases), are depicted as histograms in Figure 12. These events are characterized by low precipitation rates, typically below 1.5 mm/hr, and temperatures close to, or below, freezing. At these temperatures, the precipitation is likely mixed or solid, and hence subject to longer response delays at the observed rates. Hence, these may be cases in which the TB response to precipitation occurs during a subsequent 30 minute period. There is also potential for heating to cause evaporation or sublimation of incident precipitation at low intensities.

The conditions during 30 minute events in which the SUT detected precipitation, but the reference did not (NY cases), are shown as histograms in Figure 13. These events are also characterized by low precipitation rates, typically below 0.5 mm/hr (which corresponds to the reference accumulation threshold for a precipitation event over a 30 minute interval), and wind speeds below about 8 m/s. These events most likely result from delayed tipping bucket responses that coincide with intervals during which the reference did not detect precipitation.

Given the influence of response delays, estimates of the overall catch ratio for tipping bucket gauges should include the reference and SUT accumulations during YN and NY cases (and not just during the YY cases, as considered above). The resulting total accumulations for all YY, YN, and NY cases, and the corresponding overall catch efficiencies, are provided for the test gauges at each site in Table 6. The overall catch efficiencies for the gauges under test at both CARE and Marshall were 0.65 for 60 minute assessment intervals, demonstrating similar overall performance of the gauges tested at each site.

8. Operational consideration

The CAE PMB25R was easy to install, easy to connect, and the setup of data collection was straightforward. The heating configuration requires significant power, which may require investment in site infrastructure; the use of power cables longer than 2 m can lead to insufficient power for heating, with the potential for negative impacts on gauge performance (ice buildup, clogging).

8.1. Maintenance

Each site completed the gauge field calibration and verification as per manufacturer recommendations, at least once a year. The calibration records have been stored by each site team.

8.2. Noted issues

- The site team at CARE noted instances of communication loss with the gauge, which were attributed to the improper closing of data ports by the gauge, and resulted in data loss. Communication was restored by a hard reset of the gauge. This issue was resolved with a firmware update.

- The manufacturer visited both sites to inspect the installations. At Marshall, this inspection prompted the replacement of the power cable to the gauge in early October, 2013.

- After this replacement, the Marshall site team observed cases in which the gauge missed most or all of an event observed by the reference configuration under cold conditions. This was attributed to insufficient heating under these conditions.

- The Marshall site team also observed that clogging of the gauge orifice with debris can occur, depending on the location of the gauge and nature of the environment.

- Updating the gauge firmware at Marshall required on-site assistance from the manufacturer.

9. Performance Considerations

- Heating is critical to ensure proper functioning of the gauge under winter conditions. The voltage should be checked at the gauge location to ensure that the heaters are supplied with the recommended power. Voltage drops – for example, resulting from the use of long power cables – can impair the functioning of the gauge in winter conditions.

- Heating enables the measurement of solid precipitation, but the time required for melting can delay the time between the collection of precipitation in the funnel and the gauge response to that precipitation (response delays). Further, heating may result in the evaporation/sublimation of incident precipitation at low intensities.

- Response delays must be considered when using the gauge in operational settings. Ideally, the reporting interval (i.e. hourly observations) should exceed the maximum expected

response delay; however, the potential remains for carry-over of precipitation accumulation from previous intervals, and for delayed responses occurring in subsequent intervals.

- While not tested in SPICE, shielding of the gauge should be considered by the manufacturer and/or potential users as a means of increasing the catch efficiency at higher wind speeds.

- The collection of ancillary measurement data is recommended: air temperature and wind speed (at gauge height) to enable the application of adjustments to measurements (i.e. using transfer functions); and reports from a precipitation detector with high sensitivity to enable the identification of missed events or false alarms due to response delays.

- The application of transfer functions to gauge measurements, if available, is recommended as a post-processing step to account for reductions in catch efficiency with increasing wind speed. This recommendation comes with the caveat that the catch ratio data used in the derivation of the transfer functions will be impacted by response delays. Response delays can lead to scatter in the catch ratio data (among other factors), and hence, uncertainty in the transfer functions; however, the potential remains for the improvement of data via adjustment. Considering the catch ratios over longer time periods (e.g. 1 hour) may provide one avenue for reducing scatter in the catch ratio/wind speed relationship and the uncertainty of related transfer functions.

WMO-SPICE Instrument Performance Report
EML UPG1000

1) Technical specifications (from manufacturer provided documentation)

Instrument model:	EML UPG1000 – Universal Precipitation Gauge
Physical principle:	Heated tipping bucket (TB) gauge; tips are recorded when the bucket fills to capacity, triggering a contact closure, dual reed switch.
Bucket capacity:	0.1 mm
Collecting area:	1000 cm^2
Heating configuration:	Funnel heating (300 W, mats) and internal heating (27 W, ceramic resistors). Total heating power: 330 W. These heaters are controlled independently, triggered by separate temperature sensors, which reduces the potential for internal overheating (and subsequent evaporation).
Operating temperature range:	-40 °C to 60 °C
Measurement range:	0.1 mm/hr to 500 mm/hr
Measurement uncertainty:	Unknown
Sensitivity:	0.1 mm reporting resolution (corresponding to bucket amount)

a)

b)

Figure 1: EML UPG1000 installation at (a) Marshall, with vertical shield slats, and (b) Sodankylä, with L-shaped shield slats.

2) Data output format

Gauge data output: Contact closure Reed switch output (times of tips), corresponding to 0.1 mm precipitation. The number of tips within a given time period can be used to compute the precipitation intensity and/or accumulation within that period.

The manufacturer recommends an intensity-dependent correction for 1 minute samples:

Intensity (mm/hr) = (6.540E-4(2y)^2 + 9.562E-1(2y))/2

where y is the intensity in mm/hr computed from the number of tips:

y = number of tips/min * 0.1 mm * 60 min/1 hr

The correction equation is used for all data analyzed in this report. Minutely accumulations are derived from corrected minutely intensity reports. Analysis is based on accumulated precipitation, computed as the cumulative sum of minutely accumulation values.

3) SPICE test configuration

Shield: Shielded (configurations vary)

Test sites: Marshall (USA); Sodankylä (Finland)

Sensor provider(s): Environmental Measurements Limited (EML)

Figure 2: Map of SPICE sites where EML UPG1000 gauges were tested.

A summary of the configuration of instruments as tested, the duration of tests and availability of data reflected in these results, and the ancillary measurements used, by site, is available in Tables 1, 2, and 3, respectively.

Table 1: Summary of gauge configurations and data output, by site. Details and photos on individual site configurations are available in the respective site commissioning protocols.

	Marshall	Sodankylä
Field configuration	Shielded (vertical slats; see Fig. 1a)	Shielded (L-shaped slats; see Fig. 1b)
Height of installation (gauge rim)	1.8 m	2 m
SUT data output frequency	1 min	1 min
Data QC	SPICE QC methodology	
Data temporal resolution	1 min	
Processing interval for SPICE data analysis	30 min, 60 min	

Table 2: Data availability, by measurement season and site.

Measurement season	Marshall	Sodankylä
Season 1 (Oct. 2013 – Apr. 2014)	X	✓
Season 2 (Oct. 2014 – Apr. 2015)	X	✓

No data are available for the gauge installed at Marshall, due to an issue with the tipping mechanism that was not identified until the conclusion of the experiment. This issue is described in Section 6.

The analysis presented in this report will focus on the gauge installed at Sodankylä.

Table 3: Summary of reference and ancillary measurements, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	Marshall	**Sodankylä**
R2 site reference	Geonor 600 (DFAR)	OTT Pluvio^2 (DFAR)
R2 precip detector	Thies LPM (Site*)	OTT Parsivel2 (Site*, 2013-2014), OTT Parsivel2 (DFAR, 2014-2015)
Ancillary temp sensor	MetOne, model 060A-2/062, 2144-L	Vaisala HMP155 (2 m)
Ancillary RH sensor	Campbell Scientific CS500	Vaisala HMP155 (2 m)
Ancillary wind sensor	RM Young Wind Monitor 05103 (2 m)	Thies acoustic 2D wind sensor (3.5 m)

*A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR-fence; however, in cases where a more sensitive detector is available outside of the DFIR-fence, or there are issues with the detector within the DFIR-fence, a precipitation detector elsewhere on the site can be employed.

4) Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6.1 of the WMO-SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute assessment intervals:

- Ref 'Yes' : R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No' : R2 weighing gauge < 0.25 mm AND/OR precip detector recording < 18 min of precip;
- SUT 'Yes' : SUT accumulation > 0 mm;
- SUT 'No' : SUT accumulation = 0 mm.

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores.

4.1.2. Skill score assessment

The ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

The above scores are computed using the formulations provided in Section 3.6.1.3 of the WMO-SPICE Final Report.

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

4.2. Sensor-specific considerations: response delays

Tipping bucket gauges require that an amount of precipitation corresponding to the bucket capacity is accumulated before a tip is triggered and the gauge records precipitation. This can result in response delays relative to the reference configuration. Heated tipping bucket gauges are subject to further delays, as any solid precipitation in the funnel must be melted before reaching the bucket and potentially triggering a tip, and the heating itself can potentially evaporate incident precipitation. These response delays will impact the comparison with the reference. For this reason, the assessment of TB gauge performance relative to the reference configuration is also considered over 60 minute intervals, using the same conditions and thresholds outlined above in Section 4.1.1.

Response delays are quantified by determining the time elapsed between the onset of precipitation as determined by the reference configuration, and the first tip recorded by the TB. The assessment is based on periods with at least 30 minutes of precipitation, as identified by the reference configuration, followed by at least 180 minutes without precipitation. This extended period without precipitation is intended to allow additional time for the melting and recording of precipitation by heated TB gauges. Additional details are provided in Section 3.6.1.4.4 of the WMO-SPICE Final Report.

5) Environmental conditions

The environmental conditions at each site over the duration of the test period are expressed as probability density functions (PDFs) of mean air temperature, mean relative humidity, mean wind speed, vector mean wind direction, and precipitation rate for each component 30 minute assessment interval in Figure 3. Figure 4 presents the same parameters for all assessment intervals during which the site reference configuration detected precipitation (i.e. all Ref 'Yes' cases).

The precipitation percentage represents the number of minutes of precipitation during a 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration, expressed as a percentage. PDFs of precipitation percentage are also included in Figures 3 and 4.

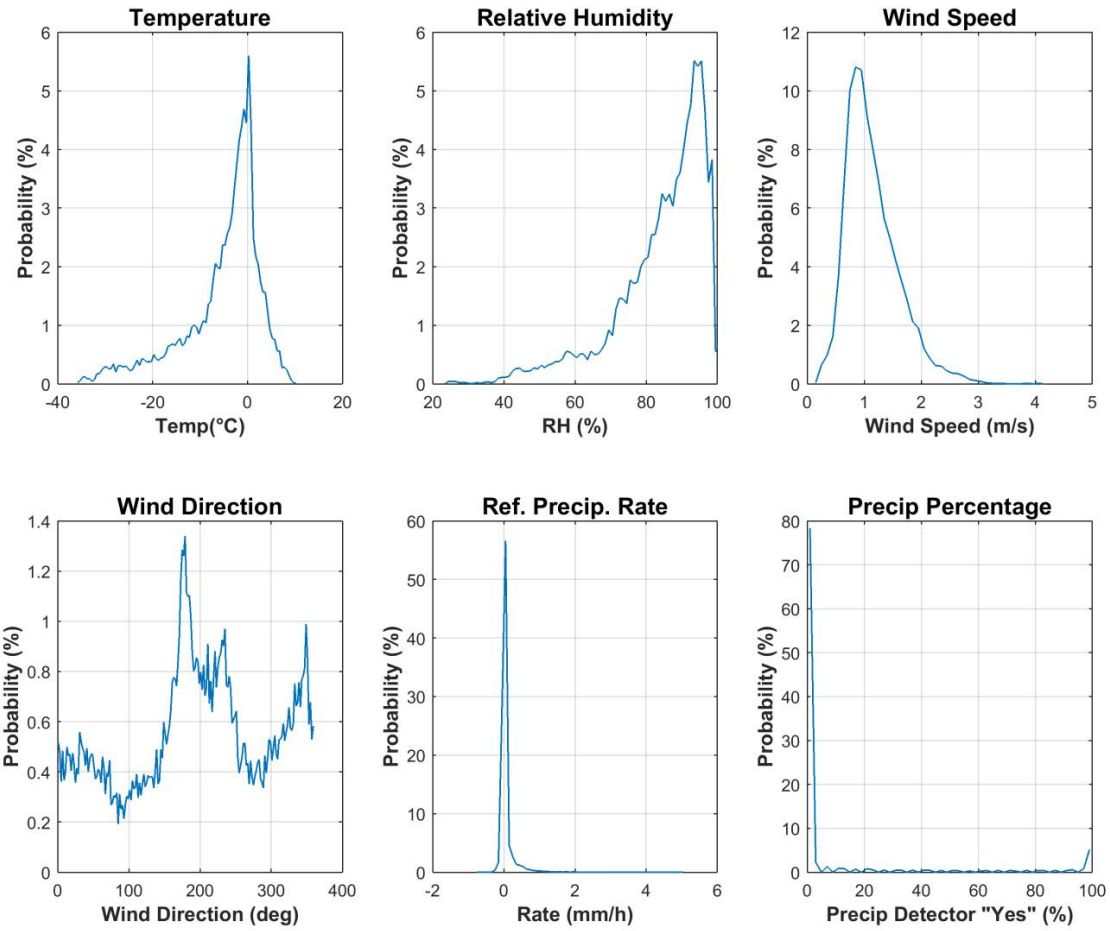


Figure 3: Summary of aggregated environmental conditions at Sodankylä, over the duration of formal tests, as per Table 2.

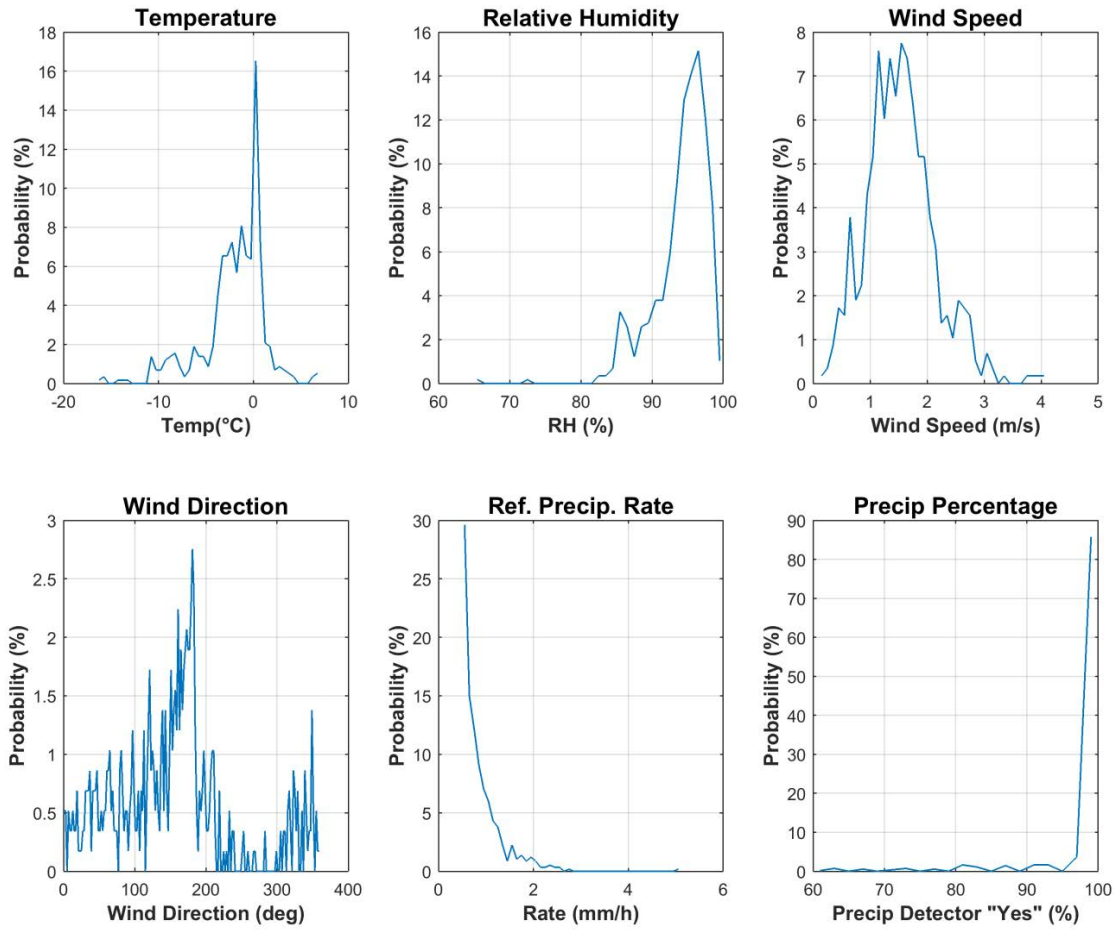


Figure 4: Summary of the aggregated environmental conditions at Sodankylä during precipitation events, as reported by the site R2 reference, during the formal tests, as per Table 2.

6) Evaluation of performance over the range of operating conditions

6.1. Ability to detect and report precipitation

6.1.1. Skill score assessment

The overall ability of the SUT to detect and report the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores (Section 4.1.2) and presented in Table 4. Scores are presented for both 30 minute and 60 minute assessment intervals. The contingency results (Section 4.1.1) corresponding to these scores are presented in Table 5.

Table 4: Skill scores for EML UPG1000 gauge at Sodankylä during the formal test periods.

Time interval	Probability of Detection, POD (%)	False Alarm Rate, FAR (%)	Bias, B (%)	Heidke Skill Score, HSS (%)
30 min	93.3	67.5	287	45.8
60 min	94.3	44.3	169	67.1

Table 5: Contingency table illustrating detection of precipitation by the EML UPG1000 relative to the specific site reference at Sodankylä, expressed as number of events over the entire test period.

Time interval	Number of Events			
	YY (hits)	YN (misses)	NY (false alarms)	NN (correct negatives)
30 min	542	39	1124	17442
60 min	629	38	501	8385

6.1.2. Characterization of response delays

Response delays were determined using the approach outlined in Section 4.2 and compiled over the entire test period. The delays are represented as a probability density function in Figure 5. This PDF provides an overall picture of response delays for the SUT in all precipitation types. The response delays are characterized further by separating the precipitation events by type/phase, and plotting the delays as a function of the mean precipitation intensity observed by the reference gauge during the delay period (Figure 6).

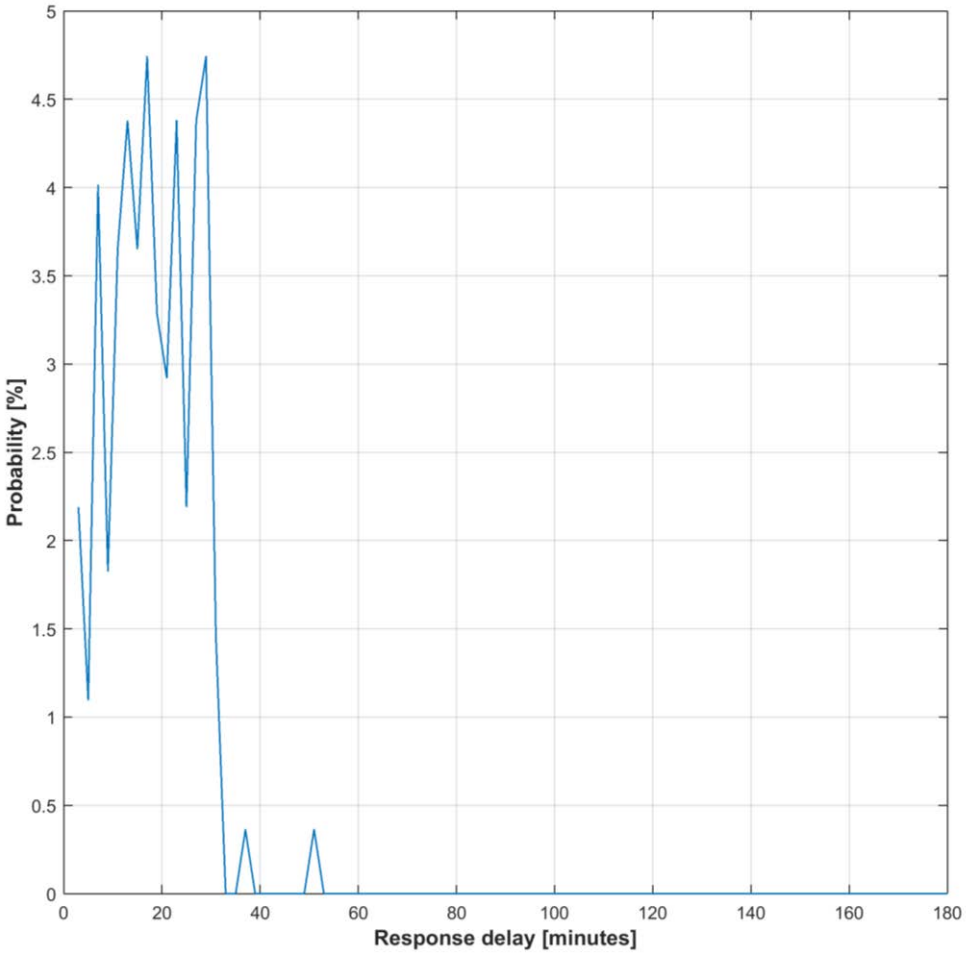


Figure 5: Response delays for heated EML UPG1000 gauge at Sodankylä relative to the R2 reference configuration for all precipitation types. The number of response assessment periods used to determine the delays was 137.

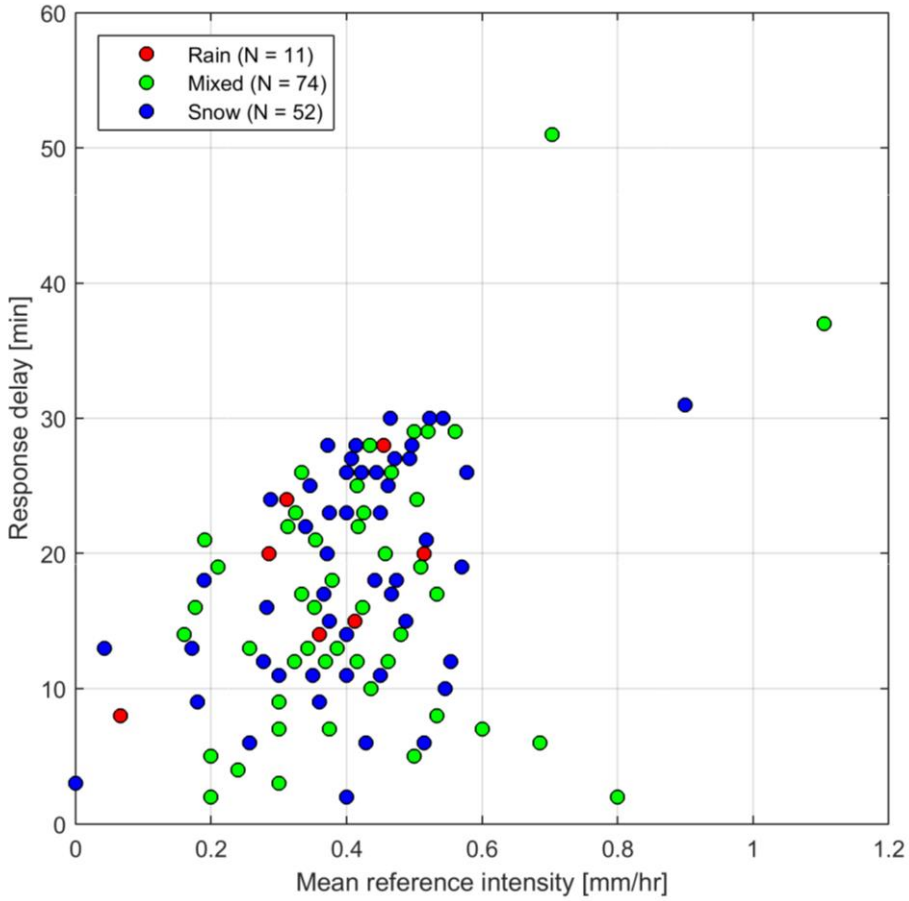


Figure 6: Response delays for EML UPG1000 at Sodankylä relative to R2 reference configuration as a function of the mean precipitation intensity observed by the reference configuration during the delay period. Results are separated by the phase of precipitation.

6.2. Ability to report accumulated precipitation

The SUT performance in terms of reporting accumulated precipitation is examined by comparing the amount reported by each sensor under test relative to the respective site reference during a given assessment interval. This is represented graphically using scatter and box plots of the catch efficiency as a function of mean wind speed at gauge height, as well as scatter plots of the amounts reported by the SUT versus the corresponding reference amounts (Figures 7 to 9). The SUT performance is also assessed in terms of the root mean square error, RMSE (Table 6).

Only assessment intervals during which the SUT and reference both reported precipitation (YY cases) are considered in this portion of the assessment. To assess the influence of the assessment interval on SUT performance, analysis was conducted using both 30 and 60 minute intervals. The plots presented in Figures 7 to 9 include only the 30 minute results, in order to better constrain the wind speed and temperature data used in the assessment, which vary to a greater extent over longer (60 minute) periods.

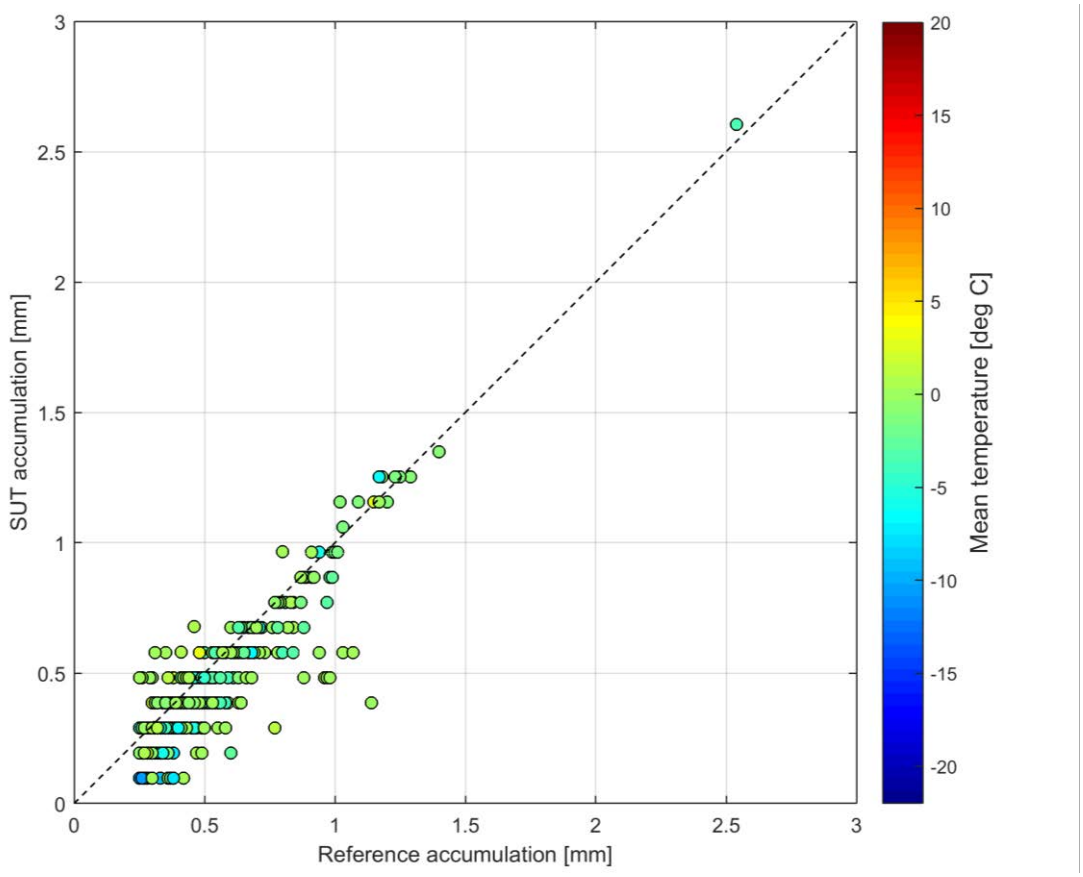


Figure 7: SUT accumulation vs. reference accumulation scatter plots for 30 minute precipitation events at Sodankylä. The mean event temperature is indicated by colour.

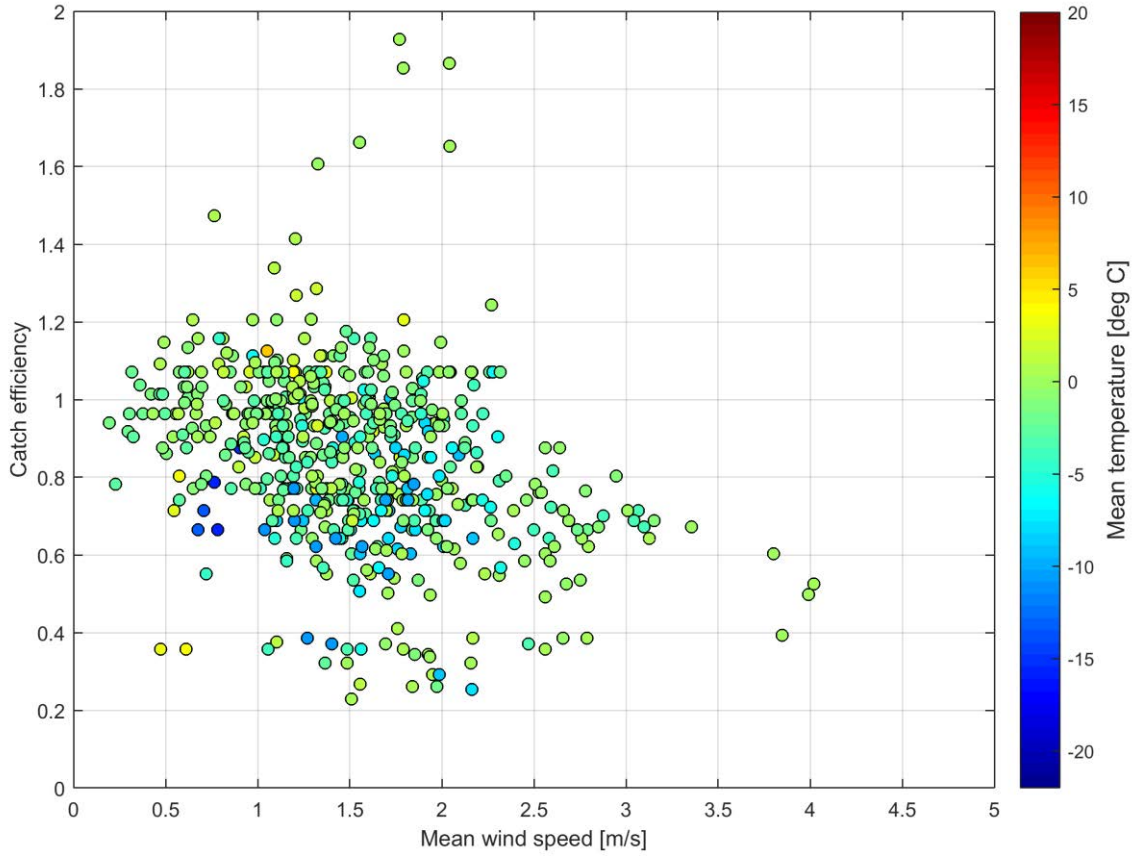


Figure 8: Catch ratio vs. mean wind speed scatter plots for 30 minute precipitation events at Sodankylä. The mean event temperature is indicated by colour.

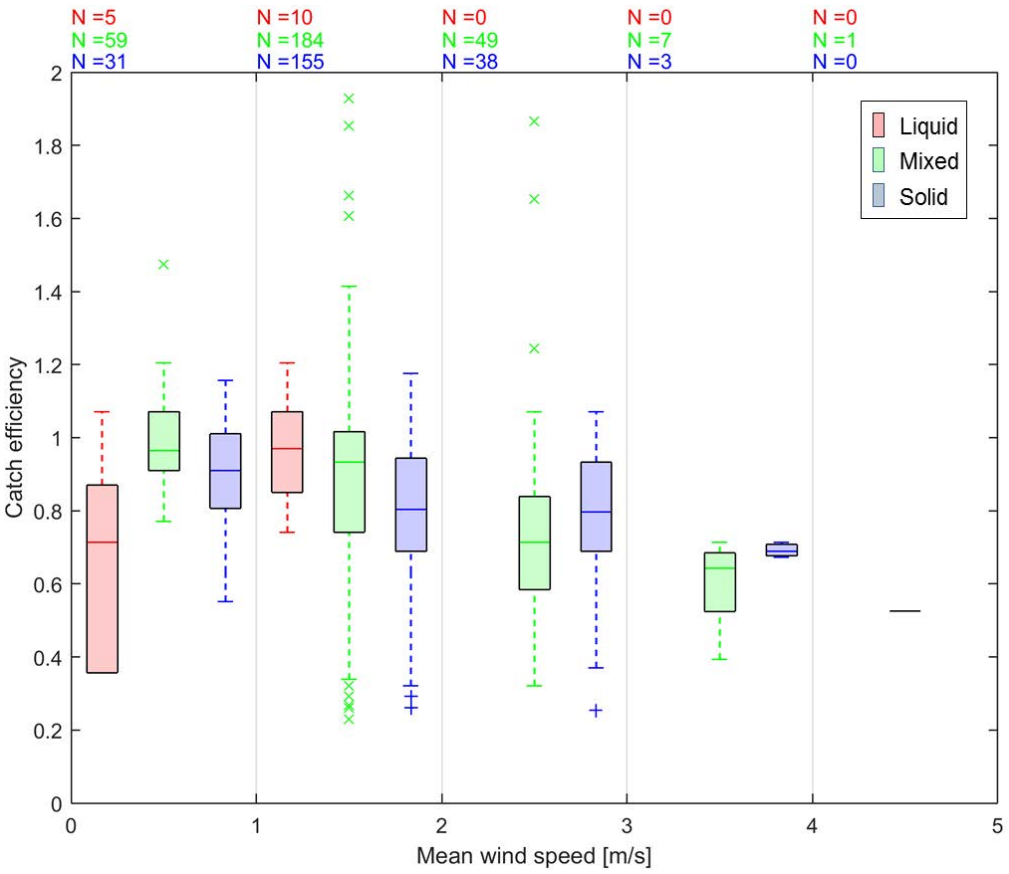


Figure 9: Catch ratio vs. mean wind speed box and whisker plots for 30 minute precipitation events at Sodankylä. The predominant precipitation type for each event is determined from the maximum and minimum reported temperature and indicated by colour.

Table 6: RMSE values by precipitation type for YY cases at Sodankylä.

Time interval	RMSE (mm)			
	Liquid	Mixed	Solid	All precip types
30 min	0.083	0.132	0.099	0.118
60 min	0.088	0.187	0.183	0.183

The total accumulation reported by each site reference and SUT, for all assessment intervals during which both detected and reported precipitation, are presented alongside the corresponding catch ratios in Table 5.

Table 7: SUT and reference accumulation and overall catch ratio, assessed for YY cases over the entire test period at Sodankylä.

Time interval	SUT accumulation	Reference accumulation	Overall catch ratio
30 min	206.1 mm	239.5 mm	0.86
60 min	294.0 mm	359.7 mm	0.82

6.3. Assessment of events when the reference and the SUT do not agree on the occurrence of precipitation

Assessment intervals during which the site reference and SUT do not agree on the occurrence of precipitation – namely, the YN and NY cases (Section 4.1.1) – are characterized using histograms in Figures 10 and 11. The histograms include accumulated precipitation, precipitation intensity as reported by the reference, and corresponding site conditions.

The total SUT and reference accumulations over the test period include contributions from YN and NY cases, which impact the overall catch efficiency. These contributions may be significant, given the response delays associated with heated tipping bucket gauges. Total accumulation and catch efficiency results presented in Table 7 (YY cases only) are expanded to include contributions from YN and NY cases in Table 8.

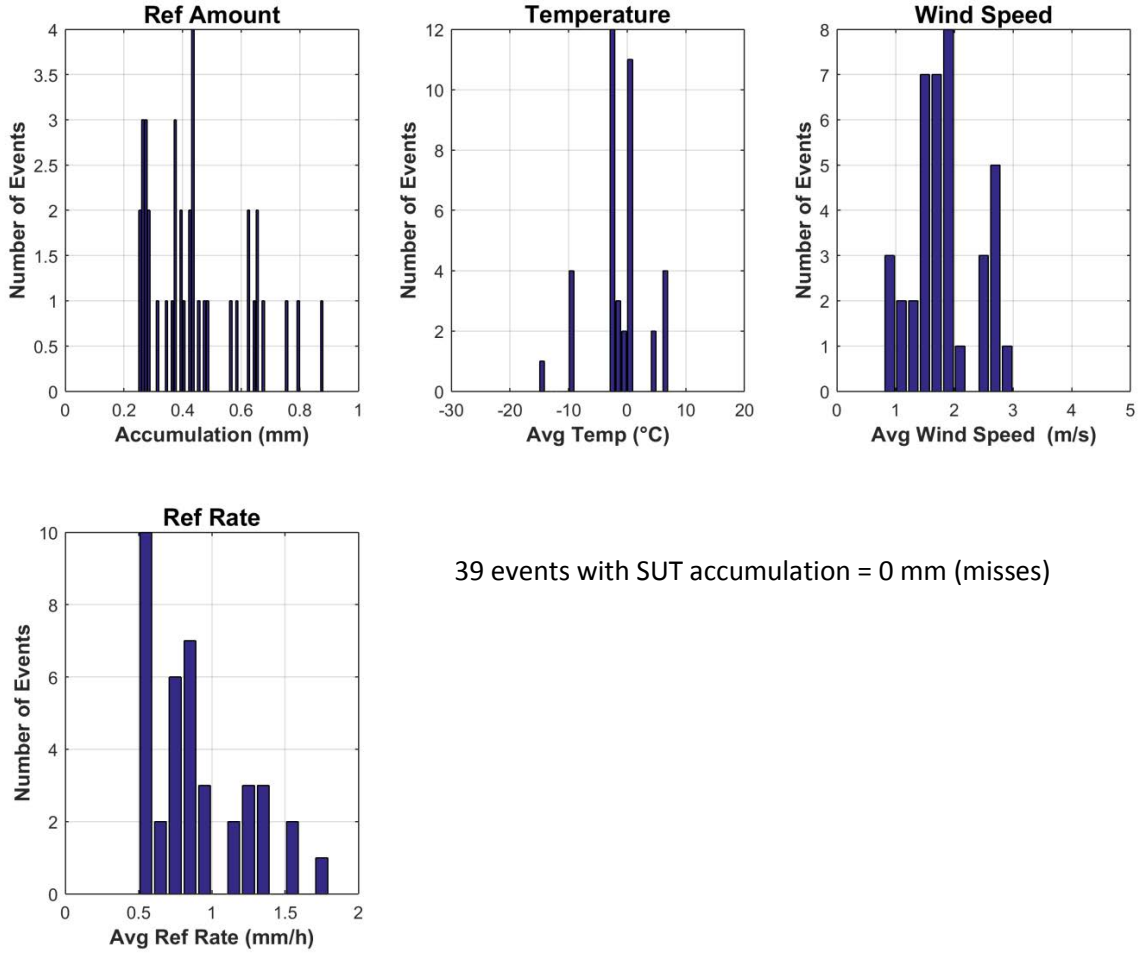


Figure 10: Histograms of reference accumulation, mean temperature, mean wind speed, and reference precipitation rate for YN cases at Sodankylä.

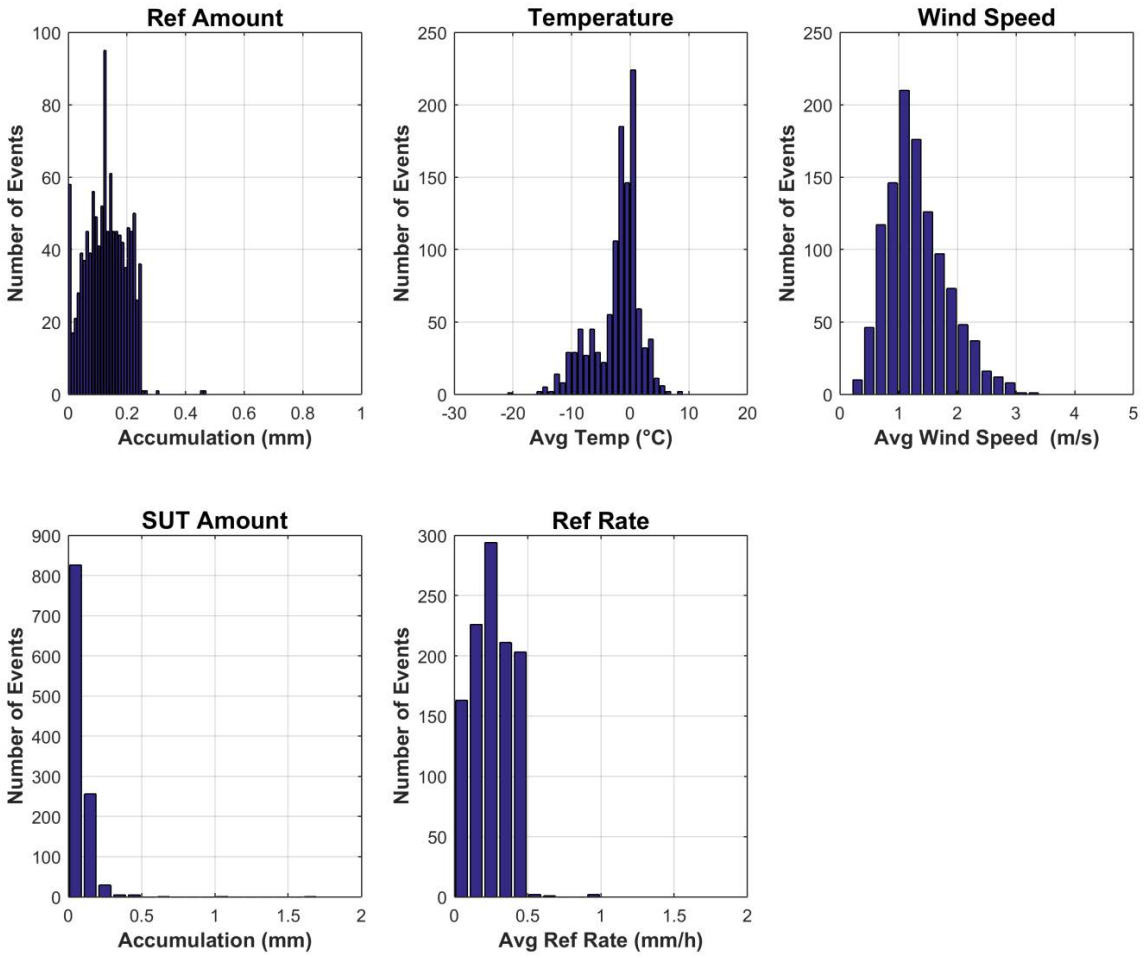


Figure 11: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, and reference precipitation rate for NY cases at Sodankylä.

Table 8: SUT and reference accumulation and overall catch ratio, assessed for the entire test period, for all YY, YN, and NY cases.

Time interval	SUT accumulation	Reference accumulation	Overall catch ratio
30 min	351.3 mm	398.2 mm	0.88
60 min	356.5 mm	443.7 mm	0.80

7. Interpretation of Results

7.1. Operating conditions

The full range of conditions under which the test gauge was operated is illustrated in Figure 3. The conditions relevant to gauge operation are as follows:

- Temperatures between approximately -40 °C and 20 °C;
- Precipitation rates within 6 mm/hr.

These conditions fall within the manufacturer's specified range of operating conditions for the UPG1000.

The conditions during precipitation events (as identified by the reference configuration) are shown in Figure 4. The events at Sodankylä (sheltered, high latitude site) are characterized by low wind speeds, within about 4 m/s, and temperatures close to, or below, freezing.

7.2. Ability to detect and report precipitation

7.2.1. Skill score assessment

The performance of the sensor under test was considered in terms of skill scores, computed for both 30 and 60 minute assessment intervals (Table 4). The Probability of Detection (POD) is slightly higher for the 60 minute intervals (94.3%) than the 30 minute intervals (93.3%). The False Alarm Rate (FAR) and Bias (B) are lower for 60 minute intervals relative to 30 minute intervals, with FAR values of 44.3% (60 min) and 67.5% (30 min) and Biases of 169% (60 min) and 287% (30 min). The Heidke Skill Score (HSS) is 67.1% for 60 minute intervals and 45.8% for 30 minute intervals, indicating greater detection skill of the SUT over the longer intervals.

The contingency results in Table 5 indicate high numbers of false alarm events relative to hits and misses for each assessment interval. The potential for accumulation of snow on the horizontal component of the wind shield (Figure 1b), and subsequent blowing of this snow into the gauge during non-precipitating conditions provides one explanation of the high numbers of false alarm events. Also of note is the fact that the number of false alarm events for 30 minute intervals (1124) is almost double that for 60 minute intervals (501), which is attributed to delays in the gauge response due to the principle of operation and heating (discussed further below).

7.2.2. Characterization of response delays

7.2.2.1. Response delay PDFs

A probability distribution function of response delays between the onset of precipitation as observed by the reference configuration and the first response by the EML UPG1000 heated tipping bucket gauge at Sodankylä is shown in Figure 5. Response times within about 30 minutes have the highest probability, though longer delays (up to about 60 minutes) were also observed.

For precipitation occurring within a given 30 or 60 minute interval (as indicated by the reference configuration), the response delay PDF in Figure 5 indicates the potential for the TB response to occur within a subsequent 30 or 60 minute interval. The delays with highest probability in the PDFs are within 30 minutes, so it is more likely that the TB gauges will respond to precipitation detected by the reference within a 60 minute period than within a 30 minute period. This is reflected by the lower number of false alarm (NY) cases for the 60 minute intervals in Table 5, and the lower FAR and Bias for these intervals in Table 4.

Another important consideration is that even when both the SUT and reference detect precipitation within a given interval, the TB may actually be reporting precipitation collected during previous intervals, or may only report a fraction of the precipitation within that interval (reporting the rest during a subsequent interval). These 'false YY cases' will impact the assessment of TB gauge performance in terms of reporting accumulated precipitation (Section 6.2).

7.2.2.2. Influence of precipitation intensity, phase

The dependence of delay times on the mean intensity of precipitation preceding the TB response, as reported by the reference configuration, is illustrated in Figure 6. For the test gauge at Sodankylä, response times are within 30 minutes for all precipitation types (liquid, mixed, solid) when the mean intensity of precipitation is within 0.6 mm/hr. This suggests that for low intensities, the TB response is independent of precipitation phase, and that heating/melting of solid precipitation does not increase delay times significantly. For mean solid and mixed precipitation intensities exceeding ~ 0.6 mm/hr, however, the range of delay times is extended to about 50 minutes, which can likely be attributed to heating/melting.

7.3. Ability to report accumulated precipitation

7.3.1. Performance when both SUT and reference detect precipitation

7.3.1.1. Temperature, phase, and wind speed influence

Focussing on the 30 minute events in which both the sensors under test and site reference configurations report precipitation (YY cases, or 'hits,' only), no clear temperature trends are observed in Figures 7 and 8, as all events fall within a fairly narrow temperature range (see Figure 4). A general decrease in catch efficiency with increasing wind speed is observed in Figure 8. Significant scatter is observed in the data below 3 m/s, with catch efficiencies < 0.4 and > 1.4.

The phase and wind speed influences on catch efficiency are more clearly illustrated in the box and whisker plots in Figure 9. For solid precipitation, the median catch efficiency is between 80 and 90% for wind speeds ≤ 3 m/s, falling to about 70% for wind speeds between 3 and 4 m/s. For mixed precipitation, the median catch efficiency is > 80% for wind speeds < 2 m/s, falling to values between 60 and 80% for wind speeds between 2 and 4 m/s. Generally, the reduction in catch efficiency with increasing wind speed would be expected to be higher for solid precipitation relative to mixed precipitation; however, the discrimination of precipitation phase by temperature, as performed in the present analysis, is not absolute, so this discrepancy is not considered to be significant.

The majority of outlying points in Figures 7 and 8 are classified as mixed precipitation. Mixed precipitation encompasses all events which are not classified as liquid or solid with a high degree of confidence; these events may be primarily rain or snow, or a transition between the two, depending on the temperature during the event. Accordingly, events in the mixed regime show the highest degree of uncertainty.

7.3.1.2. Root mean square error

The RMSE was calculated for all 30 and 60 minute events during which the reference and SUT both detected precipitation (YY cases) at Sodankylä (Table 6). The duration of the assessment interval does not significantly impact the RMSE values observed for liquid precipitation events; however, the values for mixed and liquid precipitation events are larger for 60 minute intervals relative to 30 minute intervals. These differences are attributed to the fact that mixed and solid events are impacted to a greater extent by response delays and wind-induced undercatch than liquid events. Given the broad use of hourly reports in meteorological operations, further discussion of RMSE values will be considered for 60 minute assessment intervals only.

The overall RMSE (for all precipitation types) is within about 0.18 mm over 60 minutes, suggesting that for the specific site and gauge configuration tested, in a low wind environment, the absolute uncertainty of the EML UPG1000 relative to a DFAR configuration can be considered to be within 0.18 mm over 60 minutes. The phase-specific RMSE values are approximately 0.09 mm/60 minutes for liquid events, 0.19 mm/60 minutes for mixed events, and 0.18 mm/60 minutes for solid events.

7.3.1.3. Overall catch efficiency

The total accumulated precipitation recorded by the reference and sensor under test was compiled for all 30 and 60 minute events during which both detected precipitation (YY cases) and used to calculate the overall catch ratio. The results are provided in Table 7. Focussing on the more operationally-relevant 60 minute value, the overall catch ratio for the gauge under test at Sodankylä is 0.82, indicating that the gauge performs well relative to the reference under low wind conditions.

7.3.2. Characterization of conditions when the SUT and reference do not agree on the occurrence of precipitation

The conditions during 30 minute events in which the reference detected precipitation, but the SUT did not (YN cases), are depicted as histograms in Figure 10. These events are characterized by low precipitation rates, typically below 2 mm/hr, and temperatures close to, or below, freezing. At these temperatures, the precipitation is likely mixed or solid, and hence subject to longer response delays. Hence, these may be cases in which the TB response to precipitation occurs during a subsequent 30 minute period. There is also potential for heating to cause evaporation or sublimation of incident precipitation at low intensities.

The conditions during 30 minute events in which the SUT detected precipitation, but the reference did not (NY cases), are shown as histograms in Figure 11. These events are also characterized by low precipitation rates, typically below 0.5 mm/hr (which corresponds to the reference accumulation threshold for a precipitation event over a 30 minute interval), and wind speeds below about 4 m/s.

These events may result, in part, from delayed tipping bucket responses that coincide with intervals during which the reference did not detect precipitation. Another explanation for the large number of NY events is the blowing of snow collected on the horizontal wind shield components into the gauge.

Given the influence response delays, estimates of the overall catch ratio for tipping bucket gauges should include the reference and SUT accumulations during YN and NY cases (and not just during the YY cases, as considered above). The resulting total accumulations for all YY, YN, and NY cases, and the corresponding overall catch efficiencies, are provided in Table 8. The overall catch efficiency for the gauge under test at Sodankylä was 0.80 for 60 minute assessment intervals. This value is comparable to that for YY cases only (overall catch efficiency of 0.82 for 60 minute assessment intervals in Table 7), suggesting that the total reference and SUT accumulation during YN and NY cases, respectively, are similar.

8. Operational considerations

The overall experience with the EML UPG1000 heated tipping bucket gauge at Sodankylä was positive. The gauge was easy to install and maintain, although consideration must be given to its large size and weight.

The performance of the gauge at Marshall was compromised by stiffness in the tipping mechanism during winter conditions that prevented the gauge from tipping and recording precipitation. The site team contacted the manufacturer and were able to correct the issue, but not in time for the data to be considered as part of this assessment.

8.1. Maintenance

Each site completed the gauge field calibration and verification as per manufacturer recommendations, at least once a year. The calibration records have been stored by each site team.

The gauge at Marshall required intervention by the manufacturer to address the issue of stiffness in the tipping mechanism, as noted above.

8.2. Noted issues

- The site team at Sodankylä noted the accumulation of snow on the horizontal wind shield components; the blowing of this snow into the gauge is believed to cause false reports of precipitation by the gauge.

- Stiffness in the tipping mechanism can occur under cold conditions, which can inhibit gauge operation.

8.3. Recommendations for improving the gauge or measurement

- A shield configuration with vertical slats, like that for the gauge installed at Marshall (Figure 1a), is recommended to mitigate the effects of snow accumulating on the shield.

9. Performance Considerations

- The shielded gauge has been demonstrated to perform well under low wind conditions; however, an L-shaped shield configuration with horizontal components at the top can accumulate snow, leading to the potential for blowing snow and false precipitation reports by the gauge. A shield configuration with vertical slats is now recommended by the manufacturer.

- Given the noted potential for stiffness of the tipping mechanism, the gauge should be operated alongside another gauge or a sensitive precipitation detector for comparison and validation of reports prior to its implementation in operational settings.

- Heating is critical to ensure proper functioning of the gauge under winter conditions. The voltage should be checked at the gauge location to ensure that the heaters are supplied with the recommended power. Voltage drops – for example, resulting from the use of long power cables – can impair the functioning of the gauge in winter conditions.

- Heating enables the measurement of solid precipitation, but the time required for melting can delay the time between the collection of precipitation in the funnel and the gauge response to that precipitation (response delays). Further, heating may result in the evaporation/sublimation of incident precipitation at low intensities.

- Response delays must be considered when using the gauge in operational settings. Ideally, the reporting interval (i.e. hourly observations) should exceed the maximum expected response delay; however, the potential remains for carry-over of precipitation accumulation from previous intervals, and for delayed responses occurring in subsequent intervals.

- The collection of ancillary measurement data is recommended: air temperature and wind speed (at gauge height) to enable the application of adjustments to measurements (i.e. using transfer functions) and reports from a precipitation detector with high sensitivity, to enable the identification of missed events or false alarms due to response delays or blowing snow.

- While the observed influence of wind speed on gauge catch efficiency was not significant at Sodankylä, in general, the application of transfer functions to gauge measurements (if available) is recommended to account for wind-induced undercatch. This recommendation comes with the caveat that the catch ratio data used in the derivation of the transfer functions will be impacted by response delays. Response delays can contribute to scatter in the catch ratio data, and hence, to uncertainty in the transfer functions; however, the potential remains for the improvement of data via adjustment. Considering the catch ratios over longer time periods (e.g. 1 hour) may provide a means of reducing scatter in the catch ratio/wind speed relationship and reducing the uncertainty of related transfer functions.

WMO-SPICE Instrument Performance Report
HSA TBH

1) Technical specifications (from manufacturer provided documentation)

Instrument model:	HSA TBH
Physical principle:	Heated tipping bucket (TB) gauge. Tips are recorded when the bucket fills to capacity, passing the Dual Reed Switch assembly and producing a contact closure signal.
Bucket capacity:	0.2 mm
Collecting area:	314.15 cm^2
Heating configuration:	Bucket and funnel (70 W). Snow sensor in funnel activated when ambient temperature falls below 4 °C. Heaters turned on when snow sensor detects snow for 5 seconds, continuously. Funnel maintained at 10 °C until last snow is detected, then heater is cycled for 18 minutes to melt any residual snow in funnel. In absence of snow, heaters switched off when ambient temperature falls below -20 °C or above 4.5 °C.
Operating temperature range:	-40 °C to 70 °C
Measurement range:	0 – 700 mm/hr
Measurement uncertainty:	± 2% from 0 – 250 mm/hr, ± 3% from 250 – 500 mm/hr
Sensitivity:	0.2 mm reporting resolution (corresponding to bucket amount)

a)

b)

c)

Figure 1: HSA TBH installations at (a) CARE; (b) Marshall, east; and (c) Marshall, west.

2) Data output format

Gauge data output: Gauge can report precipitation intensity, precipitation amount, and/or time of tips. For intensity reports, manufacturer notes maximum delay times of 18 minutes for solid precipitation.

Analysis is based on accumulated precipitation, computed as the cumulative sum of accumulations derived from intensity reports.

3) SPICE test configuration

Shield: Unshielded

Test Sites: CARE (Canada); Marshall (USA);

Sensor Provider(s): Hydrological Services America

Figure 2: Map of SPICE sites where HSA TBH gauges were tested.

A summary of the configuration of instruments as tested, the duration of tests and availability of data reflected in these results, and the ancillary measurements used, by site, is available in Tables 1, 2, and 3, respectively.

Table 1: Summary of gauge configurations and data output, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	CARE	Marshall (East, West gauges)
Field configuration	Unshielded	Unshielded
Height of installation (gauge rim)	1.5 m	1.97 m, 1.9 m
SUT data output frequency	1 min	6 s , 1 min (both gauges)
Data QC	SPICE QC methodology	
Data temporal resolution	1 min	
Processing interval for SPICE data analysis	30 min, 60 min	

Table 2: Data availability, by measurement season and site.

Measurement season	CARE	Marshall
Season 1 (Oct. 2013 – Apr. 2014)	✓	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓	✓

Table 3: Summary of reference and ancillary measurements, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	CARE	**Marshall**
R2 site reference	Geonor T-200B3 600 mm (DFAR)	Geonor T-200B3 600 mm (DFAR)
R2 precip detector	Thies LPM (DFAR)	Thies LPM (Site*)
Ancillary temp sensor	Vaisala HMP155 (Stevenson screen)	MetOne, model 060A-2/062, 2144-L
Ancillary RH sensor	Vaisala HMP155 (Stevenson screen)	Campbell Scientific CS500
Ancillary wind sensor	Vaisala NWS 425 (2 m)	RM Young Wind Monitor 05103 (2 m)

*A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR shield; however, in cases where a more sensitive detector is available outside of the DFIR shield, or there are issues with the detector within the DFIR shield, a precipitation detector elsewhere on the site can be employed.

4) Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6.1 of the WMO-SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute assessment intervals:

- Ref 'Yes' : R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No' : R2 weighing gauge < 0.25 mm AND/OR precip detector recording < 18 min of precip;
- SUT 'Yes' : SUT accumulation > 0 mm;
- SUT 'No' : SUT accumulation = 0 mm.

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores.

4.1.2. Skill score assessment

The ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

The above scores are computed using the formulations provided in Section 3.6.1.3 of the WMO-SPICE Final Report.

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

4.2. Sensor-specific considerations: response delays

Tipping bucket gauges require that an amount of precipitation corresponding to the bucket capacity is accumulated before a tip is triggered and the gauge records precipitation. This can result in response delays relative to the reference configuration. Heated tipping bucket gauges are subject to further delays, as any solid precipitation in the funnel must be melted before reaching the bucket and potentially triggering a tip, and the heating itself can potentially evaporate incident precipitation. These response delays will impact the comparison with the reference. For this reason, the assessment of TB gauge performance relative to the reference configuration is also considered over 60 minute intervals, using the same conditions and thresholds outlined in above in Section 4.1.1.

Response delays are quantified by determining the time elapsed between the onset of precipitation as determined by the reference configuration, and the first tip recorded by the TB. The assessment is based on periods with at least 30 minutes of precipitation, as identified by the reference configuration, followed by at least 180 minutes without precipitation. This extended period without precipitation is intended to allow additional time for the melting and recording of precipitation by heated TB gauges. Additional details are provided in Section 3.6.1.4.4 of the WMO-SPICE Final Report.

5) Environmental conditions

The environmental conditions at each site over the duration of the test period are expressed as probability density functions (PDFs) of mean air temperature, mean relative humidity, mean wind speed, vector mean wind direction, and precipitation rate for each component 30 minute assessment interval in Figure 3. Figure 4 presents the same parameters for all assessment intervals during which the site reference configuration detected precipitation (i.e. all Ref 'Yes' cases).

The precipitation percentage represents the number of minutes of precipitation during a 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration, expressed as a percentage. PDFs of precipitation percentage are also included in Figures 3 and 4.

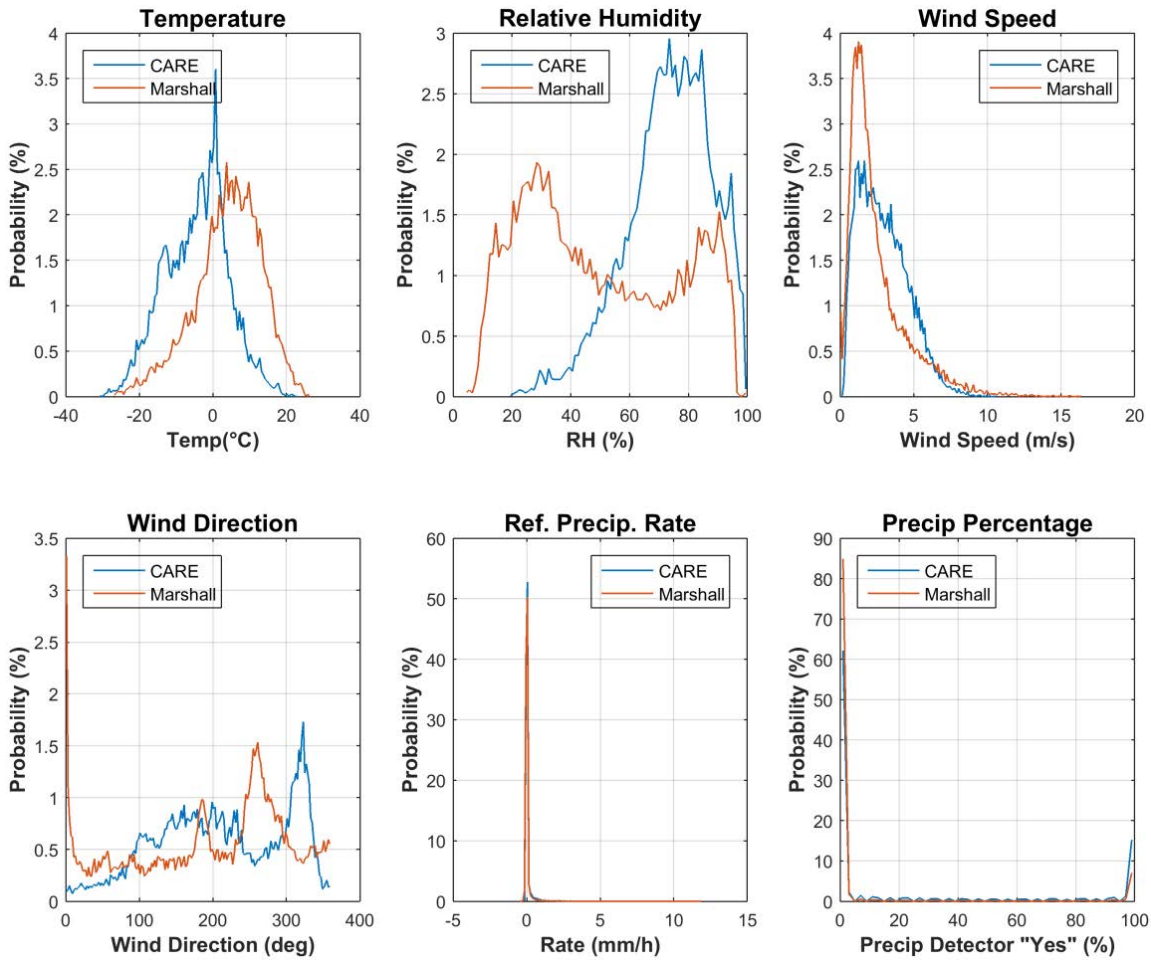


Figure 3: Summary of aggregated environmental conditions at the SPICE sites that operated HSA TBH gauges, over the entire duration of formal tests, as per Table 2.

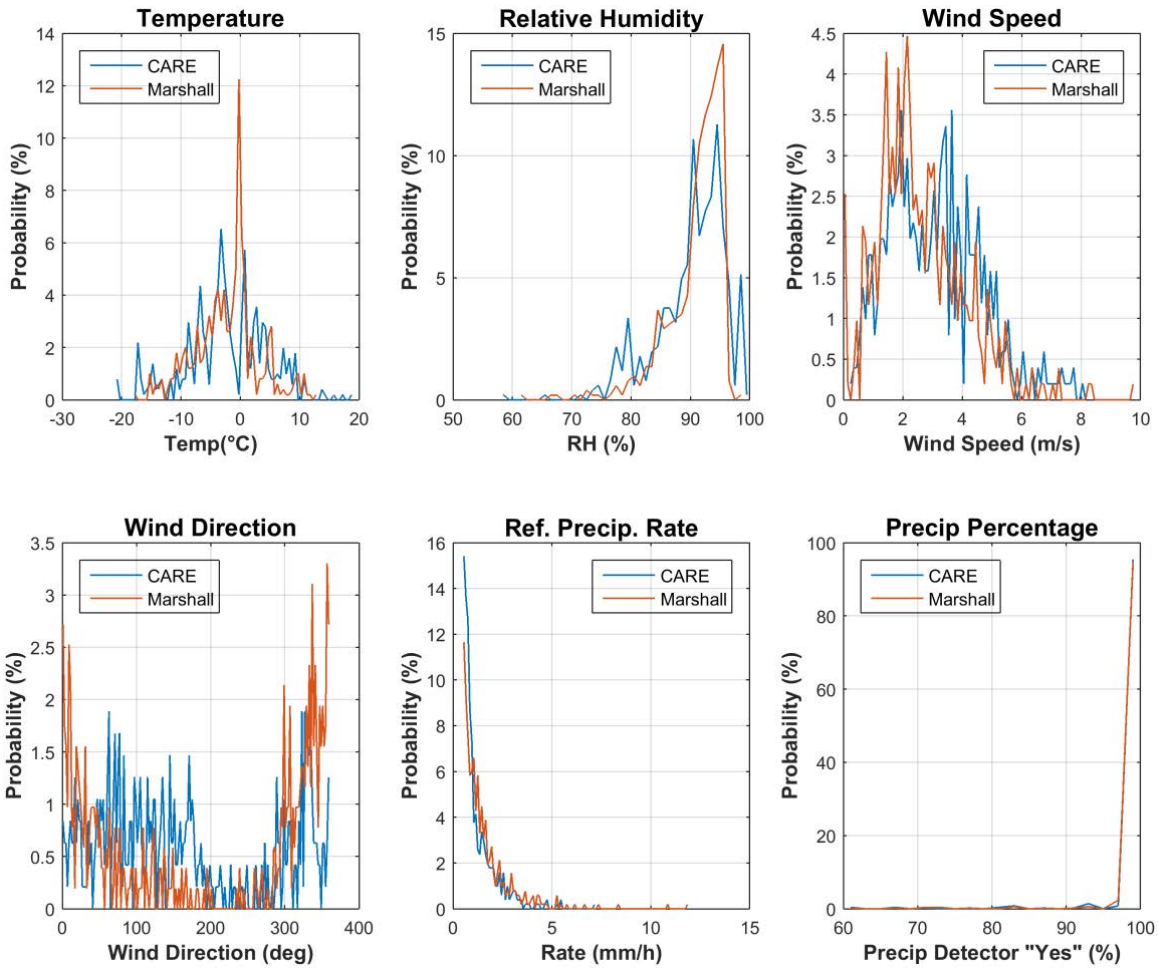


Figure 4: Summary of the aggregated environmental conditions at the SPICE sites that tested HSA TBH gauges during precipitation events, as reported by the site R2 reference, during the formal tests, as per Table 2.

6) Evaluation of performance over the range of operating conditions

6.1. Ability to detect and report precipitation

6.1.1. Skill score assessment

The overall ability of the SUT to detect and report the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores (Section 4.1.2) and presented in Figure 5. Scores are presented for both 30 minute and 60 minute assessment intervals. The contingency results (Section 4.1.1) corresponding to these scores are presented in Table 4.

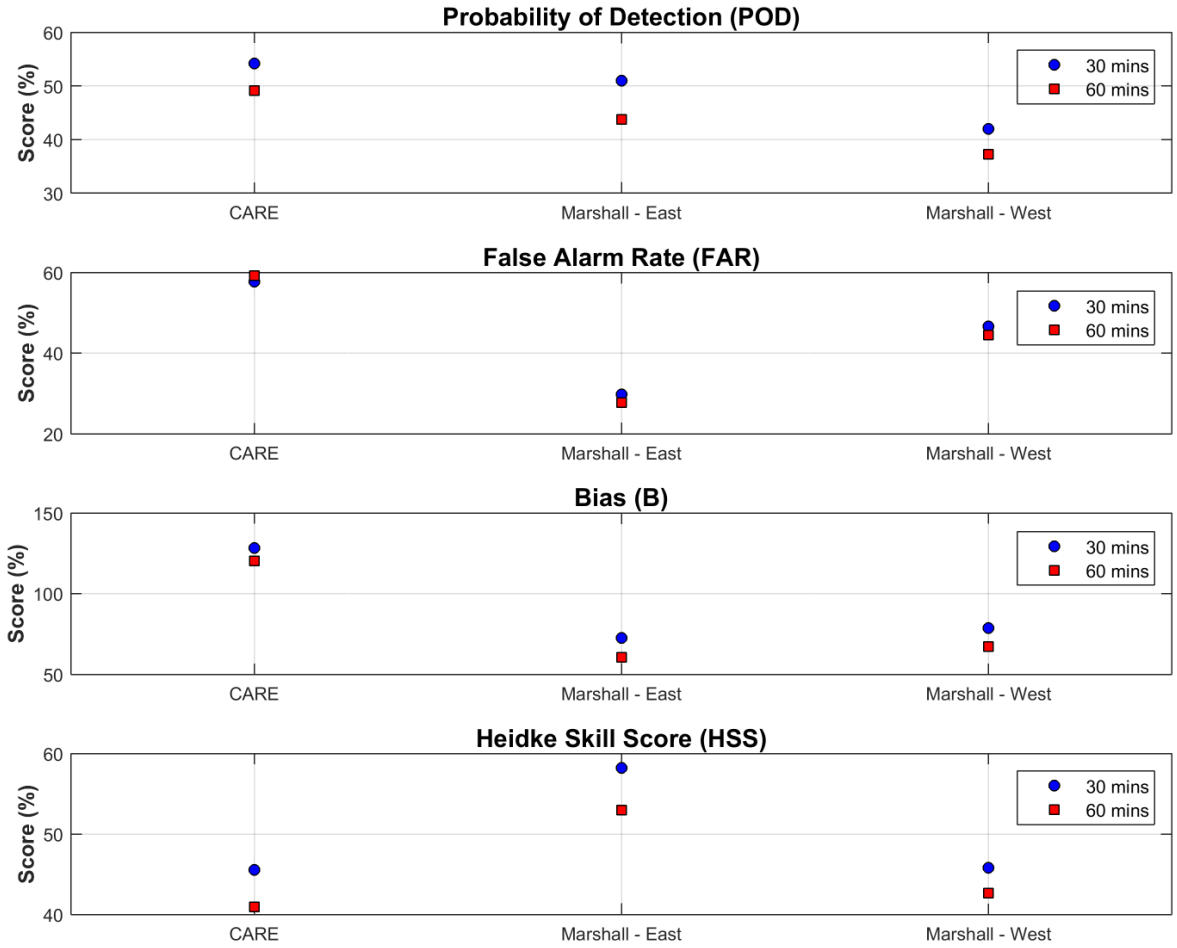


Figure 5: Skill scores for HSA TBH gauges during the formal test periods.

Table 4: Contingency table illustrating detection of precipitation by the HSA TBH relative to the specific site reference, expressed as number (and percentage) of events over the entire test period, by site.

Site	Time interval	Number of Events (% of Events)			
		YY, hits	YN, misses	NY, false alarms	NN, correct negatives
CARE	30 min	245 (1.7%)	207 (1.4%)	335 (2.3%)	13559 (94.5%)
	60 min	200 (2.8%)	207 (2.9%)	290 (4.1%)	6453 (90.3%)
Marshall (East)	30 min	255 (1.3%)	245 (1.2%)	108 (0.5%)	19069 (96.9%)
	60 min	182 (1.9%)	234 (2.4%)	70 (0.7%)	9336 (95.1%)
Marshall (West)	30 min	210 (1.1%)	290 (1.5%)	183 (0.9%)	18997 (96.5%)
	60 min	155 (1.6%)	261 (2.7%)	124 (1.3%)	9283 (94.5%)

6.1.2. Characterization of response delays

Response delays were determined for the test gauges at each site using the approach outlined in Section 4.2, and compiled over the entire test period. The delays are represented as probability density functions in Figure 6. These PDFs provide an overall picture of response delays for each sensor in all precipitation types. The response delays are characterized further by separating the precipitation events by type/phase, and plotting the response delays as a function of the mean precipitation intensity observed by the reference gauge during the delay period (Figure 7).

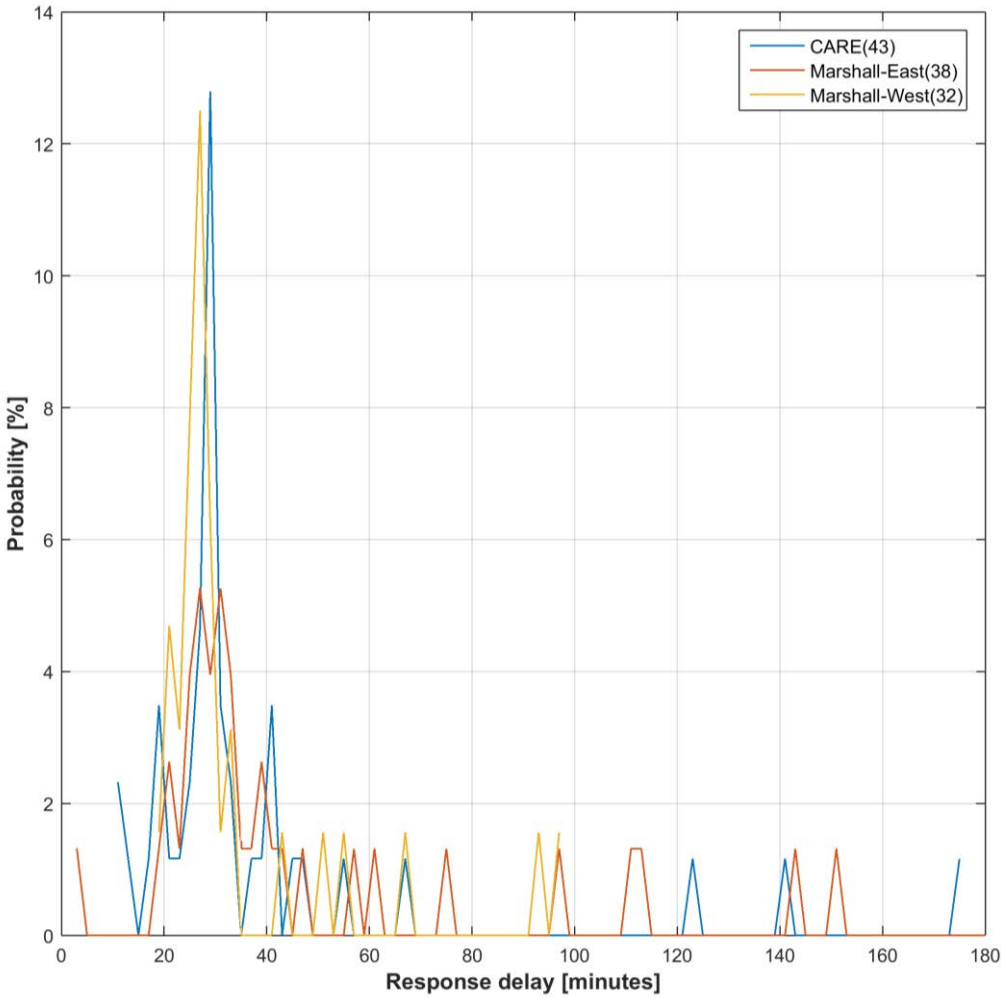
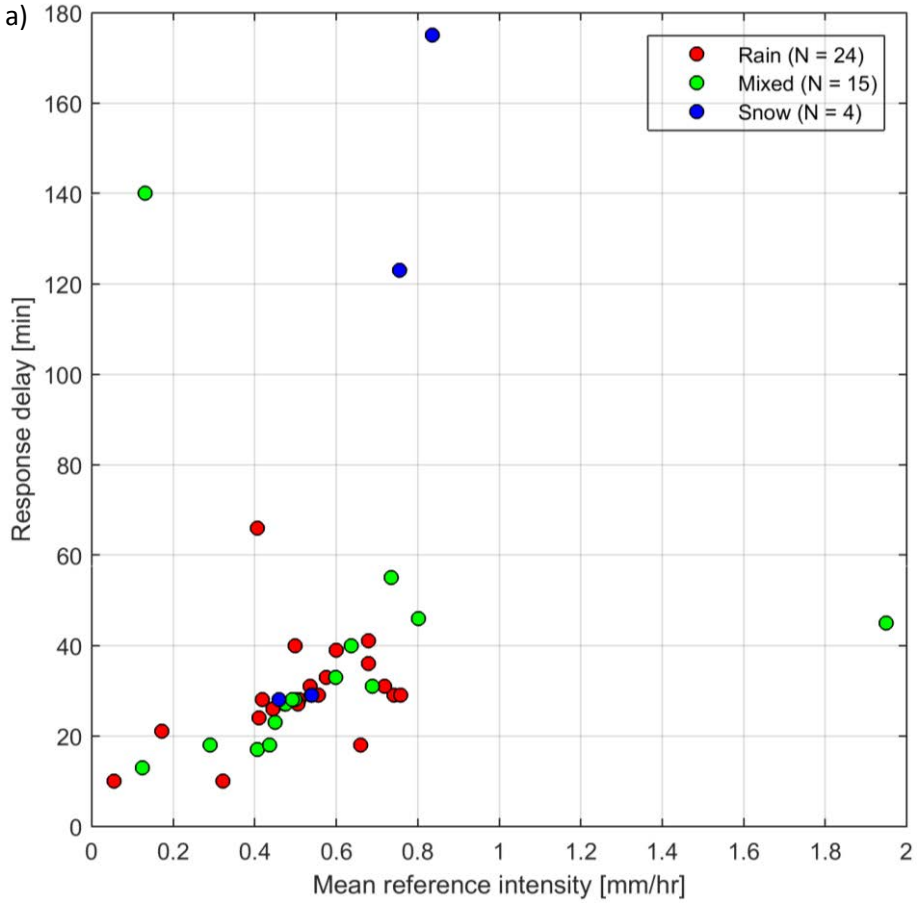
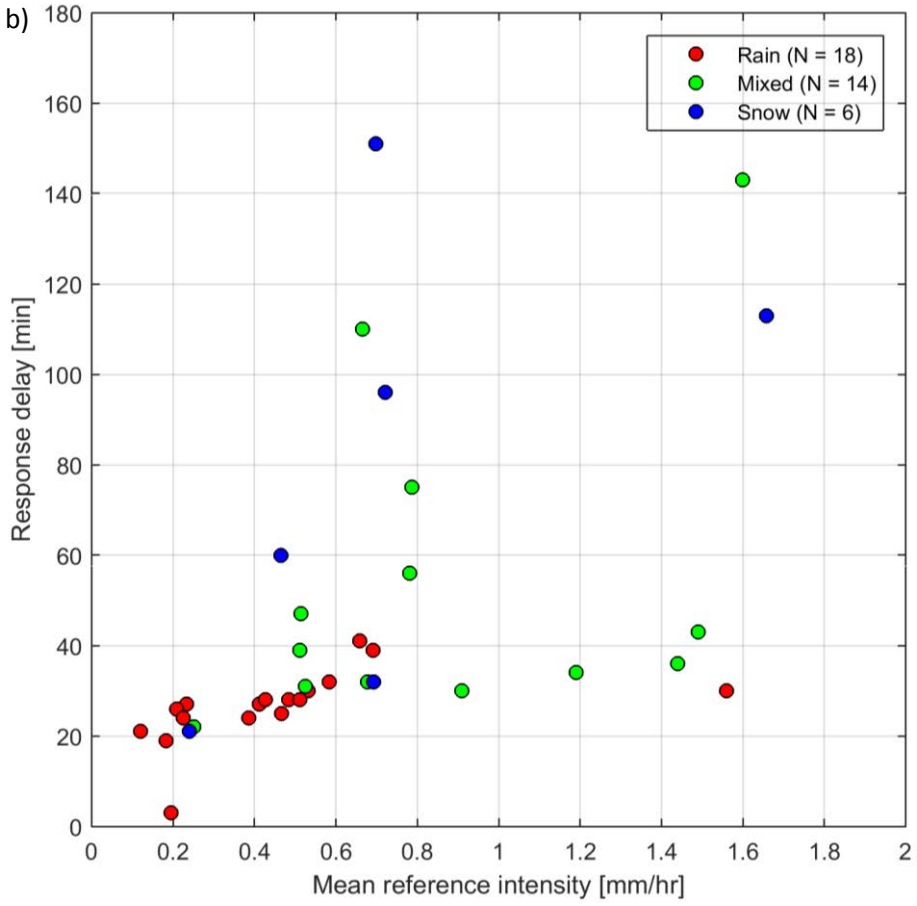


Figure 6: Response delays for heated HSA TBH gauges at CARE and Marshall relative to the R2 reference configuration at each site. The number of response assessment periods used to determine the delays for gauges at each site are indicated in parentheses in the legend.

a)


Rain (N = 24)
Mixed (N = 15)
Snow (N = 4)
Response delay [min]
Mean reference intensity [mm/hr]
0
20
40
60
80
100
120
140
160
180
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

b)


Rain (N = 18)
Mixed (N = 14)
Snow (N = 6)
Response delay [min]
Mean reference intensity [mm/hr]
0
20
40
60
80
100
120
140
160
180
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

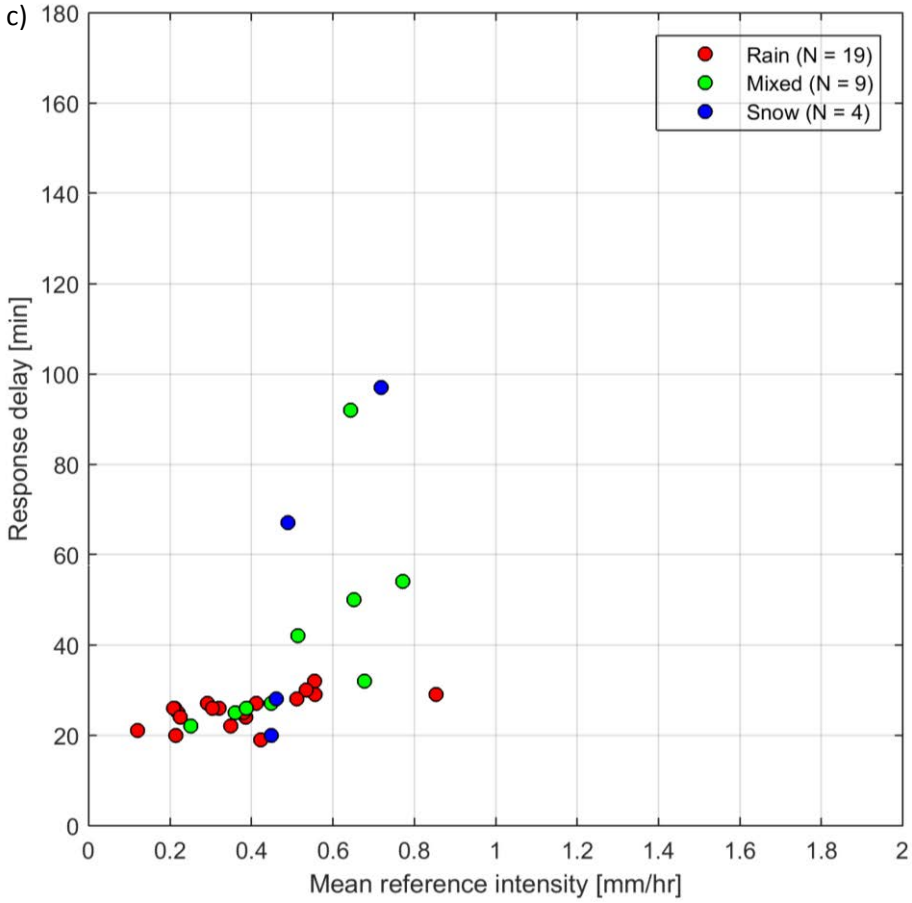


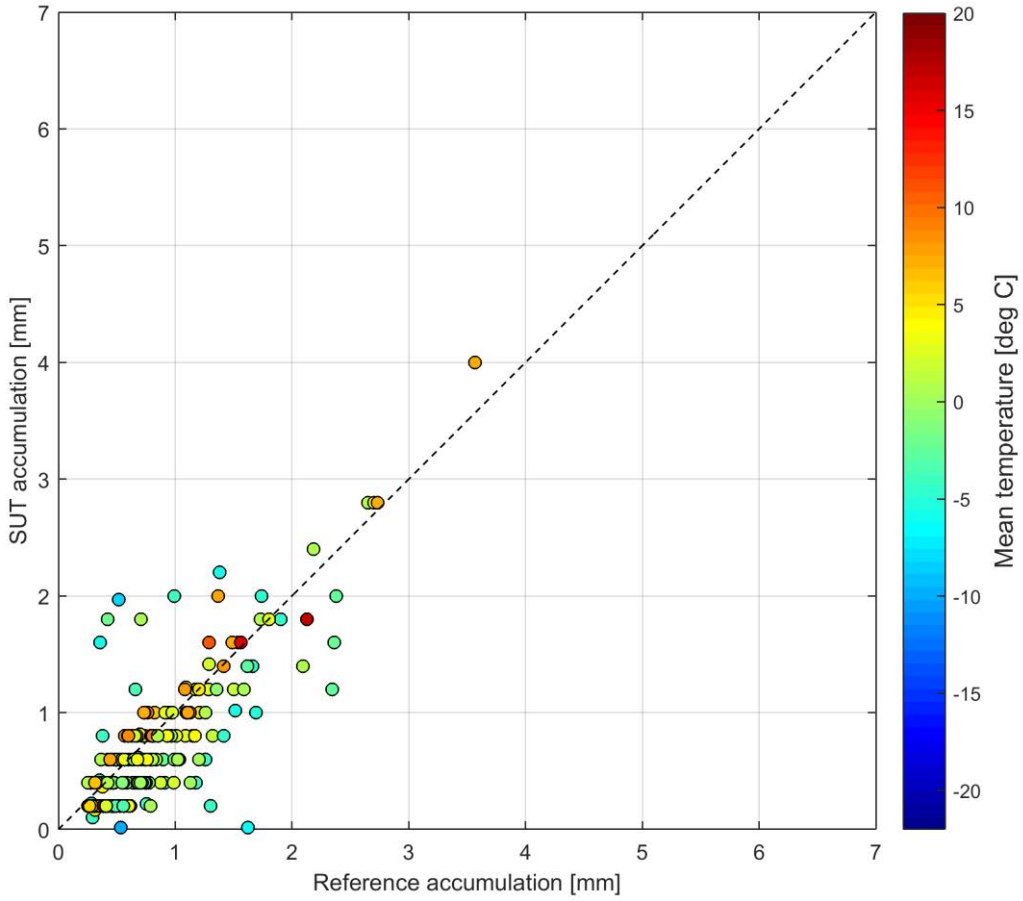
Figure 7: Response delays for HSA TBH gauges as a function of the mean precipitation intensity observed by the reference gauge during the delay period at (a) CARE, (b) Marshall, east gauge, and (c) Marshall, west gauge. The predominant precipitation type for each event is determined from the maximum and minimum reported temperature during the delay period.

6.2. Ability to report accumulated precipitation

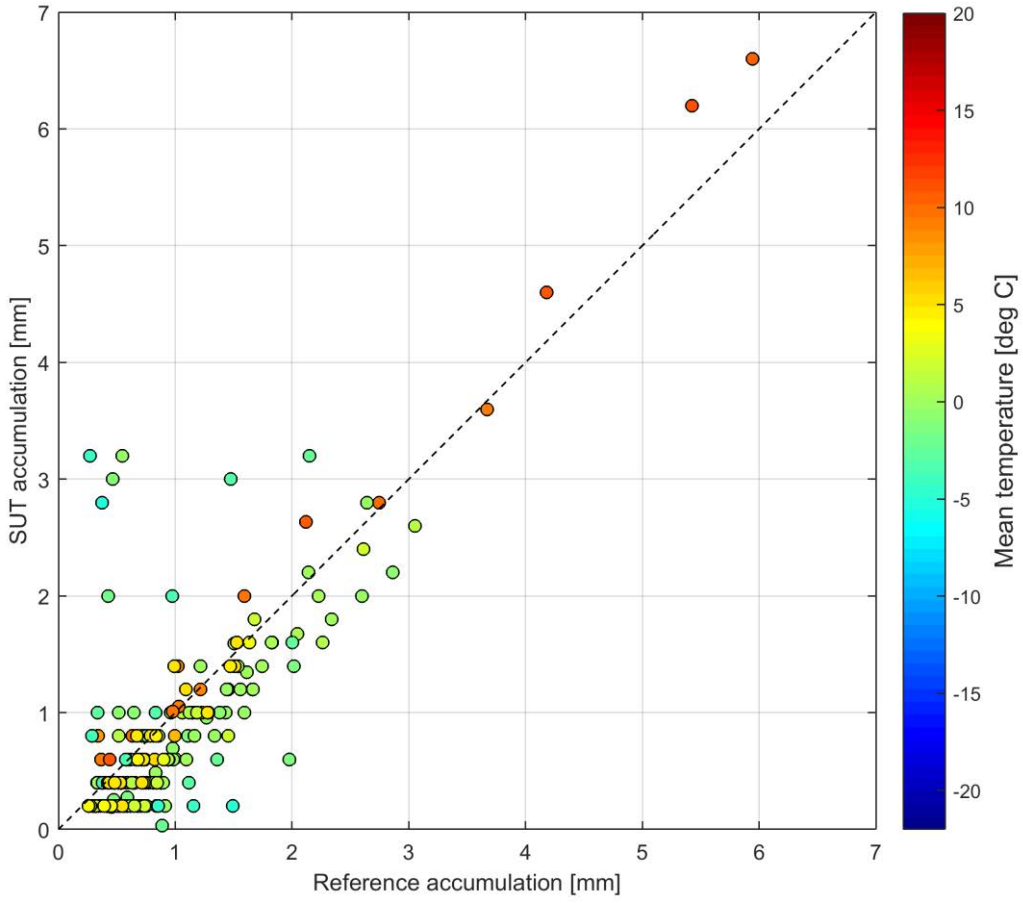
The SUT performance in terms of reporting accumulated precipitation is examined by comparing the amount reported by each sensor under test relative to the respective site reference during a given assessment interval. This is represented graphically using scatter and box plots of the catch efficiency as a function of mean wind speed at gauge height, as well as scatter plots of the amounts reported by the SUT versus the corresponding reference amounts (Figures 8 to 10). The SUT performance is also assessed in terms of the root mean square error, RMSE (Figure 11).

Only assessment intervals during which the SUT and reference both reported precipitation (YY cases) are considered in this portion of the assessment. To assess the influence of the assessment interval on SUT performance, analysis was conducted using both 30 and 60 minute intervals. The plots presented in Figures 8 to 10 include only the 30 minute results, in order to better constrain the wind speed and temperature data used in the assessment, which vary to a greater extent over longer (60 minute) periods.

a)


SUT accumulation [mm]
Reference accumulation [mm]
Mean temperature [deg C]
0
1
2
3
4
5
6
7
20
15
10
5
0
-5
-10
-15
-20

b)


SUT accumulation [mm]
Reference accumulation [mm]
Mean temperature [deg C]
0
1
2
3
4
5
6
7
20
15
10
5
0
-5
-10
-15
-20

c)

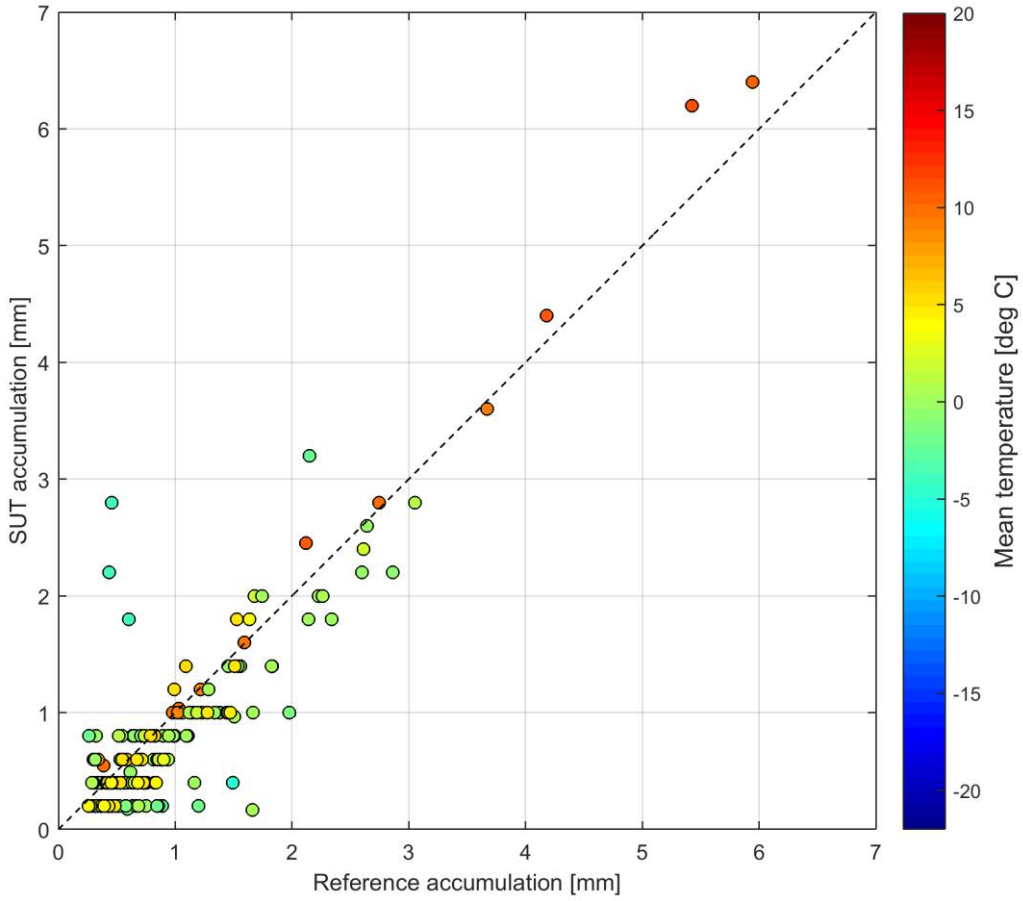
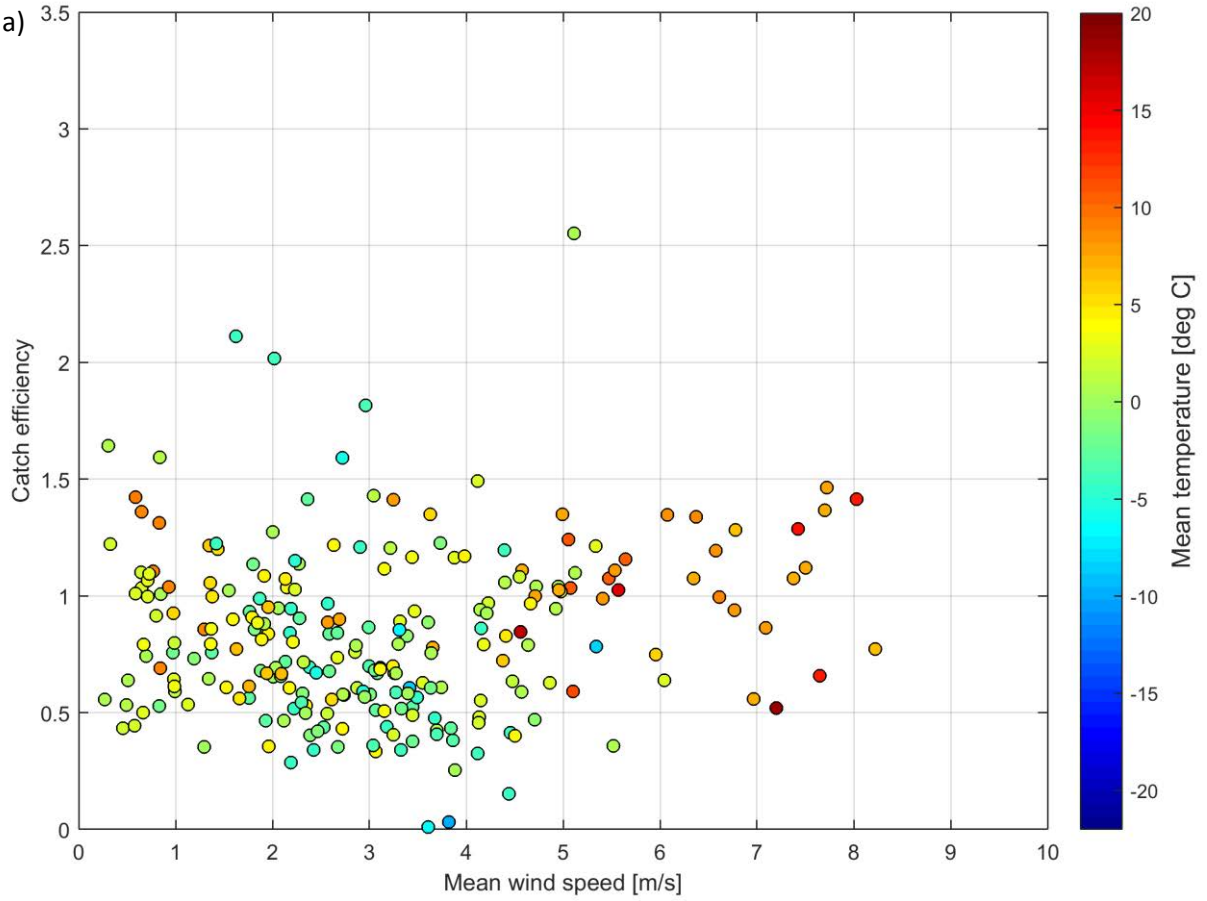
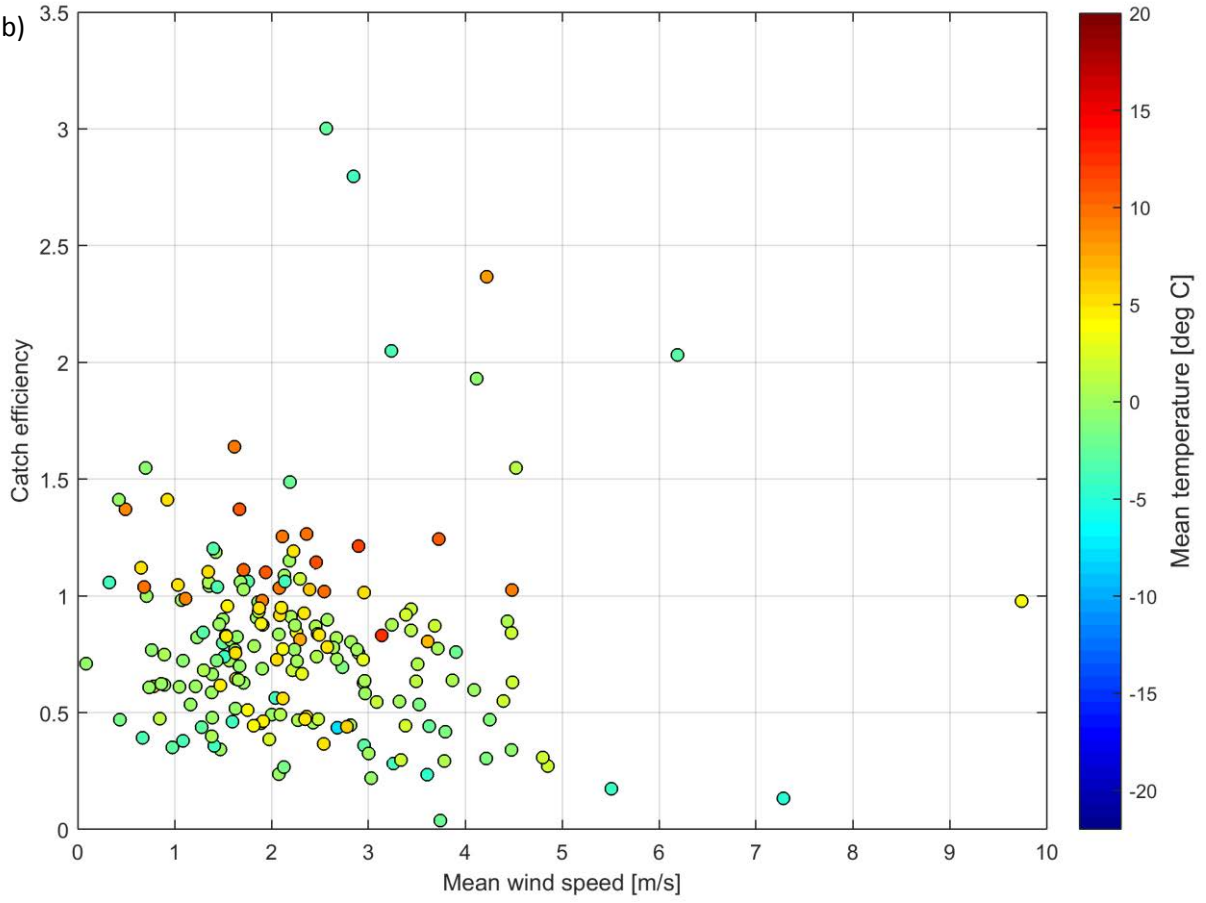


Figure 8: SUT accumulation vs. reference accumulation scatter plots for 30 minute precipitation events at (a) CARE, (b) Marshall, east gauge, and (c) Marshall, west gauge. The mean event temperature is indicated by colour.

a)


Catch efficiency
Mean wind speed [m/s]
Mean temperature [deg C]
0
0.5
1
1.5
2
2.5
3
3.5
0
1
2
3
4
5
6
7
8
9
10
20
15
10
5
0
-5
-10
-15
-20

b)


Catch efficiency
Mean wind speed [m/s]
Mean temperature [deg C]
0
0.5
1
1.5
2
2.5
3
3.5
0
1
2
3
4
5
6
7
8
9
10
20
15
10
5
0
-5
-10
-15
-20

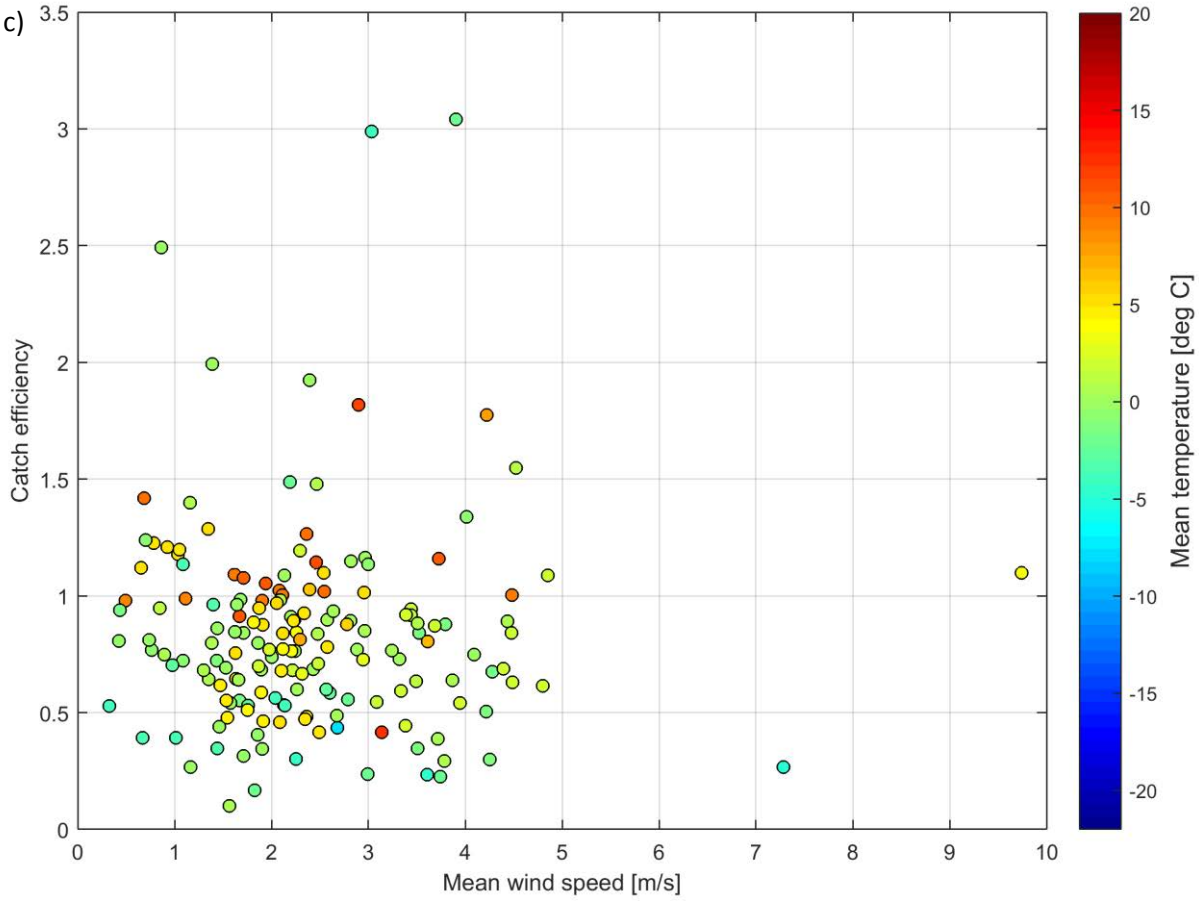
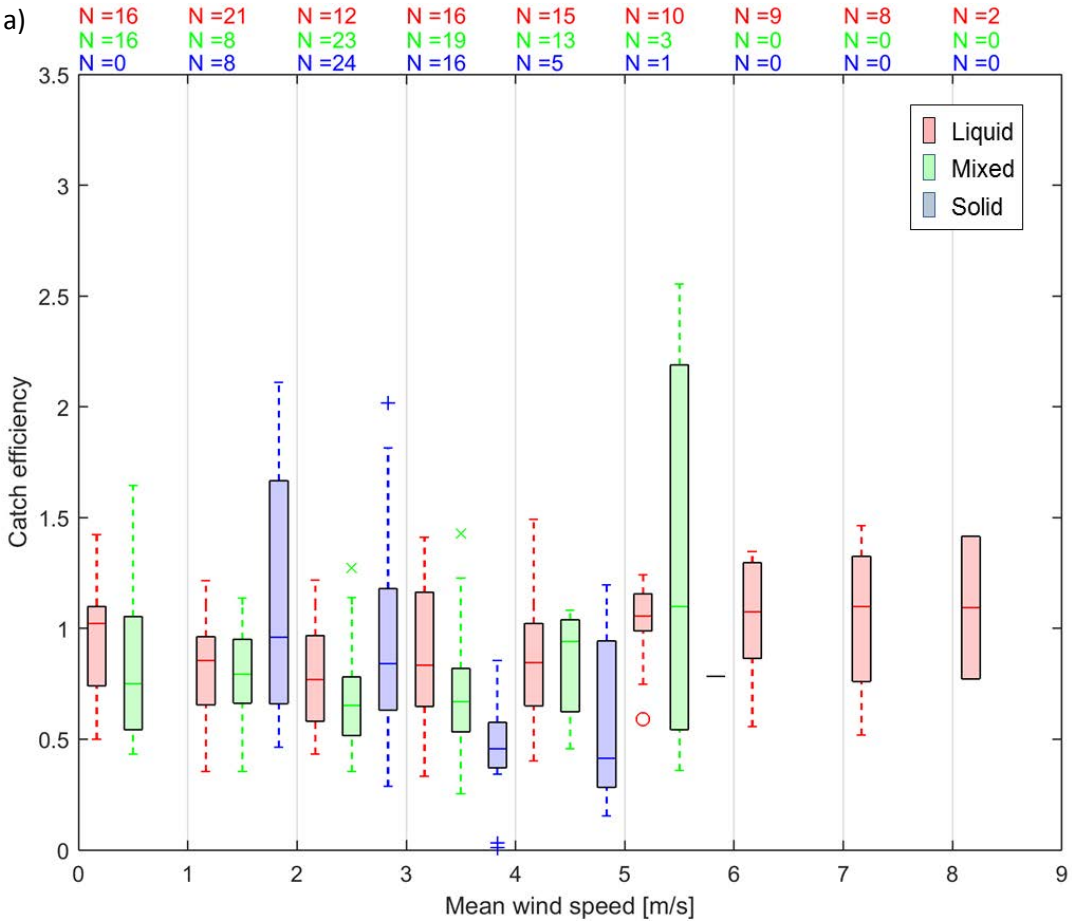
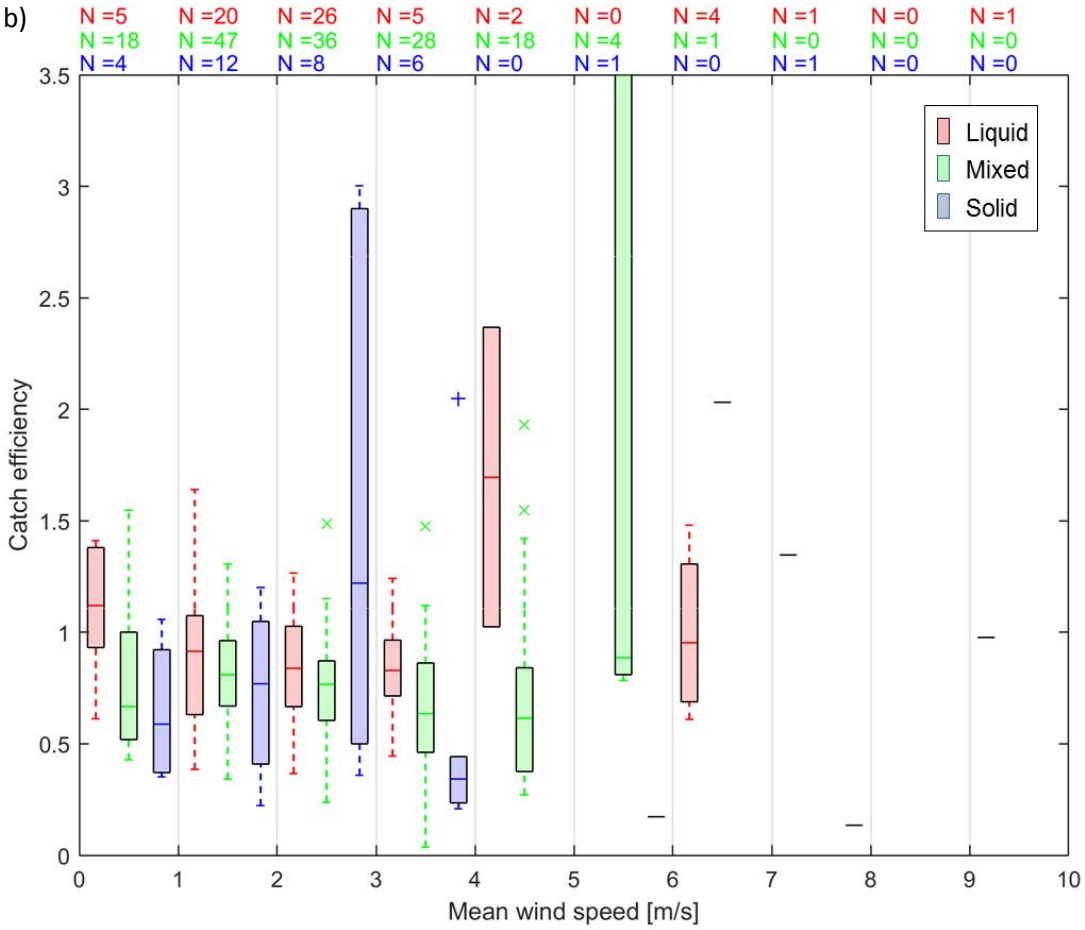


Figure 9: Catch ratio vs. mean wind speed scatter plots for 30 minute precipitation events at (a) CARE, (b) Marshall, east gauge, and (c) Marshall, west gauge. The mean event temperature is indicated by colour. The catch efficiency axes are limited to values ≤ 3.5 for clarity.

a)


N =16 N =21 N =12 N =16 N =15 N =10 N =9 N =8 N =2
N =16 N =8 N =23 N =19 N =13 N =3 N =0 N =0 N =0
N =0 N =8 N =24 N =16 N =5 N =1 N =0 N =0 N =0
Liquid
Mixed
Solid
Catch efficiency
Mean wind speed [m/s]

b)


N =5 N =20 N =26 N =5 N =2 N =0 N =4 N =1 N =0 N =1
N =18 N =47 N =36 N =28 N =18 N =4 N =1 N =0 N =0 N =0
N =4 N =12 N =8 N =6 N =0 N =1 N =0 N =1 N =0 N =0
Liquid
Mixed
Solid
Catch efficiency
Mean wind speed [m/s]

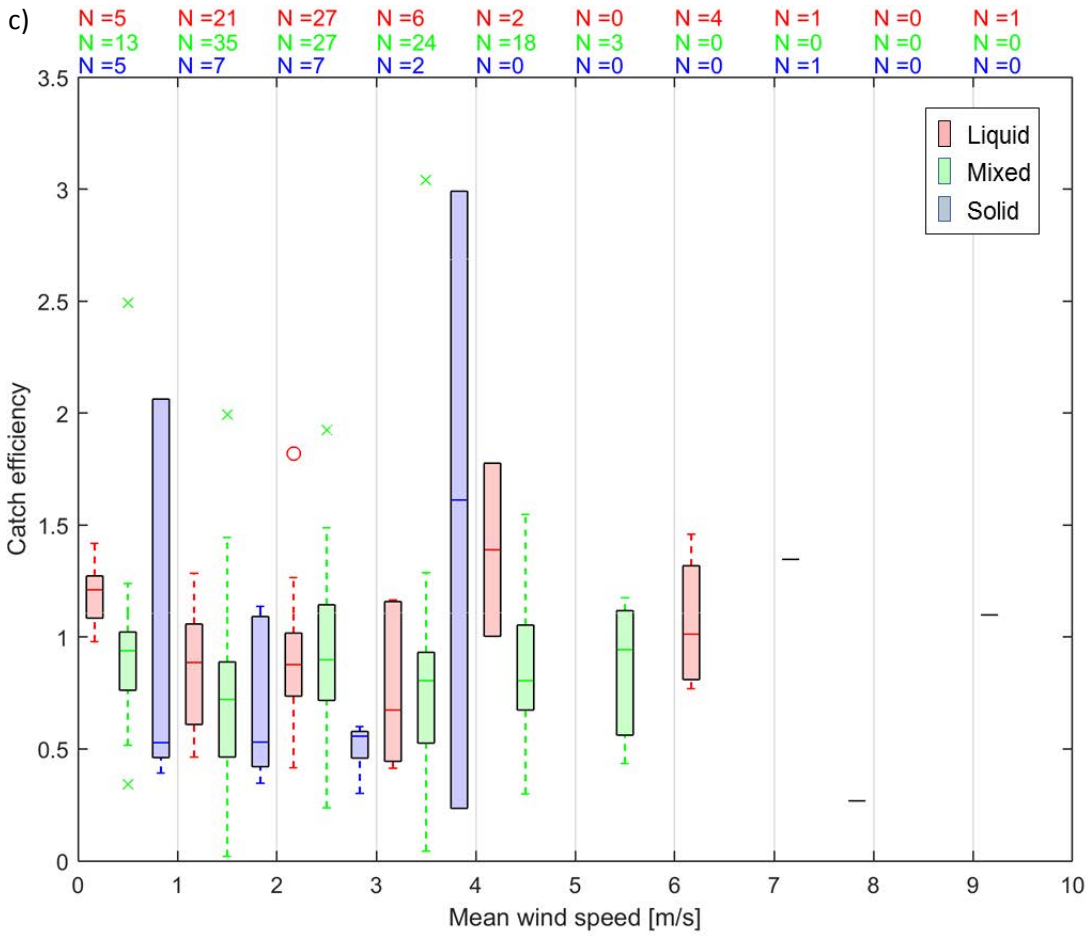


Figure 10: Catch ratio vs. mean wind speed box and whisker plots for 30 minute precipitation events at (a) CARE, (b) Marshall, east gauge, and (c) Marshall, west gauge. The predominant precipitation type for each event is determined from the maximum and minimum reported temperature and indicated by colour. The catch efficiency axes are limited to values ≤ 3.5 for clarity.

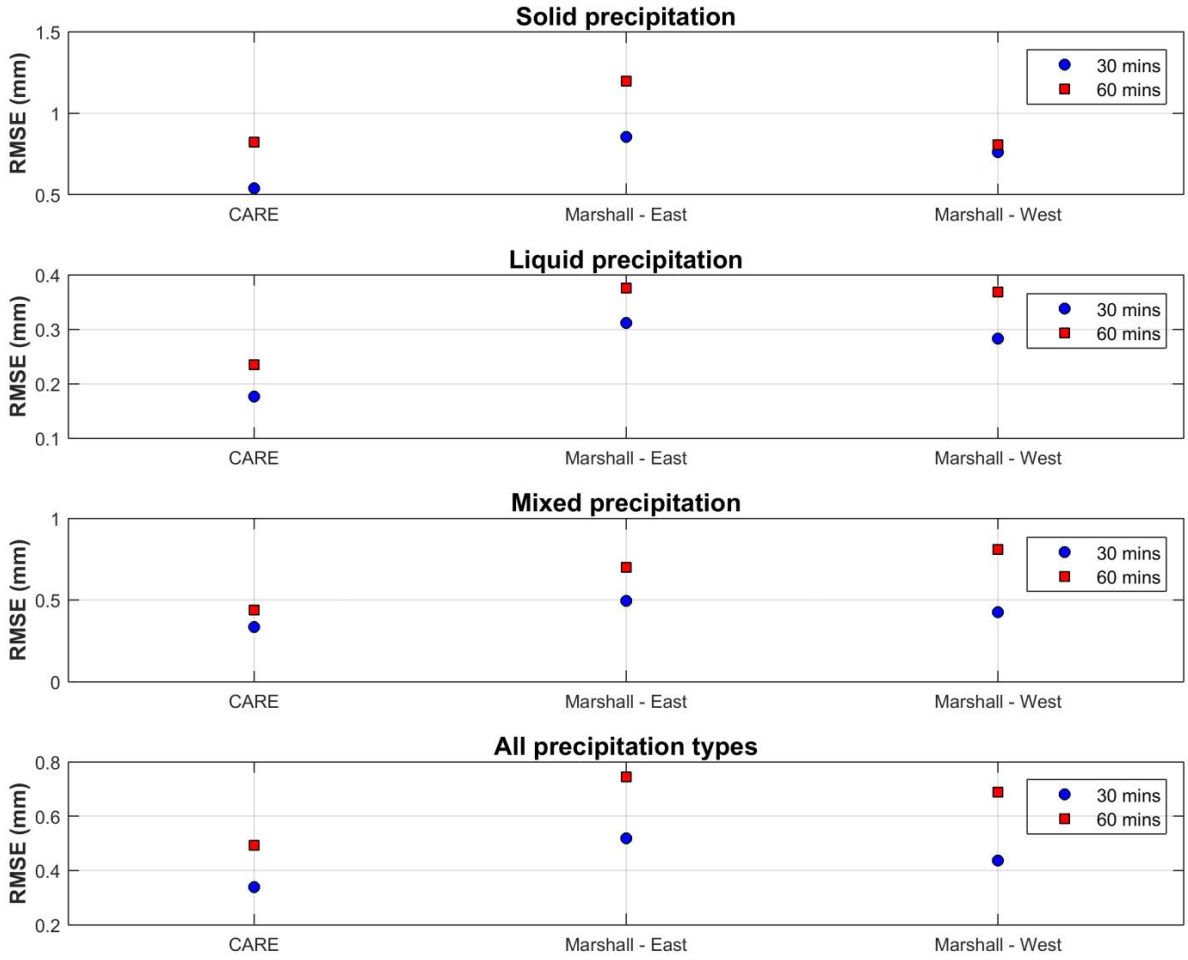


Figure 11: Graphical representation of the RMSE values, by site, and by precipitation type, for YY cases.

The total accumulation reported by each site reference and SUT, for all assessment intervals during which both detected and reported precipitation, are presented alongside the corresponding catch ratios in Table 5.

Table 5: SUT and reference accumulation and overall catch ratio, assessed for YY cases over the entire test period, by site.

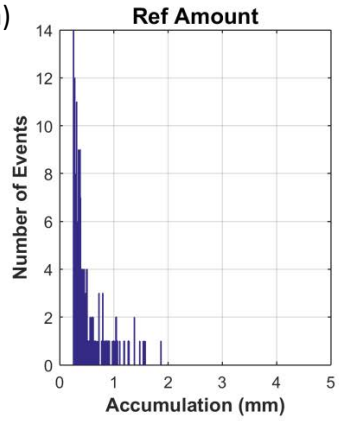
Site	Time interval	SUT accumulation	Reference accumulation	Overall catch ratio
CARE	30 min	162.4 mm	185.8 mm	0.87
	60 min	176.0 mm	218.1 mm	0.81
Marshall (East)	30 min	232.2 mm	257.7 mm	0.90
	60 min	250.7 mm	291.7 mm	0.86
Marshall (West)	30 min	177.4 mm	202.9 mm	0.87
	60 min	187.9 mm	234.7 mm	0.80

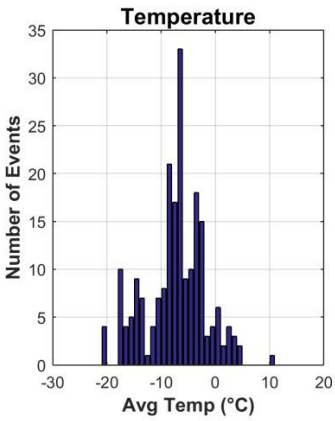
6.3. Assessment of events when the reference and the SUT do not agree on the occurrence of precipitation

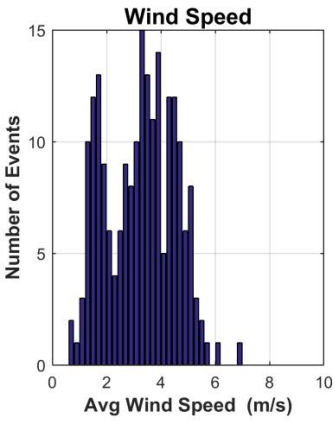
Assessment intervals during which the site reference and SUT do not agree on the occurrence of precipitation – namely, the YN and NY cases (Section 4.1.1) – are characterized using histograms in Figures 12 and 13. The histograms include accumulated precipitation, precipitation intensity as reported by the reference, and corresponding site conditions.

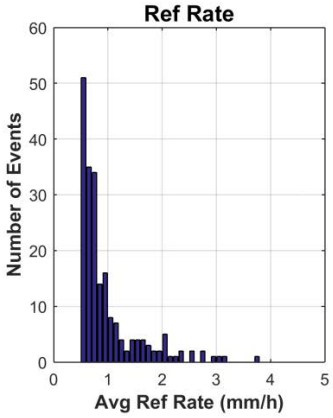
The total SUT and reference accumulations over the test period include contributions from YN and NY cases, which impact the overall catch efficiency. These contributions may be significant, given the response delays associated with heated tipping bucket gauges. Total accumulation and catch efficiency results presented in Table 5 (YY cases only) are expanded to include contributions from YN and NY cases in Table 6.

a)



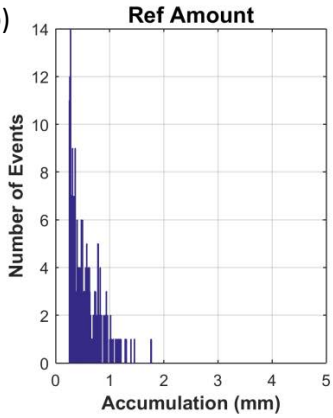


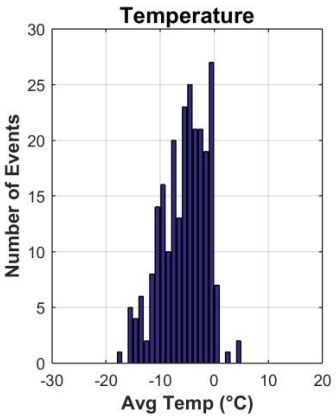


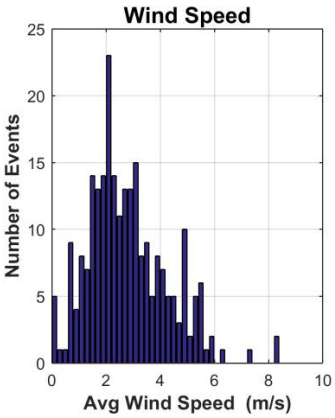


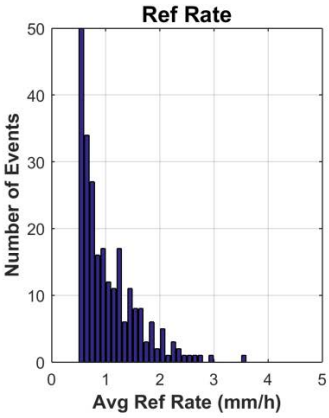
207 events with SUT accumulation = 0 mm (misses)

b)



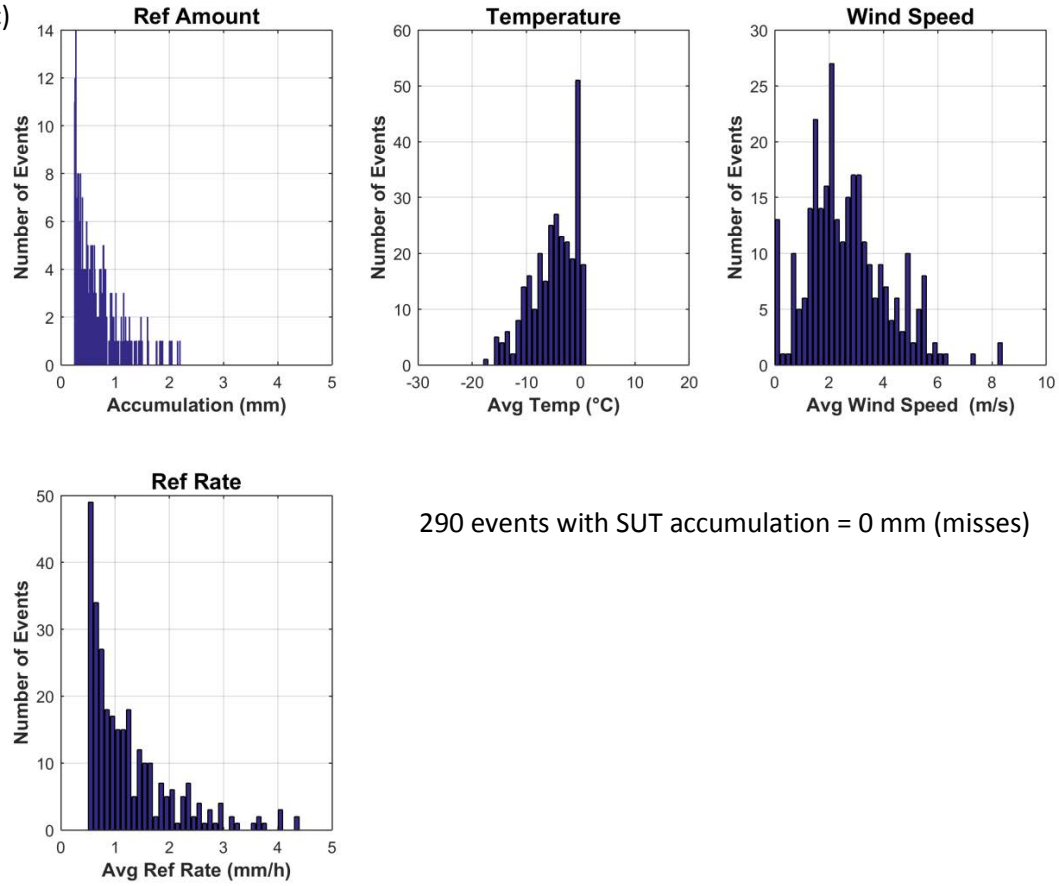






245 events with SUT accumulation = 0 mm (misses)

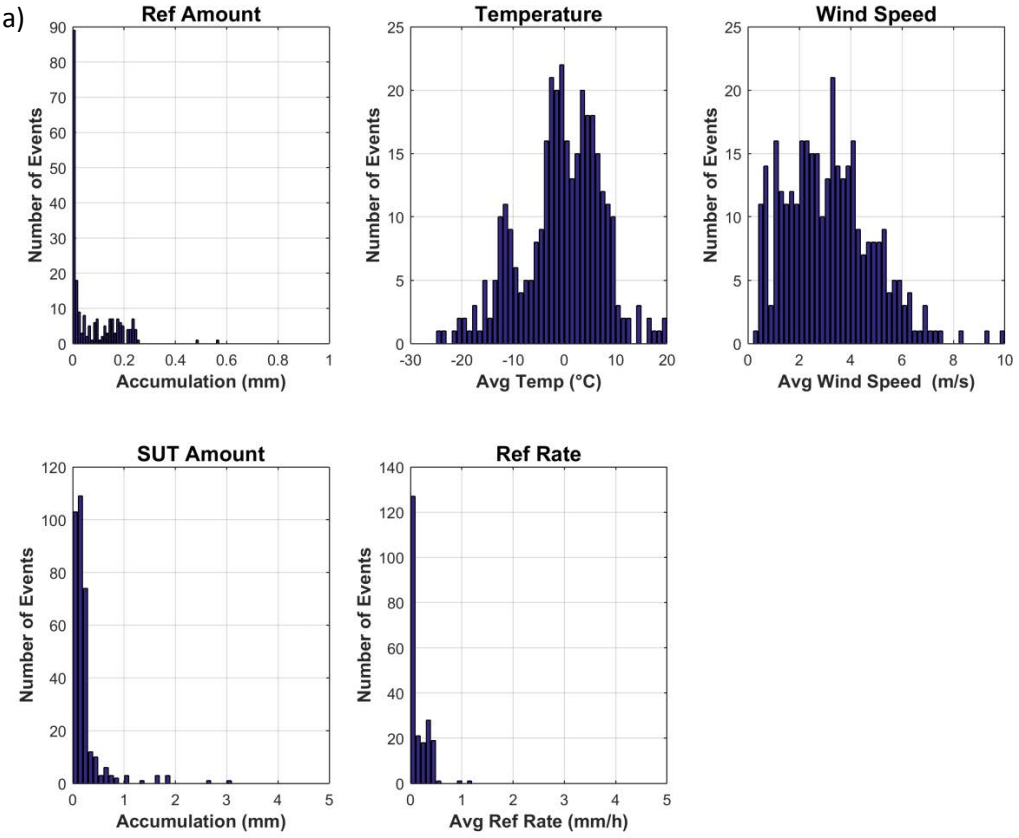
c)



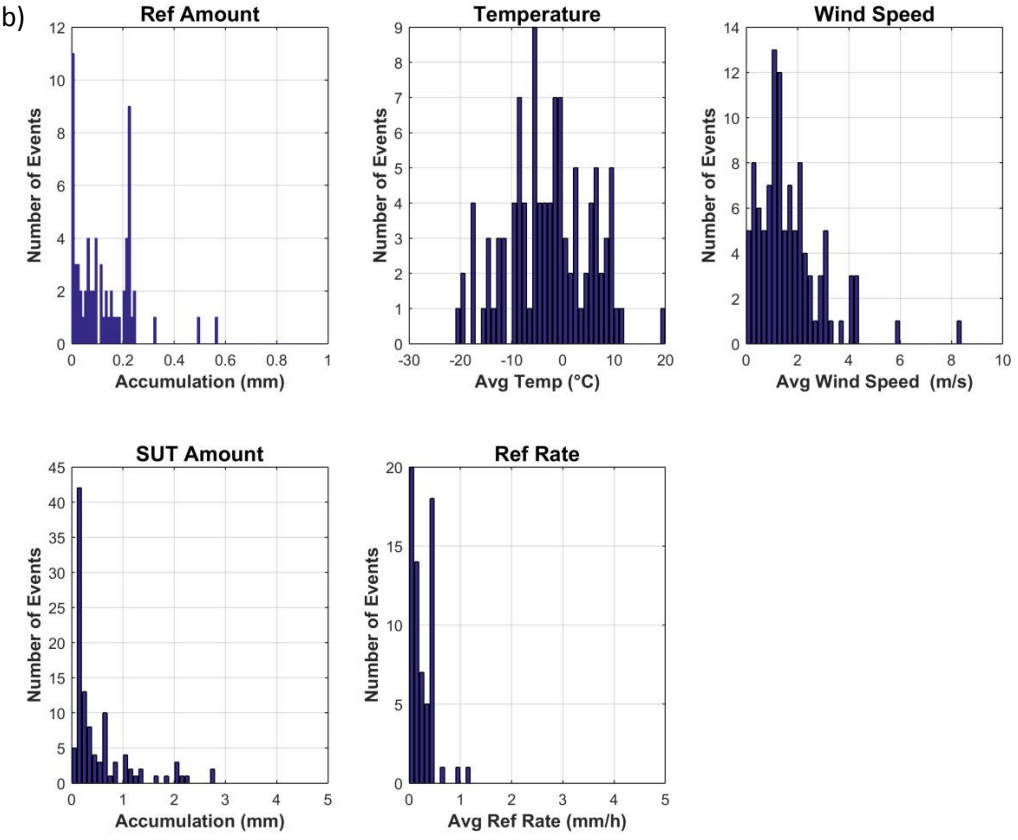
290 events with SUT accumulation = 0 mm (misses)

Figure 12: Histograms of reference accumulation, mean temperature, mean wind speed, and reference precipitation rate for YN cases at (a) CARE, (b) Marshall, east gauge, and (c) Marshall, west gauge.

a)


Ref Amount
Number of Events
Accumulation (mm)
Temperature
Number of Events
Avg Temp (°C)
Wind Speed
Number of Events
Avg Wind Speed (m/s)
SUT Amount
Number of Events
Accumulation (mm)
Ref Rate
Number of Events
Avg Ref Rate (mm/h)

b)


Ref Amount
Number of Events
Accumulation (mm)
Temperature
Number of Events
Avg Temp (°C)
Wind Speed
Number of Events
Avg Wind Speed (m/s)
SUT Amount
Number of Events
Accumulation (mm)
Ref Rate
Number of Events
Avg Ref Rate (mm/h)

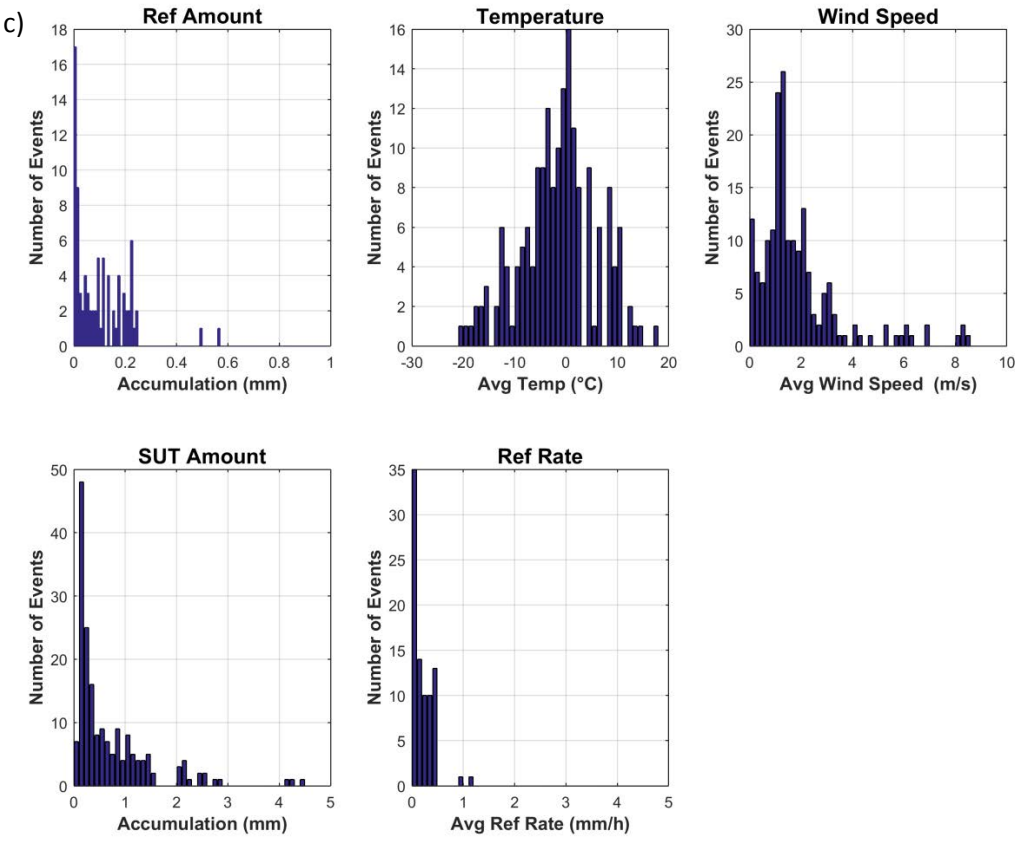


Figure 13: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, and reference precipitation rate for NY cases at (a) CARE, (b) Marshall, east gauge, and (c) Marshall, west gauge.

Table 6: SUT and reference accumulation and overall catch ratio for each site/gauge, assessed for the entire test period, for all YY, YN, and NY cases.

Site	Time interval	SUT accumulation	Reference accumulation	Overall catch ratio
CARE	30 min	248.9 mm	300.9 mm	0.83
	60 min	250.9 mm	350.2 mm	0.72
Marshall (East)	30 min	295.7 mm	396.4 mm	0.75
	60 min	301.7 mm	452.3 mm	0.67
Marshall (West)	30 min	306.4 mm	394.5 mm	0.78
	60 min	310.9 mm	450.9 mm	0.69

7. Interpretation of Results

7.1. Operating conditions

The full range of conditions under which the test gauges were operated is illustrated in Figure 3. The conditions relevant to gauge operation are as follows:

- Temperatures between -31 °C and 27 °C;
- Precipitation intensity within 12 mm/hr.

These conditions fall within the manufacturer's specified range of operating conditions.

Both sites are located in the continental interior, but the conditions at Marshall are generally drier and less windy relative to those at CARE, with a larger proportion of events with RH < 80% and a smaller proportion of events with mean wind speeds > 3 m/s (Figure 3). In general, the distributions of temperature, relative humidity, temperature, and precipitation rate are similar for the precipitation events at CARE and Marshall (Figure 4).

7.2. Ability to detect and report precipitation

7.2.1. Skill score assessment

The performance of the sensors under test was considered in terms of skill scores, computed for both 30 and 60 minute assessment intervals (Figure 5). The Probability of Detection (POD) is higher for 30 minute assessment intervals relative to 60 minute intervals, as is the Bias (B). False Alarm Rates (FAR) appear to be less sensitive to the duration of the assessment interval, with similar values for the 30 minute and 60 minute assessment intervals for each test gauge. The Heidke Skill Scores (HSS) are higher for 30 minute assessment intervals relative to the values for 60 minute assessment intervals for all gauges under test. The comparison of skill scores among the test gauges will be considered for the 60 minute assessment intervals only, as hourly reports are broadly used, operationally.

The test gauge at CARE has the highest POD (49%), but also the highest FAR (59%) and B (120%), and the lowest HSS (41%). The East and West test gauges at Marshall show different performance in terms of ability to detect precipitation, with POD of 44% and 37%, FAR of 28% and 44%, B of 61% and 67%, and HSS of 53% and 43%, respectively. The observed differences in skill scores follow from the distribution of events in the contingency table (Table 4). The biases are similar for the Marshall gauges, but the other score values differ, owing to fewer 'hits' and more 'misses' and 'false alarms' for the west gauge; these differences illustrate the potential influence of gauge siting on precipitation, as these are identical gauges in the same configuration, installed at different locations on the same site. The differences among the test gauges at Marshall and CARE are attributed to differences in siting and installation, site configuration, and the specific conditions experienced at each site.

7.2.2. Characterization of response delays

7.2.2.1. Response delay PDFs

Probability distribution functions (PDFs) of response delays between the onset of precipitation as observed by the reference configuration and the first response by the HSA TBH heated tipping bucket gauge are shown in Figure 6. The distributions are similar: for the gauge under test at CARE, response times between 10 and 40 minutes have the highest probability, while the PDFs for the gauges at Marshall indicate the highest probabilities of response times between about 20 and 40 minutes. Further, the peaks in the PDFs (indicating highest probability) are at about 20 to 30 minutes for all test gauges, and the PDFs tail off gradually, with low probabilities of response times up to 180 minutes.

For precipitation occurring within a given 30 or 60 minute interval (as indicated by the reference configuration), the response delay PDFs for the test gauges at both sites indicate the potential for the TB response to occur within a subsequent 30 or 60 minute interval. The delays with highest probability in the PDFs are within 30 minutes for all test gauges, which may explain the higher POD for 30 minute assessment intervals relative to 60 minute intervals in Figure 5. However, as there is no requirement for precipitation to begin at the beginning of an assessment interval, the shorter assessment intervals are also subject to more false alarm (NY) events (Table 4), in which the TB responds to precipitation in a subsequent interval. This is reflected by the higher FAR and B for 30 minute intervals in Figure 5.

Another important consideration is that even when both the SUT and reference detect precipitation within a given interval, the TB may actually be reporting precipitation collected during previous intervals, or may only report a fraction of the precipitation within that interval (reporting the rest during a subsequent interval). These ‘false YY cases’ will impact the assessment of TB gauge performance in terms of reporting accumulated precipitation (Section 7.3).

7.2.2.2. Influence of precipitation intensity, phase

The dependence of delay times on the mean intensity of precipitation preceding the TB response, as reported by the reference configuration, is illustrated in Figure 7. For all gauges tested, response delays are within 30 minutes for mean precipitation intensities within 0.4 mm/hr, irrespective of precipitation type. At higher mean intensities (> 0.4 mm/hr), delays are typically within 40 minutes for liquid precipitation, extending up to about 180 minutes for mixed and solid precipitation. The longer delays for mixed and solid precipitation are attributed primarily to delays in the application of heating; the snow sensor in the funnel will be activated at temperatures in the mixed and solid regimes, but heating will not be triggered until the level of precipitation in the funnel reaches the sensor height. These delays will be compounded by the time required to melt the precipitation, once heating has been initiated.

The determination of precipitation phase using temperature is not absolute, which may explain the liquid and mixed events with longer delay times, as well as the low relative numbers of snow events. The phase classification is based on the minimum and maximum temperature (see Section 4.1.4) over the period preceding the TB gauge response. For example, a predominantly solid event may

transition into the mixed temperature regime, changing the classification. Also of note is that the mean precipitation intensity values for longer delay periods may not necessarily be representative; that is, the intensity may be variable over the delay period, which could account for the outlying point with mean intensity of about 0.15 mm/hr in Figure 7a.

7.3. Ability to report accumulated precipitation

7.3.1. Performance when both SUT and reference detect precipitation

7.3.1.1. Temperature, phase, and wind speed influence

Focussing on the 30 minute events in which both the sensors under test and site reference configurations report precipitation (YY cases, or 'hits,' only), colder temperatures are generally correlated with larger differences in accumulation between the reference and SUT in Figure 8. That is, the events at colder temperatures generally fall further from the 1:1 line. Of particular note in Figure 8 are events for which the gauges under test reported more precipitation than what was reported by the reference configuration. These are believed to be 'false YY cases,' assessment intervals in which the reference responds to precipitation presently occurring, while the TB reports precipitation occurring, at least in part, during an earlier interval, due to TB response delays (Section 7.2.2.1).

Significant scatter is observed in the plots of collection efficiency vs. mean wind speed for all gauges under test (Figure 9). The catch efficiency axes are limited to values up to 3.5 for clarity. The events for which the SUT reported more precipitation than the reference configuration (false YY cases) are evident at the highest catch efficiencies. Figure 9b indicates that colder events tend to have lower catch efficiencies for the east gauge at Marshall, but this trend is not as clear for the other gauges. For all gauges, there is no apparent wind speed influence on collection efficiency.

Representing the data as box and whisker plots and stratifying the events by phase, a decrease in the median catch efficiency with increasing wind speed is observed for solid precipitation at CARE (Figure 10), with median catch efficiencies falling below 0.5 for wind speeds ≥ 3 m/s. Similar trends are not evident for the gauges under test at Marshall (Figures 8b and 8c). Here, again, the catch efficiency axes are scaled for clarity.

The response delays for mixed/solid events are a key factor in the observed variability in the catch efficiency – wind speed relationship in Figures 9 and 10. The heating of the TBH is triggered only when the precipitation in the funnel reaches the level of the snow sensor, resulting in temporal separation of the TB response from the preceding interval or intervals during which it occurred. Accordingly, the TBH is not well-suited for this type of assessment. It is important to note, however, that the TBH heating configuration has been designed to provide solid precipitation observations over extended time periods with low power requirements, and was not intended for this particular application.

7.3.1.2. Root mean square error

The RMSE was calculated for all 30 and 60 minute events during which the reference and SUT both detected precipitation (YY cases), for each site. The overall RMSE calculated for each site (all precipitation types), and RMSE values calculated for each precipitation type, are shown in Figure 11. Focussing on the operationally-relevant 60 minute events, the overall RMSE was approximately 0.5 for the gauge at CARE and 0.7 for both gauges at Marshall. Hence, for the specific site and gauge configurations tested, over the range of conditions tested, the absolute uncertainty of the HSA TBH relative to a DFAR configuration can be considered to be within approximately 0.7 mm.

The largest phase-specific RMSE values were determined for solid precipitation (within 1.2 mm over 60 minutes), with values within 0.8 mm over 60 minutes for mixed precipitation, and within about 0.4 mm over 60 minutes for liquid precipitation. Differences in the relative RMSE values among the test gauges are attributed to differences in gauge siting and/or configuration, and the specific conditions experienced at each site.

7.3.1.3. Overall catch efficiency

The total accumulated precipitation recorded by the reference and sensor under test at each site was compiled for all 30 and 60 minute events during which both detected precipitation (YY cases) and used to calculate the overall catch ratio. The results are provided in Table 5. Again focussing on the values for the more operationally-relevant 60 minute intervals, the overall catch ratios for CARE and the east and west gauges at Marshall are 0.81, 0.86, and 0.80, respectively, indicating similar overall performance by the gauges under test at both sites. The higher catch efficiency observed for the east gauge at Marshall is correlated with its higher skill in detecting precipitation as indicated by the Heidke Skill Score in Figure 5.

7.3.2. Characterization of conditions when the SUT and reference do not agree on the occurrence of precipitation

The conditions during 30 minute events in which the reference detected precipitation, but the SUT did not (YN cases), are depicted as histograms in Figure 12. These events are characterized by low to moderate precipitation rates, typically below 4 mm/hr, and temperatures close to, or below, freezing. At these low rates, it will take longer for precipitation to reach the level of the sensor in the funnel that triggers heating, and this delay will be compounded with the time required to melt the precipitation.

The conditions during 30 minute events in which the SUT detected precipitation, but the reference did not (NY cases), are shown as histograms in Figure 13. These events are also characterized by low precipitation rates, below 0.5 mm/hr, which corresponds to the reference accumulation threshold for a precipitation event over a 30 minute interval. These events most likely result from delayed tipping bucket responses that coincide with intervals during which the reference did not detect precipitation.

Given the influence of response delays, estimates of the overall catch ratio for tipping bucket gauges should include the reference and SUT accumulations during YN and NY cases (and not just during the

YY cases, as considered above). The resulting total accumulations for all YY, YN, and NY cases, and the corresponding overall catch efficiencies, are provided for the test gauges at each site in Table 6. The overall catch efficiencies for the gauge under test at CARE and the east and west test gauges at Marshall were 0.72, 0.67, and 0.69 for 60 minute assessment intervals, demonstrating similar overall performance for all gauges tested at both sites.

8. Operational considerations

The HSA TBH was easy to install, easy to connect, and the setup of data collection was straightforward. The use of a snow sensor in the funnel to trigger heating limits power consumption, which may be an asset, depending on available site infrastructure and resources.

8.1. Maintenance

Each site completed the gauge field calibration and verification as per manufacturer recommendations, at least once a year. The calibration records have been stored by each site team.

8.2. Noted issues

- The site team at Marshall noted significant problems with the heaters; while power was provided to the heaters, they were not triggered when there was snow in the funnel, and it is believed that the heating units failed.

9. Performance Considerations

- Response delays must be considered when using the gauge in operational settings. The response delays observed for this gauge are attributed primarily to the non-continuous nature of the heating configuration, and result in significant scatter in the catch efficiency data for all gauges under test, and for all precipitation types. This indicates poor agreement with the reference configuration over the time scale of events (30 minutes), and suggests that this gauge may be better suited for applications over longer time scales.

- Indeed, the overall catch ratio data indicate good performance for all gauges under test over seasonal time scales. On an event basis, however, the present results illustrate that the test gauges often over-report accumulation relative to the reference configuration.

- Ideally, the reporting interval (i.e. hourly observations) should exceed the maximum expected response delay; however, the potential remains for carry-over of precipitation accumulation from previous intervals, and for delayed responses occurring in subsequent intervals.

- While response delays due to heating impact gauge performance for different applications, heating is critical to ensure proper functioning of the gauge under winter conditions. The voltage should be checked at the gauge location to ensure that the heaters are supplied with

the recommended power. Voltage drops – for example, resulting from the use of long power cables – can impair the functioning of the gauge in winter conditions.

- The potential exists for losses due to wind pumping and/or sublimation when snow accumulates in the funnel below the level of the snow sensor that triggers heating.

- The marked scatter in catch efficiency data over 30 minute intervals confounds the investigation of wind speed influence on gauge performance, and the development and application of associated transfer functions; however, transfer functions may be feasible for data over longer time intervals (e.g. 1 hour).

- If it is possible to develop transfer functions for longer time intervals, their application to gauge measurements may help to account for any observed reductions in catch efficiency with increasing wind speed. Their utility may still, however, be mitigated by the influence of response delays, which contribute to the scatter in the catch ratio data, and hence, the uncertainty in the associated transfer functions.

- The collection of ancillary measurement data is recommended: air temperature and wind speed (at gauge height) to enable the application of adjustments to measurements (i.e. using transfer functions, if available); and reports from a precipitation detector with high sensitivity to enable the identification of missed events or false alarms due to response delays.

WMO-SPICE Instrument Performance Report
Meteoservis MR3H-FC

1) Technical specifications (from manufacturer provided documentation)

Instrument model:	Meteoservis MR3H-FC
Physical principle:	Heated tipping bucket (TB) gauge. Tips occur when the bucket fills to capacity, triggering a reed contact that outputs a pulse.
Bucket capacity:	0.1 mm
Collecting area:	500 cm^2
Heating configuration:	Heaters at 3 sections – funnel, middle part of funnel, and top of funnel/collar – controlled independently by temperature sensors on the internal side of the gauge. In the absence of recorded precipitation, the heating is limited to reduce the potential influence of evaporation. 'Shock heating' is applied to the top of funnel/collar section after precipitation is recorded to enable the removal of possible icing. As tested, equipped with additional outflow heating unit (AH-01) that is operated independently from other heaters. This unit lowers the minimum operating temperature from -30 °C to -40 °C. Total heating power: 555 W (maximum).
Operating temperature range:	-40 °C to 60 °C (as tested); -30 °C to 60 °C without AH-01 outflow heating unit
Measurement range:	0 – 400 mm/hr
Measurement uncertainty:	Within ± 2% from 0 – 400 mm/hr
Sensitivity:	0.1 mm reporting resolution (corresponding to bucket amount)

a)

b)

c)

Figure 1: Meteoservis MR3H-FC installations at (a) CARE, (b) Marshall, and (c) Sodankylä.

2) Data output format

Gauge data output: Gauge reports precipitation intensity in mm/hr, determined from modified 5 V pulses from a reed contact with constant length of 100 ms. Gauge output is corrected internally as a function of precipitation intensity.

Analysis is based on accumulated precipitation, computed as the cumulative sum of accumulations derived from intensity reports.

3) SPICE test configuration

Shield: Unshielded

Test Sites: CARE (Canada); Marshall (USA); Sodankylä (Finland)

Sensor Provider(s): Meteoservis

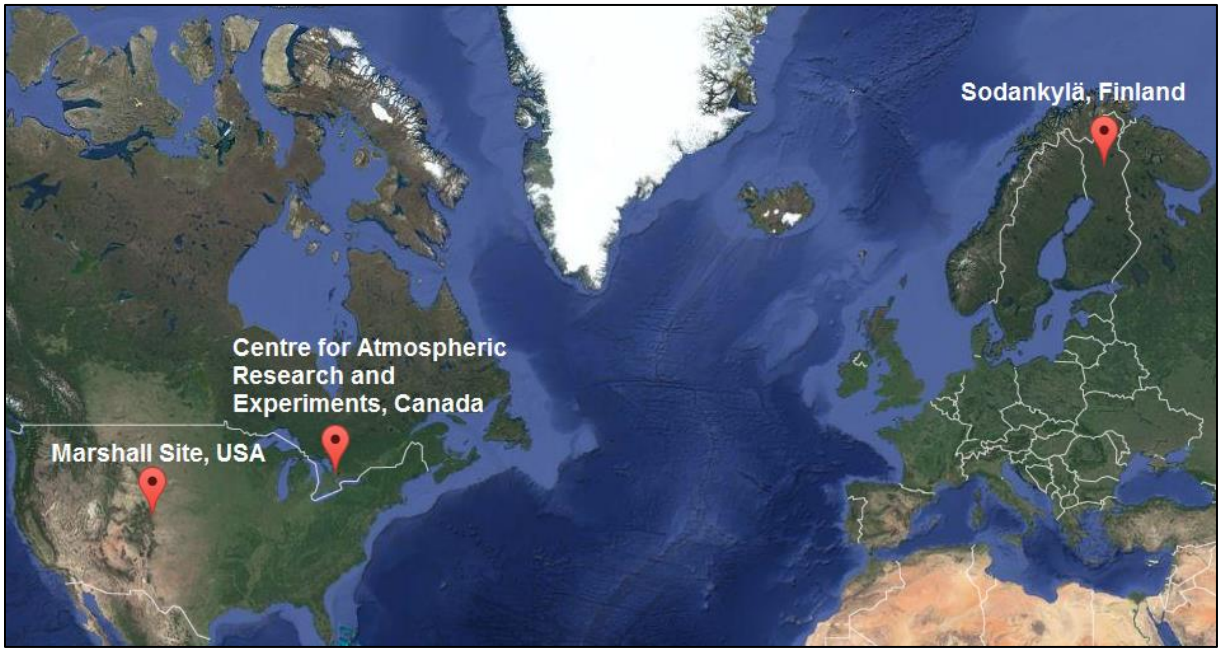


Figure 2: Map of SPICE sites testing Meteoservis MR3H-FC gauges.

A summary of the configuration of instruments as tested, the duration of tests and availability of data reflected in these results, and the ancillary measurements used, by site, is available in Tables 1, 2, and 3, respectively.

Table 1: Summary of gauge configurations and data output, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	CARE	Marshall	Sodankylä
Field configuration	Unshielded	Unshielded	Unshielded
Height of installation (gauge rim)	1.5 m	1.27 m	1.35 m
SUT data output frequency	1 min	6 s , 1 min	1 min
Data QC	SPICE QC methodology		
Data temporal resolution	1 min		
Processing interval for SPICE data analysis	30 min, 60 min		

Table 2: Data availability, by measurement season and site.

Measurement season	CARE	Marshall	Sodankylä
Season 1 (Oct. 2013 – Apr. 2014)	X	✓	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓	✓	✓

Data collected at CARE during the 2013-2014 season are excluded from the present assessment due to issues with the gauge configuration that were identified following the conclusion of the measurement season. These issues are the responsibility of the site team, and do not reflect any shortcomings of the instrument or issues with its operation.

Table 3: Summary of reference and ancillary measurements, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	CARE	Marshall	Sodankylä
R2 site reference	Geonor T-200B3 600 mm (DFAR)	Geonor T-200B3 600 mm (DFAR)	OTT Pluvio^2 (DFAR)
R2 precip detector	Thies LPM (DFAR)	Thies LPM (Site*)	OTT Parsivel2 (Site*, 2013-2014), OTT Parsivel2 (DFAR, 2014-2015)
Ancillary temp sensor	Vaisala HMP155 (Stevenson screen)	MetOne, model 060A-2/062, 2144-L	Vaisala HMP155 (2 m)
Ancillary RH sensor	Vaisala HMP155 (Stevenson screen)	Campbell Scientific CS500	Vaisala HMP155 (2 m)
Ancillary wind sensor	Vaisala NWS 425 (2 m)	RM Young Wind Monitor 05103 (2 m)	Thies acoustic 2D wind sensor (3.5 m)

*A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR-fence; however, in cases where a more sensitive detector is available outside of the DFIR-fence, or there are issues with the detector within the DFIR-fence, a precipitation detector elsewhere on the site can be employed.

4) Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6.1 of the WMO-SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute assessment intervals:

- Ref 'Yes' : R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No' : R2 weighing gauge < 0.25 mm AND/OR precip detector recording < 18 min of precip;
- SUT 'Yes' : SUT accumulation > 0 mm;
- SUT 'No' : SUT accumulation = 0 mm.

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores.

4.1.2. Skill score assessment

The ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

The above scores are computed using the formulations provided in Section 3.6.1.3 of the WMO-SPICE Final Report.

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

4.2. Sensor-specific considerations: response delays

Tipping bucket gauges require that an amount of precipitation corresponding to the bucket capacity is accumulated before a tip is triggered and the gauge records precipitation. This can result in response delays relative to the reference configuration. Heated tipping bucket gauges are subject to further delays, as any solid precipitation in the funnel must be melted before reaching the bucket and potentially triggering a tip, and the heating itself can potentially evaporate incident precipitation. These response delays will impact the comparison with the reference. For this reason, the assessment of TB gauge performance relative to the reference configuration is also considered over 60 minute intervals, using the same conditions and thresholds outlined above in Section 4.1.1.

Response delays are quantified by determining the time elapsed between the onset of precipitation as determined by the reference configuration, and the first tip recorded by the TB. The assessment is based on periods with at least 30 minutes of precipitation, as identified by the reference configuration, followed by at least 180 minutes without precipitation. This extended period without precipitation is intended to allow additional time for the melting and recording of precipitation by heated TB gauges. Additional details are provided in Section 3.6.1.4.4 of the WMO-SPICE Final Report.

5) Environmental conditions

The environmental conditions at each site over the duration of the test period are expressed as probability density functions (PDFs) of mean air temperature, mean relative humidity, mean wind speed, vector mean wind direction, and precipitation rate for each component 30 minute assessment interval in Figure 3. Figure 4 presents the same parameters for all assessment intervals during which the site reference configuration detected precipitation (i.e. all Ref 'Yes' cases).

The precipitation percentage represents the number of minutes of precipitation during a 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration, expressed as a percentage. PDFs of precipitation percentage are also included in Figures 3 and 4.

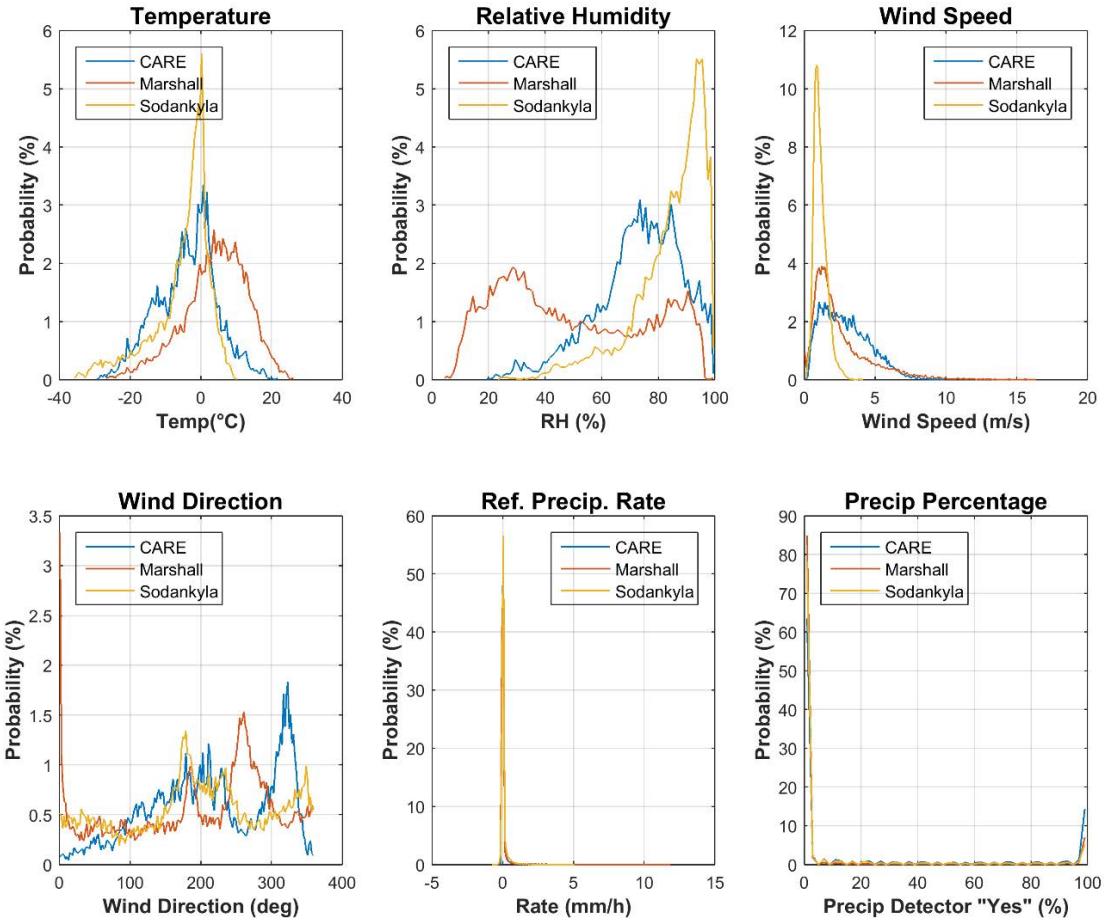


Figure 3: Summary of aggregated environmental conditions at the SPICE sites that operated Meteoservis MR3H-FC gauges, over the entire duration of formal tests, as per Table 2.

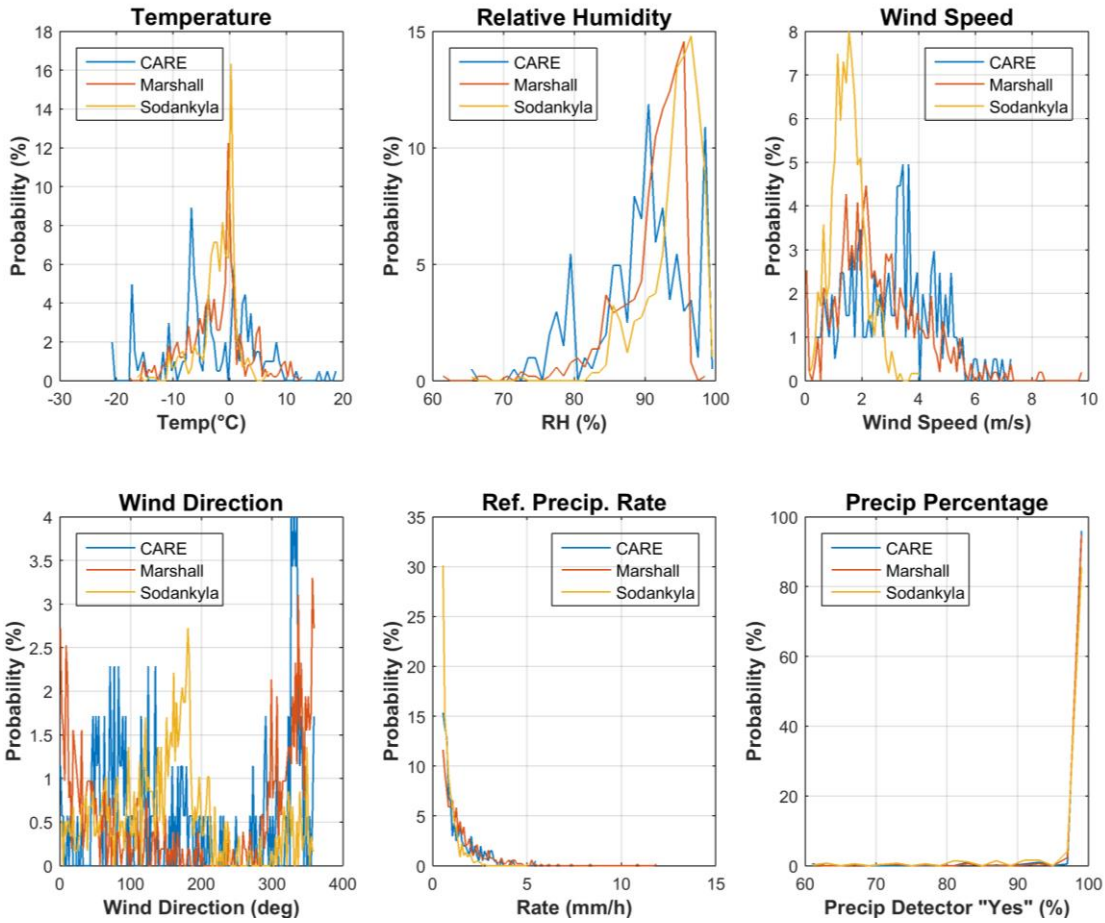


Figure 4: Summary of the aggregated environmental conditions at the SPICE sites that tested Meteoservis MR3H-FC gauges during precipitation events, as reported by the site R2 reference, during the formal tests, as per Table 2.

6) Evaluation of performance over the range of operating conditions

6.1. Ability to detect and report precipitation

6.1.1. Skill score assessment

The overall ability of the SUT to detect and report the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores (Section 4.1.2) and presented in Figure 5. Scores are presented for both 30 minute and 60 minute assessment intervals. The contingency results (Section 4.1.1) corresponding to these scores are presented in Table 4.

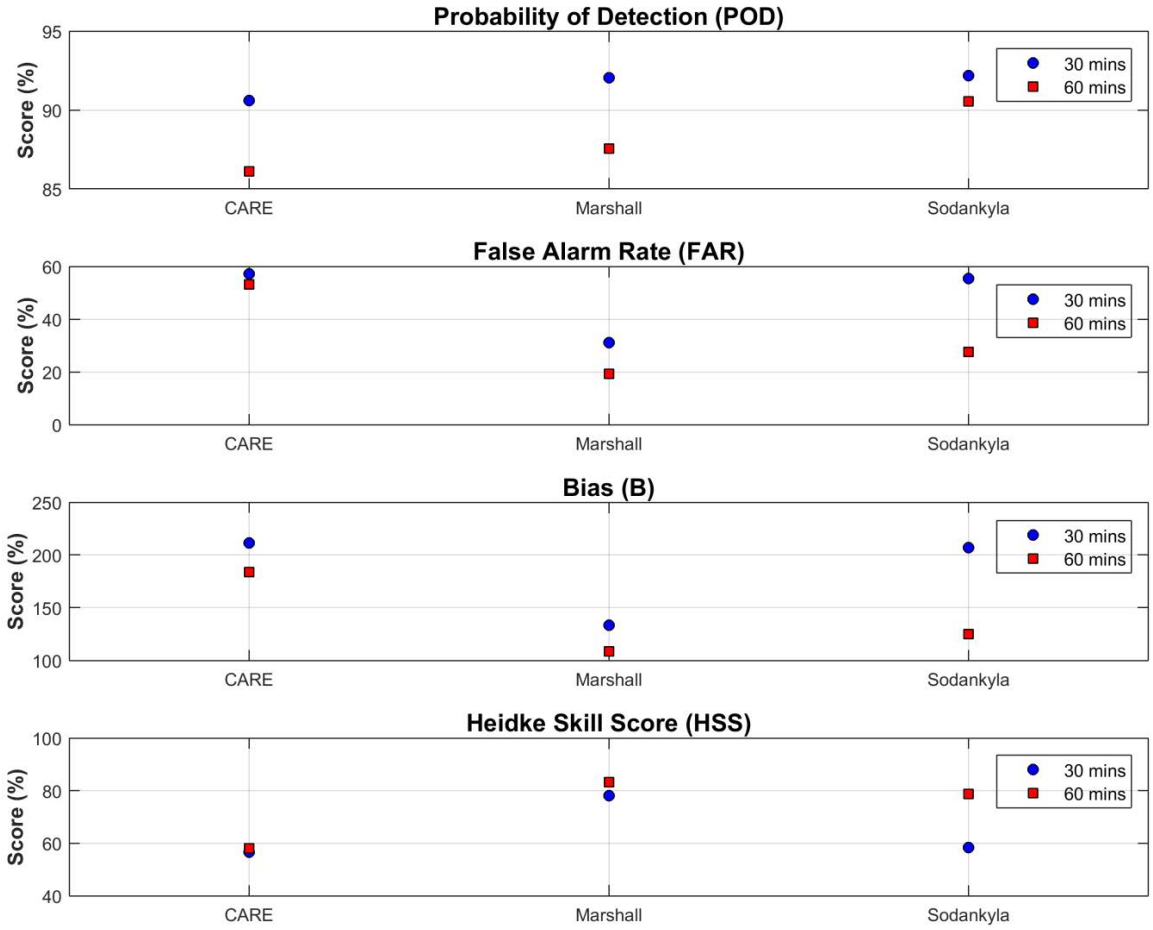


Figure 5: Skill scores for Meteoservis MR3H-FC gauges during the formal test periods.

Table 4: Contingency table illustrating detection of precipitation by the Meteoservis MR3H-FC relative to the specific site reference, expressed as number (and percentage) of events over the entire test period, by site.

Site	Time Interval	Number of Events (% of Events)			
		YY, hits	YN, misses	NY, false alarms	NN, correct negatives
CARE	30 min	183 (2.5%)	19 (0.3%)	244 (3.3%)	6917 (93.9%)
	60 min	155 (4.2%)	25 (0.7%)	176 (4.8%)	3316 (90.3%)
Marshall	30 min	474 (2.4%)	41 (0.2%)	214 (1.1%)	19246 (96.4%)
	60 min	373 (3.7%)	53 (0.5%)	89 (0.9%)	9460 (94.8%)
Sodankylä	30 min	542 (2.8%)	46 (0.2%)	674 (3.5%)	17865 (93.4%)
	60 min	613 (6.4%)	64 (0.7%)	234 (2.5%)	8619 (90.4%)

6.1.2. Characterization of response delays

Response delays were determined for the test gauges at each site using the approach outlined in Section 4.2, and compiled over the entire test period. The delays are represented as probability density functions in Figure 6. These PDFs provide an overall picture of response delays for each sensor in all precipitation types. The response delays are characterized further by separating the precipitation events by type/phase, and plotting the response delays as a function of the mean precipitation intensity observed by the reference gauge during the delay period (Figure 7).

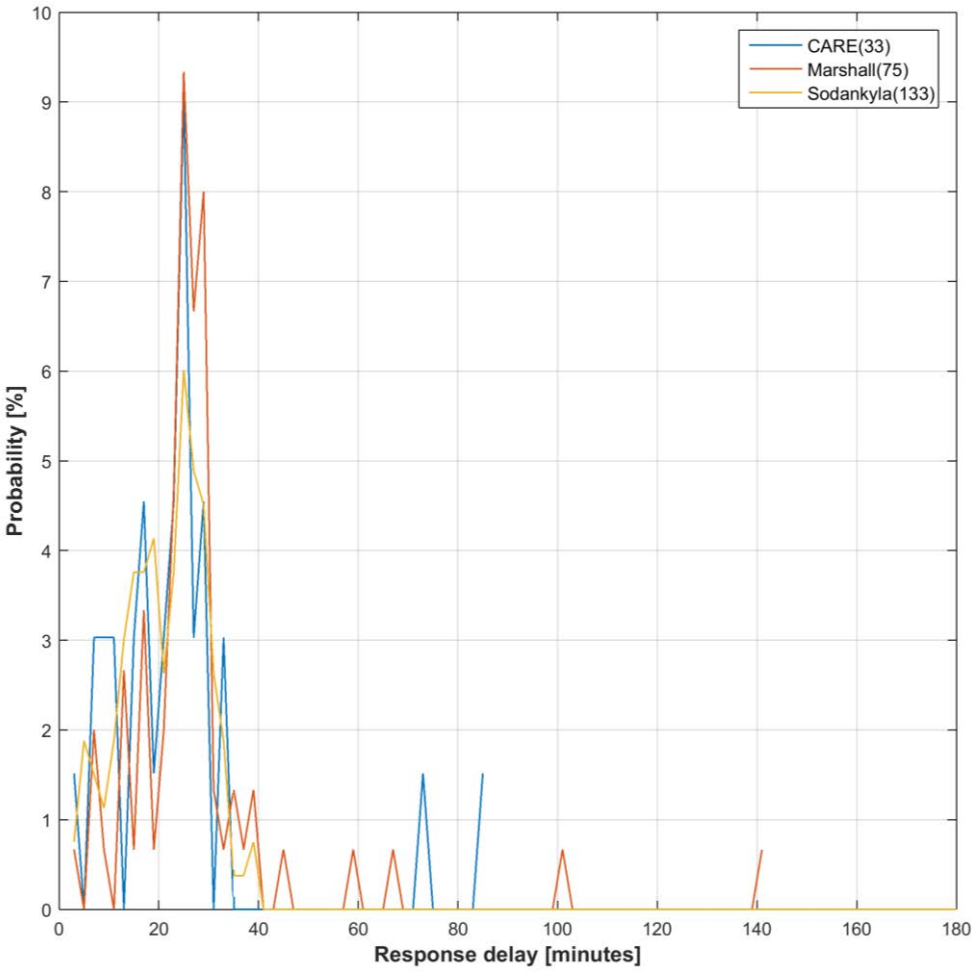
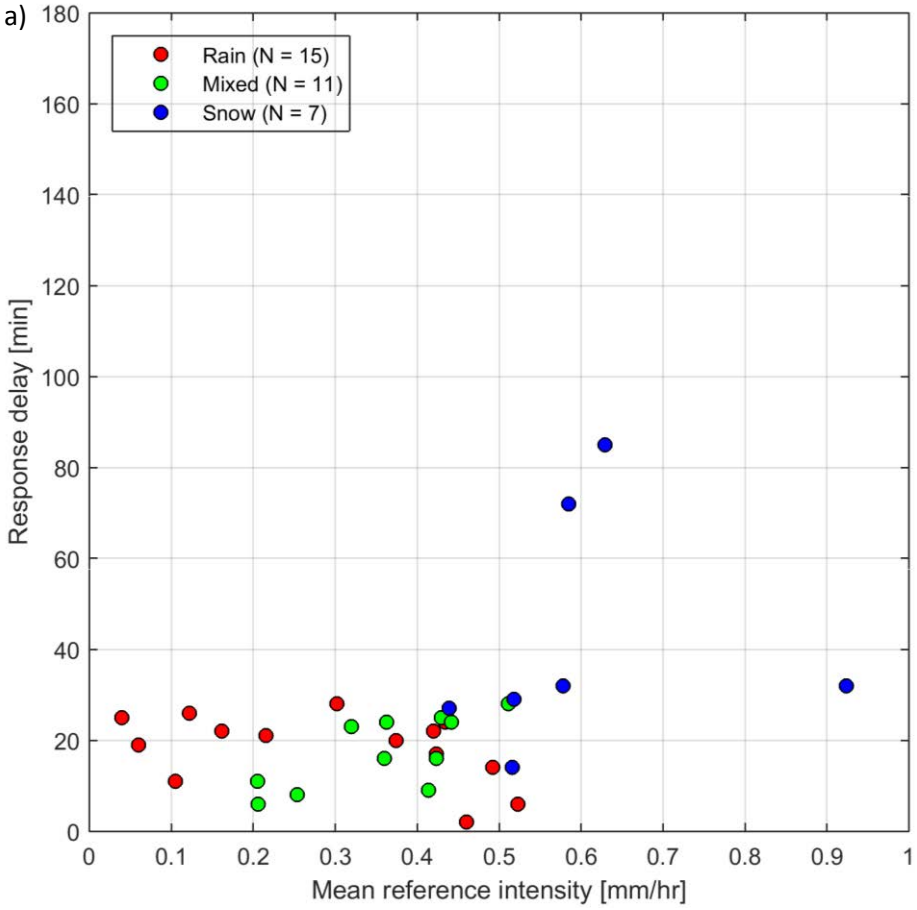
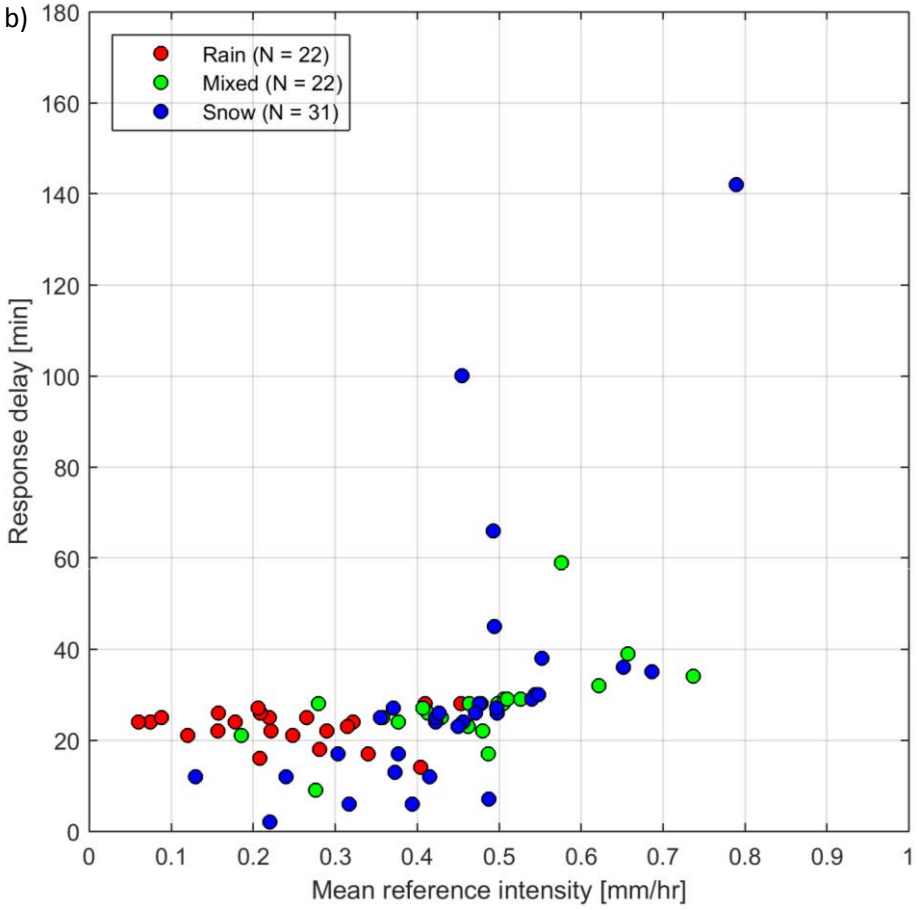


Figure 6: Response delays for heated Meteoservis MR3H-FC gauges at CARE, Marshall, and Sodankylä relative to the R2 reference configuration at each site. The number of response assessment periods used to determine the delays for gauges at each site are indicated in parentheses in the legend.

a)


Rain (N = 15)
Mixed (N = 11)
Snow (N = 7)
Response delay [min]
Mean reference intensity [mm/hr]
0
20
40
60
80
100
120
140
160
180
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

b)


Rain (N = 22)
Mixed (N = 22)
Snow (N = 31)
Response delay [min]
Mean reference intensity [mm/hr]
0
20
40
60
80
100
120
140
160
180
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

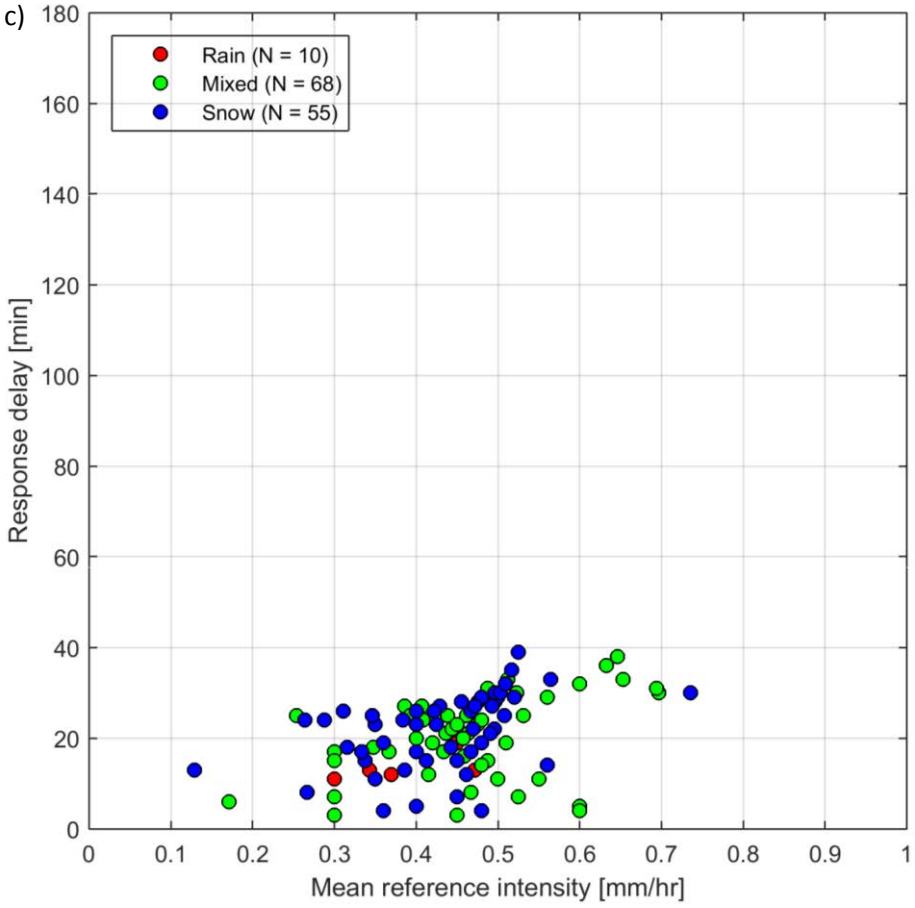


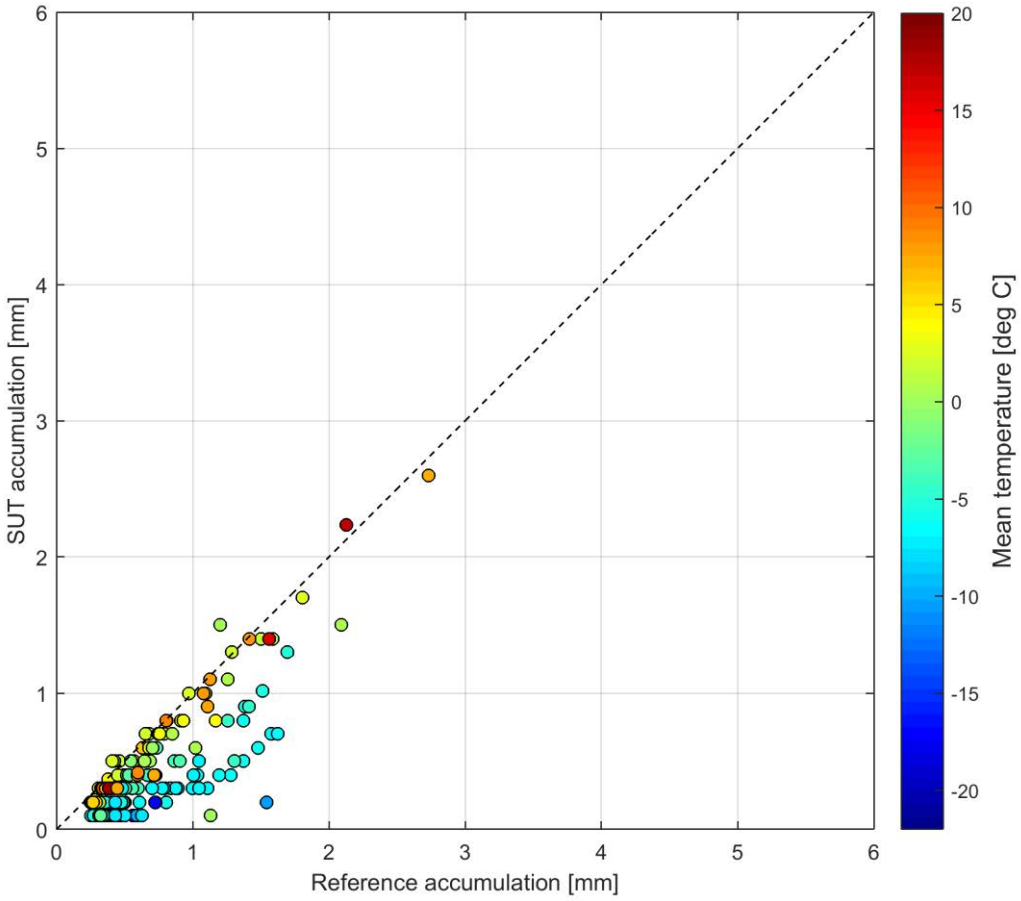
Figure 7: Response delays for Meteoservis MR3H-FC gauges as a function of the mean precipitation intensity observed by the reference gauge during the delay period at (a) CARE, (b) Marshall, and (c) Sodankylä. The predominant precipitation type for each event is determined from the maximum and minimum reported temperature during the delay period.

6.2. Ability to report accumulated precipitation

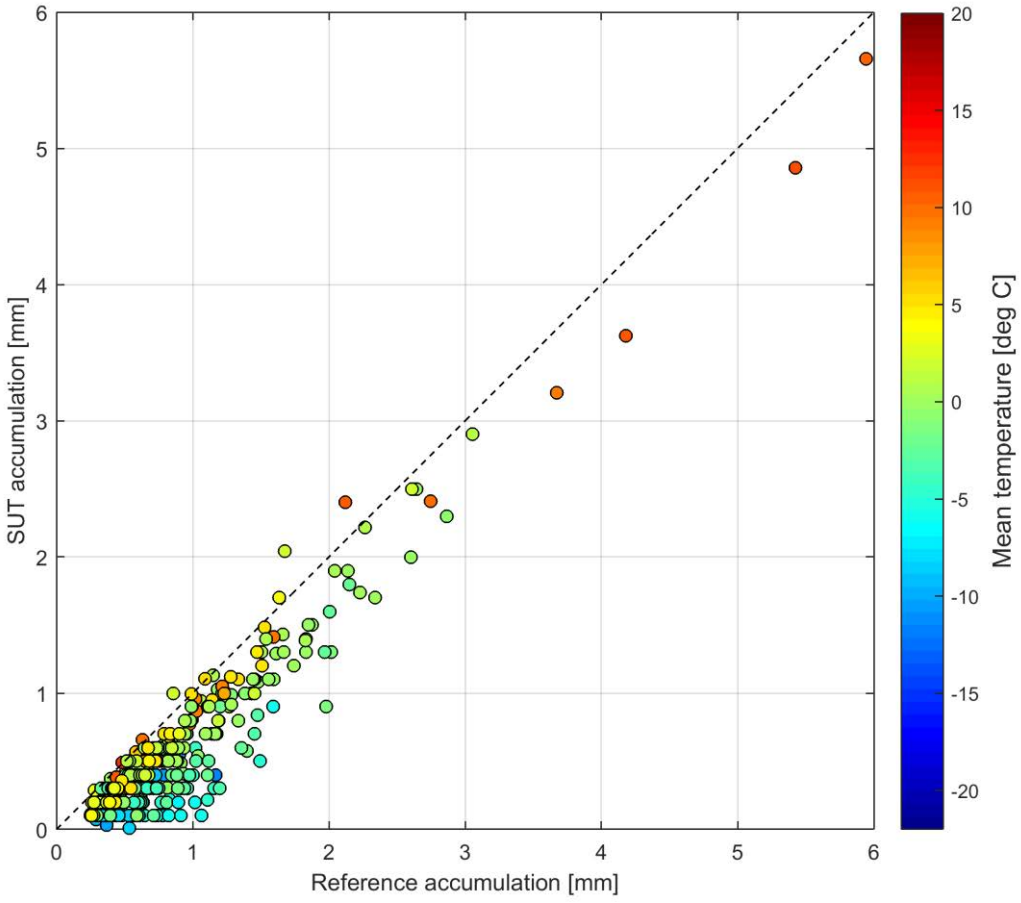
The SUT performance in terms of reporting accumulated precipitation is examined by comparing the amount reported by each sensor under test relative to the respective site reference during a given assessment interval. This is represented graphically using scatter and box plots of the catch efficiency as a function of mean wind speed at gauge height, as well as scatter plots of the amounts reported by the SUT versus the corresponding reference amounts (Figures 8 to 10). The SUT performance is also assessed in terms of the root mean square error, RMSE (Figure 11).

Only assessment intervals during which the SUT and reference both reported precipitation (YY cases) are considered in this portion of the assessment. To assess the influence of the assessment interval on SUT performance, analysis was conducted using both 30 and 60 minute intervals. The plots presented in Figures 8 to 10 include only the 30 minute results, in order to better constrain the wind speed and temperature data used in the assessment, which vary to a greater extent over longer (60 minute) periods.

a)


SUT accumulation [mm]
Reference accumulation [mm]
Mean temperature [deg C]
0
1
2
3
4
5
6
20
15
10
5
0
-5
-10
-15
-20

b)


SUT accumulation [mm]
Reference accumulation [mm]
Mean temperature [deg C]
0
1
2
3
4
5
6
20
15
10
5
0
-5
-10
-15
-20

c)

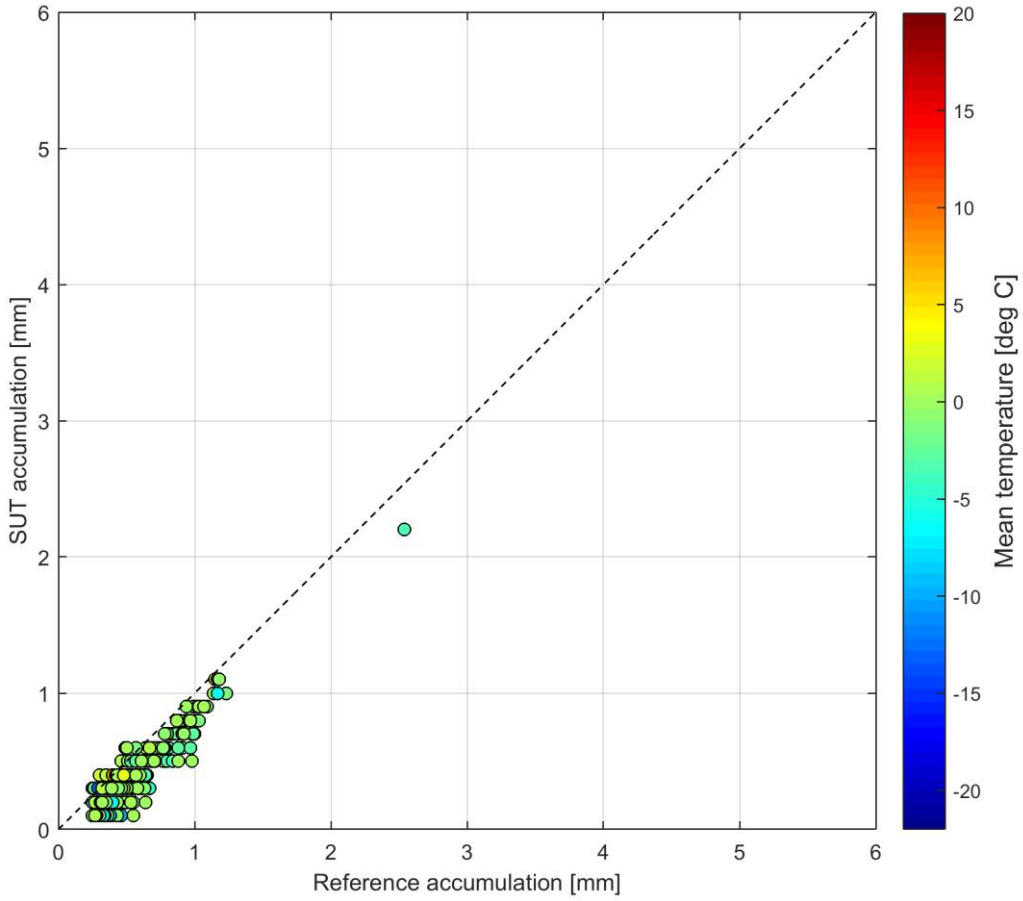
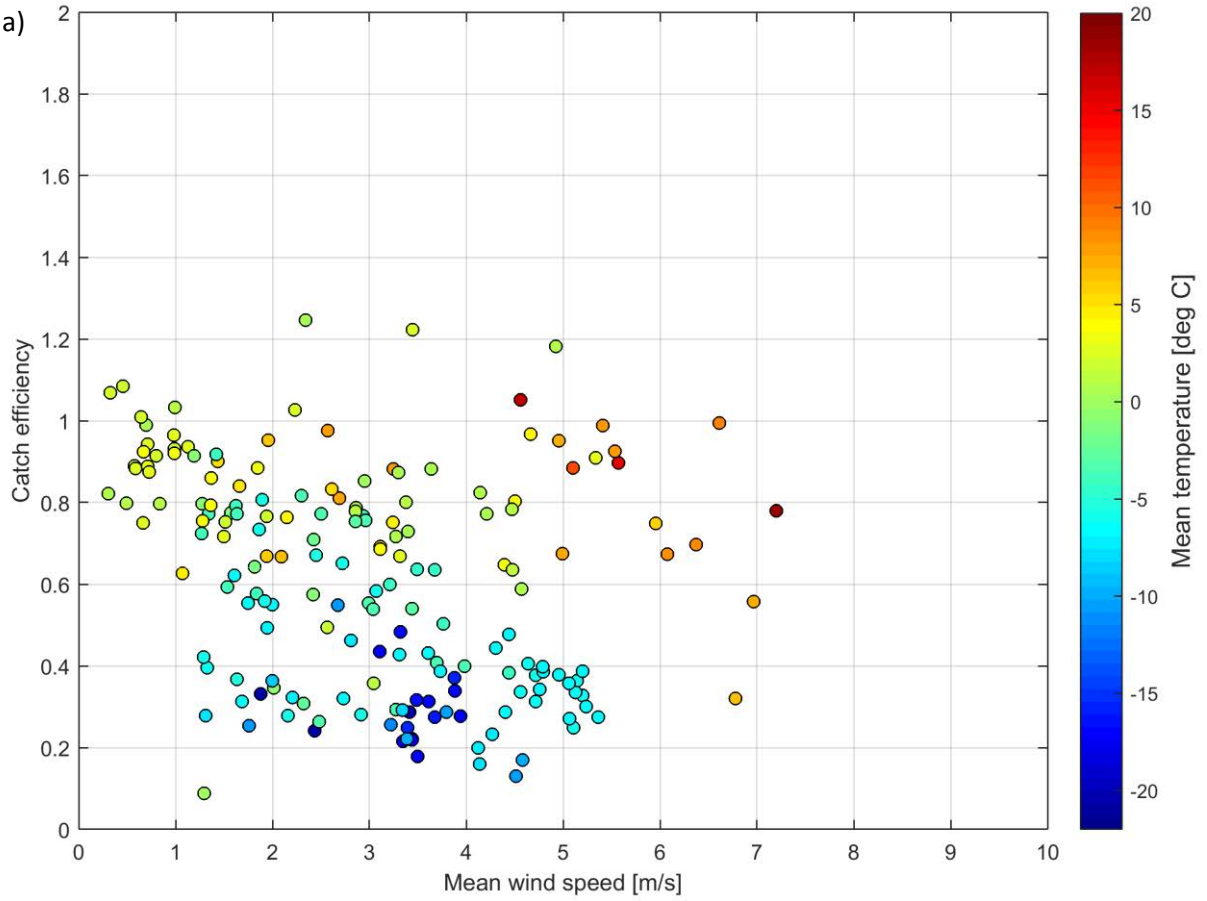
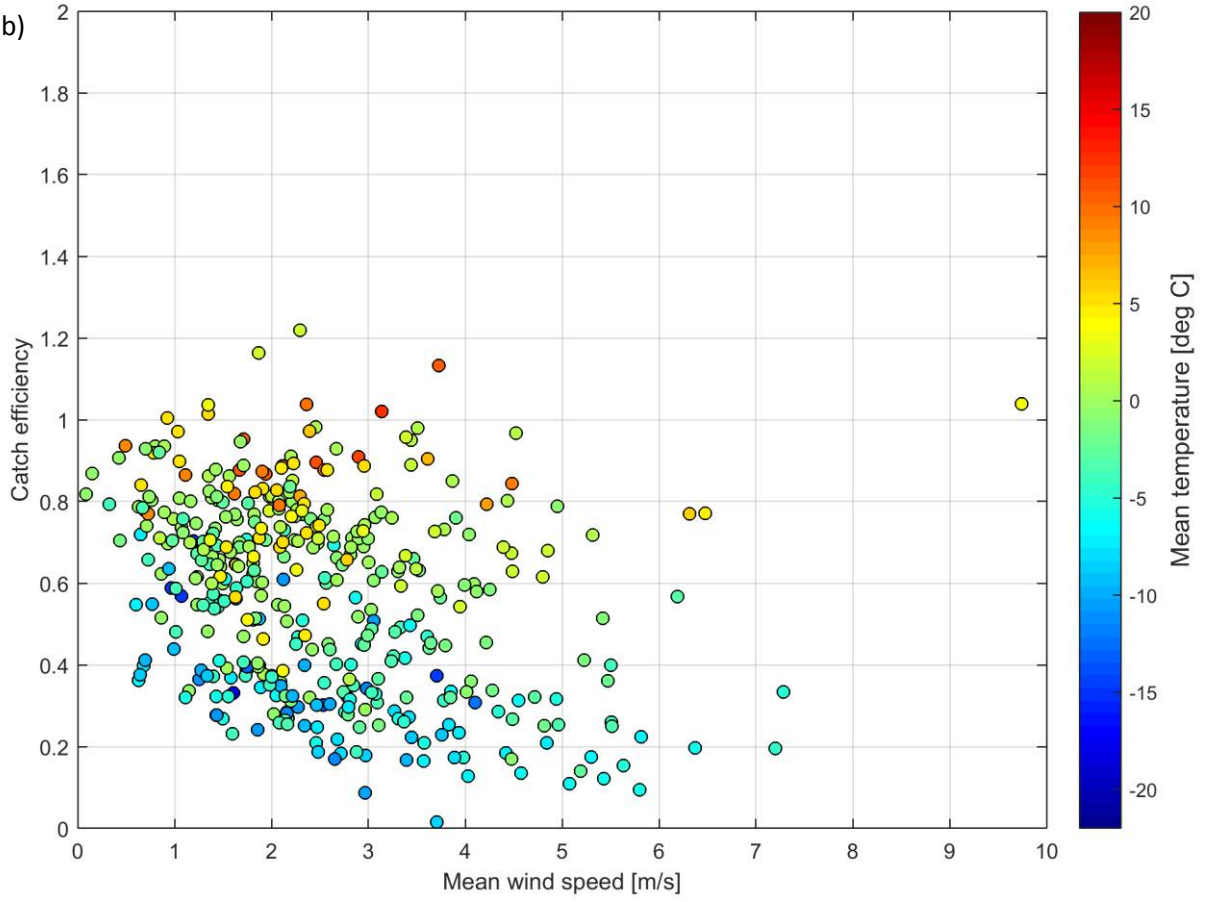


Figure 8: SUT accumulation vs. reference accumulation scatter plots for 30 minute precipitation events at (a) CARE, (b) Marshall, and (c) Sodankylä. The mean event temperature is indicated by colour.

a)


Catch efficiency
Mean wind speed [m/s]
Mean temperature [deg C]

b)


Catch efficiency
Mean wind speed [m/s]
Mean temperature [deg C]

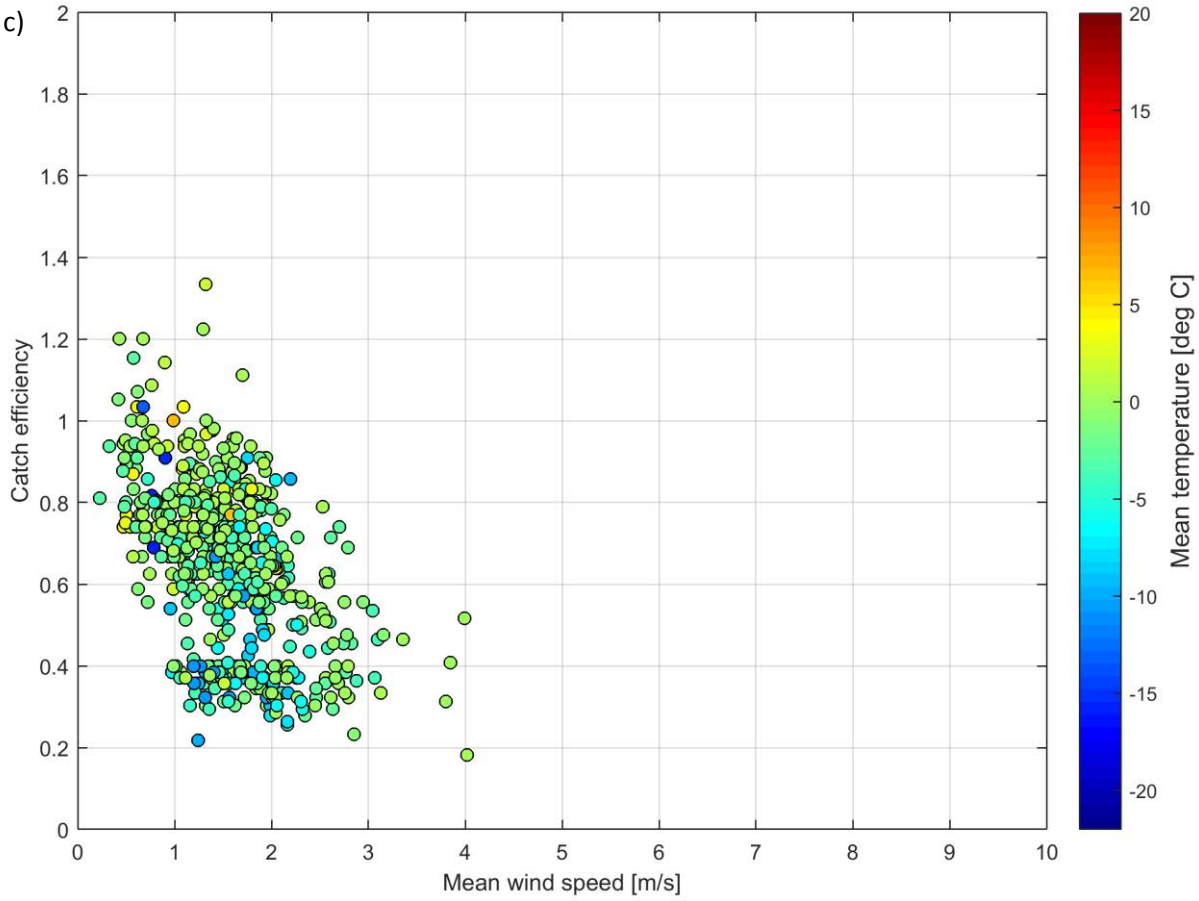
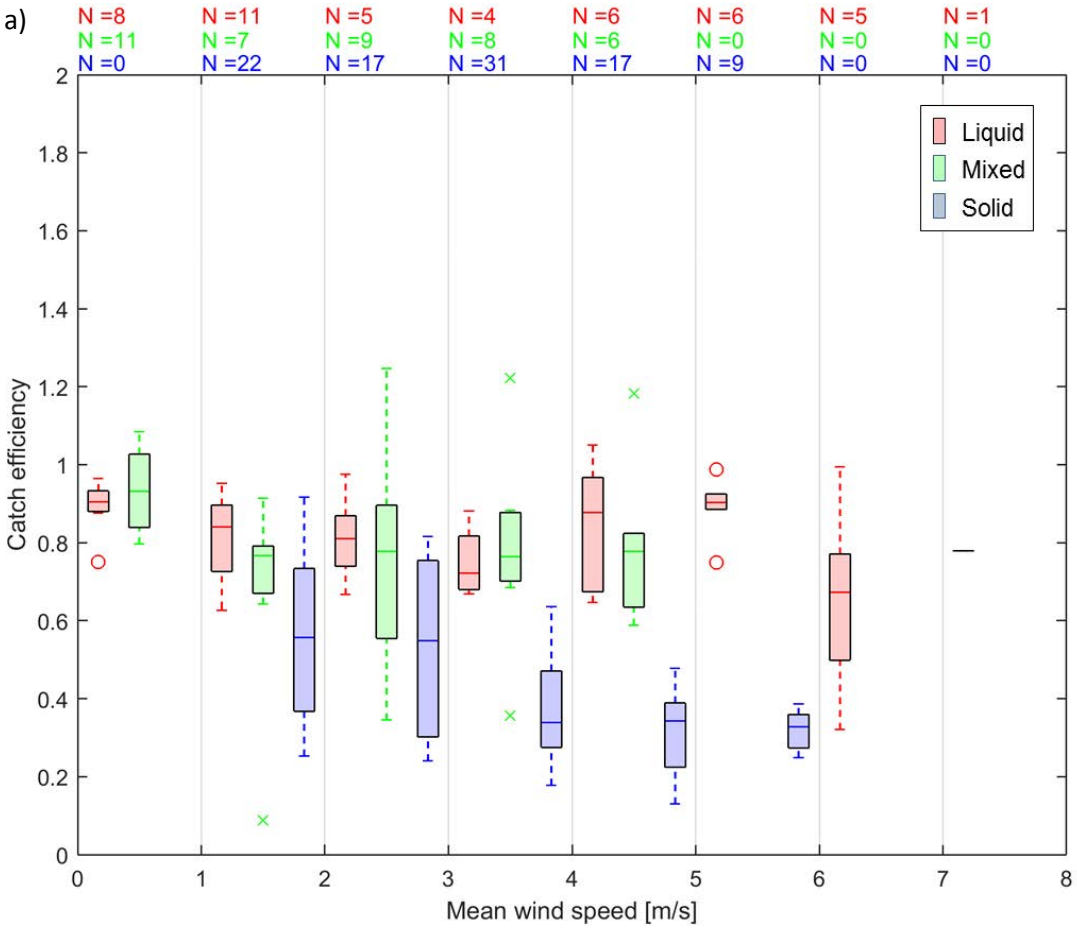
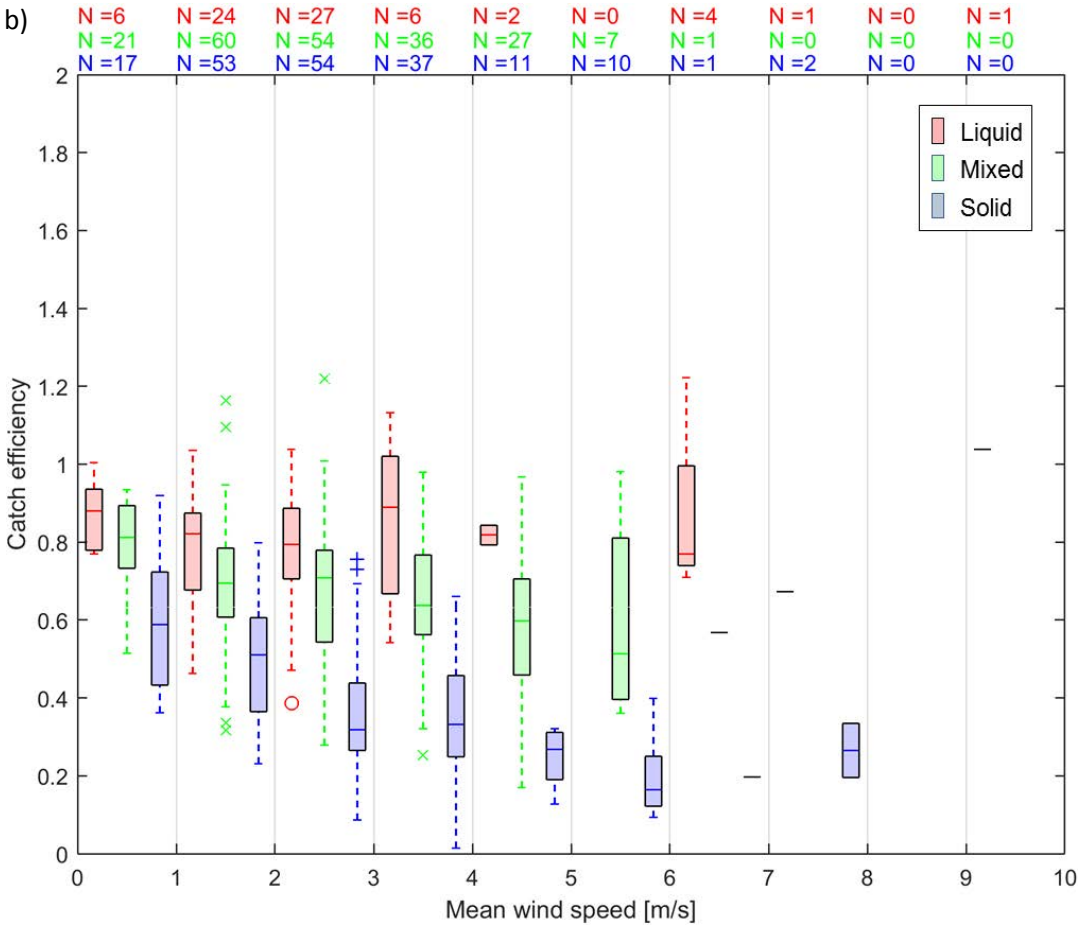


Figure 9: Catch ratio vs. mean wind speed scatter plots for 30 minute precipitation events at (a) CARE, (b) Marshall, and (c) Sodankylä. The mean event temperature is indicated by colour.

a)


N =8 N =11 N =5 N =4 N =6 N =6 N =5 N =1
N =11 N =7 N =9 N =8 N =6 N =0 N =0 N =0
N =0 N =22 N =17 N =31 N =17 N =9 N =0 N =0
Liquid
Mixed
Solid
Catch efficiency
Mean wind speed [m/s]

b)


N =6 N =24 N =27 N =6 N =2 N =0 N =4 N =1 N =0 N =1
N =21 N =60 N =54 N =36 N =27 N =7 N =1 N =0 N =0 N =0
N =17 N =53 N =54 N =37 N =11 N =10 N =1 N =2 N =0 N =0
Liquid
Mixed
Solid
Catch efficiency
Mean wind speed [m/s]

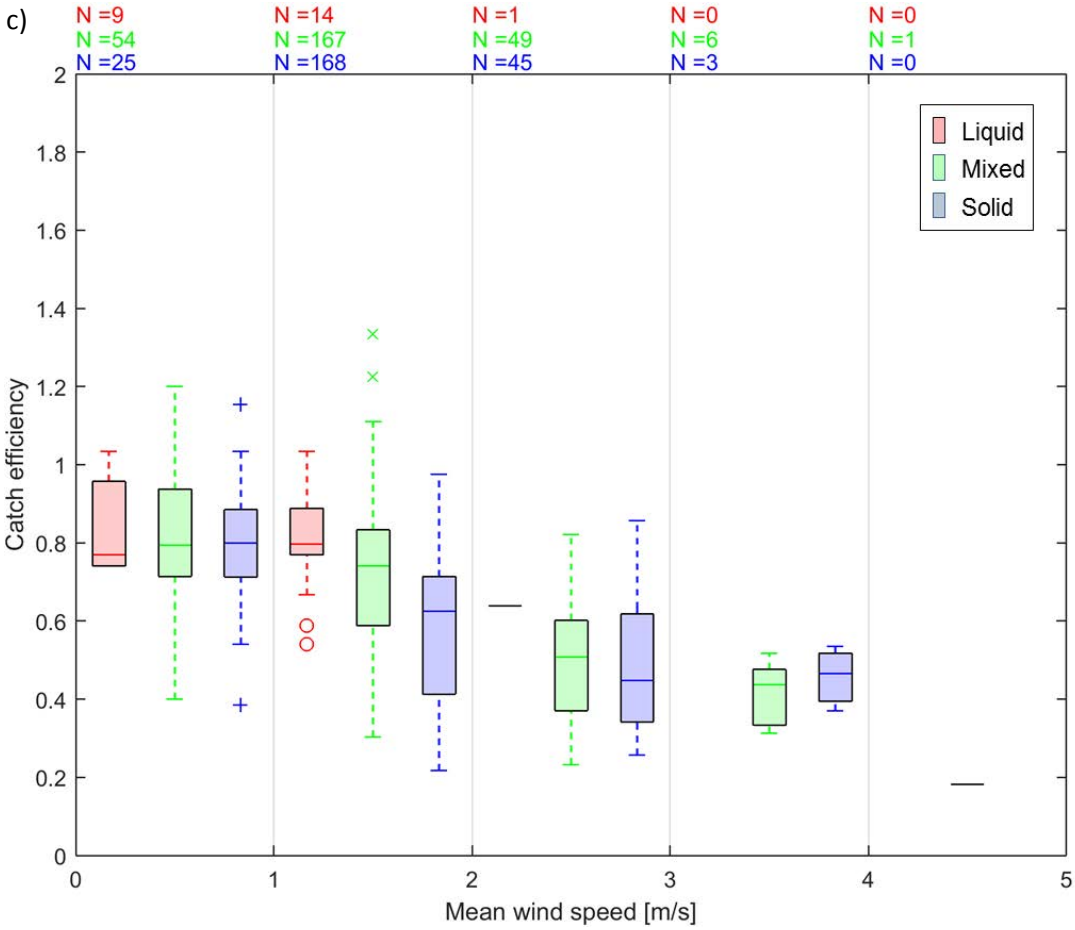


Figure 10: Catch ratio vs. mean wind speed box and whisker plots for 30 minute precipitation events at (a) CARE, (b) Marshall, and (c) Sodankylä. The predominant precipitation type for each event is determined from the maximum and minimum reported temperature and indicated by colour.

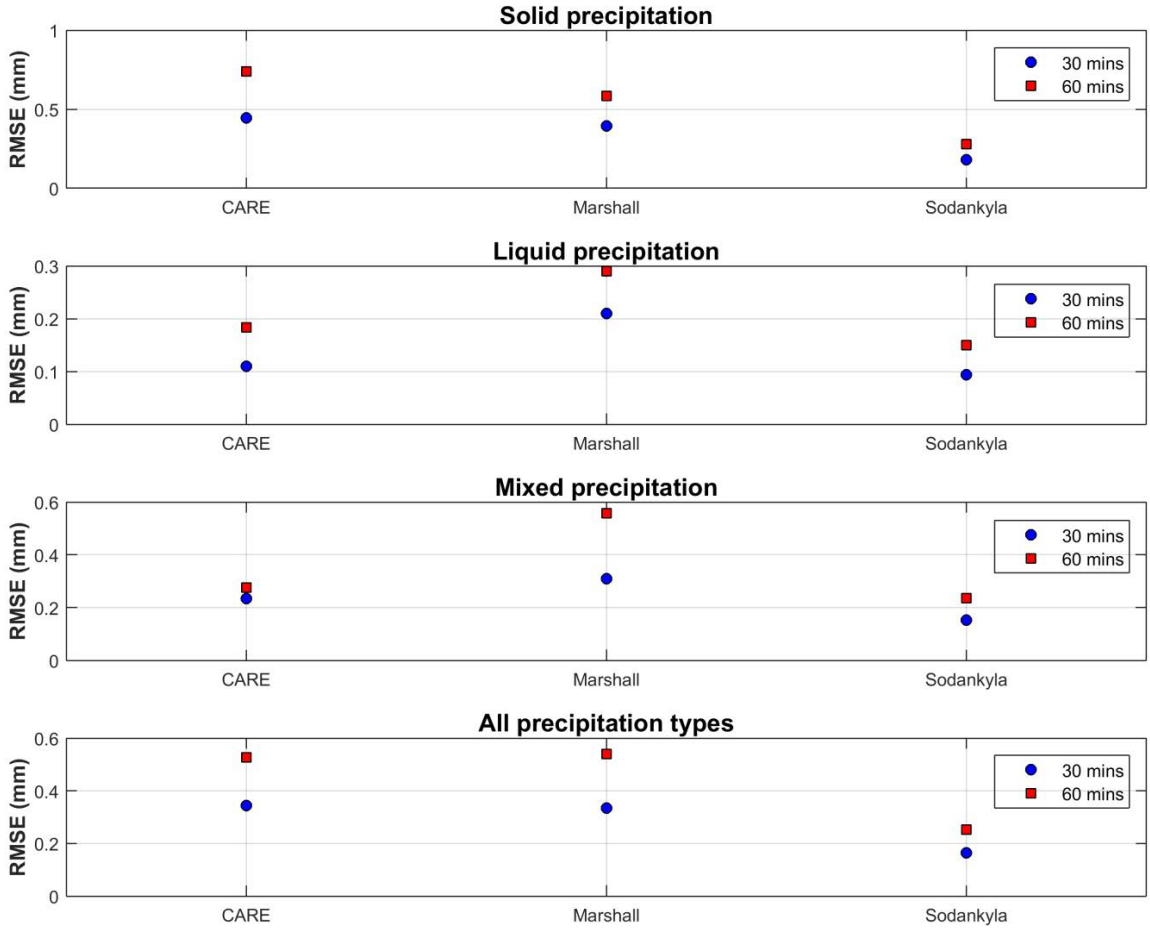


Figure 11: Graphical representation of the RMSE values, by site, and by precipitation type, for YY cases.

The total accumulation reported by each site reference and SUT, for all assessment intervals during which both detected and reported precipitation, are presented alongside the corresponding catch ratios in Table 5.

Table 5: SUT and reference accumulation and overall catch ratio, assessed for YY cases over the entire test period, by site.

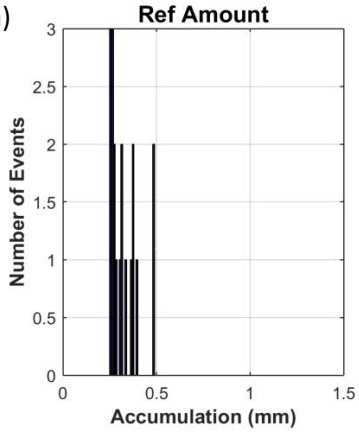
Site	Time interval	SUT accumulation	Reference accumulation	Overall catch ratio
CARE	30 min	77.3 mm	122.2 mm	0.63
	60 min	85.6 mm	141.4 mm	0.61
Marshall	30 min	254.5 mm	383.2 mm	0.66
	60 min	274.3 mm	435.6 mm	0.63
Sodankylä	30 min	162.1 mm	239.0 mm	0.68
	60 min	219.4 mm	354.7 mm	0.62

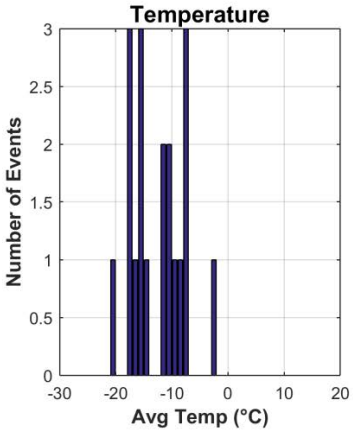
6.3. Assessment of events when the reference and the SUT do not agree on the occurrence of precipitation

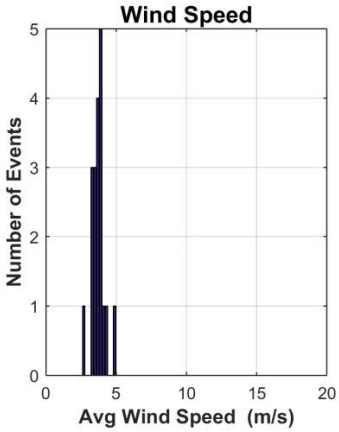
Assessment intervals during which the site reference and SUT do not agree on the occurrence of precipitation – namely, the YN and NY cases (Section 4.1.1) – are characterized using histograms in Figures 12 and 13. The histograms include accumulated precipitation, precipitation intensity as reported by the reference, and corresponding site conditions.

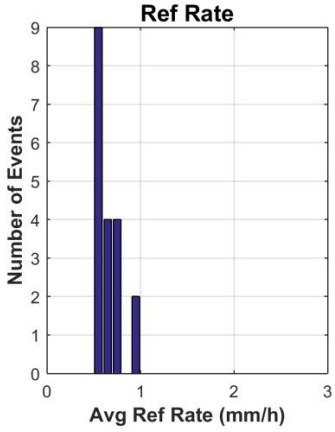
The total SUT and reference accumulations over the test period include contributions from YN and NY cases, which impact the overall catch efficiency. These contributions may be significant, given the response delays associated with heated tipping bucket gauges. Total accumulation and catch efficiency results presented in Table 5 (YY cases only) are expanded to include contributions from YN and NY cases in Table 6.

a)



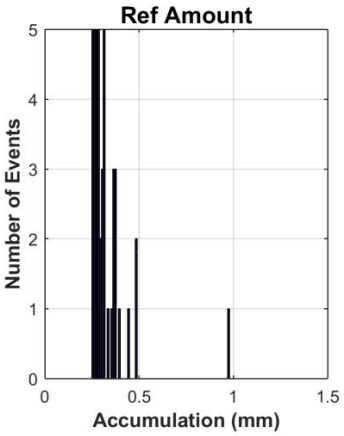


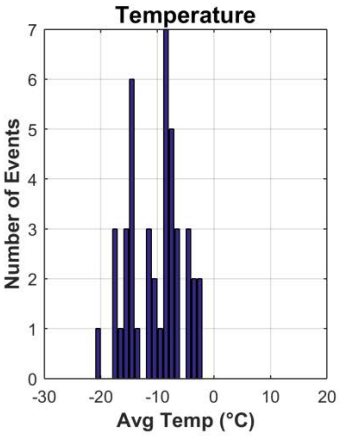


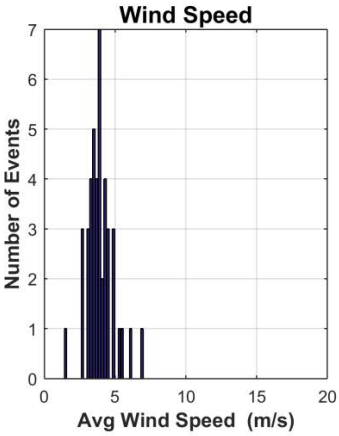


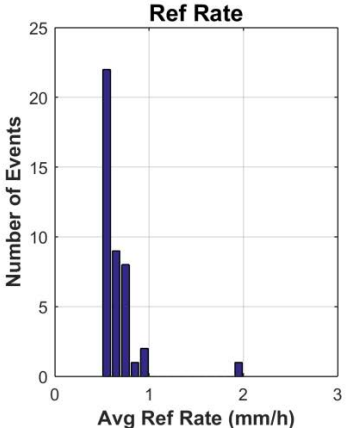
19 events with SUT accumulation = 0 mm (misses)

b)



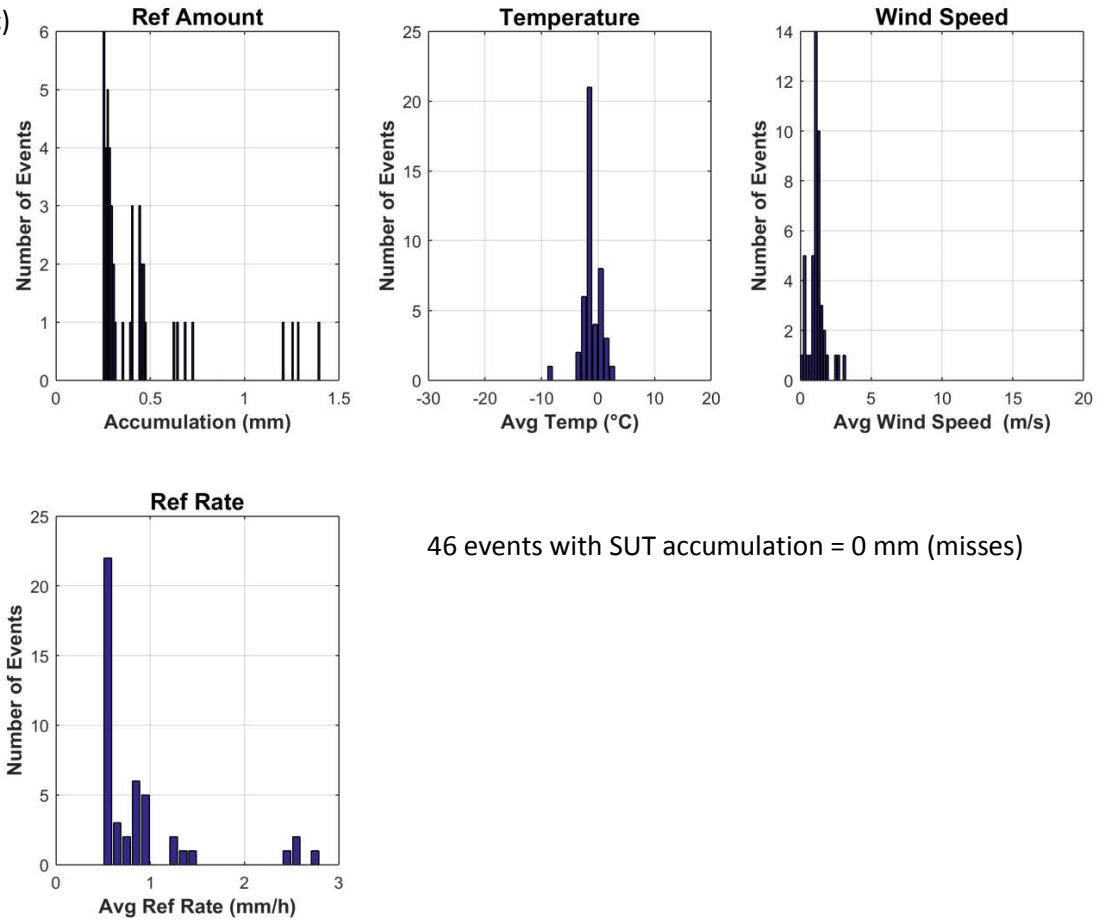






43 events with SUT accumulation = 0 mm (misses)

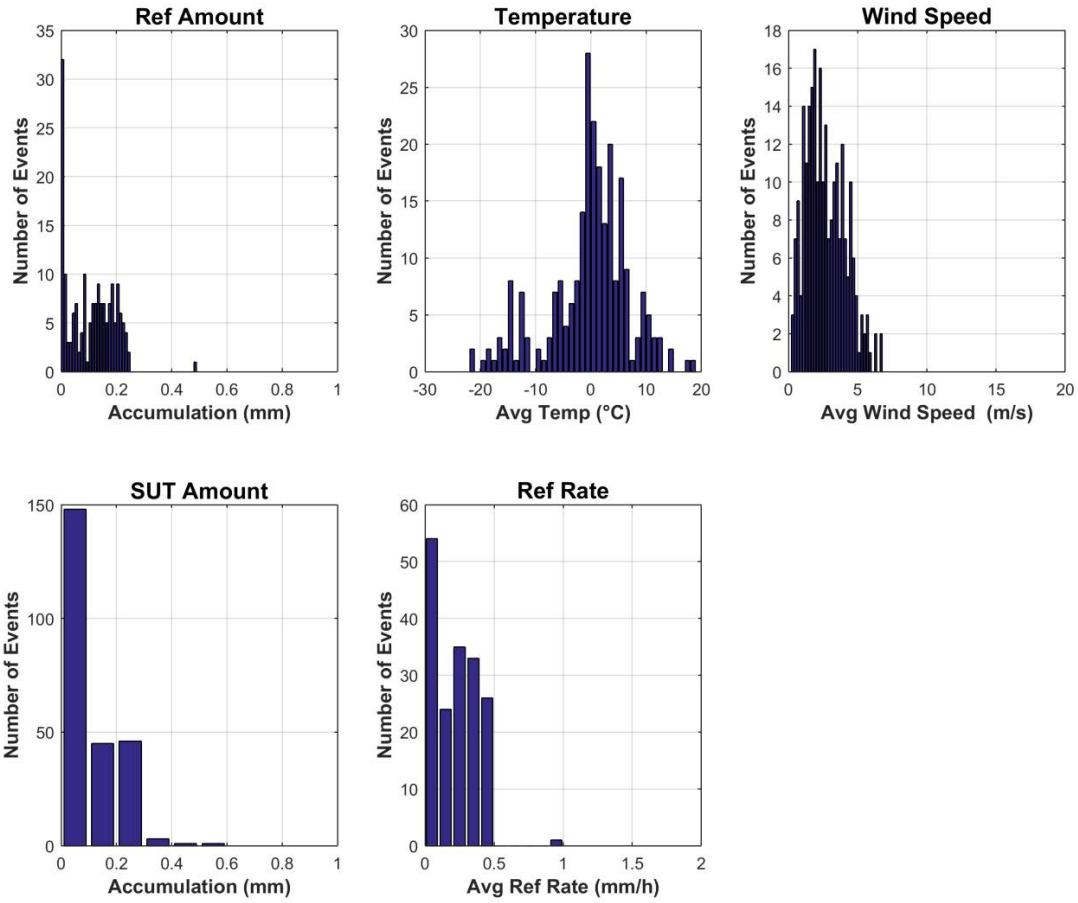
c)



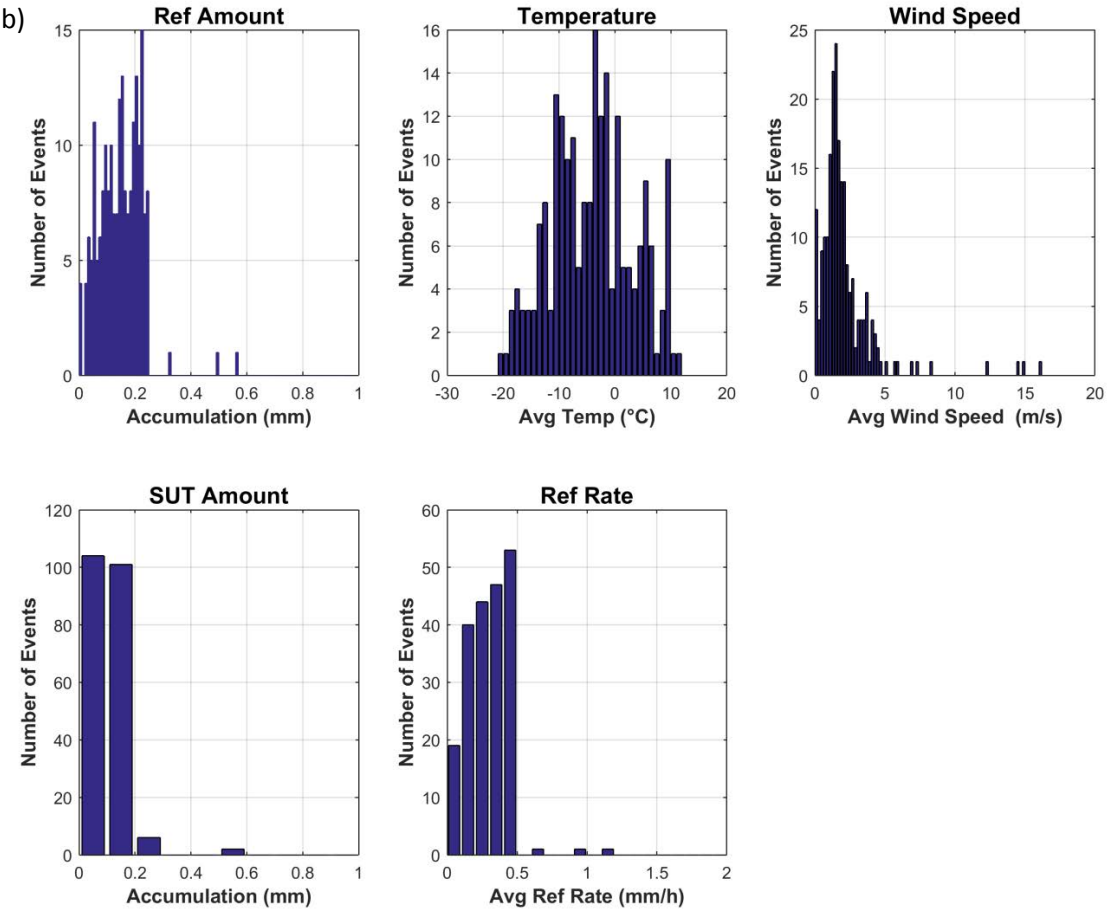
46 events with SUT accumulation = 0 mm (misses)

Figure 12: Histograms of reference accumulation, mean temperature, mean wind speed, and reference precipitation rate for YN cases at (a) CARE, (b) Marshall, and (c) Sodankylä.

a)


Ref Amount
Number of Events
Accumulation (mm)
Temperature
Number of Events
Avg Temp (°C)
Wind Speed
Number of Events
Avg Wind Speed (m/s)
SUT Amount
Number of Events
Accumulation (mm)
Ref Rate
Number of Events
Avg Ref Rate (mm/h)

b)


Ref Amount
Number of Events
Accumulation (mm)
Temperature
Number of Events
Avg Temp (°C)
Wind Speed
Number of Events
Avg Wind Speed (m/s)
SUT Amount
Number of Events
Accumulation (mm)
Ref Rate
Number of Events
Avg Ref Rate (mm/h)

c)

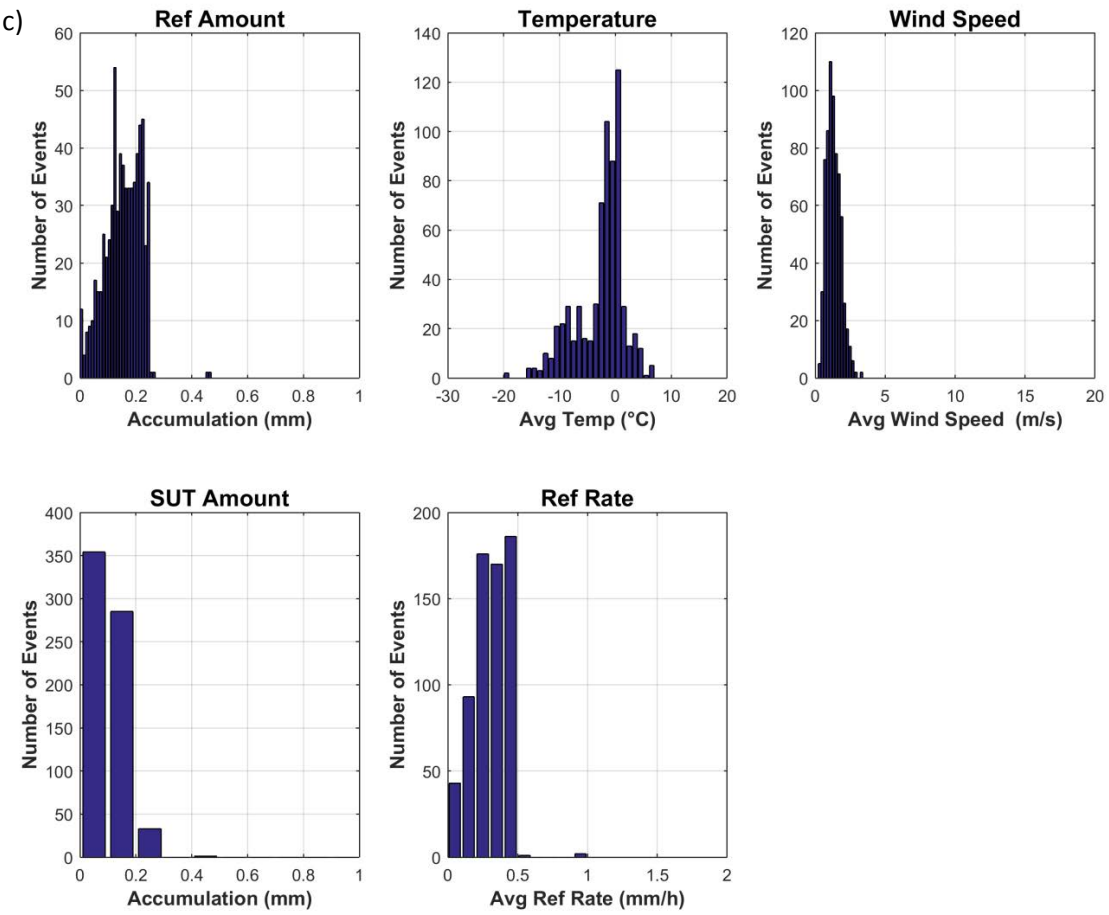


Figure 13: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, and reference precipitation rate for NY cases at (a) CARE, (b) Marshall, and (c) Sodankylä.

Table 6: SUT and reference accumulation and overall catch ratio, assessed for the entire test period, for all YY, YN, and NY cases.

Site	Time interval	SUT accumulation	Reference accumulation	Overall catch ratio
CARE	30 min	108.6 mm	146.7 mm	0.74
	60 min	109.7 mm	160.4 mm	0.68
Marshall	30 min	279.9 mm	431.0 mm	0.65
	60 min	285.1 mm	471.3 mm	0.60
Sodankylä	30 min	240.3 mm	362.2 mm	0.66
	60 min	245.3 mm	422.9 mm	0.58

7. Interpretation of Results

7.1. Operating conditions

The full range of conditions under which the test gauges were operated is illustrated in Figure 3. The conditions relevant to gauge operation are as follows:

- Temperatures from -40 °C to 30 °C;
- Precipitation rates within 12 mm/hr.

These conditions fall within the manufacturer's specified range of operating conditions. It should be noted that without the (optional) outflow heating unit installed for all sensors under test, the observed temperature conditions would extend below the operating range of the instrument (lower limit of -30 °C without the outflow heating unit).

The conditions during precipitation events (as identified by the reference configuration) are shown in Figure 4. The events at Sodankylä are characterized by low wind speeds, typically within 3 – 4 m/s, on account of it being a highly sheltered site. Those at CARE and Marshall are associated with higher wind speeds, typically within 8 m/s. The precipitation rates across all sites were typically within 5 mm/hr.

7.2. Ability to detect and report precipitation

7.2.1. Skill score assessment

The performance of the sensors under test was considered in terms of skill scores, computed for both 30 and 60 minute assessment intervals (Figure 5). Using the shorter 30 minute assessment intervals, the Probability of Detection (POD) is higher for all gauges under test relative to the scores for the longer 60 min assessment intervals, but the False Alarm Rate (FAR) and Bias (B) is also higher for both gauges. As a result, the Heidke Skill Score (HSS) values are higher for both gauges when 60 minute assessment intervals are used. The comparison of skill scores for the test gauges at each site will be considered for only the 60 minute assessment intervals, as hourly reports are broadly used, operationally.

The probability of detecting precipitation was high for the test gauges at each site, with POD values of approximately 86%, 88%, and 91% for the gauges at CARE, Marshall, and Sodankylä, respectively. This follows from the similar distributions of 'hits' and 'misses' across all sites in Table 4, with the highest ratio of hits to misses for the gauge at Sodankylä resulting in the highest POD.

The number of 'false alarms' reported by the test gauge at CARE exceeded the number of 'hits,' resulting in the highest FAR (53%), highest B (184%), and lowest HSS (58%) among the gauges tested. The test gauges at Marshall and Sodankylä reported relatively fewer 'false alarms,' resulting in more comparable FAR of 19% and 28%, B of 108% and 125%, and HSS of 83% and 79%, respectively.

7.2.2. Characterization of response delays

7.2.2.1. Response delay PDFs

Probability distribution functions (PDFs) of response delays between the onset of precipitation as observed by the reference configuration and the first response by the Meteoservis MR3H-FC heated tipping bucket gauge are shown in Figure 6. For the test gauges at all sites, response times within 40 minutes have the highest probability, with lower probabilities of longer delays for the gauges at CARE (up to about 85 minutes) and Marshall (up to about 140 minutes).

For precipitation occurring within a given 30 or 60 minute interval (as indicated by the reference configuration), the response delay PDFs for the test gauges at all sites indicate the potential for the TB response to occur within a subsequent 30 or 60 minute interval. The delays with highest probability in the PDFs are within 40 minutes, so it is more likely that the TB gauges will respond to precipitation detected by the reference within a 60 minute period than within a 30 minute period. This is reflected by the lower number of false alarm (NY) cases for the 60 minute intervals relative to the number of false alarms for the 30 minute intervals at each site in Table 4, and the lower FAR and Bias observed for the gauges at each site for 60 minute intervals in Figure 5.

Another important consideration is that even when both the SUT and reference detect precipitation within a given interval, the TB may actually be reporting precipitation collected during previous intervals, or may only report a fraction of the precipitation within that interval (reporting the rest during a subsequent interval). These ‘false YY cases’ will impact the assessment of TB gauge performance in terms of reporting accumulated precipitation (Section 6.2).

7.2.2.2. Influence of precipitation intensity, phase

The dependence of delay times on the mean intensity of precipitation preceding the TB response, as reported by the reference configuration, is illustrated in Figure 7. For the gauge under test at Sodankylä (Figure 7c), response delays are within 40 minutes for all precipitation types (liquid, mixed, solid), over the full range of reported intensities (up to about 0.75 mm/hr). For the test gauges at CARE and Marshall (Figures 7a and 7b, respectively), delay times were within 30 minutes for all precipitation types, for mean intensities within 0.5 mm/hr and 0.45 mm/hr, respectively. At higher mean intensities, longer delays are observed for solid precipitation events at CARE, and for mixed and solid precipitation events at Marshall, which are attributed to melting or evaporative delays due to heating. The mixed-phase events with longer delay times at Marshall may represent precipitation with a solid component, or a transition between phases; the classification of events by phase using temperature is not absolute.

7.3. Ability to report accumulated precipitation

7.3.1. Performance when both SUT and reference detect precipitation

7.3.1.1. Temperature, phase, and wind speed influence

Focussing on the 30 minute events in which both the sensors under test and site reference configurations report precipitation (YY cases, or 'hits,' only), general stratification of the data by temperature is observed for CARE and Marshall. In the SUT accumulation vs. reference accumulation plots (Figures 8a and 8b), the points further from the 1:1 line (which indicates perfect agreement) generally correspond to colder temperatures. Similarly, the catch efficiency vs. wind speed plots for these sites (Figures 9a and 9b) generally show lower catch efficiencies for events with lower mean temperatures. These trends are less evident in the plots for Sodankylä, in which the points are tightly clustered within a narrow range of accumulations (Figure 8c) and wind speeds (Figure 9c).

The observed trends result from the influence of wind speed on the ability of the sensors to collect and record the mixed-phase and solid precipitation that occur at colder temperatures. Acknowledging the scatter in the data, a general trend in which the catch efficiency decreases with increasing wind speed is observed in Figure 9.

The phase and wind speed influences on catch efficiency are more clearly delineated in the box and whisker plots in Figure 10. The median catch efficiencies for solid precipitation range from about 0.6 (Marshall) to 0.8 (Sodankylä) for wind speeds below 1 m/s, decreasing to 0.3 – 0.4 (all sites) for wind speeds between 1 and 4 m/s, and as low as about 0.2 (Marshall) for wind speeds > 4 m/s. These general trends for snow, while broad, characterize the maximum expected wind speed effects on catch efficiency. As observed in Figure 10, the trends for mixed precipitation fall within these limits, and are dictated by the relative proportions of liquid (less expected wind influence) and solid (greater expected wind influence) precipitation.

7.3.1.2. Root mean square error

The RMSE was calculated for all 30 and 60 minute events during which the reference and SUT both detected precipitation (YY cases), for each site. The overall RMSE calculated for each site (all precipitation types), and RMSE values calculated for each precipitation type, are shown in Figure 11. Given the broad use of hourly reports in meteorological operations, further discussion of RMSE values will be considered for 60 minute assessment intervals only.

The overall RMSE was within 0.6 mm for the test gauges at CARE and Marshall, and within 0.3 mm for the test gauge at Sodankylä. These values can be considered to represent the absolute uncertainty of the Meteoservis MR3H-FC relative to a DFAR configuration for the specific site and gauge conditions tested, over the range of conditions tested. The lower overall RMSE observed for the gauge at Sodankylä is attributed to the sheltered nature of the test site, which results in lower wind speeds and hence, less wind-induced undercatch relative to the other sites.

In general, the phase-specific RMSE values for all sites follow the trend solid > mixed > liquid, owing to the corresponding relative influence of wind-induced undercatch on precipitation of each phase.

For a given precipitation phase, the differences in RMSE among the test gauges are attributed to differences in gauge siting and configuration, as well as to differences in the specific conditions at each site.

7.3.1.3. Overall catch efficiency

The total accumulated precipitation recorded by the reference and sensor under test at each site was compiled for all 30 and 60 minute events during which both detected precipitation (YY cases) and used to calculate the overall catch ratio. The results are provided in Table 5. The more operationally-relevant 60 minute values illustrate similar gauge performance across all sites, with overall catch efficiencies of 0.61, 0.63, and 0.62 for the test gauges at CARE, Marshall, and Sodankylä, respectively.

7.3.2. Characterization of conditions when the SUT and reference do not agree on the occurrence of precipitation

The conditions during 30 minute events in which the reference detected precipitation, but the SUT did not (YN cases), are depicted as histograms in Figure 12. These events are characterized by low precipitation rates, typically below 1 mm/hr, low wind speeds, typically below 5 m/s, and temperatures close to, or below, freezing. At these precipitation rates, response delays can result from the time required for heating to melt mixed or solid precipitation at the observed temperatures. Heating may also lead to the evaporation or sublimation of incident precipitation at low intensities, increasing delay times by taking longer to fill the bucket.

The conditions during 30 minute events in which the SUT detected precipitation, but the reference did not (NY cases), are shown as histograms in Figure 13. These events are also characterized by low precipitation rates, typically below 0.5 mm/hr (which corresponds to the reference accumulation threshold for a precipitation event over a 30 minute interval), and low wind speeds, typically below 5 m/s. These events most likely result from delayed tipping bucket responses that coincide with intervals during which the reference did not detect precipitation.

Given the influence of response delays, estimates of the overall catch ratio for tipping bucket gauges should include the reference and SUT accumulations during YN and NY cases (and not just during the YY cases, as considered above). To account for the influence of response delays observed by heated TB gauges, estimates of the overall catch ratio should include the reference and SUT accumulations during these events. The resulting total accumulations for all YY, YN, and NY cases, and the corresponding overall catch efficiencies, are provided for each site in Table 6. The overall catch efficiencies for the gauges at CARE, Marshall, and Sodankylä are 0.68, 0.60, and 0.58, respectively, for the more operationally-relevant 60 minute assessment intervals.

The higher catch efficiency for CARE is attributed to the higher proportion of liquid precipitation events observed at this site. The breakdown of events by phase is not detailed here, but showed the following: 29% liquid for CARE, 16% for Marshall, 6% for Sodankylä. The similarity of the catch efficiencies for the gauges at Marshall and Sodankylä can perhaps be attributed to offsetting influences on catch efficiency, with more liquid events/higher wind speeds at Marshall and relatively fewer liquid events/lower wind speeds at Sodankylä.

8. Operational considerations

The overall experience with the Meteoservis MR3H-FC across the test sites was positive. The gauge was easy to install, easy to connect, and the setup of data collection was straightforward.

The gauge heaters can require significant power, up to 555 W (the theoretical maximum value specified by the manufacturer). The manufacturer notes that it is possible to operate the gauge long-term using only 35% of the maximum theoretical value. In general, the power source should be located as close as possible to the gauge to mitigate voltage drops and the potential reduction of heating power and efficiency. Investment in site infrastructure may be required to ensure sufficient power is provided to the gauge.

8.1. Maintenance

Each site completed the gauge field calibration and verification as per manufacturer recommendations, at least once a year. The calibration records have been stored by each site team.

8.2. Noted issues

- The site team at Sodankylä noted that the funnel can become filled with snow, resulting in the potential for incident precipitation to not be counted, or for the loss of precipitation from the funnel by wind, prior to melting (both may result in accumulation losses).

- The site team at Marshall noted issues with clogging related to the design of the spring in the orifice. The heater appeared sufficient to keep the orifice ice-free during winter precipitation events; however, ice build-up was observed on screens in the 'emptying' ports (internal icing).

- No issues with clogging were reported during the SPICE test seasons at CARE; however, the site team noted that the orifice was clogged regularly during summer operation (primarily with bird droppings), and had to be cleaned out roughly twice per month.

9. Performance Considerations

- Heating is critical to ensure proper functioning of the gauge under winter conditions. The voltage should be checked at the gauge location to ensure that the heaters are supplied with the recommended power. Voltage drops – for example, resulting from the use of long power cables – can impair the functioning of the gauge in winter conditions.

- Heating enables the measurement of solid precipitation, but the time required for melting can delay the time between the collection of precipitation in the funnel and the gauge response to that precipitation (response delays). Further, heating may result in the evaporation/sublimation of incident precipitation at low intensities.

- Response delays must be considered when using the gauge in operational settings. The reporting interval (i.e. hourly observations) should exceed the maximum expected response delay; however, the potential remains for carry-over of precipitation accumulation from previous intervals, and for delayed responses occurring in subsequent intervals.

- While not tested in SPICE, shielding of the gauge should be considered by the manufacturer and/or potential users as a means of increasing the catch efficiency at higher wind speeds.

- The collection of ancillary measurement data is recommended: air temperature and wind speed (at gauge height) to enable the application of adjustments to measurements (i.e. using transfer functions); and reports from a precipitation detector with high sensitivity to enable the identification of missed events or false alarms due to response delays.

- The application of transfer functions to gauge measurements, if available, is recommended as a post-processing step to account for reductions in catch efficiency with increasing wind speed. This recommendation comes with the caveat that the catch ratio data used in the derivation of the transfer functions will be impacted by response delays. Response delays can lead to scatter in the catch ratio data (among other factors), and hence, uncertainty in the transfer functions; however, the potential remains for the improvement of data via adjustment. Considering the catch ratios over longer time periods (e.g. 1 hour) may provide one avenue for reducing scatter in the catch ratio/wind speed relationship and the uncertainty of related transfer functions.

WMO-SPICE Instrument Performance Report
Meteoservis MR3H-FC, ZAMG version

1) Technical specifications (from manufacturer provided documentation)

Instrument model:	Meteoservis MR3H-FC, ZAMG version
Physical principle:	Heated tipping bucket (TB) gauge. Tips occur when the bucket fills to capacity, triggering a reed contact that induces an output pulse.
Bucket capacity:	0.1 mm
Collecting area:	500 cm^2
Heating configuration:	Heaters at 3 sections – funnel, middle part of funnel, and top of funnel/collar – controlled independently by temperature sensors on the internal side of the gauge. In the absence of recorded precipitation, the heating is limited to reduce the potential influence of evaporation. 'Shock heating' is applied to the top of funnel/collar section after precipitation is recorded to enable the removal of possible icing. As tested, equipped with additional outflow heating unit (AH-01) that is operated independently from other heaters. This unit lowers the minimum operating temperature from -30 °C to -40 °C. Total heating power: 555 W (maximum).
Operating temperature range:	-40 °C to 60 °C (as tested); -30 °C to 60 °C without AH-01 outflow heating unit
Measurement range:	0 – 400 mm/hr
Measurement uncertainty:	Within ± 2% from 0 – 400 mm/hr
Sensitivity:	0.1 mm reporting resolution (corresponding to bucket amount)

a)

b)

Figure 1: ZAMG MR3H-FC installations at (a) CARE and (b) Weissfluhjoch.

2) Data output format

Gauge data output: Gauge reports precipitation intensity in mm/hr, determined from modified 5 V pulses from a reed contact with constant length of 100 ms. Gauge output is corrected internally as a function of precipitation intensity.

Analysis is based on accumulated precipitation, computed as the cumulative sum of accumulations derived from intensity reports.

3) SPICE test configuration

Shield: Unshielded

Test sites: CARE (Canada); Weissfluhjoch (Switzerland)

Sensor provider(s): Zentralanstalt für Meteorologie und Geodynamik (ZAMG); Central Institute for Meteorology and Geodynamics, Austria.

Figure 2: Map of SPICE sites where ZAMG MR3H-FC gauges were tested.

A summary of the configuration of instruments as tested, the duration of tests and availability of data reflected in these results, and the ancillary measurements used, by site, is available in Tables 1, 2, and 3, respectively.

Table 1: Summary of gauge configurations and data output, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	CARE	Weissfluhjoch
Field configuration*	Unshielded	Unshielded
Height of installation (gauge rim)	1.5 m	3.5 m
SUT data output frequency	1 min	1 min
Data QC	SPICE QC methodology	
Data temporal resolution	1 min	
Processing interval for SPICE data analysis	30 min, 60 min	

**Additional configuration details:*

Serial numbers of instruments tested: 856 (CARE); 762 (Weissfluhjoch).
Serial numbers of control units tested: CAR 250/07 (CARE); CAR 133/06, CAR 246/07 (Weissfluhjoch).
Control parameter settings tested: 20090327 (both sites).

Table 2: Data availability, by measurement season and site.

Measurement season	CARE	Weissfluhjoch
Season 1 (Oct. 2013 – Apr. 2014)	X	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓	✓

Data collected at CARE during the 2013-2014 season are excluded from the present assessment due to issues with the gauge configuration that were identified following the conclusion of the measurement season. These issues are the responsibility of the site team, and do not reflect any shortcomings of the instrument or issues with its operation.

Table 3: Summary of reference and ancillary measurements, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	CARE	Weissfluhjoch
R2 site reference	Geonor T-200B3 600 mm (DFAR)	OTT Pluvio^2 – Bucket RT (DFAR)
R2 precip detector	Thies LPM (DFAR)	Thies LPM (DFAR)
Ancillary temp sensor	Vaisala HMP155 (Stevenson screen)	Meteolabor AG VTP6 Thygan (5 m)
Ancillary RH sensor	Vaisala HMP155 (Stevenson screen)	Meteolabor AG VTP6 Thygan (5 m)
Ancillary wind sensor	Vaisala NWS 425 (2 m)	R.M. Young 05103 wind monitor (5.5 m)

4) Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6.1 of the WMO-SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute assessment intervals:

- Ref 'Yes' : R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No' : R2 weighing gauge < 0.25 mm AND/OR precip detector recording < 18 min of precip;
- SUT 'Yes' : SUT accumulation > 0 mm;
- SUT 'No' : SUT accumulation = 0 mm.

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores.

4.1.2. Skill score assessment

The ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

The above scores are computed using the formulations provided in Section 3.6.1.3 of the WMO-SPICE Final Report.

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

4.2. Sensor-specific considerations: response delays

Tipping bucket gauges require that an amount of precipitation corresponding to the bucket capacity is accumulated before a tip is triggered and the gauge records precipitation. This can result in response delays relative to the reference configuration. Heated tipping bucket gauges are subject to further delays, as any solid precipitation in the funnel must be melted before reaching the bucket and potentially triggering a tip, and the heating itself can potentially evaporate incident precipitation. These response delays will impact the comparison with the reference. For this reason, the assessment of TB gauge performance relative to the reference configuration is also considered over 60 minute intervals, using the same conditions and thresholds outlined above in Section 4.1.1.

Response delays are quantified by determining the time elapsed between the onset of precipitation as determined by the reference configuration, and the first tip recorded by the TB. The assessment is based on periods with at least 30 minutes of precipitation, as identified by the reference configuration, followed by at least 180 minutes without precipitation. This extended period without precipitation is intended to allow additional time for the melting and recording of precipitation by heated TB gauges. Additional details are provided in Section 3.6.1.4.4 of the WMO-SPICE Final Report.

5) Environmental conditions

The environmental conditions at each site over the duration of the test period are expressed as probability density functions (PDFs) of mean air temperature, mean relative humidity, mean wind speed, vector mean wind direction, and precipitation rate for each component 30 minute assessment interval in Figure 3. Figure 4 presents the same parameters for all assessment intervals during which the site reference configuration detected precipitation (i.e. all Ref 'Yes' cases).

The precipitation percentage represents the number of minutes of precipitation during a 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration, expressed as a percentage. PDFs of precipitation percentage are also included in Figures 3 and 4.

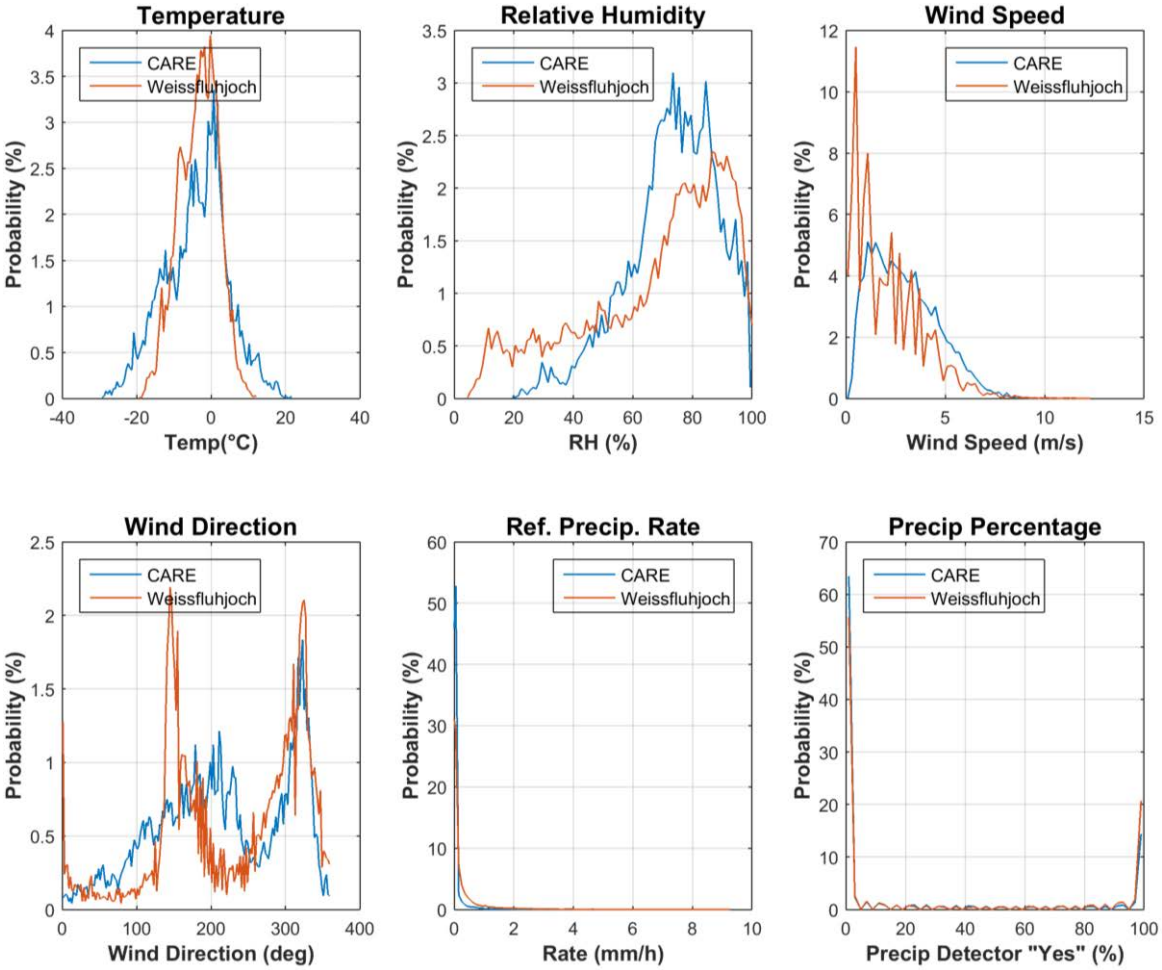


Figure 3: Summary of aggregated environmental conditions at the SPICE sites that operated ZAMG MR3H-FC gauges, over the entire duration of formal tests, as per Table 2.

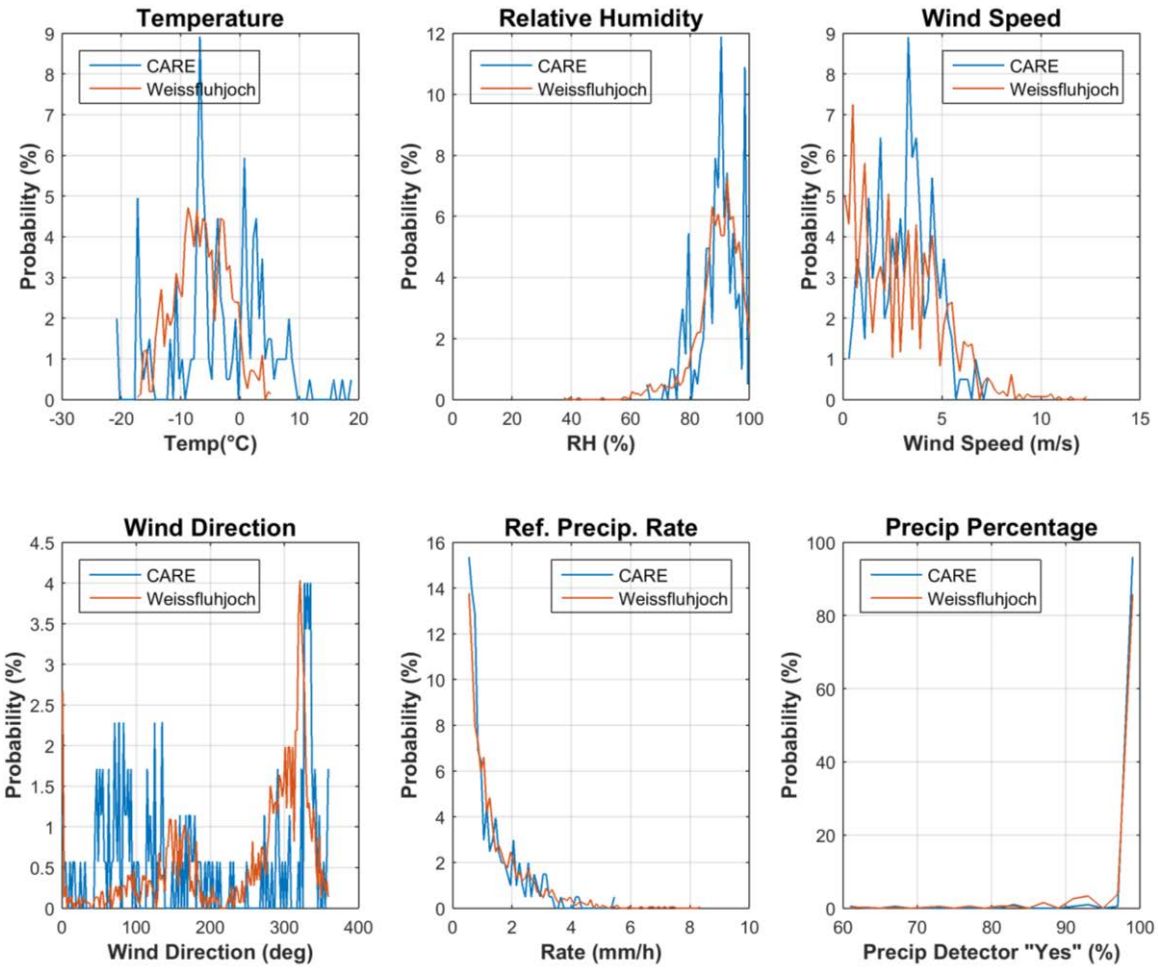


Figure 4: Summary of the aggregated environmental conditions at the SPICE sites that tested ZAMG MR3H-FC gauges during precipitation events, as reported by the site R2 reference, during the formal tests, as per Table 2.

6) Evaluation of performance over the range of operating conditions

6.1. Ability to detect and report precipitation

6.1.1. Skill score assessment

The overall ability of the SUT to detect and report the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores (Section 4.1.2) and presented in Figure 5. Scores are presented for both 30 minute and 60 minute assessment intervals. The contingency results (Section 4.1.1) corresponding to these scores are presented in Table 4.

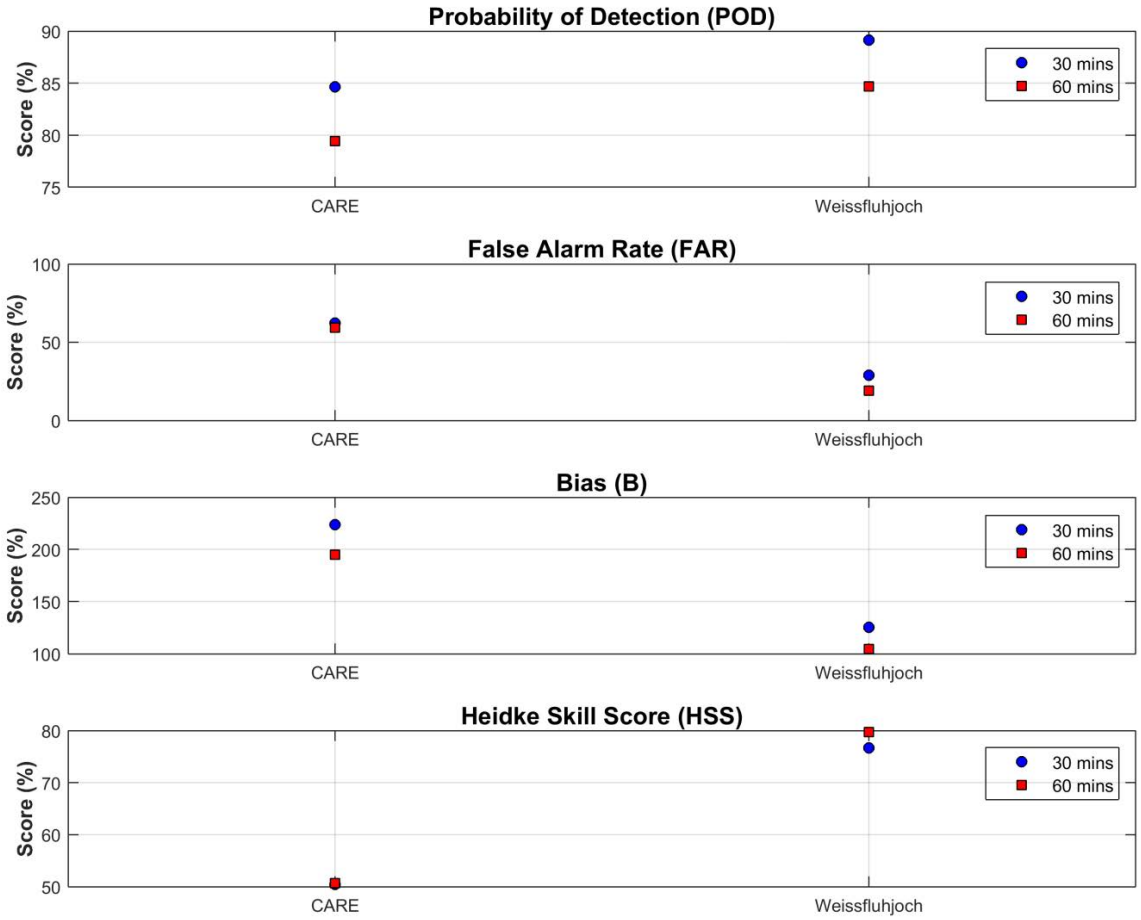


Figure 5: Skill scores for ZAMG MR3H-FC gauges during the formal test periods.

Table 4: Contingency table illustrating detection of precipitation by the ZAMG MR3H-FC relative to the specific site reference, expressed as number of events over the entire test period, by site.

Site	Time interval	Number of Events (% of Events)			
		YY, hits	YN, misses	NY, false alarms	NN, correct negatives
CARE	30 min	171 (2.3%)	31 (0.4%)	281 (3.8%)	6881 (93.4%)
	60 min	143 (3.9%)	37 (1.0%)	208 (5.7%)	3285 (89.4%)
Weissfluhjoch	30 min	1417 (8.4%)	173 (1.0%)	575 (3.4%)	14622 (87.1%)
	60 min	1045 (12.5%)	189 (2.3%)	247 (2.9%)	6912 (82.4%)

6.1.2. Characterization of response delays

Response delays were determined for the test gauges at each site using the approach outlined in Section 4.2, and compiled over the entire test period. The delays are represented as probability density functions in Figure 6. These PDFs provide an overall picture of response delays for each sensor in all precipitation types. The response delays are characterized further by separating the precipitation events by type/phase, and plotting the response delays as a function of the mean precipitation intensity observed by the reference gauge during the delay period (Figure 7).

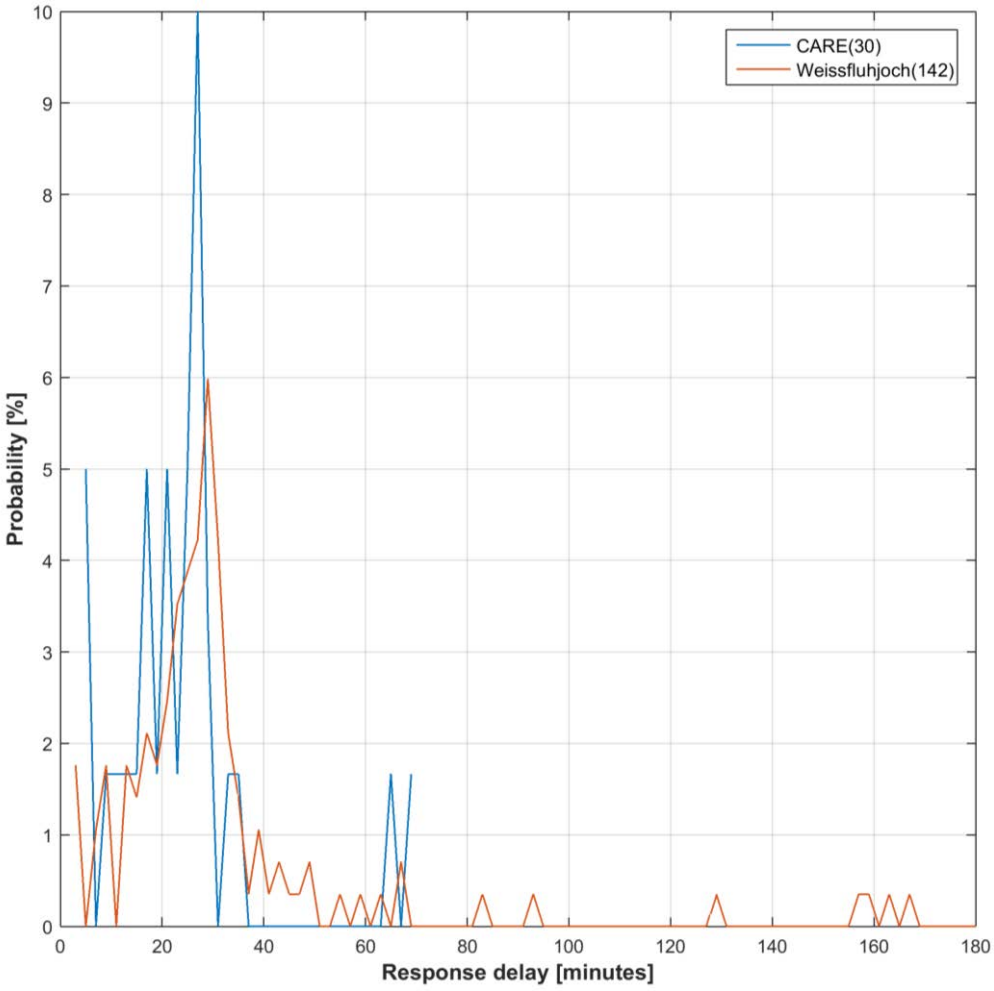


Figure 6: Response delays for heated ZAMG MR3H-FC gauges at CARE and Weissfluhjoch relative to the R2 reference configuration at each site. The number of response assessment periods used to determine the delays for gauges at each site are indicated in parentheses in the legend.

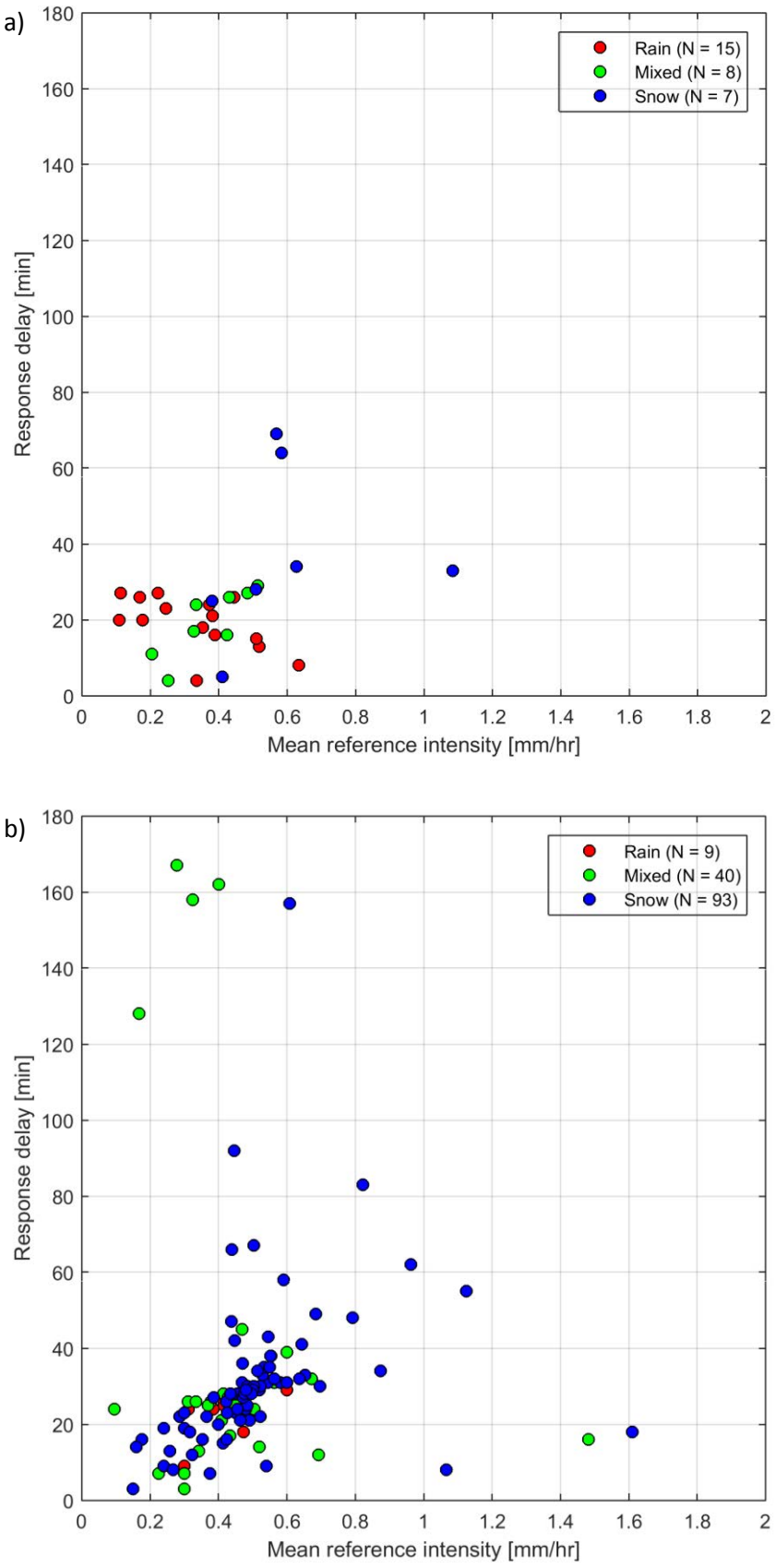


Figure 7: Response delays as a function of the mean precipitation intensity observed by the reference gauge during the delay period at (a) CARE and (b) Weissfluhjoch. The predominant precipitation type for each event is determined from the maximum and minimum reported temperature during the delay period.

6.2. Ability to report accumulated precipitation

The SUT performance in terms of reporting accumulated precipitation is examined by comparing the amount reported by each sensor under test relative to the respective site reference during a given assessment interval. This is represented graphically using scatter and box plots of the catch efficiency as a function of mean wind speed at gauge height, as well as scatter plots of the amounts reported by the SUT versus the corresponding reference amounts (Figures 8 to 10). The SUT performance is also assessed in terms of the root mean square error, RMSE (Figure 11).

Only assessment intervals during which the SUT and reference both reported precipitation (YY cases) are considered in this portion of the assessment. To assess the influence of the assessment interval on SUT performance, analysis was conducted using both 30 and 60 minute intervals. The plots presented in Figures 8 to 10 include only the 30 minute results, in order to better constrain the wind speed and temperature data used in the assessment, which vary to a greater extent over longer (60 minute) periods.

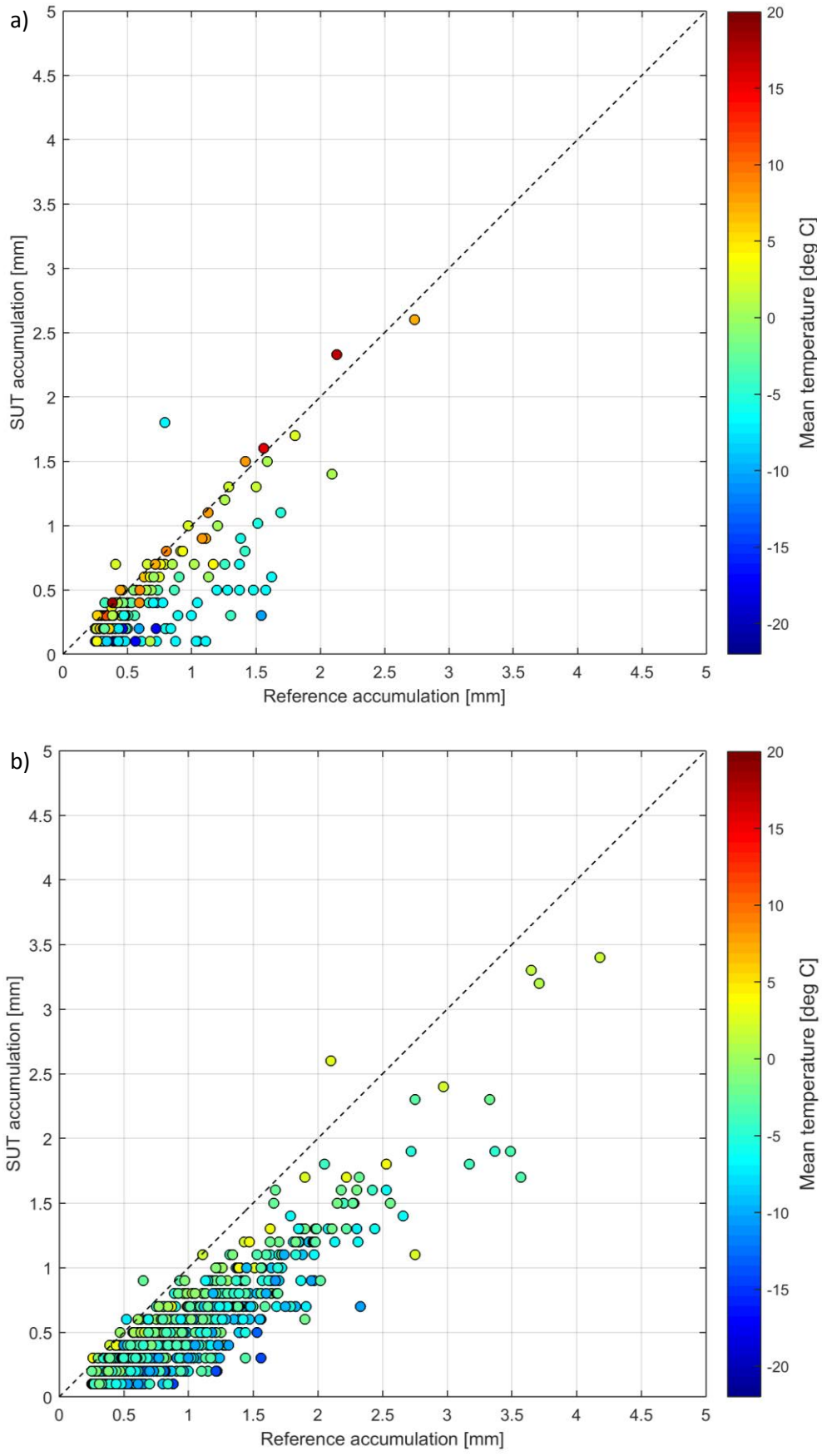


Figure 8: SUT accumulation vs. reference accumulation scatter plots for 30 minute precipitation events at (a) CARE and (b) Weissfluhjoch. The mean event temperature is indicated by colour. Note that black markers represent events with no valid temperature data, which are not considered in the analysis based on precipitation phase (Figure 8).

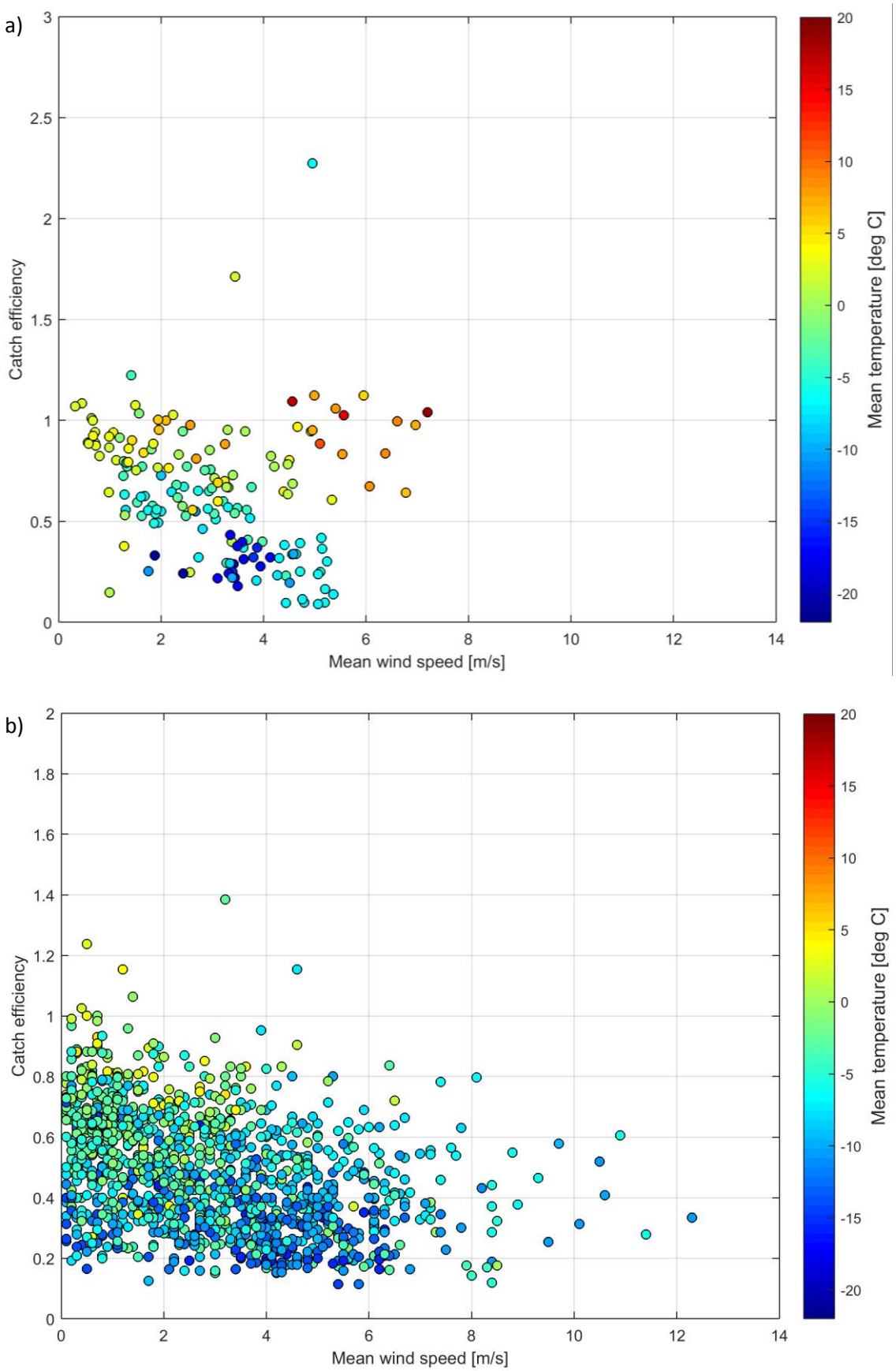


Figure 9: Catch ratio vs. mean wind speed scatter plots for 30 minute precipitation events at (a) CARE and (b) Weissfluhjoch. The mean event temperature is indicated by colour.

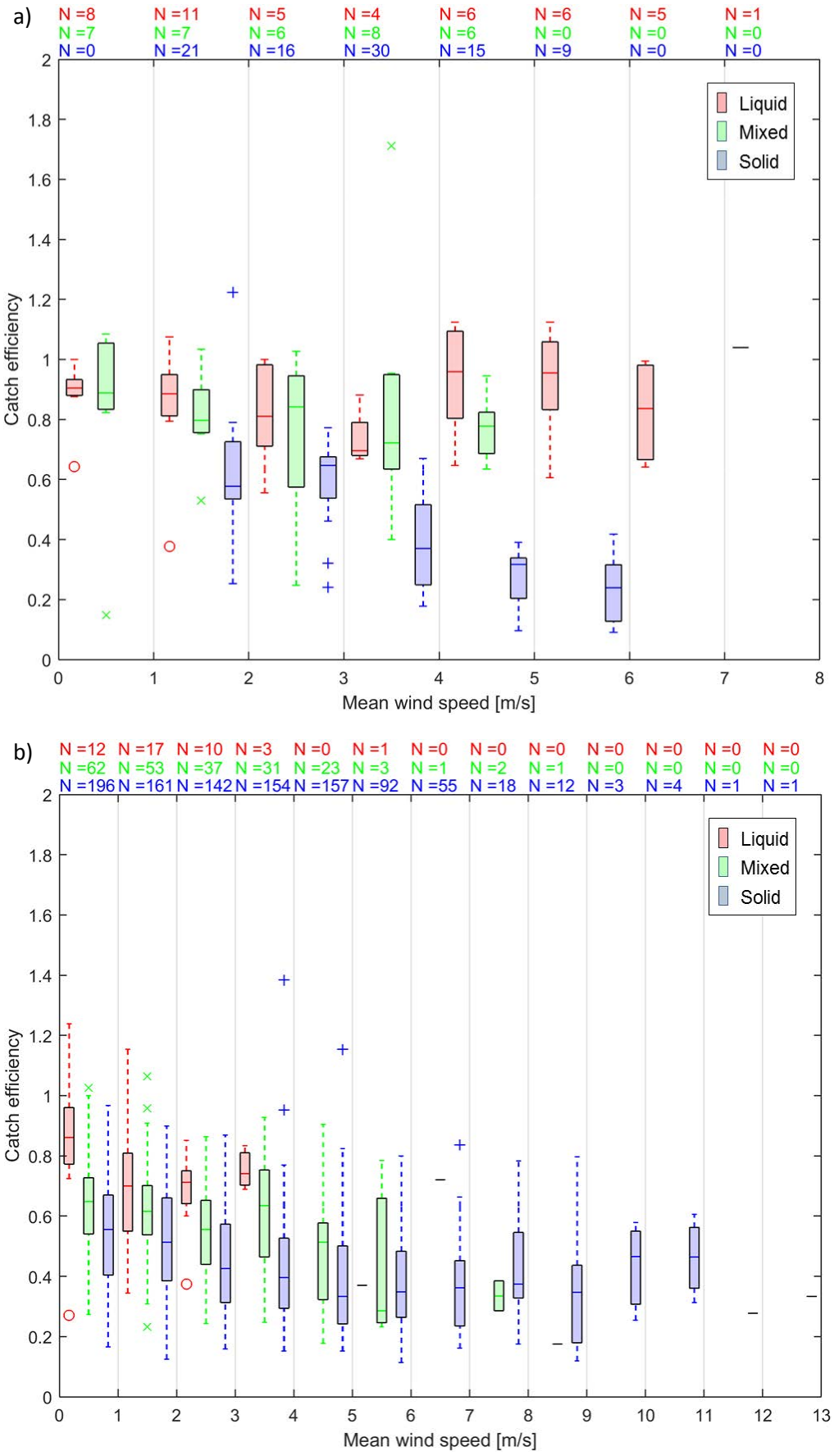


Figure 10: Catch ratio vs. mean wind speed box and whisker plots for 30 minute precipitation events at (a) CARE and (b) Weissfluhjoch. The predominant precipitation type for each event is determined from the maximum and minimum reported temperature and indicated by colour.

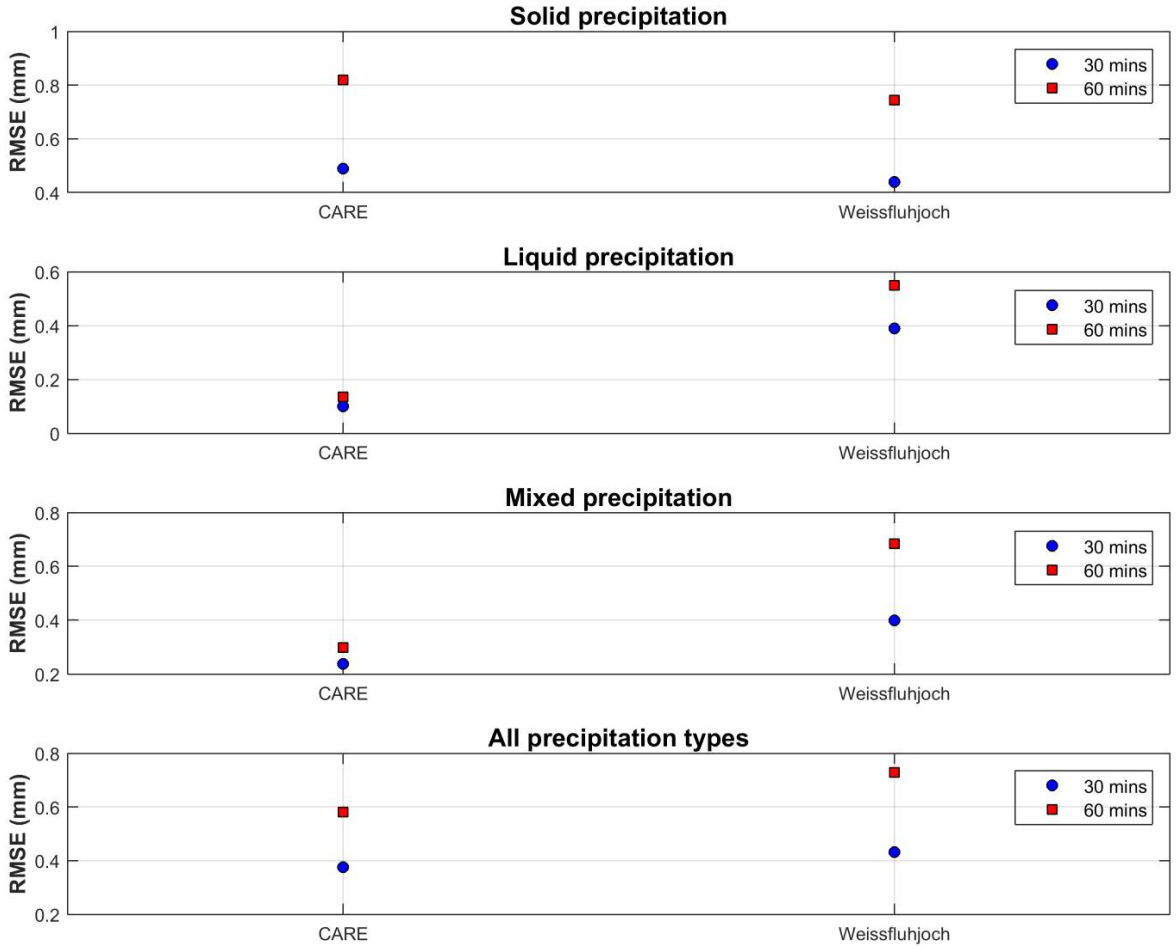


Figure 11: Graphical representation of the RMSE values, by site, and by precipitation type, for YY cases.

The total accumulation reported by each site reference and SUT, for all assessment intervals during which both detected and reported precipitation, are presented alongside the corresponding catch ratios in Table 5.

Table 5: SUT and reference accumulation and overall catch ratio, assessed for YY cases over the entire test period, by site.

Site	Time interval	SUT accumulation	Reference accumulation	Overall catch ratio
CARE	30 min	75.1 mm	116.7 mm	0.64
	60 min	82.5 mm	136.1 mm	0.61
Weissfluhjoch	30 min	557.8 mm	1065.2 mm	0.52
	60 min	607.2 mm	1217.2 mm	0.50

6.3. Assessment of events when the reference and the SUT do not agree on the occurrence of precipitation

Assessment intervals during which the site reference and SUT do not agree on the occurrence of precipitation – namely, the YN and NY cases (Section 4.1.1) – are characterized using histograms in Figures 12 and 13. The histograms include accumulated precipitation, precipitation intensity as reported by the reference, and corresponding site conditions.

The total SUT and reference accumulations over the test period include contributions from YN and NY cases, which impact the overall catch efficiency. These contributions may be significant, given the response delays associated with heated tipping bucket gauges. Total accumulation and catch efficiency results presented in Table 5 (YY cases only) are expanded to include contributions from YN and NY cases in Table 6.

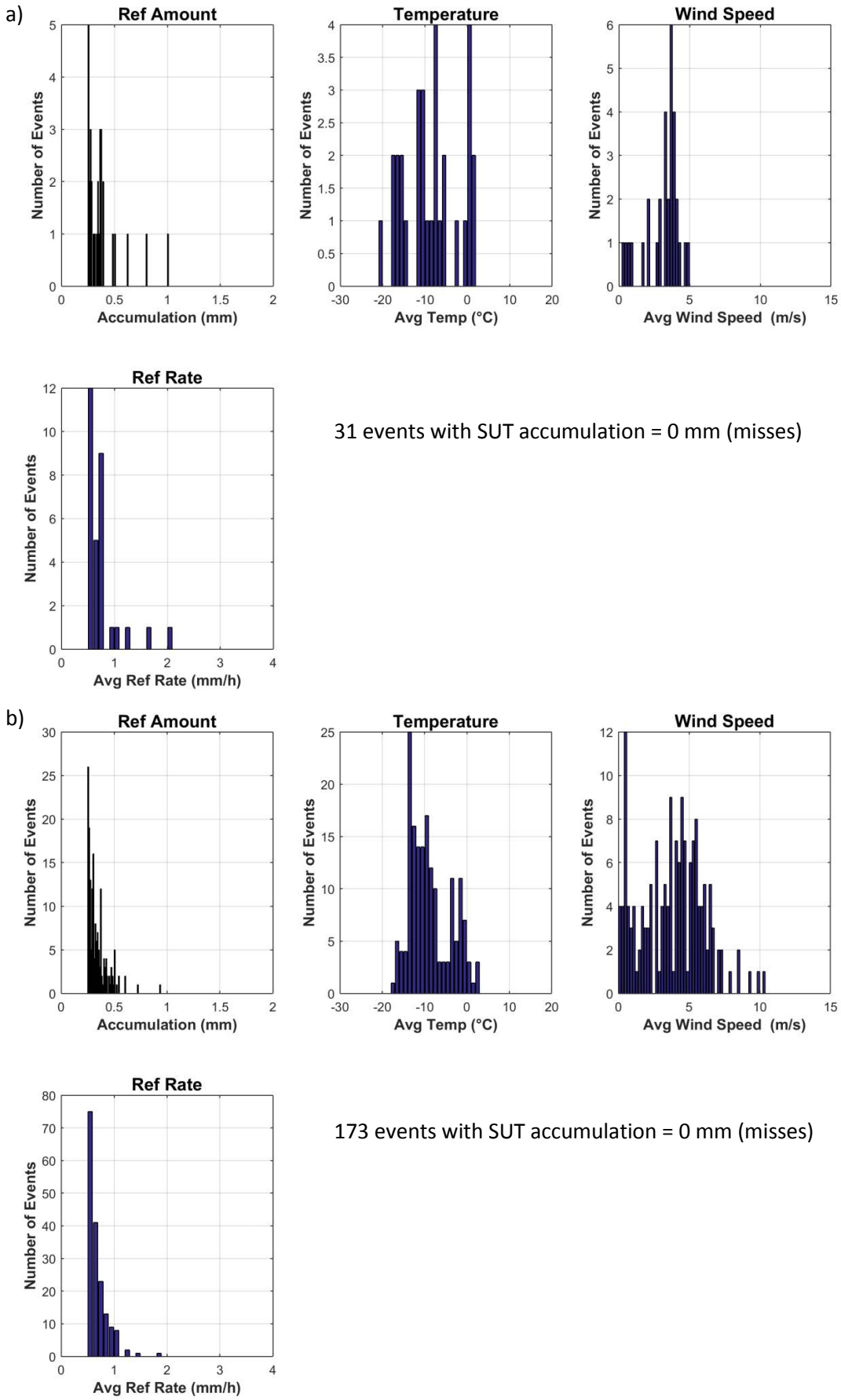


Figure 12: Histograms of reference accumulation, mean temperature, mean wind speed, and reference precipitation rate for YN cases at (a) CARE and (b) Weissfluhjoch.

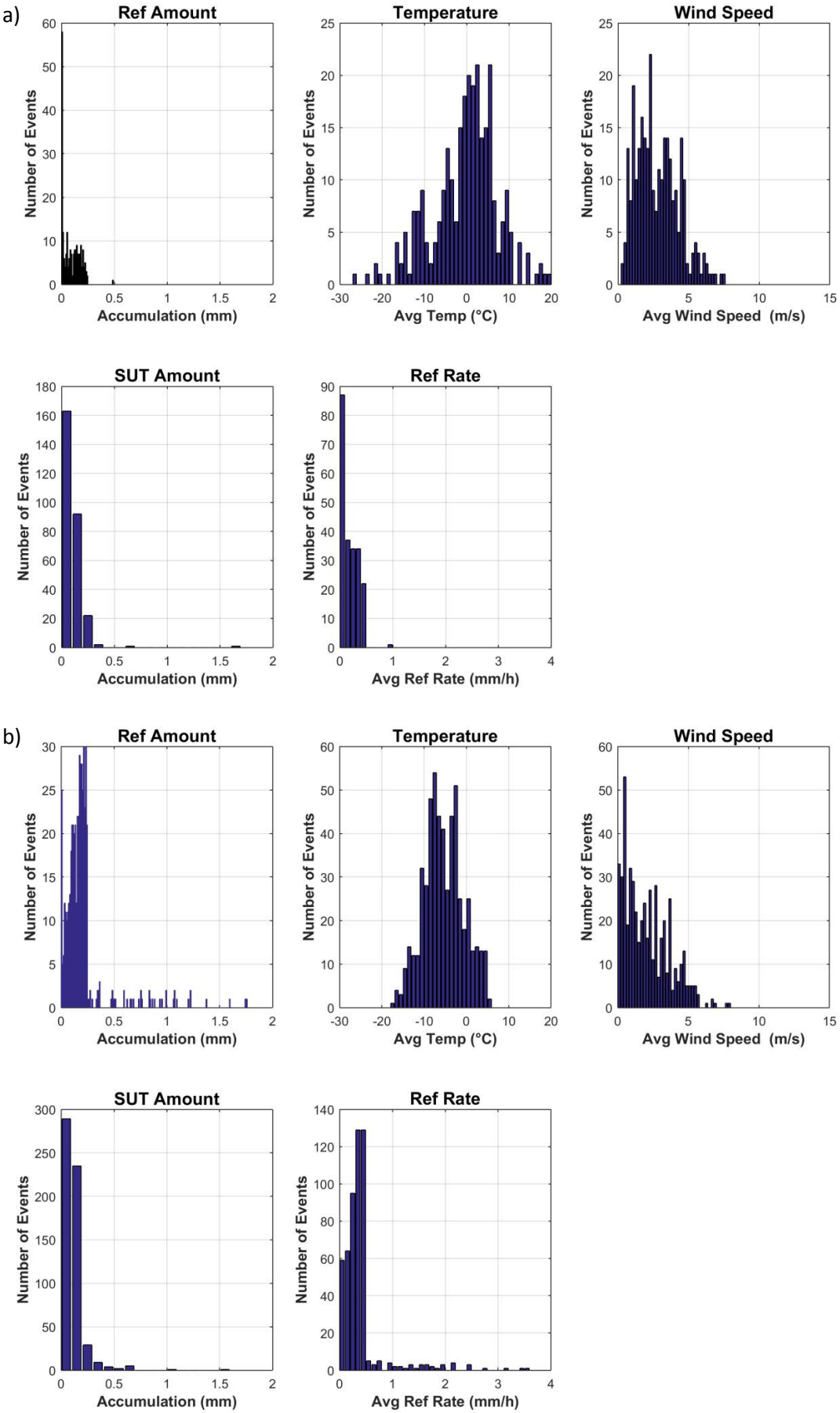


Figure 13: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, and reference precipitation rate for NY cases at (a) CARE (b) Weissfluhjoch.

Table 6: SUT and reference accumulation and overall catch ratio, assessed for the entire test period, for all YY, YN, and NY cases.

Site	Time interval	SUT accumulation	Reference accumulation	Overall catch ratio
CARE	30 min	108.3 mm	147.2 mm	0.74
	60 min	108.3 mm	162.4 mm	0.67
Weissfluhjoch	30 min	631.7 mm	1233.3 mm	0.51
	60 min	643.1 mm	1343.6 mm	0.48

7. Interpretation of Results

7.1. Operating conditions

The full range of conditions under which the test gauges were operated is illustrated in Figure 3. The conditions relevant to gauge operation are as follows:

- Temperatures from -31 °C to 23 °C;
- Precipitation rates within about 7 mm/hr.

These conditions fall within the manufacturer's specified range of operating conditions.

The conditions during precipitation events (as identified by the reference configuration) are shown in Figure 4. The events at Weissfluhjoch, an alpine site, are characterized by higher wind speeds (larger proportion of events with mean wind speeds > 6 m/s) and generally colder temperatures relative to CARE, a continental site. The snow events at CARE and Weissfluhjoch are characterized by mean intensities of 1.2 and 1.4 mm/hr, respectively, and maximum intensities of 3.4 and 7.1 mm/hr, respectively.

7.2. Ability to detect and report precipitation

7.2.1. Skill score assessment

The performance of the sensors under test with respect to the detection of precipitation (relative to the reference configuration) was considered in terms of skill scores, computed for both 30 and 60 minute assessment intervals (Figure 5). Using the shorter 30 minute assessment intervals, the Probability of Detection (POD) is higher for both gauges under test relative to the scores for the longer 60 min assessment intervals, but the False Alarm Rate (FAR) and Bias (B) is also higher for both gauges. As a result, the Heidke Skill Score (HSS) values are higher for both gauges when 60 minute assessment intervals are used. The comparison of skill scores for the test gauges at each site will be considered for only the 60 minute assessment intervals, as hourly reports are broadly used, operationally.

The POD was high for both gauges under test, with values of approximately 79% and 85% for CARE and Weissfluhjoch, respectively. Higher FAR and B are observed for the CARE data, on the order of 59% and 195%, respectively, relative to Weissfluhjoch, with values on the order of 19% and 105%, respectively. This is attributed to differences in the relative proportions of 'hits' and 'false alarms' in the data from each site; more false alarms than hits are observed for the test gauge at CARE, while approximately four times as many hits as false alarms are observed for the test gauge at Weissfluhjoch (Table 4). These differences give rise to the higher Heidke Skill Score (HSS) of about 80% for Weissfluhjoch, relative to the 51% value for CARE.

7.2.2. Characterization of response delays

7.2.2.1. Response delay PDFs

Probability distribution functions of response delays (PDFs) between the onset of precipitation as observed by the reference configuration and the first response by the ZAMG MR3H-FC heated tipping bucket gauge are shown in Figure 6. For CARE, response times within 40 minutes have the highest probability, with lower probability for delays between 60 and 70 minutes. For Weissfluhjoch, the highest probabilities for delays are also within about 40 minutes, but the distribution tails off more gradually to longer delays (60 to 70 minutes). These differences can likely be attributed to differences in heating efficacy in different environments and conditions.

For precipitation occurring within a given 30 or 60 minute interval (as indicated by the reference configuration), the response delay PDFs for the test gauges at both sites indicate the potential for the TB response to occur within a subsequent 30 or 60 minute interval. The delays with highest probability in the PDFs are within about 40 minutes, so it is more likely that the TB gauges will respond to precipitation detected by the reference within a 60 minute period than within a 30 minute period. This is reflected by the lower number of false alarm (NY) cases for the 60 minute intervals relative to the number of false alarms for the 30 minute intervals at each site in Table 4, and the lower FAR and Bias observed for the gauges at each site for 60 minute intervals in Figure 5.

Another important consideration is that even when both the SUT and reference detect precipitation within a given interval, the TB may actually be reporting precipitation collected during previous intervals, or may only report a fraction of the precipitation within that interval (reporting the rest during a subsequent interval). These ‘false YY cases’ will impact the assessment of TB gauge performance in terms of reporting accumulated precipitation (Section 7.3).

7.2.2.2. Influence of precipitation intensity, phase

The dependence of delay times on the mean intensity of precipitation preceding the TB response, as reported by the reference configuration, is illustrated in Figure 7. For the test gauge at CARE (Figure 7a), response times are within 30 minutes for all precipitation types (liquid, mixed, solid) when the mean intensity of precipitation is within about 0.55 mm/hr. This suggests that for low intensities, the TB response is independent of precipitation phase, and that heating/melting of solid precipitation does not increase delay times significantly. For mean solid precipitation intensities exceeding ~ 0.55 mm/hr, however, the delay times extending to about 60 – 70 minutes. This can perhaps be attributed to heating/melting, as delay times remain within 30 minutes for liquid precipitation intensities > 0.55 mm/hr; however, the number of events upon which to base this attribution is limited.

For the test gauge at Weissfluhjoch (Figure 7b), response delays are typically within 30 minutes for all precipitation types when the mean intensity is within 0.4 mm/hr. At higher mean intensities (> 0.4 mm/hr), mixed and solid precipitation events show longer response delays, up to ~ 160 minutes, while liquid events show delays within 30 minutes. As proposed above, this can be attributed to heating/melting delays for the mixed and solid events at higher intensities. The mixed events with delay times > 120 minutes at mean intensities < 0.4 mm/hr are likely events with variable intensity

over a longer duration; that is, while the mean intensity is low, there may have been periods of higher intensity that impacted the gauge response.

7.3. Ability to measure and report precipitation

7.3.1. Performance when both SUT and reference detect precipitation

7.3.1.1. Temperature, phase, and wind speed influence

Focussing on the 30 minute events in which both the sensors under test and site reference configurations report precipitation (YY cases, or 'hits,' only), general stratification of the data by temperature is observed for CARE and Weissfluhjoch in Figures 8 and 9. The coldest temperatures are generally correlated with larger differences in accumulation between the reference and SUT (Figure 8), and hence lower catch efficiencies (Figure 9), owing to the enhanced influence of wind speed on the collection of mixed and solid precipitation, relative to liquid precipitation.

The phase and wind speed influences on catch efficiency are more clearly illustrated in the box and whisker plots in Figure 10. For CARE, the median catch efficiencies for solid precipitation range from about 0.5 to 0.7 for wind speeds < 3 m/s, decreasing to values between 0.2 and 0.4 for wind speeds between 3 and 6 m/s. Similar trends are observed for Weissfluhjoch, with catch efficiencies for solid precipitation between 0.4 and 0.6 for wind speeds < 4 m/s and values between 0.2 and 0.4 for wind speeds > 4 m/s. (The higher median values for solid precipitation observed for Weissfluhjoch at wind speeds > 9 m/s are not believed to be representative, given the relatively small numbers of events at these wind speeds.)

These general trends for snow, while broad, characterize the maximum expected wind speed effects on catch efficiency for the gauges tested at CARE and Weissfluhjoch, under the specific conditions tested. As observed in Figure 10, the trends for mixed and liquid precipitation fall within these limits.

7.3.1.2. Root mean square error

The RMSE was calculated for all 30 and 60 minute events during which the reference and SUT both detected precipitation (YY cases), for each site. The overall RMSE calculated for each site (all precipitation types), and RMSE values calculated for each precipitation type, are shown in Figure 11. Given the broad use of hourly reports in meteorological operations, further discussion of RMSE values will be considered for 60 minute assessment intervals only.

The overall RMSE, for all precipitation types, was 0.58 mm for the test gauge at CARE and 0.73 mm for the test gauge at Weissfluhjoch. These values can be considered to represent the absolute uncertainty of the ZAMG MR3H-FC relative to a DFAR configuration for the specific site and gauge conditions tested, over the range of conditions tested. Weissfluhjoch is a site with complex orography, which may impact the flow field above reference and/or test gauges, and is characterized by higher winds relative to CARE. The relative proportions of 60 minute events by precipitation type also differ by site: 78% solid events at Weissfluhjoch compared to 43% at CARE; 18% mixed events at Weissfluhjoch compared to 26% for CARE; and 4% liquid events at Weissfluhjoch compared to 31% at CARE.

The differences outlined above may explain the relative differences in RMSE between the test gauges at each site in Figure 11. The overall RMSE is larger for the test gauge at Weissfluhjoch, as are the RMSE values for liquid and mixed precipitation, while the RMSE for solid precipitation is lower. Among the different precipitation types, the RMSE values for solid precipitation are the largest, owing to the larger wind-induced differences in reference and SUT accumulation relative to mixed and liquid precipitation. The RMSE for solid precipitation at Weissfluhjoch (0.74) is close to the value for mixed precipitation (0.68), suggesting that the mixed precipitation may be predominated by solid precipitation over the duration of experiments.

7.3.1.3. Overall catch efficiency

The total accumulated precipitation recorded by the reference and sensor under test at each site was compiled for all 30 and 60 minute events during which both detected precipitation (YY cases) and used to calculate the overall catch ratio. The results are provided in Table 5. Focussing on the more operationally-relevant 60 minute values, the overall catch ratio for CARE, 0.61, is higher than that for Weissfluhjoch, 0.50, which is attributed to the complex orography, higher winds, and larger proportion of solid precipitation events at Weissfluhjoch.

7.3.2. Characterization of conditions when the SUT and reference do not agree on the occurrence of precipitation

The conditions during 30 minute events in which the reference detected precipitation, but the SUT did not (YN cases), are depicted as histograms in Figure 12. These events are characterized by low precipitation rates, typically below 2 mm/hr, and temperatures close to, or below, freezing. At these temperatures, the precipitation is likely mixed or solid, and hence subject to longer response delays relative to liquid precipitation at the observed intensities. Hence, these may be cases in which the TB response to precipitation occurs during a subsequent 30 minute period. There is also potential for heating to cause evaporation or sublimation of incident precipitation at low intensities.

The conditions during 30 minute events in which the SUT detected precipitation, but the reference did not (NY cases), are shown as histograms in Figure 13. These events are also characterized by low precipitation rates, typically below 0.5 mm/hr (which corresponds to the reference accumulation threshold for a precipitation event over a 30 minute interval), and wind speeds below about 8 m/s. These events most likely result from delayed tipping bucket responses that coincide with intervals during which the reference did not detect precipitation.

Given the influence of response delays, estimates of the overall catch ratio for tipping bucket gauges should include the reference and SUT accumulations during YN and NY cases (and not just during the YY cases, as considered above). The resulting total accumulations for all YY, YN, and NY cases, and the corresponding overall catch efficiencies, are provided for the test gauges at each site in Table 8. The overall catch efficiencies for the gauges under test at CARE and Weissfluhjoch were 0.67 and 0.48 for 60 minute assessment intervals, respectively. The lower catch efficiencies at Weissfluhjoch are attributed to the complex terrain, higher winds, and larger relative proportion of solid precipitation events, as discussed above.

8. Operational considerations

The ZAMG MR3H-FC gauge was easy to install, easy to connect, and the setup of data collection was straightforward.

The gauge heaters require significant power, the source of which should be located as close as possible to the gauge to mitigate voltage drops and the potential reduction of heating power and efficiency. Investment in site infrastructure may be required to operate the gauge as specified by the manufacturer.

8.1. Maintenance

Each site completed the gauge field calibration and verification as per manufacturer recommendations, at least once a year. The calibration records have been stored by each site team.

8.2. Noted issues

- The site team at Weissfluhjoch indicated that the power of the heater might not be high enough for their site conditions, or that the algorithm to handle the trade-off between capping and evaporation might not be ideally configured for their conditions. The manufacturer noted that standard heating settings were used for the gauges under test, but that the settings can be adapted to suit the specific site conditions. Further, the manufacturer indicated that the heating configuration has been changed since providing the instruments for SPICE, and now features pulse heating by section.

- No issues with clogging were reported during the SPICE test seasons at CARE; however, the site team noted that the orifice was clogged regularly during summer operation (primarily with bird droppings), and had to be cleaned out roughly twice per month.

- The manufacturer indicated that blocking/clogging of the gauge outflow with ice can occur, depending on the details of how the gauge is installed. No blocking/clogging of the gauge outflow was observed for the gauges under test in SPICE.

- At the conclusion of the SPICE measurement period, the gauges were sent back to the manufacturer. The buckets were inspected visually and tested for symmetry. Inspection of the gauge from Weissfluhjoch revealed that the bucket coating had almost been removed; consultation with the site team did not reveal any reasons for this removal, and it is believed that this may result from a manufacturing defect. Further testing revealed that this bucket was no longer symmetric (off by > 17%), which may have resulted from the coating removal.

9. Performance Considerations

- Heating is critical to ensure proper functioning of the gauge under winter conditions. The voltage should be checked at the gauge location to ensure that the heaters are supplied with

the recommended power. Voltage drops – for example, resulting from the use of long power cables – can impair the functioning of the gauge in winter conditions.

- Heating enables the measurement of solid precipitation, but the time required for melting can delay the time between the collection of precipitation in the funnel and the gauge response to that precipitation (response delays). Further, heating may result in the evaporation/sublimation of incident precipitation at low intensities.

- Response delays must be considered when using the gauge in operational settings. Ideally, the reporting interval (i.e. hourly observations) should exceed the maximum expected response delay; however, the potential remains for carry-over of precipitation accumulation from previous intervals, and for delayed responses occurring in subsequent intervals.

- While not tested in SPICE, shielding of the gauge should be considered by the manufacturer and/or potential users as a means of increasing the catch efficiency at higher wind speeds.

- The collection of ancillary measurement data is recommended: air temperature and wind speed (at gauge height); to enable the application of adjustments to measurements (i.e. using transfer functions); and reports from a precipitation detector with high sensitivity, to enable the identification of missed events or false alarms due to response delays.

- The application of transfer functions to gauge measurements, if available, is recommended as a post-processing step to account for reductions in catch efficiency with increasing wind speed. This recommendation comes with the caveat that the catch ratio data used in the derivation of the transfer functions will be impacted by response delays. Response delays can lead to scatter in the catch ratio data (among other factors), and hence, uncertainty in the transfer functions; however, the potential remains for the improvement of data via adjustment. Considering the catch ratios over longer time periods (e.g. 1 hour) may provide one avenue for reducing scatter in the catch ratio/wind speed relationship and the uncertainty of related transfer functions.

WMO-SPICE Instrument Performance Report
Thies Precipitation Transmitter/Ombrometer

1) Technical specifications (from manufacturer provided documentation)

Instrument model:	Thies Precipitation Transmitter, models 5.4032.35.228 and 5.4032.45.008
Physical principle:	Heated tipping bucket (TB) gauge. The bucket tips when filled to capacity; tips trigger a Reed switch, inducing an output pulse.
Bucket capacity:	0.2 mm (model 5.4032.35.228); 0.1 mm (model 5.4032.45.008)
Collecting area:	200 cm^2
Heating configuration:	Electronically regulated heater, activated when ambient temperature sensor records values below 5 °C. Heating power of 49 W (model 5.4032.35.228) or 113 W (model 5.4032.45.008), 24 V AC/DC.
Operating temperature range:	-25 °C to 60 °C (model 5.4032.35.228); -35 °C to 60 °C (model 5.4032.45.008)
Measurement range:	0 – 15 mm/min (model 5.4032.35.228); 0 – 11 mm/min (model 5.4032.45.008)
Measurement uncertainty:	± 3 % over the full measurement range
Sensitivity:	Corresponds to bucket amount; 0.2 mm for model 5.4032.35.228, 0.1 mm for model 5.4032.45.008.

a)

b)

Figure 1: Thies Precipitation Transmitter installation at (a) Formigal and (b) Marshall.

2) Data output format

Gauge data output: Tips are detected by a Reed-switch, inducing an output pulse for each 0.1 mm of precipitation. The relationship between the number of tips and the precipitation intensity is not linear; the gauge electronics perform an intensity-dependent pulse number correction (linearization procedure) for intensities between 0 and 7 mm/min. The corrected number of pulses per sampling period is used to determine the precipitation amount or intensity during that period.

Analysis is based on accumulated precipitation, computed as the cumulative sum of accumulations from each sampling period.

3) SPICE test configuration

Shield: Unshielded

Test sites: Formigal (Spain); Marshall (USA)

Sensor provider(s): AEMet, Spanish State Agency of Meteorology (Formigal); Thies GmbH & Co. KG (Marshall)

Figure 2: Map of SPICE sites where Thies Precipitation Transmitter gauges were tested.

A summary of the configuration of instruments as tested, the duration of tests and availability of data reflected in these results, and the ancillary measurements used, by site, is available in Tables 1, 2, and 3, respectively.

Table 1: Summary of gauge configurations and data output, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	Formigal	Marshall
Field configuration	Unshielded	Unshielded
Height of installation (gauge rim)	3.5 m	1.4 m
Model	5.4032.35.228	5.4032.45.008
Heating configuration*	49 W, 24 V AC/DC	113 W, 24 V AC/DC
Operating range	-25 °C to 60 °C	-35 °C to 60 °C
Reporting resolution	0.2 mm	0.1 mm
SUT data output frequency	1 min	6 s, 1 min
Data QC	SPICE QC methodology	
Data temporal resolution	1 min	
Processing interval for SPICE data analysis	30 min, 60 min	

* The manufacturer recommends choosing a heating configuration to suit the expected range of conditions. The heating configuration employed at Formigal is that used in the Spanish operational network, and may not represent the manufacturer's recommendation.

Table 2: Data availability, by measurement season and site.

Duration of experiments	Formigal	Marshall
Season 1 (Oct. 2013 – Apr. 2014)	X	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓	✓

No data are available for the 2013-2014 season at Formigal, as the R2 reference configuration (DFAR) had not yet been installed.

Table 3: Summary of reference and ancillary measurements, by site. Details and photos of individual site configurations are available in the respective site commissioning protocols.

	Formigal	Marshall
R2 site reference	OTT Pluvio^2 – Bucket RT (DFAR)	Geonor T-200B3 600 mm (DFAR)
R2 precip detector	Thies LPM (DFAR)	Thies LPM (Site*)
Ancillary temp sensor	Thies Thermo Hygrometer (4 m)	MetOne, model 060A-2/062, 2144-L
Ancillary RH sensor	Thies Thermo Hygrometer (4 m)	Campbell Scientific CS500
Ancillary wind sensor	RM Young Alpine 05103-45 (10 m)	RM Young Wind Monitor 05103 (2 m)

*A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR-fence; however, in cases where a more sensitive detector is available outside of the DFIR-fence, or there are issues with the detector within the DFIR-fence, a precipitation detector elsewhere on the site can be employed.

4) Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6.1 of the WMO-SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute assessment intervals:

- Ref 'Yes' : R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No' : R2 weighing gauge < 0.25 mm AND/OR precip detector recording < 18 min of precip;
- SUT 'Yes' : SUT accumulation > 0 mm;
- SUT 'No' : SUT accumulation = 0 mm.

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores.

4.1.2. Skill score assessment

The ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

The above scores are computed using the formulations provided in Section 3.6.1.3 of the WMO-SPICE Final Report.

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

4.2. Sensor-specific considerations: response delays

Tipping bucket gauges require that an amount of precipitation corresponding to the bucket capacity is accumulated before a tip is triggered and the gauge records precipitation. This can result in response delays relative to the reference configuration. Heated tipping bucket gauges are subject to further delays, as any solid precipitation in the funnel must be melted before reaching the bucket and potentially triggering a tip, and the heating itself can potentially evaporate incident precipitation. These response delays will impact the comparison with the reference. For this reason, the assessment of TB gauge performance relative to the reference configuration is also considered over 60 minute intervals, using the same conditions and thresholds outlined above in Section 4.1.1.

Response delays are quantified by determining the time elapsed between the onset of precipitation as determined by the reference configuration, and the first tip recorded by the TB. The assessment is based on periods with at least 30 minutes of precipitation, as identified by the reference configuration, followed by at least 180 minutes without precipitation. This extended period without precipitation is intended to allow additional time for the melting and recording of precipitation by heated TB gauges. Additional details are provided in Section 3.6.1.4.4 of the WMO-SPICE Final Report.

5) Environmental conditions

The environmental conditions at each site over the duration of the test period are expressed as probability density functions (PDFs) of mean air temperature, mean relative humidity, mean wind speed, vector mean wind direction, and precipitation rate for each component 30 minute assessment interval in Figure 3. Figure 4 presents the same parameters for all assessment intervals during which the site reference configuration detected precipitation (i.e. all Ref 'Yes' cases).

The precipitation percentage represents the number of minutes of precipitation during a 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration, expressed as a percentage. PDFs of precipitation percentage are also included in Figures 3 and 4.

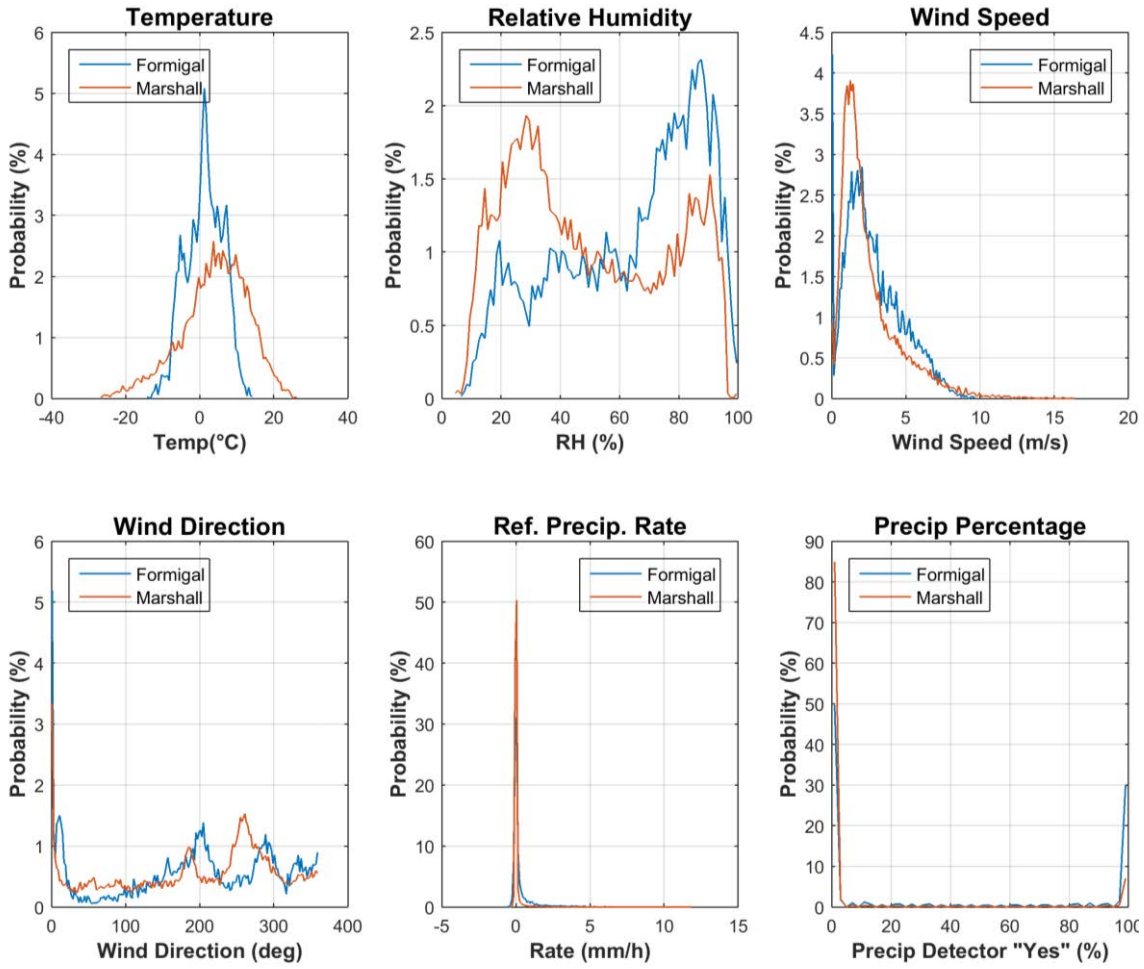


Figure 3: Summary of aggregated environmental conditions at the SPICE sites that operated Thies Precipitation Transmitter gauges, over the entire duration of formal tests, as per Table 2.

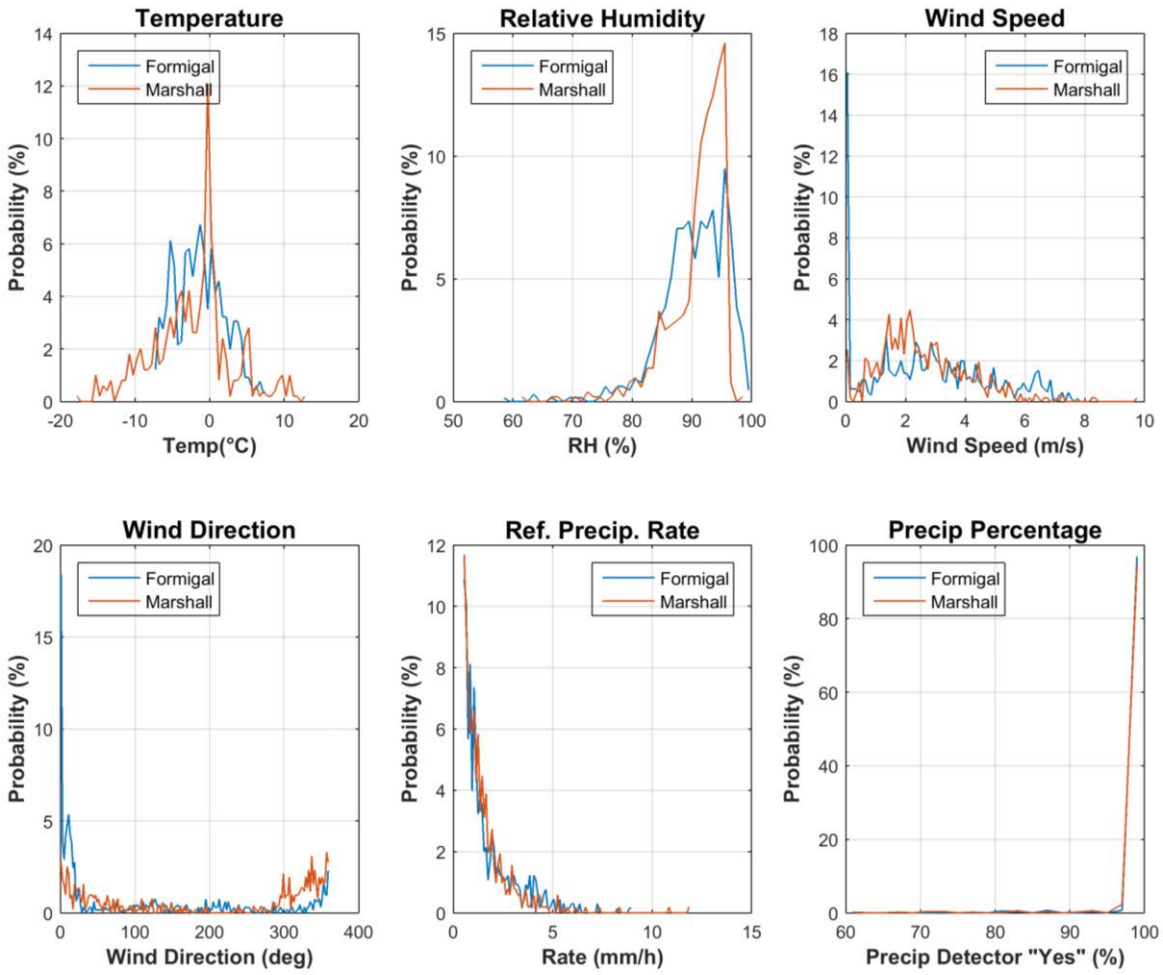


Figure 4: Summary of the aggregated environmental conditions at the SPICE sites that tested Thies Precipitation Transmitter gauges during precipitation events, as reported by the site R2 reference, during the formal tests, as per Table 2.

6) Evaluation of performance over the range of operating conditions

6.1. Ability to detect and report precipitation

6.1.1. Skill score assessment

The overall ability of the SUT to detect and report the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores (Section 4.1.2) and presented in Table 4. Scores are presented for both 30 minute and 60 minute assessment intervals. The contingency results (Section 4.1.1) corresponding to these scores are presented in Table 5.

Table 4: Skill scores for Thies Precipitation Transmitter gauges during the formal test periods.

Site (bucket resolution)	Time interval	Skill Score (%)			
		Probability of Detection, POD	False Alarm Rate, FAR	Bias, B	Heidke Skill Score, HSS
Formigal (0.2 mm)	30 min	84.1	21.1	107	78.9
	60 min	82.9	12.4	94.6	81.9
Marshall (0.1 mm)	30 min	85.2	29.1	120	76.7
	60 min	78.8	16.7	94.6	80.2

Table 5: Contingency table illustrating detection of precipitation by the Thies Precipitation Transmitter relative to the specific site reference, expressed as number (and percentage) of events over the entire test period, by site.

Site (bucket resolution)	Time interval	Number of Events (% of Events)			
		YY, hits	YN, misses	NY, false alarms	NN, correct negatives
Formigal (0.2 mm)	30 min	549 (10.8%)	104 (2.0%)	147 (2.9%)	4303 (84.3%)
	60 min	387 (15.2%)	80 (3.1%)	55 (2.2%)	2027 (79.5%)
Marshall (0.1 mm)	30 min	438 (2.2%)	76 (0.4%)	180 (0.9%)	19348 (96.5%)
	60 min	335 (3.3%)	90 (0.9%)	67 (0.7%)	9517 (95.1%)

6.1.2. Characterization of response delays

Response delays were determined for the test gauges at each site using the approach outlined in Section 4.2, and compiled over the entire test period. The delays are represented as probability density functions in Figure 5. These PDFs provide an overall picture of response delays for each sensor in all precipitation types. The response delays are characterized further by separating the precipitation events by type/phase, and plotting the response delays as a function of the mean precipitation intensity observed by the reference gauge during the delay period (Figure 6).

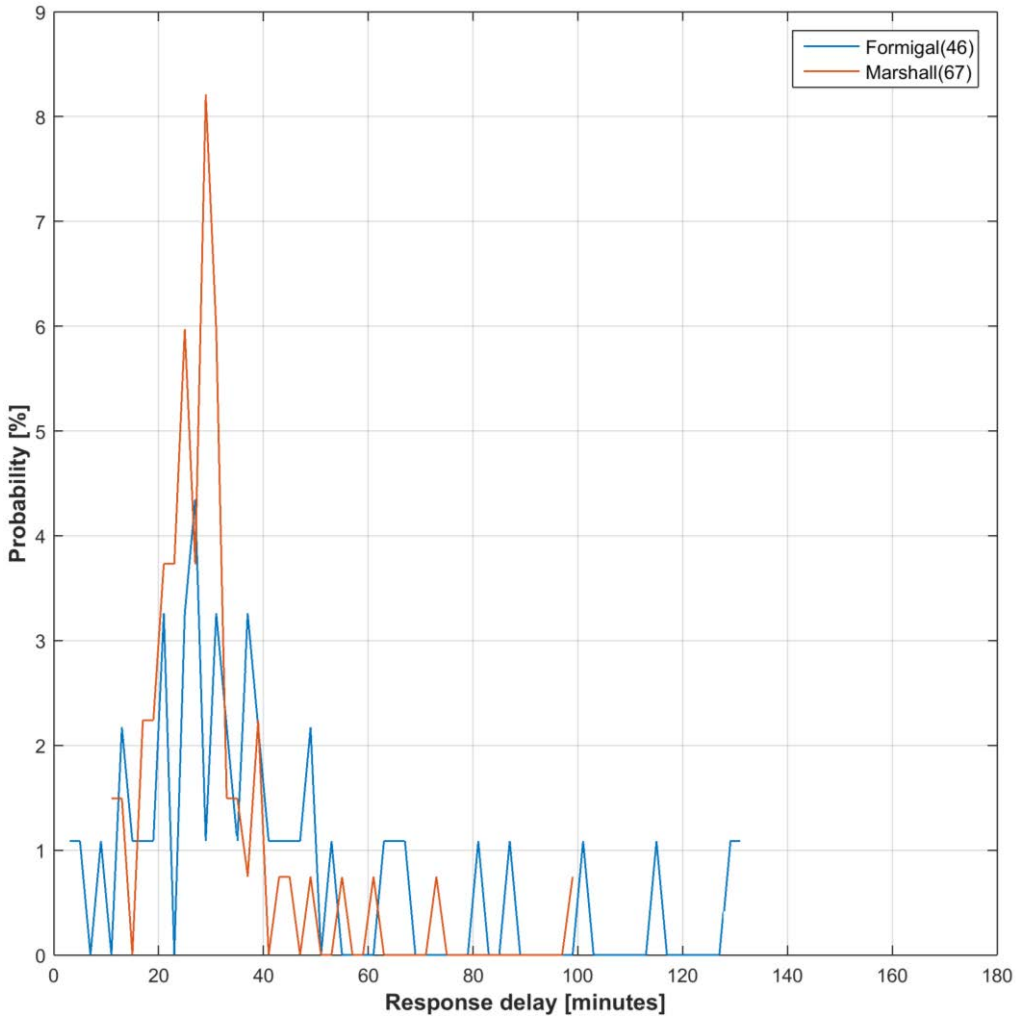


Figure 5: Response delays for heated Thies Precipitation Transmitter gauges at Formigal and Marshall relative to the R2 reference configuration at each site. The number of response assessment periods used to determine the delays for gauges at each site are indicated in parentheses in the legend.

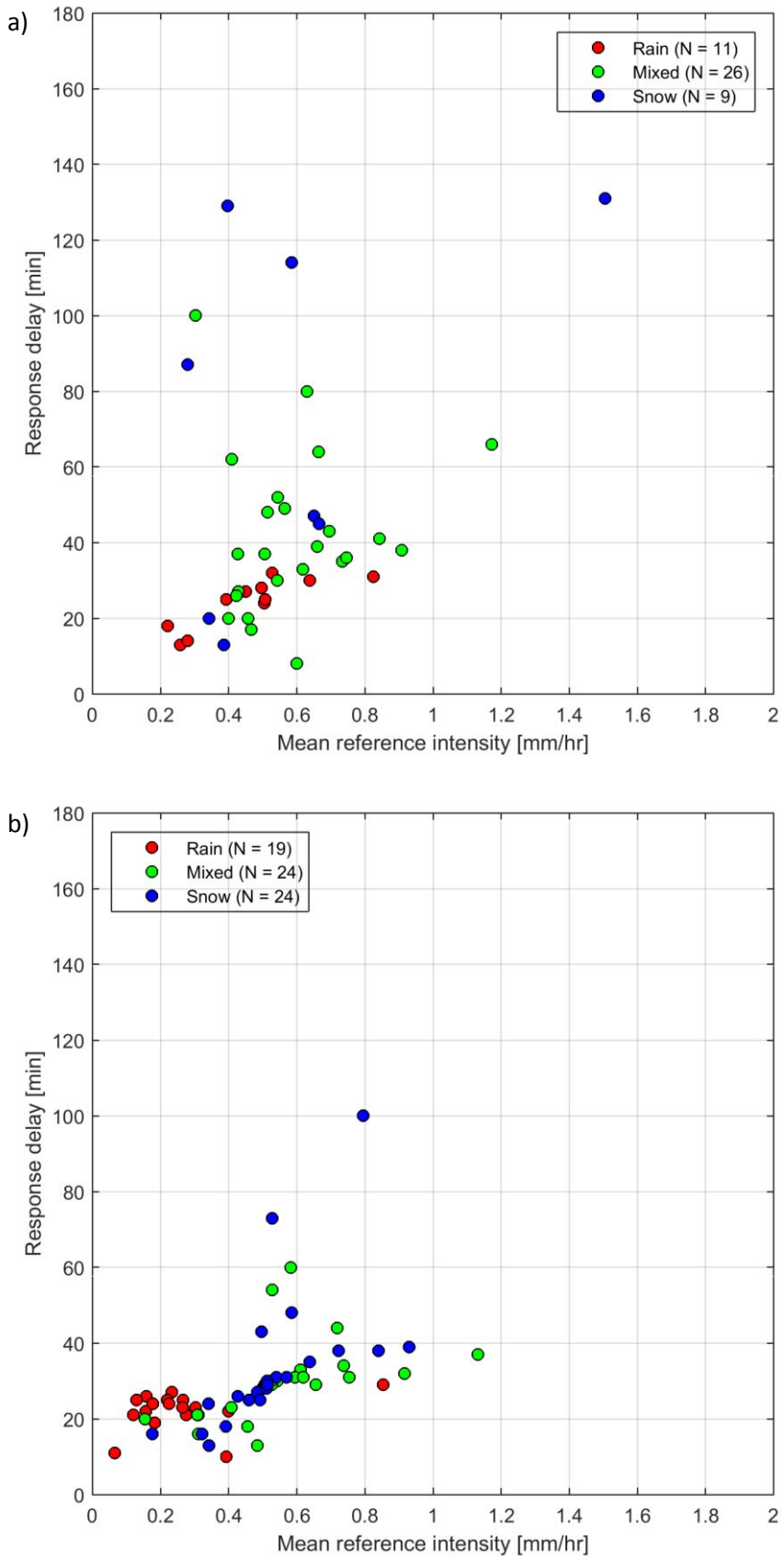


Figure 6: Response delays as a function of the mean precipitation intensity observed by the reference gauge during the delay period at (a) Formigal and (b) Marshall. The predominant precipitation type for each event is determined from the maximum and minimum reported temperature during the delay period.

6.2. Ability to report accumulated precipitation

The SUT performance in terms of reporting accumulated precipitation is examined by comparing the amount reported by each sensor under test relative to the respective site reference during a given assessment interval. This is represented graphically using scatter and box plots of the catch efficiency as a function of mean wind speed at gauge height, as well as scatter plots of the amounts reported by the SUT versus the corresponding reference amounts (Figures 7 to 9). The SUT performance is also assessed in terms of the root mean square error, RMSE (Table 6).

Only assessment intervals during which the SUT and reference both reported precipitation (YY cases) are considered in this portion of the assessment. To assess the influence of the assessment interval on SUT performance, analysis was conducted using both 30 and 60 minute intervals. The plots presented in Figures 8 to 10 include only the 30 minute results, in order to better constrain the wind speed and temperature data used in the assessment, which vary to a greater extent over longer (60 minute) periods.

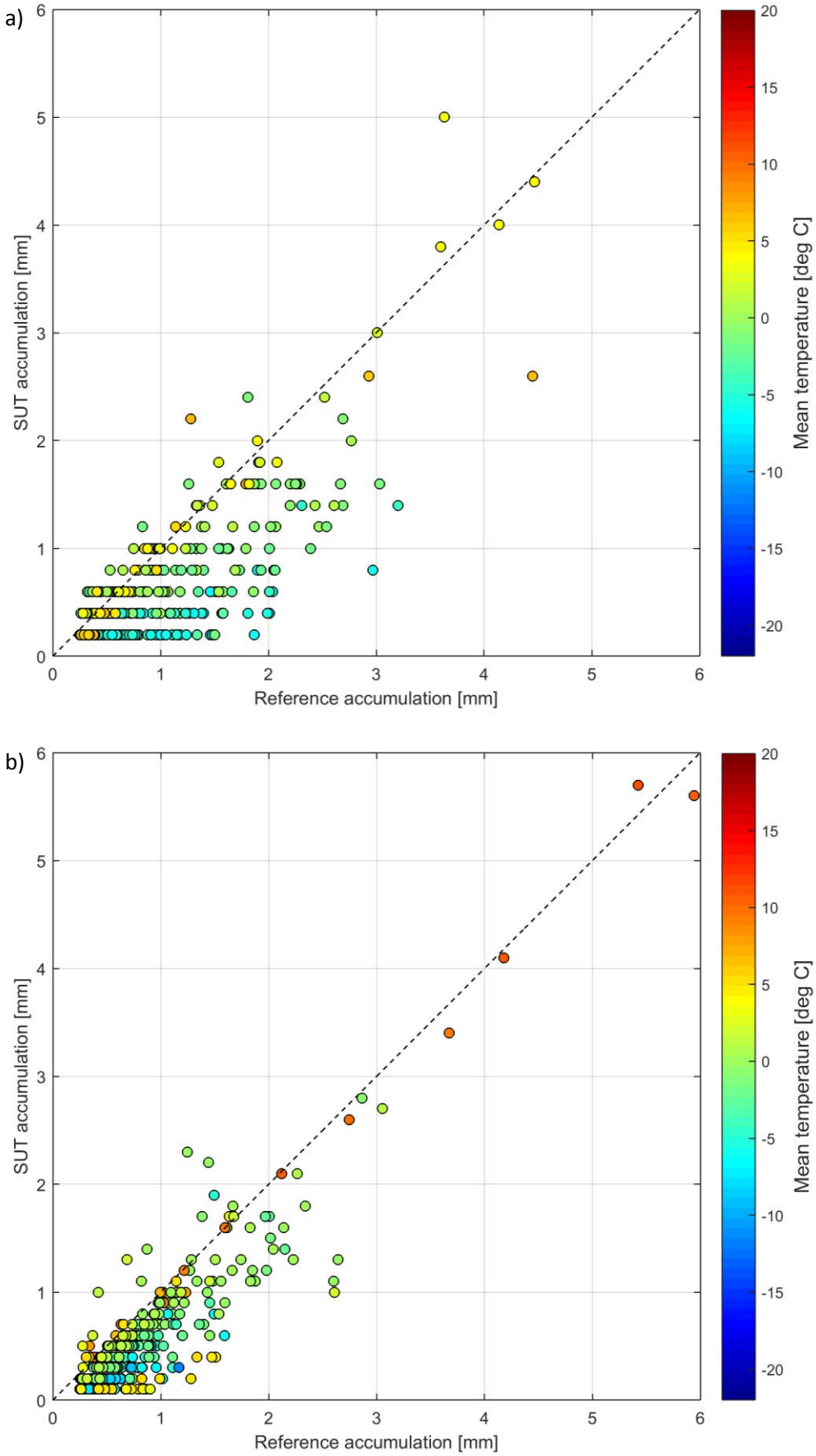


Figure 7: SUT accumulation vs. reference accumulation scatter plots for 30 minute precipitation events at (a) Formigal and (b) Marshall. The mean event temperature is indicated by colour.

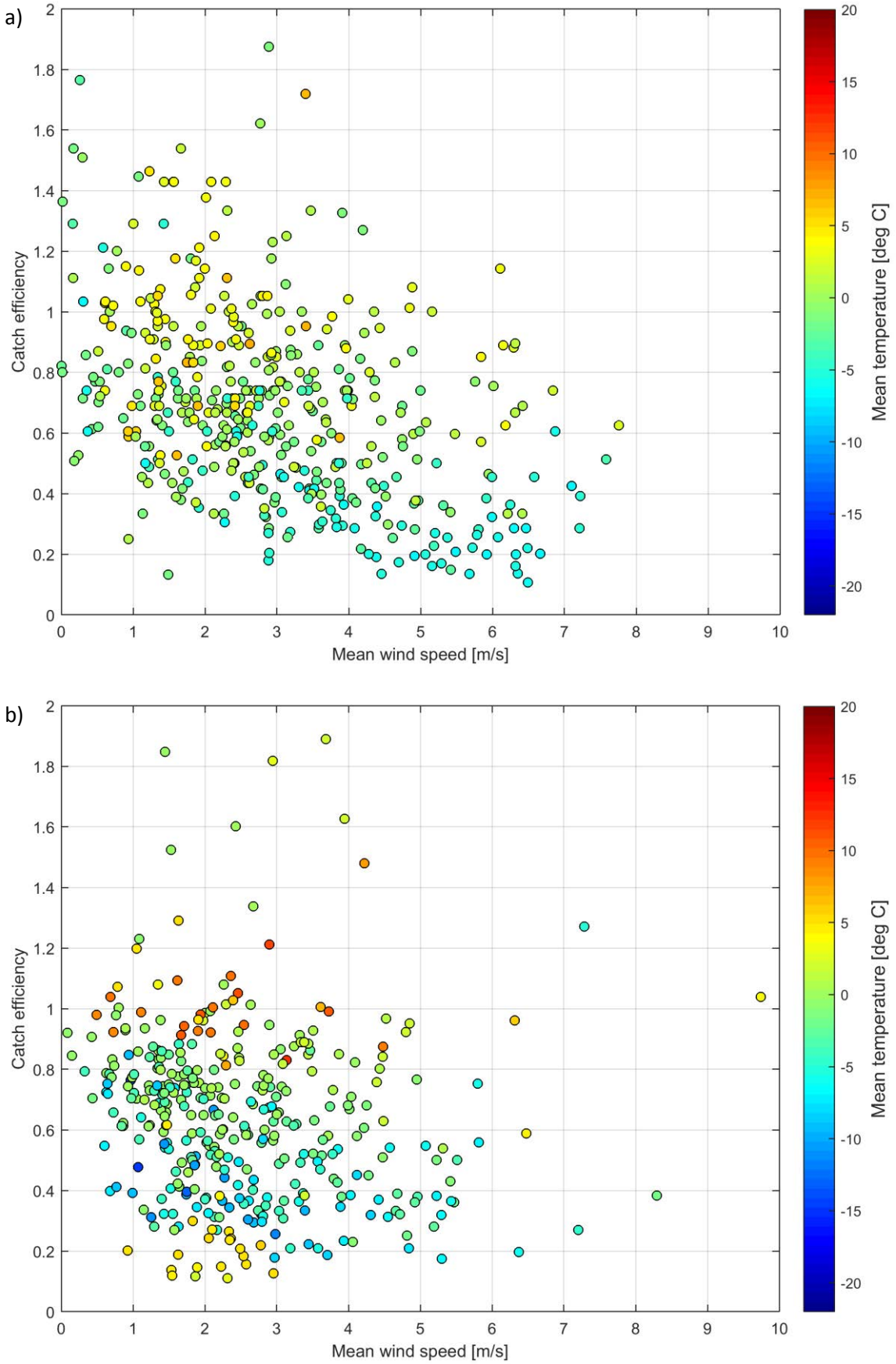


Figure 8: Catch ratio vs. mean wind speed scatter plots for 30 minute precipitation events at (a) Formigal and (b) Marshall. The mean event temperature is indicated by colour.

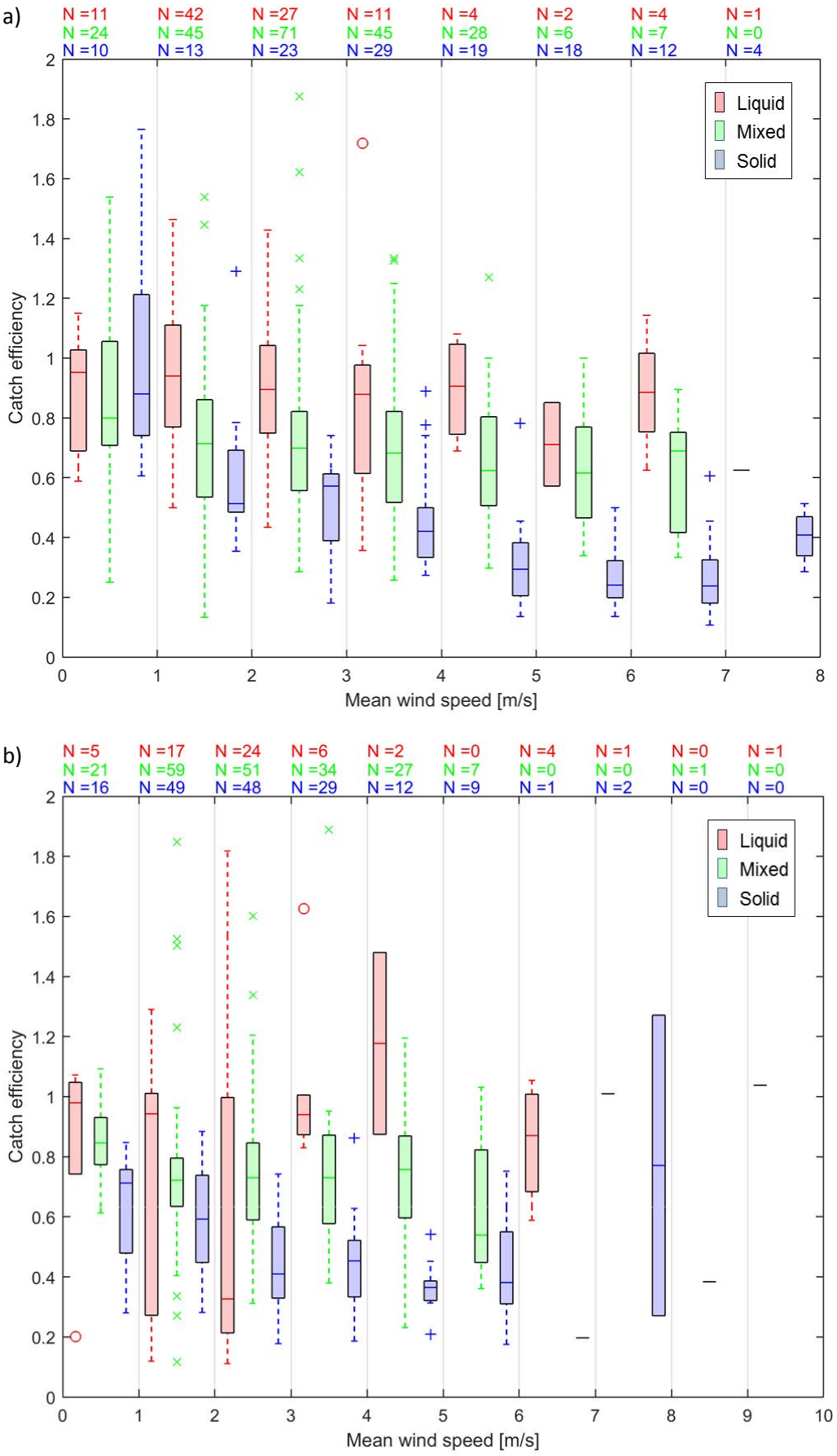


Figure 9: Catch ratio vs. mean wind speed box and whisker plots for 30 minute precipitation events at (a) Formigal and (b) Marshall. The predominant precipitation type for each event is determined from the maximum and minimum reported temperature and indicated by colour.

Table 6: RMSE values, by site, and by precipitation type, for YY cases.

Site (bucket resolution)	Time interval	RMSE (mm)			
		Liquid	Mixed	Solid	All precip types
Formigal (0.2 mm)	30 min	0.276	0.532	0.735	0.577
	60 min	0.226	0.869	1.269	0.955
Marshall (0.1 mm)	30 min	0.404	0.369	0.327	0.359
	60 min	0.630	0.532	0.508	0.538

The total accumulation reported by each site reference and SUT, for all assessment intervals during which both detected and reported precipitation, are presented alongside the corresponding catch ratios in Table 7.

Table 7: SUT and reference accumulation and overall catch ratio, assessed for YY cases over the entire test period, by site.

Site (bucket resolution)	Time interval	SUT accumulation	Reference accumulation	Overall catch ratio
Formigal	30 min	325.0 mm	521.0 mm	0.62
	60 min	354.4 mm	578.1 mm	0.61
Marshall	30 min	255.8 mm	361.5 mm	0.71
	60 min	272.1 mm	406.6 mm	0.67

6.3. Assessment of events when the reference and the SUT do not agree on the occurrence of precipitation

Assessment intervals during which the site reference and SUT do not agree on the occurrence of precipitation – namely, the YN and NY cases (Section 4.1.1) – are characterized using histograms in Figures 10 and 11. The histograms include accumulated precipitation, precipitation intensity as reported by the reference, and corresponding site conditions.

The total SUT and reference accumulations over the test period include contributions from YN and NY cases, which impact the overall catch efficiency. These contributions may be significant, given the response delays associated with heated tipping bucket gauges. Total accumulation and catch efficiency results presented in Table 7 (YY cases only) are expanded to include contributions from YN and NY cases in Table 8.

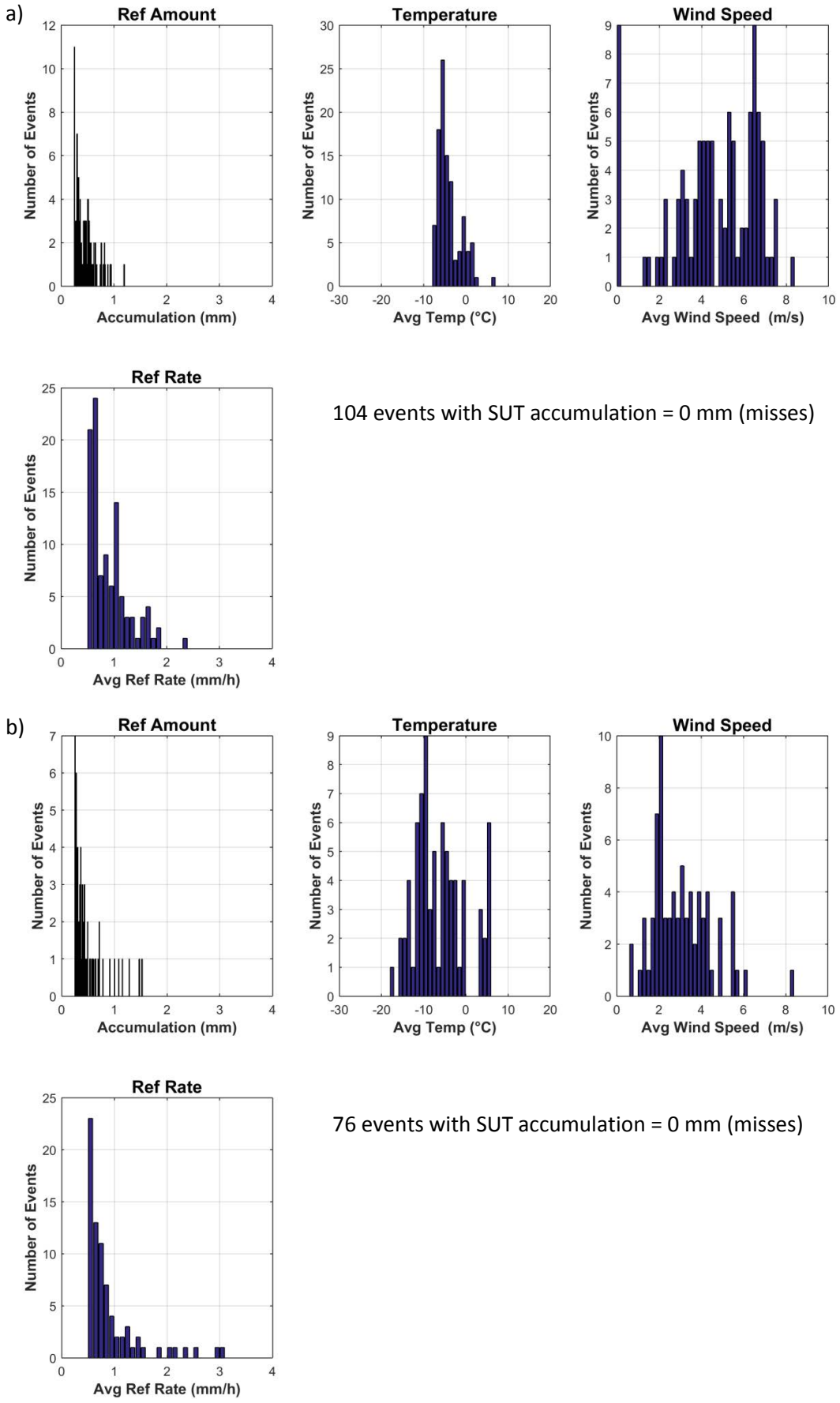


Figure 10: Histograms of reference accumulation, mean temperature, mean wind speed, and reference precipitation rate for YN cases at (a) Formigal and (b) Marshall.

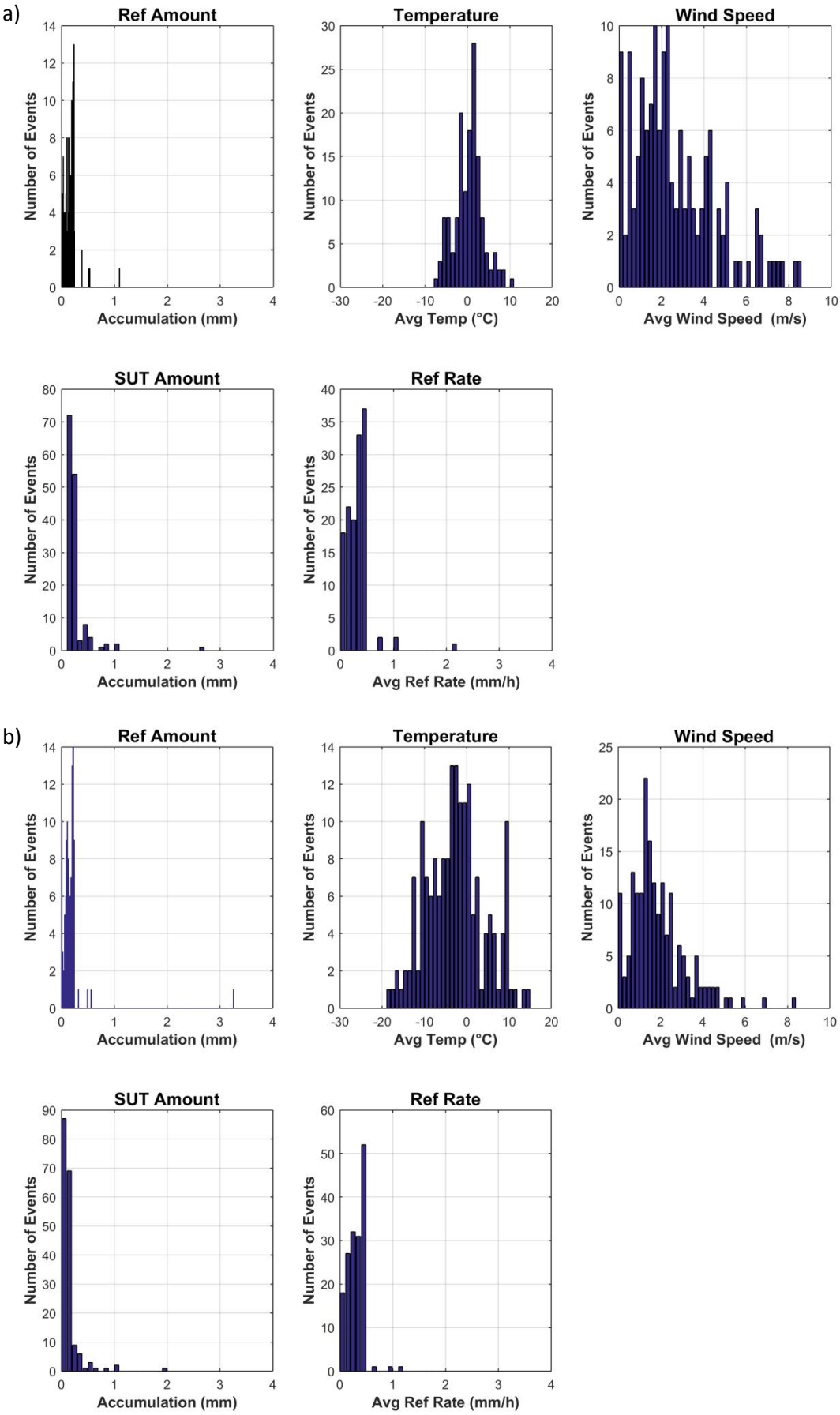


Figure 11: Histograms of reference accumulation, mean temperature, mean wind speed, SUT accumulation, and reference precipitation rate for NY cases at (a) Formigal and (b) Marshall.

Table 8: SUT and reference accumulation and overall catch ratio, assessed for the entire test period, for all YY, YN, and NY cases.

Site (bucket resolution)	Time interval	SUT accumulation	Reference accumulation	Overall catch ratio
Formigal	30 min	364.0 mm	589.3 mm	0.62
	60 min	371.2 mm	626.3 mm	0.59
Marshall	30 min	284.4 mm	423.9 mm	0.67
	60 min	287.1 mm	464.0 mm	0.62

7. Interpretation of Results

7.1. Operating conditions

The full range of conditions under which the test gauges were operated is illustrated in Figure 3. The conditions relevant to gauge operation are as follows:

- Temperatures between approximately -30 °C and 25 °C;
- Precipitation rates within 15 mm/hr.

These conditions fall within the manufacturer's specified range of operating conditions for the specific gauge models operated at each site (see Section 1).

The conditions during precipitation events (as identified by the reference configuration) are shown in Figure 4. The events at Formigal (valley in alpine region) are characterized by a narrower, and warmer, range of temperatures, with the distribution of wind speeds shifted toward higher values relative to Marshall (dry continental region). It should be noted that the wind speed observations at Marshall are made at 2 m, which is close to gauge height, while those at Formigal are made at 10 m, which is the standard configuration in the Spanish operational network.

7.2. Ability to detect and report precipitation

7.2.1. Skill score assessment

The performance of the sensors under test with respect to the detection of precipitation (relative to the reference configuration at each site) was considered in terms of skill scores, computed for both 30 and 60 minute assessment intervals (Table 4). Using the shorter 30 minute assessment intervals, the Probability of Detection (POD) is higher for both gauges under test relative to the scores for the longer 60 min assessment intervals, but the False Alarm Rate (FAR) and Bias (B) are also higher for both gauges. As a result, the Heidke Skill Score (HSS) values are higher for both gauges when 60 minute assessment intervals are used. The comparison of skill scores for the test gauges at each site will be considered for only the 60 minute assessment intervals, as hourly reports are broadly used, operationally.

The skill scores for 60 minute intervals in Table 4 illustrate similar overall performance for the different models of gauge under test at each site: POD of 82.9% for Formigal and 78.8% for Marshall; FAR of 12.4% for Formigal and 16.7% for Marshall; B of 94.6% for Formigal and 94.6% for Marshall; and Heidke Skill Scores (HSS) of 81.9% for Formigal and 80.2% for Marshall.

7.2.2. Characterization of response delays

Probability distribution functions (PDFs) of response delays between the onset of precipitation as observed by the reference configuration and the first response by the Thies Precipitation Transmitter heated tipping bucket gauge are shown in Figure 5. For the test gauge at Formigal, response times within 50 minutes have the highest probability, though delays up to 130 minutes were observed. For

the gauge at Marshall, response times within 40 minutes have the highest probability, with longer delays observed at lower probability, up to about 100 minutes. These differences can be attributed to the different gauge models employed, and to the different conditions at each site. In particular, evaporation of precipitation from the bucket in the 0.2 mm gauge (Formigal) is expected to be more significant than that from the bucket in the 0.1 mm gauge (Marshall). More time is required to fill the 0.2 mm bucket (for a given precipitation intensity), resulting in more time for heating and subsequent evaporation, and longer response times. In addition, the lower heating power for the test gauge at Formigal (49 W) relative to the gauge at Marshall (113 W), could result in longer times required to melt mixed and solid precipitation, resulting in longer response delays.

For precipitation occurring within a given 30 or 60 minute interval (as indicated by the reference configuration), the response delay PDFs for the test gauges at both sites indicate the potential for the TB response to occur within a subsequent 30 or 60 minute interval. The delays with highest probability in the PDFs are within 40 to 50 minutes, so it is more likely that the TB gauges will respond to precipitation detected by the reference within a 60 minute period than within a 30 minute period. This is reflected by the lower numbers of false alarm (NY) cases for the 60 minute intervals in Table 5, and the lower FAR and Bias for these intervals in Table 4.

Another important consideration is that even when both the SUT and reference detect precipitation within a given interval, the TB may actually be reporting precipitation collected during previous intervals, or may only report a fraction of the precipitation within that interval (reporting the rest during a subsequent interval). These ‘false YY cases’ will impact the assessment of TB gauge performance in terms of reporting accumulated precipitation (Section 6.2).

7.2.2.2. Influence of precipitation intensity, phase

The dependence of delay times on the mean intensity of precipitation preceding the TB response, as reported by the reference configuration, is illustrated in Figure 6. For the test gauge at Marshall (Figure 6b), response times are generally within 30 minutes for all precipitation types (liquid, mixed, solid) when the mean intensity of precipitation is within about 0.5 mm/hr. This suggests that for low intensities, the TB response is independent of precipitation phase, and that heating/melting of solid precipitation does not increase delay times significantly. For mean mixed and solid precipitation intensities exceeding ~ 0.5 mm/hr, however, the delay times are longer, extending to about 100 minutes, which can likely be attributed to heating/melting.

For the gauge under test at Formigal, response delays are within about 35 minutes for liquid precipitation across the range of observed intensities, but vary significantly for both mixed and solid precipitation, with delays of up to about 130 minutes. This variability is attributed to the lower heating power of the test gauge at Formigal, as well as the larger bucket capacity, both of which may exacerbate response delays relative to the test gauge with higher heating power and lower bucket capacity tested at Marshall.

7.3. Ability to measure and report precipitation

7.3.1. Performance when both SUT and reference detect precipitation

7.3.1.1. Temperature, phase, and wind speed influence

Focussing on the 30 minute events in which both the sensors under test and site reference configurations report precipitation (YY cases, or 'hits,' only), general stratification of the data by temperature is observed for Formigal and Marshall in Figures 7 and 8. Colder temperatures are generally correlated with larger differences in accumulation between the reference and SUT, and hence lower catch efficiencies, with increasing wind speed. This results from enhanced influence of wind speed on the collection of mixed and solid precipitation, relative to liquid precipitation. Also evident in Figure 7 is the influence of bucket resolution on the SUT accumulation values: 0.2 mm for Formigal (Figure 7a) and 0.1 mm for Marshall (Figure 7b).

Figure 8b shows a cluster of points in the Marshall data with warmer temperatures (about 3 to 4 °C; yellow points in plot) and low catch efficiencies that defies this general trend. An in-depth assessment of these specific events is beyond the scope of the present work; however, they may be attributed to events with mixed-phase precipitation containing a solid precipitation component or rain accelerating the melting process within the bucket.

The phase and wind speed influences on catch efficiency are more clearly illustrated in the box and whisker plots in Figure 8. For Formigal, the median catch efficiencies for solid precipitation range from about 0.5 to 0.7 for wind speeds ≤ 3 m/s, decreasing to values between 0.2 and 0.4 for wind speeds > 3 m/s. For Marshall, the median catch efficiencies for solid precipitation range from 0.5 to 0.7 for wind speed ≤ 2 m/s, falling to values between 0.3 and 0.5 for wind speeds > 2 m/s.

These general trends for solid precipitation characterize the maximum expected wind speed effects on catch efficiency for the test gauges at Formigal and Marshall under the conditions tested. As observed in Figure 9, the trends for mixed precipitation fall within these limits.

Figure 9b indicates significant scatter in events classified as 'liquid' at wind speeds < 3 m/s for the gauge at Marshall. These events correspond to the cluster of warmer events with low catch efficiencies noted above. The classification of precipitation phase by temperature is not definite; accordingly, mixed-phase events may be classified as liquid, accounting for the scatter observed.

7.3.1.2. Root mean square error

The RMSE was calculated for all 30 and 60 minute events during which the reference and SUT both detected precipitation (YY cases), for each site. The overall RMSE calculated for each site (all precipitation types), and RMSE values calculated for each precipitation type, are shown in Table 6. Given the broad use of hourly reports in meteorological operations, further discussion of RMSE values will be considered for 60 minute assessment intervals only.

The overall RMSE was approximately 1 mm for the test gauge at Formigal and 0.5 mm for the test gauge at Marshall. For the specific site and gauge configurations/models tested, over the range of

conditions tested, these values can be considered to represent the absolute uncertainty of the Thies Precipitation Transmitter relative to a DFAR configuration. The larger overall RMSE for the test gauge at Formigal is attributed to the longer response delays due to the larger bucket capacity and lower heating power.

The largest phase-specific RMSE values in Table 6 were determined for solid precipitation, owing to the larger wind-induced differences in reference and SUT accumulation relative to mixed and liquid precipitation. Differences in results between the sites can be attributed to differences in the configuration of gauges (bucket resolution, heating configuration) and differences in the conditions experienced at each site.

7.3.1.3. Overall catch efficiency

The total accumulated precipitation recorded by the reference and sensor under test at each site was compiled for all 30 and 60 minute events during which both detected precipitation (YY cases) and used to calculate the overall catch ratio. The results are provided in Table 7. Focussing on the more operationally-relevant 60 minute values, the overall catch ratios for the test gauges at Formigal and Marshall are 0.61 and 0.67, respectively. This difference can likely be attributed to the higher wind speeds observed at Formigal (which reduce the catch efficiency); however, these values demonstrate similar overall performance for the gauges operated at each site. An investigation of the influence of using 10 m wind speeds instead of gauge height wind speeds on the results for Formigal is beyond the scope of this work.

7.3.2. Characterization of conditions when the SUT and reference do not agree on the occurrence of precipitation

The conditions during 30 minute events in which the reference detected precipitation, but the SUT did not (YN cases), are depicted as histograms in Figure 10. These events are characterized by low precipitation rates, typically below 3 mm/hr, and temperatures close to, or below, freezing. At these temperatures, the precipitation is likely mixed or solid, and hence subject to longer response delays. Hence, these may be cases in which the TB response to precipitation occurs during a subsequent 30 minute period. There is also potential for heating to cause evaporation or sublimation of incident precipitation at low intensities.

The conditions during events in which the SUT detected precipitation, but the reference did not (NY cases), are shown as histograms in Figure 11. These events are also characterized by low precipitation rates, typically below 0.5 mm/hr (which corresponds to the reference accumulation threshold for a precipitation event over a 30 minute interval), and wind speeds below about 8 m/s. These events most likely result from delayed tipping bucket responses that coincide with intervals during which the reference did not detect precipitation.

Given the influence of response delays, estimates of the overall catch ratio for tipping bucket gauges should include the reference and SUT accumulations during YN and NY cases (and not just during the YY cases, as considered above). The resulting total accumulations for all YY, YN, and NY cases, and the corresponding overall catch efficiencies, are provided for the test gauges at each site in Table 8. The overall catch efficiencies for the gauges under test at both Formigal and Marshall were 0.59 and 0.62

for 60 minute assessment intervals, respectively, demonstrating similar overall performance of the gauges tested at each site.

8. Operational considerations

The overall experience with the Thies Precipitation Transmitter was positive at both Formigal and Marshall.

8.1. Maintenance

Each site completed the gauge field calibration and verification as per manufacturer recommendations, at least once a year. The calibration records have been stored by each site team.

8.2. Noted issues

- It was noted that heating was sometimes triggered only when snow touched the location of the funnel thermistor, resulting in response delays. This was corroborated by the manufacturer, and suggests that precipitation intensity could be a factor in determining the response time, as the snow level would reach the thermistor location more quickly for higher intensities.

- It was also noted and corroborated by the manufacturer that the Reed switch could be triggered under high wind conditions, particularly when the bucket was close to full, leading to false reports.

- Finally, the Formigal site team noted that under cold, windy conditions, the heating was sometimes insufficient to melt snow in the funnel using the 49 W configuration. As noted in Section 3, this configuration was selected to correspond with that used in the Spanish operational network, and may not be best configuration for the site conditions.

9. Performance Considerations

- Heating is critical to ensure proper functioning of the gauge under winter conditions. The voltage should be checked at the gauge location to ensure that the heaters are supplied with the recommended power. Voltage drops – for example, resulting from the use of long power cables – can impair the functioning of the gauge in winter conditions.

- Heating enables the measurement of solid precipitation, but the time required for melting can delay the time between the collection of precipitation in the funnel and the gauge response to that precipitation (response delays). Further, heating may result in the evaporation/sublimation of incident precipitation at low intensities.

- Response delays must be considered when using the gauge in operational settings. Ideally, the reporting interval (i.e. hourly observations) should exceed the maximum expected response delay; however, the potential remains for carry-over of precipitation accumulation from previous intervals, and for delayed responses occurring in subsequent intervals.

- While not tested in SPICE, shielding of the gauge should be considered by the manufacturer and/or potential users as a means of increasing the catch efficiency at higher wind speeds.

- The collection of ancillary measurement data is recommended: air temperature and wind speed (at gauge height) to enable the application of adjustments to measurements (i.e. using transfer functions); and reports from a precipitation detector with high sensitivity to enable the identification of missed events or false alarms due to response delays.

- The application of transfer functions to gauge measurements, if available, is recommended to account for reductions in catch efficiency with increasing wind speed. This recommendation comes with the caveat that the catch ratio data used in the derivation of the transfer functions will be impacted by response delays. Response delays can lead to scatter in the catch ratio data (among other factors), and hence, uncertainty in the transfer functions; however, the potential remains for the improvement of data via adjustment. Considering the catch ratios over longer time periods (e.g. 1 hour) may provide one avenue for reducing scatter in the catch ratio/wind speed relationship and the uncertainty of related transfer functions.

WMO-SPICE Instrument Performance Report
Campbell Scientific PWS100

1. Technical specifications

Instrument model:	Campbell PWS100 Present Weather Sensor
Measuring area:	40 cm^2
Physical principle:	Optical near-infrared laser based present weather sensor capable of determining particle type, precipitation and visibility parameters from hydrometeors size and velocity measurements and from the structure of the received signal.
Operating temperature range:	-25 to +50°C
Measurement uncertainty:	± 10 %
Sensitivity:	0.0001 mm (for rainfall)

Note: Specifications from manufacturer provided documentation.

Figure 1: Campbell PWS100 at Marshall (left) and at Haukeliseter (right) test sites.

2. Data output format

The present weather sensor PWS100 is a sensor outputting precipitation information, as intensity, accumulation or weather code, derived from raw data (particles size and fall velocity), which are also available to the user. Table 1 summarizes the main output parameters from the instrument. The firmware version of the PWS100 tested during this experiment was the version available at the time, before December 2014.

Table 1: Summary of main instrument outputs, as recorded by the sites during the experiment.

Measured Parameters	Units
Average Visibility	[m]
Present Weather Code (WMO)	-
Present Weather Code (METAR)	-
Present Weather Code (NWS)	-
Precipitation Intensity	[mm/h]
Precipitation Accumulation	[mm]
Average Particle Velocity	[m/s]
Average Particle Size	[mm]
Fault Code	-
Air Temperature	[°C]
Wet bulb Temperature	[°C]
Relative Humidity	[%]
# particles classified by precipitation type	[#/min]

This document reports on the ability of the PWS100 to derive solid precipitation. The results are consequently computed using the 'Precipitation Accumulation' output that corresponds to a 1 min accumulation report. The analysis has been performed on the cumulative sum of these 1 min accumulation reports.

3. SPICE test configuration

The Campbell PWS100, as sensor under test (SUT), has been tested on two different sites:

Test Sites: Haukeliseter (Norway); Marshall (USA)

Sensor Provider(s): All instruments evaluated were provided by the manufacturer (Campbell Scientific)

Figure 2: Map of SPICE sites testing Campbell PWS100 instruments.

A summary on the configuration of instruments as tested, the duration of tests and availability reflected in these results, and the ancillary measurements used, by site, is available in Table 2, Table 3, and Table 4, respectively.

Table 2: Summary of instrument configurations and data output, by site. Details and photos on individual site configurations are available in the respective site commissioning protocols.

	Haukeliseter	Marshall
Main prevailing wind directions	NE and E	btw N and E (during pcp)
Sensor Orientation	N-S	NNE
Height of installation	6 m	3 m
Heating	Yes, as recommended	
Shield	No	No
Data QC	SPICE QC methodology	
Data temporal resolution	1 min	
Processing interval for SPICE data analysis	30 min	

Table 3: Data availability, by measurement season, by site.

Measurement season	Haukeliseter	Marshall
Season 1 (Oct. 2013 – Apr. 2014)	✓	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓	✓

Table 4: Summary of reference and ancillary measurements, by site, with measurement height.

	Haukeliseter	Marshall
R2 Site Reference	Geonor 1000 (DFAR) (4.5 m, rim height)	Geonor 600 (DFAR) (3 m, rim height)
R2 Precip Detector	Thies LPM X2 (Site*, 2013-2014) (6 m) Thies LPM X5 (DFAR, 2014-2015) (4.5 m)	Thies LPM (Site*) (3 m)
Ancillary Temp Sensor	PT100 (4.5 m)	MetOne, model 060A-2/062, 2144-L (2 m)
Ancillary RH Sensor	PWS100 (6 m)	Campbell Scientific CS500 (3 m)
Ancillary Wind Sensor	Thies ultrasonic anemometer 3D (4.5 m) (2013-2014) RM Young Wind Monitor 05103 (4.5 m) (2014-2015)	RM Young Wind Monitor 05103 (3 m)

**A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR-fence; however, in cases where a more sensitive detector is available outside of the DFIR-fence, or there are issues with the detector within the DFIR-fence, a precipitation detector elsewhere on the site can be employed.*

4. Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6 of the SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute intervals:

- Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip;
- SUT 'Yes': SUT accumulation > 0 mm;
- SUT 'No': SUT accumulation = 0 mm.

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores.

4.1.2. Skill score assessment

The ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

The above scores are computed using the formulations provided in Section 3.6 of the SPICE Final Report.

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

5. Environmental conditions

The environmental conditions over the entire duration of the test period, at each site, are expressed as the probability density functions (PDFs) for air temperature, relative humidity, wind speed, wind direction, and precipitation rate in Figure 3. Figure 4 presents the same parameters during the precipitation events reported by the corresponding site reference, R2. The precipitation percentage in Figure 3 and Figure 4 represents the number of minutes of precipitation over a standard 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration at each site.

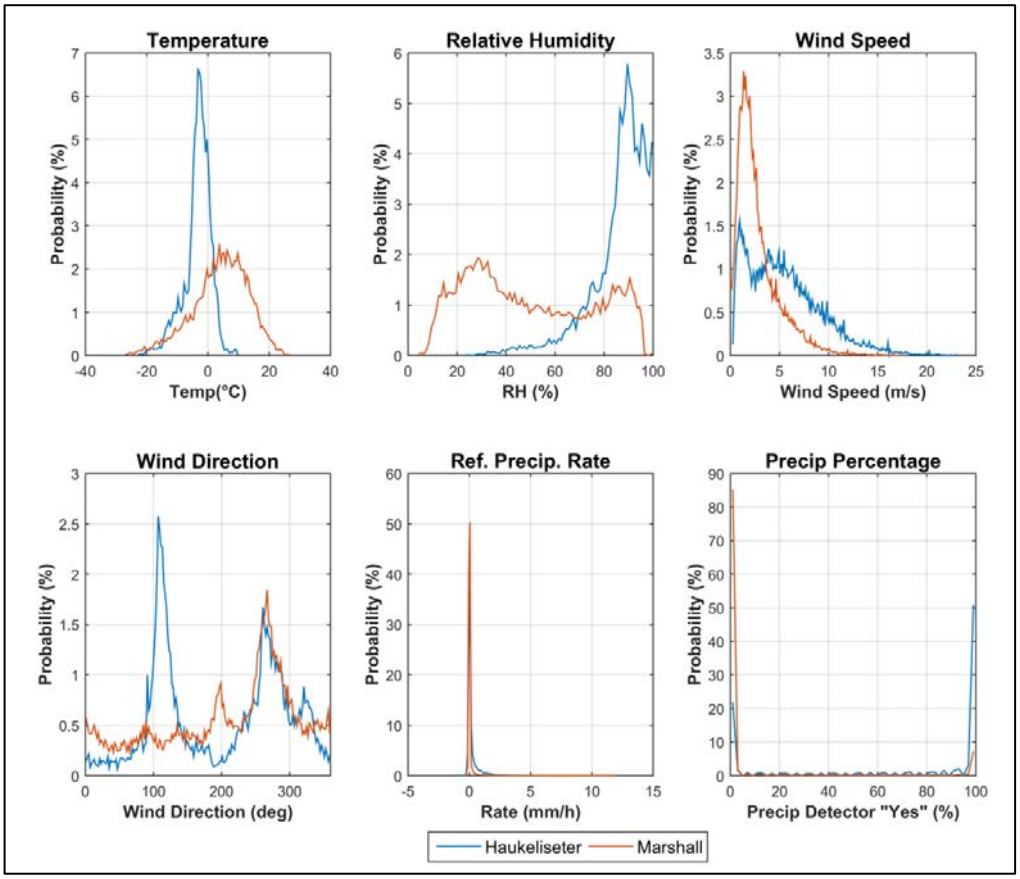


Figure 3: Summary of aggregated environmental conditions on the SPICE sites that operated Campbell PWS100, over the entire duration of tests, as per Table 3, above.

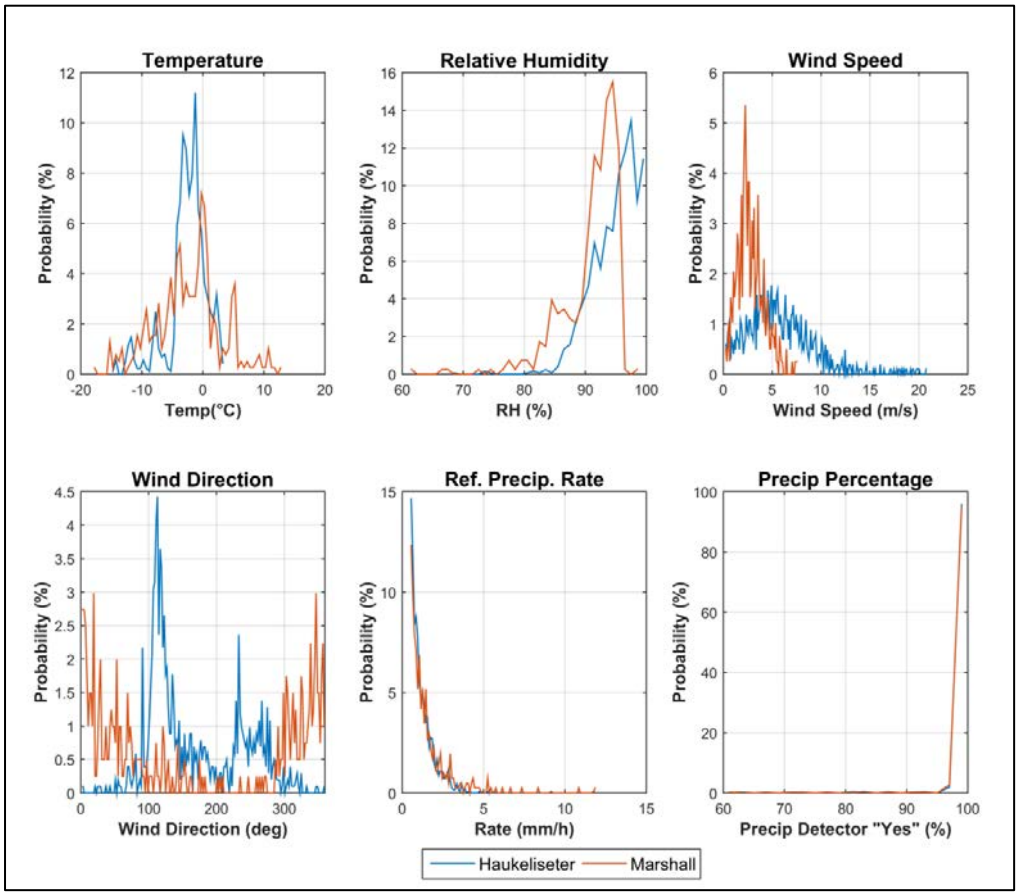


Figure 4: Summary of the aggregated environmental conditions on the SPICE sites that tested Campbell PWS100, corresponding to precipitation events, as reported by the site R2, reference, during the tests, as per Table 3 above.

6. Evaluation of performance over the range of operating conditions

6.1. Skill scores assessment

The ability of the SUT to represent precipitation similarly to the site field reference R2, is assessed using contingency tables (Section 4.1.1) and derived selected skill scores (Section 4.1.2). To better understand the potential influence of threshold choices on the derived results, two cases are considered here (see note below). The contingency results related to these two cases are given in Table 5 and the respective skill scores in Table 6, for both testing sites.

Note : Following the data derivation explained in Section 4.1.1, the conditions required to have a 'Yes' or a 'No' event over the 30 min interval, for the reference and the SUT, for the two different cases treated here, are:

CASE 1 *(as defined in Section 4.1.1)*:

- *Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip*
- *Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip*
- *SUT 'Yes': SUT accumulation > 0 mm*
- *SUT 'No': SUT accumulation = 0 mm*

CASE 2:

- *Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip*
- *Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip*
- *SUT 'Yes': SUT accumulation ≥ 0.1 mm*
- *SUT 'No': SUT accumulation < 0.1 mm*

Results presented in this report are based on Case 1.

Table 5: Contingency Tables: detection of precipitation of the Campbell PWS100 relative to the specific site reference, expressed as number of events over the entire test period, by site. The skill scores associated with these events are given in Table 6.

Haukeliseter Ref Geo1000DFAR **Case 1**

SUT PWS100	Yes	No	Total
Yes	1102	0	1102
No	162	716	878
Total	1264	716	1980

Marshall Ref Geo600DFAR **Case 1**

SUT PWS100	Yes	No	Total
Yes	395	17	412
No	11	13277	13288
Total	406	13294	13700

Haukeliseter Ref Geo1000DFAR **Case 2**

SUT PWS100	Yes	No	Total
Yes	1027	0	1027
No	237	716	953
Total	1264	716	1980

Marshall Ref Geo600DFAR **Case 2**

SUT PWS100	Yes	No	Total
Yes	363	1	364
No	43	13293	13336
Total	406	13294	13700

Table 6: Skill Scores for the Campbell PWS100, by site. POD: Probability Of Detection, FAR: False Alarm Rate, B: Bias, HSS: Heidke Skill Score (see Section 4.1.2 for more details).

Campbell PWS100, Skill Scores				
	POD	**FAR**	**B**	**HSS**
Case 1				
Haukeliseter	87.2%	0%	87.2%	83.1%
Marshall	97.3%	4.13%	101%	96.5%
Case 2				
Haukeliseter	81.3%	0%	81.3%	75.8%
Marshall	89.4%	0.275%	89.7%	94.1%

6.2. Assessment of SUT performance during non-precipitating events

The performance of the SUT in the absence of precipitation (when the reference precipitation detector recorded 30 minutes without precipitation) is represented in Figure 5 and Table 7, reflecting the distribution of the sensor response, as measured during the interval.

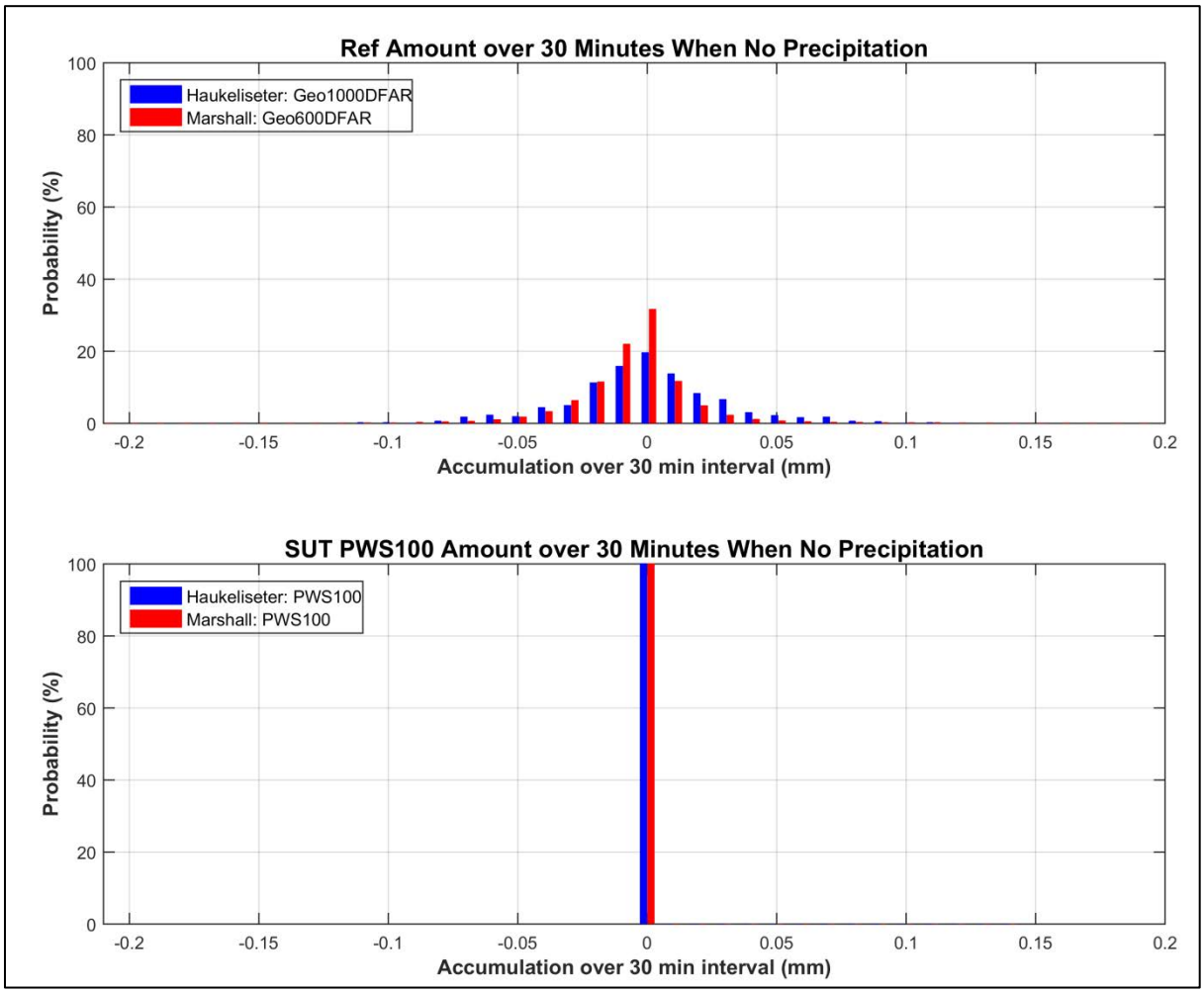


Figure 5: Probability of occurrence of a response during a 30 min interval in the absence of precipitation, represented by the signal output from (top) the R2 reference and (bottom) the SUT Campbell PWS100, by site. Statistics associated with these graphs are given in Table 7.

Table 7: Reference and SUT statistics of response signal when no precipitation was occurring, as plotted in Figure 5; Average (Avg), standard deviation (STD), maximum (Max) and minimum (Min) of the response signal, together with the number of events (Num) over the test period is given by site.

No Precip Statistics - Reference: Geo1000 DFAR (Haukeliseter) and Geo600 DFAR (Marshall)

	Ref Avg [mm]	Ref STD [mm]	Ref Max [mm]	Ref Min [mm]	Ref Num
Haukeliseter	0.004	0.030	0.119	-0.100	717
Marshall	-0.001	0.023	0.196	-0.207	13353

No Precip Statistics - SUT: PWS100

	SUT Avg [mm]	SUT STD [mm]	SUT Max [mm]	SUT Min [mm]	SUT Num
Haukeliseter	0.000	0.000	0.000	0.000	717
Marshall	0.000	0.001	0.137	0.000	13353

6.3. Ability of the SUT to measure precipitation

6.3.1. Yes-Yes cases

Quantitatively, the performance of the SUT to derive and report precipitation is assessed relative to the site reference in several graphs and tables illustrated in this section, using only the cases where both instruments reported precipitation over the 30 min interval, according to the criteria used in Case 1 of Table 5 (cases 'Yes-Yes', or shorter 'YY').

6.3.1.1. Time series plots

The time series (cumulative sum of 30 min YY events accumulation) of each individual SUT is plotted against their corresponding reference for the two seasons, by precipitation type (see Section 4.1.4) and for each site in Figure 6.

The corresponding seasonal accumulations are given in Table 8.

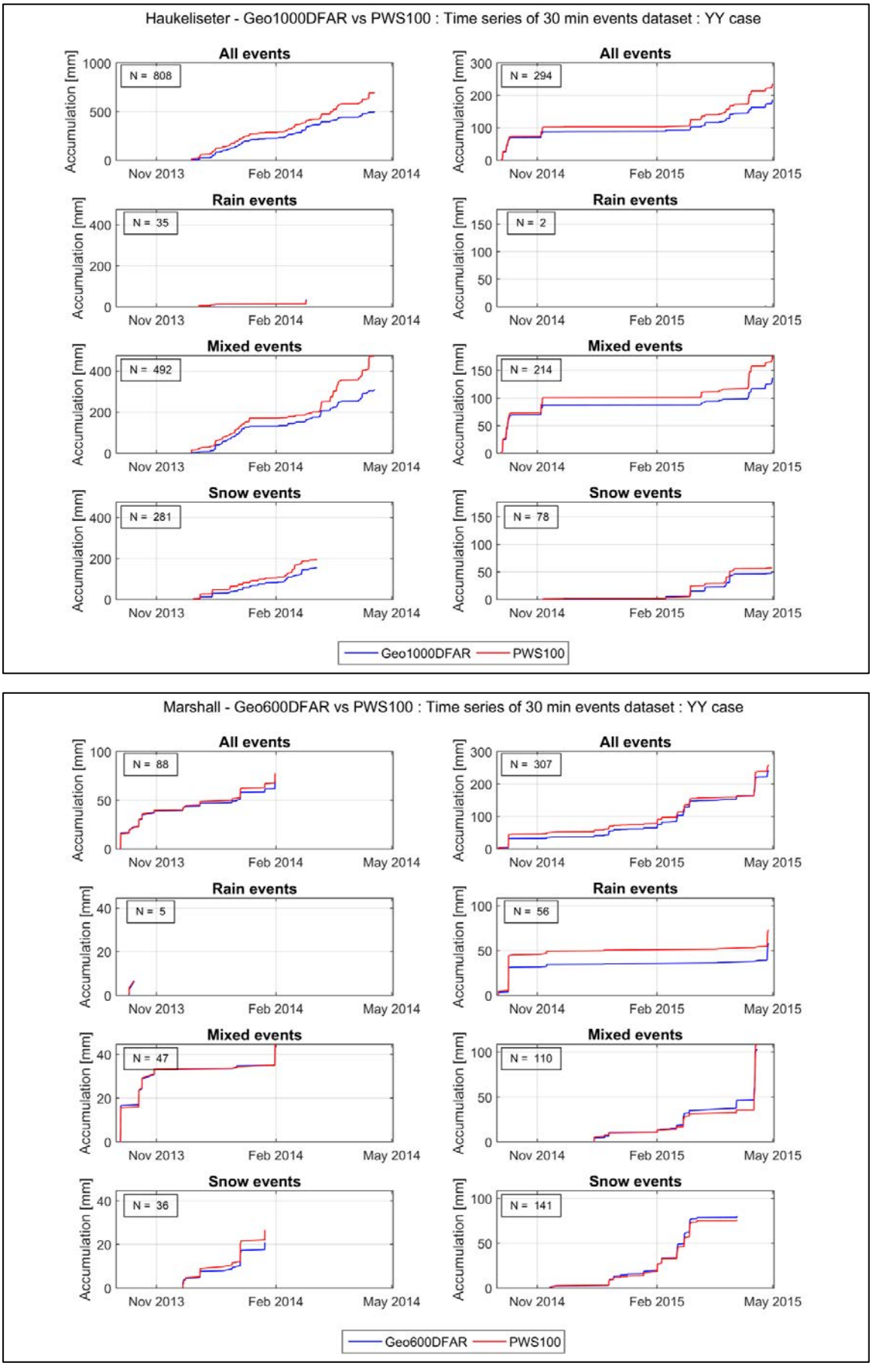


Figure 6: Time series based on 30 min YY events of the SUT Campbell PWS100 against the corresponding site reference, discriminated by precipitation type (Rain, Mixed, Snow), for both seasons (2013/14 on the left, 2014/15 on the right), for (top) Haukeliseter and (bottom) Marshall test sites.

Table 8: Seasonal accumulation [mm] for (a) Haukeliseter and (b) Marshall test sites based on the sum of YY events from the SUT Campbell PWS100 and the corresponding site field references R2: (a) Ref Geo1000 DFAR, (b) Ref Geo600 DFAR.

(a) Haukeliseter

	Season 2013/14		Season 2014/15	
[mm]	Ref	SUT	Ref	SUT
All events	496.78	696.79	184.31	234.39
Rain events	34.10	28.22	0.91	0.35
Mixed events	308.21	474.55	135.46	176.77
Snow events	154.48	194.02	47.94	57.27

(b) Marshall

	Season 2013/14		Season 2014/15	
[mm]	Ref	SUT	Ref	SUT
All events	70.19	77.43	240.03	256.96
Rain events	6.12	6.50	57.45	72.59
Mixed events	43.42	44.53	102.71	108.58
Snow events	20.65	26.39	79.86	75.79

6.3.1.2. Scatter plots and RMSE values

Scatter plots of the amount derived by the SUT versus the corresponding reference amount, and discriminated by precipitation type, is given for both sites in Figure 7.

Quantitatively, the SUT performance is assessed using the Root Mean Square Error (RMSE) also known in practice as Operational Comparability. The results are available for both sites in Table 9.

Table 9: Root Mean Square Error (RMSE) statistics, in mm, for the SUT Campbell PWS100 with respect to the corresponding site reference, by precipitation type and by site, including both seasons data.

RMSE [mm]	All	Rain	Mixed	Snow
Haukeliseter	0.740	0.314	0.691	0.817
Marshall	0.558	0.688	0.697	0.343

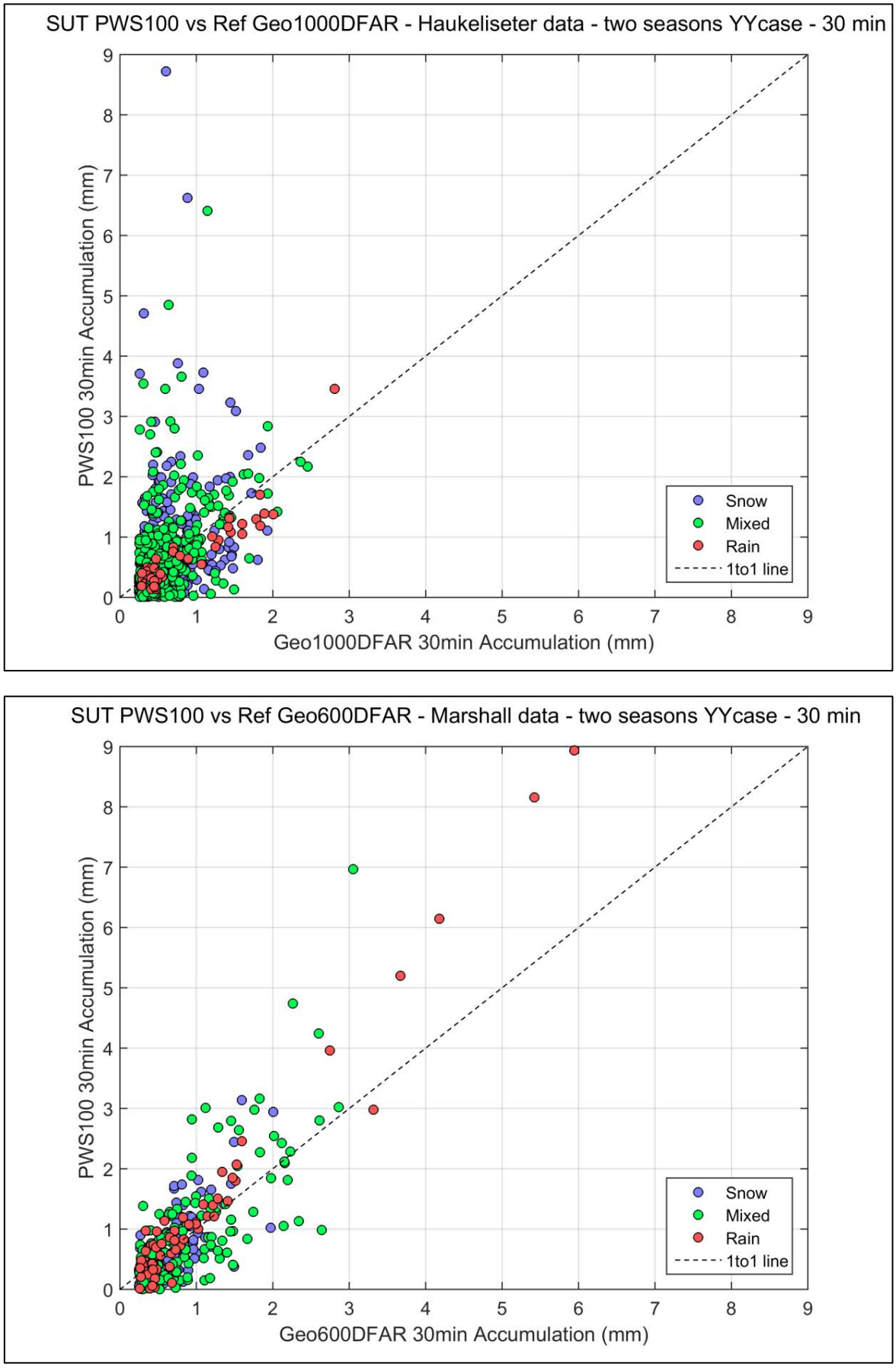


Figure 7: Scatter plots based on 30 min YY events accumulation from the SUT Campbell PWS100 against the corresponding site reference, for the two seasons, discriminated by precipitation types, for (top) Haukeliseter and (bottom) Marshall test sites.

6.3.1.3. Catch efficiency evaluation by precipitation type

The Catch Efficiency (CE) of the SUT is represented by histograms (Figure 8) and boxplots (Figure 9), both discriminated by precipitation type, and representing both sites.

The quantitative evaluation of the CE is provided in Table 10. The mean catch efficiency is given in the first line of this table for each site, considering both seasons data and for each category of precipitation type as well as for all the events together.

Note: All events with a CE greater than 3, if any, are included in one category named '3 and more' or '≥3' in the upcoming graphs. Additionally, for all graphs representing the CE, a dashed black line is added at CE = 1, which represents the ideal case where the SUT reports exactly the same precipitation amount as the reference.

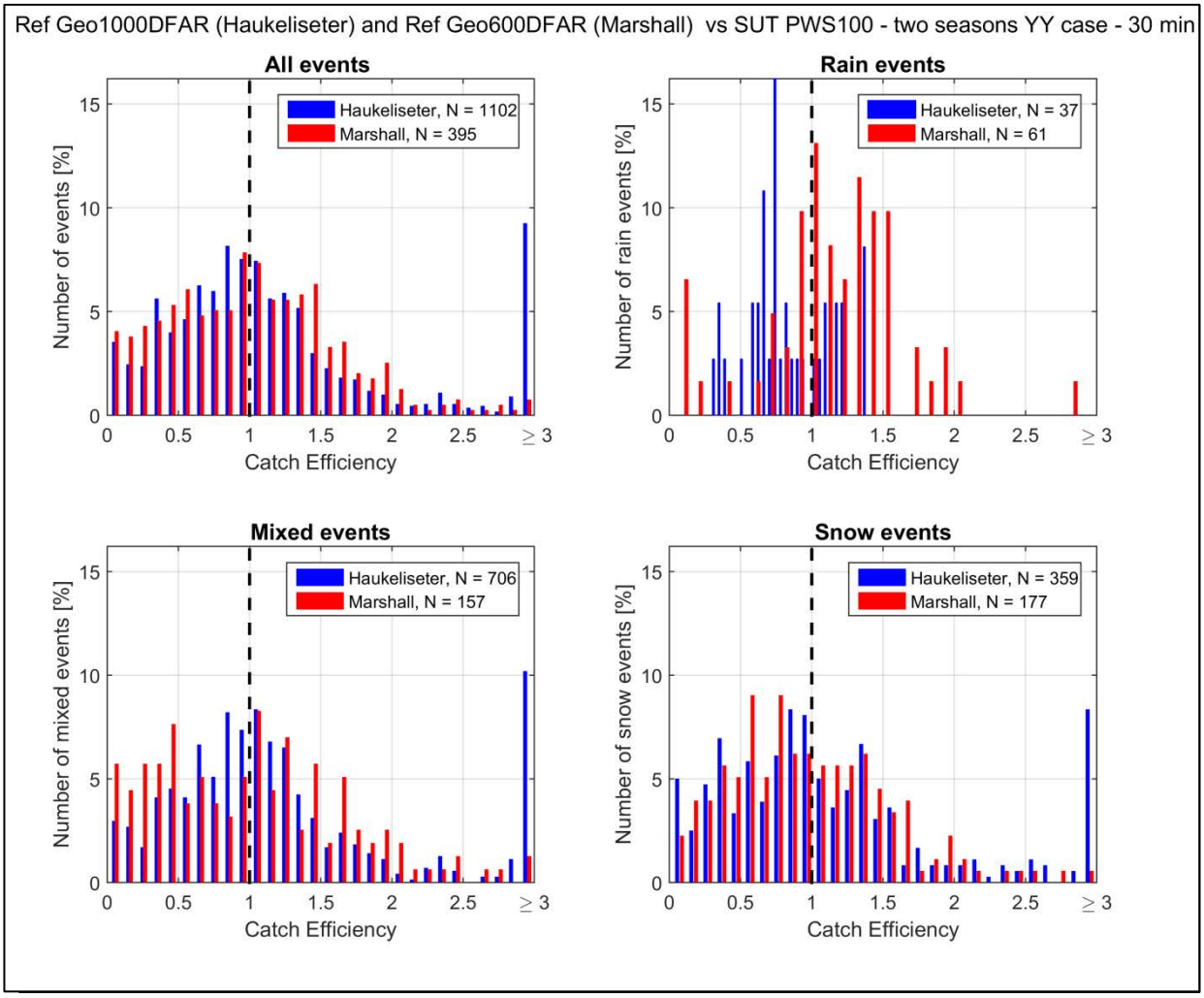


Figure 8: Histograms based on 30 min YY events for the two seasons, representing the distribution of the catch efficiency of the SUT Campbell PWS100 against the corresponding site reference (SUT/Ref), discriminated by precipitation type, with number of events given in the legend for each category, and represented for both sites. The dashed black line at CE = 1 represents the ideal case.

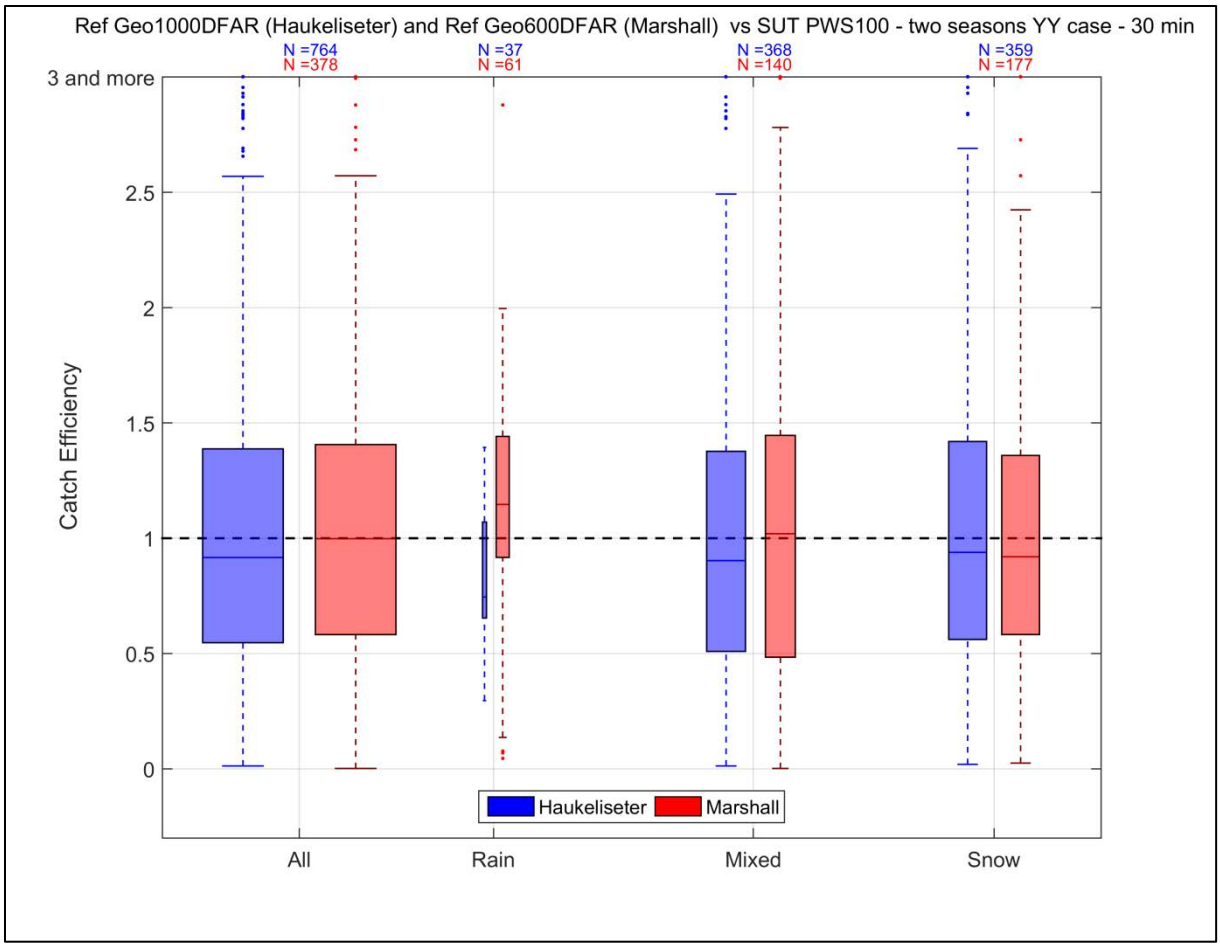


Figure 9: Boxplot based on 30 min YY events of the two seasons, representing the catch efficiency of the SUT Campbell PWS100 against the corresponding site reference (SUT/Ref), discriminated by precipitation type, with number of events given at the top of each category, and represented for both sites. The width of the boxes is proportional to the percentage of events in each category ('All' boxes represents 100% data for each site, making them having the same width despite the difference in number of events). The dashed black line at CE = 1 represents the ideal case.

Table 10: Statistics related to the CE of SUT Campbell PWS100 against the corresponding site reference (SUT/Ref, see Figure 9), discriminated by precipitation type.

Catch Efficiency Statistics: **Haukeliseter**

CE Boxplot Parameters	All	Rain	Mixed	Snow
Mean	1.25	0.83	1.24	1.30
Median	0.92	0.75	0.90	0.94
75 percentile	1.39	1.07	1.38	1.42
25 percentile	0.55	0.65	0.51	0.56
Upper Whisker	2.57	1.39	2.49	2.69
Lower Whisker	0.01	0.30	0.01	0.02
Maximum	15.14	1.39	11.46	15.14
Minimum	0.01	0.30	0.01	0.02
# Outliers	70	1	35	32
# Outliers ≥ 3	57	0	29	28
# Events	764	37	368	359

Catch Efficiency Statistics: **Marshall**

CE Boxplot Parameters	All	Rain	Mixed	Snow
Mean	1.04	1.15	1.05	0.99
Median	1.00	1.15	1.02	0.92
75 percentile	1.41	1.44	1.45	1.36
25 percentile	0.58	0.92	0.48	0.58
Upper Whisker	2.57	2.00	2.78	2.42
Lower Whisker	0.00	0.14	0.00	0.03
Maximum	4.59	2.88	4.59	3.33
Minimum	0.00	0.05	0.00	0.03
# Outliers	7	4	2	3
# Outliers ≥ 3	2	0	1	1
# Events	378	61	140	177

6.3.1.4. Catch efficiency dependency on wind speed

The variation of the SUT catch efficiency (SUT/Ref) with wind speed is illustrated in a scatter plot in Figure 10, together with a boxplot in Figure 11, both discriminated by precipitation type and for each site.

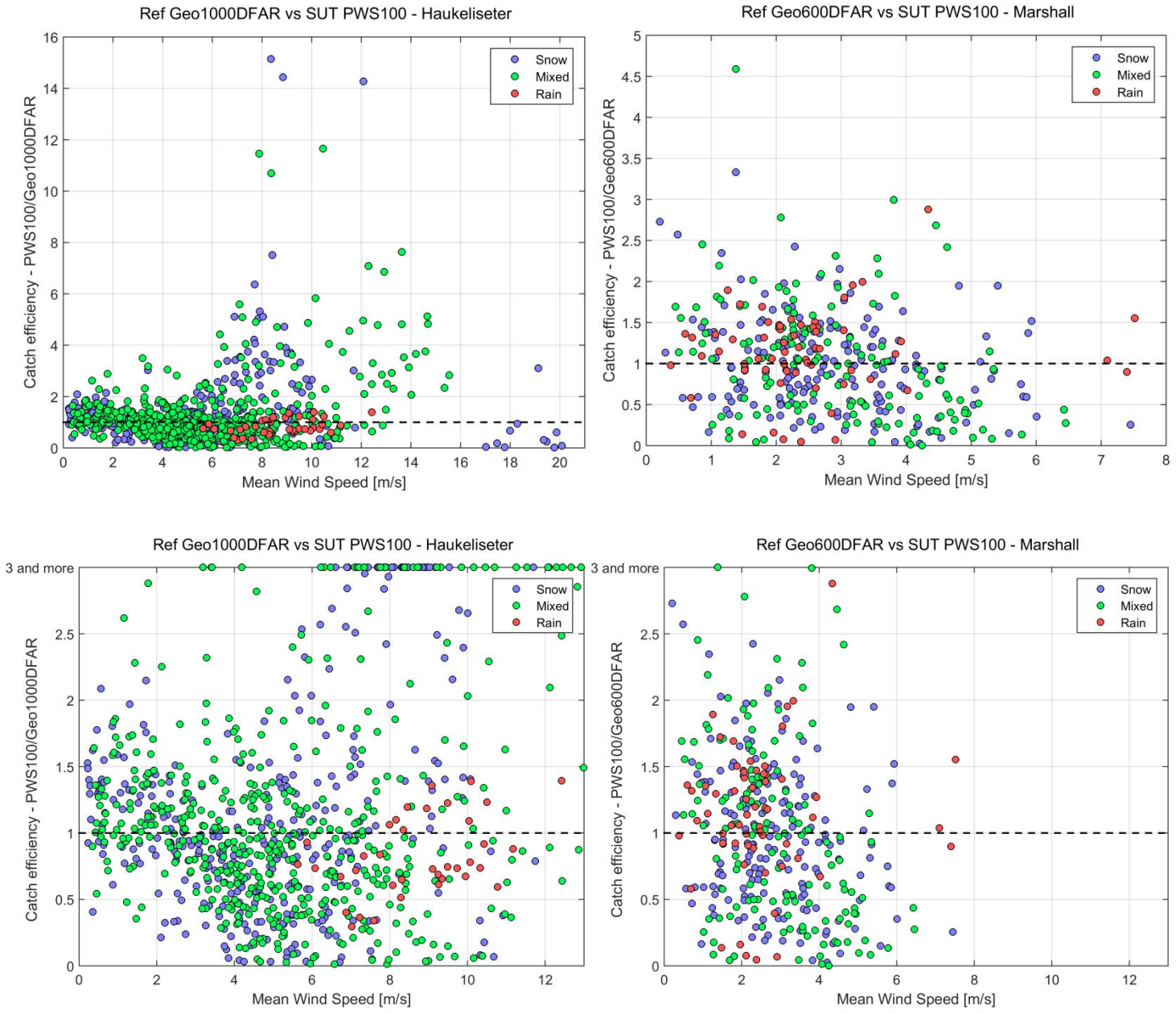
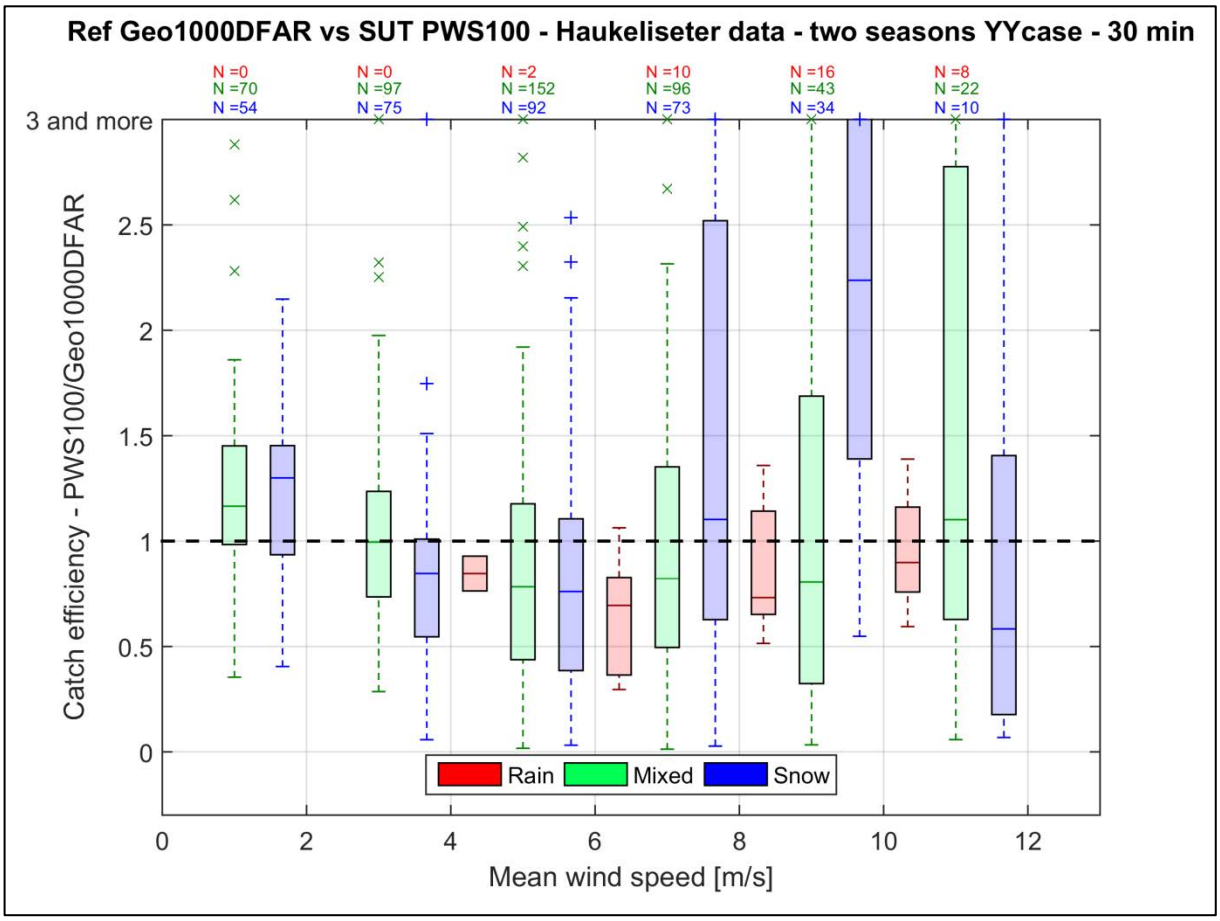


Figure 10: Scatter plots based on 30 min YY events for the two seasons, representing the catch efficiency of the SUT Campbell PWS100 with respect to the corresponding site reference (SUT/Ref), against wind speed and discriminated by precipitation type. Full x and y scale range (top) and constrained axes (bottom) is given to allow comparison between the two sites. On the left column, graphs are for Haukeliseter site, on the right, they represent Marshall site. The dashed black line at CE = 1 represents the ideal case.



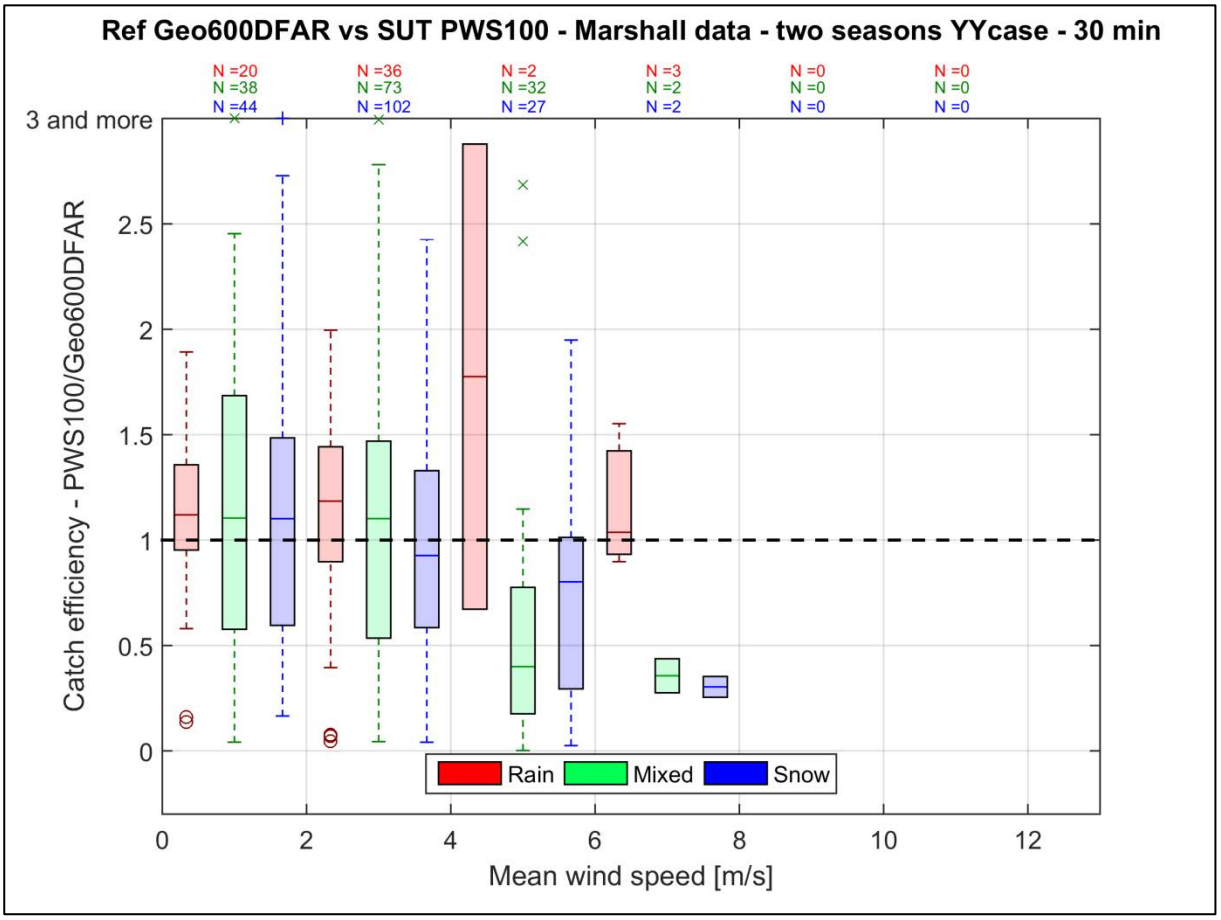


Figure 11: Boxplots based on 30 min YY events for the two seasons, representing the catch efficiency of SUT Campbell PWS100 with respect to the corresponding site reference (SUT/Ref), against wind speed and discriminated by precipitation type for (top) Haukeliseter and (bottom) Marshall test sites. The dashed black line at CE = 1 represents the ideal case.

6.3.1.5. Catch efficiency dependency on wind direction

In order to assess the dependency of the CE with wind direction, a wind rose is produced (Figure 12) representing, for each site, the wind data of the two seasons, binned by catch efficiency in order to represent undercatch (CE < 0.8), overcatch (CE > 1.2) and catch efficiency of 1 ± 20% of the SUT.

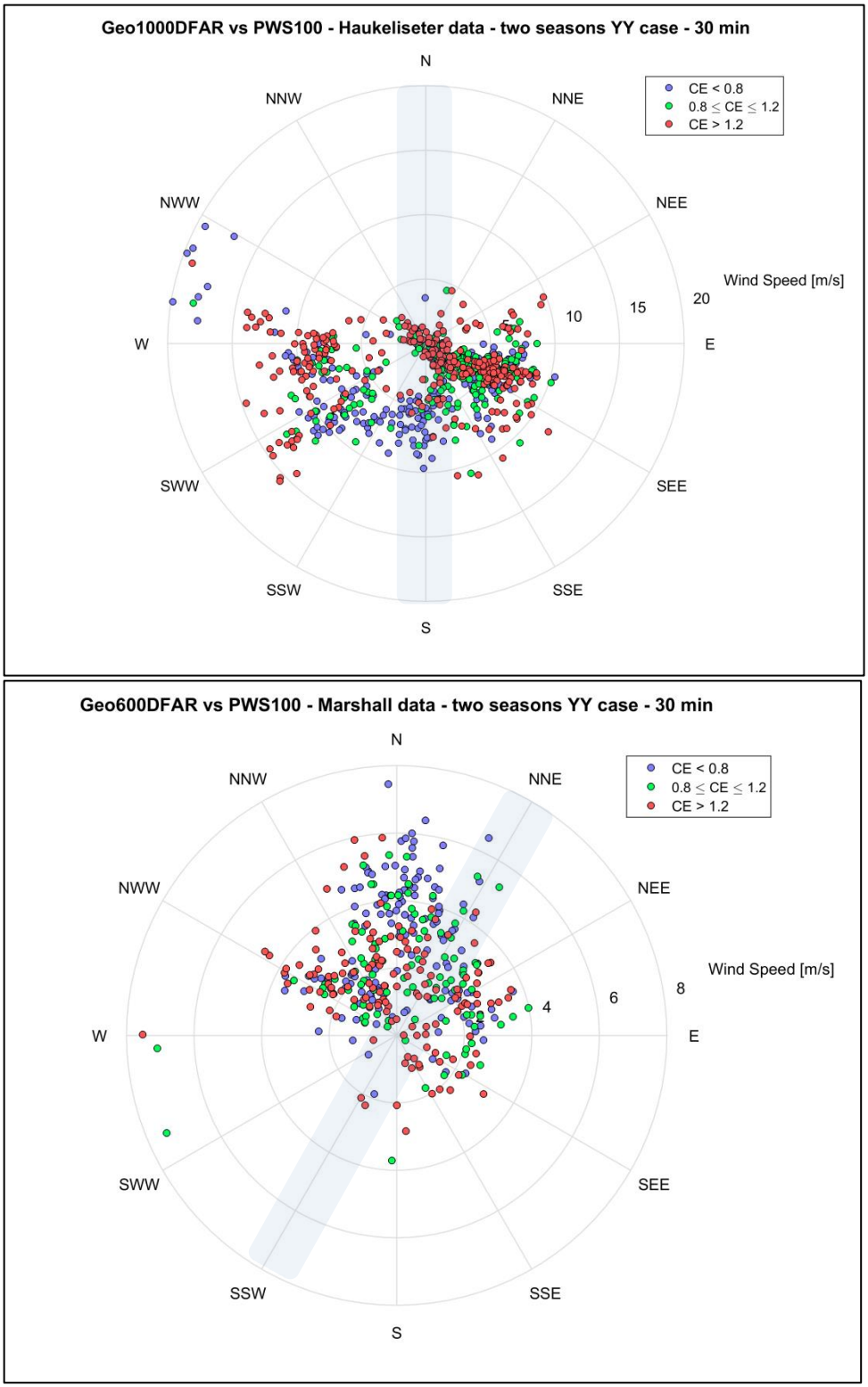


Figure 12: Precipitation events (YY cases) as function of wind speed and direction of (top) Haukeliseter and (bottom) Marshall test sites, binned by catch efficiency (CE). The grey zone indicates the SUT orientation.

6.3.2. Yes-No and No-Yes cases

Events when the site reference and SUT do not agree on the occurrence of precipitation includes two categories of cases (Section 4.1.1): (1) when the field reference reported a precipitation event, while the SUT did not (Yes-No cases, 'YN'), and (2) when the field reference did not report a precipitation event while the SUT did (No-Yes cases, 'NY').

Histograms illustrating field reference and SUT reports and associated site conditions for all YN and NY cases of both sites during the test period are provided in Figure 13 and Figure 14, respectively.

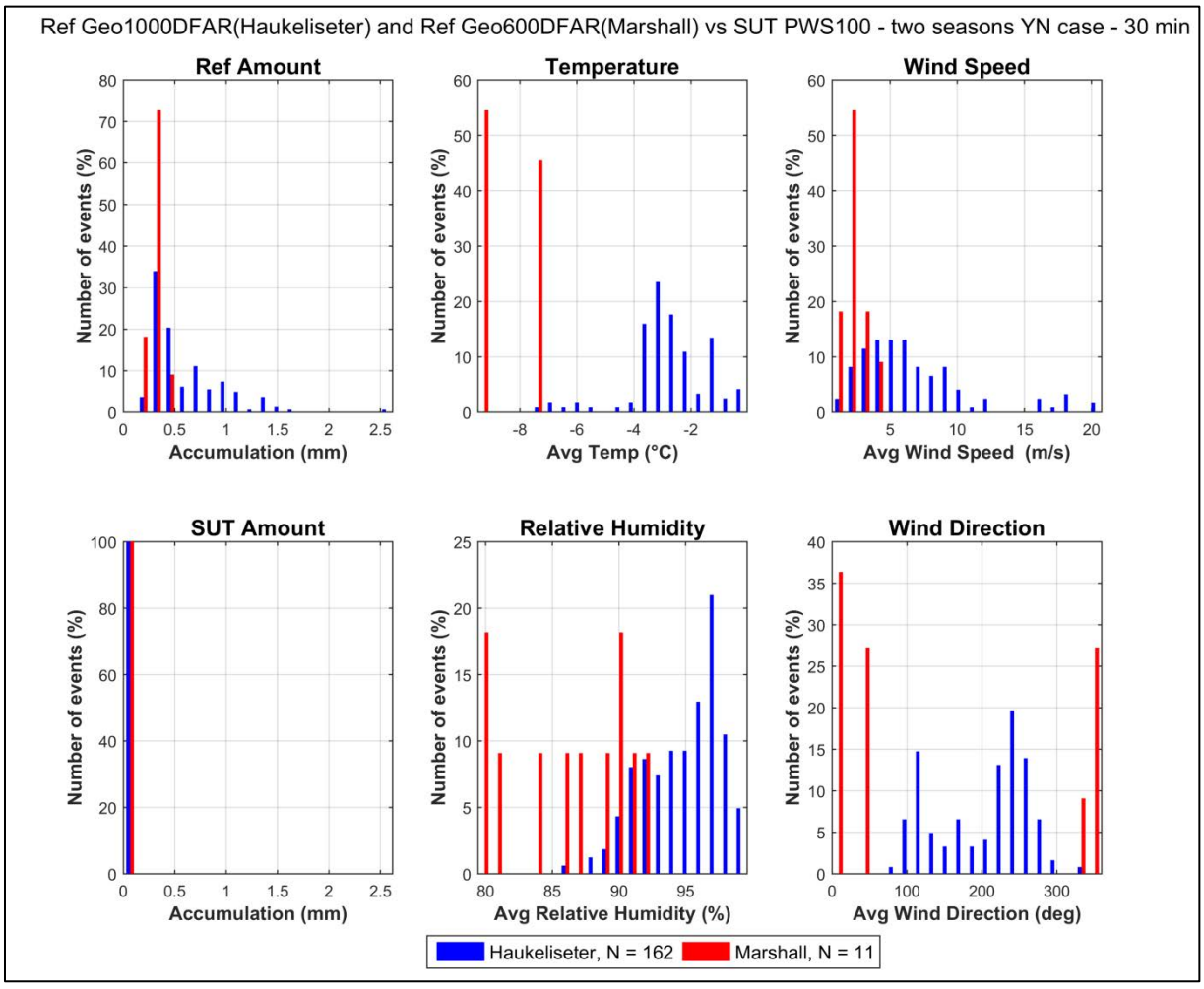


Figure 13: Histograms of SUT Campbell PWS100 and corresponding site reference accumulations (left column), along with distributions of mean temperature, mean relative humidity, mean wind speed, and mean wind direction for all YN cases (number indicated in the legend) of the 30 min intervals, as reported by Case 1 of Table 5, at the two different test sites.

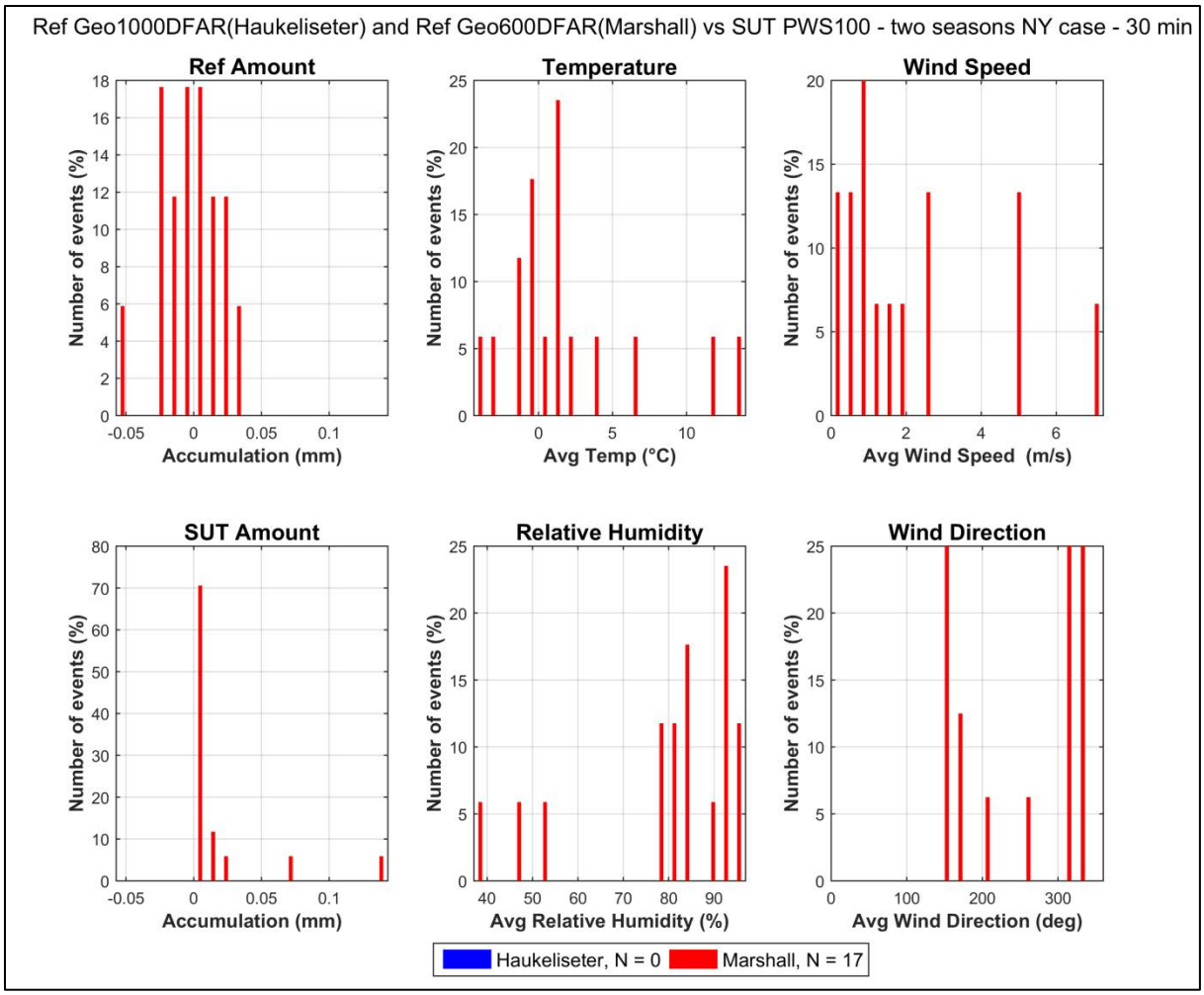


Figure 14: Histograms of SUT Campbell PWS100 and corresponding site reference accumulations (left column), along with distributions of mean temperature, mean relative humidity, mean wind speed, and mean wind direction for all NY cases (number indicated in the legend) of the 30 min intervals, as reported by Case 1 of Table 5, at the two different test sites.

7. Interpretation of results

7.1. Operating conditions

The two sites hosting the PWS100 account for different climates, Haukeliseter being characterized by an alpine regime, while Marshall is led by a continental climate. The main difference between the two is the wind speed distribution, as shown in Figure 3 and Figure 4. Wind speed is generally higher at Haukeliseter, with maximum ranging up to 15 m/s (5 m/s for Marshall). Haukeliseter is also showing lower temperature (no event above 2-3°C) and higher relative humidity (more than 95% of the events with RH above 90%).

7.2. Reliability in detecting precipitation

The PWS100 was efficient in detecting precipitation when the reference reported precipitation, as presented in Case 1 of Table 6 (when no threshold is applied to the SUT accumulation), with high POD (87% for Haukeliseter and 97% for Marshall) and low FAR (0% for Haukeliseter and 4% for Marshall). The lower POD for Haukeliseter results from the YN missed cases of the PWS100 compared to the reference and discussed further in detail in Section 7.5.
Applying a threshold of 0.1 mm over the 30 min events to the SUT (Case 2) reduces the POD of both sites due to a higher number of 'artificial' YN events (see Section 7.5 for more details), but still keeping a high value of more than 80% and the FAR from Marshall gets closer to 0%. As the number of YN events is higher in this case, the bias and HSS scores get slightly degraded accordingly. As a conclusion, applying a threshold to the SUT accumulation seems, overall, to not help giving better results for the PWS100, except freeing the user from the few NY events of unknown origin (either real or artefact).

The low total number of events selected for Haukeliseter over the two seasons (only 1980, see Table 5) is due to intermittent operation of the SUT during the second season (see Section 8.2 below).

7.3. Performance of SUT during no-precipitation events

The PWS100 has a stable output signal, showing no noise during no-precipitation events (see Figure 5 and Table 7). Note that the no-precipitation study is performed using only the reference precipitation detector data (here Thies LPM sensors for both sites) with the criteria ensuring that 30 min of 'no-precipitation' was recorded. For these cases, the PWS100 indicates 0 mm 100% of the time for Haukeliseter and 99.9% of the time for Marshall (see Figure 5), indicating a good agreement with the reference precipitation detector on the absence of precipitation.

7.4. Performance of SUT during precipitation events

When both the reference and the PWS100 report precipitation (YY cases), the results indicate different performance of the SUT, on the two sites, to derive solid precipitation, with an average catch efficiency near 1 (0.99) for Marshall, and a clear overcatch for Haukeliseter with an average CE of 1.30 (see Table 10). This tendency is also valid for mixed precipitation. Catch efficiency during rain events (small number of events on both sites) shows another pattern, with undercatch at Haukeliseter (0.83) and overcatch at Marshall (1.15).

From Figure 7 and Figure 10 it can be seen that a lot of snow and mixed events at Haukeliseter occurring under high winds (> 6 m/s) result in a large overcatch, with CE of 3 and more. For wind speed up to 6 m/s, the behaviour of the sensor is similar for the two sites, showing a very large scatter below and above the ideal case of a CE equal to 1, with CE varying randomly between 0 and 2 (see Figure 8 and Figure 10). There seems to be no relation with specific environmental conditions to explain this scatter (i.e. to relate under and over catch with typical environmental conditions).

This large scatter is also quantitatively shown by the RMSE (Table 9), which reflects the variability of measurements from the SUT against the site reference, by precipitation type. Here again, the large number of outliers for Haukeliseter at high winds causes the RMSE for snow to be much higher than for Marshall (0.817 mm vs 0.343 mm), reflecting the higher scatter for snow events on Figure 7.

Overall, the SUT seems to be a reliable instrument to account for the total accumulation over a longer period (e.g. one season), with a mean catch ratio around 1 (see Figure 9, Figure 11 and Table 10). This is especially true for sites with maximum wind speed at around 6 m/s (30 min average). The large scatter on event based statistics (typically 30 min interval), occurring at all wind speeds (see Figure 10), tends to show, though, that the SUT is less reliable to derive solid and mixed precipitation accumulation over near real time periods. This behavior makes it hard to correct or adjust the data, because the bias is inconsistent for a given wind speed or condition.

Even if better in deriving accumulation for longer time periods, the PWS100, as other non-catchment type instruments, doesn't have an assurance of the continuity in the measurements, though, which is critical to long term data collection. If power is off or signal transmission is interrupted, no data is recorded and there is no possibility to know what has fallen during this time, thus affecting the long term data reports of the PWS100.

The results from both sites in Figure 11 show a continuous and consistent decrease of the catch efficiency with increasing wind speed up to 6 m/s under snow and mixed precipitations. As mentioned above, the behavior of the instrument for higher winds (Haukeliseter) shows much higher scatter and generally outliers with catch ratio of 3 and more.

The wind roses (see Figure 12) show no clear dependency of the catch efficiency with wind direction for the Marshall site. For Haukeliseter, it seems that most of the undercatch cases are registered during Southerly wind situations. As the two sensor heads were positioned in the South direction, a shadowing of the measurement area may have happen for Southerly winds that may explain this result. In both cases, the sensor was installed perpendicular to the prevailing wind direction during precipitation events. Impact on catch efficiency for a sensor parallel to main flow has not been assessed.

7.5. Assessment of Yes-No, No-Yes events

The assessment of YN, NY events completes the picture of the performance of the SUT. As it can be seen in Table 5, when no threshold is applied to the SUT accumulation (Case 1), on both sites, there are very few events when the reference and the SUT do not agree on the occurrence of precipitation (except for the 162 YN events from Haukeliseter, explained further below). This confirms the results from Section 7.2 above. When applying a threshold to the SUT (Case 2), the number of YN events

from both sites increases (especially for Haukeliseter with 237-162 = 75 more YN events for Haukeliseter and 43-11 = 32 more YN events for Marshall), but this augmentation (compared to Case 1) reflects more cases of undercatch than real misses, since the PWS100 recorded, for these, less than 0.1 mm, but more than 0 mm, and the reference system recorded more than 0.25 mm and 18 minutes of precipitation over the 30 min intervals. These 'new' YN events (compared to Case 1) are therefore 'artificial' and a direct consequence of the choice of the SUT threshold. On the other hand, applying a threshold to the SUT has a positive impact on the few NY cases recorded for Marshall, which are diminished to only one event, and prevents to deal with NY events of unknown origin (either real or artefacts).

Looking more closely at Case 1, the number of misses for Haukeliseter (YN cases) is relatively high (162 vs 11 for Marshall). All these events were tracked back, and appear to be "real" misses from the SUT, where it indicated 0 mm of precipitation when the reference collected more than 0.25 mm and where the precipitation detector recorded more than 18 min of precipitation within a 30 min interval. These cases are associated with high relative humidity and lightly negative temperatures (Figure 13). They appeared mainly during the first season, in December 2013 and March 2014 and correspond to periods of around 2 days where the PWS100 outputted suddenly only 0 values and a "Required Maintenance" status value (Haukeliseter being a remote site, no immediate maintenance would have been possible), then came back to normal again once the temperature rose up. Examples of such behavior are given in Figure 15. From this analysis, it is more likely that the unit at Haukeliseter seems to have encountered a specific issue that has been recognized by the sensor itself, which outputted a warning status for these periods. A possible explanation could be that, at these temperatures and during precipitations, the sensor lenses were (partially) covered by icing, wet sticky snow or condensation, which prevented the sensor to measure properly snowfall, giving the warning status to come out about maintenance requirement (to clean the lenses) and as soon as the temperature rose, the lenses got free by natural evaporation and the sensor could start to work properly again. However, this assumption cannot be confirmed and it occurred only at one site and for one season, such that we can't conclude that it could be a general sensor issue. Such 'blackouts' should need further investigations from the manufacturer, but it is worthwhile to know that if it happens, the sensor status output gives the appropriate warning information and hence the possibility to filter out impacted data or, if possible, go directly on site to clean the lenses.

The amount of YN and NY cases at Marshall are marginally low, and they cannot be related to any specific atmospheric conditions differing from normal conditions for that site (see Figure 13 and Figure 14). Following an information from the manufacturer (document sent after meeting discussion in Brussels, 2015) noting that their sensor had an issue (fixed in newer firmware releases since) about false hail reports, it has been verified that no hail was reported by the PWS100 during these 17 NY events at Marshall.

Overall, these results indicate that the SUT is a reliable instrument to detect precipitation events, as already mentioned in Section 7.2.

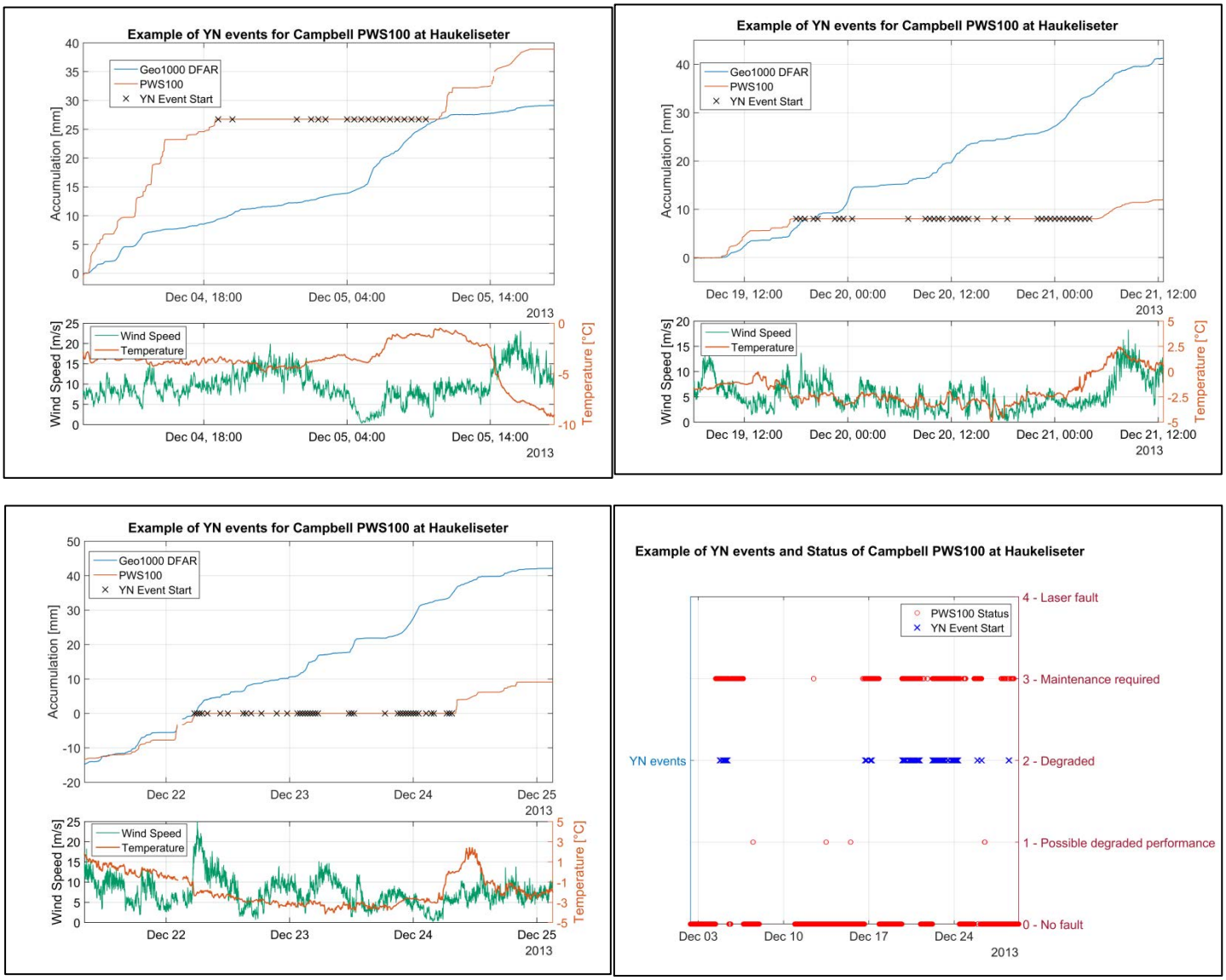


Figure 15: Three examples of YN events for the SUT PWS100 tested at Haukeliseter during the first season, compared to the site reference (Geo1000 DFAR), together with weather conditions illustrated by wind speed and temperature in the bottom subplot. Bottom-right panel illustrates YN events from December 2013, together with sensor status output.

7.6. Threshold selection

According to the results presented above, especially high POD, low FAR and zero noise signal during no-precipitation events, and following the methodology defined in Section 3.6.1.3.2 of the SPICE Final Report, the PWS100 doesn't need any minimum threshold over a 30 min interval to be able to report the occurrence of precipitation adequately.

8. Operational considerations

The overall experience with the Campbell PWS100 at Marshall was positive. At Haukeliseter, however, an up to now unexplained behavior of the sensor led to intermittent periods of missing data (see Section 8.2).

It has been recognize as an advantage that the PWS100 has its own temperature and humidity sensor, in a standard radiation screen, such that it can be used as a backup for air temperature measurement. Moreover, site managers acknowledged the utility of the battery provided to secure the data collection during short term power outages.

On the other hand, it has been noted that it requires quite a few pieces to install (Sensor, Temperature Radiation Screen, Control unit, Power unit) and that the power unit operates with 220V and needs consequently to be extra secured for non-authorized access. Concerning the output of the sensor, it has been noted that no total accumulation calculation is provided, such that the cumulative sum must be calculated from the discrete accumulation reports.

8.1. Maintenance

Some parts have to be replaced regularly. At Haukeliseter, for instance, after two seasons, the provider asked to change the temperature/humidity sensor, as part of a regular maintenance. Furthermore, desiccant in control unit must be changed often (every few months). In the case of remote site (like Haukeliseter), the control unit being installed at 6 m above the ground, this was experienced as a severe limitation of operation.

8.2. Noted issue

At Haukeliseter, the PWS100 experienced operational issue during the second season of the experiment. The sensor worked fine during two seasons (installation on January 2013), then it started to produce garbled telegrams for periods of time (of different length), which could not be processed by the logger (the instrument itself seemed to work fine). Due to this data-telegram problems, resulting in intermittent operation of the SUT during weeks (Haukeliseter is a remote site, with low accessibility and no personal on site), the site was advised by the manufacturer to apply two extra resistances and to extra ground the signal cables, but it did not seem to have any effect. In the meantime, however, the sensor got back to a more stable state by himself, with reasonable data again, which can also be due to the warmer and dryer weather in that period.

9. Performance considerations

- The PWS100 may have periods of missed precipitation events in unfavorable conditions (as experienced with the YN events from season 1 at Haukeliseter) with the presumed cause being lenses potentially contaminated by icing, sticky wet snow or condensation. Its ‘Fault Code’ status output is acknowledged to be a good indicator at informing that maintenance is required during these periods and is therefore reliable to detect such cases and get rid of the wrong associated data.
- Overall, the PWS100 catch efficiency consistently decreases with increasing wind speed up to 8 m/s under snow and mixed precipitations. Wind speeds above 8 m/s seem to impact severely the performance of the PWS100 in deriving properly snow and mixed accumulation, with a CE that has a high variability with values that could greatly exceeding 3. This result needs therefore to be taken into consideration when PWS100 is operated in windy sites.
- High scatter of CE on an event basis (30 min), but good mean CE over the seasons, makes the PWS100 more suitable for deriving solid precipitation measurements over long periods of time, giving the condition that the sensor operates continuously.

WMO-SPICE Instrument Performance Report
Yankee Environmental Systems Hotplate

1. Technical specifications

Instrument model:	YES Total Precipitation Sensor-3100 Hotplate
Measuring area:	132.7 cm^2 (13 cm diameter)
Physical principle:	Thermodynamic precipitation sensor head and electronics enabling the measurement of one minute average liquid and frozen precipitation rate by measuring the evaporative cooling of the plate at a constant temperature.
Operating temperature range:	-50 to +50°C
Measurement uncertainty:	± 0.25 mm/h (liquid equivalent rate accuracy)
Sensitivity:	0.1 mm/h (resolution)

Note: Specifications from manufacturer provided documentation.

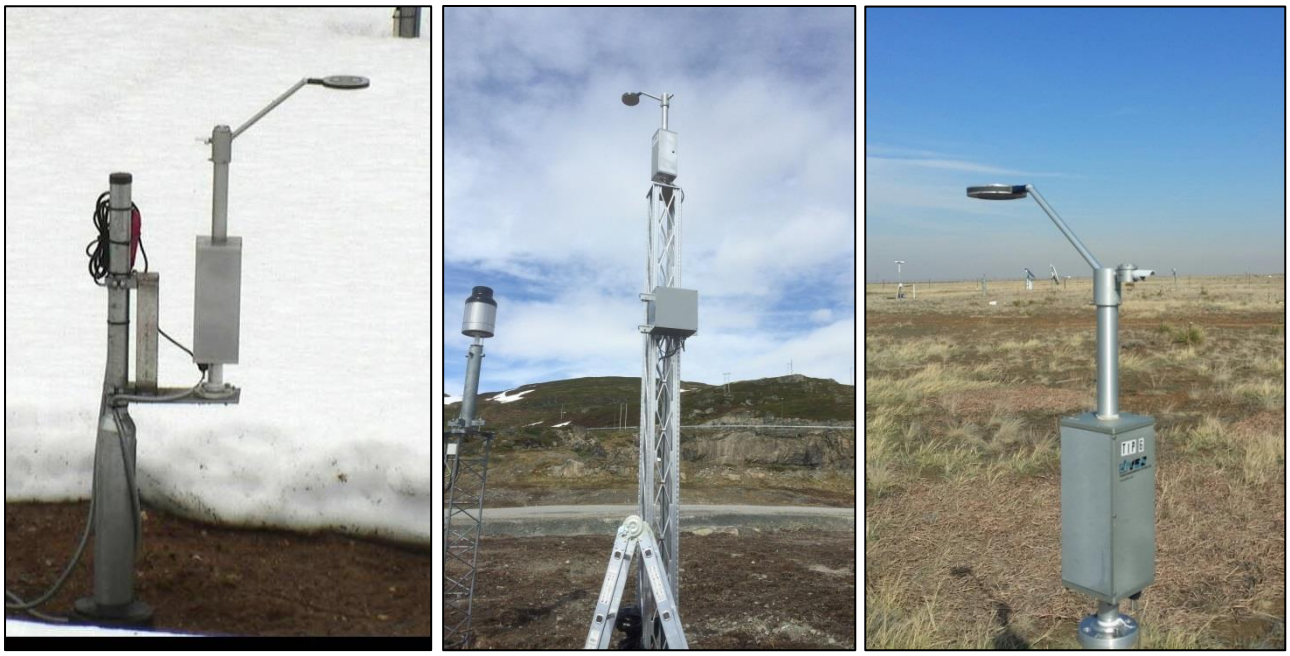


Figure 1: YES Hotplate at Sodankylä (left), at Haukeliseter (middle) and at Marshall (right) test sites.

2. Data output format

The Hotplate total precipitation sensor TPS-3100 is a sensor outputting both precipitation information, as intensity every minute or total accumulation, as well as ancillary measurements, such as air temperature, wind speed, relative humidity or solar radiation. Table 1 summarizes the main output parameters from the instrument. The firmware versions of the Hotplate tested during this experiment were V3.1.1 and V3.1.2 depending on the test site (see Table 2).

Table 1: Summary of main instrument outputs, as recorded by the sites during the experiment.

Measured Parameters	Units
Current Precipitation Rate	[mm/h]
1 min Raw Precipitation Rate average	[mm/h]
5 min Raw Precipitation Rate average	[mm/h]
Total Accumulated Liquid Precipitation	[mm]
Ambient Temperature	[°C]
Wind Speed	[m/s]
Solar Radiation	[W/m^2]
Net IR radiation ground to sky	[W/m^2]
Barometric Pressure	[mbar]
Relative Humidity	[%]
Housekeeping	-

This document reports on the ability of the Hotplate to measure solid precipitation. The results are consequently computed using the 'Total Accumulated Liquid Precipitation' output.

3. SPICE test configuration

The Hotplate, as sensor under test (SUT), has been tested on three different sites:

Test Sites: Haukeliseter (Norway); Sodankylä (Finland); Marshall (USA)

Sensor Provider(s): Haukeliseter and Sodankylä Hotplates were provided by the manufacturer (Yankee Environmental System, Inc.) and Marshall Hotplate was provided by the site host.

Figure 2: Map of SPICE sites testing Hotplate instruments.

A summary on the configuration of instruments as tested, the duration of tests and availability reflected in these results, and the ancillary measurements used, by site, is available in Table 2, Table 3, and Table 4, respectively.

Table 2: Summary of instrument configurations and data output, by site. Details and photos on individual site configurations are available in the respective site commissioning protocols.

	Haukeliseter	Sodankylä	Marshall
Main prevailing wind directions	NE and E	South	btw N and E (during pcp)
Sensor orientation	North	North	North
Height of installation	4.5 m	1.5 m	2 m
Firmware version	V3.1.1	V3.1.2	V3.1.2
Heating	Yes, internal, as recommended		
Shield	No	No	No
Data QC	SPICE QC methodology		
Data temporal resolution	1 min		
Processing interval for SPICE data analysis	30 min		

Table 3: Data availability, by measurement season, by site.

Measurement season	Haukeliseter	Sodankylä	Marshall
Season 1 (Oct. 2013 – Apr. 2014)	X	✓	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓	✓	✓

Data from Haukeliseter during the 2013-2014 season are not available for the analysis due to the late mid-season delivery of the sensor and some difficulties at the first installation attempt.

Table 4: Summary of reference and ancillary measurements, by site, with measurement height.

	Haukeliseter	Sodankylä	Marshall
R2 Site Reference	Geonor 1000 (DFAR) (4.5 m, rim height)	OTT Pluvio^2 1500mm (DFAR) (4 m, rim height)	Geonor 600 (DFAR) (3 m, rim height)
R2 Precip Detector	Thies LPM X5 (DFAR, 2014/15) (4.5 m)	DRD11A (Site*, 2013/14) (1 m) OTT Parsivel^2 (DFAR, 2014/15) (2.7 m)	Thies LPM (Site*) (3 m)
Ancillary Temp Sensor	PT100 (4.5 m)	Vaisala HMP155 (2 m)	MetOne, model 060A-2/062, 2144-L (2 m)
Ancillary RH Sensor	PWS100 (6 m)	Vaisala HMP155 (2 m)	Campbell Scientific CS500 (3 m)
Ancillary Wind Sensor	RM Young Wind Monitor 05103 (4.5 m)	Thies acoustic 2D wind sensor (3.5 m)	RM Young Wind Monitor 05103 (3 m)

**A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR-fence; however, in cases where a more sensitive detector is available outside of the DFIR-fence, or there are issues with the detector within the DFIR-fence, a precipitation detector elsewhere on the site can be employed.*

4. Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6 of the SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute intervals:

- Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip;
- SUT 'Yes': SUT accumulation > 0 mm;
- SUT 'No': SUT accumulation = 0 mm.

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores.

4.1.2. Skill score assessment

The ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

The above scores are computed using the formulations provided in Section 3.6 of the SPICE Final Report.

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

5. Environmental conditions

The environmental conditions over the entire duration of the test period, at each site, are expressed as the probability density functions (PDFs) for air temperature, relative humidity, wind speed, wind direction, and precipitation rate in Figure 3. Figure 4 presents the same parameters during the precipitation events reported by the corresponding site reference, R2. The precipitation percentage in Figure 3 and Figure 4 represents the number of minutes of precipitation over a standard 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration at each site.

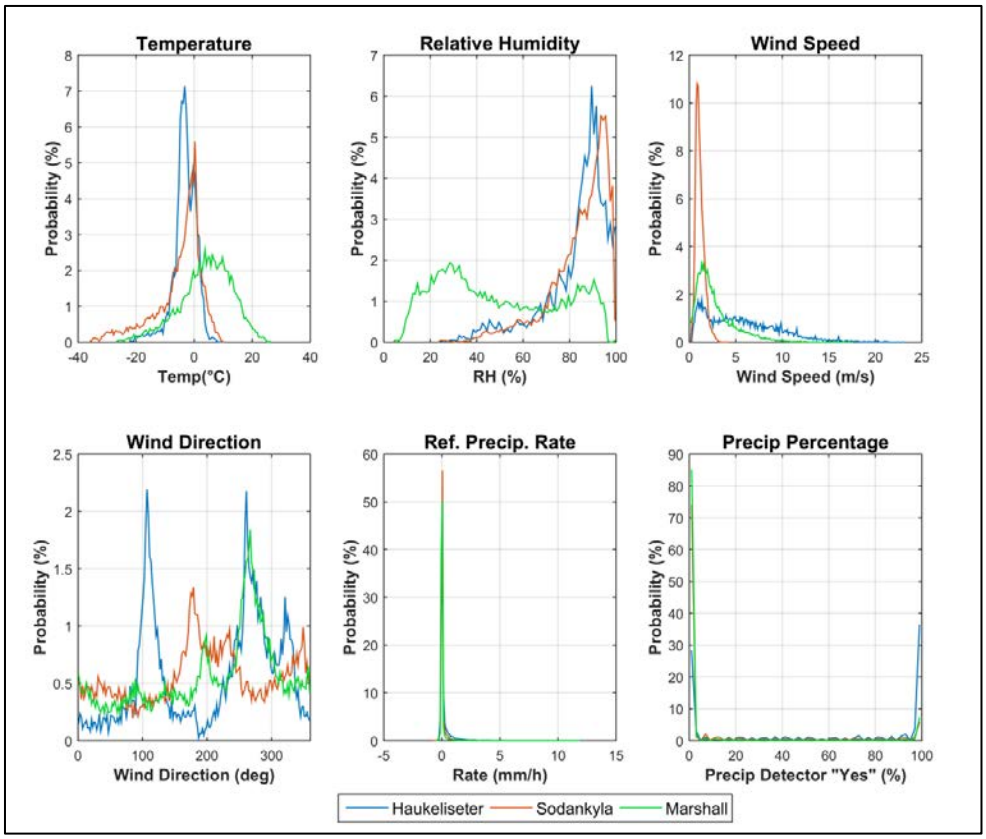


Figure 3: Summary of aggregated environmental conditions on the SPICE sites that operated Hotplate, over the entire duration of tests, as per Table 3, above.

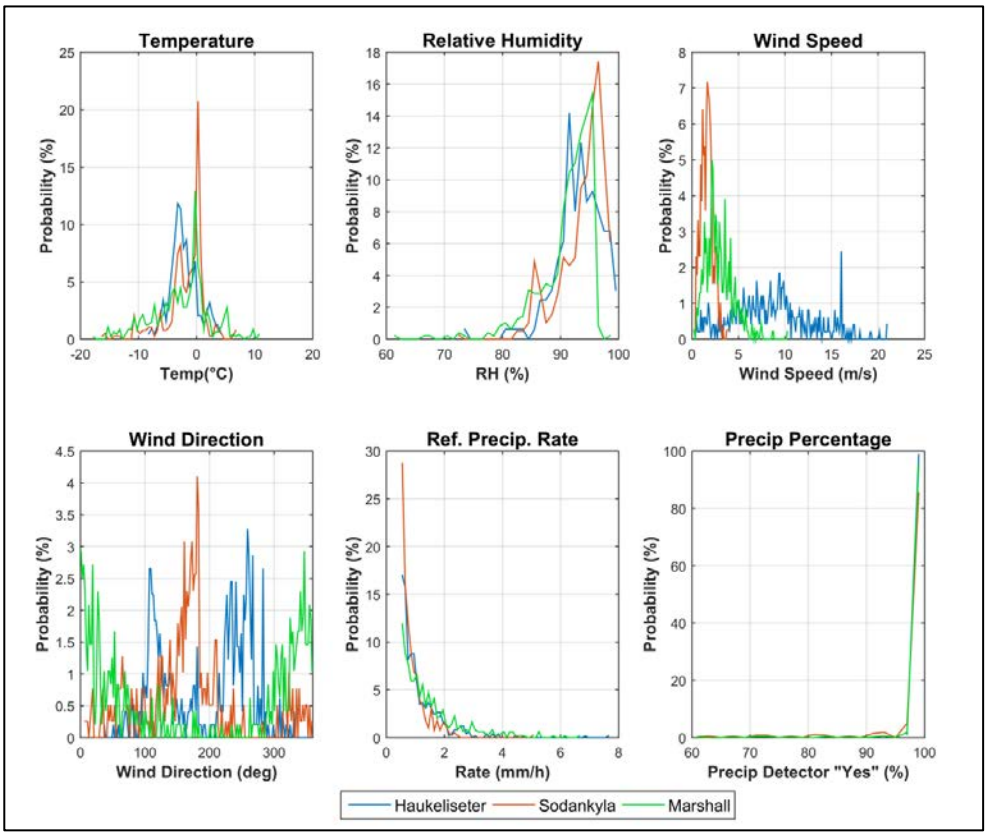


Figure 4: Summary of the aggregated environmental conditions on the SPICE sites that tested Hotplate, corresponding to precipitation events, as reported by the site R2, reference, during the tests, as per Table 3 above.

6. Evaluation of performance over the range of operating conditions

6.1. Skill score assessment

The ability of the SUT to represent precipitation similarly to the site field reference R2, is assessed using contingency tables (Section 4.1.1) and derived selected skill scores (Section 4.1.2). To better understand the potential influence of threshold choices on the derived results, two cases are considered here (see note below). The contingency results related to these two cases are given in Table 5 and the respective skill scores in Table 6, for both testing sites.

Note: Following the data derivation explained in Section 4.1.1, the conditions required to have a 'Yes' or a 'No' event over the 30 min interval, for the reference and the SUT, for the two different cases treated here, are:

CASE 1 *(as defined in Section 4.1.1)*:

- *Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip*
- *Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip*
- *SUT 'Yes': SUT accumulation > 0 mm*
- *SUT 'No': SUT accumulation = 0 mm*

CASE 2:

- *Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip*
- *Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip*
- *SUT 'Yes': SUT accumulation ≥ 0.1 mm*
- *SUT 'No': SUT accumulation < 0.1 mm*

Results presented in this report are based on Case 1.

Table 5: Contingency Tables: detection of precipitation of the Hotplate relative to the specific site reference, expressed as number of events over the entire test period, by site. The skill scores associated with these events are given in Table 6.

Case 1

Haukeliseter Ref Geo1000DFIR			
SUT Hotplate	Yes	No	Total
Yes	368	3	371
No	121	266	387
Total	489	269	758

Sodankylä Ref Plv2DFIR			
SUT Hotplate	Yes	No	Total
Yes	390	1894	2284
No	0	8357	8357
Total	390	10251	10641

Marshall Ref Geo600DFIR			
SUT Hotplate	Yes	No	Total
Yes	486	26	512
No	2	14701	14703
Total	488	14727	15215

Case 2

Haukeliseter Ref Geo1000DFIR			
SUT Hotplate	Yes	No	Total
Yes	337	0	337
No	152	269	421
Total	489	269	758

Sodankylä Ref Plv2DFIR			
SUT Hotplate	Yes	No	Total
Yes	389	846	1235
No	1	9405	9406
Total	390	10251	10641

Marshall Ref Geo600DFIR			
SUT Hotplate	Yes	No	Total
Yes	484	5	489
No	4	14722	14726
Total	488	14727	15215

Table 6: Skill Scores for the Hotplate, by site. POD: Probability Of Detection, FAR: False Alarm Rate, B: Bias, HSS: Heidke Skill Score (see Section 4.1.2 for more details).

Hotplate, Skill Scores				
	POD	**FAR**	**B**	**HSS**
Case 1				
Haukeliseter	75.3%	0.809%	75.9%	67.5%
Sodankylä	100%	82.9%	586%	24.4%
Marshall	99.6%	5.08%	105%	97.1%
Case 2				
Haukeliseter	68.9%	0%	68.9%	61.1%
Sodankylä	99.7%	68.5%	317%	44.8%
Marshall	99.2%	1.02%	100%	99%

6.2. Assessment of SUT performance during non-precipitating events

The performance of the SUT in the absence of precipitation (when the reference precipitation detector recorded 30 minutes without precipitation) is represented in Figure 5 and Table 7, reflecting the distribution of the sensor response, as measured during the interval.

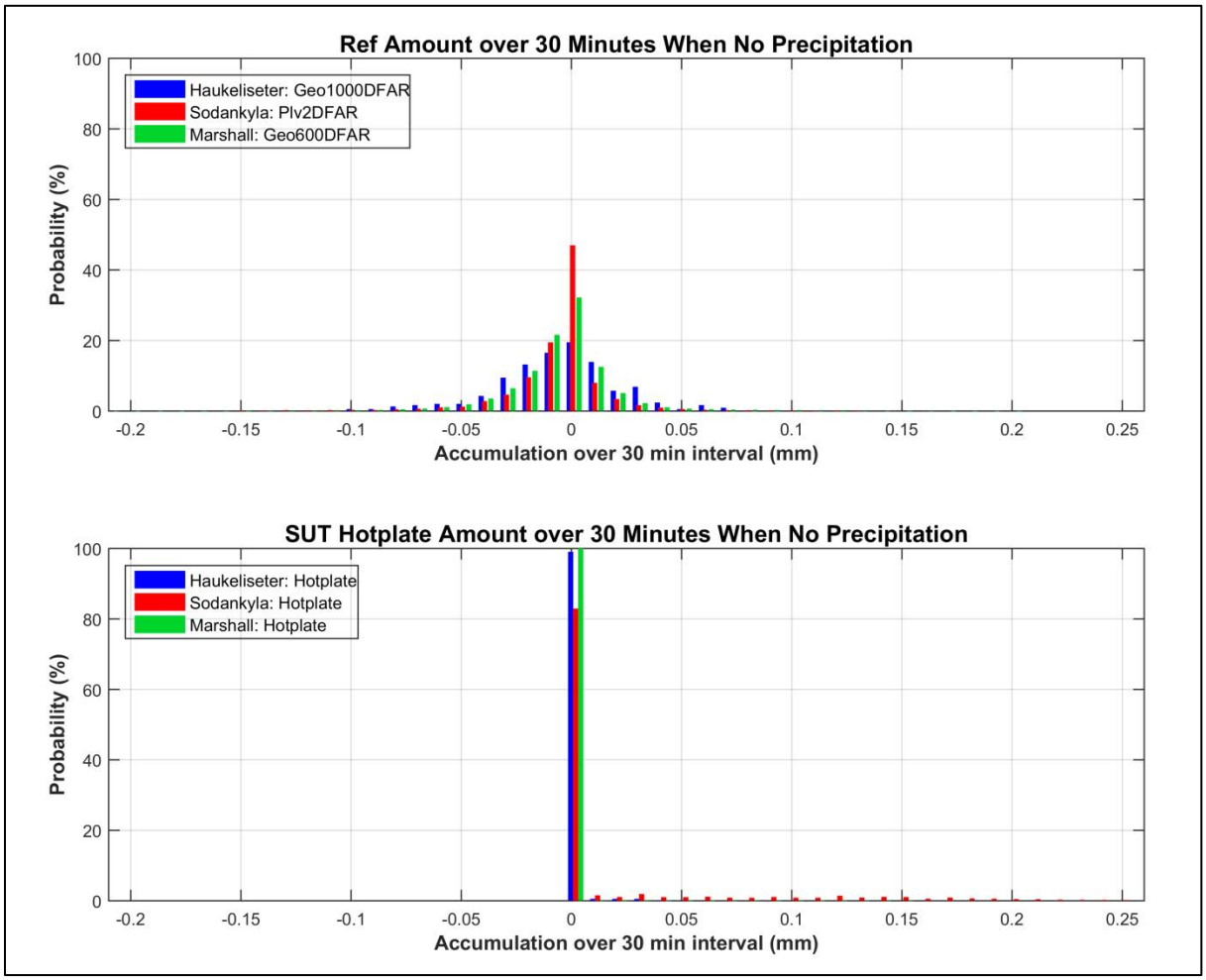


Figure 5: Probability of occurrence of a response during a 30 min interval in the absence of precipitation, represented by the signal output from (top) the R2 reference and (bottom) the SUT Hotplate, by site. Statistics associated with these graphs are given in Table 7.

Table 7: Reference and SUT statistics of response signal when no precipitation was occurring, as plotted in Figure 5; Average (Avg), standard deviation (STD), maximum (Max) and minimum (Min) of the response signal, together with the number of events (Num) over the test period is given by site.

No Precip Statistics - Reference:
Geo1000 DFAR (Haukeliseter), Plv2 DFAR (Sodankylä) and Geo600DFAR(Marshall)

	Ref Avg [mm]	Ref STD [mm]	Ref Max [mm]	Ref Min [mm]	Ref Num
Haukeliseter	-0.002	0.027	0.074	-0.091	269
Sodankylä	-0.003	0.020	0.120	-0.140	10254
Marshall	-0.001	0.022	0.196	-0.207	14776

No Precip Statistics - SUT: Hotplate

	SUT Avg [mm]	SUT STD [mm]	SUT Max [mm]	SUT Min [mm]	SUT Num
Haukeliseter	0.000	0.003	0.040	0.000	269
Sodankylä	0.017	0.045	0.260	0.000	10254
Marshall	0.000	0.003	0.142	0.000	14776

6.3. Ability of the SUT to measure precipitation

6.3.1. Yes-Yes cases

Quantitatively, the performance of the SUT to derive and report precipitation is assessed relative to the site reference in several graphs and tables illustrated in this section, using only the cases where both instruments reported precipitation over the 30 min interval, according to the criteria used in Case 1 of Table 5 (cases 'Yes-Yes', or shorter 'YY').

6.3.1.1. Time series plots

The time series (cumulative sum of 30 min YY events accumulation) of each individual SUT is plotted against their corresponding reference for the testing period, by precipitation types (see Section 4.1.4) and for each site in Figure 6.

The corresponding seasonal accumulations are given in Table 8.

Table 8: Seasonal accumulation [mm] for (a) Haukeliseter (b) Sodankylä and (c) Marshall test sites based on the sum of YY events from the SUT Hotplate and the corresponding site field references R2: (a) Ref Geo1000 DFAR, (b) Ref Plv2 DFAR, (c) Ref Geo600 DFAR.

(a) Haukeliseter

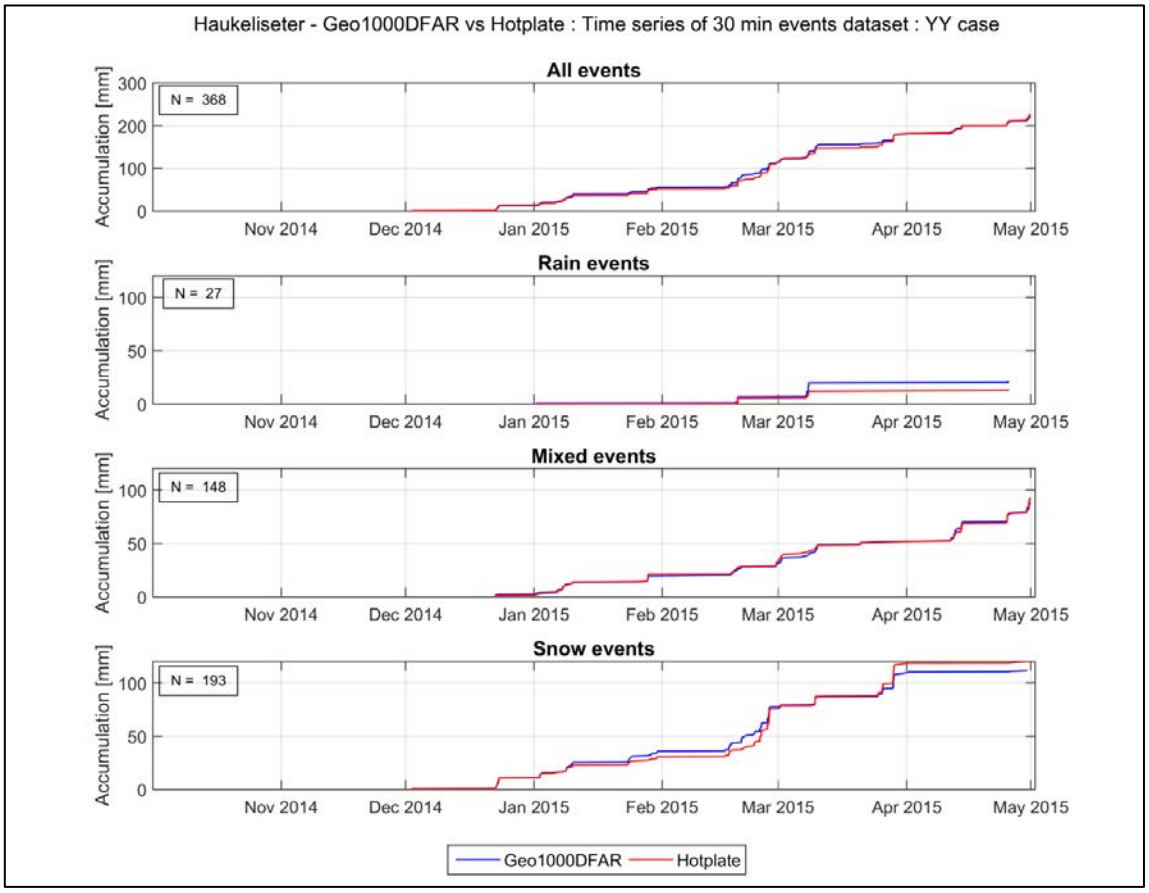
	Season 2014/15	
[mm]	Ref	SUT
All events	220.54	226.00
Rain events	20.59	13.37
Mixed events	88.38	92.34
Snow events	111.57	120.29

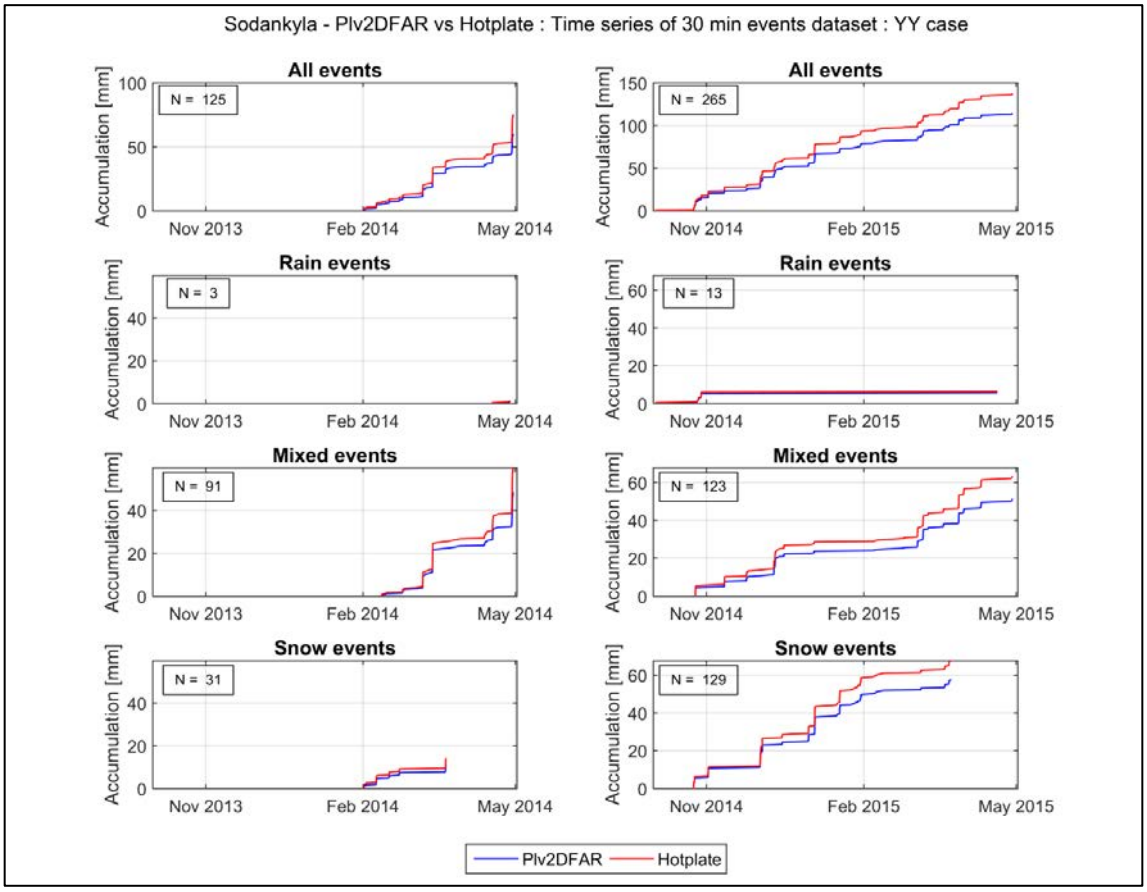
(b) Sodankylä

	Season 2013/14		Season 2014/15	
[mm]	Ref	SUT	Ref	SUT
All events	59.83	74.79	114.03	137.12
Rain events	0.85	1.10	5.60	6.32
Mixed events	47.99	59.83	51.04	63.13
Snow events	10.99	13.86	57.39	67.67

(c) Marshall

	Season 2013/14		Season 2014/15	
[mm]	Ref	SUT	Ref	SUT
All events	155.96	175.27	205.86	223.50
Rain events	14.64	15.99	25.64	29.15
Mixed events	100.10	116.97	101.56	110.74
Snow events	41.22	42.31	78.65	83.61


Haukeliseter - Geo1000DFAR vs Hotplate : Time series of 30 min events dataset : YY case
All events
N = 368
Rain events
N = 27
Mixed events
N = 148
Snow events
N = 193
Accumulation [mm]
Nov 2014
Dec 2014
Jan 2015
Feb 2015
Mar 2015
Apr 2015
May 2015
Geo1000DFAR
Hotplate


Sodankyla - Plv2DFAR vs Hotplate : Time series of 30 min events dataset : YY case
All events
N = 125
All events
N = 265
Rain events
N = 3
Rain events
N = 13
Mixed events
N = 91
Mixed events
N = 123
Snow events
N = 31
Snow events
N = 129
Accumulation [mm]
Nov 2013
Feb 2014
May 2014
Nov 2014
Feb 2015
May 2015
Plv2DFAR
Hotplate

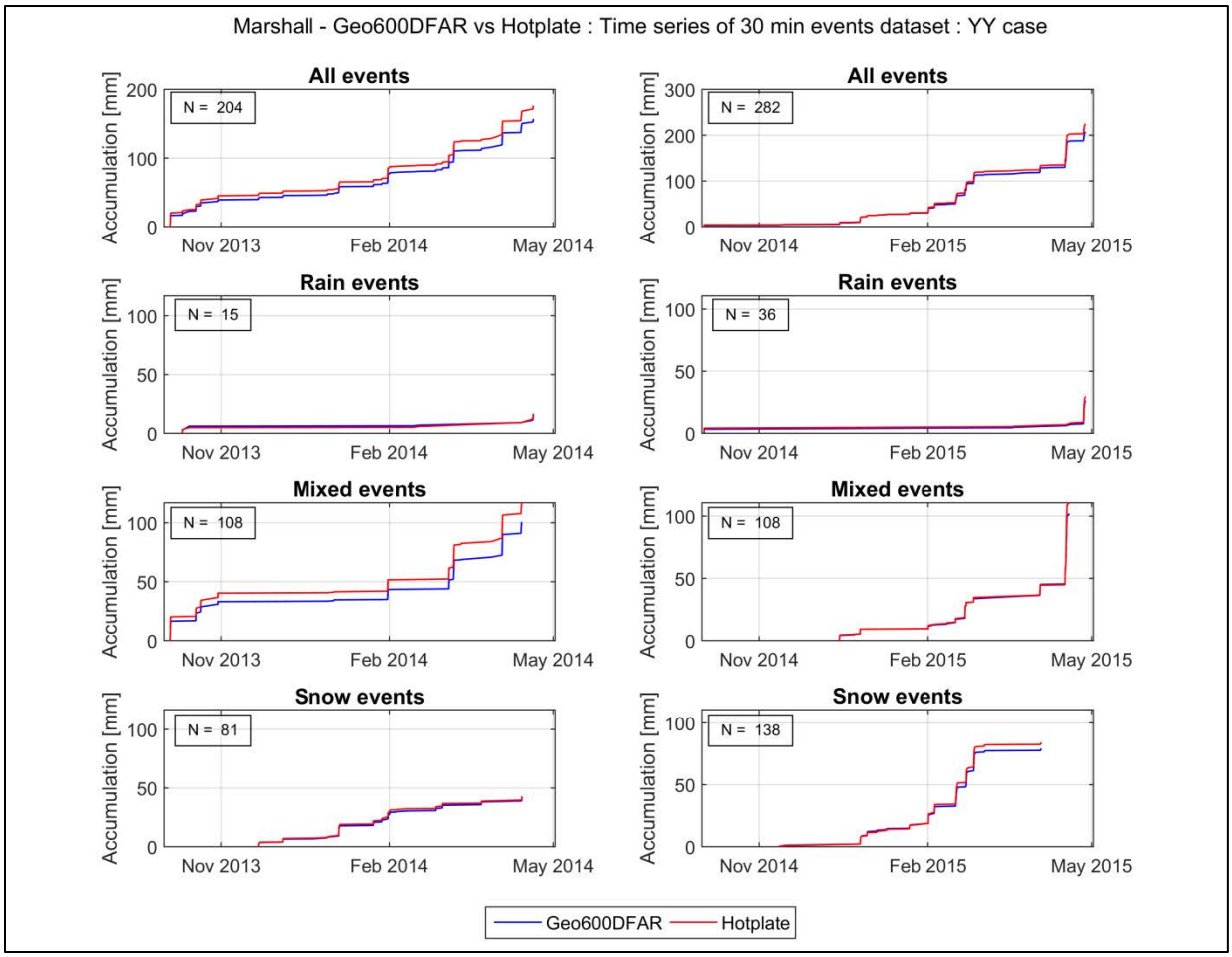


Figure 6: Time series based on 30 min YY events of the SUT Hotplate against the corresponding site reference, discriminated by precipitation type (Rain, Mixed, Snow), for (top previous page) season 2014/15 for Haukeliseter test site, and both seasons (2013/14 on the left, 2014/15 on the right) for Sodankylä (bottom previous page) and Marshall (above) test sites respectively.

6.3.1.2. Scatter plots and RMSE values

Scatter plots of the amount derived by the SUT versus the corresponding reference amount, and discriminated by precipitation type, is given for the three sites in Figure 7.

Quantitatively, the SUT performance is assessed using the Root Mean Square Error (RMSE) also known in practice as Operational Comparability. The results are available for all sites in Table 9.

Table 9: Root Mean Square Error (RMSE) statistics, in mm, for the SUT Hotplate with respect to the corresponding site reference, by precipitation type and by site, including 2014/15 season data for Haukeliseter and both seasons data for Sodankylä and Marshall.

RMSE [mm]	All	Rain	Mixed	Snow
Haukeliseter	0.333	0.409	0.360	0.306
Sodankylä	0.129	0.094	0.142	0.114
Marshall	0.232	0.344	0.283	0.121

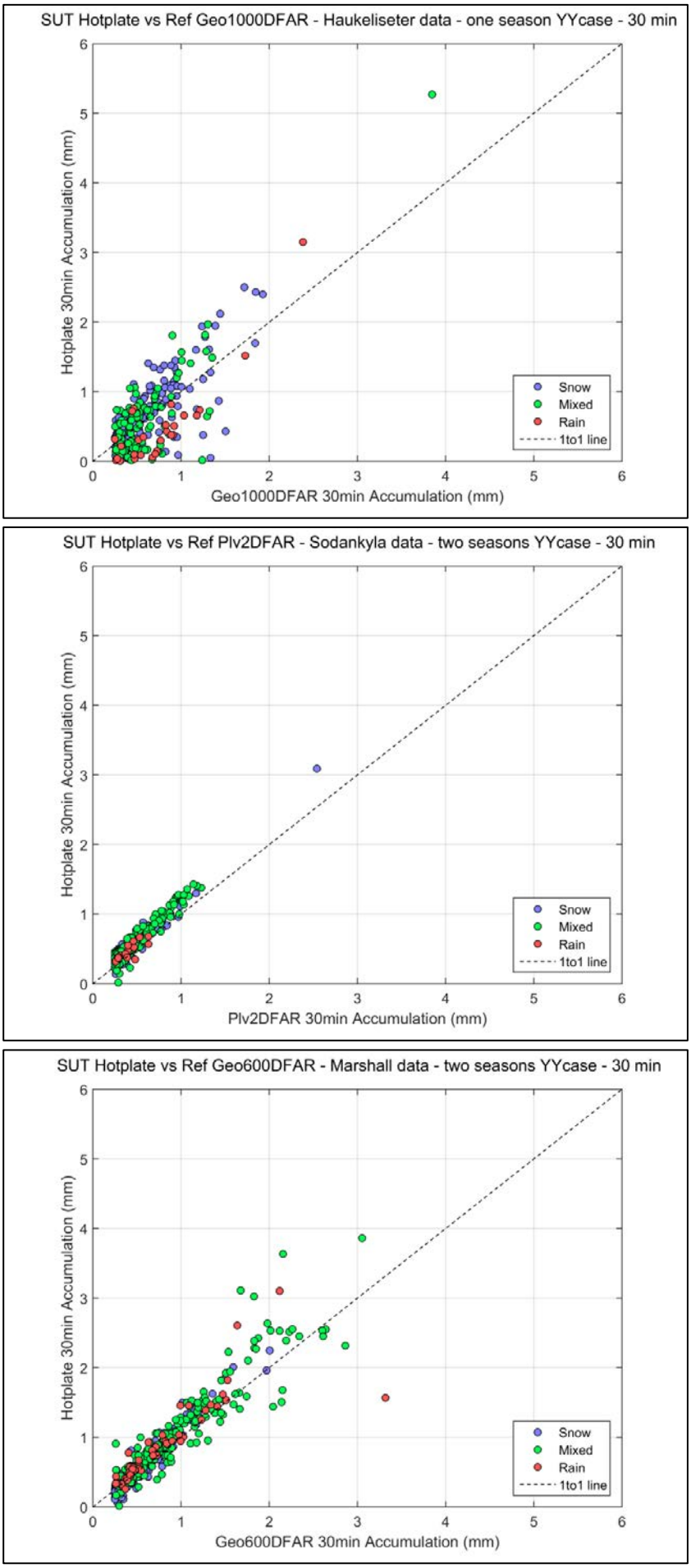


Figure 7: Scatter plots based on 30 min YY events accumulation from the SUT Hotplate against the corresponding site reference, over the test period, discriminated by precipitation type, for (top) Haukeliseter, (middle) Sodankylä and (bottom) Marshall test sites.

6.3.1.3. Catch efficiency evaluation by precipitation type

The Catch Efficiency (CE) of the SUT is represented by histograms (Figure 8) and boxplots (Figure 9), both discriminated by precipitation type, and representing the three sites.

The quantitative evaluation of the CE is provided in Table 10. The mean catch efficiency is given in the first line of this table for each site, considering test period data and for each category of precipitation type as well as for all the events together.

Note: All the events with a CE greater than 3, if any, are included in one category named '3 and more' or '≥3' in the upcoming graphs. Additionally, for all graphs representing the CE, a dashed black line is added at CE = 1, which represents the ideal case where the SUT reports exactly the same precipitation amount as the reference.

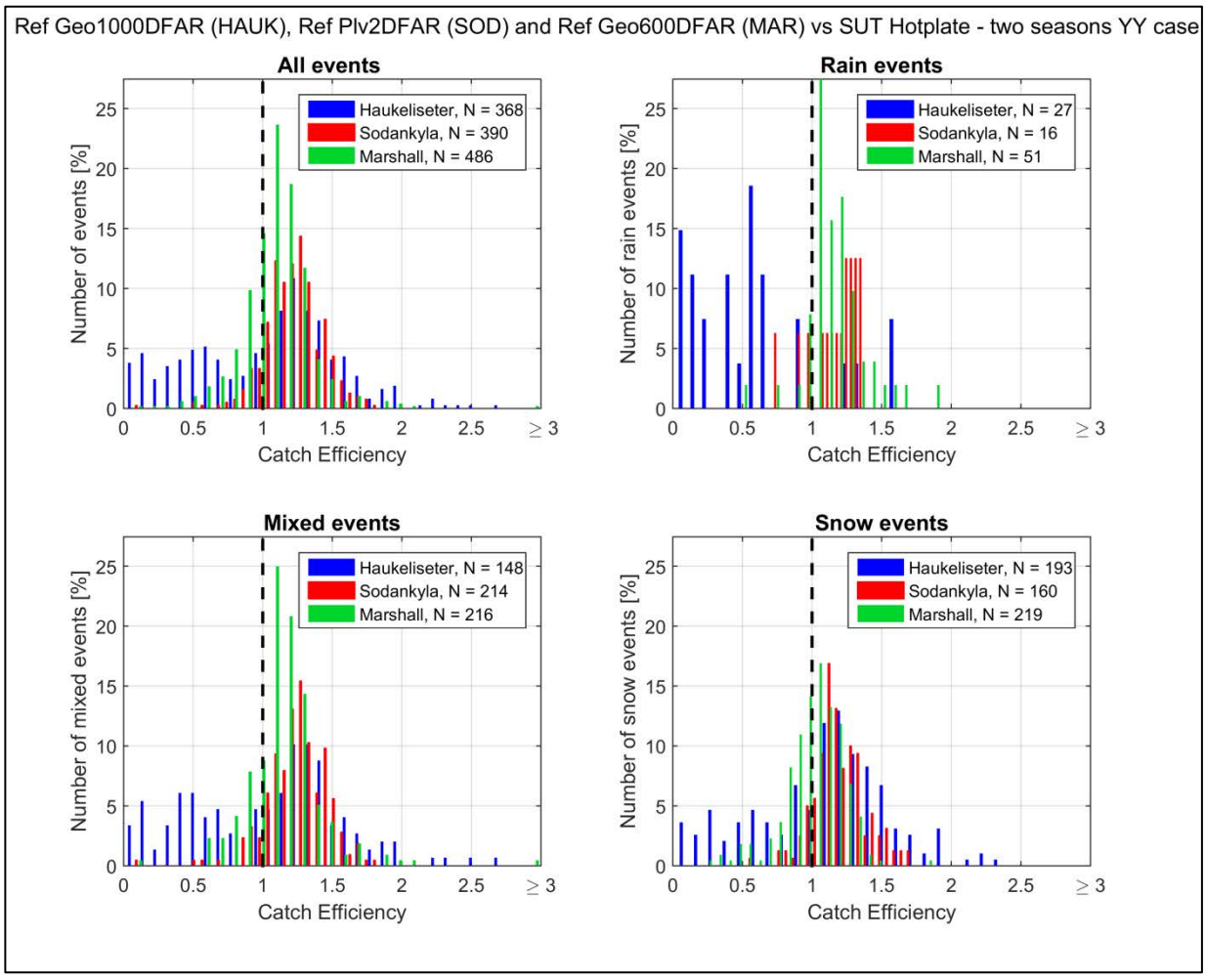


Figure 8: Histograms based on 30 min YY events for the test period, representing the distribution of the catch efficiency of the SUT Hotplate against the corresponding site reference (SUT/Ref), discriminated by precipitation type, with number of events given in the legend for each category, and represented for the three sites. The dashed black line at CE = 1 represents the ideal case.

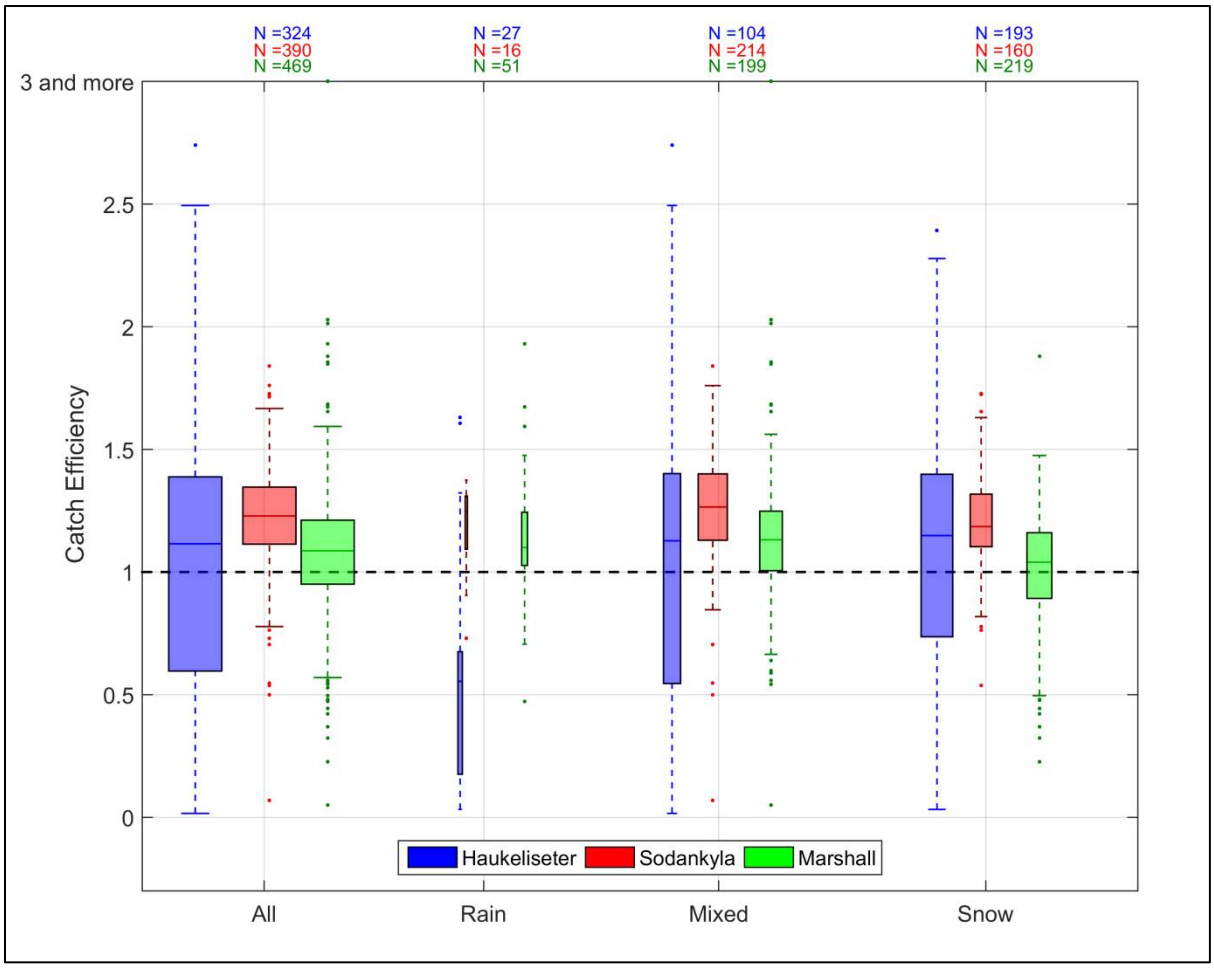


Figure 9: Boxplot based on 30 min YY events from the test period, representing the catch efficiency of the SUT Hotplate against the corresponding site reference (SUT/Ref), discriminated by precipitation type, with number of events given at the top of each category, and represented for the three sites. The width of the boxes is proportional to the percentage of events in each category ('All' boxes represent 100% data for each site, making them having the same width despite the difference in number of events). The dashed black line at CE = 1 represents the ideal case.

Table 10: Statistics related to the CE of SUT Hotplate against the corresponding site reference (SUT/Ref, see Figure 9), discriminated by precipitation type.

	Haukeliseter				Sodankylä				Marshall			
CE Boxplot Parameters	All	Rain	Mixed	Snow	All	Rain	Mixed	Snow	All	Rain	Mixed	Snow
Mean	1.03	0.57	1.05	1.08	1.23	1.18	1.25	1.21	1.08	1.16	1.14	1.02
Median	1.12	0.55	1.13	1.15	1.23	1.25	1.27	1.19	1.09	1.1	1.13	1.04
75 percentile	1.39	0.68	1.4	1.4	1.35	1.31	1.4	1.32	1.21	1.24	1.25	1.16
25 percentile	0.6	0.18	0.55	0.74	1.11	1.09	1.13	1.1	0.95	1.03	1.01	0.89
Upper Whisker	2.49	1.32	2.49	2.28	1.67	1.37	1.76	1.63	1.59	1.48	1.56	1.48
Lower Whisker	0.02	0.03	0.02	0.03	0.78	0.9	0.85	0.82	0.57	0.71	0.66	0.5
Maximum	2.74	1.63	2.74	2.39	1.84	1.37	1.84	1.73	3.48	1.93	3.48	1.88
Minimum	0.02	0.03	0.02	0.03	0.07	0.73	0.07	0.54	0.05	0.47	0.05	0.23
# Outliers	1	2	1	1	12	1	5	6	26	4	14	8
# Outliers ≥ 3	0	0	0	0	0	0	0	0	1	0	1	0
# Events	324	27	104	193	390	16	214	160	469	51	199	219

6.3.1.4. Catch efficiency dependency on wind speed

The variation of the SUT catch efficiency (SUT/Ref) with wind speed is illustrated in a scatter plot in Figure 10, together with a boxplot in Figure 11, both discriminated by precipitation type and for each site.

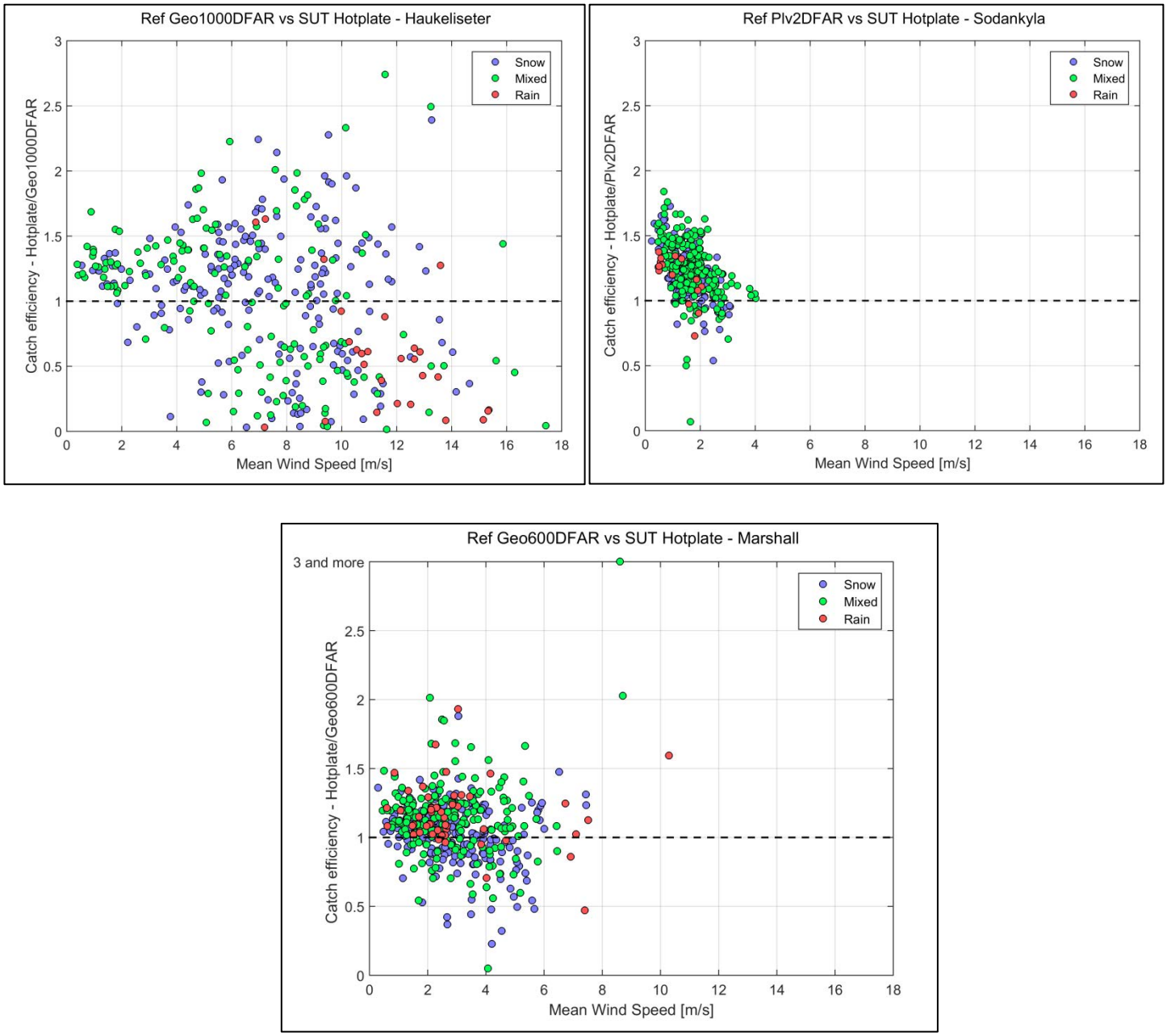


Figure 10: Scatter plots based on 30 min YY events for the test period, representing the catch efficiency of the SUT Hotplate with respect to the corresponding site reference (SUT/Ref), against wind speed and discriminated by precipitation type, for (top left) Haukeliseter, (top right) Sodankylä and (bottom) Marshall test sites. The dashed black line at CE = 1 represents the ideal case.

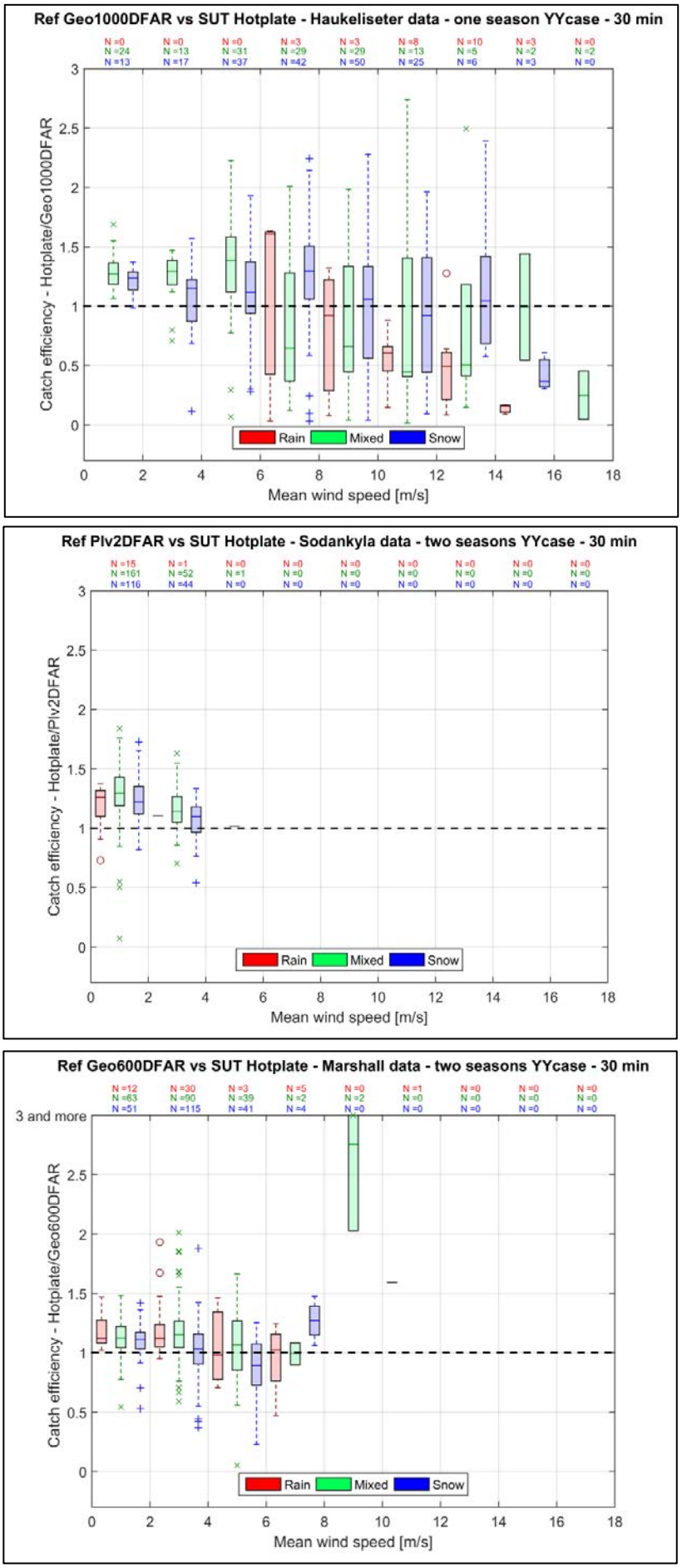


Figure 11: Boxplots based on 30 min YY events for the test period, representing the catch efficiency of SUT Hotplate with respect to the corresponding site reference (SUT/Ref), against wind speed and discriminated by precipitation type for (top) Haukeliseter, (middle) Sodankylä and (bottom) Marshall test sites. The dashed black line at CE = 1 represents the ideal case.

6.3.1.5. Catch efficiency dependency on wind direction

In order to assess the dependency of the CE with wind direction, wind roses are produced (Figure 12) representing, for each site, the wind data of the testing period, binned by catch efficiency in order to represent undercatch (CE < 0.8), overcatch (CE > 1.2) and catch efficiency of 1 ± 20% of the SUT.

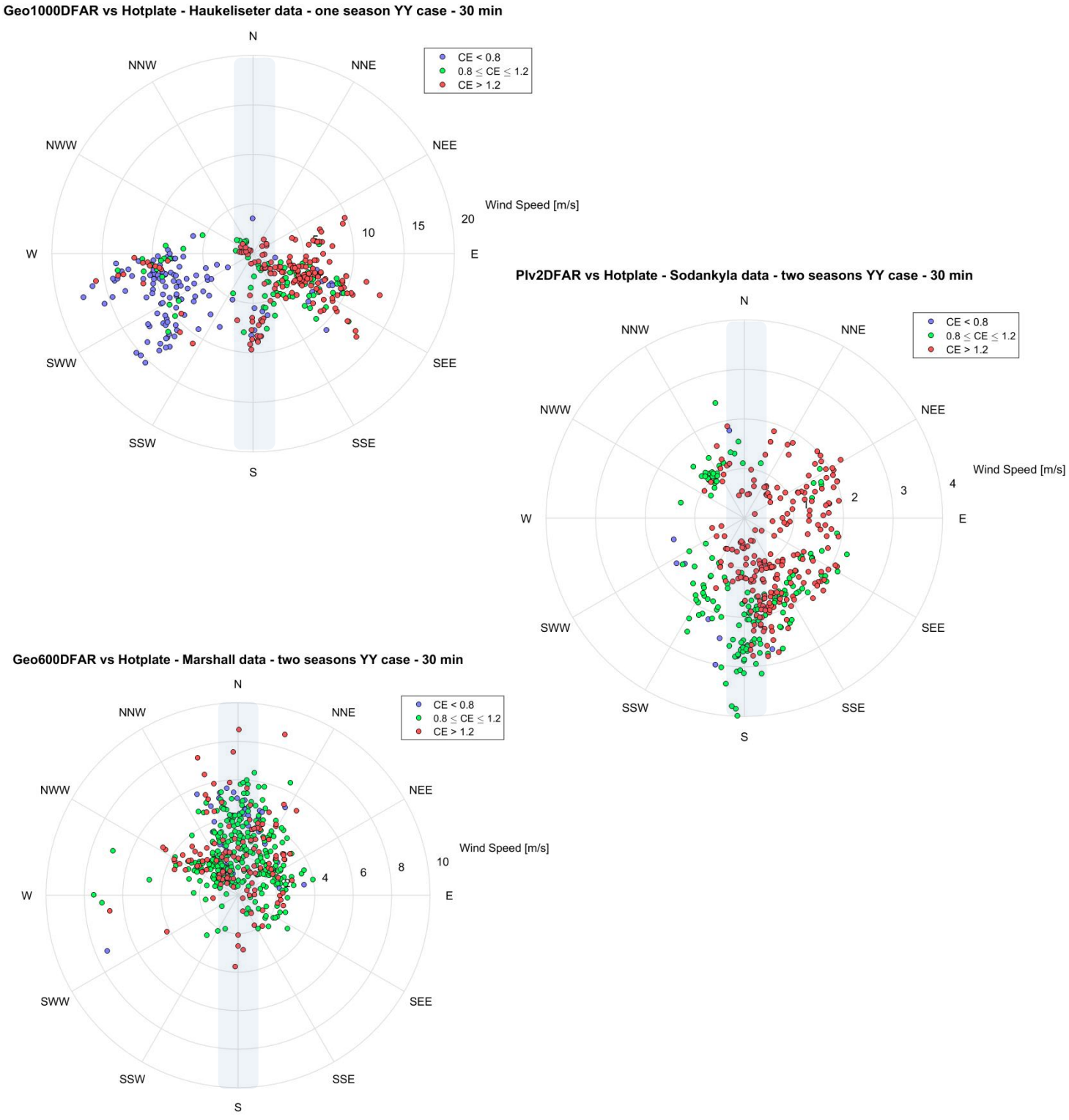


Figure 12: Precipitation events (YY cases) as function of wind speed and direction of (top) Haukeliseter, (middle) Sodankylä and (bottom) Marshall test sites, binned by catch efficiency (CE). The grey zone indicates the SUT orientation.

6.3.2. Yes-No and No-Yes cases

Events when the site reference and SUT do not agree on the occurrence of precipitation includes two categories of cases (Section 4.1.1): (1) when the field reference reported a precipitation event, while the SUT did not (Yes-No cases, 'YN'), and (2) when the field reference did not report a precipitation event while the SUT did (No-Yes cases, 'NY').

Histograms illustrating field reference and SUT reports and associated site conditions for all YN and NY cases of the three sites during the test period are provided in Figure 13 and Figure 14, respectively.

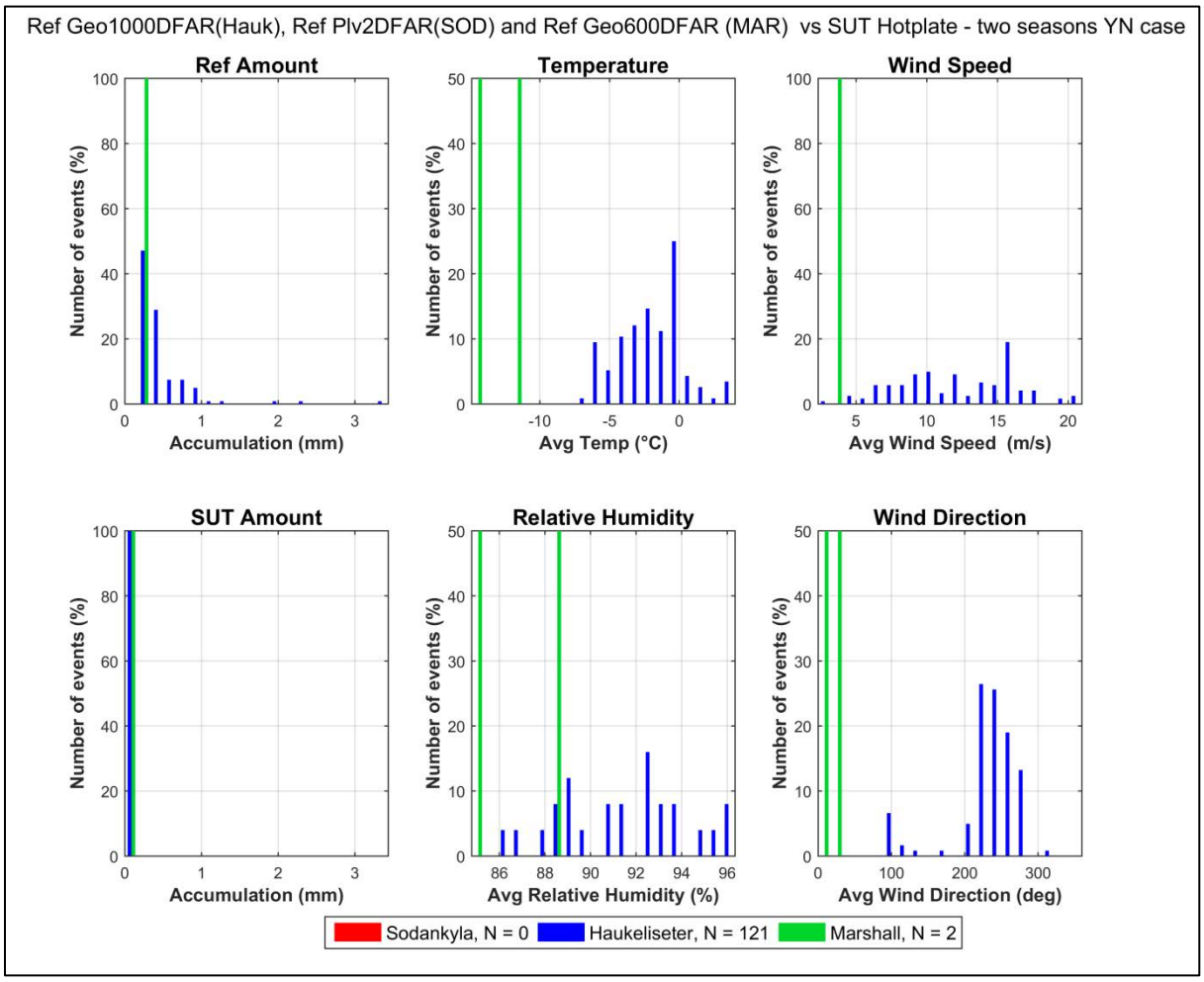


Figure 13: Histograms of SUT Hotplate and corresponding site reference accumulations (left column), along with distributions of mean temperature, mean relative humidity, mean wind speed, and mean wind direction for all YN cases (number indicated in the legend) of the 30 min intervals over the test period, as reported by Case 1 of Table 5, at the three different test sites.

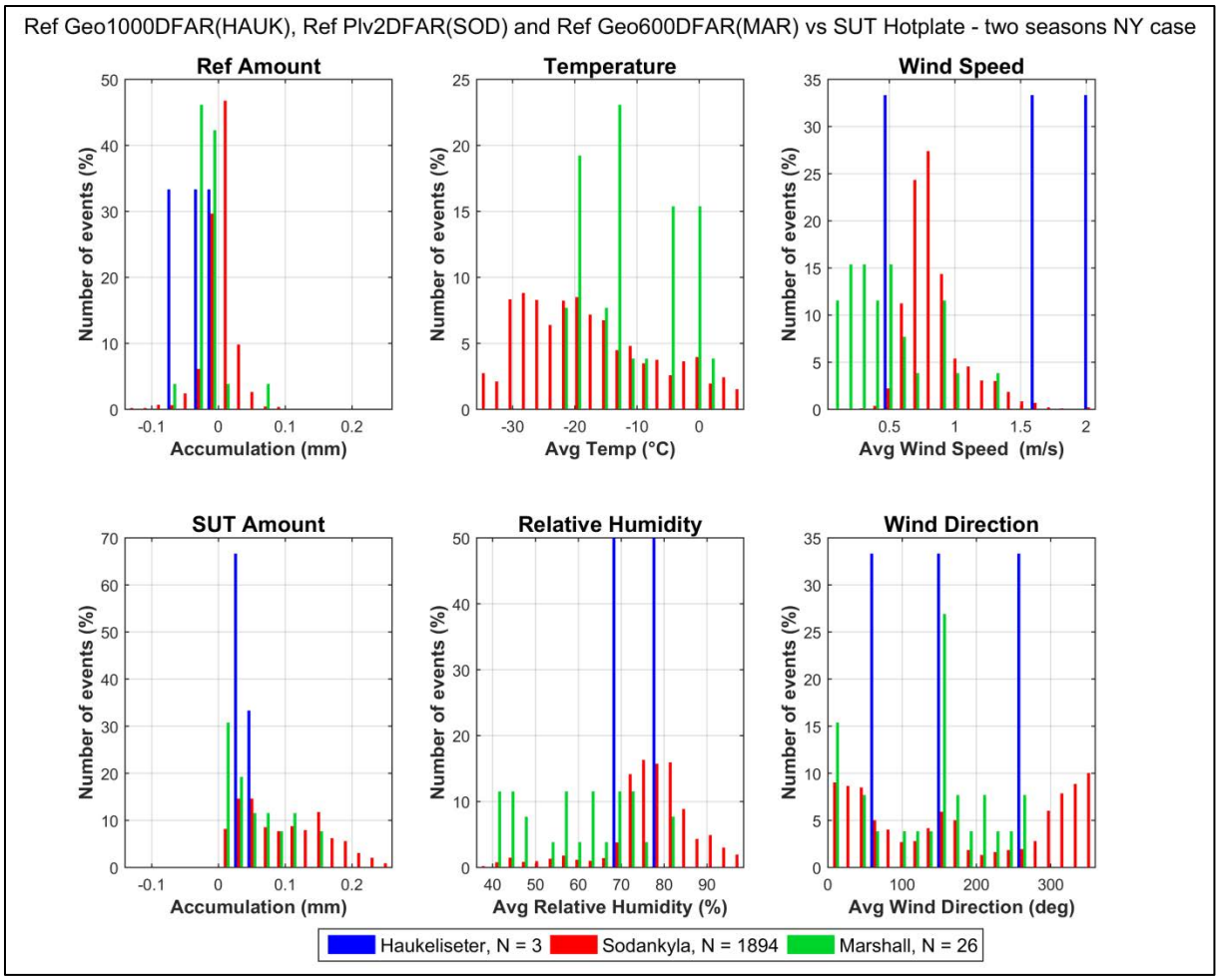


Figure 14: Histograms of SUT Hotplate and corresponding site reference accumulations (left column), along with distributions of mean temperature, mean relative humidity, mean wind speed, and mean wind direction for all NY cases (number indicated in the legend) of the 30 min intervals over the test period, as reported by Case 1 of Table 5, at the three different test sites.

7. Interpretation of results

7.1. Operating conditions

The main difference between the three sites hosting the Hotplate is the wind speed distribution, as shown in Figure 3 and Figure 4. Wind speed is higher at Haukeliseter, with maximum ranging up to 15-20 m/s, whereas the maximum wind speed for Sodankylä doesn't exceed 4 m/s. Marshall is in between, with typical wind speed during precipitation events ranging from 1 to 6 m/s. The other relevant environmental parameters are very similar for all three sites (during precipitation), with temperature ranging mainly from -10 to 0°C and RH between 80% and 100%.

7.2. Reliability in detecting precipitation

The Hotplate has a high capability of detecting precipitation when the reference reported precipitation, as presented in Case 1 of Table 6, when no threshold is applied to the SUT, with high POD values: 99.6% for Marshall, 100% for Sodankylä and 75.3% for Haukeliseter. The cases "missed" by the Hotplate in Haukeliseter represent conditions with high wind, high relative humidity, and negative temperature (see Figure 13). These YN cases are studied in more detail in section 7.5.

This result is also confirmed with a relatively low FAR, less than 1% for Haukeliseter and 5% for Marshall. However, Sodankylä shows a very high FAR (83%). These false report events from Sodankylä occurred mainly at very low temperature (most of the events with temperature ranging from -30°C to -10°C) and high relative humidity (70% and above), as shown in Figure 14. The analysis of these NY cases is performed in detail in section 7.5.

When a threshold of 0.1 mm over 30 min events is applied to the SUT (Case 2), it can be noted that the POD of the three sites is slightly deteriorated, but still keep high values. On the other hand, the FAR scores are improved, even if Sodankylä FAR is still very high for the reasons explained under the section 7.5. Overall, applying a threshold on the SUT accumulation gives also better results for the bias and HSS scores.

7.3. Performance of SUT during no-precipitation events

As already suggested by the high FAR in Sodankylä, the signal during no-precipitation events is much noisier for the Hotplate at this site than the ones installed at Haukeliseter and Marshall (see Figure 5). Note that the no-precipitation study is performed using only the site reference precipitation detector data with the criteria ensuring that 30 min of 'no-precipitation' was recorded. For Haukeliseter and Marshall, the signal output of the Hotplate is stable and shows almost no noise, reporting zero value when there is no precipitation for 98.9% and 99.8% of the time, respectively, indicating a good agreement with the reference precipitation detector on the absence of precipitation. For Sodankylä, it is only true for 81.5% of the time, noise ranging up to 0.26 mm (see Table 7). This is most likely due to a heat plume being asymmetric from the bottom and top plates when the winds are light and the temperature low, leading to a false reading under these conditions. Once the wind exceeds 1.5 m/s, the conditions on the top and bottom plates are more similar, enabling the hotplate to use this difference to determine if precipitation is occurring or not.

7.4. Performance of SUT during precipitation events

When both the reference and the Hotplate report precipitation (YY cases), the results indicate, overall, a consistent behavior of the Hotplate across the sites and across the different precipitation types. Figure 6, for instance, shows that the Hotplate tracks well the reference accumulation with a tendency to overcatch, especially for mixed and snow precipitations, for all three sites. Figure 8 and Figure 9 give also a good glance at this results consistency across the sites when looking at the distribution of the catch efficiency, which is globally centered at the same place for all precipitation types and for the three sites, even if Haukeliseter shows a bigger dispersion for all precipitation types. The low catch efficiency for rain conditions at Haukeliseter cannot be interpreted here due to the small amount of events.

Quantitatively, the CE statistics indicate some differences in the performance of the SUT under snow conditions, with an average catch efficiency near 1 for Marshall (1.02) and Haukeliseter (1.08), while Sodankylä shows an overcatch of around 20% (CE = 1.21), as shown in Table 10. This tendency is also valid for mixed precipitation. Catch efficiency during rain events (small number of events on all three sites) shows another pattern, with a clear undercatch in Haukeliseter (0.57) and overcatch in Sodankylä (1.18) and Marshall (1.16). The low CE for rain in Haukeliseter (only 27 events out of 324 in total) is caused by events characterized with high wind speed (the CE is more centered around 1 for events with lower wind speed, as shown in Figure 11), coming from SW direction (see Figure 12). One reason for this could be the presence of a 10m mast located 5m Southwest from the Hotplate on Haukeliseter site that would have shaded the measurement area for precipitation coming from that direction.

The box plots in Figure 11 show that the scatter is very similar for all three sites for wind up to 8 m/s (up to 4 m/s at Sodankyla, its maximum wind speed) for snow, staying relatively low for all types of precipitation. The scatter increases significantly for wind speeds higher than 8 m/s at Haukeliseter for snow, varying from 0.1 to 2.5, and, to a lesser extent, in Marshall, as shown in Figure 10 and Figure 11. The nearly constant CE for snow for the wind speed range between 1 - 8 m/s suggests that no snow transfer function is needed for these conditions. For mixed precipitation types, the CE is close to one and RMSE relatively small out to 8 m/s for Marshall and 6 m/s for Haukeliseter. The increase in RMSE for Haukeliseter beyond 6 m/s for mixed conditions may be related to the high RMSE for rain as well. Further investigation is necessary to determine the cause for this behavior.

Even if better in deriving accumulation for longer time periods, the Hotplate, as the non-catchment type instruments, doesn't have an assurance of the continuity in the measurements, though, which is critical to long term data collection. If power is off or signal transmission is interrupted, no data is recorded and there is no possibility to know what has fallen during this time, thus affecting the long term data reports of the Hotplate.

Another way to assess the scatter of the measurement (and hence its repeatability) is to look at the RMSE (see Table 9). The RMSE, which reflects here the variability of measurements from the SUT against the site reference, shows lower values for Sodankylä, for all precipitation types. On the contrary, Haukeliseter shows three times more dispersion than Sodankylä due to its tendency to strong winds. This is also nicely shown in the scatter plots of the SUT vs Reference accumulation in Figure 7. From the Marshall results depicted on this figure it could be suggested that a higher

intensity may lead to higher variability in the derivation of precipitation quantity of the Hotplate, but the small number of events of higher intensity (≥ 2 mm/30 min), including those from the other sites, prevents to draw any robust conclusion.

The catch efficiency for snow shows a slight decrease with increasing wind speed for Sodankylä (see Figure 11), but the analysis is limited, since maximum wind speed reaches only 4 m/s at this site. The performance of the Hotplate under higher wind conditions, assessed with data from Haukeliseter, shows no clear dependency of the catch efficiency with wind speed. Indeed, the catch efficiency for snow oscillates around 1 for wind speed up to 14 m/s and Marshall results corroborate this observation until 8 m/s. Note that the high mean CE (above 2.5) for mixed precipitation at wind speed above 8 m/s in Marshall is due to only two events to which it shouldn't be given too much importance.

The wind roses (see Figure 12) for Sodankylä and Marshall don't show any relevant dependency of the Hotplate with wind direction. Results for Haukeliseter, however, show a clear separation, with Westerly wind causing undercatch, and Easterly wind resulting in overcatch. As already said, the undercatch may be due to the presence of the 10m mast situated 5m South-West of the Hotplate. These results highlight the importance of an appropriate siting of the sensor when installed at a site. The overcatch for Esterly winds is more difficult to explain. In Haukeliseter, the sensor was installed perpendicular to the prevailing wind direction during precipitation events, while at the two other sites the Hotplate was installed parallel to the prevailing winds. It is not known, however, if this difference in sensor orientation compared to prevailing winds could be a reason of such a dependency visible for Haukeliseter results.

7.5. Assessment of Yes-No, No-Yes events

The assessment of YN, NY events completes the picture of the performance of the SUT. As it can be seen in Table 5, when no threshold is applied to the SUT (Case 1), for the YN situation (where the reference reported precipitation and the SUT missed it, according to the criteria defined for Case 1), only Haukeliseter reported a significant number of cases; 121 out of 489 events where the reference indicated precipitation, or about 25% of the Ref Y cases. The Hotplate in Sodankylä (0 YN event) and Marshall (2 YN events) showed no issue with "missed events". The YN events in Haukeliseter tend to have occurred for temperature between -6 and 0°C, and for relative humidity around 90% (see Figure 13). The accumulation of the site weighing gauge reference during these 30 min events ranged from 0.25 mm (lower threshold set for the event selection) up to 3.4 mm and according selection criteria, the reference precipitation detector recorded precipitation more than 60% of the time for all YN events. Further investigation confirmed that these events were real precipitation events caught by the reference (i.e. not false alarm due to noise from the reference), as demonstrated in Figure 15, where some of these YN events are represented on the reference weighing gauge accumulation curve as well as on the Hotplate accumulation curve. The Hotplate seems to have really missed these periods of precipitation, and no log from either the site Manager, or the status of the sensor itself (called 'Fault Indicator') could explain this (status of the sensor always reported no issues during these periods). The fact that almost all YN events are characterized by South-Westerly wind (see Figure 13), supposes that the 10m mast located at South-West of the Hotplate may have, here again, played a role and could have shaded the Hotplate measurement area in some ways. It could also be

related to a high wind effect (see the regular increase of wind in Figure 15 from 5 to 20 m/s), which induced undercatch in the measurements.

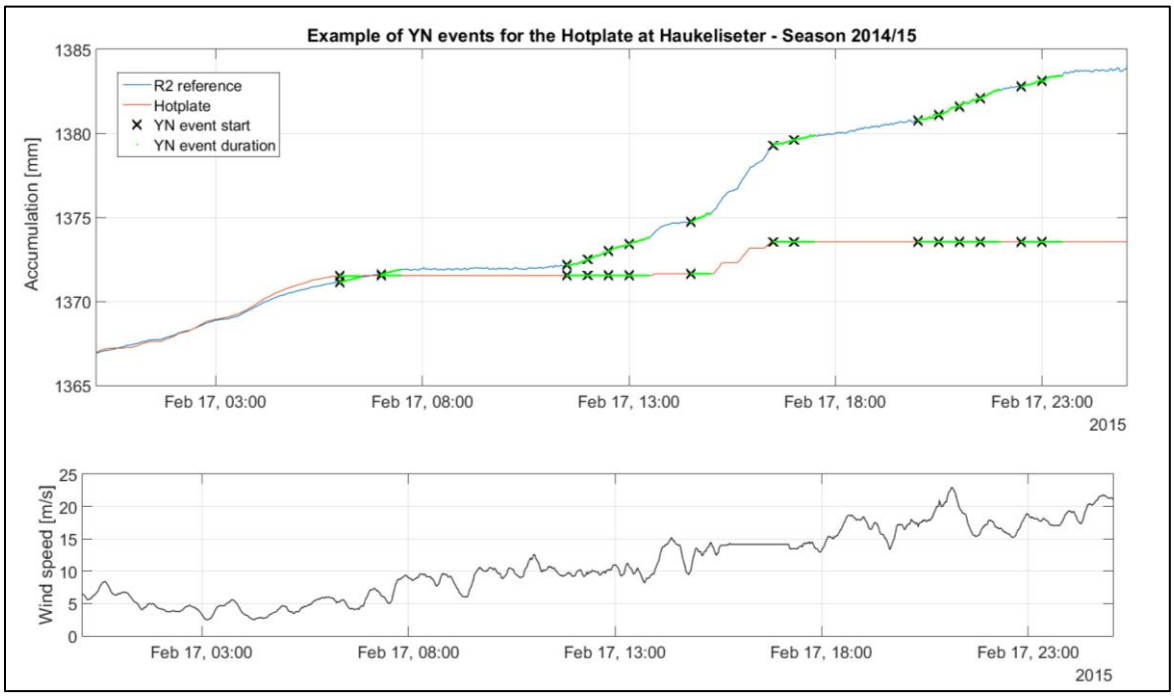


Figure 15: Example of some YN events for the SUT Hotplate compared to the R2 reference at Haukeliseter test site.

Looking at the NY events (when the reference did not record any precipitation, but the Hotplate did), only the Hotplate in Sodankylä showed a significant number of NY cases as indicated by Case 1 of Table 5; 1894 out of 10'251 events where the reference indicated no precipitation, or about 18% of the Ref N cases. This high number of "false alarms" (reflected by the high FAR of 83% for Sodankylä in Table 6) contrasts with the results of the other two sites, where this didn't happen as frequently with only 3 NY events for Haukeliseter and 26 NY events for Marshall. The weather condition characterizing these cases in Sodankylä shows low temperature (-30 to 0°C), high relative humidity (70-90%) and very low wind speeds (below 2 m/s), as shown in Figure 14, and for Hotplate accumulation ranging from 0 mm (no minimum threshold set for SUT, see Section 4.1.1) up to 0.26 mm. More in-depth analysis showed that these false reports seem to be mostly related to important temperature and humidity gradients, most of them appearing when temperature drops below -10 °C, on a regular (daily) basis for cold periods. Figure 16 gives time series examples of such behavior in two different time periods where NY events occurred for the Hotplate at Sodankylä; the NY event starts are represented on both the reference and the Hotplate accumulation curves as grey crosses and temperature and relative humidity (RH) are given on a secondary axis, with the -10°C highlighted in a dashed line for comparison. This observation leads to the conclusion that these NY events are actual real false reports of the Hotplate and need more attention from the manufacturer. The reason why they appeared at Sodankylä and not on the other two sites could be explained by the formation of heat plumes over the instrument under the especially low wind speed conditions at Sodankylä.

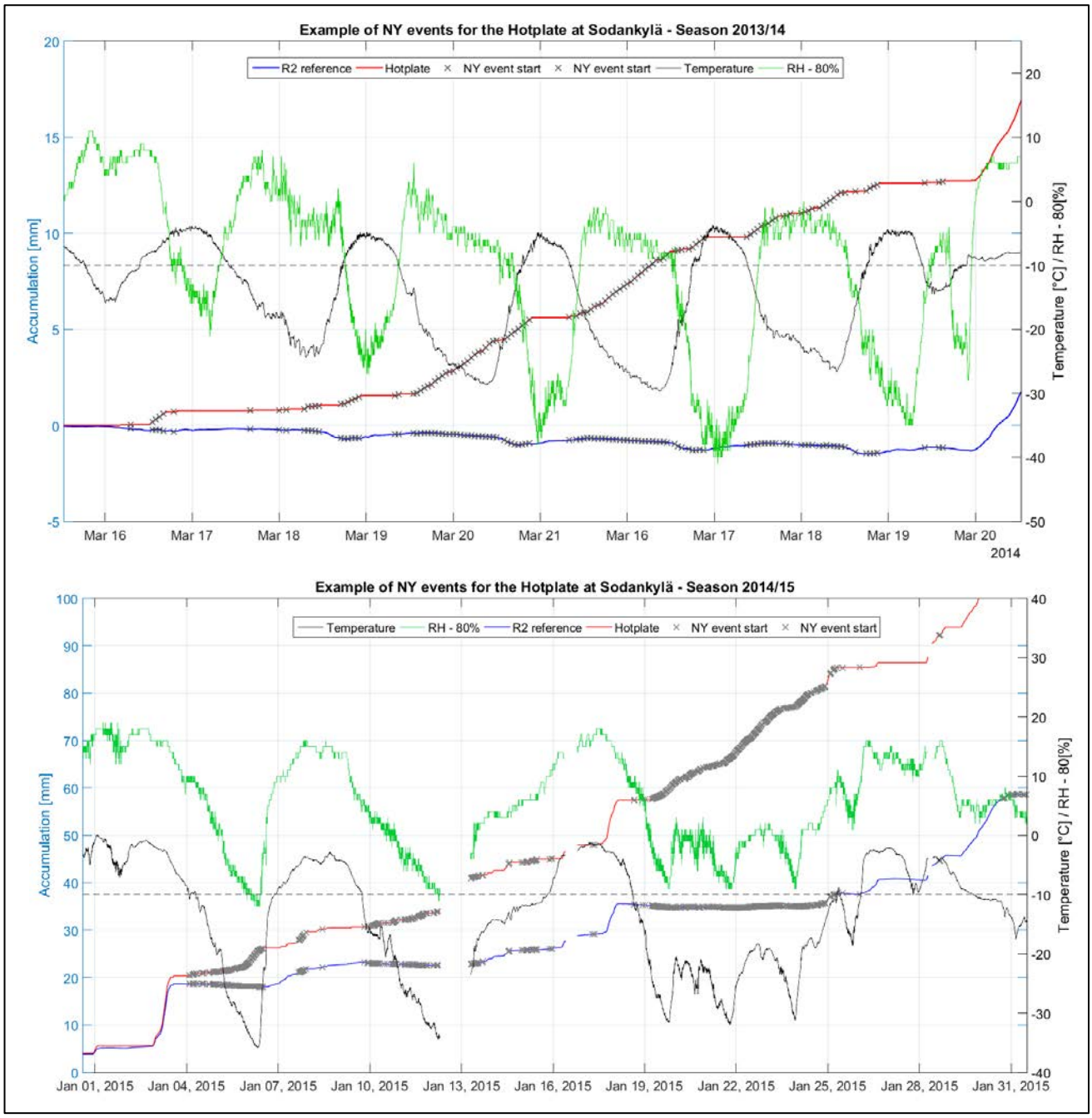


Figure 16: Examples of YN events for the SUT Hotplate compared to the R2 reference at Sodankylä test site for (top) 2013/14 and (bottom) 2014/15 seasons.

7.6. Threshold selection

The threshold to be set for the Hotplate in order to report precipitation adequately over a 30 min interval (3 STD, according to the methodology defined in Section 3.6.1.3.2 of the SPICE Final Report of the SPICE Report) corresponds to 0.009 mm (3 x 0.003 mm, from Table 7), or 0.135 mm (3 x 0.045 mm) if we include the results from Sodankylä, which comprise the higher noise level induced by the NY events. As an example of what can be applied, looking at the contingency tables for Case 2 (Table 5), a threshold of 0.1 mm/30 min enables to reduce the number of false reports and improve bias and HSS scores, while keeping high POD values.

8. Operational considerations

The overall experience with the Hotplate at the three sites was positive.

Overall, the Hotplate was found easy to install and maintain. Its relatively small and light construction was given as an advantage, together with the fact that it comprises a complete weather station with interesting data products. It was considered to be operationally reliable, with no breaks in the data. One key strength that was raised is that it didn't accumulate snow on its mounting parts, nor on the sensor itself (thanks to its high temperature directly related to the measurement principle), as illustrated in Figure 17, in Sodankylä, over a quiet period (in terms of wind) of important accumulation, where other sensors experienced capping or accumulation on devices.

Following points were raised by site managers as potential improvements:

- The high power consumption requires stable AC power supply and surge protection. The system would benefit by having this onboard the instrument.
- The extended output are only available using the hyperterminal to collect the data and not available when using the manufacturer's software package.
- As set up by the manufacturer, the distance between instrument and PC is limited by communication type.
- A calibration (or verification) method for precipitation amount missed in the documentation.

Figure 17: No snow accumulation observed on the Hotplate device at Sodankylä site despite an important snowfall, 2015-02-15, 11:00 UTC.

8.1. Maintenance

The hotplate does not need any regular maintenance, making it really suitable for remote sites. Precipitation tends to keep the plates relatively clean and the very warm surface temperature of the plates tends to prevent birds and insects from landing on them.

8.2. Noted issue

At Haukeliseter, the Hotplate data encountered some issues for half of the season (December 2014 to February 2015). This was due to a format incompatibility between the logger configuration and the Hotplate. This site used the optional RS485 connection instead of the conventional LAN-port recommended by the manufacturer and had some difficulties to connect correctly the sensor. The data impacted (around 3% of the entire season) were removed from the present analysis by the manual and automatic SPICE quality-control procedures. This issue explains partly why Haukeliseter, additionally to the fact that the Hotplate was run only for one season, has so few events to analyze compared to the other two sites.

9. Performance considerations

- High scatter of CE for wind speed above 6-8 m/s for snow on an event basis (30 min), but good mean CE over the seasons, suggests that the hotplate should be mainly deployed under conditions in which the winds are < 8 m/s during precipitation.
- Overall, the Hotplate has a general tendency to overestimate mixed and solid precipitation as compared with the reference and it doesn't seem to have a CE wind speed dependency.
- When no threshold is applied to the SUT accumulation (Case 1), a high number of false alarms (NY) events were reported by the Hotplate in Sodankylä and also several 'missed' (YN) events appeared at Haukeliseter, when comparing with the respective field reference. These events are probably linked with a heat plume over the instrument and high wind conditions, respectively. Applying a threshold of 0.1 mm/30 min showed to reduce these YN/NY events and therefore ensures better derivation of solid precipitation accumulation.
- An advantage of the hotplate is the low RMSE between 1 to 8 m/s for snow.
- The larger scatter for mixed precipitation above 8 m/s at Haukeliseter is not well understood and should be the subject of additional studies. This may be due to blockage by other instruments, but currently beyond the current analysis approach.

WMO-SPICE Instrument Performance Report
OTT Hydromet Parsivel²

1. Technical specifications

Instrument model:	OTT Parsivel^2
Measuring area:	54 cm^2
Physical principle:	Laser-optical disdrometer for capturing hydrometeor size and fall velocity
Operating temperature range:	-40 to +70°C
Measurement uncertainty:	+/- 5 %
Sensitivity:	0.001 mm/h (for drizzle)

Note: Specifications from manufacturer provided documentation.

Figure 1: Parsivel² at Sodankylä test site.

2. Data output format

The disdrometer OTT Parsivel^2 is a sensor outputting precipitation information as intensity, accumulation or weather code. The raw data (particle size and fall velocity) are also available for the user. Table 1 summarizes the main output parameters from the instrument that are derived from the raw data. The firmware version of the OTT Parsivel^2 tested during this experiment was the version V2.02.1.

Table 1: Summary of main instrument outputs, as recorded by the site during the experiment.

Measured Parameters	Units
Rain Intensity	[mm/h]
Rain Amount accumulated	[mm]
Weather Code SYNOP, Tab 4680	-
Weather Code SYNOP, Tab 4677	-
Weather Code SYNOP, Tab 4678	-
MOR Visibility in the precipitation	[m]
Radar reflectivity	[dBz]
Number of detected particles	[#/min]
Housekeeping	-

This document reports on the ability of the Parsivel[2] to derive solid precipitation accumulation. The results should have consequently been derived from the 'Rain Amount accumulated' output, but firmware version V2.02.1 experienced an issue with this parameter (fixed in subsequent firmware versions, confirmed by the manufacturer, October 2015), that prevented a fair comparison with respect to the field reference (refer to section 6.3.3 of the present report for more details). The analysis has therefore been performed using the cumulative sum of the 'Rain Intensity' parameter.

3. SPICE test configuration

The Parsivel[2], as sensor under test (SUT), has been tested on one site:

Test Site: Sodankylä (Finland)

Sensor Provider(s): The instrument evaluated was provided by the manufacturer (OTT Hydromet)

Figure 2: Map of SPICE site testing Parsivel[2] instrument.

A summary on the configuration of the instrument as tested, the duration of tests and availability reflected in these results, and the ancillary measurements used, is available in Table 2, Table 3, and Table 4, respectively.

Table 2: Summary of instrument configuration and data output. Details and photos on site configuration are available in the site commissioning protocol.

	Sodankylä
Main prevailing wind directions	South
Sensor orientation	East – West
Height of installation	2.1 m
Heating	Heating of sensor heads, as recommended
Shield	No
Data QC	SPICE QC methodology
Data temporal resolution	1 min
Processing interval for SPICE data analysis	30 min

Table 3: Data availability, by measurement season.

Measurement season	**Sodankylä**
Season 1 (Oct. 2013 – Apr. 2014)	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓

Table 4: Summary of reference and ancillary measurements, with measurement height.

	Sodankylä
R2 Site Reference	OTT Pluvio^2 1500mm (DFAR) (4 m, rim height)
R2 Precip Detector	DRD11A (Site*, 2013-2014) (1 m) OTT Parsivel^2 (DFAR, 2014-2015) (2.7 m)
Ancillary Temp Sensor	Vaisala HMP155 (2 m)
Ancillary RH Sensor	Vaisala HMP155 (2 m)
Ancillary Wind Sensor	Thies acoustic 2D wind sensor (3.5 m)

**A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR-fence; however, in cases where a more sensitive detector is available outside of the DFIR-fence, or there are issues with the detector within the DFIR-fence, a precipitation detector elsewhere on the site can be employed.*

4. Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6 of the SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute intervals:

- Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip;
- SUT 'Yes': SUT accumulation > 0 mm;
- SUT 'No': SUT accumulation = 0 mm.

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores.

4.1.2. Skill score assessment

The ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

The above scores are computed using the formulations provided in Section 3.6 of the SPICE Final Report.

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

5. Environmental conditions

The environmental conditions at the site over the duration of the test period are expressed as probability density functions (PDFs) for air temperature, relative humidity, wind speed, wind direction, and precipitation rate in Figure 3. Both the entire period data, and the data corresponding only to the precipitation events reported by the site reference, R2, are shown. The precipitation percentage in Figure 3 represents the number of minutes of precipitation over a standard 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration.

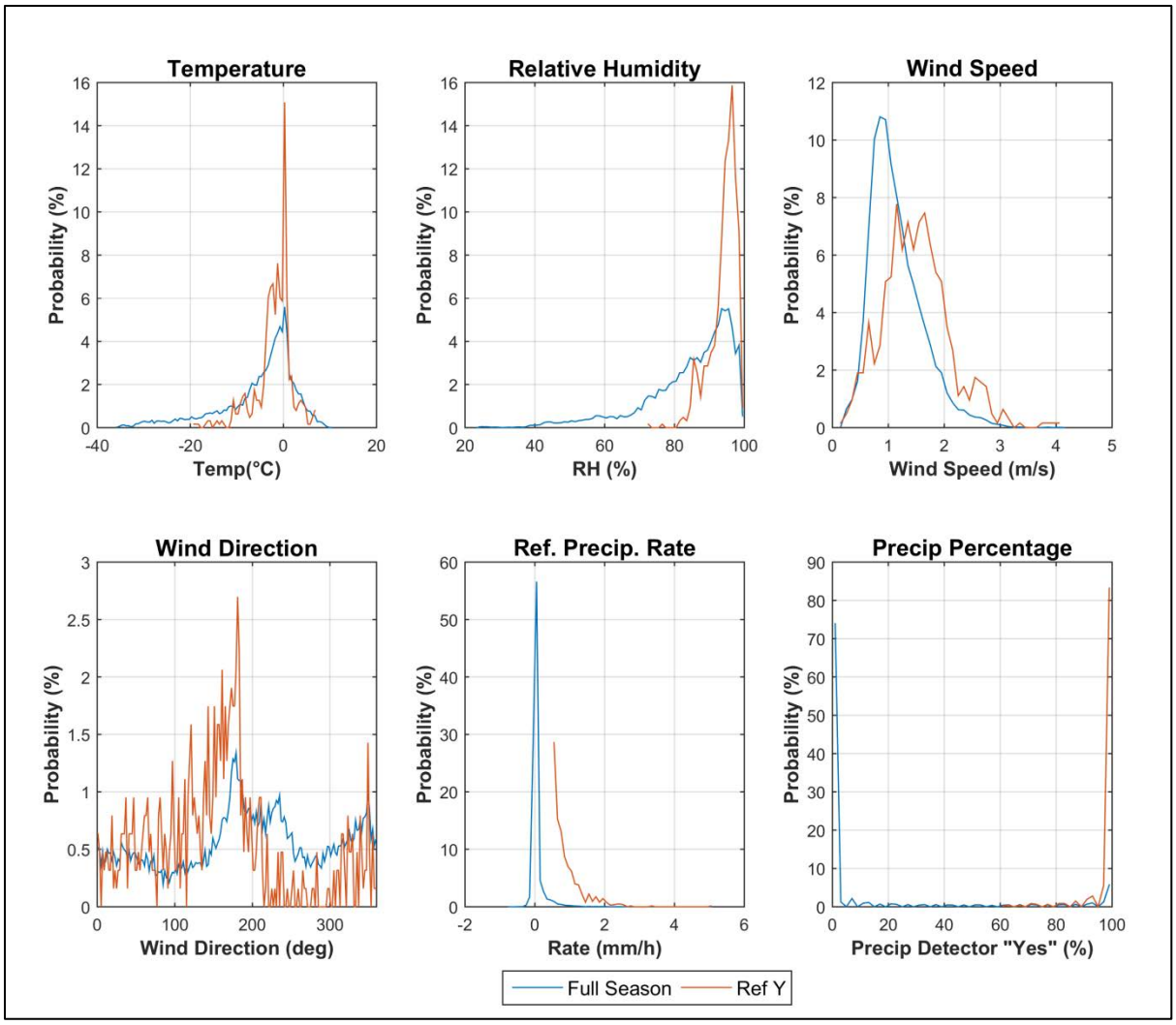


Figure 3: Summary of aggregated environmental conditions on the SPICE site that operated Parsivel[2], over the entire duration ('Full Season', in blue) and for data corresponding to precipitation events, as reported by the site R2 reference ('Ref Y', in red), during the tests, as per Table 3 above.

6. Evaluation of performance over the range of operating conditions

6.1. Skill score assessment

The ability of the SUT to represent precipitation similarly to the site field reference R2, is assessed using contingency tables (Section 4.1.1) and derived selected skill scores (Section 4.1.2). To better understand the potential influence of threshold choices on the derived results, two cases are considered here (see note below). The contingency results related to these two cases are given in Table 5 and the respective skill scores in Table 6.

Note: Following the data derivation explained in Section 4.1.1, the conditions required to have a 'Yes' or a 'No' event over the 30 min interval, for the reference and the SUT, for the two different cases treated here, are:

CASE 1 *(as defined in Section 4.1.1)*:

- *Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip*
- *Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip*
- *SUT 'Yes': SUT accumulation > 0 mm*
- *SUT 'No': SUT accumulation = 0 mm*

CASE 2:

- *Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip*
- *Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip*
- *SUT 'Yes': SUT accumulation ≥ 0.1 mm*
- *SUT 'No': SUT accumulation < 0.1 mm*

Results of this report are based on Case 1.

Table 5: Contingency Tables: detection of precipitation of the Parsivel² relative to the field reference, expressed as number of events over the entire test period. The skill scores associated with these events are given in Table 6.

Case 1

SUT Parsivel²	Ref Plv2DFAR Yes	No	Total
Yes	615	675	1290
No	0	10847	10847
Total	615	11522	12137

Case 2

SUT Parsivel²	Ref Plv2DFAR Yes	No	Total
Yes	614	1	615
No	1	11521	11522
Total	615	11522	12137

Table 6: Skill Scores for the Parsivel². POD: Probability Of Detection, FAR: False Alarm Rate, B: Bias, HSS: Heidke Skill Score (see Section 4.1.2 for more details).

	Parsivel², Skill Scores			
	POD	**FAR**	**B**	**HSS**
Case 1	100%	52.3%	210%	62.0%
Case 2	99.8%	0.2%	100%	99.8%

6.2. Assessment of SUT performance during non-precipitating events

The performance of the SUT in the absence of precipitation (when the reference precipitation detector recorded 30 minutes without precipitation) is represented in Figure 4 and Table 7, reflecting the distribution of the sensor response, as measured during the interval.

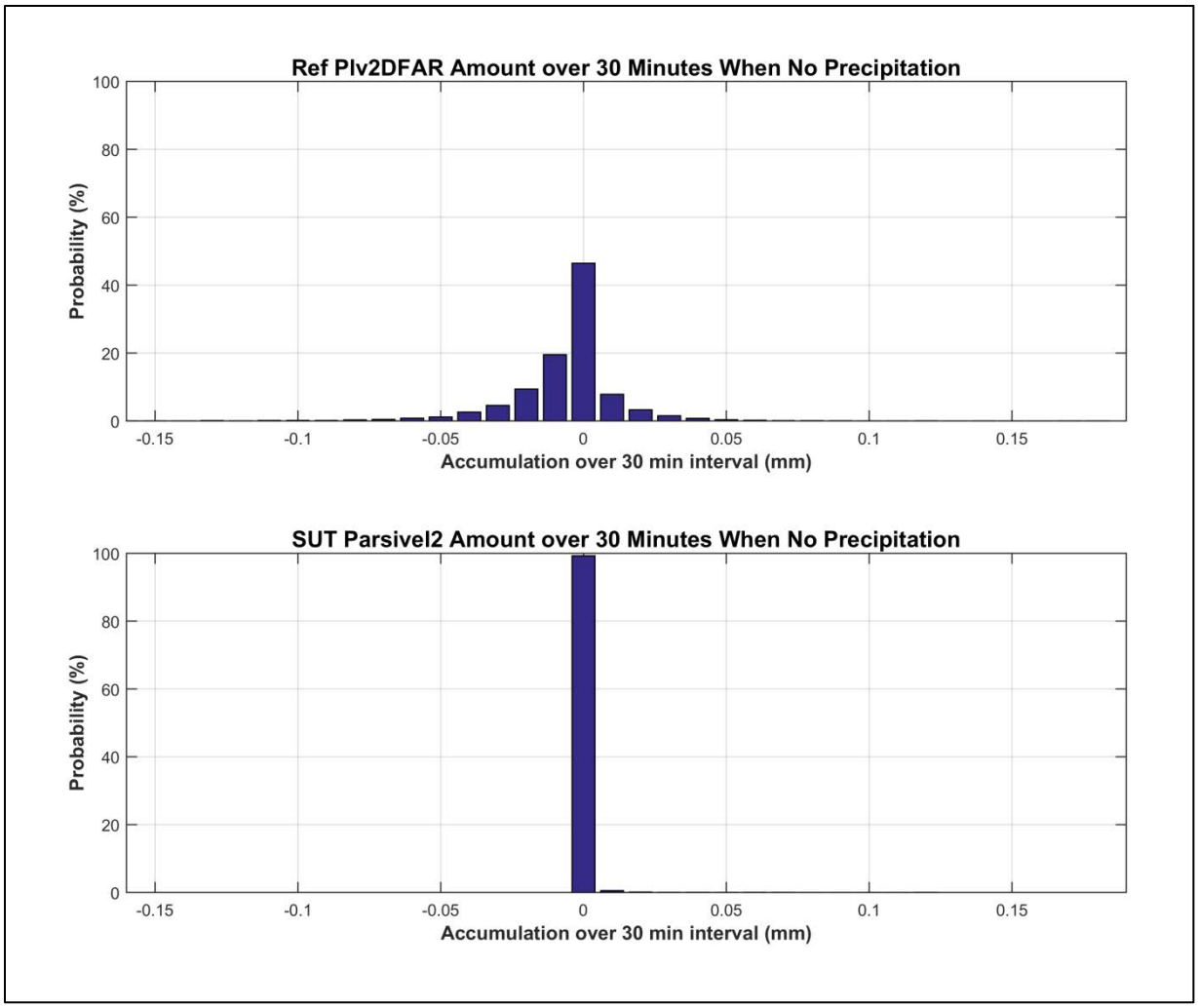


Figure 4: Probability of occurrence of a response during a 30 min interval in the absence of precipitation, represented by the signal output from (top) the R2 reference and (bottom) the SUT Parsivel^2. Statistics associated with these graphs are given in Table 7.

Table 7: Reference and SUT statistics of response signal when no precipitation was occurring, as plotted in Figure 4; Average (Avg), standard deviation (STD), maximum (Max) and minimum (Min) of the response signal, together with the number of events (Num) over the test period is given.

No Precip Statistics – Reference: Plv2DFAR

	Ref Avg [mm]	Ref STD [mm]	Ref Max [mm]	Ref Min [mm]	Ref Num
Sodankylä	-0.002	0.020	0.190	-0.160	11531

No Precip Statistics – SUT: Parsivel^2

	SUT Avg [mm]	SUT STD [mm]	SUT Max [mm]	SUT Min [mm]	SUT Num
Sodankylä	0.000	0.003	0.125	0.000	11531

6.3. Ability of the SUT to measure precipitation

6.3.1. Yes-Yes cases

Quantitatively, the performance of the SUT to derive and report precipitation is assessed relative to the site reference in several graphs and tables illustrated in this section, using only the cases where both instruments reported precipitation over the 30 min interval, according to the criteria used in Case 1 of Table 5 (cases 'Yes-Yes', or shorter 'YY').

6.3.1.1. Time series plots

The time series (cumulative sum of 30 min YY events accumulation) of the SUT is plotted against the reference for the two seasons, by precipitation type (see Section 4.1.4) in Figure 5.

The corresponding seasonal accumulations are given in Table 8.

Table 8: Seasonal accumulation [mm] for Sodankylä test site based on the sum of YY events from the SUT Parsivel^2 and the field reference R2.

Sodankylä				
	Season 2013/14		Season 2014/15	
[mm]	Ref	SUT	Ref	SUT
All events	160.47	177.63	114.03	139.05
Rain events	17.67	16.00	5.60	5.70
Mixed events	92.12	101.01	51.04	63.22
Snow events	50.68	60.63	57.39	70.14

6.3.1.2. Scatter plots and RMSE values

Scatter plot of the amount derived by the SUT versus the reference amount, and discriminated by precipitation type, is given in Figure 6.

Quantitatively, the SUT performance is assessed using the Root Mean Square Error (RMSE) also known in practice as Operational Comparability. The results are available in Table 9.

Table 9: Root Mean Square Error (RMSE) statistics, in mm, for the SUT Parsivel^2 with respect to the field reference, by precipitation type, including both seasons data.

RMSE [mm]	All	Rain	Mixed	Snow
Sodankylä	0.208	0.075	0.192	0.241

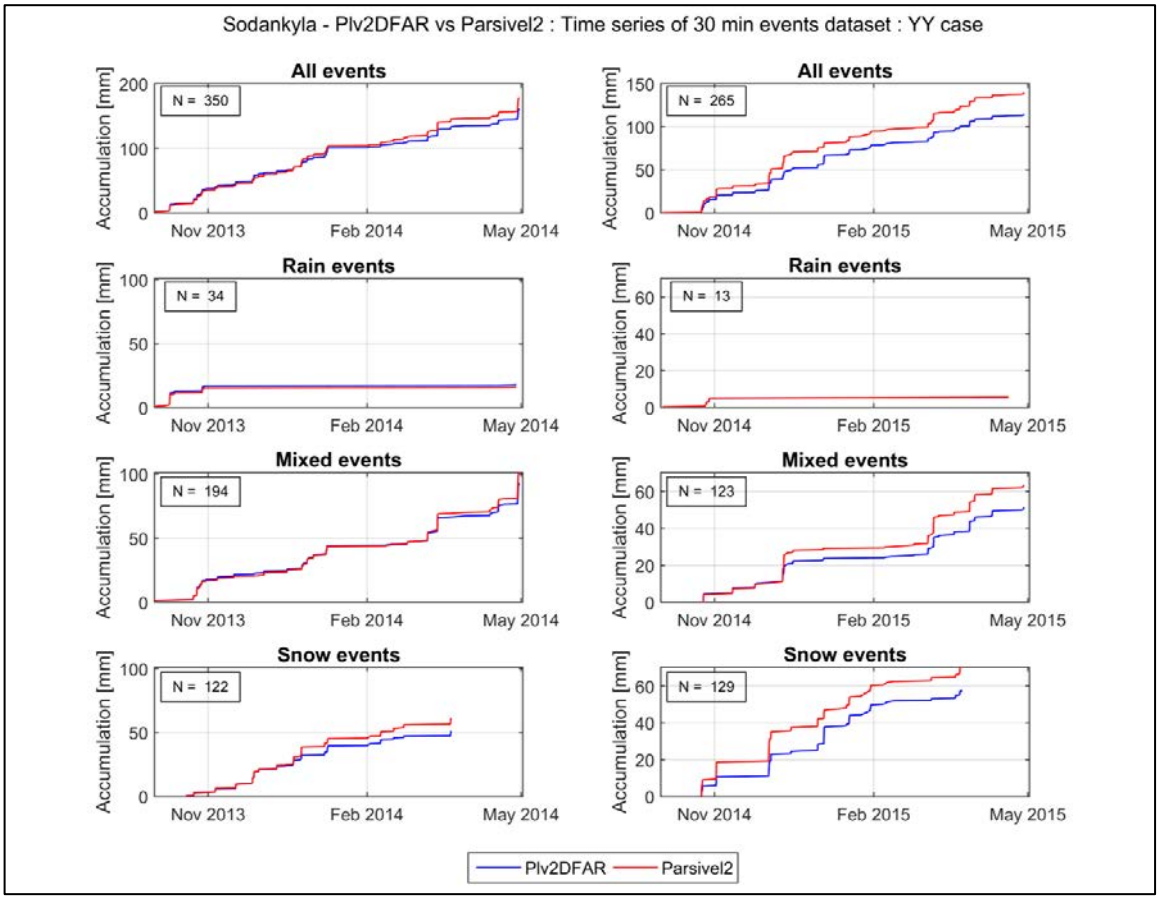


Figure 5: Time series based on 30 min YY events of the SUT Parsivel[2] against the field reference, discriminated by precipitation type (Rain, Mixed, Snow), for both seasons (2013/14 on the left, 2014/15 on the right).

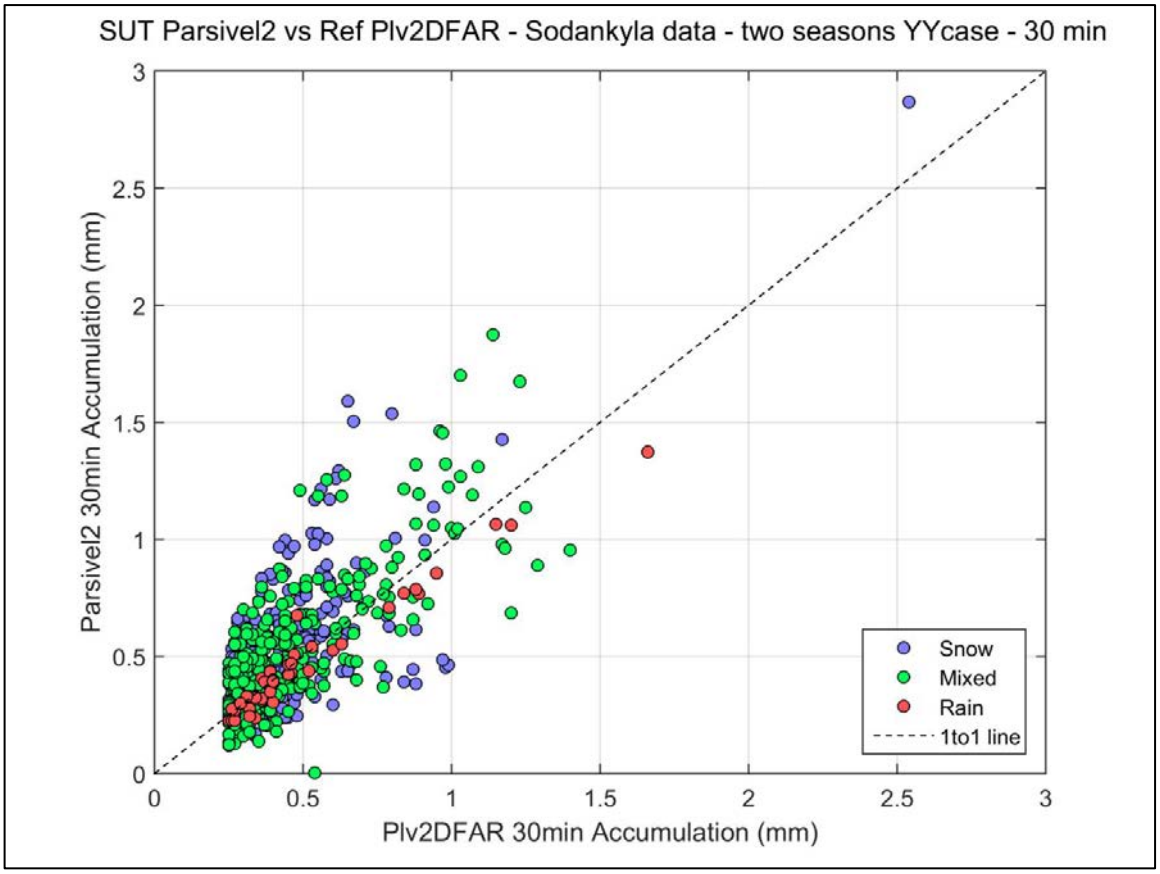


Figure 6: Scatter plot based on 30 min YY events accumulation from the SUT Parsivel[2] against the field reference, for the two seasons, discriminated by precipitation type.

6.3.1.3. Catch efficiency evaluation by precipitation type

The Catch Efficiency (CE) of the SUT is represented by histograms (Figure 7) and boxplots (Figure 8), discriminated by precipitation type.

The quantitative evaluation of the CE is provided in Table 10. The mean catch efficiency is given in the first line of this table, considering both seasons data and for each category of precipitation type as well as for all the events together.

Note: All events with a CE greater than 3, if any, are included in one category named '3 and more' or '≥3' in the upcoming graphs. Additionally, for all graphs representing the CE, a line is added at CE = 1, which represents the ideal case where the SUT reports exactly the same precipitation amount as the reference.

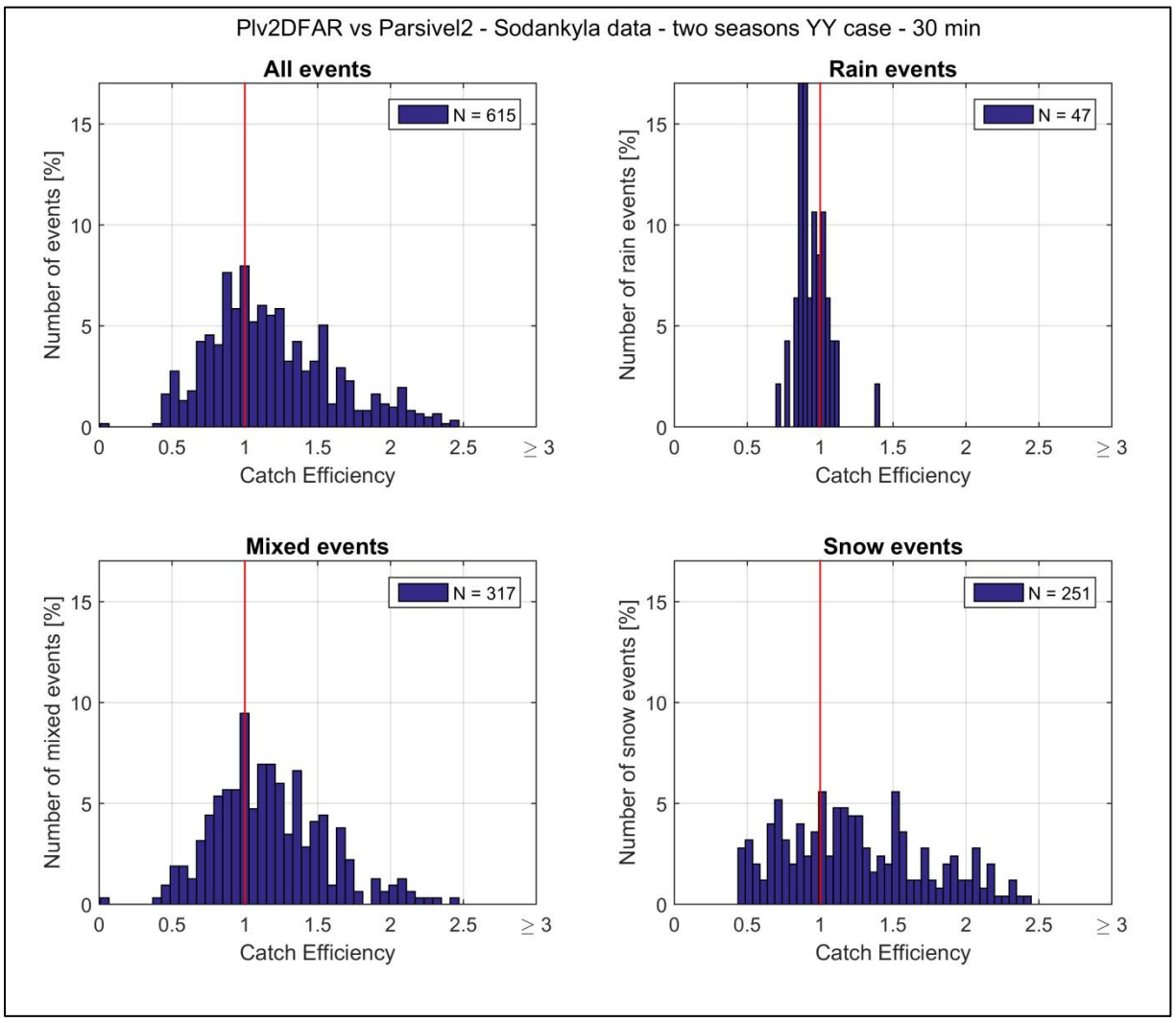


Figure 7: Histograms based on 30 min YY events for the two seasons, representing the distribution of the catch efficiency of the SUT Parsivel^2 against the field reference (SUT/Ref), discriminated by precipitation type, with number of events given in the legend for each category. The red line at CE = 1 represents the ideal case.

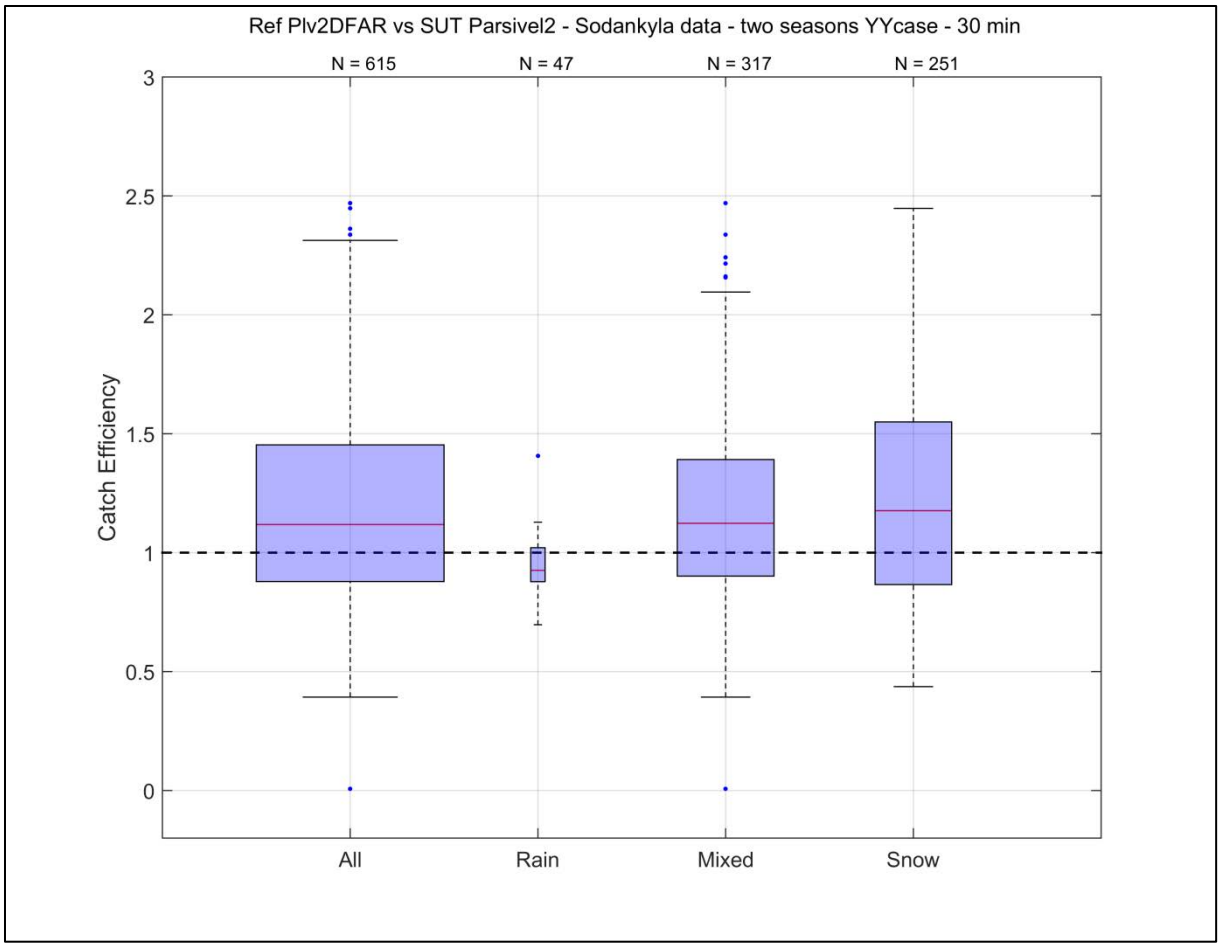


Figure 8: Boxplot based on 30 min YY events of the two seasons, representing the catch efficiency of the SUT Parsivel^2 against the field reference (SUT/Ref), discriminated by precipitation type, with number of events given at the top of each category. The width of the boxes is proportional to the percentage of events in each category ('All' box represents 100% data). The dashed black line at CE = 1 represents the ideal case.

Table 10: Statistics related to the CE of SUT Parsivel^2 against the field reference (SUT/Ref, see Figure 8), discriminated by precipitation type.

Catch Efficiency Statistics: **Parsivel^2**

CE Boxplot Parameters	All	Rain	Mixed	Snow
Mean	1.18	0.95	1.17	1.24
Median	1.12	0.93	1.12	1.18
75 percentile	1.45	1.02	1.39	1.55
25 percentile	0.88	0.88	0.90	0.87
Upper Whisker	2.31	1.13	2.10	2.45
Lower Whisker	0.39	0.70	0.39	0.44
Maximum	2.47	1.41	2.47	2.45
Minimum	0.01	0.70	0.01	0.44
# Outliers	5	1	7	1
# Outliers ≥ 3	0	0	0	0
# Events	615	47	317	251

6.3.1.4. Catch efficiency dependency on wind speed

The variation of the SUT catch efficiency (SUT/Ref) with wind speed is illustrated in a scatter plot together with a boxplot in Figure 9, both discriminated by precipitation type.

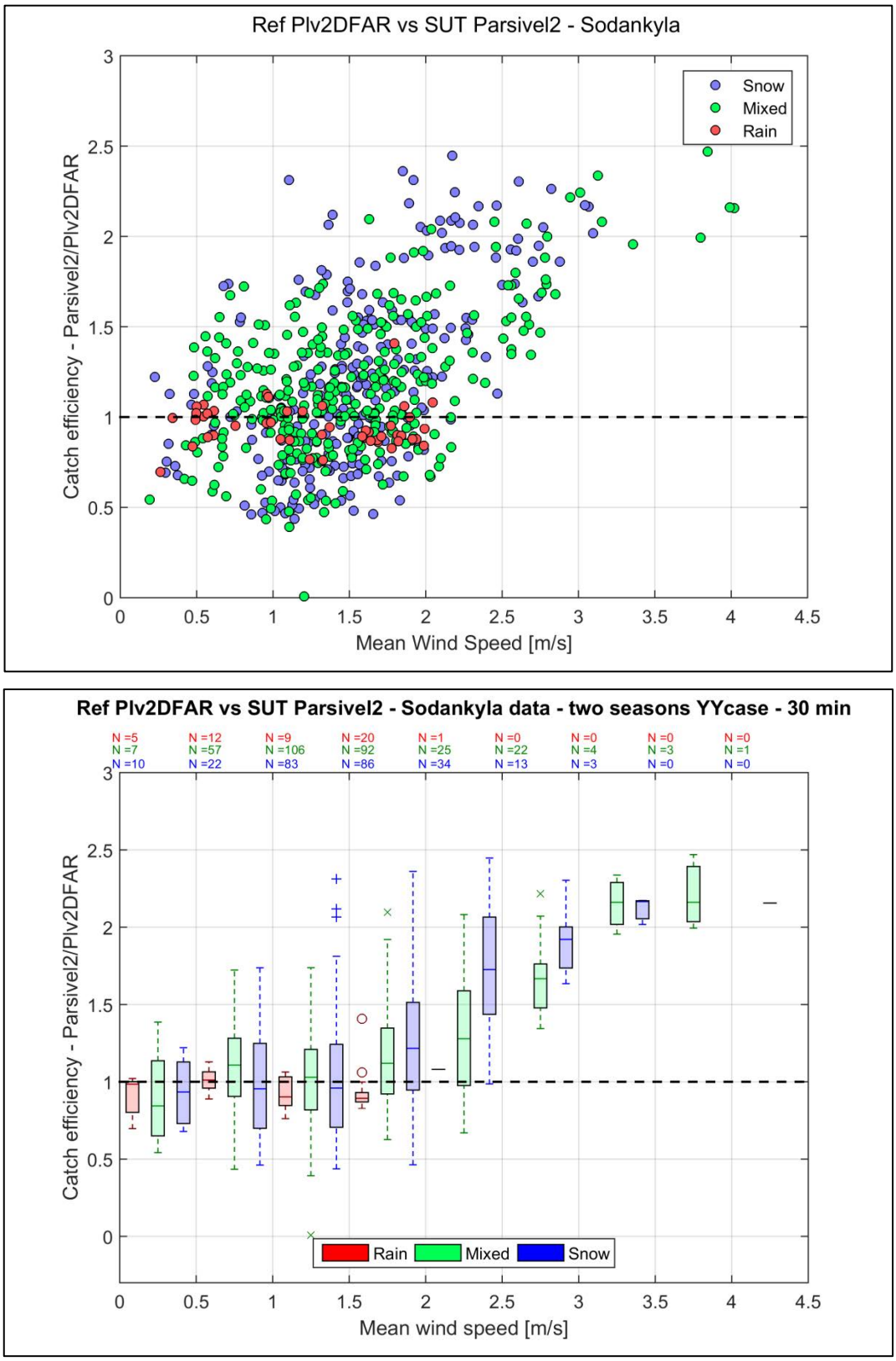


Figure 9: Scatter plot (top) and boxplot (bottom) based on 30 min YY events for the two seasons, representing the catch efficiency of the SUT Parsivel^2 with respect to the field reference (SUT/Ref), against wind speed and discriminated by precipitation type. The dashed black line at CE = 1 represents the ideal case.

6.3.1.5. Catch efficiency dependency on wind direction

In order to assess the dependency of the CE with wind direction, a wind rose is produced (Figure 10) representing the wind data of the two seasons, binned by catch efficiency in order to represent undercatch (CE < 0.8), overcatch (CE > 1.2) and catch efficiency of 1 ± 20% of the SUT.

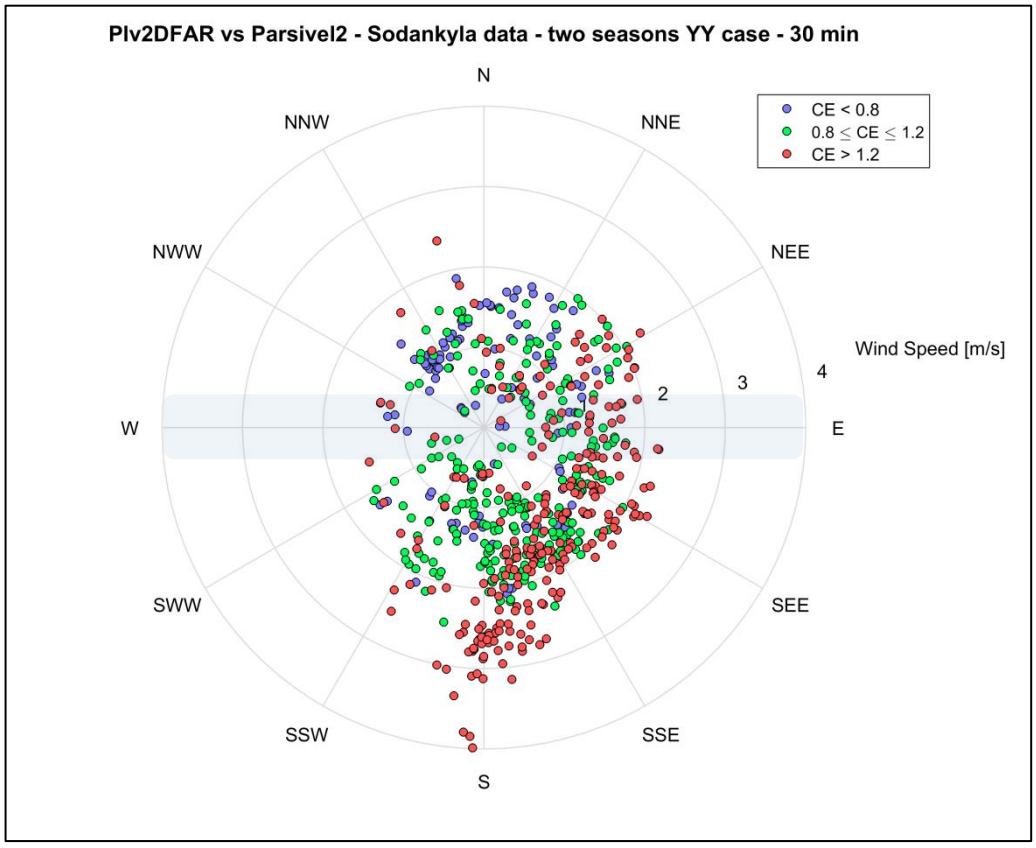


Figure 10: Precipitation events (YY cases) as function of wind speed and direction, and binned by catch efficiency (CE). The grey zone indicates the SUT orientation.

6.3.2. Yes-No and No-Yes cases

Events when the site reference and SUT do not agree on the occurrence of precipitation includes two categories of cases (Section 4.1.1): (1) when the field reference reported a precipitation event, while the SUT did not (Yes-No cases, 'YN'), and (2) when the field reference did not report a precipitation event while the SUT did (No-Yes cases, 'NY').

Histograms illustrating field reference and SUT reports and associated site conditions for all NY cases (no YN cases were reported for the Parsivel^2) during the test period are provided in Figure 11.

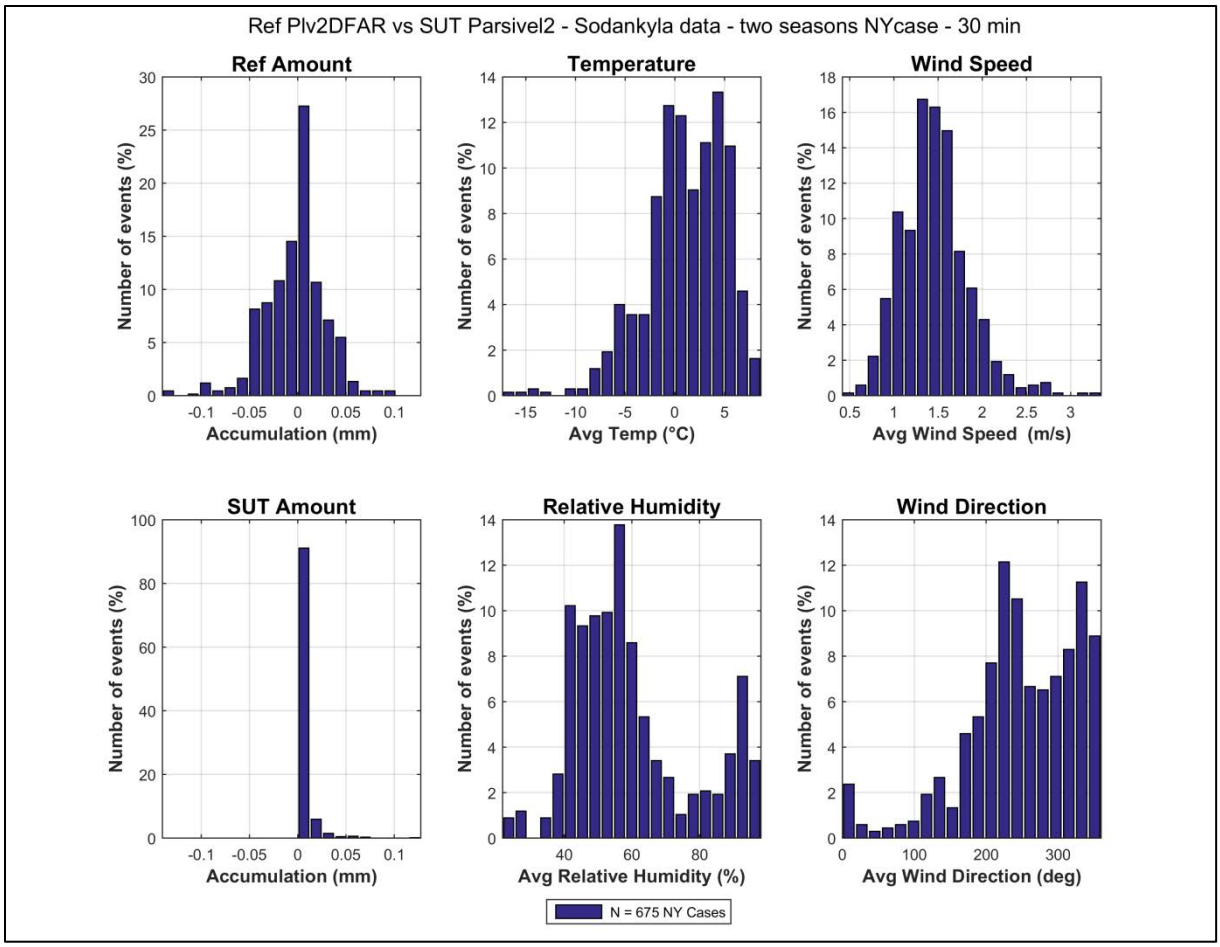


Figure 11: Histograms of SUT Parsivel^2 and field reference accumulations (left column), along with distributions of mean temperature, mean relative humidity, mean wind speed, and mean wind direction for all NY cases (number indicated in the legend) of the 30 min intervals, as reported by Case 1 of Table 5.

6.3.3. Reporting of 'Rain Amount accumulated' in firmware V2.0.2.1

The Parsivel^2 has two output parameters reporting precipitation quantity, the 'Rain Intensity' and the 'Rain Amount accumulated' parameters, as mentioned in Table 1.

Figure 12 shows time series of these two parameters for both seasons (cumulative sum for the 'Rain Intensity' parameter). It was confirmed with the manufacturer that the firmware version 2.02.1 had an error in deriving the 'Rain Amount accumulated' (likely a factor 2), affecting measurements during both seasons, as shown in Figure 12. According to the manufacturer, this factor issue has been fixed in newly firmware releases (Brussels meeting, October 2015).

To prevent this factor to impact SPICE results and to allow a fair assessment of the performance of the Parsivel^2, the analysis has been done only based on the cumulative sum of the 'Rain intensity' parameter.

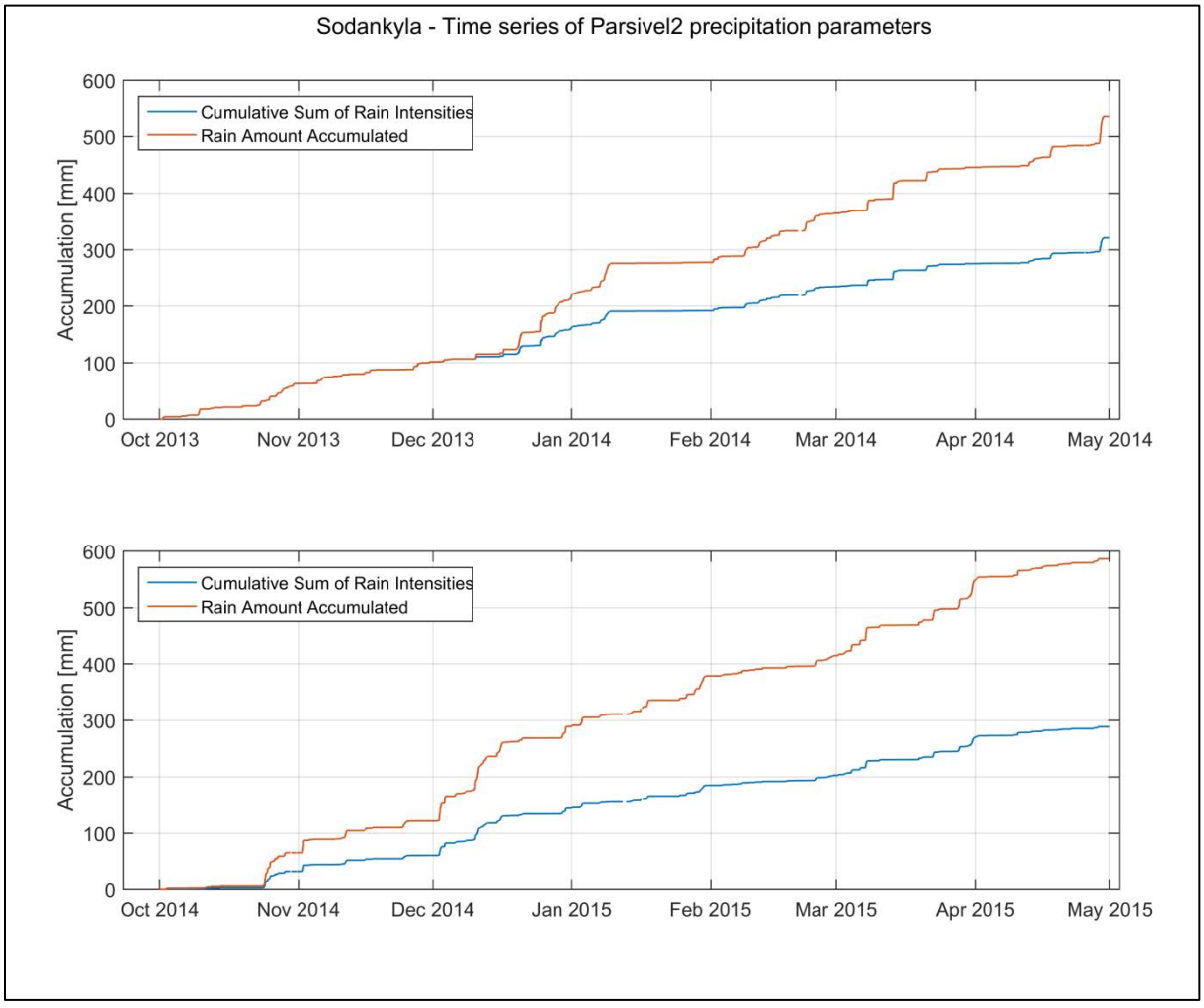


Figure 12: Time series of Parsivel^2 precipitation amount parameters for (top) season 2013/14 and (bottom) season 2014/15. The rain intensities have been turned into a cumulative sum for comparison.

7. Interpretation of results

7.1. Operating conditions

Sodankylä is a site characterized by low wind speeds, as shown in Figure 3, with a majority of precipitation events occurring below 3 m/s. Regarding temperatures, Sodankylä recorded most of the precipitation events at temperature between -10 and 0°C, with extreme by -20°C. The site shows fairly humid climate conditions as expected from a Northern Boreal climate site.

7.2. Reliability in detecting precipitation

The Parsivel^2 was efficient in detecting precipitation when the reference reported precipitation, as presented in Table 6, with a POD of 99% or more, looking at the two different threshold cases. When no threshold is applied to the SUT (Case 1), the FAR of 52.3%, the bias around 210% and the HSS of 62% are a direct consequence of the large number of NY events (see Section 7.5 for more details). However, it can be noted that the FAR score is significantly reduced to less than 1%, and the bias and HSS scores improved, when a threshold of 0.1 mm over 30 min events is applied to the SUT (Case 2), indicating that the majority of the NY events of Case 1 are for very low accumulations.

7.3. Performance of SUT during no-precipitation events

The Parsivel^2 has a stable output signal, showing almost no noise during no-precipitation events (see Figure 4 and Table 7). Note that the no-precipitation study is performed using only the reference precipitation detector data (a DRD11A for the first season and a Parsivel^2 for the second season) with the criteria ensuring that 30 min of 'no-precipitation' was recorded. For these cases, the Parsivel^2 indicates 0 mm for more than 94% of the time, indicating a good agreement with the reference precipitation detector on the absence of precipitation. Most of the remaining cases (where the reference precipitation detector doesn't detect any precipitation while the SUT does) correspond to very low accumulations reported by the SUT (less than 0.1 mm/30 min) and can be related to the NY events mentioned above and analyzed in Section 7.5.

7.4. Performance of SUT during precipitation events

The ability of the SUT to derive correct precipitation accumulation (according to the reference) is assessed in this section, using the 'YY' cases.

The RMSE in Table 9 shows different results according to precipitation type (0.075 mm for rain, 0.192 mm for mixed, and 0.241 mm for snow over the 30 min events). The higher RMSE values are reflected in the larger scatter noticeable for snow and mixed precipitation (compared to rain), as shown in Figure 6, as well as in Figure 7, where the catch efficiency varies between 0.5 and 2.5.

The mean catch efficiency has consequently also a different behavior according to precipitation type, with 0.95 for rain, 1.17 for mixed, and 1.24 for snow, leading, overall, to an overcatch with respect to the reference (see Table 10 and Figure 8), mainly influenced by the higher number of mixed and snow events than the number of rain events.

Assessing the catch efficiency as a function of wind speed shows that it is very close to 1 for wind speed below 2 m/s, and for all precipitation types (Figure 9). For higher wind speed, as measured up to 4 m/s (maximum wind speed in Sodankylä), the catch ratio increases for mixed and snow

precipitation (no rain events were recorded at these wind speeds). The reason for this increase could not be identified with the information available at the time of the analysis, and the number of events at large wind speed is not sufficient to draw a robust conclusion. An assessment of the instrument under higher wind conditions would be valuable to complement this analysis.

According to the results discussed above, the Parsivel^2 seems to be a reliable instrument to account for the total accumulation over a longer period (e.g. one season), with a catch ratio close to 1, especially for wind speed below 2 m/s, as shown in Figure 8 or Figure 9 (bottom). The relatively high scatter visible in Figure 9 (top), on event based statistics (typically 30 min interval), occurring at all wind speeds (up to 4 m/s), tends to show that the Parsivel^2 is less reliable to derive solid and mixed precipitation accumulation over near real time periods (for rain precipitation, the scatter remains low).

Even if better in deriving accumulation for longer time periods, the Parsivel^2, as other non-catchment type instruments, doesn't have an assurance of the continuity in the measurements, though, which is critical to long term data collection. Indeed, the sensor reports data only when powered. If power is off or signal transmission is interrupted, no data will be recorded, thus affecting the long term data reports of the Parsivel^2.

The wind rose (see Figure 10) shows that all cases with higher catch ratios, representing events with wind speed between 2 and 4 m/s (as discussed above), are related to southerly and, to a lesser extent, easterly wind. The events with southerly wind (main prevailing wind direction of the site) correspond to cases where the SUT is perpendicular to the flow. Impact on catch efficiency for a sensor parallel to main flow has not been assessed, since no such configuration was tested.

7.5. Assessment of Yes-No, No-Yes events

The assessment of YN, NY events completes the picture of the performance of the SUT. As it can be seen in Table 5 and Figure 11, when no threshold is applied to the SUT (Case 1), the Parsivel^2 shows a large amount of NY cases, but no YN cases were reported, the latter indicating the good agreement of the sensor to detect precipitation whenever the reference reported precipitation. When a threshold is applied to the SUT (Case 2) - here a threshold of 0.1 mm/30 min – the number of YN and NY events is close to 0. Therefore, applying a threshold enables to ensure the good agreement of the Parsivel^2 with the reference, in all cases (YN, NY, YY, NN), and prevents the user to deal with NY events of unknown origin (either real or artefacts).

Looking more closely at Case 1 and according to Table 5, 675 NY events occurred during the two seasons of the experiment. It is useful to remind that these events are characterized, over the 30 minutes of an event, by an accumulation from the reference below the defined threshold of 0.1 mm, the reference precipitation detector having recorded 0 min of precipitation, and the SUT indicating more than 0 mm (see Section 4.1.1). Figure 11 shows that this high number of NY events appears to happen mainly for very low accumulation reported by the Parsivel^2 (0.05 mm/30 min and below), during events characterized by low relative humidity (around 50%) and temperature around and above 0°C.

Additional analysis, illustrated in Figure 13: Analysis of the 675 NY events from Case 1 in Table 5., shows that most of these NY events occurred at the end of the winter season (March/April),

identically in both seasons, during sunny mornings (08:00 to 14:00 local time), when there is an important positive temperature gradient and a negative relative humidity gradient. During these specific events, the reference shows weak signal around zero value (mainly noise), the reference precipitation detector confirms that there was no precipitation (corroborated by the increase of temperature and decrease of relative humidity), and the SUT indicates small accumulation.

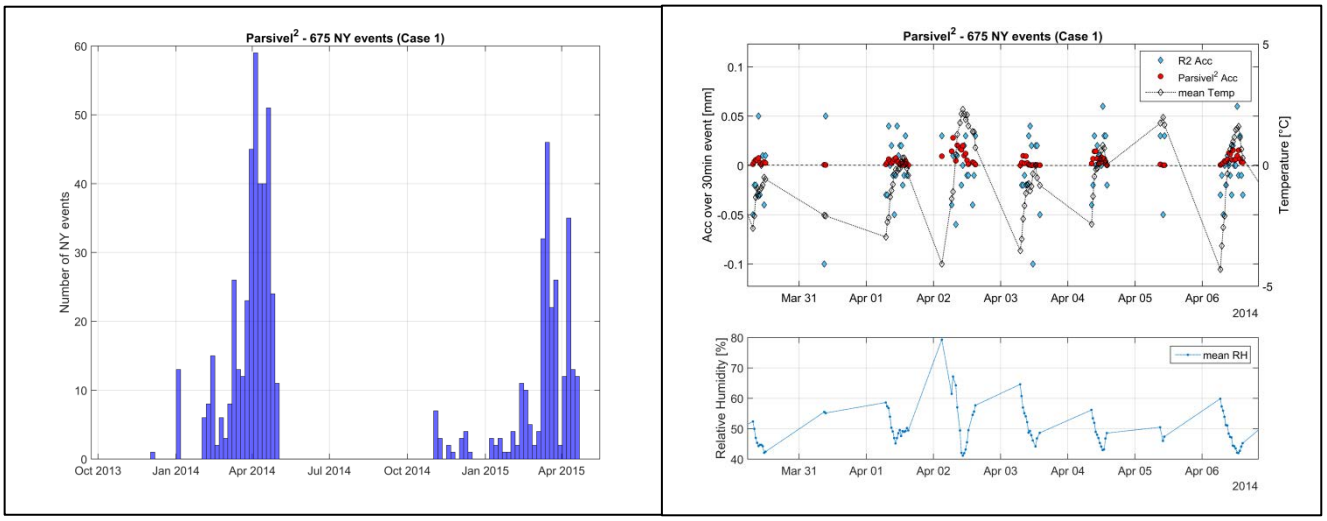


Figure 13: Analysis of the 675 NY events from Case 1 in Table 5. Left: distribution of the NY events over the seasons. Right: part of a time series representing (up) reference R2 and SUT 30 min accumulations together with mean temperature and (down) mean relative humidity, for the 675 NY events.

The real origin of these NY events remains unknown. These cases might be related to the high sensitivity of the SUT in detecting very light snow precipitation, but could also be due to a systematic measurement error (as the cyclic behavior visible for spring times in Figure 13 seems to show). According to the User Manual, intense sun (refraction effect) associated with wind, as well as vibration, associated with high wind, can cause false report of precipitation. It is unlikely that wind could be the source of error, though, as wind is very low at Sodankylä site. The sun, however, could be very low over the horizon during spring, and sunbeams could reach the receiver detector, causing disturbance in the received signal.

Apart from potential wrong measurements, it is known that the Parsivel^2, as other non-catchment type instruments, is more sensitive than the site reference. This is due to the different measurement physical principles of the instruments involved. Some of the NY events may thus also be an indicator of the better sensitivity of the Parsivel^2 to detect very light precipitation events that the reference couldn't detect.

7.6. Threshold selection

The threshold to be set for the Parsivel^2 in order to report precipitation adequately over a 30 min interval (3 STD, according to the methodology defined in Section 3.6.1.3.2 of the SPICE Final Report) corresponds to 0.009 mm (3 x 0.003 mm, from Table 7). Additionally, as shown in the contingency tables (Table 5), a threshold of 0.1 mm/30 min enables to get rid of events whose origin is not confirmed or well established.

8. Operational considerations

The Parsivel^2 disdrometer was considered having a reliable operation by the site manager. No major data breaks were recorded in the data flow during the SPICE campaign. It was easy to install and calibrate, making it easy to take into operation. The instrument's heating prevented the accumulation of snow on the instrument.

8.1. Maintenance

The Parsivel^2 requires minimum maintenance. The manufacturer recommends to clean the laser optics every six months, independently from the message of the instruments about the status of the optics. No issue were encountered concerning the maintenance of the Parsivel^2.

9. Performance considerations

- The firmware version V2.02.1 of the Parsivel^2 has an inaccurate 'Rain Amount accumulated' output (see section 6.3.3) that is to be considered (later version are not impacted, according to the manufacturer).
- When no threshold is applied to the SUT accumulation (Case 1), a high number of false alarms events (NY) were reported by the Parsivel^2, but it has to be noted that these are not well understood. Further investigations are recommended to be undertaken by the manufacturer and care must be taken by the users concerning the light snow precipitation reports of the Parsivel^2. However, when applying a threshold (of 0.1 mm/30 min for instance in Case 2) to the Parsivel^2 accumulation output, the FAR is significantly reduced.
- Overall, the Parsivel^2 reports overestimate solid precipitation accumulation and increasingly overestimates for wind speeds higher than 2 m/s for mixed and snow precipitation. To confirm this tendency and fully assess the dependency of the Parsivel^2 with wind speed, it is recommended to test the sensor in windier conditions (≥ 4 m/s) and conduct an assessment on the need of a transfer function to compensate this effect.
- High scatter of CE on an event basis (30 min), but good mean CE over the seasons, makes the Parsivel^2 more suitable for deriving solid precipitation accumulation over long periods of time (giving the condition that the sensor operates continuously), but makes it less reliable to derive solid precipitation accumulation over near real-time periods (as the 30 min events studied in this experiment).

WMO-SPICE Instrument Performance Report
Thies Clima Laser Precipitation Monitor

1. Technical specifications

Instrument model:	Thies Laser Precipitation Monitor (LPM) 5.4110.01.200 V2.50 STD
Measuring area:	40 - 47 cm^2 (instrument-specific)
Physical principle:	Optical laser based disdrometer enabling the acquisition of types of precipitation, intensity and spectrum based on the information of hydrometeors size and fall velocity.
Operating temperature range:	-40 to +70°C
Measurement uncertainty:	≤ 15 % (rain) (for wind speed < 3 m/s) ≤ 30 % (snow) (for wind speed < 3 m/s)
Sensitivity:	< 0.005 mm/h (for drizzle)

Note: Specifications from manufacturer provided documentation.

Figure 1: Thies LPM at Marshall (up) and at Weissfluhjoch (bottom) test sites. In both sites, the instrument was installed with a shield provided by the manufacturer.

2. Data output format

The disdrometer Thies LPM is a sensor outputting precipitation information, as intensity, accumulation or weather code, derived from the raw data. The raw data (particles size and fall velocity) are also available for the user. Table 1 summarizes the main output parameters from the instrument. The firmware version of the Thies LPM tested during this experiment was the version V2.5.

Table 1: Summary of main instrument outputs, as recorded by the sites during the experiment.

Measured Parameters	Units
1M Intensity (total precip)	[mm/h]
1M Intensity (liquid precip)	[mm/h]
1M Intensity (solid precip)	[mm/h]
Precipitation Amount	[mm]
1M SYNOP Tab. 4677	-
1M SYNOP Tab. 4680	-
1M METAR Tab. 4678	-
1M Visibility in precipitation	[m]
Ambient Temperature	[°C]
1M Radar Reflectivity	[dBZ]
1M Measuring quality	[%]
Housekeeping	-
# particles classified by precipitation type	[#/min]

This document reports on the ability of the Thies LPM to derive solid precipitation. The results are consequently computed using the 'Precipitation Amount' output.

3. SPICE test configuration

The Thies LPM, as sensor under test (SUT), has been tested on two different sites:

Test Sites: Weissfluhjoch (Switzerland); Marshall (USA)

Sensor Provider(s): All instruments evaluated were provided by the manufacturer (Thies Clima)

Figure 2: Map of SPICE sites testing Thies LPM instruments.

A summary on the configuration of instruments as tested, the duration of tests and availability reflected in these results, and the ancillary measurements used, by site, is available in Table 2, Table 3, and Table 4, respectively.

Table 2: Summary of instrument configurations and data output, by site. Details and photos on individual site configurations are available in the respective site commissioning protocols.

	Weissfluhjoch	Marshall
Main prevailing wind directions	SSE and NNW	btw N and E (during pcp)
Sensor orientation	SW	NNW
Height of installation	5 m	3 m
Heating	Yes, as recommended	
Shield	Wind Protection Element model 5.4200.00.000	
Data QC	SPICE QC methodology	
Data temporal resolution	1 min	
Processing interval for SPICE data analysis	30 min	

Table 3: Data availability, by measurement season, by site.

Measurement season	Weissfluhjoch	Marshall
Season 1 (Oct. 2013 – Apr. 2014)	✓	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓	✓

Table 4: Summary of reference and ancillary measurements, by site, with measurement height.

	Weissfluhjoch	Marshall
R2 Site Reference	OTT Pluvio^2 1500mm (DFAR) (3.5 m, rim height)	Geonor 600 (DFAR) (3 m, rim height)
R2 Precip Detector	Thies LPM (DFAR) (3.5 m)	OTT Parsivel^2 (Site*) (2 m)
Ancillary Temp Sensor	Thygan VTP 6 (5 m)	MetOne, model 060A-2/062, 2144-L (2 m)
Ancillary RH Sensor	Thygan VTP 6 (5 m)	Campbell Scientific CS500 (3 m)
Ancillary Wind Sensor	RM Young Wind Monitor 05103 (5.5 m)	RM Young Wind Monitor 05103 (3 m)

**A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR-fence; however, in cases where a more sensitive detector is available outside of the DFIR-fence, or there are issues with the detector within the DFIR-fence, a precipitation detector elsewhere on the site can be employed. For Marshall site, the precipitation detector installed in the DFAR showed some deficiencies during the field experiment. Therefore, the SUT Thies LPM served as precipitation detector for the reference system and, hence, for the evaluation of all SUT in Marshall. It is obvious that it cannot be used for the evaluation of the Thies LPM itself. For that purpose, a Parsivel^2, installed and managed by the site (i.e. not a SUT from SPICE), was used.*

4. Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6 of the SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute intervals:

- Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip;
- SUT 'Yes': SUT accumulation > 0 mm;
- SUT 'No': SUT accumulation = 0 mm.

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores.

4.1.2. Skill score assessment

The ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

The above scores are computed using the formulations provided in Section 3.6 of the SPICE Final Report.

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

5. Environmental conditions

The environmental conditions over the entire duration of the test period, at each site, are expressed as the probability density functions (PDFs) for air temperature, relative humidity, wind speed, wind direction, and precipitation rate in Figure 3. Figure 4 presents the same parameters during the precipitation events reported by the corresponding site reference, R2. The precipitation percentage in Figure 3 and Figure 4 represents the number of minutes of precipitation over a standard 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration at each site.

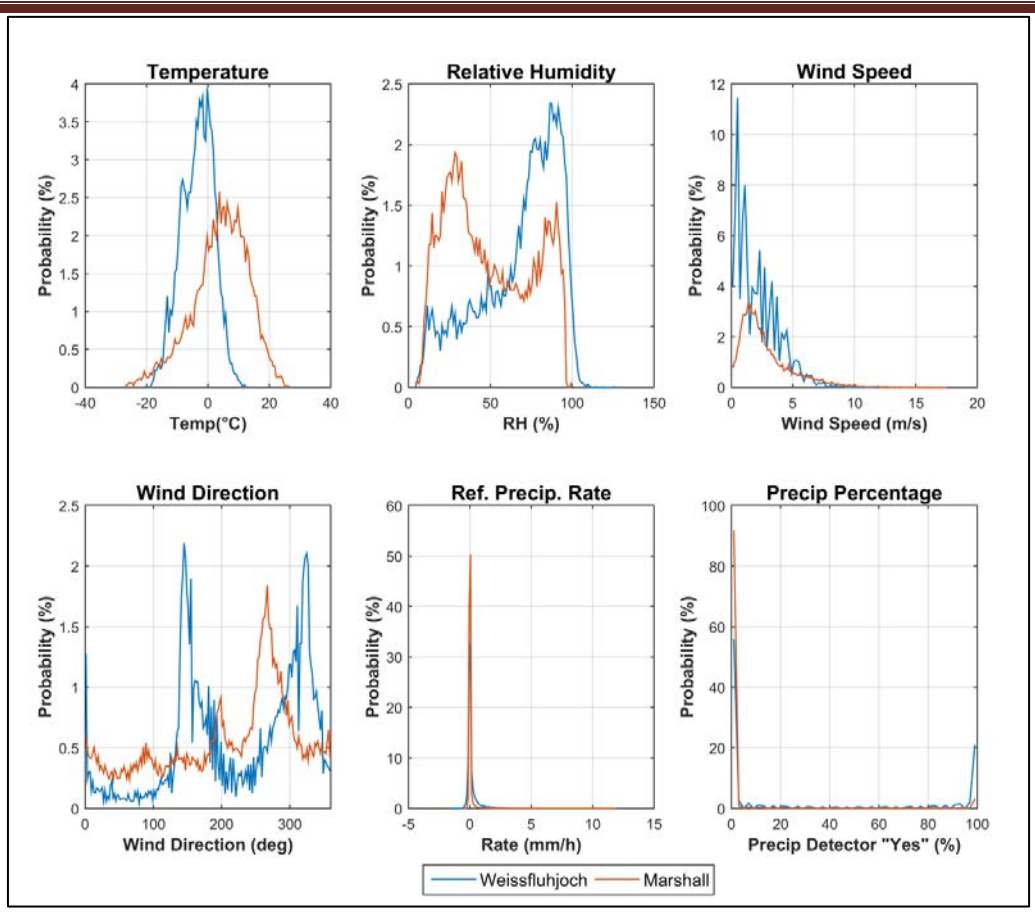


Figure 3: Summary of aggregated environmental conditions on the SPICE sites that operated Thies LPM, over the entire duration of tests, as per Table 3, above.

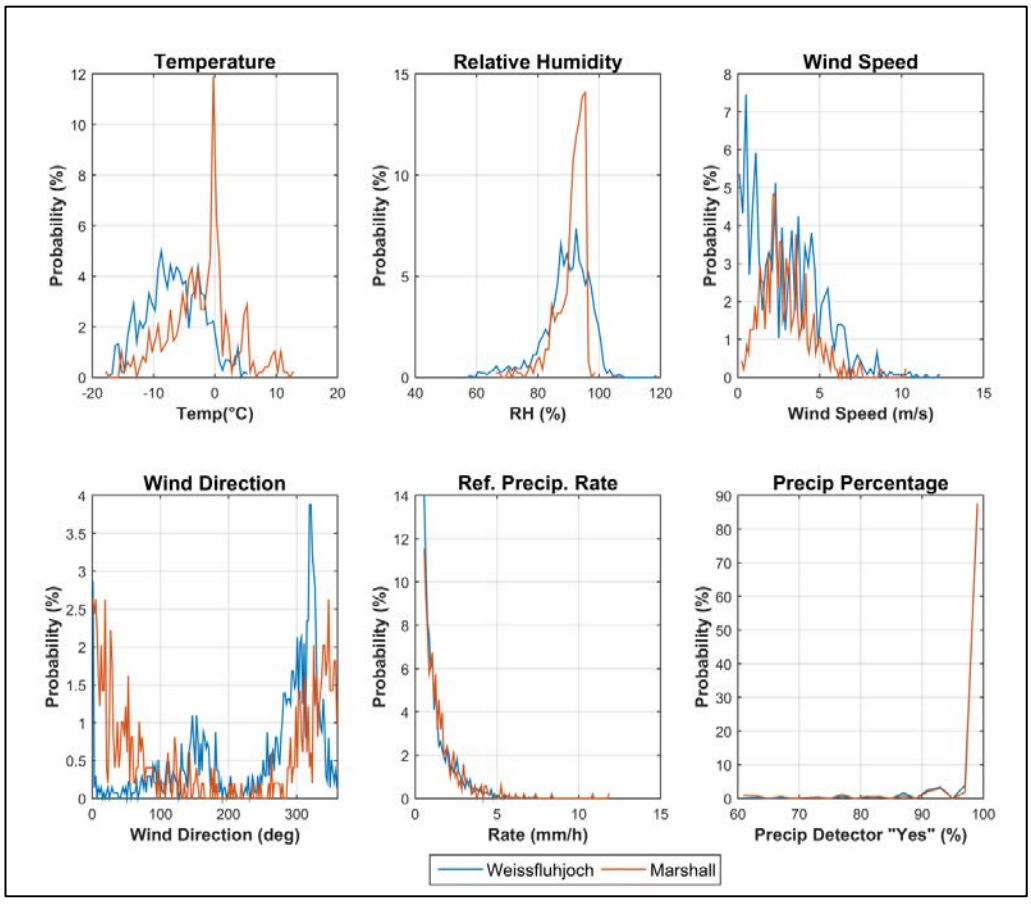


Figure 4: Summary of the aggregated environmental conditions on the SPICE sites that tested Thies LPM, corresponding to precipitation events, as reported by the site R2, reference, during the tests, as per Table 3 above.

6. Evaluation of performance over the range of operating conditions

6.1. Skill score assessment

The ability of the SUT to represent precipitation similarly to the site field reference R2, is assessed using contingency tables (Section 4.1.1) and derived selected skill scores (Section 4.1.2). To better understand the potential influence of threshold choices on the derived results, two cases are considered here (see note below). The contingency results related to these two cases are given in Table 5 and the respective skill scores in Table 6, for both testing sites.

Note: Following the data derivation explained in Section 4.1.1, the conditions required to have a 'Yes' or a 'No' event over the 30 min interval, for the reference and the SUT, for the two different cases treated here, are:

CASE 1 *(as defined in Section 4.1.1)*:

- *Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip*
- *Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip*
- *SUT 'Yes': SUT accumulation > 0 mm*
- *SUT 'No': SUT accumulation = 0 mm*

CASE 2:

- *Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip*
- *Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip*
- *SUT 'Yes': SUT accumulation ≥ 0.1 mm*
- *SUT 'No': SUT accumulation < 0.1 mm*

Results presented in this report are based on Case 1.

Table 5: Contingency Tables: detection of precipitation of the Thies LPM relative to the specific site reference, expressed as number of events over the entire test period, by site. The skill scores associated with these events are given in Table 6.

Weissfluhjoch Ref Plv2DFAR **Case 1**

SUT ThiesLPM	Yes	No	Total
Yes	1472	3	1475
No	13	7700	7713
Total	1485	7703	9188

Marshall Ref Geo600DFAR **Case 1**

SUT ThiesLPM	Yes	No	Total
Yes	504	384	888
No	0	17804	17804
Total	504	18188	18692

Weissfluhjoch Ref Plv2DFAR **Case 2**

SUT ThiesLPM	Yes	No	Total
Yes	1210	0	1210
No	275	7703	7978
Total	1485	7703	9188

Marshall Ref Geo600DFAR **Case 2**

SUT ThiesLPM	Yes	No	Total
Yes	471	2	473
No	33	18186	18219
Total	504	18188	18692

Table 6: Skill Scores for the Thies LPM, by site. POD: Probability Of Detection, FAR: False Alarm Rate, B: Bias, HSS: Heidke Skill Score (see Section 4.1.2 for more details).

Thies LPM, Skill Scores				
	POD	**FAR**	**B**	**HSS**
Case 1				
Weissfluhjoch	99.1%	0.203%	99.3%	99.4%
Marshall	100%	43.2%	176%	71.4%
Case 2				
Weissfluhjoch	81.5%	0%	81.5%	88.1%
Marshall	93.5%	0.423%	93.8%	96.3%

6.2. Assessment of SUT performance during non-precipitating events

The performance of the SUT in the absence of precipitation (when the reference precipitation detector recorded 30 minutes without precipitation) is represented in Figure 5 and Table 7, reflecting the distribution of the sensor response, as measured during the interval.

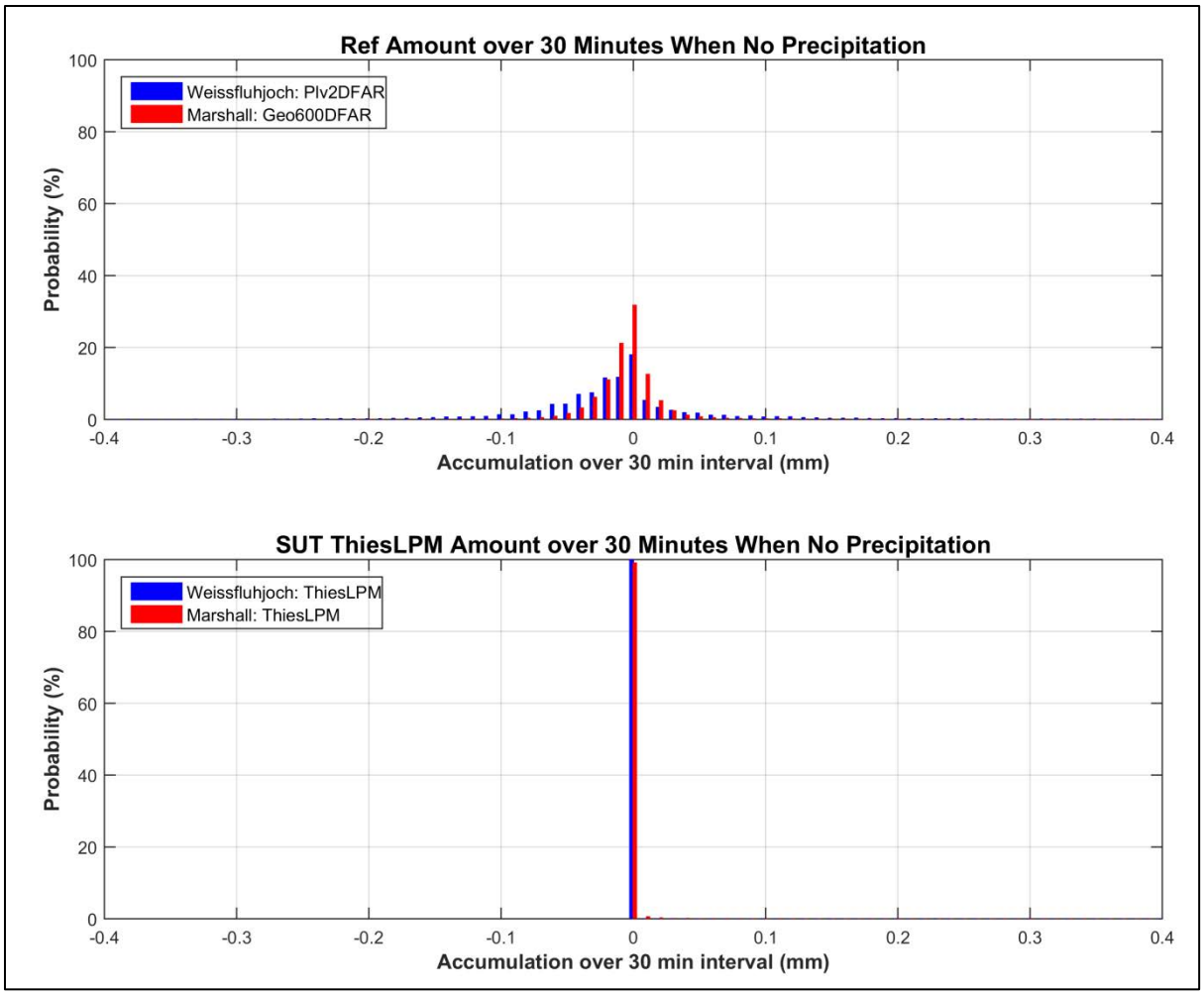


Figure 5: Probability of occurrence of a response during a 30 min interval in the absence of precipitation, represented by the signal output from (top) the R2 reference and (bottom) the SUT Thies LPM, by site. Statistics associated with these graphs are given in Table 7.

Table 7: Reference and SUT statistics of response signal when no precipitation was occurring, as plotted in Figure 5; Average (Avg), standard deviation (STD), maximum (Max) and minimum (Min) of the response signal, together with the number of events (Num) over the test period is given by site.

No Precip Statistics - Reference: Plv2DFAR (Weissfluhjoch) and Geo600 DFAR (Marshall)

	Ref Avg [mm]	Ref STD [mm]	Ref Max [mm]	Ref Min [mm]	Ref Num
Weissfluhjoch	-0.007	0.074	1.740	-0.750	8177
Marshall	0.000	0.029	1.876	-0.207	18271

No Precip Statistics - SUT: ThiesLPM

	SUT Avg [mm]	SUT STD [mm]	SUT Max [mm]	SUT Min [mm]	SUT Num
Weissfluhjoch	0.001	0.045	3.440	0.000	8177
Marshall	0.001	0.026	3.070	0.000	18271

6.3. Ability of the SUT to measure precipitation

6.3.1. Yes-Yes cases

Quantitatively, the performance of the SUT to derive and report precipitation is assessed relative to the site reference in several graphs and tables illustrated in this section, using only the cases where both instruments reported precipitation over the 30 min interval, according to the criteria used in Case 1 of Table 5 (cases 'Yes-Yes', or shorter 'YY').

6.3.1.1. Time series plots

The time series (cumulative sum of 30 min YY events accumulation) of each individual SUT is plotted against their corresponding reference for the two seasons, by precipitation type (see Section 4.1.4) and for each site in Figure 6.

The corresponding seasonal accumulations are given in Table 8.

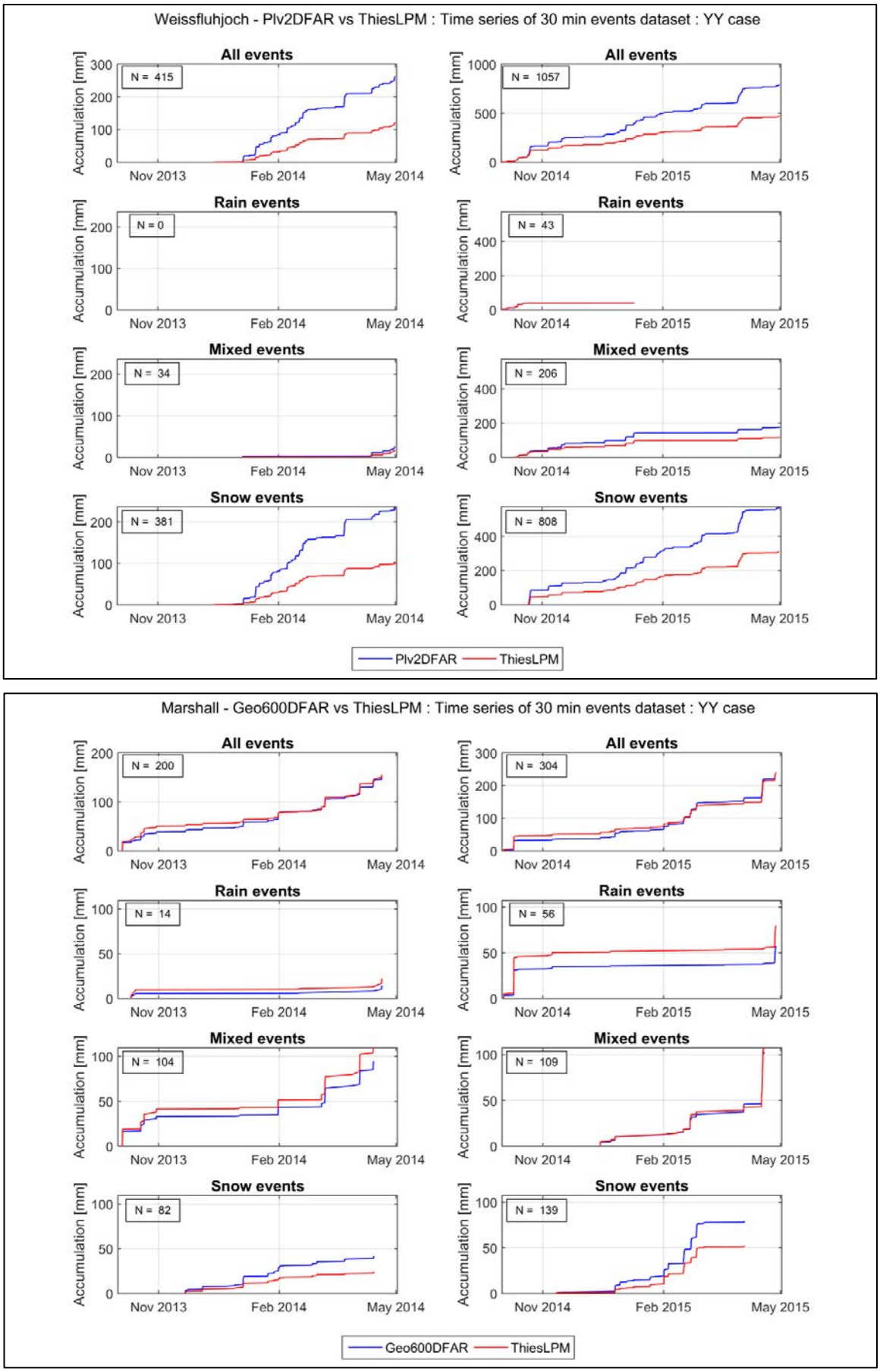


Figure 6: Time series based on 30 min YY events of the SUT Thies LPM against the corresponding site reference, discriminated by precipitation type (Rain, Mixed, Snow), for both seasons (2013/14 on the left, 2014/15 on the right), for (top) Weissfluhjoch and (bottom) Marshall test sites.

Table 8: Seasonal accumulation [mm] for (a) Weissfluhjoch and (b) Marshall test sites based on the sum of YY events from the SUT Thies LPM and the corresponding site field references R2: (a) Ref Plv2 DFAR, (b) Ref Geo600 DFAR.

(a) Weissfluhjoch

	Season 2013/14		Season 2014/15	
[mm]	Ref	SUT	Ref	SUT
All events	261.48	121.05	790.68	466.32
Rain events	0.00	0.41	40.97	40.25
Mixed events	25.20	17.55	177.17	117.20
Snow events	236.28	103.50	572.54	308.87

(b) Marshall

	Season 2013/14		Season 2014/15	
[mm]	Ref	SUT	Ref	SUT
All events	149.31	154.75	238.40	238.61
Rain events	13.95	22.00	57.02	79.48
Mixed events	93.96	109.38	102.42	107.26
Snow events	41.39	23.37	78.97	51.87

6.3.1.2. Scatter plots and RMSE values

Scatter plots of the amount derived by the SUT versus the corresponding reference amount, and discriminated by precipitation type, is given for both sites in Figure 7.

Quantitatively, the SUT performance is assessed using the Root Mean Square Error (RMSE) also known in practice as Operational Comparability. The results are available for both sites in Table 9.

Table 9: Root Mean Square Error (RMSE) statistics, in mm, for the SUT Thies LPM with respect to the corresponding site reference, by precipitation type and by site, including both seasons data.

RMSE [mm]	All	Rain	Mixed	Snow
Weissfluhjoch	0.483	0.248	0.505	0.486
Marshall	0.488	0.767	0.526	0.305

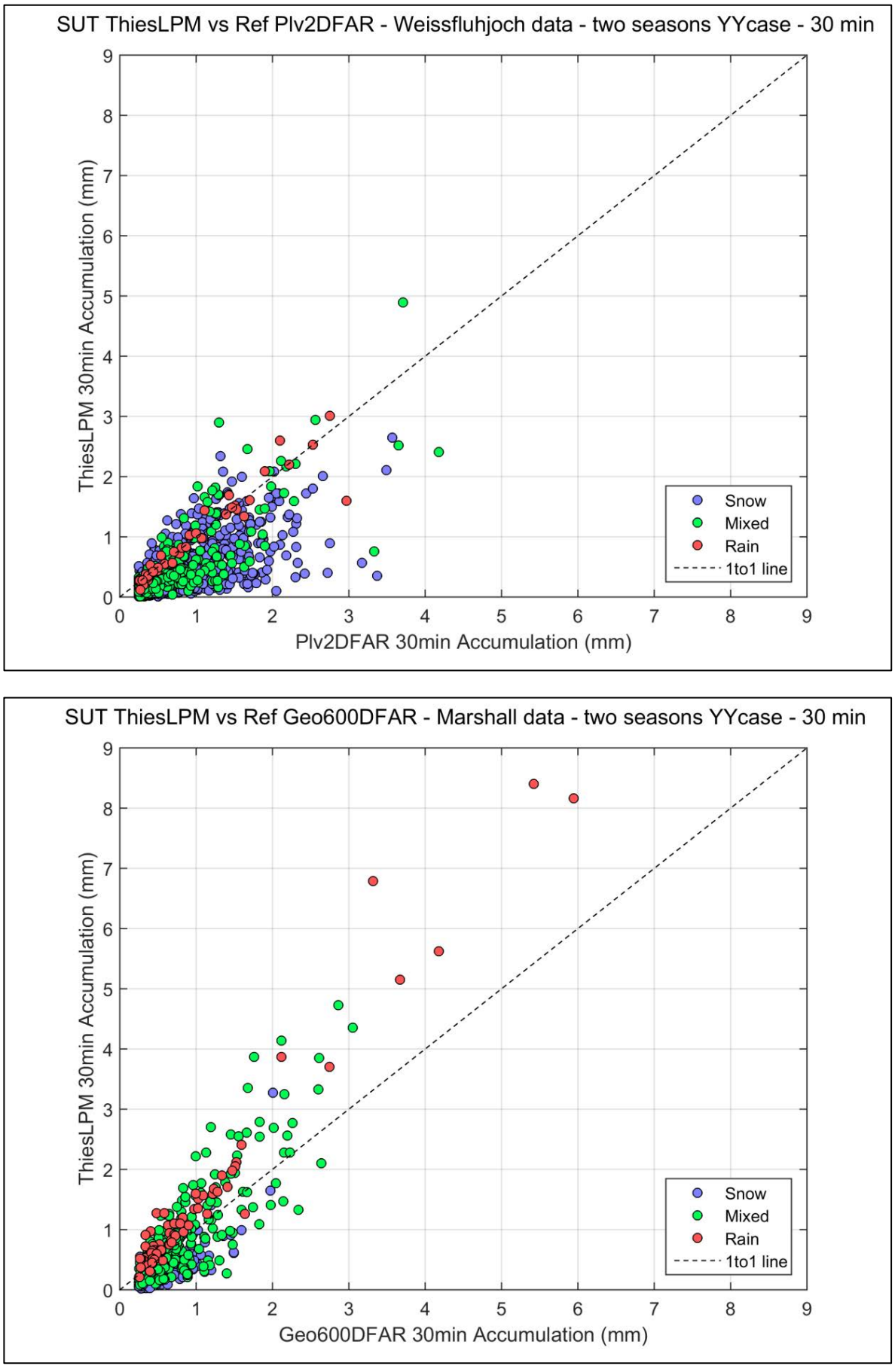


Figure 7: Scatter plots based on 30 min YY events accumulation from the SUT Thies LPM against the corresponding site reference, for the two seasons, discriminated by precipitation type, for (top) Weissfluhjoch and (bottom) Marshall test sites.

6.3.1.3. Catch efficiency evaluation by precipitation type

The Catch Efficiency (CE) of the SUT is represented by histograms (Figure 8) and boxplots (Figure 9), both discriminated by precipitation type, and representing both sites.

The quantitative evaluation of the CE is provided in Table 10. The mean catch efficiency is given in the first line of this table for each site, considering both seasons data and for each category of precipitation type as well as for all the events together.

Note: All events with a CE greater than 3, if any, are included in one category named '3 and more' or '≥3' in the upcoming graphs. Additionally, for all graphs representing the CE, a dashed black line is added at CE = 1, which represents the ideal case where the SUT reports exactly the same precipitation amount as the reference.

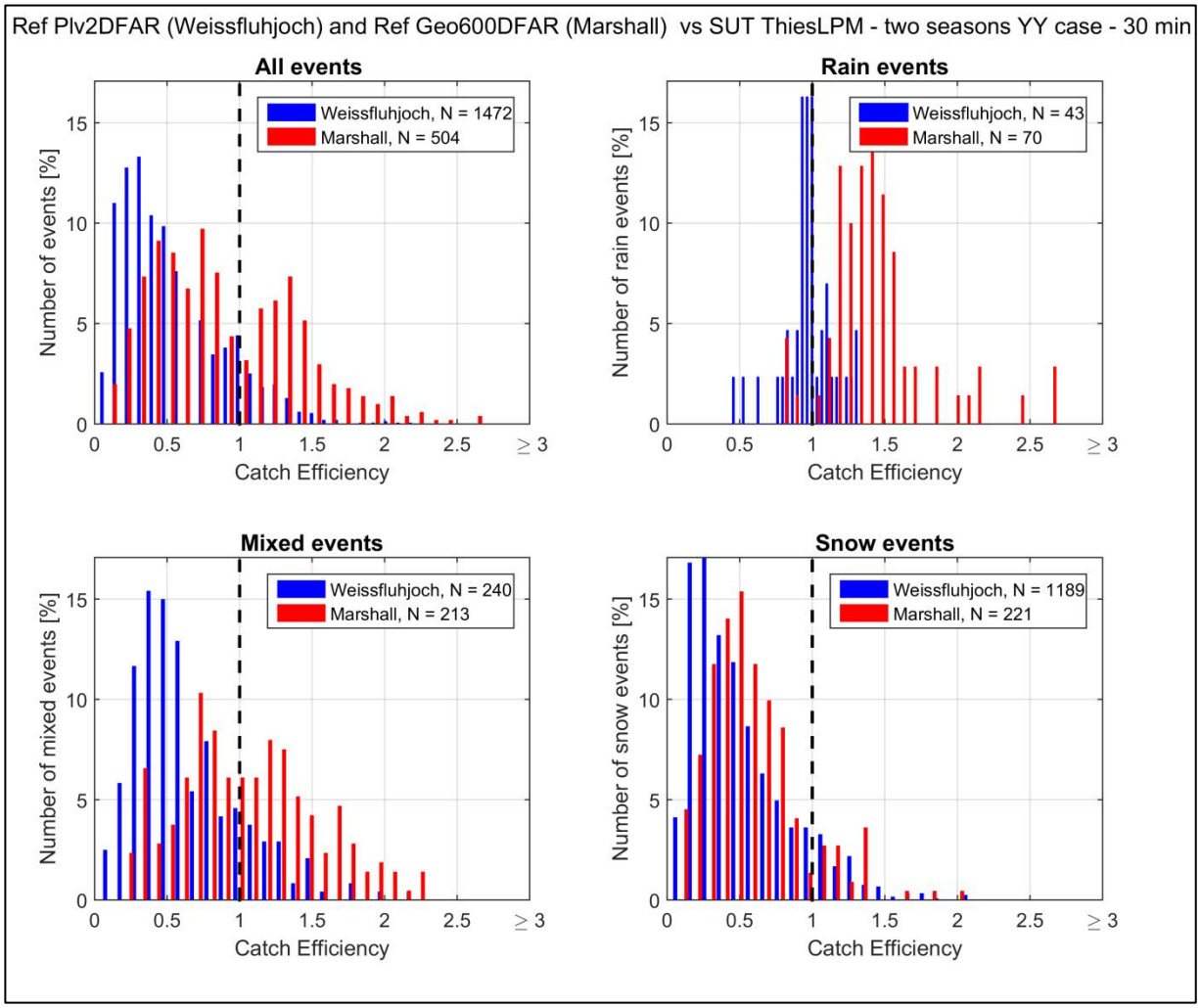


Figure 8: Histograms based on 30 min YY events for the two seasons, representing the distribution of the catch efficiency of the SUT Thies LPM against the corresponding site reference (SUT/Ref), discriminated by precipitation type, with number of events given in the legend for each category, and represented for both sites. The dashed black line at CE = 1 represents the ideal case.

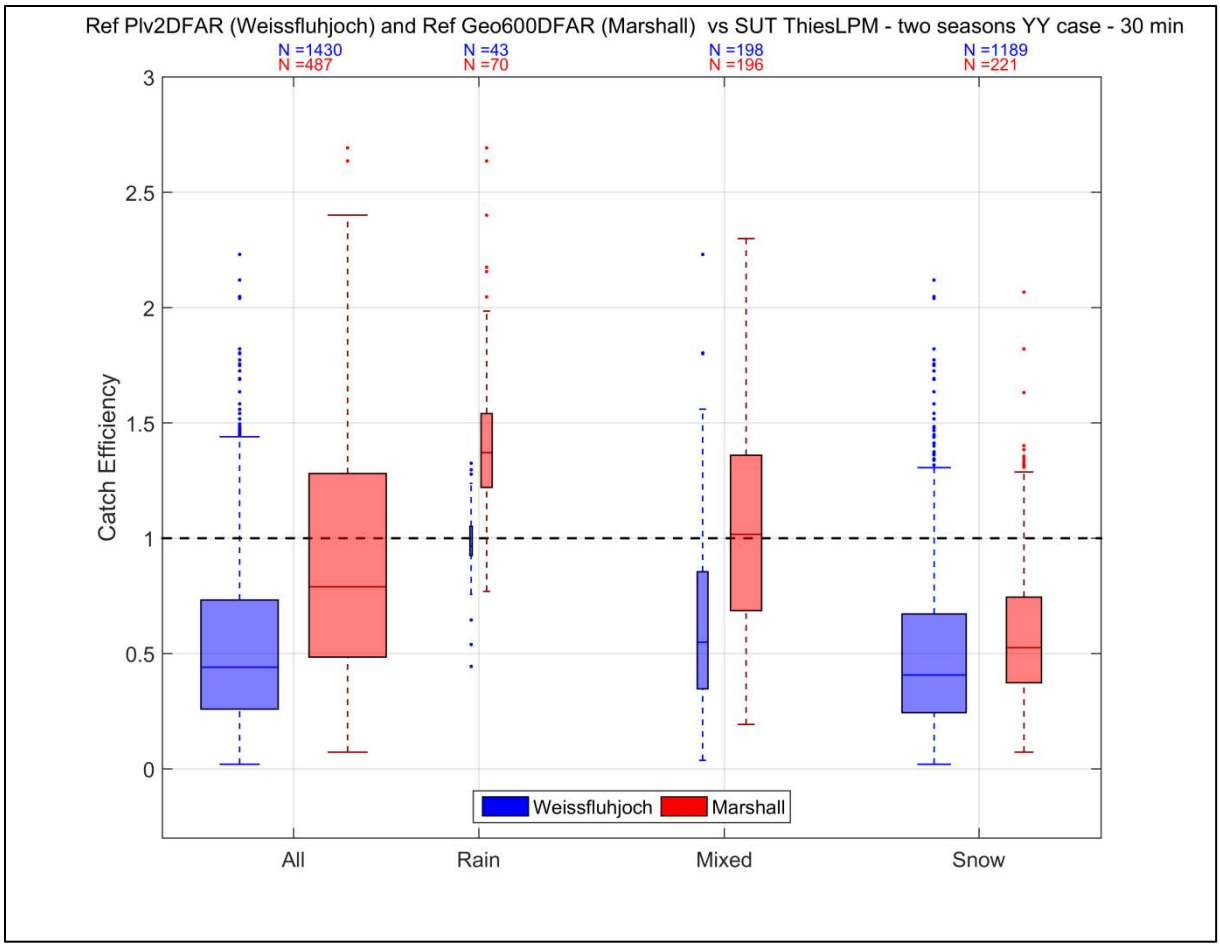


Figure 9: Boxplot based on 30 min YY events of the two seasons, representing the catch efficiency of the SUT Thies LPM against the corresponding site reference (SUT/Ref), discriminated by precipitation type, with number of events given at the top of each category, and represented for both sites. The width of the boxes is proportional to the percentage of events in each category ('All' boxes represent 100% data for each site, making them having the same width despite the difference in number of events). The dashed black line at CE = 1 represents the ideal case.

Table 100: Statistics related to the CE of SUT Thies LPM against the corresponding site reference (SUT/Ref, see Figure 9), discriminated by precipitation type.

Catch Efficiency Statistics: **Weissfluhjoch**

CE Boxplot Parameters	All	Rain	Mixed	Snow
Mean	0.54	0.97	0.65	0.50
Median	0.44	0.97	0.55	0.41
75 percentile	0.73	1.05	0.85	0.67
25 percentile	0.26	0.93	0.35	0.24
Upper Whisker	1.44	1.24	1.56	1.31
Lower Whisker	0.02	0.76	0.04	0.02
Maximum	2.23	1.32	2.23	2.12
Minimum	0.02	0.44	0.04	0.02
# Outliers	29	6	3	32
# Outliers ≥ 3	0	0	0	0
# Events	1430	43	198	1189

Catch Efficiency Statistics: **Marshall**

CE Boxplot Parameters	All	Rain	Mixed	Snow
Mean	0.90	1.43	1.06	0.59
Median	0.79	1.37	1.02	0.53
75 percentile	1.28	1.54	1.36	0.74
25 percentile	0.48	1.22	0.69	0.37
Upper Whisker	2.40	1.98	2.30	1.29
Lower Whisker	0.07	0.77	0.19	0.07
Maximum	2.69	2.69	2.30	2.07
Minimum	0.07	0.77	0.19	0.07
# Outliers	2	6	1	11
# Outliers ≥ 3	0	0	0	0
# Events	487	70	196	221

6.3.1.4. Catch efficiency dependency on wind speed

The variation of the SUT catch efficiency (SUT/Ref) with wind speed is illustrated in a scatter plot in Figure 10, together with a boxplot in Figure 11, both discriminated by precipitation type and for each site.

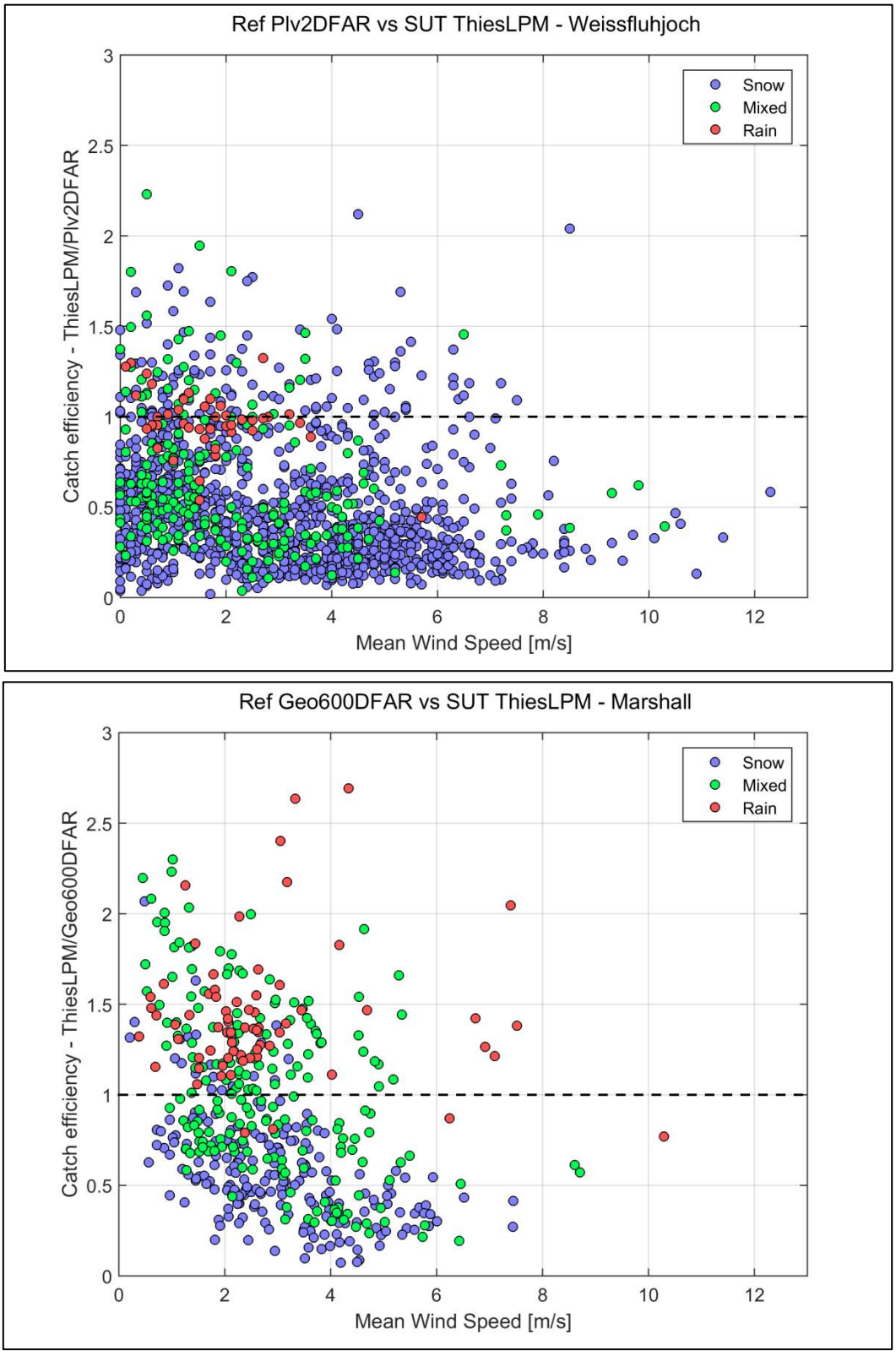


Figure 10: Scatter plots based on 30 min YY events for the two seasons, representing the catch efficiency of the SUT Thies LPM with respect to the corresponding site reference (SUT/Ref), against wind speed and discriminated by precipitation type, for (top) Weissfluhjoch and (bottom) Marshall test sites. The dashed black line at CE = 1 represents the ideal case.

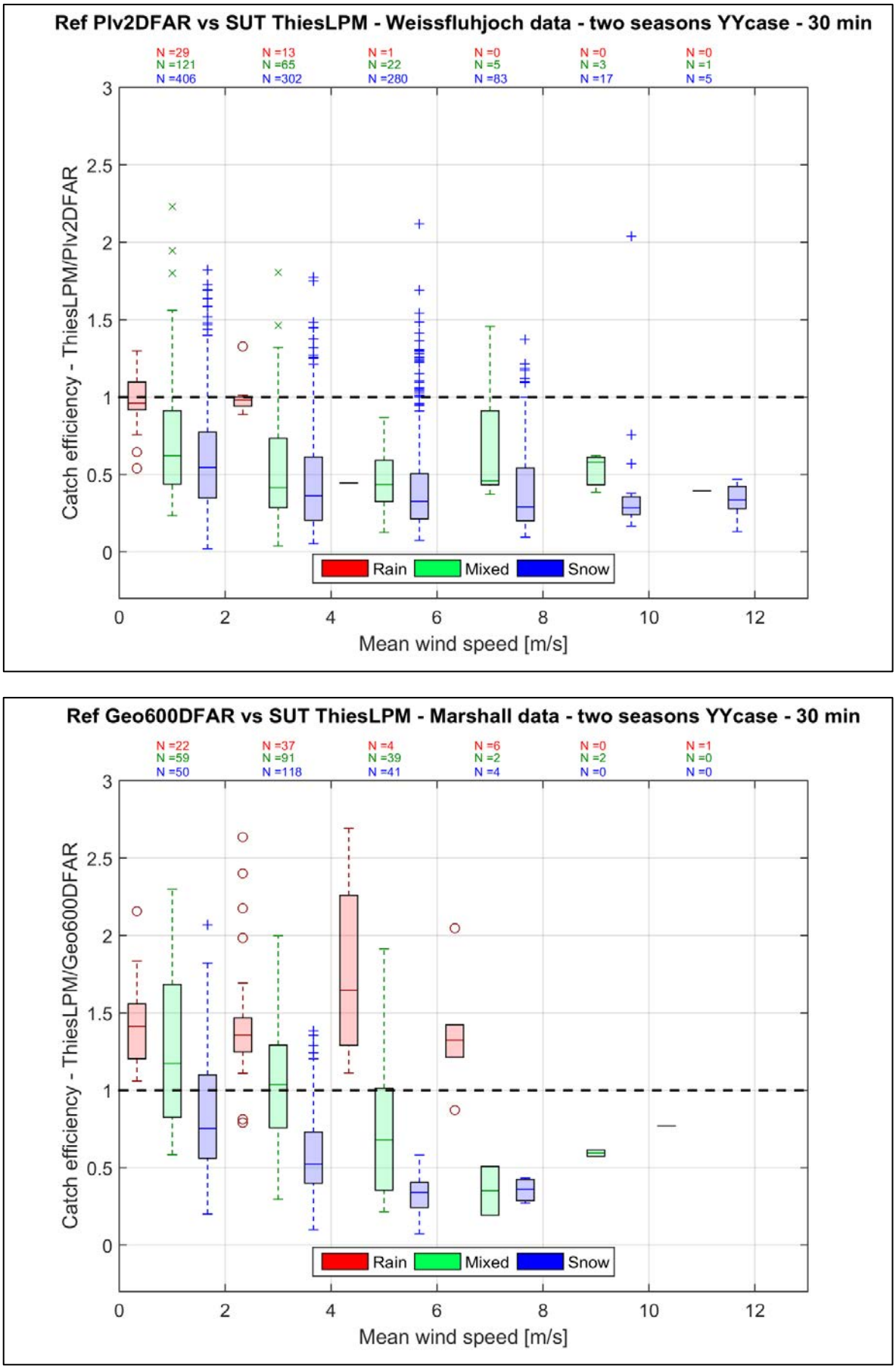


Figure 11: Boxplots based on 30 min YY events for the two seasons, representing the catch efficiency of SUT Thies LPM with respect to the corresponding site reference (SUT/Ref), against wind speed and discriminated by precipitation type for (top) Weissfluhjoch and (bottom) Marshall test sites. The dashed black line at CE = 1 represents the ideal case.

6.3.1.5. Catch efficiency dependency on wind direction

In order to assess the dependency of the CE with wind direction, a wind rose is produced (Figure 12) representing, for each site, the wind data of the two seasons, binned by catch efficiency in order to represent undercatch (CE < 0.8), overcatch (CE > 1.2) and catch efficiency of 1 ± 20% of the SUT.

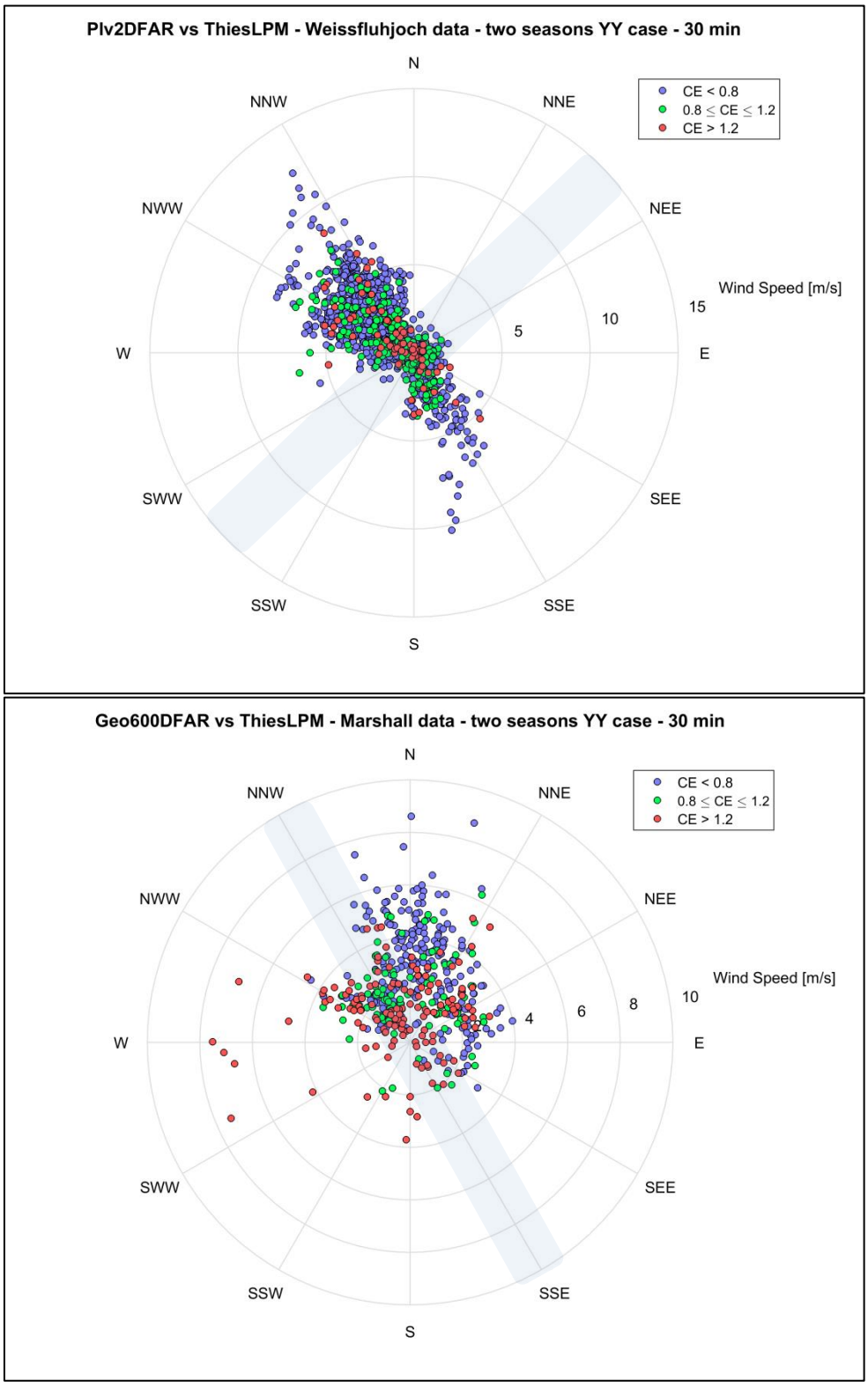


Figure 12: Precipitation events (YY cases) as function of wind speed and direction of (top) Weissfluhjoch and (bottom) Marshall test sites, binned by catch efficiency (CE). The grey zone indicates the SUT orientation.

6.3.2. Yes-No and No-Yes cases

Events when the site reference and SUT do not agree on the occurrence of precipitation includes two categories of cases (Section 4.1.1): (1) when the field reference reported a precipitation event, while the SUT did not (Yes-No cases, 'YN'), and (2) when the field reference did not report a precipitation event while the SUT did (No-Yes cases, 'NY').

Histograms illustrating field reference and SUT reports and associated site conditions for all YN and NY cases of both sites during the test period are provided in Figure 13 and Figure 14, respectively.

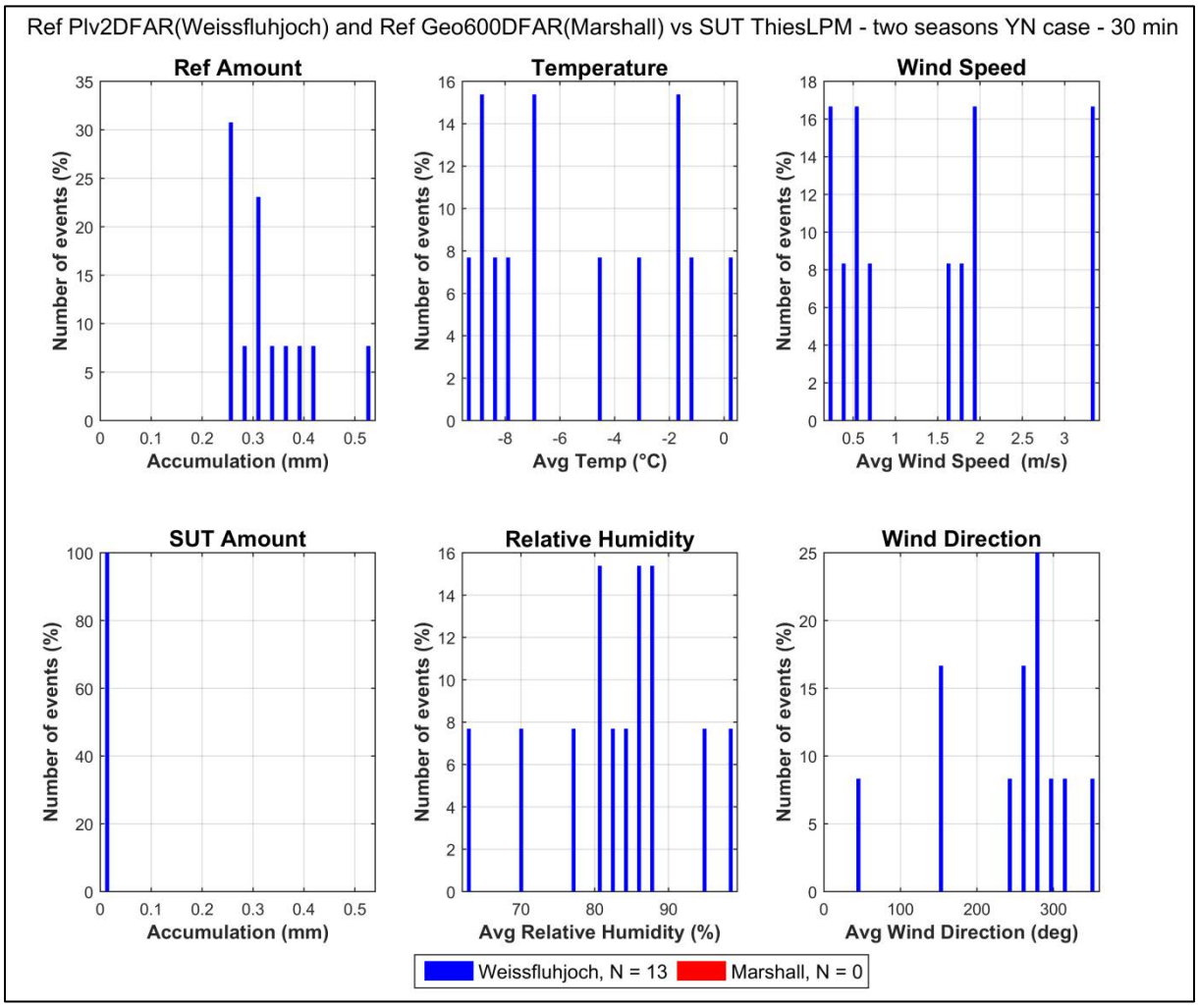


Figure 13: Histograms of SUT Thies LPM and corresponding site reference accumulations (left column), along with distributions of mean temperature, mean relative humidity, mean wind speed, and mean wind direction for all YN cases (number indicated in the legend) of the 30 min intervals, as reported by Case 1 of Table 5, at the two different test sites.

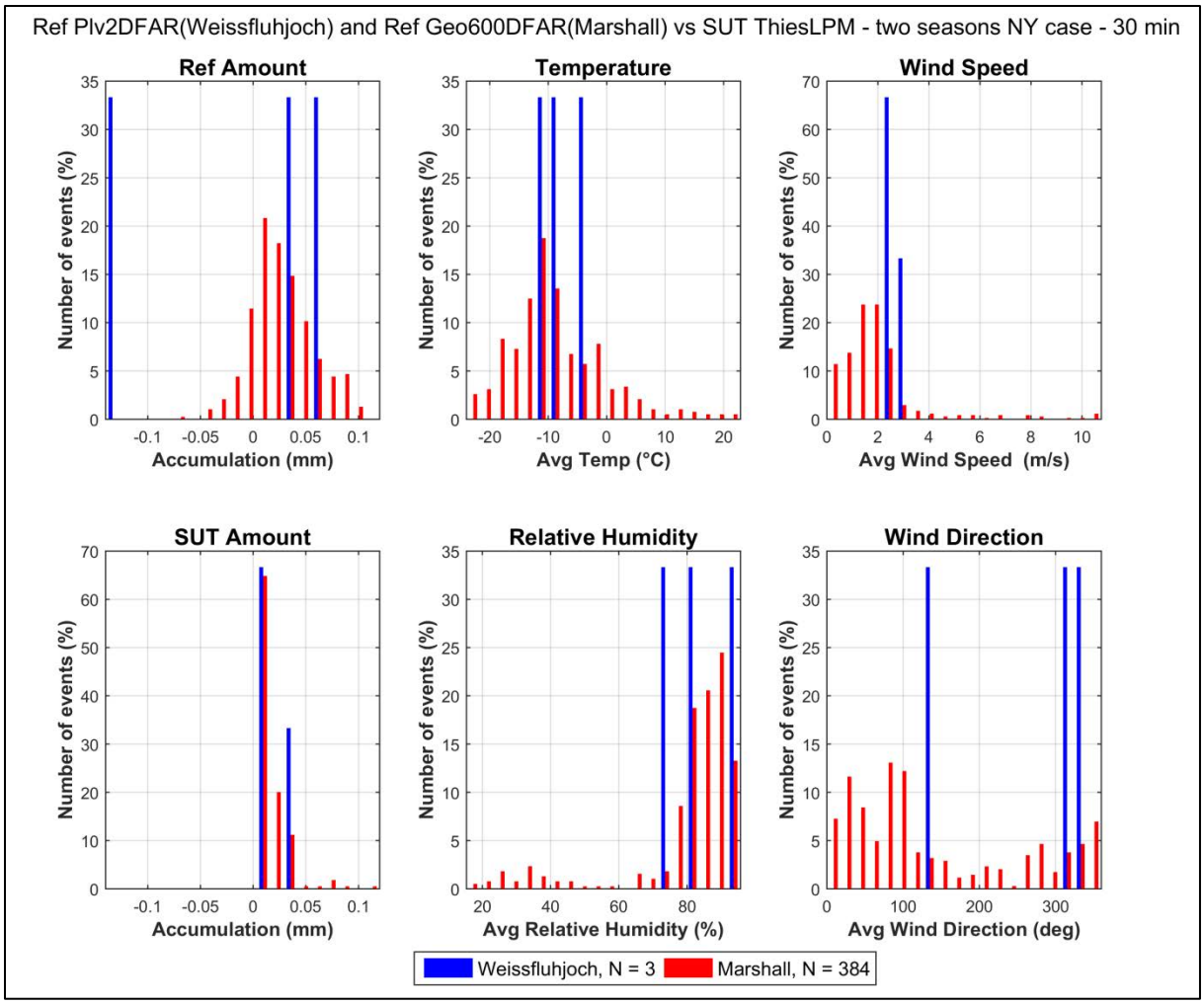


Figure 14: Histograms of SUT Thies LPM and corresponding site reference accumulations (left column), along with distributions of mean temperature, mean relative humidity, mean wind speed, and mean wind direction for all NY cases (number indicated in the legend) of the 30 min intervals, as reported by Case 1 of Table 5, at the two different test sites.

7. Interpretation of results

7.1. Operating conditions

The two sites hosting the Thies LPM account for different climates, Weissfluhjoch being characterized by an alpine regime, while Marshall is led by a continental climate. These differences can be seen in Figure 3 and Figure 4, especially in terms of temperature, with cooler conditions for Weissfluhjoch and most of the precipitation events occurring at temperature around -5°C, against 0°C for Marshall. The wind regime is also different, with higher wind speed at Weissfluhjoch (up to 10 m/s, against around 5 m/s for Marshall), most of the events occurring at wind speeds up to 7 m/s for both sites.

7.2. Reliability in detecting precipitation

The Thies LPM was efficient in detecting precipitation when the reference reported precipitation, as presented in Case 1 of Table 6, when no threshold is applied to the SUT, with a high POD score of 99% and above for both sites. The Weissfluhjoch site shows low FAR (less than 1%), but Marshall shows a higher rate of 43% due to a large number of NY events characterized by very small intensities (see section 7.5 for more details).
When a threshold of 0.1 mm over 30 min events is applied to the SUT (Case 2), it can be noted that the POD of both sites get a bit deteriorated (see Section 7.5), but still keep high values (more than 81%), while the two FAR scores are reaching (or approaching) the best score of 0%, the value from Marshall having experienced the most significant reduction from 43% to less than 1%. Even if for Weissfluhjoch the scores seem to get generally slightly worse in Case 2, we can note that in this case, the two sites score results show more consistency across themselves. Applying a threshold reduces therefore the FAR percentage and gives more consistency on the skill scores results of the instrument on different sites.

7.3. Performance of SUT during no-precipitation events

The Thies LPM has a stable output signal, showing almost no noise during no-precipitation events (see Figure 5 and Table 7). Note that the no-precipitation study is performed using only the reference precipitation detector data (here a Thies LPM for Weissfluhjoch site and a Parsivel^2 for Marshall) with the criteria ensuring that 30 min of 'no-precipitation' was recorded. For these cases, the Thies LPM under test indicates 0 mm for more than 98% of the time for Marshall, and 99.9% of the time for Weissfluhjoch, indicating a good agreement with the reference precipitation detector on the absence of precipitation. Most of the remaining cases (where the reference precipitation detector doesn't detect any precipitation while the SUT does) correspond to very low accumulations reported by the SUT (less than 0.1 mm/30 min) and can be related to the Marshall NY events mentioned above and analyzed in Section 7.5.

7.4. Performance of SUT during precipitation events

When both the reference and the Thies LPM report precipitation (YY cases), the overall results show a different behavior of the SUT according to precipitation type, described in the following points. Figure 6 shows a clear undercatch of the Thies LPM over the time series under snow conditions, which has a repercussion on the "All events" category, proportionally to the number of snow events

involved (higher proportion of snow events in Weissfluhjoch than in Marshall, leading to an 'All events' category with a larger undercatch). Figure 8 depicts nicely the bimodal feature of the CE under the 'All events' category (especially true for Marshall), being the aggregation of the two rain and snow modes characterized by overcatch and undercatch, respectively.

Quantitatively, the CE statistics indicate a similar performance of the instruments under test, on the two sites, to measure solid precipitation, with an average catch ratio of 0.50 for Weissfluhjoch and 0.59 for Marshall (see Table 10 and Figure 9), showing a clear and consistent undercatch of the instrument against the site reference for the measurement of snow. The performance for rain in terms of catch efficiency is closer to 1, with a catch ratio of 0.97 for Weissfluhjoch. Low number of rain events prevent to draw a robust conclusion, but these results generally confirm what has been found in previous studies for liquid precipitation (e.g. Lanzinger et al., 2006). For Marshall, the catch ratio for rain is 1.43. From Figure 7 it can be seen that this overcatch is mainly due to a certain number of rain and mixed events with high intensities (> 3 mm/30 min). Again, the small number of events involved prevent to make a thorough conclusion, but the fact that these events seem to follow a line with a slope greater than the 1to1 line, may indicate a calibration or internal calculation issue for the sensor at the Marshall site.

The RMSE reflects the variability of measurements from the SUT against the site reference, by precipitation type (see Table 9). It was found to be double as high for snow than rain at Weissfluhjoch (0.48 mm and 0.24 mm, respectively) reflecting the higher scatter for snow events on Figure 7. The high RMSE for rain at Marshall reflects events with a constant relative offset shown in Figure 7 and already mentioned above. With exception of this issue, the RMSE values are very close for the two sites, indicating similar and consistent behaviour of the SUT in two different environments.

The boxplots on Figure 9 present a similar CE behavior from both instruments according to precipitation types, but with the results from Weissfluhjoch being slightly shifted to lower CE. This difference may result from the difference in the site climatology, with Weissfluhojch site being colder and higher in altitude. Therefore an alpine site may result in less good performance of the Thies LPM than a continental site.

Figure 10 depicts the large scatter of the Thies LPM CE on the 30 min basis, the CE for snow and mixed precipitation varying from 0.1 to 1.5 for wind speeds up to 8 m/s, making the derivation of a transfer function difficult, if not impossible. This high scatter also leads to the conclusion that this sensor is not suitable to derive accurate precipitation amount under mixed and snow conditions over near real time periods as the 30 min intervals used in this study.

The results from both sites show globally a decrease of the catch ratio with increasing wind speed (for snow and mixed precipitation), with catch ratio for snow already at 0.50 (Weissfluhjoch) and 0.75 (Marshall) for winds less than 2 m/s (see Figure 11). As the Thies LPM sensors were shielded on both sites, the performance of the sensor related to wind without shield was not assessed during the SPICE experiment.

The wind roses (see Figure 12) show no clear dependency of the catch efficiency with wind direction (no asymmetrical distribution of colored points considering the prevailing wind direction), for both

sites. In both cases, the sensor was installed perpendicular to the prevailing wind direction during precipitation events. Impact on catch efficiency for a sensor parallel to main flow has not been assessed.

7.5. Assessment of Yes-No, No-Yes events

The assessment of YN, NY events completes the picture of the performance of the SUT. As it can be seen in Table 5, when no threshold is applied to the SUT (Case 1), on both sites, there are very few events (except for NY events from Marshall, explained further down) when the reference and the SUT do not agree on the occurrence of precipitation. When applying a threshold to the SUT (Case 2), the number of YN events from both sites increases (especially for Weissfluhjoch), but these are all events for which the Thies LPM recorded less than 0.1 mm (but more than 0 mm, except for the 13 events of Case 1) and the R2 reference more than 0.25 mm, with 60% of precipitating minutes recorded by the precipitation detector during the 30 min interval, which means that these YN events are cases of undercatch from the Thies LPM, but not real misses. On the other hand, looking at Case 2, the high number of NY events appearing for Marshall in Case 1 is drastically reduced. Therefore, applying a threshold prevents the user to deal with NY events of unknown origin (either real or artefacts).

Looking more closely at Case 1, only Weissfluhjoch shows some YN cases (when the reference indicated an event, and the SUT did not), but they cannot be related to any specific atmospheric conditions (see Figure 13). Also the number of events is very low (13), and prevents from any robust conclusion.

The NY cases from Case 1 (when the accumulation from the reference was below the defined threshold of 0.1 mm, the precipitation detector recorded 0 min of precipitation, and the SUT indicated more than 0 mm) seem to occur at very low temperatures (-10 to -20°C) and light winds (0 to 2 m/s), as shown in Figure 14. The influence of high relative humidity might also play a role.

With a more in-depth analysis on the NY events from Marshall (for Case 1), it was found that they all show very low intensities recorded by the Thies LPM (with a mean over all NY events of 0.017 mm over the 30 min interval), as Figure 14 shows, and are correlated with high relative humidity, leading to the thought that it might be actual real light precipitation events. Indeed, the different instrument sensitivity might play a role, as illustrated by the example of Figure 15, where some of the NY events of season 2 are shown over the accumulation curve of the R2 reference weighing gauge, together with the precipitation detection measured by the Parsivel^2 (the reference precipitation detector), a DRD11A and the Thies LPM under test. It can be noted that this light precipitation was well detected by the DRD11A and the Thies LPM, but nothing was measured by the Parsivel^2. This led to the idea that the NY results could change if the choice of the reference precipitation detector was different.

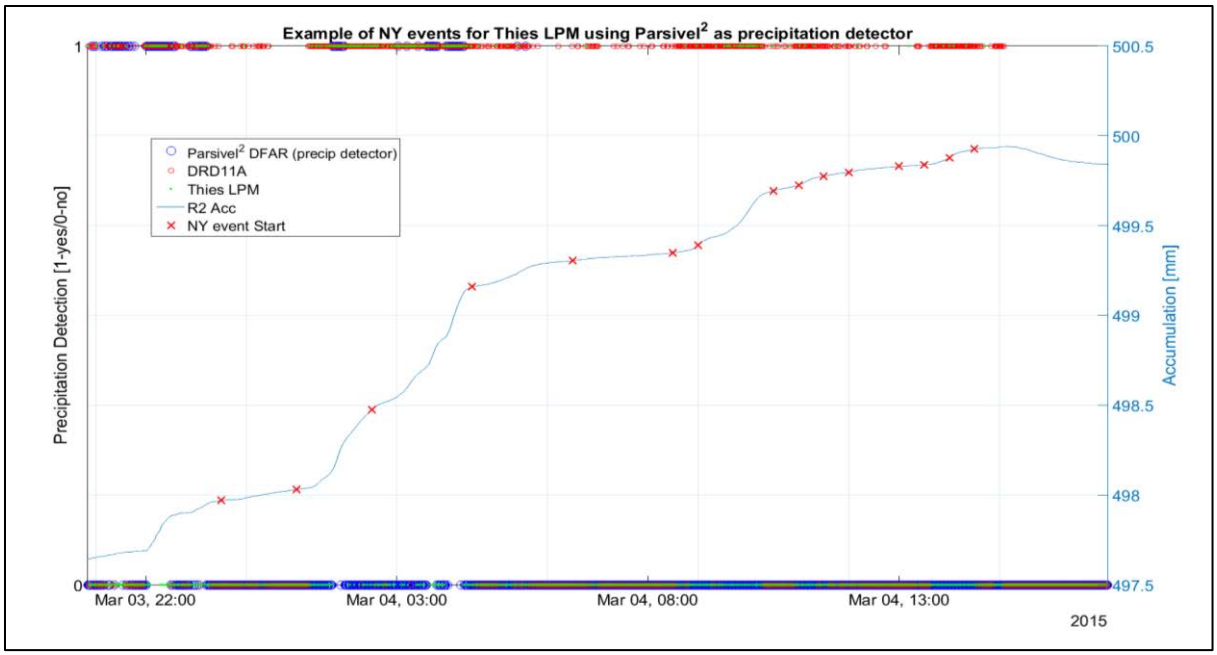


Figure 15: Example of NY events for the SUT Thies LPM using the Parsivel^2 as the reference precipitation detector during the second season.

To better address these differences in precipitation sensitivity, and to give a sense of the results dependency on the selected reference precipitation detector, a second study has been made using the DRD11A capacitive sensor for the precipitation detection as part of the R2 reference (instead of the Parsivel^2). This sensor was located inside the DFIR-fence, located about 25 meters from the Thies LPM. The same number of YY and YN events was found, but the number of NY events was significantly reduced (109 against 384 with the Parsivel^2), leading to a FAR of only 17.8% (against 43.2% with the Parsivel^2), a bias of 82.2% and an HSS of 89.9%. Even if capacitive precipitation detectors, such as the DRD11A, are less sensitive than laser-based precipitation detectors, like the Parsivel^2 (see Section 4.1.3.5.2 of the SPICE Final Report), the DRD11A seems to better match the R2 field reference and Thies LPM results. A similar graph to Figure 14, but using the DRD11A as the reference precipitation detector, is given in Figure 16 for Marshall NY events from Case 1. This figure shows wider distribution over the relative humidity range, as well as over the temperature range. Snow events are mainly related to high relative humidity, while rain events occurred under low RH conditions. It is still possible that some of these remaining 109 NY events are real light events that were not detected by the DRD11A (which measured 30 min of no precipitation for all of them), but this remains uncertain and no other parameters or photos are available to confirm or infirm this statement.

To conclude, the Thies LPM NY events of either study (with Parsivel^2 or DRD11A as reference precipitation detector) reflect therefore rather a reliable higher sensitivity of the Thies LPM compared to the reference system. Indeed, on one hand the reference precipitation detector could have a lower sensitivity than the Thies LPM, on the other hand the reference weighing gauge has a different physical principle of measurement implying an expected lower sensitivity than the Thies LPM.

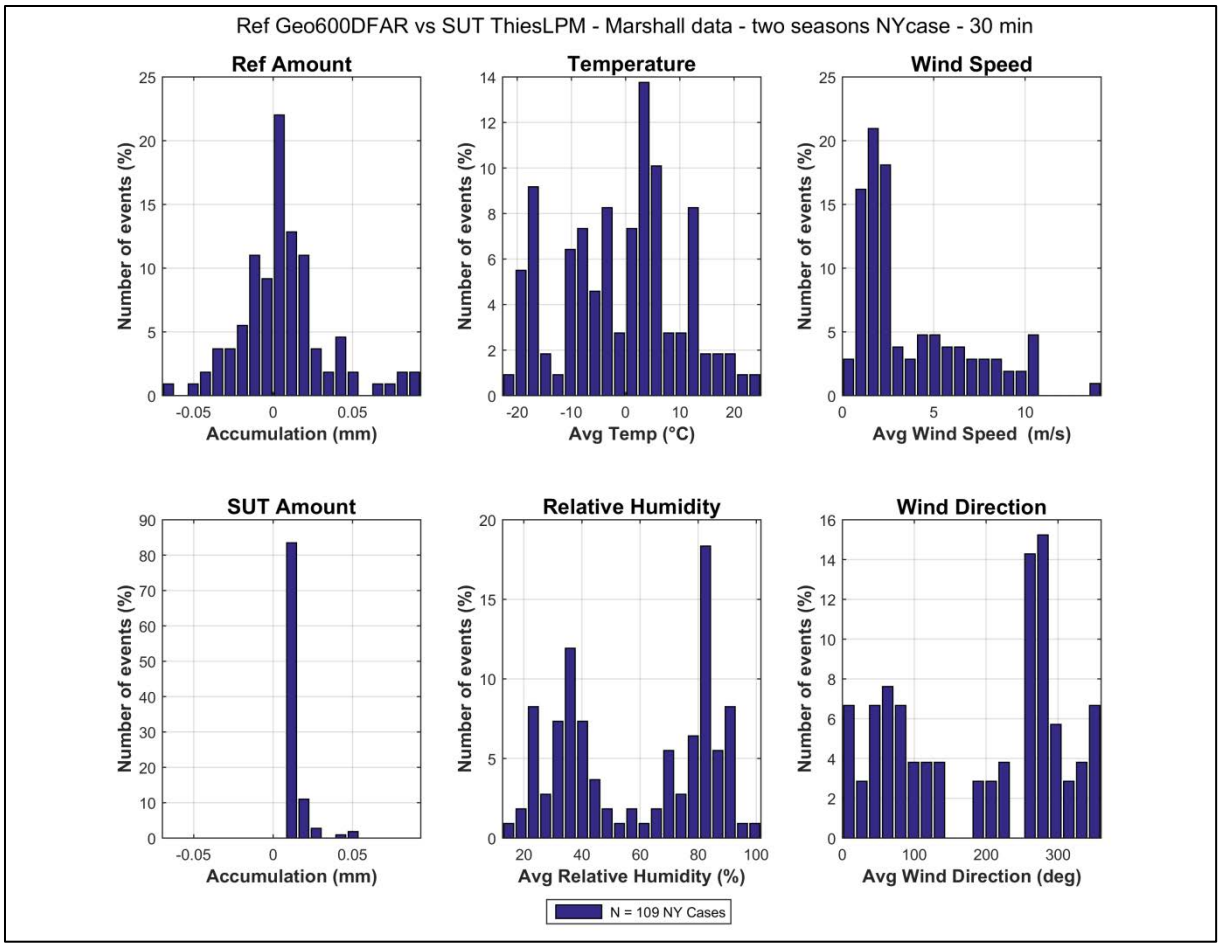


Figure 16: Histograms of NY events from Marshall similar to results of Figure 14, but using a DRD11A as the precipitation detector (instead of the Parsivel^2) for the R2 reference.

7.6. Threshold selection

The threshold to be set for the Thies LPM in order to report precipitation adequately over a 30 min interval (3 STD, according to the methodology defined in Section 3.6.1.3.2 of the SPICE Final Report) corresponds to 0.135 mm (3 x 0.045 mm, from Table 7). As shown in the contingency tables (Table 5), a threshold of 0.1 mm/30 min enables to get rid of NY events whose origin is not confirmed or well established.

8. Operational considerations

The overall experience with the Thies LPM at both sites was positive. The sensor appeared to be robust and reliable operationally, with almost no data breaks during the two seasons (or, if so, there weren't due to the sensor). Moreover, the Thies LPM was found easy to install, with an install kit very user friendly. A note was however made on the difficulty to access the sensor port (no connector outside), leading to the necessity to open the instrument to plug or unplug it. Adding a connector on the cover of the instrument is therefore advisable.

The Thies LPM was installed in a heated configuration. No capping issue has been noticed during the experiment (difficult to assess, since the shield prevented webcam to take full picture of the

instrument). Some snow can accumulate on the horizontal arms of the sensor. This could result in blowing snow and, consequently, false event report (less likely to happen with the shield, reducing wind on and around the sensor).

8.1. Maintenance

The Thies LPM requires almost no maintenance, which makes it suitable for remote locations. A cleaning of the glass panes is however recommended every 3 months (see User Manual).

9. Performance considerations

- When no threshold is applied to the SUT accumulation (Case 1), a high number of false alarms events (NY) were reported by the Thies LPM at Marshall site when comparing with the site reference. These seem to be mainly due to the higher sensitivity of the SUT compared to the field reference system, but it remains that it has not been completely understood if some are real or artefacts. Applying a threshold (of 0.1 mm/30 min for instance in Case 2) to the Thies LPM accumulation output, though, significantly reduces the FAR and ensures to derive more reliable precipitation events.
- Overall, the Thies LPM underestimates solid precipitation accumulation and increasingly underestimates with wind speed for mixed and snow precipitation. It is supposed to be mainly due to the shield surrounding and shadowing the measurement area.
- Colder sites, as alpine sites, may result in less good performance of the Thies LPM for the derivation of precipitation amount, than continental sites.
- High scatter of CE on an event basis (30 min) makes the Thies LPM not suitable to derive accurate precipitation amount under mixed and snow conditions over near real time periods, as the 30 min intervals used in this study.

WMO-SPICE Instrument Performance Report
Vaisala FS11P

1. Technical specifications

Instrument model:	Vaisala FS11P (FS11 / PWD32 combination)
Sample volume:	0.1 L
Physical principle:	Sensor combining the functions of a forward scatter visibility meter and a present weather sensor. It can measure the intensity and amount of both liquid and solid precipitation.
Operating temperature range:	-40 to +65°C
Measurement uncertainty:	Scatter measurement accuracy of ± 3 %
Sensitivity:	0.05 mm/h or less, within 10 min

Note: Specifications from manufacturer provided documentation.

Figure 1: FS11P at Sodankylä test site.

2. Data output format

The Vaisala present weather sensor FS11P provides precipitation information as intensity, accumulation or weather code. Table 1 summarizes the main output parameters from the instrument. The firmware version of the FS11P tested during this experiment was the version V2.26.14.95.

Table 1: Summary of main instrument outputs, as recorded by the site during the experiment.

Measured Parameters	Units
Water Intensity	[mm/h]
Cumulative Water Sum	[mm]
Cumulative Snow Sum	[mm]
Visibility	[m]
Present Weather Code (WMO)	-
Present Weather Code (NWS)	-
Present Weather Code (METAR)	-
Housekeeping	-

This document reports on the ability of the FS11P to derive solid precipitation. Although the 'Cumulative Snow Sum' parameter could seem appropriate to derive the results, it is actually not meant to give a water content measurement, but rather a coarse estimation of the new snow accumulation, i.e. new snow depth (see FSS11P User Manual). The analysis has been performed using the 'Cumulative Water Sum' parameter, which gives the water content of snow and is therefore the suitable parameter to compare fairly the sensor with the field reference.

3. SPICE test configuration

The FS11P, as sensor under test (SUT), has been tested on one site:

Test Site: Sodankylä (Finland)

Sensor Provider(s): The instrument evaluated was provided by the manufacturer (Vaisala)

Figure 2: Map of SPICE site testing FS11P instrument.

A summary on the configuration of the instrument as tested, the duration of tests and availability reflected in these results, and the ancillary measurements used, is available in Table 2, Table 3, and Table 4, respectively.

Table 2: Summary of instrument configuration and data output. Details and photos on site configuration are available in the site commissioning protocol.

	Sodankylä
Main prevailing wind direction	South
Sensor orientation	East – West
Height of installation	PWD32 2.15 m, FS11 2.90 m
Heating	Yes, as recommended
Shield	No
Data QC	SPICE QC methodology
SUT data output frequency	15 sec
Data temporal resolution	1 min
Processing interval for SPICE data analysis	30 min

Table 3: Data availability, by measurement season.

Measurement season	**Sodankylä**
Season 1 (Oct. 2013 – Apr. 2014)	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓

Table 4: Summary of reference and ancillary measurements, with measurement height.

	Sodankylä
R2 Site Reference	OTT Pluvio^2 1500mm (DFAR) (4 m, rim height)
R2 Precip Detector	DRD11A (Site*, 2013-2014) (1 m) OTT Parsivel^2 (DFAR, 2014-2015) (2.7 m)
Ancillary Temp Sensor	Vaisala HMP155 (2 m)
Ancillary RH Sensor	Vaisala HMP155 (2 m)
Ancillary Wind Sensor	Thies acoustic 2D wind sensor (3.5 m)

**A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR-fence; however, in cases where a more sensitive detector is available outside of the DFIR-fence, or there are issues with the detector within the DFIR-fence, a precipitation detector elsewhere on the site can be employed.*

4. Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6 of the SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute intervals:

- Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip;
- SUT 'Yes': SUT accumulation > 0 mm;
- SUT 'No': SUT accumulation = 0 mm.

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores.

4.1.2. Skill score assessment

The ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

The above scores are computed using the formulations provided in Section 3.6 of the SPICE Final Report.

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

5. Environmental conditions

The environmental conditions at the site over the duration of the test period are expressed as probability density functions (PDFs) for air temperature, relative humidity, wind speed, wind direction, and precipitation rate in Figure 3. Both the entire period data, and the data corresponding only to the precipitation events reported by the site reference, R2, are shown. The precipitation percentage in Figure 3 represents the number of minutes of precipitation over a standard 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration.

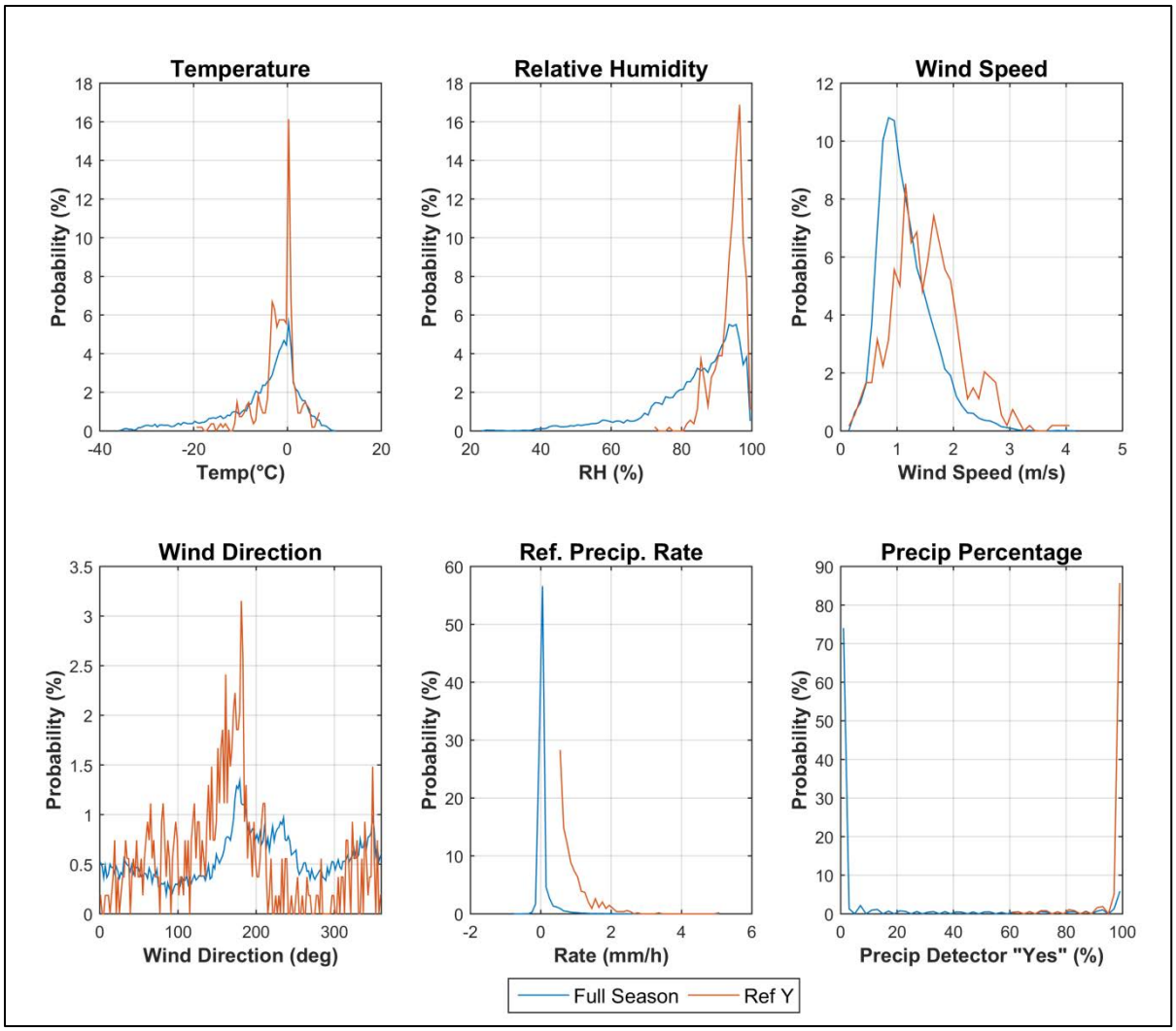


Figure 3: Summary of aggregated environmental conditions on the SPICE site that operated FS11P, over the entire duration ('Full Season', in blue) and for data corresponding to precipitation events, as reported by the site R2 reference ('Ref Y', in red), during the tests, as per Table 3 above.

6. Evaluation of performance over the range of operating conditions

6.1. Skill score assessment

The overall ability of the SUT to represent precipitation similarly to the site field reference R2, is assessed using contingency tables (Section 4.1.1) and derived selected skill scores (Section 4.1.2). To better understand the potential influence of threshold choices on the derived results, two cases are considered here (see note below). The contingency results related to these two cases are given in Table 5 and the respective skill scores in Table 6.

Note: Following the data derivation explained in Section 4.1.1, the conditions required to have a 'Yes' or a 'No' event over the 30 min interval, for the reference and the SUT, for the two different cases treated here, are:

CASE 1 *(as defined in Section 4.1.1)*:

- *Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip*
- *Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip*
- *SUT 'Yes': SUT accumulation > 0 mm*
- *SUT 'No': SUT accumulation = 0 mm*

CASE 2:

- *Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip*
- *Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip*
- *SUT 'Yes': SUT accumulation ≥ 0.1 mm*
- *SUT 'No': SUT accumulation < 0.1 mm*

Results presented in this report are based on Case 1.

Table 5: Contingency Tables: detection of precipitation of the FS11P relative to the field reference, expressed as number of events over the entire test period. The skill scores associated with these events are given in Table 6.

Case 1 Ref Plv2DFAR

SUT FS11P	Yes	No	Total
Yes	524	598	1122
No	0	9729	9729
Total	524	10327	10851

Case 2 Ref Plv2DFAR

SUT FS11P	Yes	No	Total
Yes	515	9	524
No	9	10318	10327
Total	524	10327	10851

Table 6: Skill Scores for the FS11P. POD: Probability Of Detection, FAR: False Alarm Rate, B: Bias, HSS: Heidke Skill Score (see Section 4.1.2 for more details).

FS11P, Skill Scores				
	POD	**FAR**	**B**	**HSS**
Case 1	100%	53.3%	214%	61.1%
Case 2	98.3%	1.7%	100%	98.2%

6.2. Assessment of SUT performance during non-precipitating events

The performance of the SUT in the absence of precipitation (when the reference precipitation detector recorded 30 minutes without precipitation) is represented in Figure 4 and Table 7, reflecting the distribution of the sensor response, as measured during the interval.

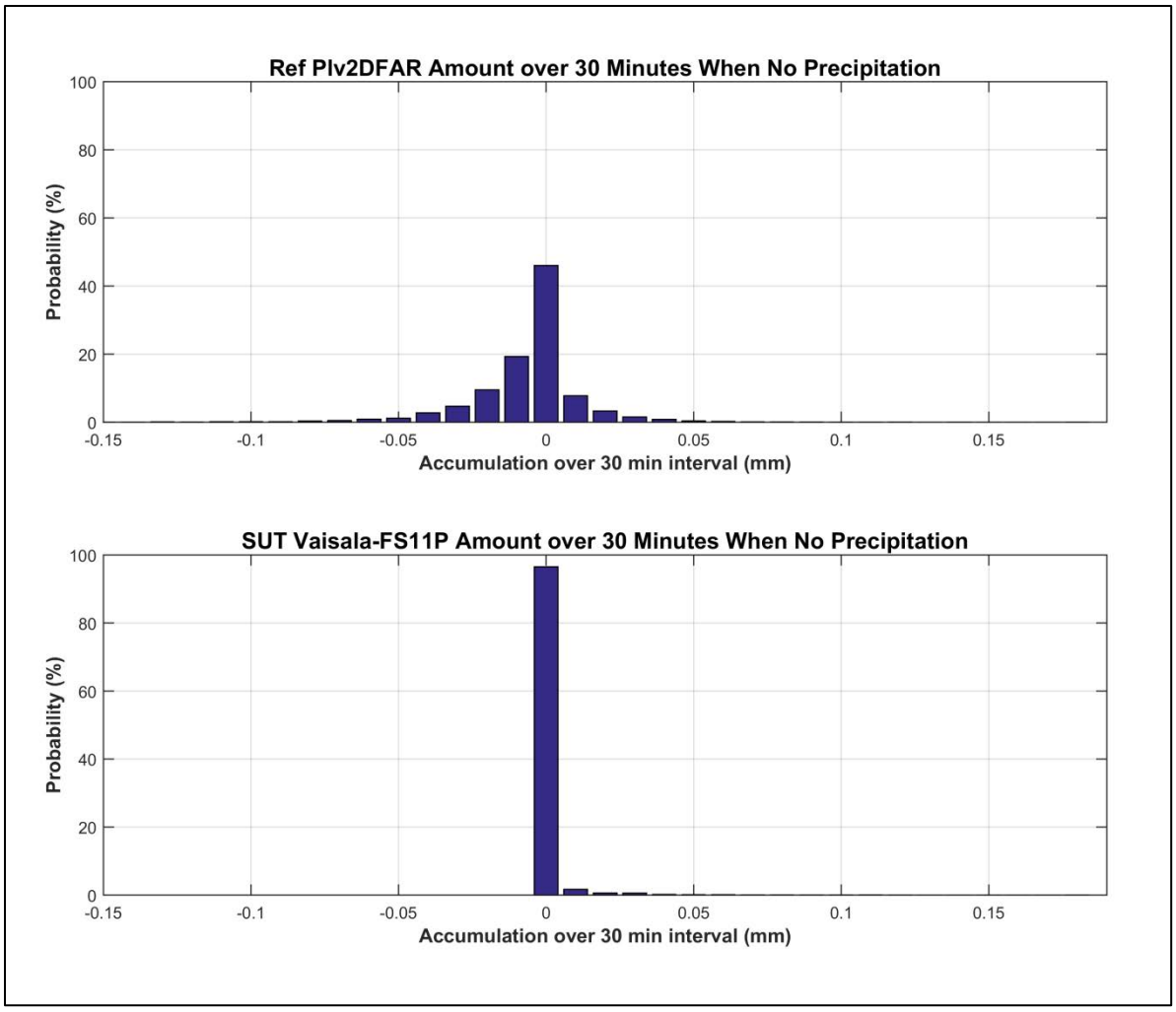


Figure 4: Probability of occurrence of a response during a 30 min interval in the absence of precipitation, represented by the signal output from (top) the R2 reference and (bottom) the SUT FS11P. Statistics associated with these graphs are given in Table 7.

Table 7: Reference and SUT statistics of response signal when no precipitation was occurring, as plotted in Figure 4; Average (Avg), standard deviation (STD), maximum (Max) and minimum (Min) of the response signal, together with the number of events (Num) over the test period is given.

No Precip Statistics – Reference: Plv2DFAR

	Ref Avg [mm]	Ref STD [mm]	Ref Max [mm]	Ref Min [mm]	Ref Num
Sodankylä	-0.003	0.021	0.190	-0.140	10335

No Precip Statistics – SUT: Vaisala-FS11P

	SUT Avg [mm]	SUT STD [mm]	SUT Max [mm]	SUT Min [mm]	SUT Num
Sodankylä	0.001	0.007	0.190	0.000	10335

6.3. Ability of the SUT to measure precipitation

6.3.1. Yes-Yes cases

Quantitatively, the performance of the SUT to derive and report precipitation is assessed relative to the site reference in several graphs and tables illustrated in this section, using only the cases where both instruments reported precipitation over the 30 min interval, according to the criteria used in Table 5 (cases 'Yes-Yes', or shorter 'YY').

6.3.1.1. Time series plots

The time series (cumulative sum of 30 min YY events accumulation) of the SUT is plotted against the reference for the two seasons, by precipitation type (see Section 4.1.4) in Figure 5.

The corresponding seasonal accumulations are given in Table 8.

Table 8: Seasonal accumulation [mm] for Sodankylä test site based on the sum of YY events from the SUT FS11P and the field reference R2.

Sodankylä				
	Season 2013/14		Season 2014/15	
[mm]	Ref	SUT	Ref	SUT
All events	124.49	102.59	114.03	110.10
Rain events	17.67	17.77	5.60	5.68
Mixed events	72.26	57.53	51.04	52.42
Snow events	34.56	27.29	57.39	51.99

6.3.1.2. Scatter plots and RMSE values

Scatter plot of the amount derived by the SUT versus the reference amount, and discriminated by precipitation type, is given in Figure 6.

Quantitatively, the SUT performance is assessed using the Root Mean Square Error (RMSE) also known in practice as Operational Comparability. The results are available in Table 9.

Table 9: Root Mean Square Error (RMSE) statistics, in mm, for the SUT FS11P with respect to the field reference, by precipitation type, including both seasons data.

RMSE [mm]	All	Rain	Mixed	Snow
Sodankylä	0.146	0.137	0.157	0.133

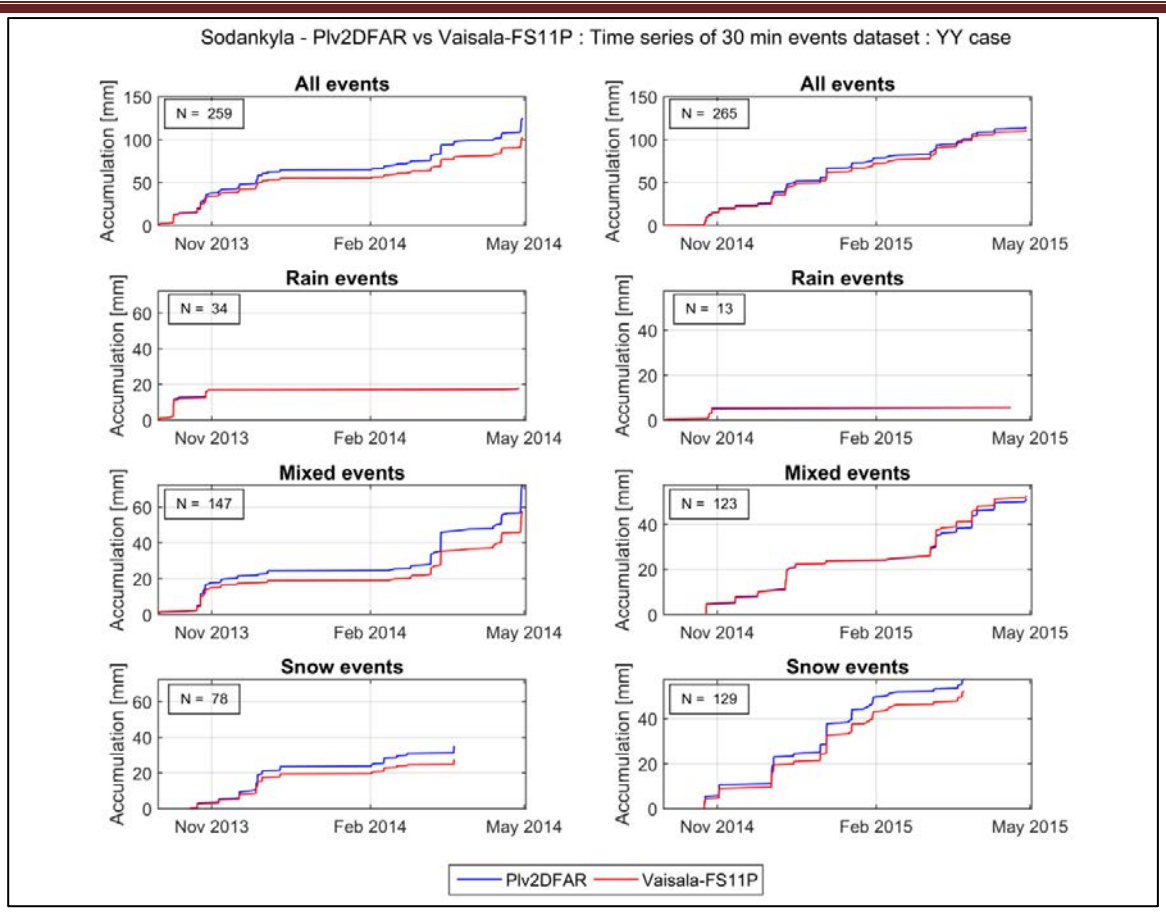


Figure 5: Time series based on 30 min YY events of the SUT FS11P against the field reference, discriminated by precipitation type (Rain, Mixed, Snow), for both seasons (2013/14 on the left, 2014/15 on the right).

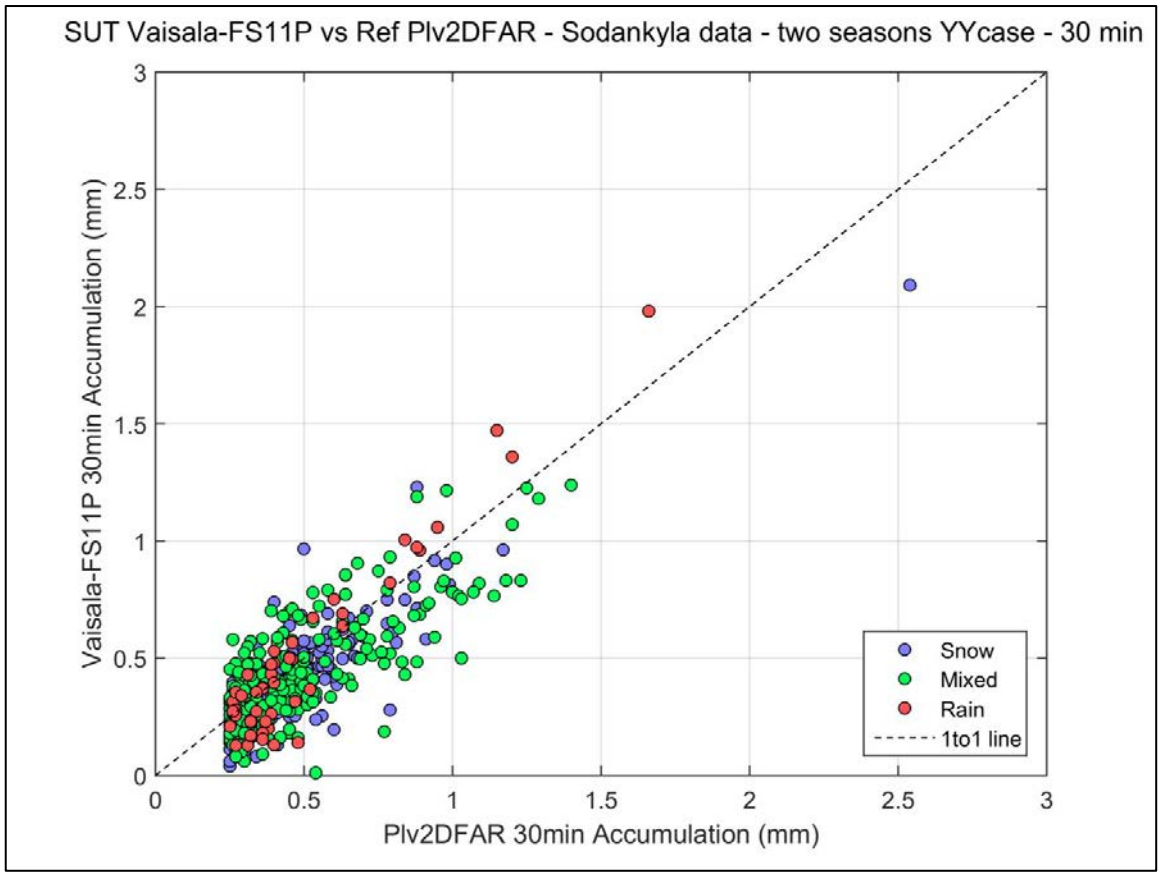


Figure 6: Scatter plot based on 30 min YY events accumulation from the SUT FS11P against the field reference, for the two seasons, discriminated by precipitation type.

6.3.1.3. Catch efficiency evaluation by precipitation type

The Catch Efficiency (CE) of the SUT is represented by histograms (Figure 7) and boxplots (Figure 8), both discriminated by precipitation type.

The quantitative evaluation of the CE is provided in Table 10. The mean catch efficiency is given in the first line of this table, considering both seasons data and for each category of precipitation type as well as for all the events together.

Note: All events with a CE greater than 3, if any, are included in one category named '3 and more' or '≥3' in the upcoming graphs. Additionally, for all graphs representing the CE, a line is added at CE = 1, which represents the ideal case where the SUT reports exactly the same precipitation amount as the reference.

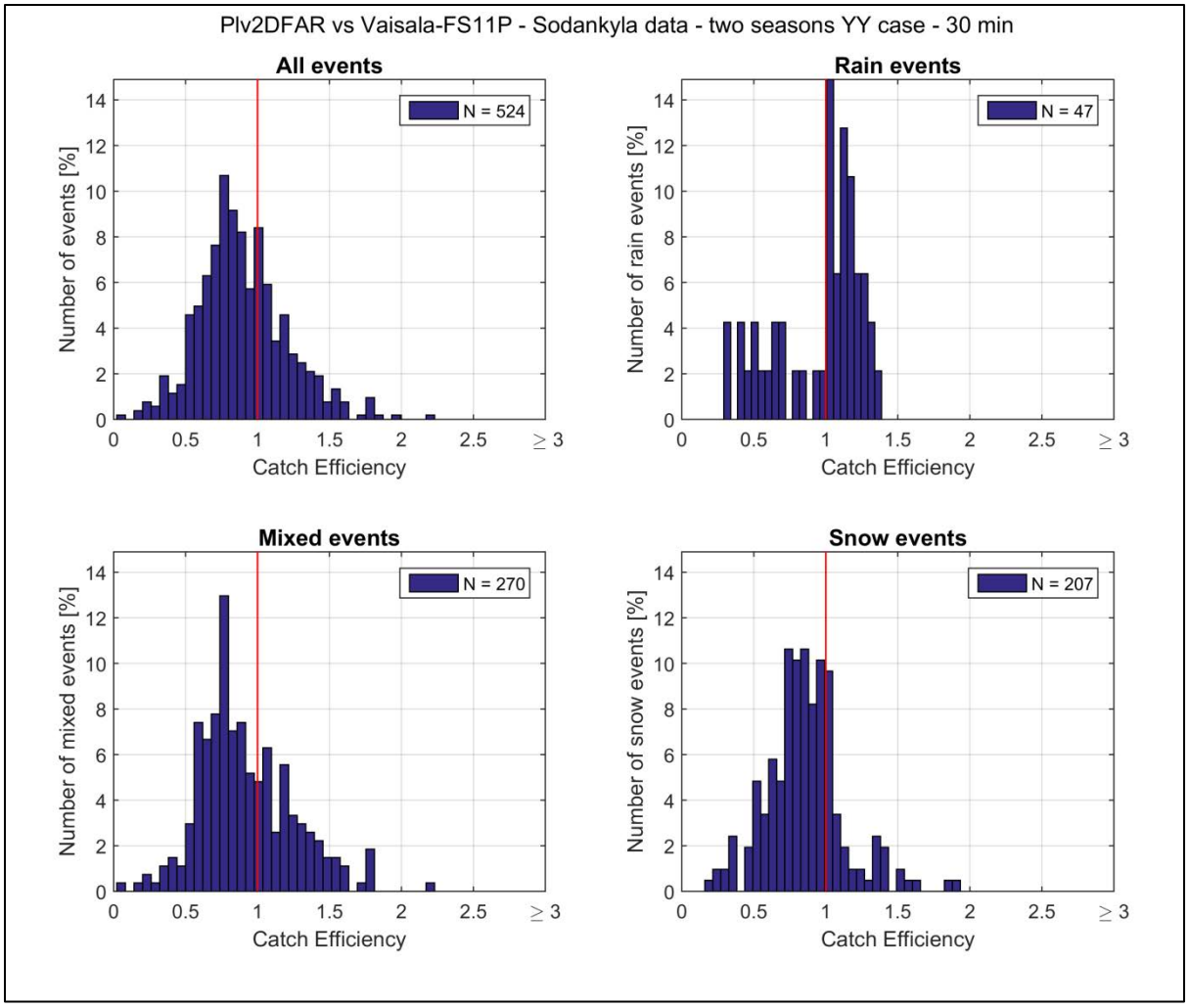


Figure 7: Histograms based on 30 min YY events for the two seasons, representing the distribution of the catch efficiency of the SUT FS11P against the field reference (SUT/Ref), discriminated by precipitation type, with number of events given in the legend for each category. The red line at CE = 1 represents the ideal case.

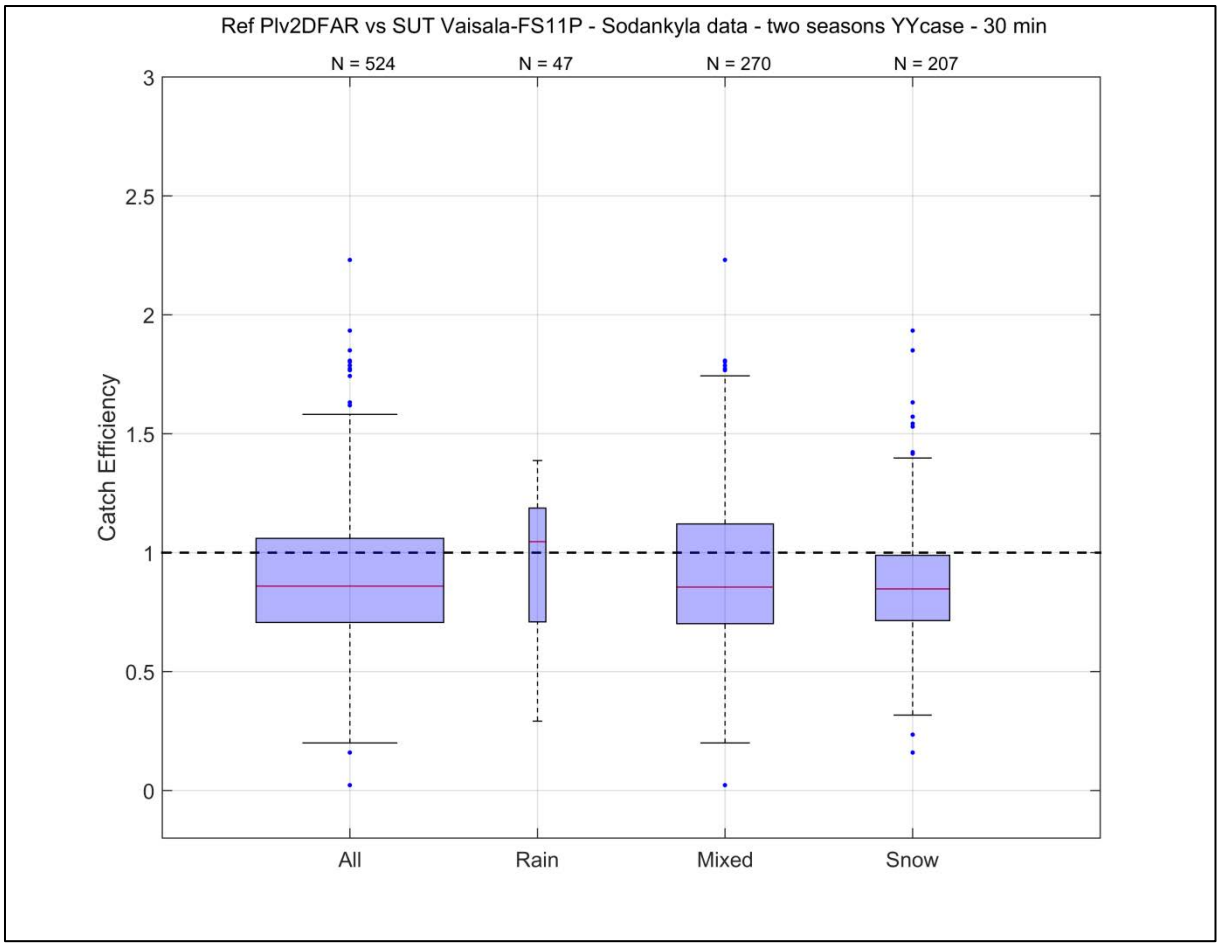


Figure 8: Boxplot based on 30 min YY events of the two seasons, representing the catch efficiency of the SUT FS11P against the field reference (SUT/Ref), discriminated by precipitation type, with number of events given at the top of each category. The width of the boxes is proportional to the percentage of events in each category ('All' box represents 100% data). The dashed black line at CE = 1 represents the ideal case.

Table 10: Statistics related to the CE of SUT FS11P against the field reference (SUT/Ref, see Figure 8), discriminated by precipitation type.

Catch Efficiency Statistics: **FS11P**

CE Boxplot Parameters	All	Rain	Mixed	Snow
Mean	0.90	0.96	0.92	0.86
Median	0.86	1.05	0.86	0.85
75 percentile	1.06	1.19	1.12	0.99
25 percentile	0.71	0.71	0.70	0.71
Upper Whisker	1.58	1.39	1.74	1.40
Lower Whisker	0.20	0.29	0.20	0.32
Maximum	2.23	1.39	2.23	1.93
Minimum	0.02	0.29	0.02	0.16
# Outliers	13	1	7	10
# Outliers ≥ 3	0	0	0	0
# Events	524	47	270	207

6.3.1.4. Catch efficiency dependency on wind speed

The variation of the SUT catch efficiency (SUT/Ref) with wind speed is illustrated in a scatter plot together with a boxplot in Figure 9, both discriminated by precipitation type.

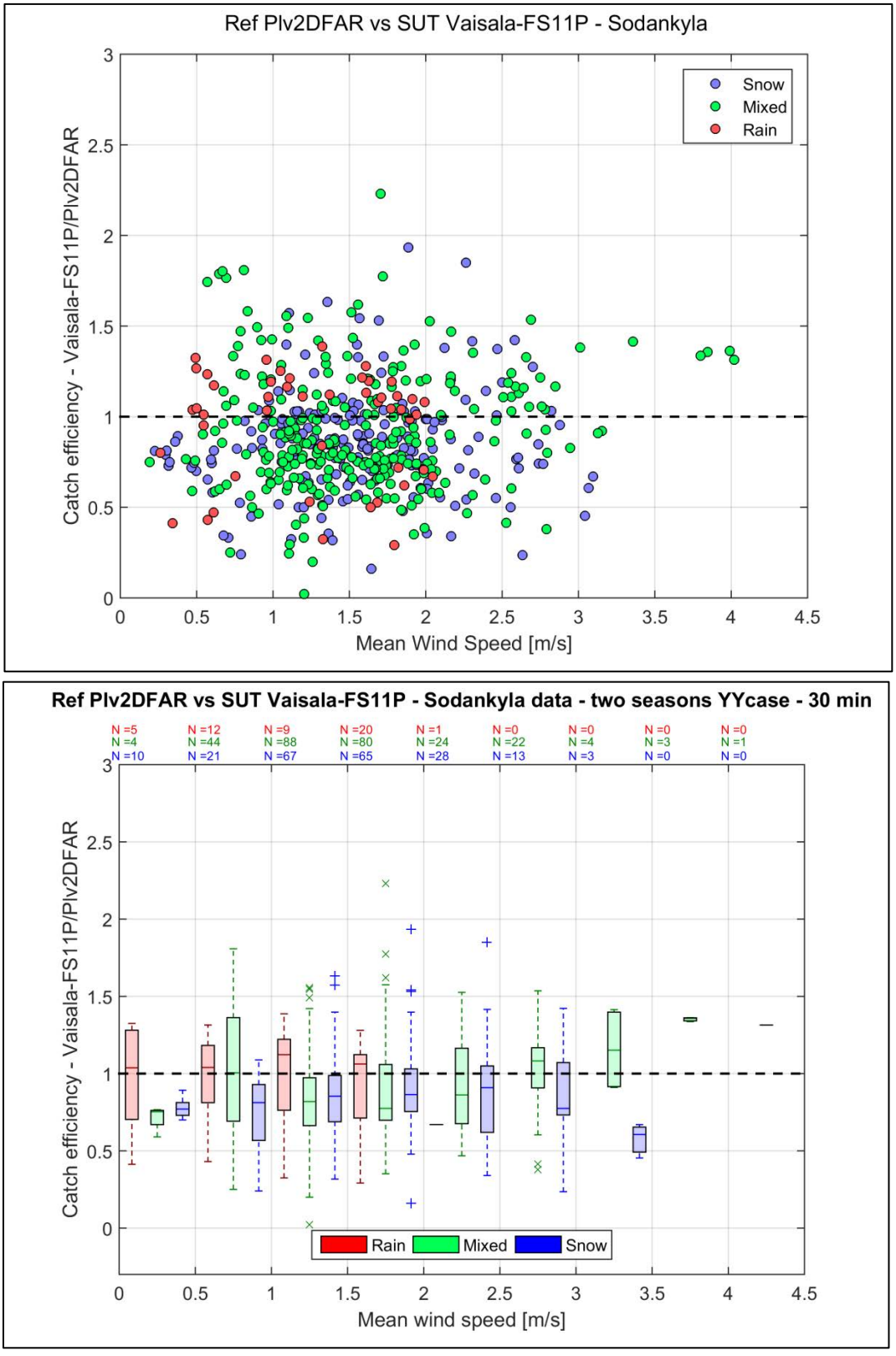


Figure 9: Scatter plot (top) and boxplot (bottom) based on 30 min YY events for the two seasons, representing the catch efficiency of the SUT FS11P with respect to the field reference (SUT/Ref), against wind speed and discriminated by precipitation type. The dashed black line at CE = 1 represents the ideal case.

6.3.1.5. Catch efficiency dependency on wind direction

In order to assess the dependency of the CE with wind direction, a wind rose is produced (Figure 10) representing the wind data of the two seasons, binned by catch efficiency in order to represent undercatch (CE < 0.8), overcatch (CE > 1.2) and catch efficiency of 1 ± 20% of the SUT.

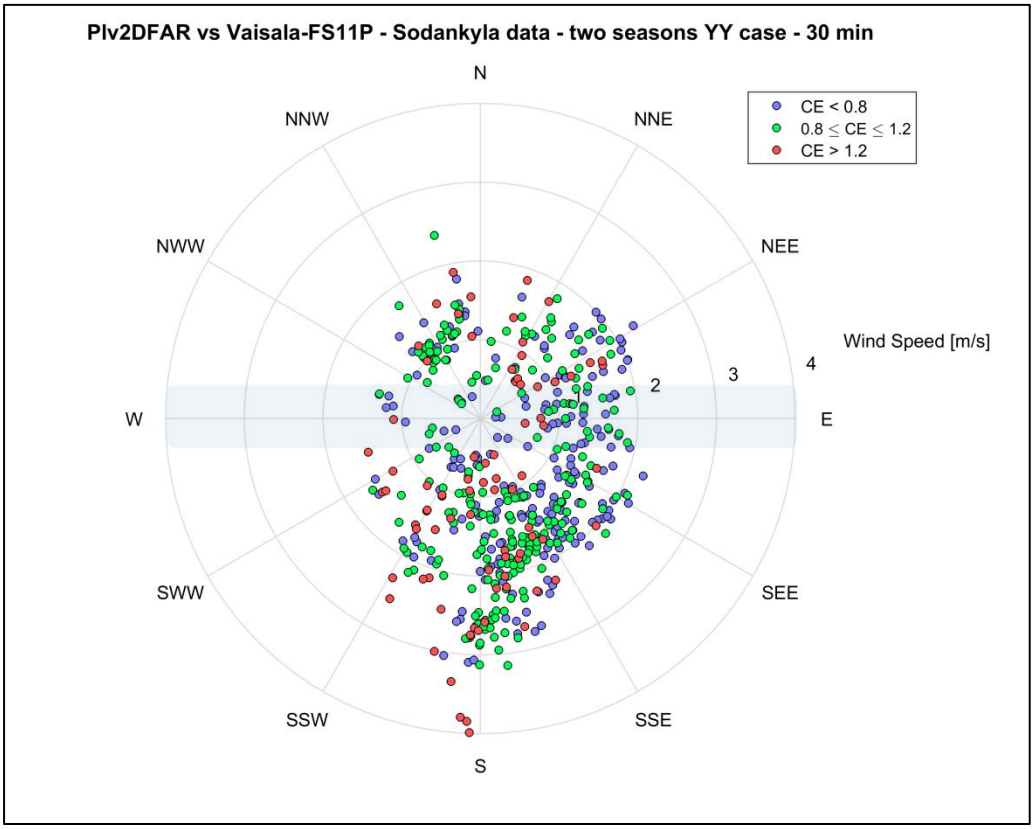


Figure 10: Precipitation events (YY cases) as function of wind speed and direction, and binned by catch efficiency (CE). The grey zone indicates the SUT orientation.

6.3.2. Yes-No and No-Yes cases

Events when the site reference and SUT do not agree on the occurrence of precipitation includes two categories of cases (Section 4.1.1): (1) when the field reference reported a precipitation event, while the SUT did not (Yes-No cases, 'YN'), and (2) when the field reference did not report a precipitation event while the SUT did (No-Yes cases, 'NY').

Histograms illustrating field reference and SUT reports and associated site conditions for all NY cases (No YN cases were reported for the FS11P) during the test period are provided in Figure 11.

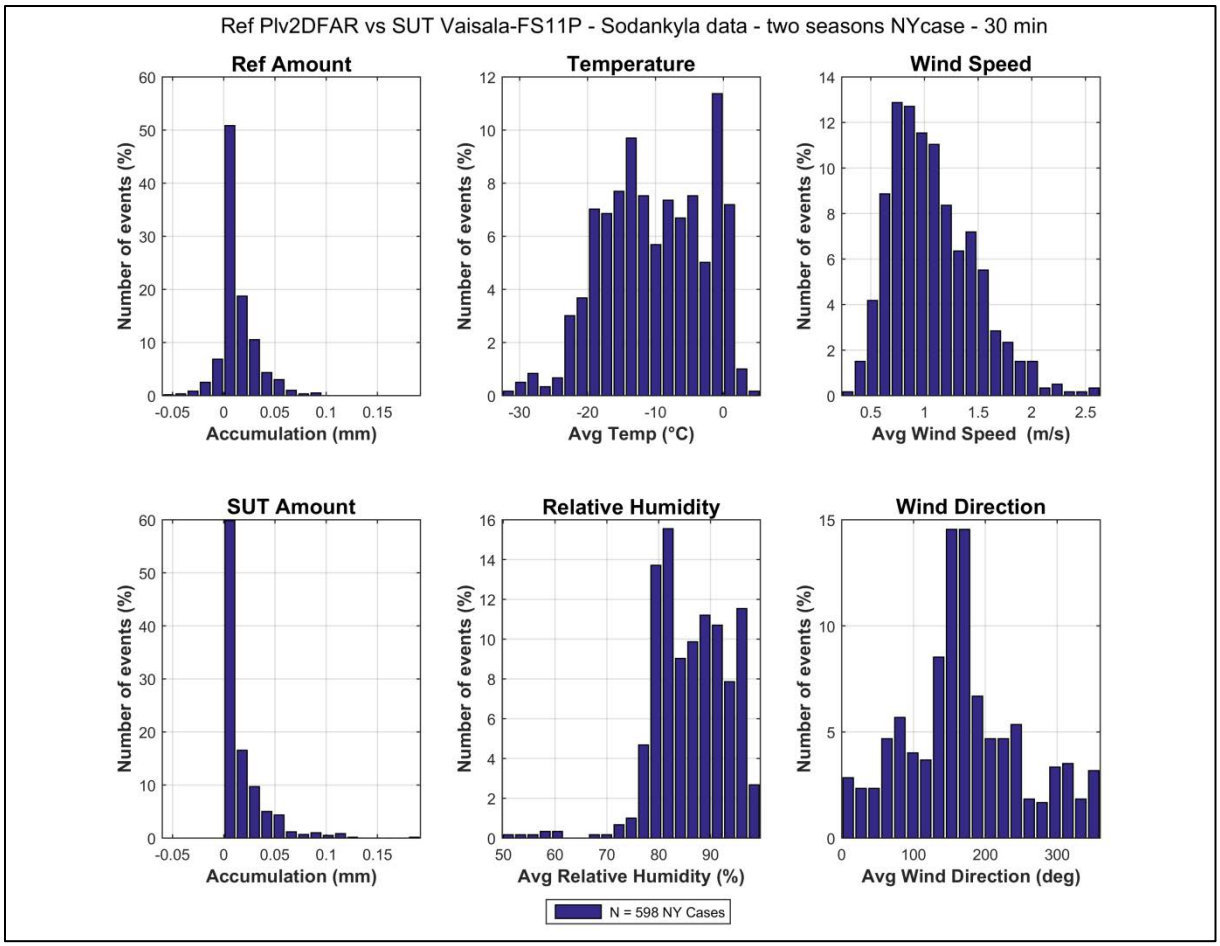


Figure 11: Histograms of SUT FS11P and field reference accumulations (left column), along with distributions of mean temperature, mean relative humidity, mean wind speed, and mean wind direction for all NY cases (number indicated in the legend) of the 30 min intervals, as reported by Table 5.

7. Interpretation of results

7.1. Operating conditions

Sodankylä is a site characterized by very low wind speeds, as shown in Figure 3, with a majority of precipitation events occurring below 3 m/s. Regarding temperatures, Sodankylä recorded most of the precipitation events at temperature between -10 and 0°C, with extreme by -20°C. The site shows fairly humid climate conditions as expected from a Northern Boreal climate site.

7.2. Reliability in detecting precipitation

The Vaisala FS11P was efficient in detecting precipitation when the reference reported precipitation, as presented in Table 6, with a POD of 98% or more, looking at the two different threshold cases. When no threshold is applied to the SUT (Case 1), the FAR of 53%, the bias around 200% and the HSS of 61% are a direct consequence of the large number of NY events (see Section 7.5 for more details). However, it can be noted that the FAR score is reduced to 1 to 2%, and the bias and HSS scores improved, when a threshold of 0.1 mm over 30 min events is applied to the SUT (Case 2), indicating that the majority of the NY events of Case 1 are for very low accumulations. This could be the consequence of a difference in sensitivity between the weighing gauge used as the reference and the non-catchment type SUT, related to their different measurement principle.

7.3. Performance of SUT during no-precipitation events

The FS11P has a stable output signal, showing almost no noise during no-precipitation events (see Figure 4 and Table 7). Note that the no-precipitation study is performed using only the reference precipitation detector data (here a DRD11A for the first season and a Parsivel^2 for the second season) with the criteria ensuring that 30 min of 'no-precipitation' was recorded. For these cases, the FS11P indicates 0 mm for more than 94% of the time, indicating a good agreement with the reference precipitation detector on the absence of precipitation. Most of the remaining cases (where the reference precipitation detector doesn't detect any precipitation while the SUT does) correspond to very low accumulations reported by the SUT (less than 0.1 mm/30 min) and can be related to the NY events mentioned above and analyzed in Section 7.5.

7.4. Performance of SUT during precipitation events

When both the reference and the FS11P report precipitation (YY cases), the overall results show a very similar behavior of the sensor for all precipitation types.

The RMSE in Table 9 shows almost identical values for rain, mixed and snow precipitations, with 0.15 ± 0.01 mm. It can be noted from Figure 6, that the rain events with accumulation greater than 0.5 mm show a slight overcatch.

Despite this slight overcatch for some rain events, the FS11P has a tendency in overall to undercatch, as represented by Figure 7 andFigure 8 and Table 10, with a mean catch efficiency over the two seasons of 0.9, dropping to 0.86 for snow. It has to be reminded, however, that the number of snow events is significantly higher than that for rain events (few during winter seasons), and have hence a stronger impact over the global assessment.

Overall, the FS11P seems to be a reliable instrument to account for the total accumulation over a longer period (e.g. one season), with a mean catch ratio close to the 1-line, as shown in Figure 8 or Figure 9 (bottom). The relatively high scatter visible in Figure 9 (top), with catch ratio varying randomly from 0.3 to 1.5, on event based statistics (typically 30 min interval), occurring at all wind speeds (up to 4 m/s, maximum wind speed in Sodankylä), tends to show that the FS11P is less reliable to derive solid and mixed precipitation accumulation over near real time periods.

Even if better in deriving accumulation for longer time periods, the FS11P, as other non-catchment type instruments, doesn't have an assurance of the continuity in the measurements, though, which is critical to long term data collection. If power is off or signal transmission is interrupted, no data is recorded and there is no possibility to know what has fallen during this time, thus affecting the long term data reports of the FS11P.

The results in Figure 9 indicate no clear dependency of the catch efficiency with increasing wind speed, for all precipitation types. The performance of the sensor for higher wind speed cannot be assessed, since Sodankylä is a site with low wind speed (up to 4 m/s), and the SUT was not tested in other environmental conditions.

The wind rose (see **Error! Reference source not found.**) shows no dependency of the catch efficiency with wind direction. The sensor was installed perpendicular to the prevailing wind direction during precipitation events. An assessment with a sensor parallel to the main flow has not been assessed, though, since no such configuration was tested.

7.5. Assessment of Yes-No, No-Yes events

The assessment of YN, NY events completes the picture of the performance of the SUT. As it can be seen in Table 5 and Figure 11, when no threshold is applied to the SUT (Case 1), the FS11P shows a large amount of NY cases, but no YN cases were reported, the latter indicating the good agreement of the sensor to detect precipitation whenever the reference reported precipitation. When a threshold is applied to the SUT (Case 2) - here a threshold of 0.1 mm/30 min – the number of NY events is significantly reduced. Therefore, applying a threshold enables to ensure the good agreement of the FS11P with the reference, in all cases (YN, NY, YY, NN), and prevents the user to deal with NY events of unknown origin (either real or artefacts).

Looking more closely at Case 1 and according to Table 5, almost 600 NY events occurred during the two seasons of the experiment. It could be useful to remind that these NY events are characterized, over the 30 minutes of an event, by an accumulation from the reference below the defined threshold of 0.1 mm, the reference precipitation detector having recorded 0 min of precipitation, and the SUT indicating more than 0 mm (see Section 4.1.1). This high number of NY events appears to happen for very low accumulations (mostly less than 0.1 mm/30 min) reported by the FS11P and for low temperatures and wind speeds (see Figure 11). Moreover more in-depth analysis showed that the NY events occurred primarily in the second season (537 for season 2 compared to 61 for season 1), where the precipitation detector being part of the reference was changed from the DRD11A capacitive sensor by a Parsivel^2 (see Table 4). Figure 12 shows some of the NY events of season 2 over the accumulation curve of the R2 reference weighing gauge, together with the precipitation detection measured by the Parsivel^2, the DRD11A and the FS11P. It can be noted that this light precipitation was well detected by the DRD11A and the FS11P, but nothing was measured by the

Parsivel^2. This led to the idea that the NY results could change if the choice of the reference precipitation detector was different for the second season.

To better address these differences in precipitation sensitivity and to give a sense of the results dependency on the selected reference precipitation detector, a second study has been made using the DRD11A as the reference precipitation detector for the second season as well (instead of the Parsivel^2). The same number of YY and YN events was found, but the number of NY events was drastically reduced (162 against 598 with the Parsivel^2), leading to a FAR of only 23.6% (against 53.3% with the Parsivel^2), a bias of 131% and an HSS of 85.8%. Even if capacitive precipitation detectors, such as the DRD11A, are less sensitive than laser-based precipitation detectors, like the Parsivel^2 (see Section 4.1.3.5.2 of the SPICE Final Report), the DRD11A seems to better match the FS11P results in terms of precipitation detection (as it could be seen in Figure 12). A similar graph to Figure 11, but this time using the DRD11A as the reference precipitation detector for both seasons, is given in Figure 13 for the FS11P NY events. This figure shows a wider distribution over the wind speed range, but a distribution of temperatures closer to 0°C. It is still possible that these remaining 162 NY events are real light events that were not detected by the DRD11A (which measured 30 min of no precipitation for all of them), but this remains uncertain and no other parameters or photos are available to confirm or infirm this statement.

To conclude, the FS11P NY events of either study (with DRD11A at season 1 and Parsivel^2 at season 2 or only DRD11A for both seasons) reflect therefore rather a reliable higher sensitivity of the FS11P compared to the reference system. Indeed, on one hand the reference precipitation detector could have a lower sensitivity than the FS11P, on the other hand the reference weighing gauge has a different physical principle of measurement implying an expected lower sensitivity than the FS11P.

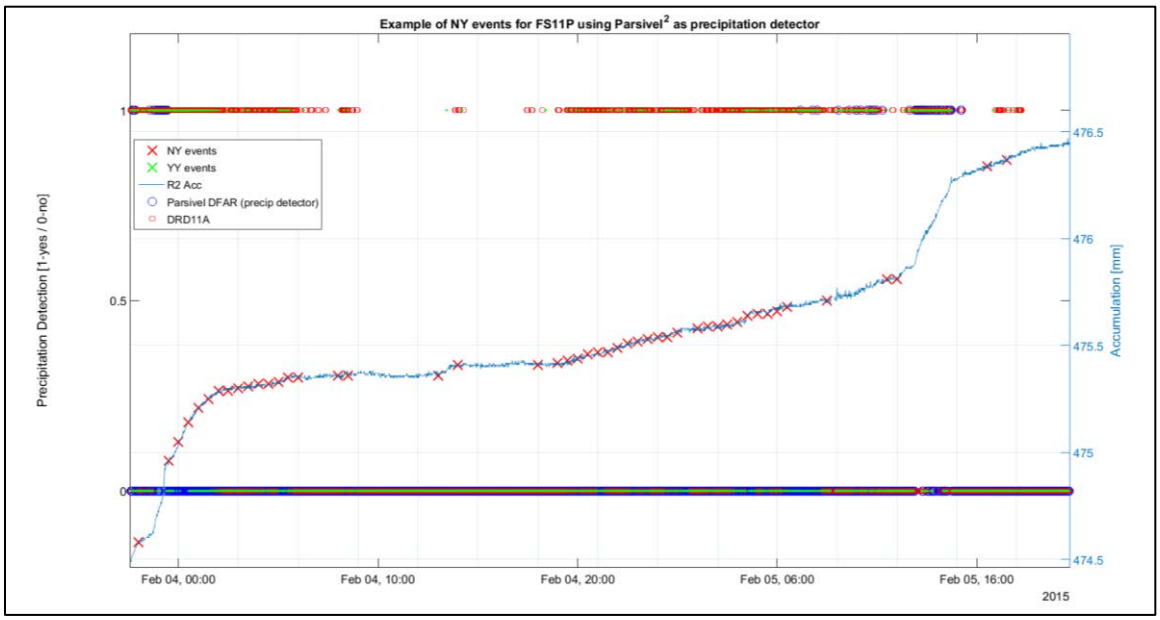


Figure 12 : Example of NY events for the SUT FS11P using the Parsivel^2 as the reference precipitation detector for the second season.

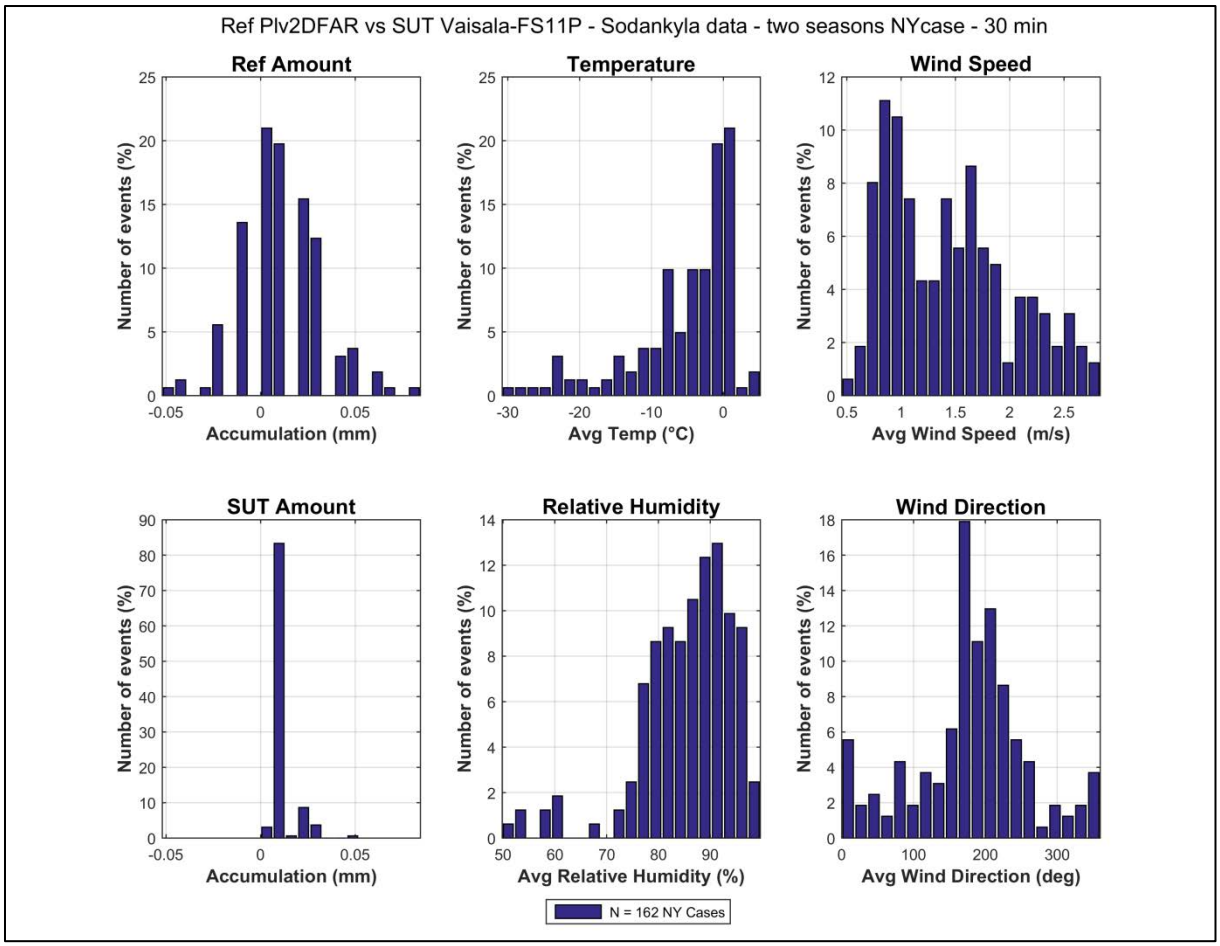


Figure 13: Histograms of NY events similar to results of Figure 11, but using a DRD11A as the reference precipitation detector (instead of the Parsivel[2]) for the second season.

7.6. Threshold selection

The threshold to be set for the FS11P in order to report precipitation adequately over a 30 min interval (3 STD, according to the methodology defined in Section 3.6.1.3.2 of the SPICE Final Report) corresponds to 0.02 mm (3 x 0.007 mm, from Table 7). Additionally, as shown in the contingency tables (Table 5), a threshold of 0.1 mm/30 min enables to get rid of events whose origin is not confirmed or well established.

8. Operational considerations

The overall experience with the FS11P at Sodankylä was positive. The sensor appeared to be robust operationally with almost no data breaks due to the instrument (there was one major data gap of $1^{1/2}$ month in December 2013, but this was due to an issue with the data collection program).

During the experiment, some snow accumulation has been observed on the instrument device and mounting parts (see Figure 14). This is expected to happen on sites with calm weather conditions, as the Sodankylä site. According to the manufacturer, the probability that blowing snow from these accumulations could be counted as precipitation is very unlikely, though, because it would require that the precipitation detector 'RAINCAP' surface and the optical measurement volume, both part of the FS11P, are hit by the snow particles at the same time and with sufficient strength.

Figure 14: Snow accumulation on FS11P device, 2015-02-15, 11:03 UTC.

8.1. Maintenance

The FS11P requires almost no maintenance. According to the User Manual, the sensor's lenses should be cleaned every six months. Before using it operationally, though, the User Manual indicates also that an individual calibration on site may be needed to derive correctly the liquid water content of precipitation.

9. Performance considerations

- When no threshold is applied to the SUT accumulation (Case 1), a high number of false alarms events (NY) were reported by the FS11P when comparing with the reference. These seem to be mainly due to the higher sensitivity of the SUT compared to the reference, but it remains that it has not been completely understood if some are real or artefacts. Applying a threshold (of 0.1 mm/30 min for instance in Case 2) to the FS11P accumulation output, though, significantly reduces the FAR and ensures to derive reliable precipitation events.
- High scatter of CE on an event basis (30 min), but good mean CE over the seasons, makes the FS11P more suitable for deriving solid precipitation amount over long periods of time, giving the condition that the sensor operates continuously.

WMO-SPICE Instrument Performance Report
Vaisala PWD33 EPI

1. Technical specifications

Instrument model:	Vaisala PWD33 EPI
Sample volume:	0.1 L
Physical principle:	Sensor combining the functions of a forward scatter, a capacitive rain, a particle impact and a temperature sensors. These four independent measurements lead to the derivation of precipitation intensity and amount.
Operating temperature range:	-40 to +55°C
Sensitivity:	0.05 mm/h or less, within 10 min

Note: Specifications from manufacturer provided documentation.

Figure 1: PWD33 EPI at Sodankylä test site.

2. Data output format

The Vaisala Enhanced Precipitation Identifier (EPI) PWD33 is a sensor providing precipitation information as intensity, accumulation or weather code. Table 1 summarizes the main output parameters from the instrument. The firmware version of the PWD33 EPI tested during this experiment was the version V2.13N.

Table 1: Summary of main instrument outputs, as recorded by the site during the experiment.

Measured Parameters	Units
Water Intensity	[mm/h]
Cumulative Water Sum	[mm]
Cumulative Snow Sum	[mm]
Visibility	[m]
Present Weather Code (WMO)	-
Present Weather Code (NWS)	-
Present Weather Code (METAR)	-
Housekeeping	-

This document reports on the ability of the PWD33 EPI to derive solid precipitation. Although the 'Cumulative Snow Sum' parameter could seem appropriate to derive the results, it is actually not meant to give a water content measurement, but rather a coarse estimation of the new snow accumulation, i.e. new snow depth (see PWD33 EPI User Manual). The analysis has been performed using the 'Cumulative Water Sum' parameter, which gives the water content of snow and is therefore the suitable parameter to compare fairly the sensor with the field reference.

3. SPICE test configuration

The PWD33 EPI, as sensor under test (SUT), has been tested on one site:

Test Site: Sodankylä (Finland)

Sensor Provider(s): The instrument evaluated was provided by the manufacturer (Vaisala)

Figure 2: Map of SPICE site testing PWD33 EPI instrument.

A summary on the configuration of the instrument as tested, the duration of tests and availability reflected in these results, and the ancillary measurements used, is available in Table 2, Table 3, and Table 4, respectively.

Table 2: Summary of instrument configuration and data output. Details and photos on site configuration are available in the site commissioning protocol.

	Sodankylä
Main prevailing wind directions	South
Sensor orientation	North
Height of installation	2.7 m
Heating	Yes, as recommended
Shield	No
Data QC	SPICE QC methodology
SUT data output frequency	15 sec
Data temporal resolution	1 min
Processing interval for SPICE data analysis	30 min

Table 3: Data availability, by measurement season.

Measurement season	**Sodankylä**
Season 1 (Oct. 2013 – Apr. 2014)	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓

Table 4: Summary of reference and ancillary measurements, with measurement height.

	Sodankylä
R2 Site Reference	OTT Pluvio^2 1500mm (DFAR) (4 m, rim height)
R2 Precip Detector	DRD11A (Site*, 2013-2014) (1 m) OTT Parsivel^2 (DFAR, 2014-2015) (2.7 m)
Ancillary Temp Sensor	Vaisala HMP155 (2 m)
Ancillary RH Sensor	Vaisala HMP155 (2 m)
Ancillary Wind Sensor	Thies acoustic 2D wind sensor (3.5 m)

**A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR-fence; however, in cases where a more sensitive detector is available outside of the DFIR-fence, or there are issues with the detector within the DFIR-fence, a precipitation detector elsewhere on the site can be employed.*

4. Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6 of the SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute intervals:

- Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip;
- SUT 'Yes': SUT accumulation > 0 mm;
- SUT 'No': SUT accumulation = 0 mm.

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores.

4.1.2. Skill score assessment

The ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

The above scores are computed using the formulations provided in Section 3.6 of the SPICE Final Report.

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

5. Environmental conditions

The environmental conditions at the site over the duration of the test period are expressed as probability density functions (PDFs) for air temperature, relative humidity, wind speed, wind direction, and precipitation rate in Figure 3. Both the entire period data, and the data corresponding only to the precipitation events reported by the site reference, R2, are shown. The precipitation percentage in Figure 3 represents the number of minutes of precipitation over a standard 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration.

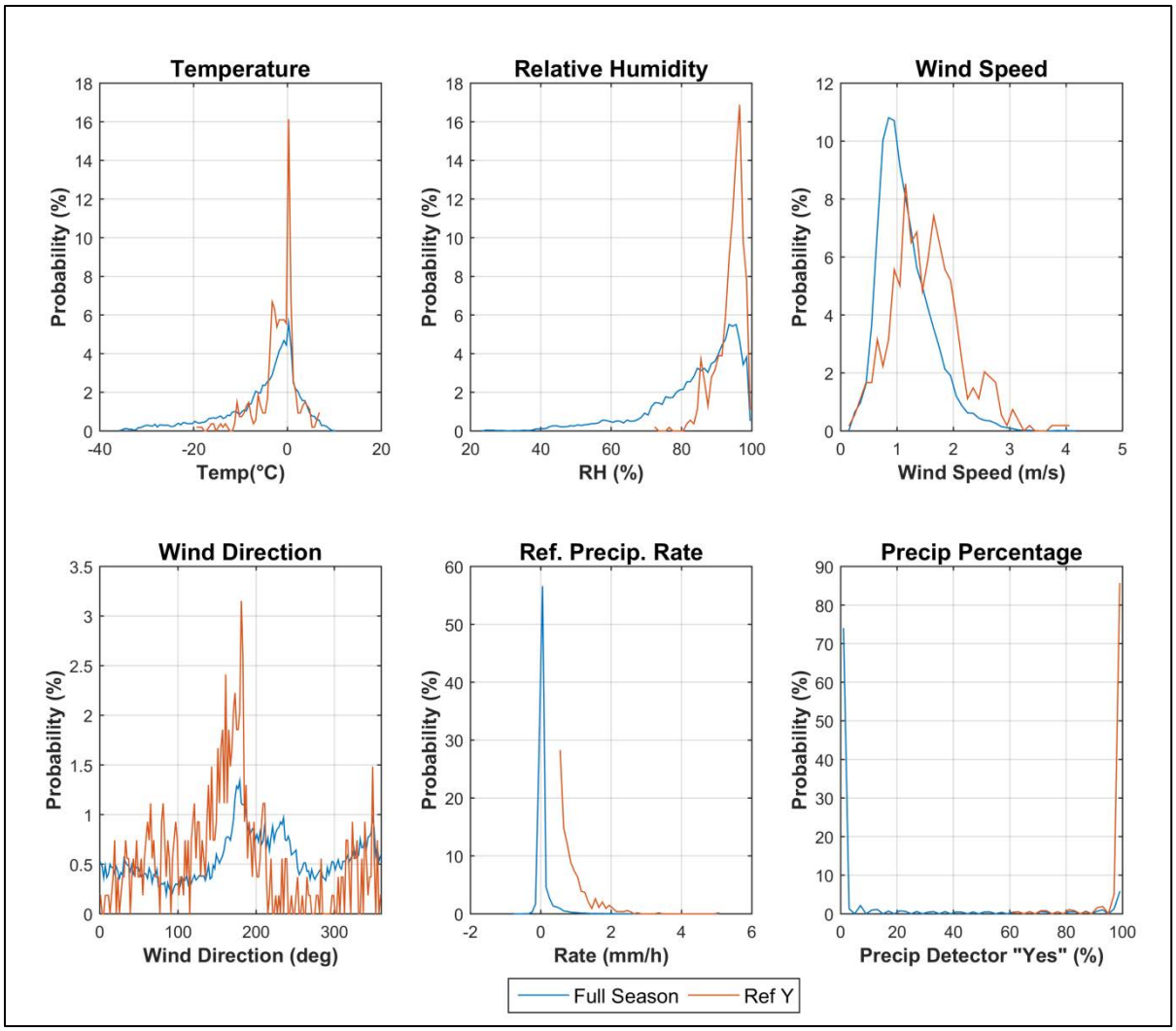


Figure 3: Summary of aggregated environmental conditions on the SPICE site that operated PWD33 EPI, over the entire duration ('Full Season', in blue) and for data corresponding to precipitation events, as reported by the site R2 reference ('Ref Y', in red), during the tests, as per Table 3 above.

6. Evaluation of performance over the range of operating conditions

6.1. Skill score assessment

The overall ability of the SUT to represent precipitation similarly to the site field reference R2, is assessed using contingency tables (Section 4.1.1) and derived selected skill scores (Section 4.1.2). To better understand the potential influence of threshold choices on the derived results, two cases are considered here (see note below). The contingency results related to these two cases are given in Table 5 and the respective skill scores in Table 6.

Note: Following the data derivation explained in Section 4.1.1, the conditions required to have a 'Yes' or a 'No' event over the 30 min interval, for the reference and the SUT, for the two different cases treated here, are:

CASE 1 *(as defined in Section 4.1.1)*:

- *Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip*
- *Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip*
- *SUT 'Yes': SUT accumulation > 0 mm*
- *SUT 'No': SUT accumulation = 0 mm*

CASE 2:

- *Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip*
- *Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip*
- *SUT 'Yes': SUT accumulation ≥ 0.1 mm*
- *SUT 'No': SUT accumulation < 0.1 mm*

Results presented in this report are based on Case 1.

Table 5: Contingency Tables: detection of precipitation of the PWD33 EPI relative to the field reference, expressed as number of events over the entire test period. The skill scores associated with these events are given in Table 6.

Case 1 Ref Plv2DFAR

SUT PWD33 EPI	Yes	No	Total
Yes	523	868	1391
No	0	9459	9459
Total	523	10327	10850

Case 2 Ref Plv2DFAR

SUT PWD33 EPI	Yes	No	Total
Yes	520	19	539
No	3	10308	10311
Total	523	10327	10850

Table 6: Skill Scores for the PWD33 EPI. POD: Probability Of Detection, FAR: False Alarm Rate, B: Bias, HSS: Heidke Skill Score (see Section 4.1.2 for more details).

PWD33 EPI, Skill Scores				
	POD	**FAR**	**B**	**HSS**
Case 1	100%	62.4%	266%	51.2%
Case 2	99.4%	3.53%	103%	97.8%

6.2. Assessment of SUT performance during non-precipitating events

The performance of the SUT in the absence of precipitation (when the reference precipitation detector recorded 30 minutes without precipitation) is represented in Figure 4 and Table 7, reflecting the distribution of the sensor response, as measured during the interval.

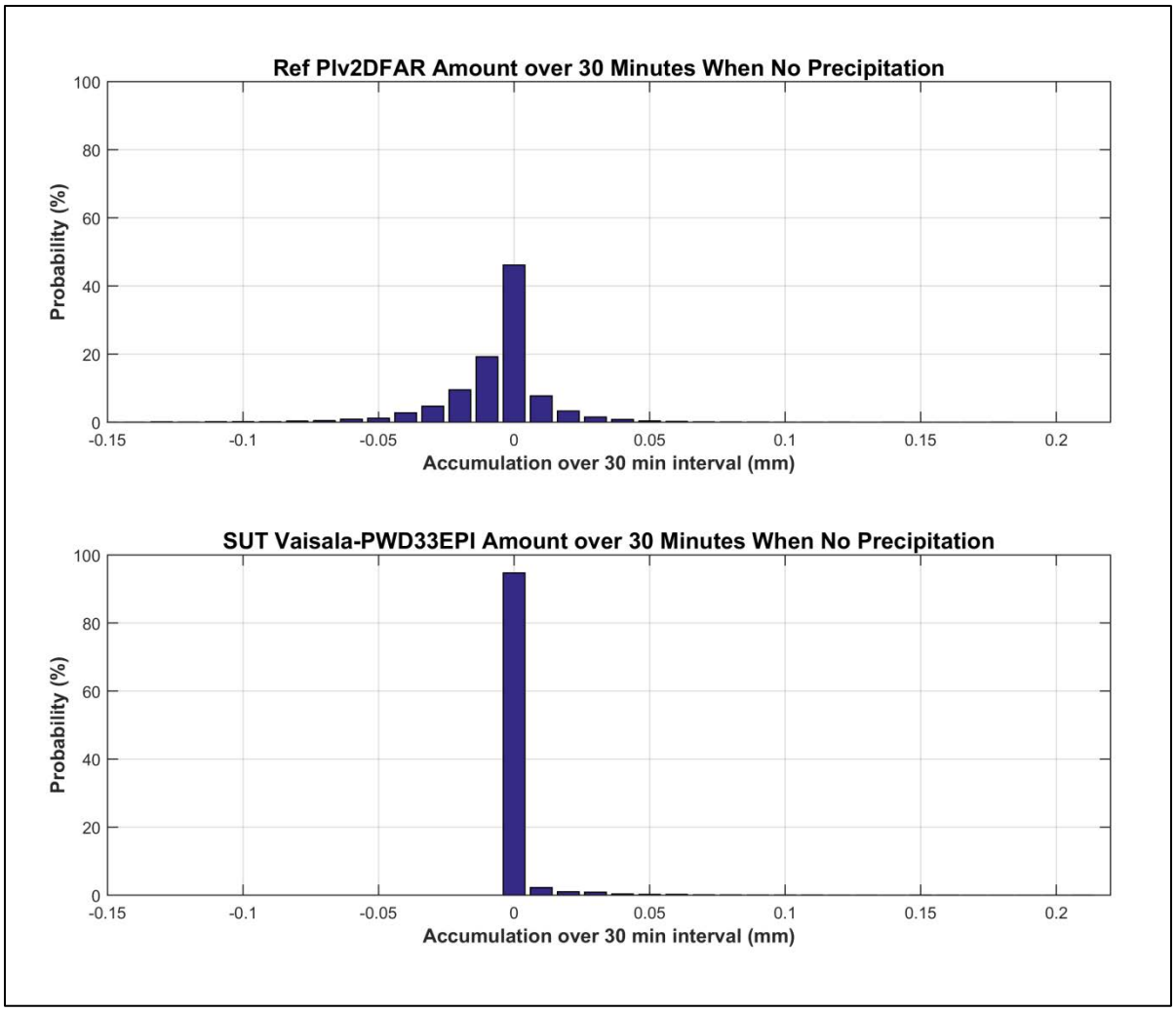


Figure 4: Probability of occurrence of a response during a 30 min interval in the absence of precipitation, represented by the signal output from (top) the R2 reference and (bottom) the SUT PWD33 EPI. Statistics associated with these graphs are given in Table 7.

Table 7: Reference and SUT statistics of response signal when no precipitation was occurring, as plotted in Figure 4; Average (Avg), standard deviation (STD), maximum (Max) and minimum (Min) of the response signal, together with the number of events (Num) over the test period is given.

No Precip Statistics – Reference: Plv2DFAR

	Ref Avg [mm]	Ref STD [mm]	Ref Max [mm]	Ref Min [mm]	Ref Num
Sodankylä	-0.003	0.020	0.190	-0.140	10335

No Precip Statistics – SUT: Vaisala-PWD33 EPI

	SUT Avg [mm]	SUT STD [mm]	SUT Max [mm]	SUT Min [mm]	SUT Num
Sodankylä	0.002	0.010	0.215	0.000	10335

6.3. Ability of the SUT to measure precipitation

6.3.1. Yes-Yes cases

Quantitatively, the performance of the SUT to derive and report precipitation is assessed relative to the site reference in several graphs and tables illustrated in this section, using only the cases where both instruments reported precipitation over the 30 min interval, according to the criteria used in Case 1 of Table 5 (cases 'Yes-Yes', or shorter 'YY').

6.3.1.1. Time series plots

The time series (cumulative sum of 30 min YY events accumulation) of the SUT is plotted against the reference for the two seasons, by precipitation type (see Section 4.1.4) in Figure 5.

The corresponding seasonal accumulations are given in Table 8.

Table 8: Seasonal accumulation [mm] for Sodankylä test site based on the sum of YY events from the SUT PWD33 EPI and the field reference R2.

Sodankylä				
	Season 2013/14		Season 2014/15	
[mm]	Ref	SUT	Ref	SUT
All events	124.49	125.27	113.77	137.57
Rain events	17.67	18.98	5.60	7.26
Mixed events	72.26	75.41	50.78	72.45
Snow events	34.56	30.88	57.39	57.86

6.3.1.2. Scatter plots and RMSE values

Scatter plot of the amount derived by the SUT versus the reference amount, and discriminated by precipitation type, is given in Figure 6.

Quantitatively, the SUT performance is assessed using the Root Mean Square Error (RMSE) also known in practice as Operational Comparability. The results are available in Table 9.

Table 9: Root Mean Square Error (RMSE) statistics, in mm, for the SUT PWD33 EPI with respect to the field reference, by precipitation type, including both seasons data.

RMSE [mm]	All	Rain	Mixed	Snow
Sodankylä	0.363	0.176	0.485	0.143

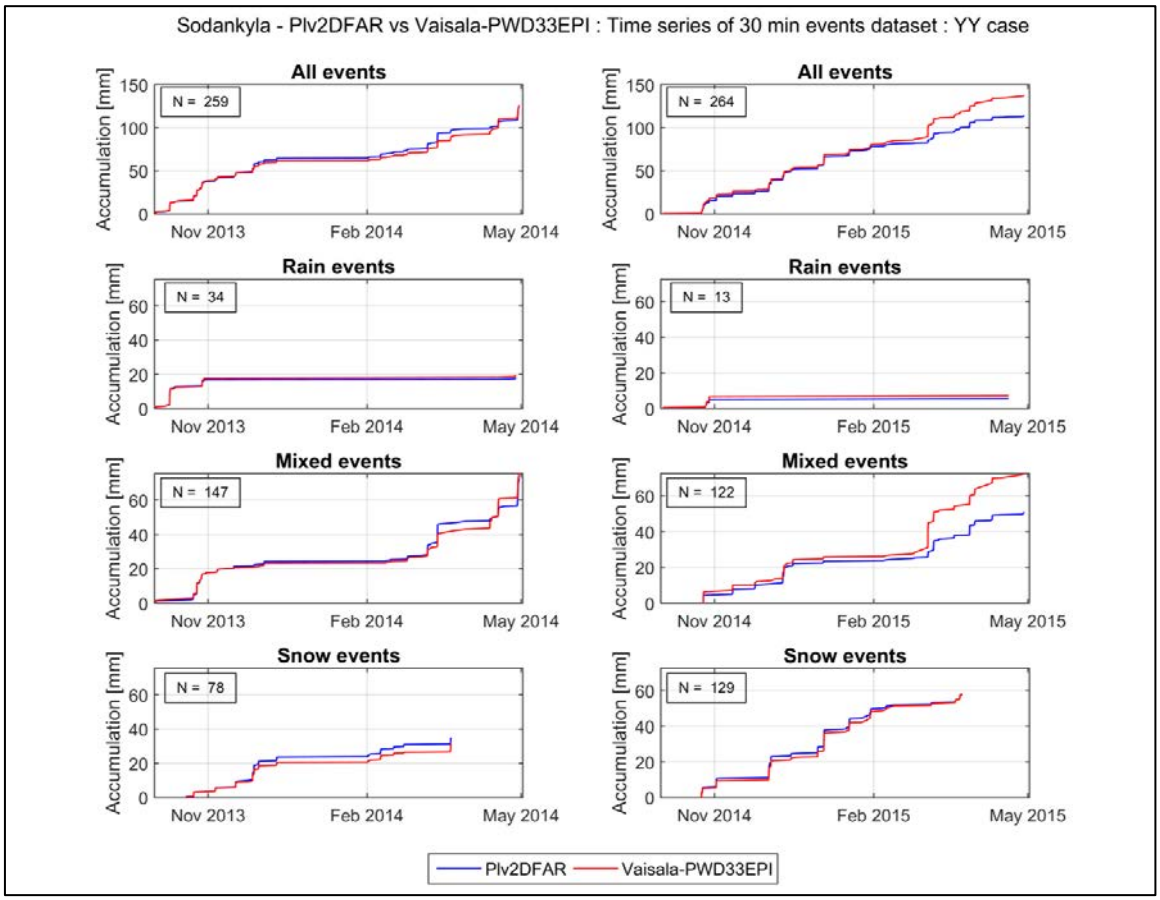


Figure 5: Time series based on 30 min YY events of the SUT PWD33 EPI against the field reference, discriminated by precipitation type (Rain, Mixed, Snow), for both seasons (2013/14 on the left, 2014/15 on the right).

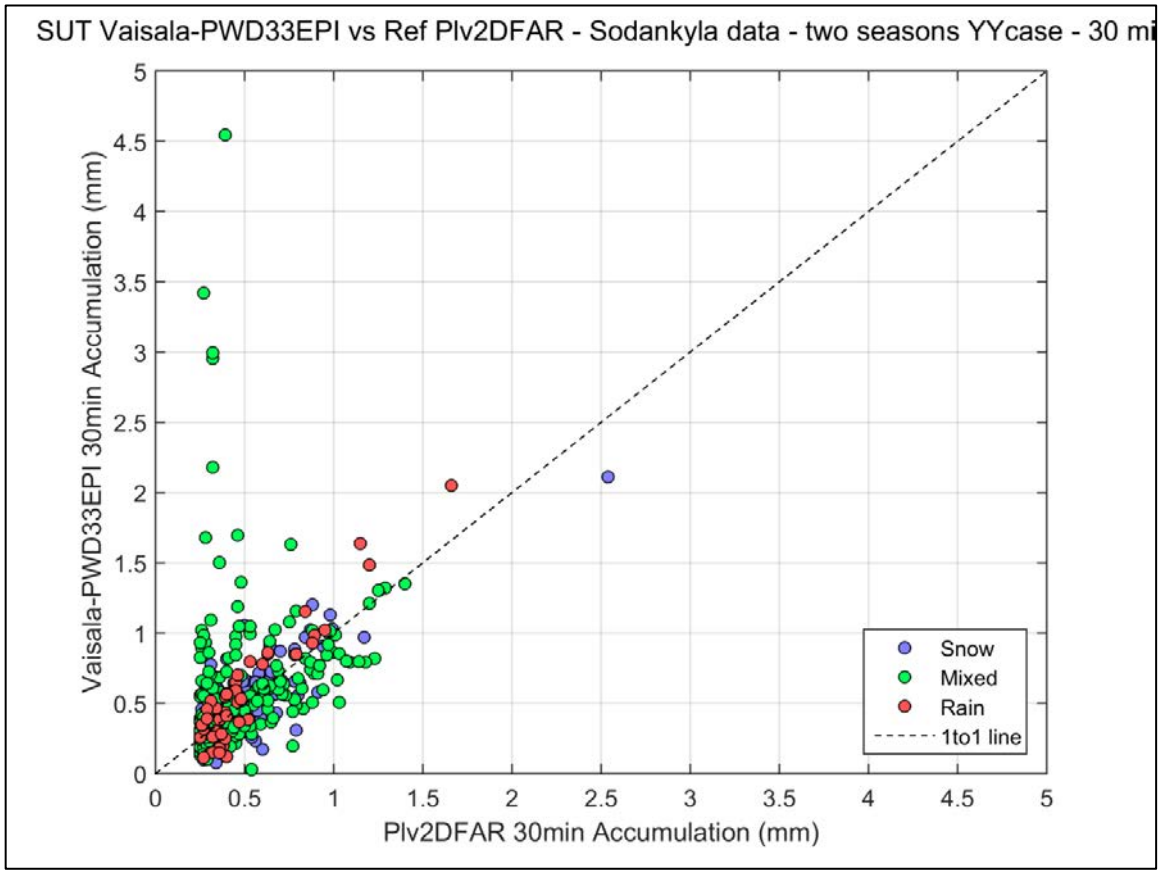


Figure 6: Scatter plot based on 30 min YY events accumulation from the SUT PWD33 EPI against the field reference, for the two seasons, discriminated by precipitation type.

6.3.1.3. Catch efficiency evaluation by precipitation type

The Catch Efficiency (CE) of the SUT is represented by histograms (Figure 7) and boxplots (Figure 8), both discriminated by precipitation type.

The quantitative evaluation of the CE is provided in Table 10. The mean catch efficiency is given in the first line of this table, considering both seasons data and for each category of precipitation type as well as for all the events together.

Note: All the events with a CE greater than 3, if any, are included in one category named '3 and more' or ' ≥3 ' in the upcoming graphs. Additionally, for all graphs representing the CE, a line is added at CE = 1, which represents the ideal case where the SUT reports exactly the same precipitation amount as the reference.

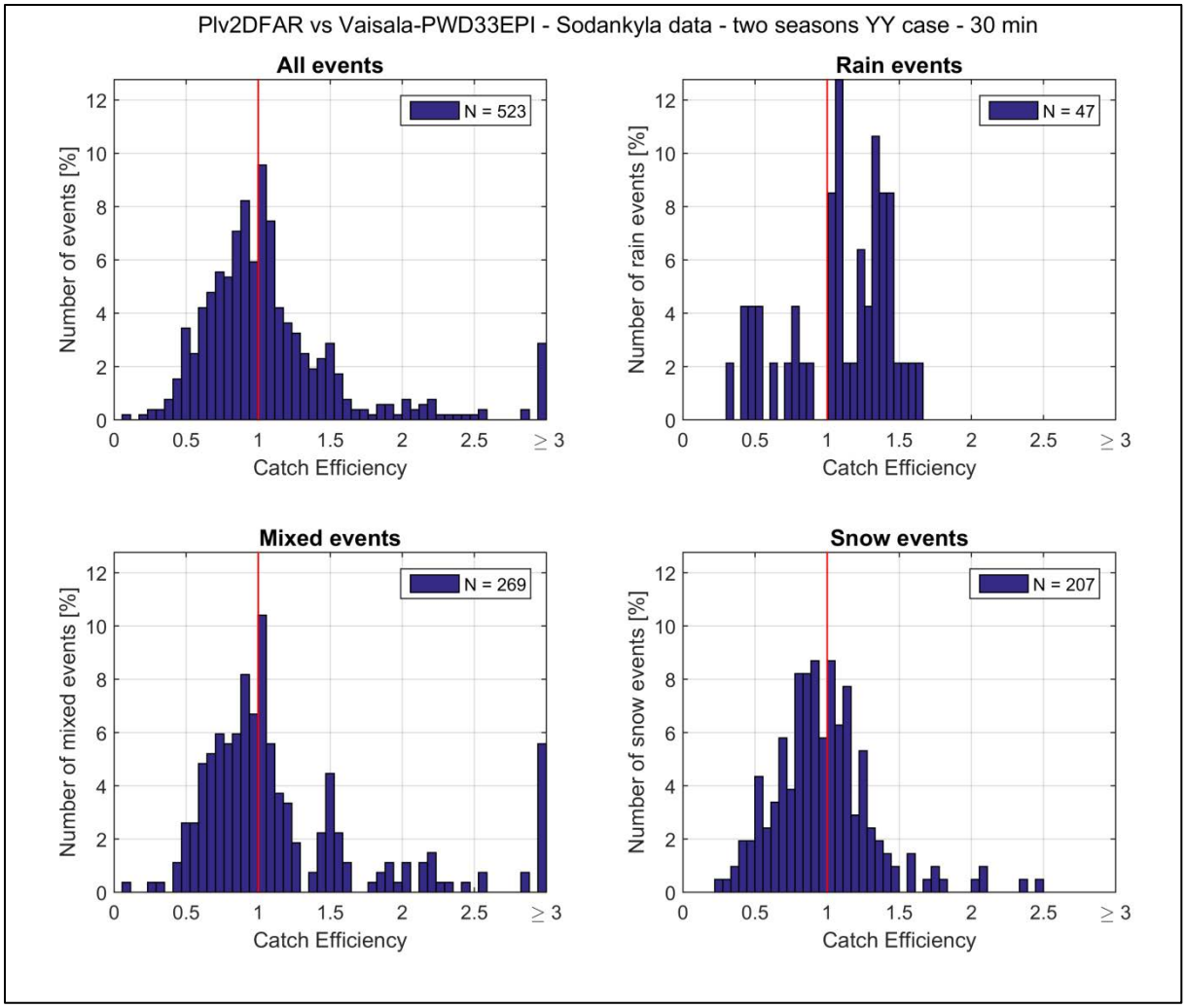


Figure 7: Histograms based on 30 min YY events for the two seasons, representing the distribution of the catch efficiency of the SUT PWD33 EPI against the field reference (SUT/Ref), discriminated by precipitation type, with number of events given in the legend for each category. The red line at CE = 1 represents the ideal case.

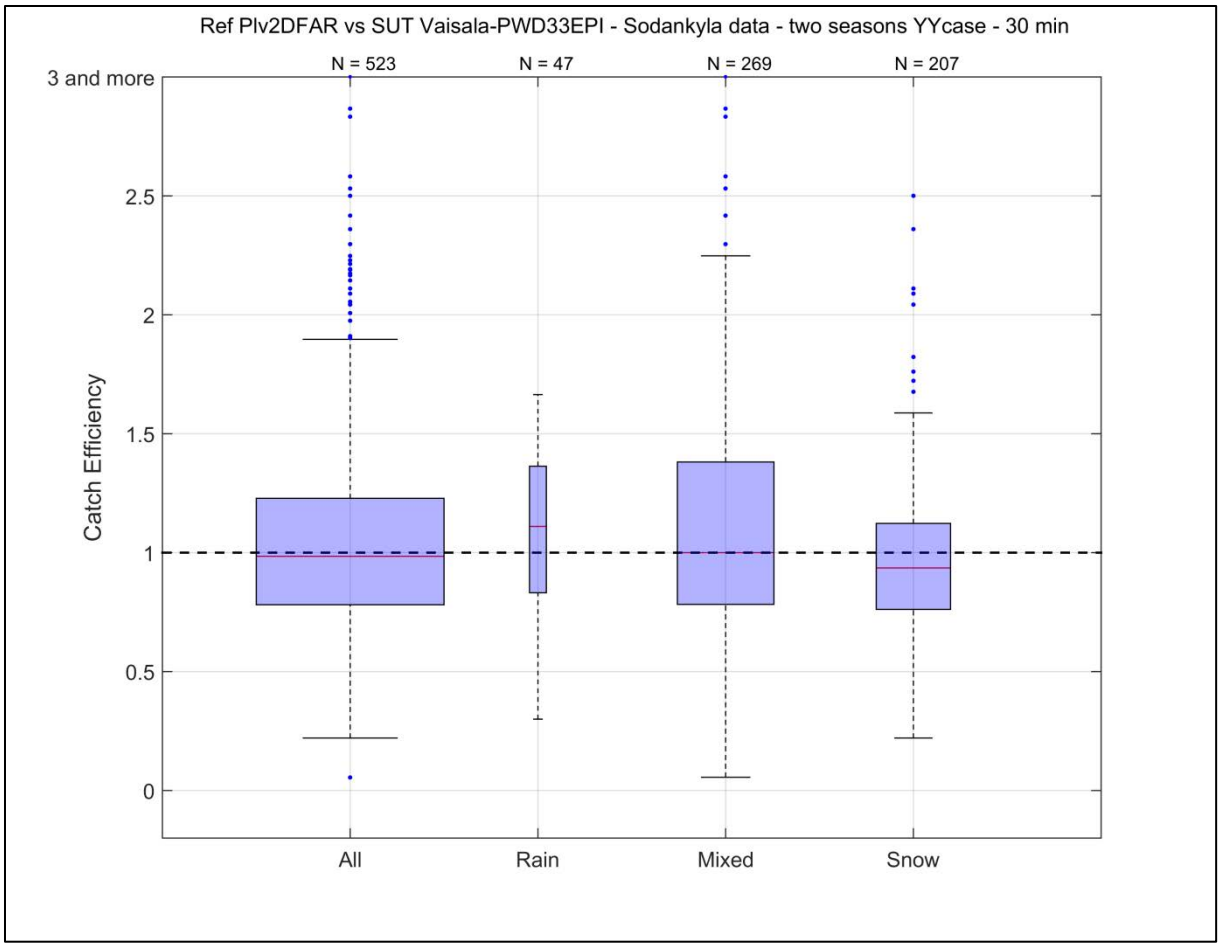


Figure 8: Boxplot based on 30 min YY events of the two seasons, representing the catch efficiency of the SUT PWD33 EPI against the field reference (SUT/Ref), discriminated by precipitation type, with number of events given at the top of each category. The width of the boxes is proportional to the percentage of events in each category ('All' box represents 100% data). The dashed black line at CE = 1 represents the ideal case.

Table 10: Statistics related to the CE of SUT PWD33 EPI against the field reference (SUT/Ref, see Figure 8), discriminated by precipitation type.

Catch Efficiency Statistics: **PWD33 EPI**

CE Boxplot Parameters	All	Rain	Mixed	Snow
Mean	1.17	1.09	1.34	0.97
Median	0.98	1.11	1.00	0.94
75 percentile	1.23	1.36	1.38	1.12
25 percentile	0.78	0.83	0.78	0.76
Upper Whisker	1.90	1.66	2.25	1.59
Lower Whisker	0.22	0.30	0.06	0.22
Maximum	12.67	1.66	12.67	2.50
Minimum	0.06	0.30	0.06	0.22
# Outliers	41	1	21	9
# Outliers ≥ 3	15	0	15	0
# Events	523	47	269	207

6.3.1.4. Catch efficiency dependency on wind speed

The variation of the SUT catch efficiency (SUT/Ref) with wind speed is illustrated in a scatter plot together with a boxplot in Figure 9, both discriminated by precipitation type.

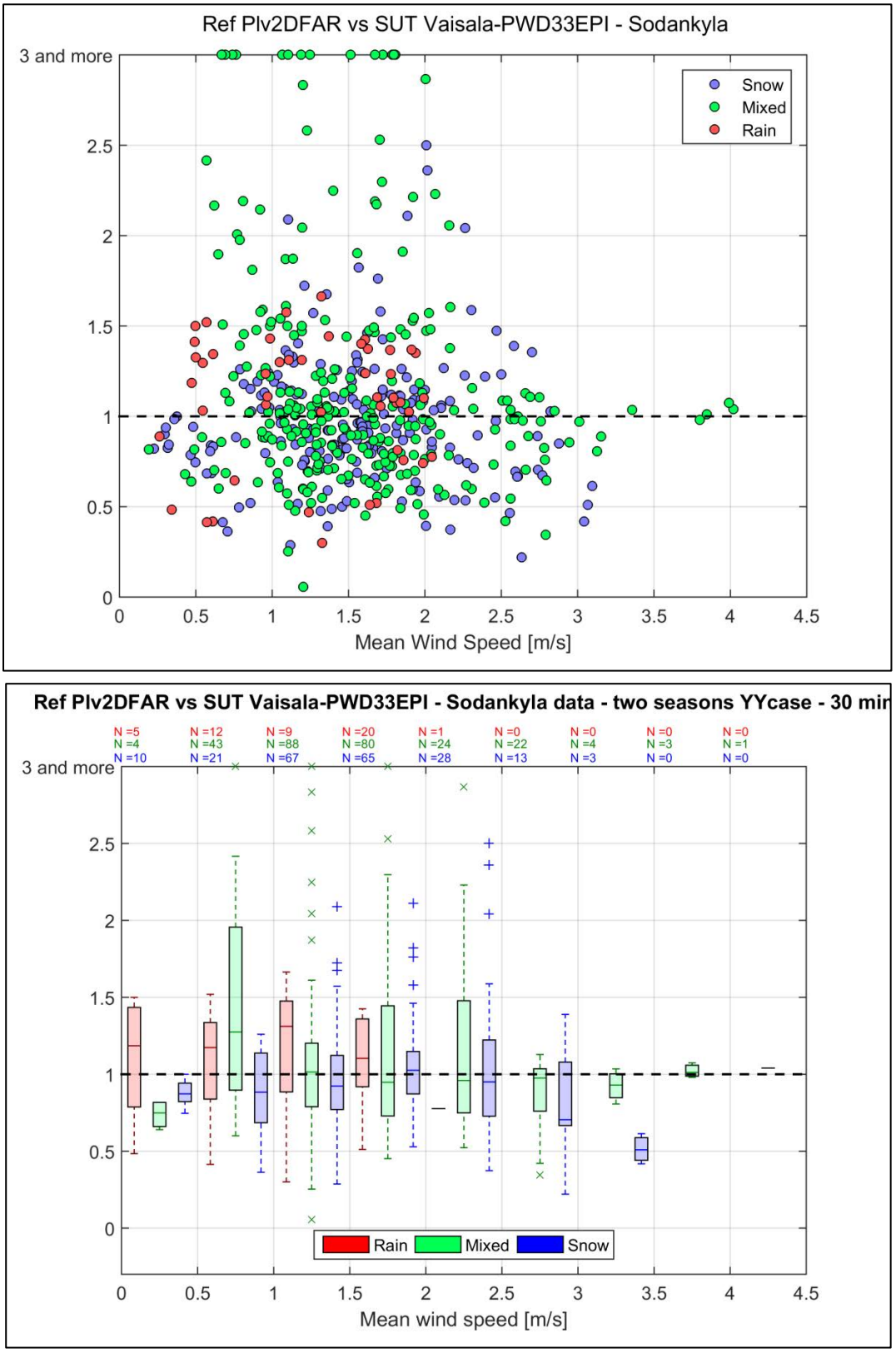


Figure 9: Scatter plot (top) and boxplot (bottom) based on 30 min YY events for the two seasons, representing the catch efficiency of the SUT PWD33 EPI with respect to the field reference (SUT/Ref), against wind speed and discriminated by precipitation type. The dashed black line at CE = 1 represents the ideal case.

6.3.1.5. Catch efficiency dependency on wind direction

In order to assess the dependency of the CE with wind direction, a wind rose is produced (Figure 10) representing the wind data of the two seasons, binned by catch efficiency in order to represent undercatch (CE < 0.8), overcatch (CE > 1.2) and catch efficiency of 1 ± 20% of the SUT.

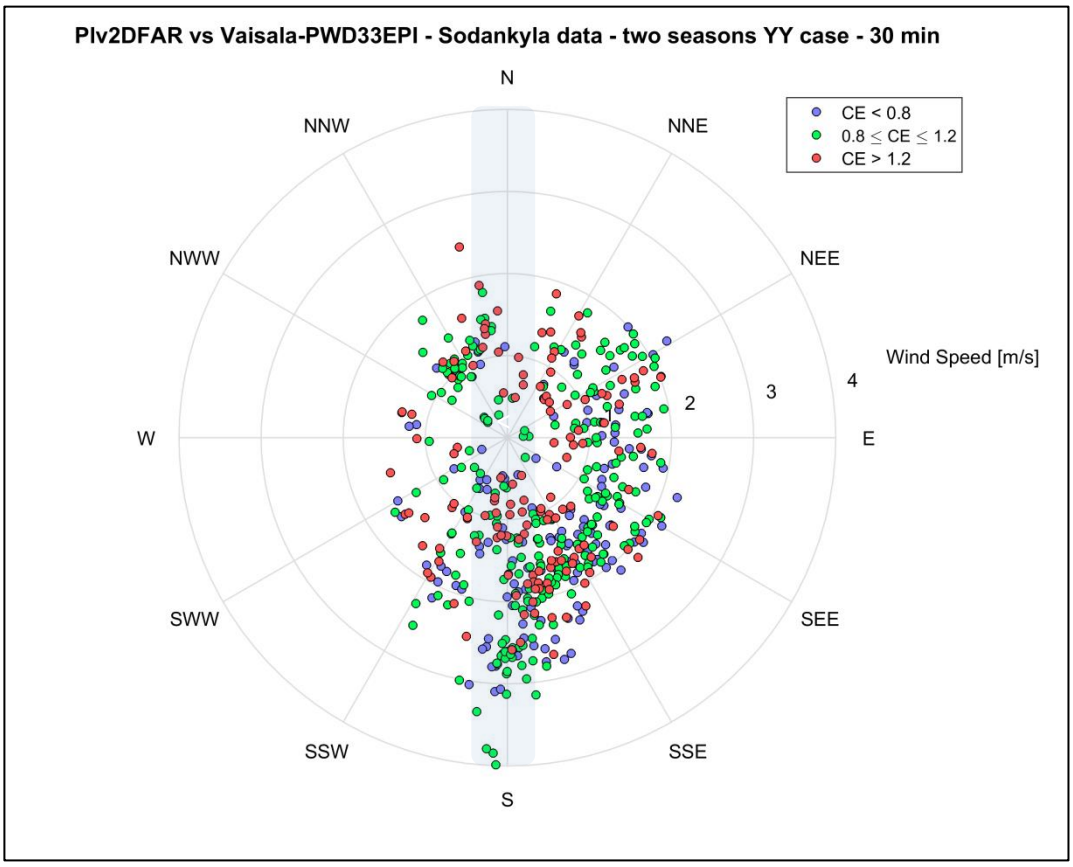


Figure 10: Precipitation events (YY cases) as function of wind speed and direction, and binned by catch efficiency (CE). The grey zone indicates the SUT orientation.

6.3.2. Yes-No and No-Yes cases

Events when the site reference and SUT do not agree on the occurrence of precipitation includes two categories of cases (Section 4.1.1): (1) when the field reference reported a precipitation event, while the SUT did not (Yes-No cases, 'YN'), and (2) when the field reference did not report a precipitation event while the SUT did (No-Yes cases, 'NY').

Histograms illustrating field reference and SUT reports and associated site conditions for all NY cases (no YN cases were reported for the PWD33 EPI) during the test period are provided in Figure 11.

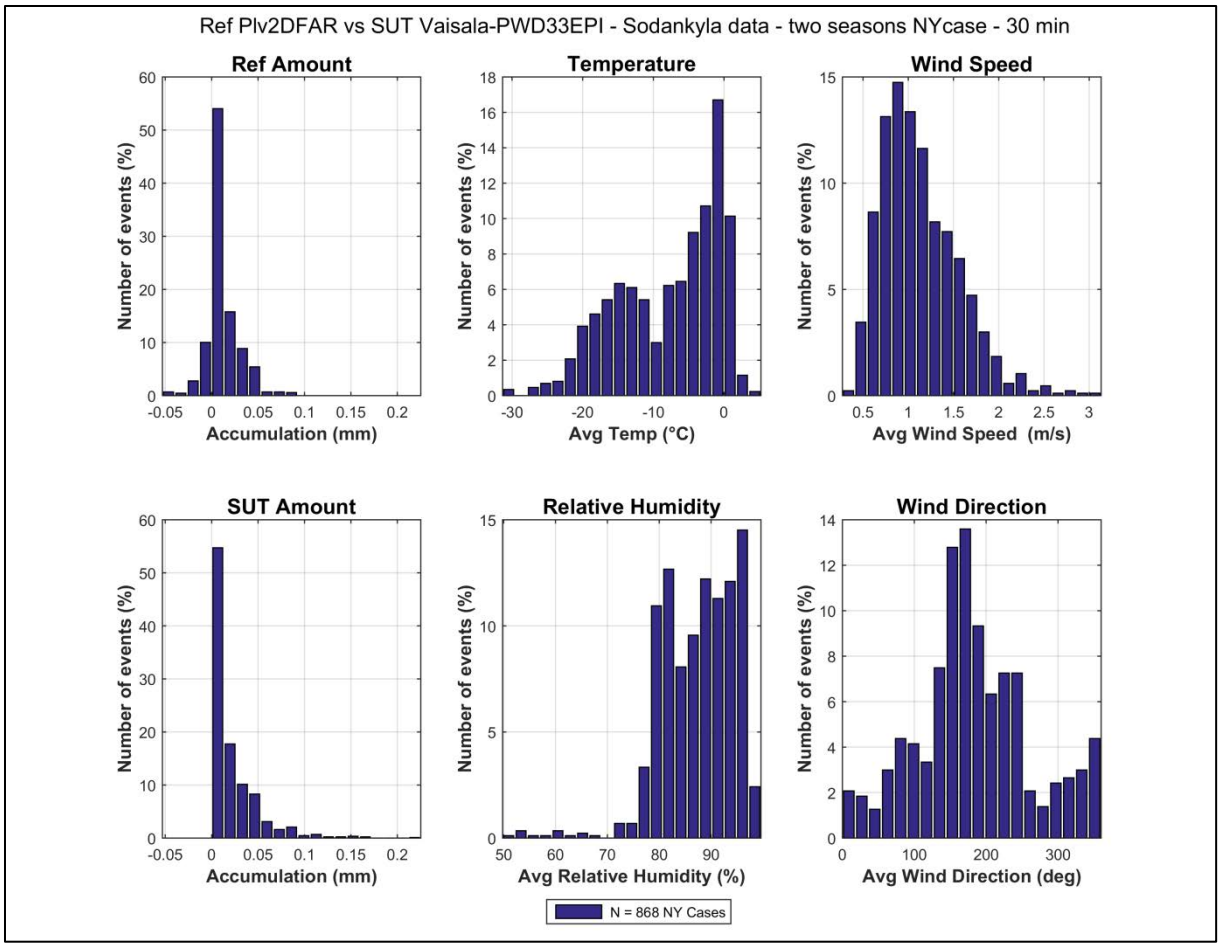


Figure 11: Histograms of SUT PWD33 EPI and field reference accumulations (left column), along with distributions of mean temperature, mean relative humidity, mean wind speed, and mean wind direction for all NY cases (number indicated in the legend) of the 30 min intervals, as reported by Case 1 of Table 5.

7. Interpretation of results

7.1. Operating conditions

Sodankylä is a site characterized by very low wind speeds, as shown in Figure 3, with a majority of precipitation events occurring below 3 m/s. Regarding temperatures, Sodankylä recorded most of the precipitation events at temperature between -10 and 0°C, with extreme by -20°C. The site shows fairly humid climate conditions as expected from a Northern Boreal climate site.

7.2. Reliability in detecting precipitation

The Vaisala PWD33 EPI was efficient in detecting precipitation when the reference reported precipitation, as presented in Table 6, with a POD of more than 99%, looking at the two different threshold cases. When no threshold is applied to the SUT (Case 1), the high FAR of 62%, the bias of 266% and the HSS of 51% are a direct consequence of the large number of NY events (see Section 7.5 for more details). However, it can be noted that the FAR score is reduced to less than 4%, and the bias and HSS scores improved, when a threshold of 0.1 mm over 30 min events is applied to the SUT (Case 2), indicating that the majority of the NY events of Case 1 are for very low accumulations. This could be the consequence of a difference in sensitivity between the weighing gauge used as the reference and the non-catchment type SUT, related to their different measurement principle.

7.3. Performance of SUT during no-precipitation events

The PWD33 EPI has a stable output signal, showing almost no noise during no-precipitation events (see Figure 4 and Table 7). Note that the no-precipitation study is performed using only the reference precipitation detector data (here a DRD11A for the first season and a Parsivel^2 for the second season) with the criteria ensuring that 30 min of 'no-precipitation' was recorded. For these cases, the PWD33 EPI indicates 0 mm for more than 92% of the time, indicating a good agreement with the reference precipitation detector on the absence of precipitation. Most of the remaining cases (where the reference precipitation detector doesn't detect any precipitation while the SUT does) correspond to very low accumulations reported by the SUT (less than 0.1 mm/30 min) and can be related to the NY events mentioned above and analyzed in Section 7.5.

7.4. Performance of SUT during precipitation events

When both the reference and the PWD33 EPI report precipitation (YY cases), the overall results show that the sensor behaves slightly differently according to precipitation types and especially for mixed events.

The RMSE in Table 9 shows almost identical values for rain and snow precipitations, with 0.18 and 0.14 mm, respectively, while the mixed events show a higher RMSE value of 0.49 mm. This higher variability for mixed events (mainly overcatch) is illustrated both in time series of Figure 5 (especially for second season) and as points with high SUT accumulation in Figure 6.

These specific cases have been analyzed more thoroughly by the manufacturer (Vaisala document shared on November 27, 2015, following discussions in Brussels), using time series and SYNOP code outputted by the PWD33 EPI and by the other two Vaisala sensors tested on site (PWD52 and FS11P). The outcomes showed that these events were not correctly categorized by the PWD33 EPI, which reported rain instead of snow, leading to a bigger liquid water content measurement and thus to an overcatch compared to the reference (corresponding, in Figure 7Figure 9, to the events with a catch

efficiency greater than 3). According to the manufacturer, this misinterpretation is related to events with temperature between 0 and 6°C (end of season 2014/15), for which the precipitation type differentiation algorithm was changed prior to SPICE. Indeed, following a previous comparison of this instrument with manual observations, it was observed that the PWD33 EPI reported snow instead of rain over that range of temperature. The algorithm parameters were then changed accordingly, but this caused the opposite situation during the SPICE campaign, described above. As a conclusion to their study, the manufacturer noted that the original algorithm (used for PWD52 and FS11P) provides a more reliable solid-liquid differentiation and recognized that the PWD33 EPI algorithm needs to be harmonized with PWD52 and FS11P with respect to the liquid-solid precipitation differentiation.

From Figure 6, it can be seen that the rain events with accumulation greater than 0.5 mm show a slight overcatch. Despite the overcatch for rain and mixed precipitation influencing the overall result, due to some particular events, the PWD33 EPI has a slight tendency to undercatch for snow, as represented by Figure 7 andFigure 8 and Table 10, with a mean overall catch efficiency over the two seasons of 1.17, dropping to 0.97 for snow.

Overall and without the differentiation issue mentioned above, the PWD33 EPI seems to be a reliable instrument to account for the total accumulation over a longer period (e.g. one season), with a mean catch ratio very close to the 1-line, as shown in Figure 8 or Figure 9 (bottom). The relatively high scatter visible in Figure 9 (top), with catch ratio varying randomly from 0.3 to 2.5, on event based statistics (typically 30 min interval), occurring at all wind speeds (up to 4 m/s, maximum wind speed in Sodankylä), tends to show that the PWD33 EPI is less reliable to derive solid and mixed precipitation accumulation over near real time periods.

Even if better in deriving accumulation for longer time periods, the PWD33 EPI, as other non-catchment type instruments, doesn't have an assurance of the continuity in the measurements, though, which is critical to long term data collection. If power is off or signal transmission is interrupted, no data is recorded, thus affecting the long term data reports of the PWD33 EPI.

The results in Figure 9 indicate no clear dependency of the catch efficiency with increasing wind speed, for all precipitation types. The performance of the sensor for higher wind speed cannot be assessed, since Sodankylä is a site with low wind speed (up to 4 m/s), and the SUT was not tested in other environmental conditions.

The wind rose (see Figure 10) shows no dependency of the catch efficiency with wind direction. The sensor was installed parallel to the prevailing wind direction during precipitation events. An assessment with a sensor perpendicular to the main flow has not been assessed, though, since no such configuration was tested.

7.5. Assessment of Yes-No, No-Yes events

The assessment of YN, NY events completes the picture of the performance of the SUT. As it can be seen in Table 5 and Figure 11, when no threshold is applied to the SUT (Case 1), the PWD33 EPI shows a large amount of NY cases, but no YN cases were reported, the latter indicating the good reliability of the sensor to detect precipitation whenever the reference reported precipitation. When a threshold is applied to the SUT (Case 2) - here a threshold of 0.1 mm/30 min – the number of NY events is significantly reduced (19 NY events compared to the 868 NY events without any threshold).

Therefore, applying a threshold enables to ensure the good agreement of the PWD33 EPI with the reference, in all cases (YN, NY, YY, NN), and prevents the user to deal with NY events of unknown origin (either real or artefacts).

Looking more closely at Case 1 and according to Table 5, a little less than 700 NY events occurred during the two seasons of the experiment. It could be useful to remind that these NY events are characterized, over the 30 minutes of an event, by an accumulation from the reference below the defined threshold of 0.1 mm, the reference precipitation detector having recorded 0 min of precipitation, and the SUT indicating more than 0 mm (see Section 4.1.1). This high number of NY events appears to happen for very low accumulations (mostly less than 0.1 mm/30 min) reported by the PWD33 EPI and for low temperatures and wind speeds (see Figure 11).

The NY events occurred primarily in the second season (698 for season 2 compared to 170 for season 1), where the precipitation detector being part of the reference was changed from the DRD11A capacitive sensor by a Parsivel^2 (see Table 4). Figure 12 shows some of the NY events of season 2 over the accumulation curve of the R2 reference weighing gauge, together with the precipitation detection measured by the Parsivel^2, the DRD11A and the PWD33 EPI. It can be noted that this light precipitation was well detected by the DRD11A and the PWD33 EPI, but nothing was measured by the Parsivel^2. This led to the idea that the NY results could change if the choice of the reference precipitation detector was different for the second season.

To better address these differences in precipitation sensitivity and to give a sense of the results dependency on the selected reference precipitation detector, a second study has been made using the DRD11A as the reference precipitation detector for the second season as well (instead of the Parsivel^2). The same number of YY and YN events was found, but the number of NY events was reduced (380 against 868 with the Parsivel^2), leading to a FAR of 42.1% (against 62.4% with the Parsivel^2), a bias of 173% and an HSS of 71.7%. Even if capacitive precipitation detectors, such as the DRD11A, are less sensitive than laser-based precipitation detectors, like the Parsivel^2 (see Section 4.1.3.5.2 of the SPICE Final Report), the DRD11A seems to better match the PWD33 EPI results in terms of precipitation detection (as it could be seen in Figure 12). A similar graph to Figure 11, but this time using the DRD11A as the reference precipitation detector for both seasons, is given in Figure 13 for the PWD33 EPI NY events. This figure shows a wider distribution over the wind speed range, but a distribution of temperatures closer to 0°C. It is still possible that some of these remaining 380 NY events are real light events that were not detected by the DRD11A (which measured 30 min of no precipitation for all of them), as suggest by Figure 14, where a light precipitation is reported by the PWD33 EPI and not by the DRD11A, leading to NY events that are actually real light events. Extending the fact that all the 380 NY events are real events would not be reasonable, though, since they remain uncertain and no other parameters or photos are available to confirm or infirm this statement.

To conclude, the PWD33 EPI NY events of either study (with DRD11A at season 1 and Parsivel^2 at season 2 or only DRD11A for both seasons) reflect therefore rather a reliable higher sensitivity of the PWD33 EPI compared to the reference system. Indeed, on one hand the reference precipitation detector could have a lower sensitivity than the PWD33 EPI, on the other hand the reference weighing gauge has a different physical principle of measurement implying an expected lower sensitivity than the PWD33 EPI.

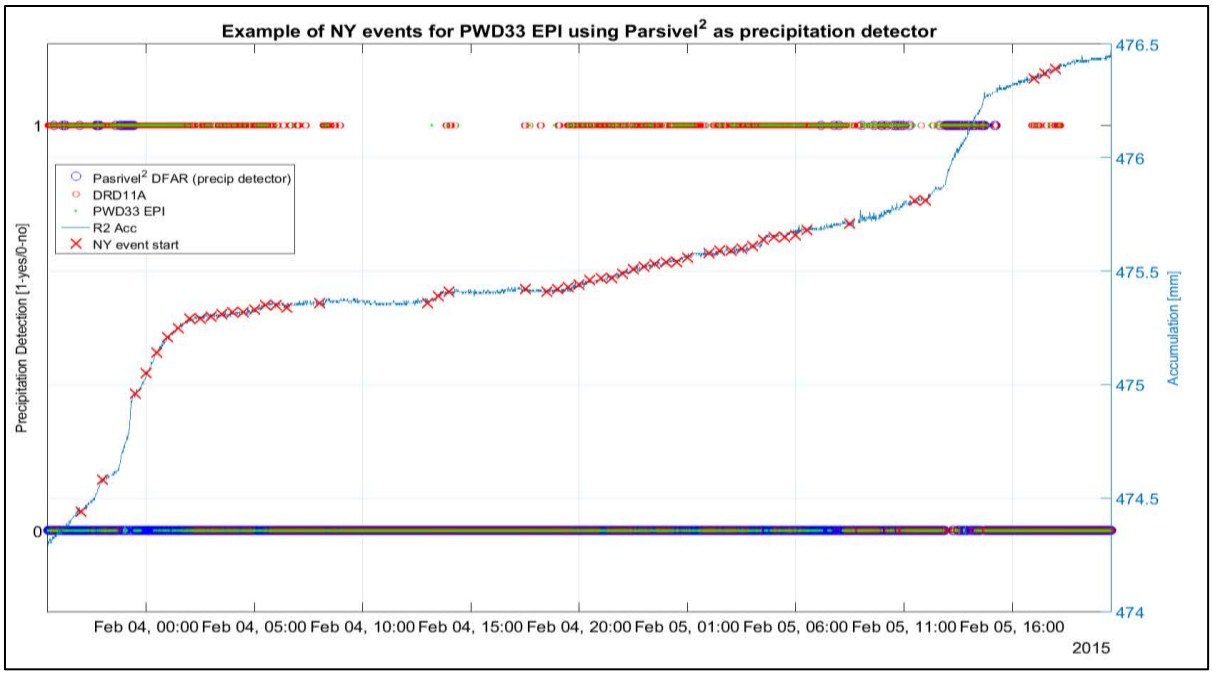


Figure 12: Example of NY events for the SUT PWD33 EPI using the Parsivel^2 as the reference precipitation detector for the second season.

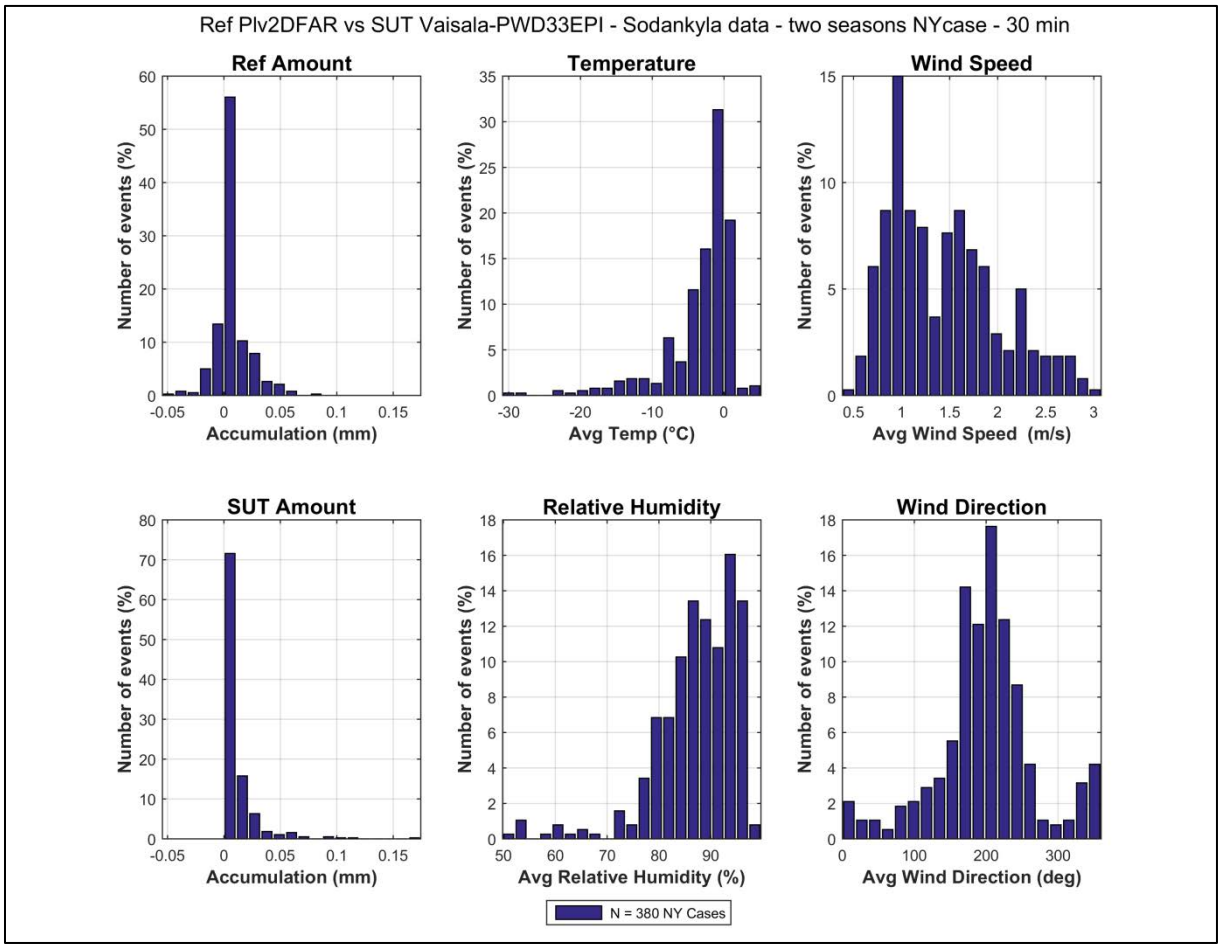


Figure 13: Histograms of NY events similar to results of Figure 11, but using a DRD11A as the reference precipitation detector (instead of the Parsivel^2) for the second season.

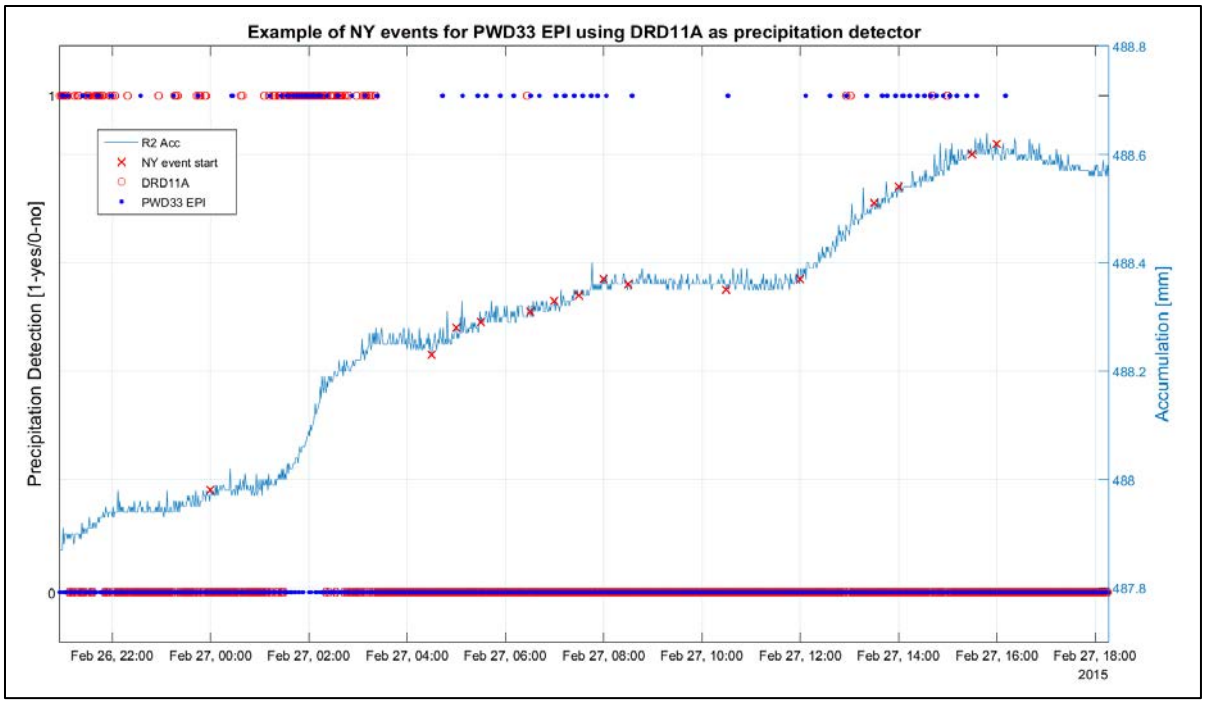


Figure 14: Example of NY events for the SUT PWD33 EPI using the DRD11A as the reference precipitation detector for the second season.

7.6. Threshold selection

The threshold to be set for the PWD33 EPI in order to report precipitation adequately over a 30 min interval (3 STD, according to the methodology defined in Section 3.6.1.3.2 of the SPICE Final Report) corresponds to 0.03 mm (3 x 0.01 mm, from Table 7). Additionally, as shown in the contingency tables (Table 5), a threshold of 0.1 mm/30 min enables to get rid of events whose origin is not confirmed or well established.

8. Operational considerations

The overall experience with the PWD33 EPI at Sodankylä was positive. The sensor appeared to be robust operationally with almost no data breaks due to the instrument (there was one major data gap of $1^{1/2}$ month in December 2013, but this was due to an issue with the data collection program).

During the experiment, some snow accumulation has been observed on the instrument device and mounting parts (see Figure 15). This is expected to happen on sites with calm weather conditions, as the Sodankylä site. According to the manufacturer, the probability that blowing snow from these accumulations could be counted as precipitation is very unlikely, though, because it would require that the precipitation detector 'RAINCAP' surface and the optical measurement volume, both part of the PWD33 EPI, are hit by the snow particles at the same time and with sufficient strength.

Figure 15: Snow accumulation on PWD33 EPI device, 2015- 02-15, 11:02 UTC.

8.1. Maintenance

The PWD33 EPI requires almost no maintenance. According to the User Manual, the sensor's lenses and hoods and the PWR211 rain sensor and PWS111 impact sensor sensing surfaces should be cleaned every six months. Before using it operationally, though, the User Manual indicates also that an individual calibration on site may be needed to derive correctly the liquid water content of precipitation.

9. Performance considerations

- When no threshold is applied to the SUT accumulation (Case 1), a high number of false alarms events (NY) were reported by the PWD33 EPI when comparing with the reference. These seem to be mainly due to the higher sensitivity of the SUT compared to the reference, but it remains that it has not been completely understood if some are real or artefacts. Applying a threshold (of 0.1 mm/30 min for instance in Case 2) to the PWD33 EPI accumulation output, though, significantly reduces the FAR and ensures to derive reliable precipitation events.
- The issue about liquid-solid differentiation algorithm for specific temperature conditions, explained in Section 7.4, is another performance consideration to mention for the PWD33 EPI sensor. The manufacturer has noted and recognized that issue and is intending to fix it.
- High scatter of CE on an event basis (30 min), but good mean CE over the seasons, makes the PWD33 EPI more suitable for deriving solid precipitation measurements over long periods of time, giving the condition that the sensor operates continuously.

WMO-SPICE Instrument Performance Report
Vaisala PWD52

1. Technical specifications

Instrument model:	Vaisala PWD52
Sample volume:	0.1 L
Physical principle:	Sensor combining the functions of a forward scatter visibility meter and a present weather sensor. It can measure the intensity and amount of both liquid and solid precipitation.
Operating temperature range:	-40 to +60°C
Sensitivity:	0.05 mm/h or less, within 10 min

Note: Specifications from manufacturer provided documentation.

Figure 1: PWD52 at Sodankylä test site.

2. Data output format

The Vaisala present weather sensor PWD52 provides precipitation information as intensity, accumulation or weather code. Table 1 summarizes the main output parameters from the instrument. The firmware version of the PWD52 tested during this experiment was the version V2.05.

Table 1: Summary of main instrument outputs, as recorded by the site during the experiment.

Measured Parameters	Units
Water Intensity	[mm/h]
Cumulative Water Sum	[mm]
Cumulative Snow Sum	[mm]
Visibility	[m]
Present Weather Code (WMO)	-
Present Weather Code (NWS)	-
Present Weather Code (METAR)	-
Housekeeping	-

This document reports on the ability of the PWD52 to derive solid precipitation. Although the 'Cumulative Snow Sum' parameter could seem appropriate to derive the results, it is actually not meant to give a water content measurement, but rather a coarse estimation of the new snow accumulation, i.e. new snow depth (see PWD52 User Manual). The analysis has been performed using the 'Cumulative Water Sum' parameter, which gives the water content of snow and is therefore the suitable parameter to compare fairly the sensor with the field reference.

3. SPICE test configuration

The PWD52, as sensor under test (SUT), has been tested on one site:

Test Site: Sodankylä (Finland)

Sensor Provider(s): The instrument evaluated was provided by the manufacturer (Vaisala)

Figure 2: Map of SPICE site testing PWD52 instrument.

A summary on the configuration of the instrument as tested, the duration of tests and availability reflected in these results, and the ancillary measurements used, is available in Table 2, Table 3, and Table 4, respectively.

Table 2: Summary of instrument configuration and data output. Details and photos on site configuration are available in the site commissioning protocol.

	Sodankylä
Main prevailing wind directions	South
Sensor orientation	North
Height of installation	2.7 m
Heating	Yes, as recommended
Shield	No
Data QC	SPICE QC methodology
SUT data output frequency	15 sec
Data temporal resolution	1 min
Processing interval for SPICE data analysis	30 min

Table 3: Data availability, by measurement season.

Measurement season	**Sodankylä**
Season 1 (Oct. 2013 – Apr. 2014)	✓
Season 2 (Oct. 2014 – Apr. 2015)	✓

Table 4: Summary of reference and ancillary measurements, with measurement height.

	Sodankylä
R2 Site Reference	OTT Pluvio^2 1500mm (DFAR) (4 m, rim height)
R2 Precip Detector	DRD11A (Site*, 2013-2014) (1 m) OTT Parsivel^2 (DFAR, 2014-2015) (2.7 m)
Ancillary Temp Sensor	Vaisala HMP155 (2 m)
Ancillary RH Sensor	Vaisala HMP155 (2 m)
Ancillary Wind Sensor	Thies acoustic 2D wind sensor (3.5 m)

**A sensitive precipitation detector is a required component of the SPICE R2 reference configuration. Ideally, the precipitation detector should be located within the DFIR-fence; however, in cases where a more sensitive detector is available outside of the DFIR-fence, or there are issues with the detector within the DFIR-fence, a precipitation detector elsewhere on the site can be employed.*

4. Assessment approach

4.1. Methods

Readers are encouraged to review the methodology used for the assessment of the sensor under test relative to the reference detailed in Section 3.6 of the SPICE Final Report. Elements of the methodology that are critical to the interpretation of results in this report are summarized below.

4.1.1. Data derivation

The assessment data are derived over 30 minute intervals (unless otherwise specified) and predicated on the detection of precipitation by the site reference R2 ('Ref') and the SUT. Precipitation detection is considered in terms of the following 'yes' (Y) or 'no' (N) conditions for the reference and SUT over 30 minute intervals:

- Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip;
- Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip;
- SUT 'Yes': SUT accumulation > 0 mm;
- SUT 'No': SUT accumulation = 0 mm.

For a given assessment interval, there are four possible detection contingencies: Ref 'Yes', SUT 'Yes' (YY); Ref 'Yes', SUT 'No' (YN); Ref 'No', SUT 'Yes' (NY); Ref 'No', SUT 'No' (NN). The numbers of events in each contingency are used in the computation of skill scores.

4.1.2. Skill score assessment

The ability of the SUT to detect the occurrence of precipitation relative to the site field reference R2 is expressed using selected skill scores:

- *Probability of Detection (POD):* percentage of the total number of 'Yes' events identified by the reference that are also identified as precipitation events by the SUT (ideal value = 100%);
- *False Alarm Rate (FAR):* percentage of the total number of 'Yes' events reported by the SUT that are not identified as precipitation events by the reference (ideal value = 0%);
- *Bias (B):* percentage of total SUT 'Yes' events relative to total reference 'Yes' events (ideal value = 100%, for which the SUT detects the same number of 'Yes' events as the Ref);
- *Heidke Skill Score (HSS):* percentage that considers the number of correct 'Yes' and 'No' events from the SUT relative to the reference, accounting for the number of expected correct responses due to chance alone (a sensor that is always correct has a value of 100%, while a sensor with no skill has a value of 0%).

The above scores are computed using the formulations provided in Section 3.6 of the SPICE Final Report.

4.1.3. Catch efficiency

For assessment intervals during which the reference and SUT both detect precipitation, the accumulation reported by the SUT, relative to that reported by the reference configuration, can be expressed in terms of the catch efficiency, or catch ratio.

$$\text{Catch efficiency} = \frac{\text{SUT accumulation}}{\text{Reference accumulation}}$$

The ideal value for catch efficiency is 1.

4.1.4. Precipitation type

To assess the influence of the predominant precipitation type (phase) on SUT performance relative to the reference configuration, the ambient temperature during the assessment interval is used to stratify the data by precipitation type.

- Liquid precipitation: minimum temperature over the 30 min interval ≥ 2 °C;
- Solid precipitation: maximum temperature over the 30 min interval ≤ -2 °C;
- Mixed precipitation: all precipitation events not classified as liquid or solid.

5. Environmental conditions

The environmental conditions at the site over the duration of the test period are expressed as probability density functions (PDFs) for air temperature, relative humidity, wind speed, wind direction, and precipitation rate in Figure 3. Both the entire period data, and the data corresponding only to the precipitation events reported by the site reference, R2, are shown. The precipitation percentage in Figure 3 represents the number of minutes of precipitation over a standard 30 minute interval, as recorded by the precipitation detector in the R2 reference configuration.

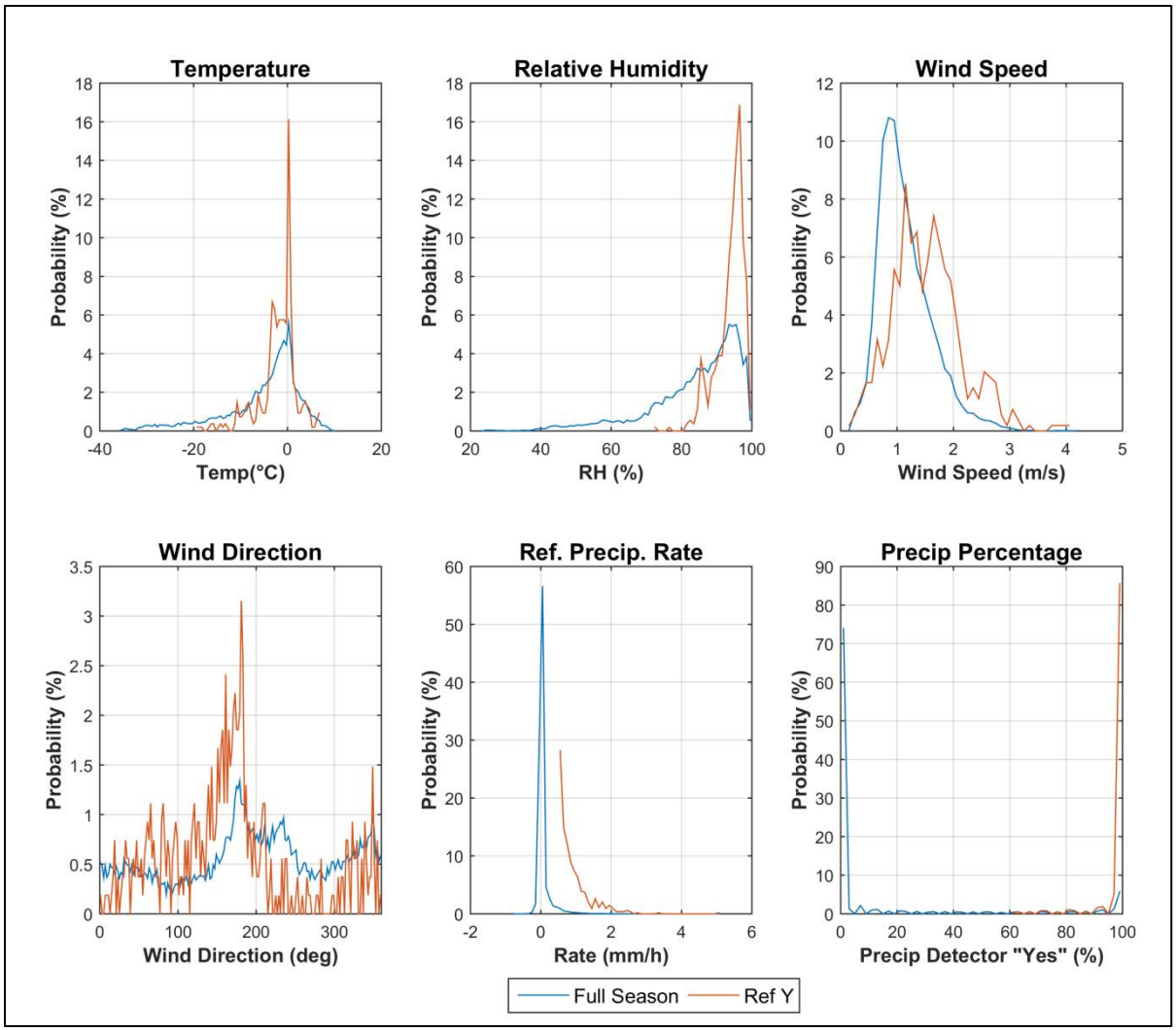


Figure 3: Summary of aggregated environmental conditions on the SPICE site that operated PWD52, over the entire duration ('Full Season', in blue) and for data corresponding to precipitation events, as reported by the site R2 reference ('Ref Y', in red), during the tests, as per Table 3 above.

6. Evaluation performance over the range of operating conditions

6.1. Skill score assessment

The overall ability of the SUT to represent precipitation similarly to the site field reference R2, is assessed using contingency tables (Section 4.1.1) and derived selected skill scores (Section 4.1.2). To better understand the potential influence of threshold choices on the derived results, two cases are considered here (see note below). The contingency results related to these two cases are given in Table 5 and the respective skill scores in Table 6.

Note: Following the data derivation explained in Section 4.1.1, the conditions required to have a 'Yes' or a 'No' event over the 30 min interval, for the reference and the SUT, for the two different cases treated here, are:

CASE 1 *(as defined in Section 4.1.1)*:

- *Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip*
- *Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip*
- *SUT 'Yes': SUT accumulation > 0 mm*
- *SUT 'No': SUT accumulation = 0 mm*

CASE 2:

- *Ref 'Yes': R2 weighing gauge ≥ 0.25 mm AND precip detector recording ≥ 18 min of precip*
- *Ref 'No': R2 weighing gauge < 0.1 mm AND precip detector recording 0 min of precip*
- *SUT 'Yes': SUT accumulation ≥ 0.1 mm*
- *SUT 'No': SUT accumulation < 0.1 mm*

Results presented in this report are based on Case 1.

Table 5: Contingency Tables: detection of precipitation of the PWD52 relative to the field reference, expressed as number of events over the entire test period. The skill scores associated with these events are given in Table 6.

Case 1 Ref Plv2DFAR

SUT PWD52	Yes	No	Total
Yes	524	707	1231
No	0	9639	9639
Total	524	10346	10870

Case 2 Ref Plv2DFAR

SUT PWD52	Yes	No	Total
Yes	519	24	543
No	5	10322	10327
Total	524	10346	10870

Table 6: Skill Scores for the PWD52. POD: Probability Of Detection, FAR: False Alarm Rate, B: Bias, HSS: Heidke Skill Score (see Section 4.1.2 for more details).

PWD52, Skill Scores				
	POD	**FAR**	**B**	**HSS**
Case 1	100%	57.4%	235%	56.8%
Case 2	99.1%	4.4%	104%	97.1%

6.2. Assessment of SUT performance during non-precipitating events

The performance of the SUT in the absence of precipitation (when the reference precipitation detector recorded 30 minutes without precipitation) is represented in Figure 4 and Table 7, reflecting the distribution of the sensor response, as measured during the interval.

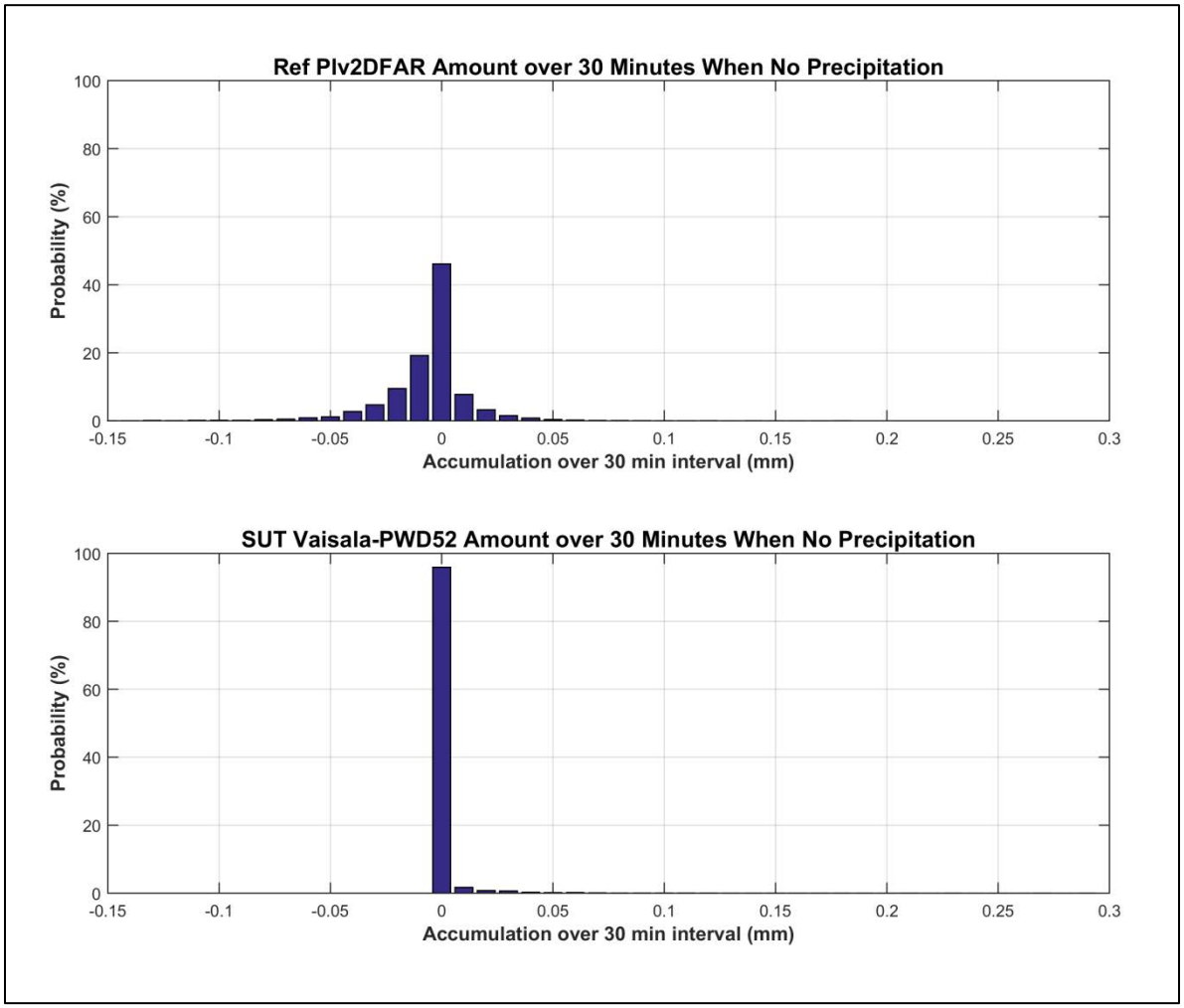


Figure 4: Probability of occurrence of a response during a 30 min interval in the absence of precipitation, represented by the signal output from (top) the R2 reference and (bottom) the SUT PWD52. Statistics associated with these graphs are given in Table 7.

Table 7: Reference and SUT statistics of response signal when no precipitation was occurring, as plotted in Figure 4; Average (Avg), standard deviation (STD), maximum (Max) and minimum (Min) of the response signal, together with the number of events (Num) over the test period is given.

No Precip Statistics – Reference: Plv2DFAR

	Ref Avg [mm]	Ref STD [mm]	Ref Max [mm]	Ref Min [mm]	Ref Num
Sodankylä	-0.003	0.020	0.190	-0.140	10354

No Precip Statistics – SUT: Vaisala-PWD52

	SUT Avg [mm]	SUT STD [mm]	SUT Max [mm]	SUT Min [mm]	SUT Num
Sodankylä	0.002	0.010	0.293	0.000	10354

6.3. Ability of the SUT to measure precipitation

6.3.1. Yes-Yes cases

Quantitatively, the performance of the SUT to derive and report precipitation is assessed relative to the site reference in several graphs and tables illustrated in this section, using only the cases where both instruments reported precipitation over the 30 min interval, according to the criteria used in Case 1 of Table 5 (cases 'Yes-Yes', or shorter 'YY').

6.3.1.1. Time series plots

The time series (cumulative sum of 30 min YY events accumulation) of the SUT is plotted against the reference for the two seasons, by precipitation type (see Section 4.1.4) in Figure 5.

The corresponding seasonal accumulations are given in Table 8.

Table 8: Seasonal accumulation [mm] for Sodankylä test site based on the sum of YY events from the SUT PWD52 and the field reference R2.

Sodankylä				
	Season 2013/14		Season 2014/15	
[mm]	Ref	SUT	Ref	SUT
All events	124.49	111.10	114.03	118.77
Rain events	17.67	18.48	5.60	5.81
Mixed events	72.26	62.07	51.04	55.78
Snow events	34.56	30.55	57.39	57.18

6.3.1.2. Scatter plots and RMSE values

Scatter plot of the amount derived by the SUT versus the reference amount, and discriminated by precipitation type, is given in Figure 6.

Quantitatively, the SUT performance is assessed using the Root Mean Square Error (RMSE) also known in practice as Operational Comparability. The results are available in Table 9.

Table 9: Root Mean Square Error (RMSE) statistics, in mm, for the SUT PWD52 with respect to the field reference, by precipitation type, including both seasons data.

RMSE [mm]	All	Rain	Mixed	Snow
Sodankylä	0.138	0.133	0.149	0.124

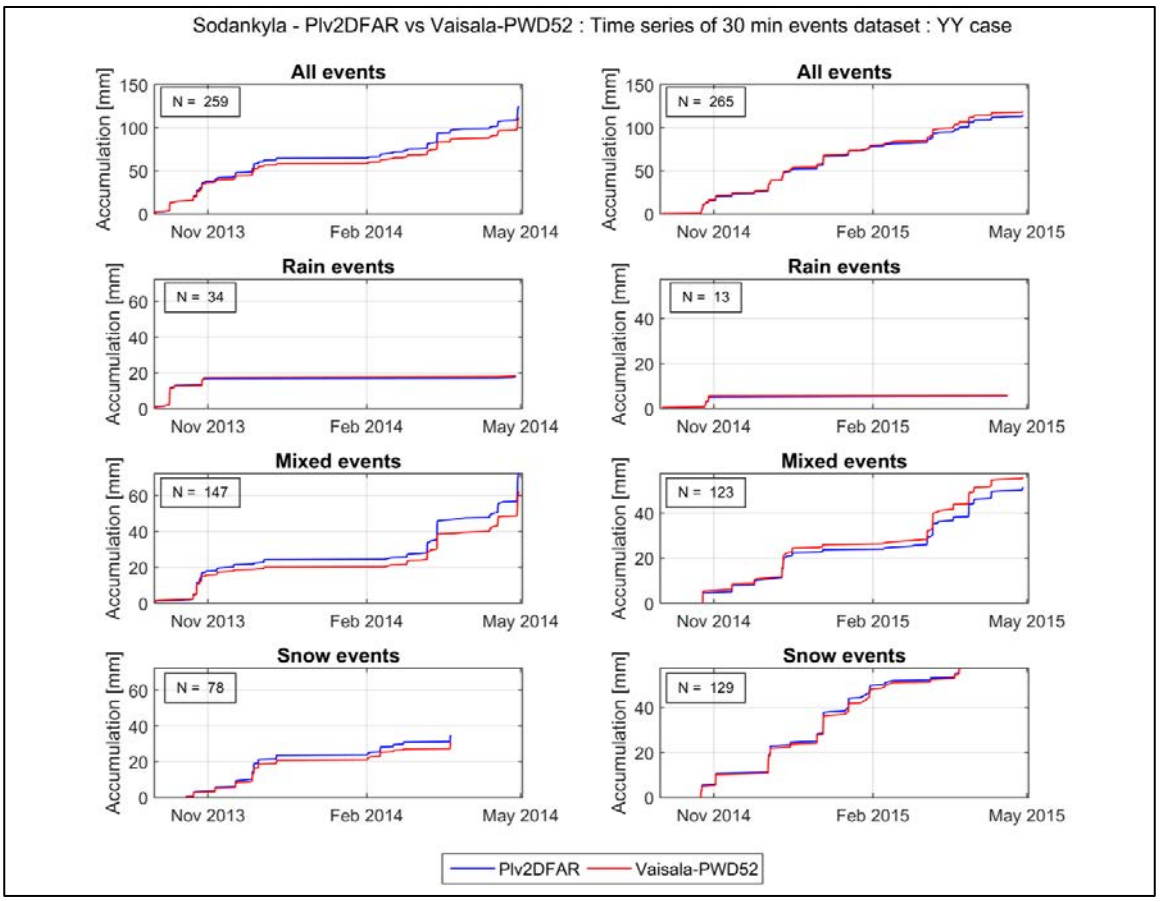


Figure 5: Time series based on 30 min YY events of the SUT PWD52 against the field reference, discriminated by precipitation type (Rain, Mixed, Snow), for both seasons (2013/14 on the left, 2014/15 on the right).

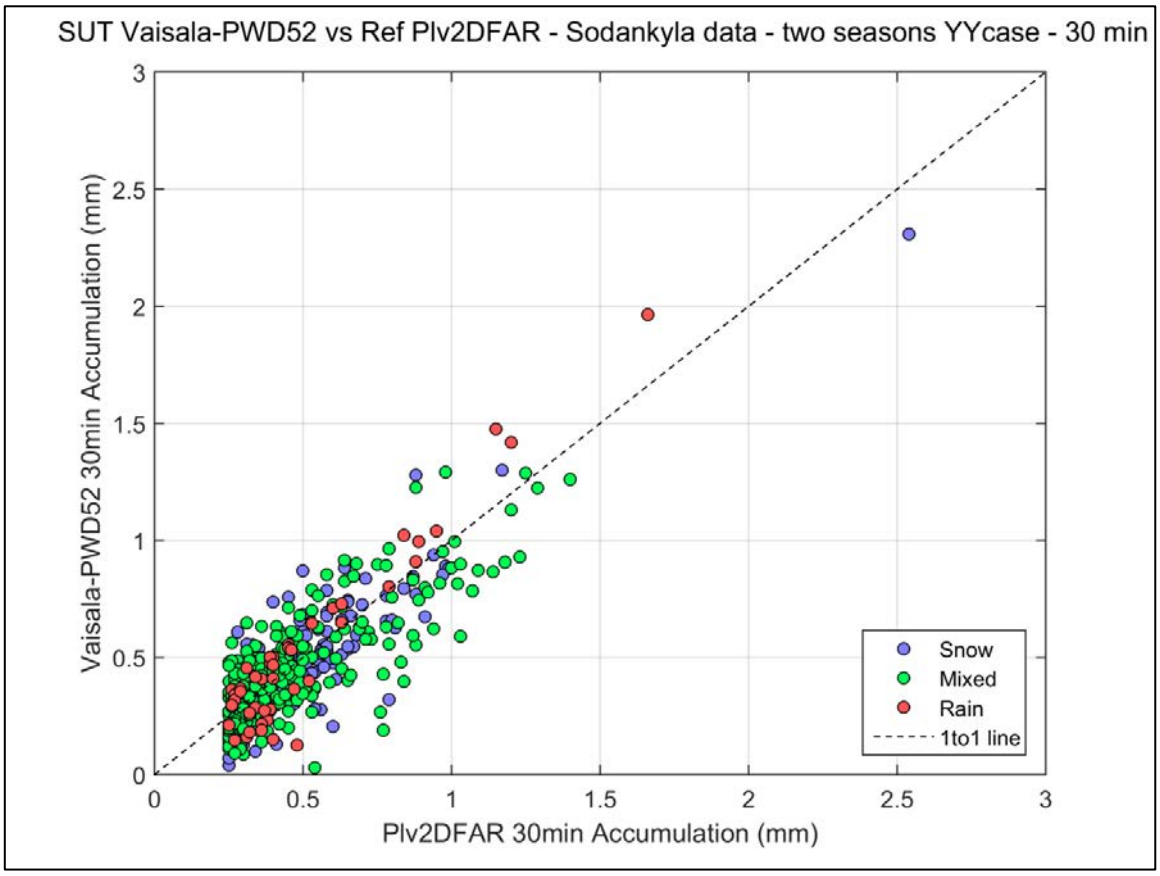


Figure 6: Scatter plot based on 30 min YY events accumulation from the SUT PWD52 against the field reference, for the two seasons, discriminated by precipitation type.

6.3.1.3. Catch efficiency evaluation by precipitation type

The Catch Efficiency (CE) of the SUT is represented by histograms (Figure 7) and boxplots (Figure 8), both discriminated by precipitation type.

The quantitative evaluation of the CE is provided in Table 10. The mean catch efficiency is given in the first line of this table, considering both seasons data and for each category of precipitation type as well as for all the events together.

Note: All events with a CE greater than 3, if any, are included in one category named '3 and more' or '≥3' in the upcoming graphs. Additionally, for all graphs representing the CE, a line is added at CE = 1, which represents the ideal case where the SUT reports exactly the same precipitation amount as the reference.

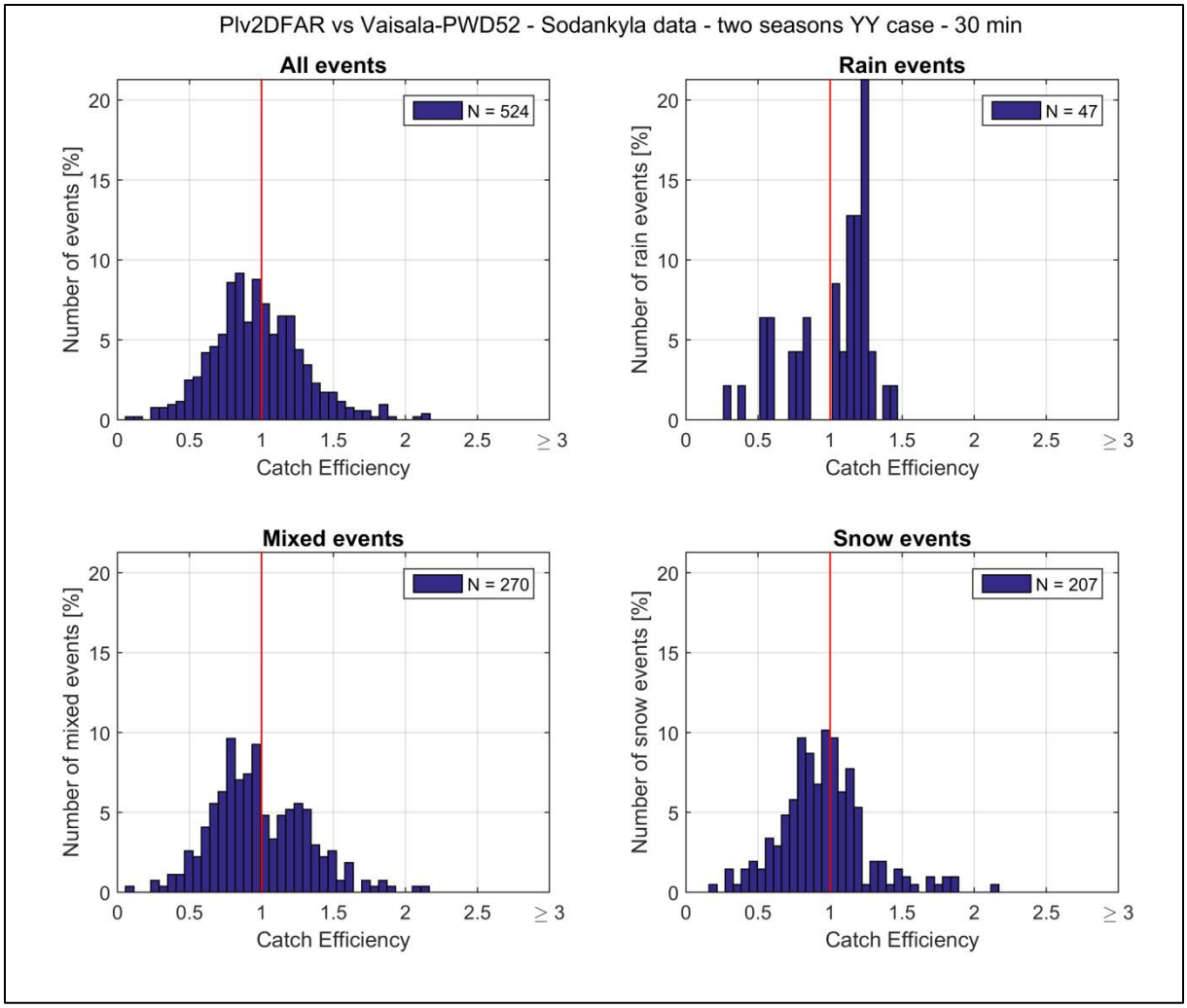


Figure 7: Histograms based on 30 min YY events for the two seasons, representing the distribution of the catch efficiency of the SUT PWD52 against the field reference (SUT/Ref), discriminated by precipitation type, with number of events given in the legend for each category. The red line at CE = 1 represents the ideal case.

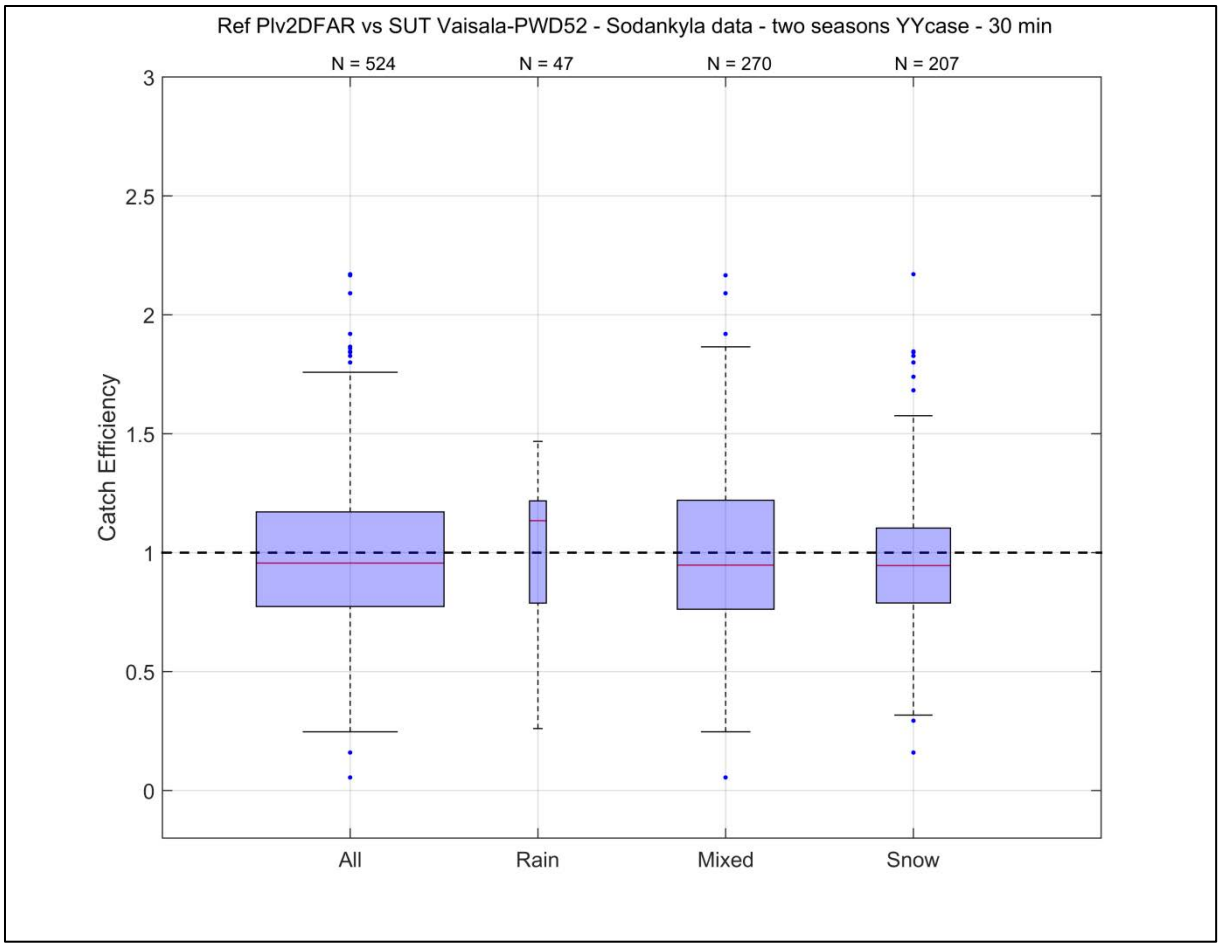


Figure 8: Boxplot based on 30 min YY events of the two seasons, representing the catch efficiency of the SUT PWD52 against the field reference (SUT/Ref), discriminated by precipitation type, with number of events given at the top of each category. The width of the boxes is proportional to the percentage of events in each category ('All' box represents 100% data). The dashed black line at CE = 1 represents the ideal case.

Table 10: Statistics related to the CE of SUT PWD52 against the field reference (SUT/Ref, see Figure 8), discriminated by precipitation type.

Catch Efficiency Statistics: **PWD52**

CE Boxplot Parameters	All	Rain	Mixed	Snow
Mean	0.98	1.01	0.99	0.95
Median	0.96	1.13	0.95	0.95
75 percentile	1.17	1.22	1.22	1.10
25 percentile	0.77	0.79	0.76	0.79
Upper Whisker	1.76	1.47	1.87	1.58
Lower Whisker	0.25	0.26	0.25	0.32
Maximum	2.17	1.47	2.17	2.17
Minimum	0.06	0.26	0.06	0.16
# Outliers	12	1	4	10
# Outliers ≥ 3	0	0	0	0
# Events	524	47	270	207

6.3.1.4. Catch efficiency dependency on wind speed

The variation of SUT catch efficiency (SUT/Ref) with wind speed is illustrated in a scatter plot together with a boxplot in Figure 9, both discriminated by precipitation type.

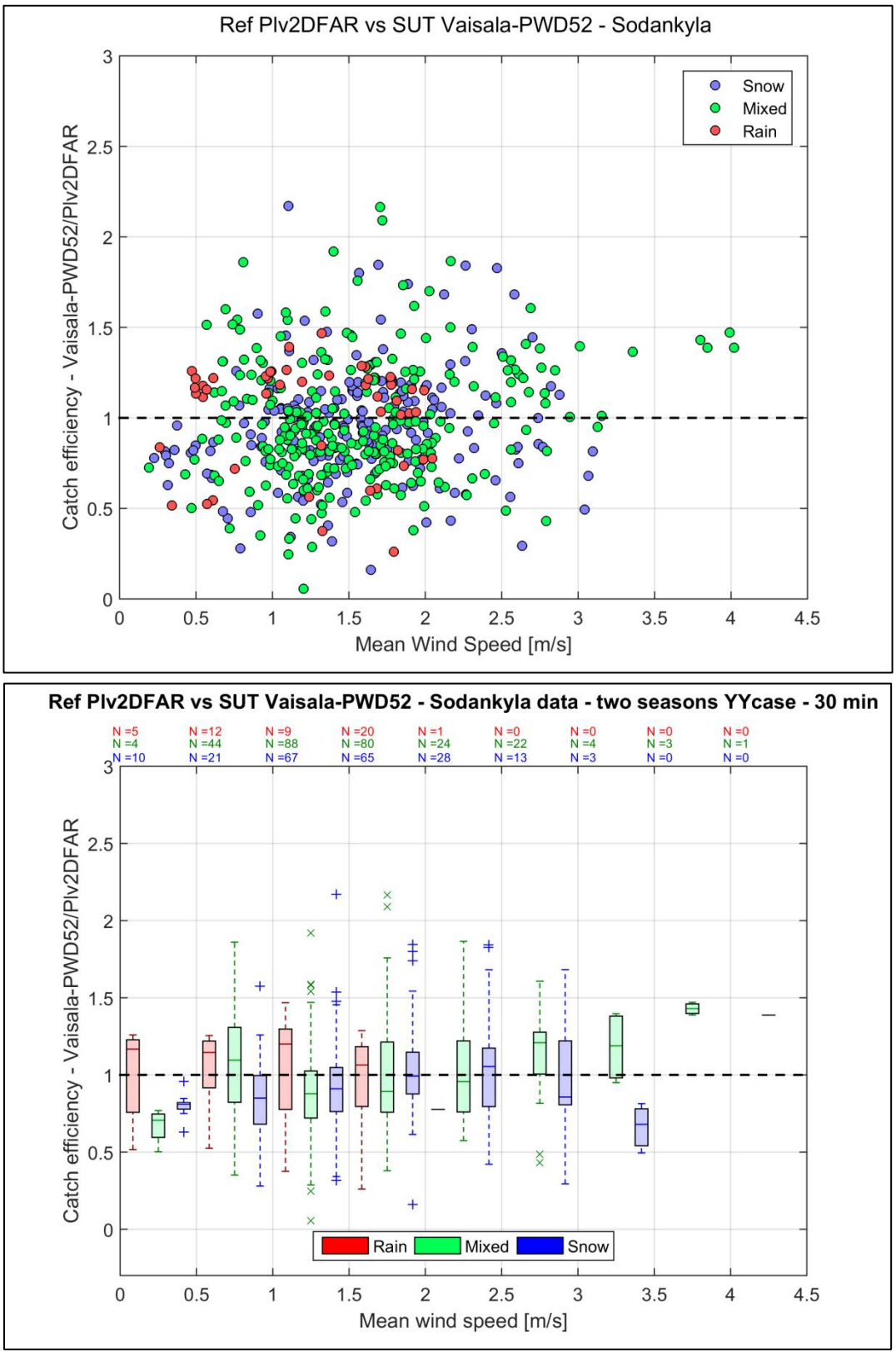


Figure 9: Scatter plot (top) and boxplot (bottom) based on 30 min YY events fpr the two seasons, representing the catch efficiency of the SUT PWD52 with respect to the field reference (SUT/Ref), against wind speed and discriminated by precipitation type. The dashed black line at CE = 1 represents the ideal case.

6.3.1.5. Catch efficiency dependency on wind direction

In order to assess the dependency of the CE with wind direction, a wind rose is produced (Figure 10) representing the wind data of the two seasons, binned by catch efficiency in order to represent undercatch (CE < 0.8), overcatch (CE > 1.2) and catch efficiency of 1 ± 20% of the SUT.

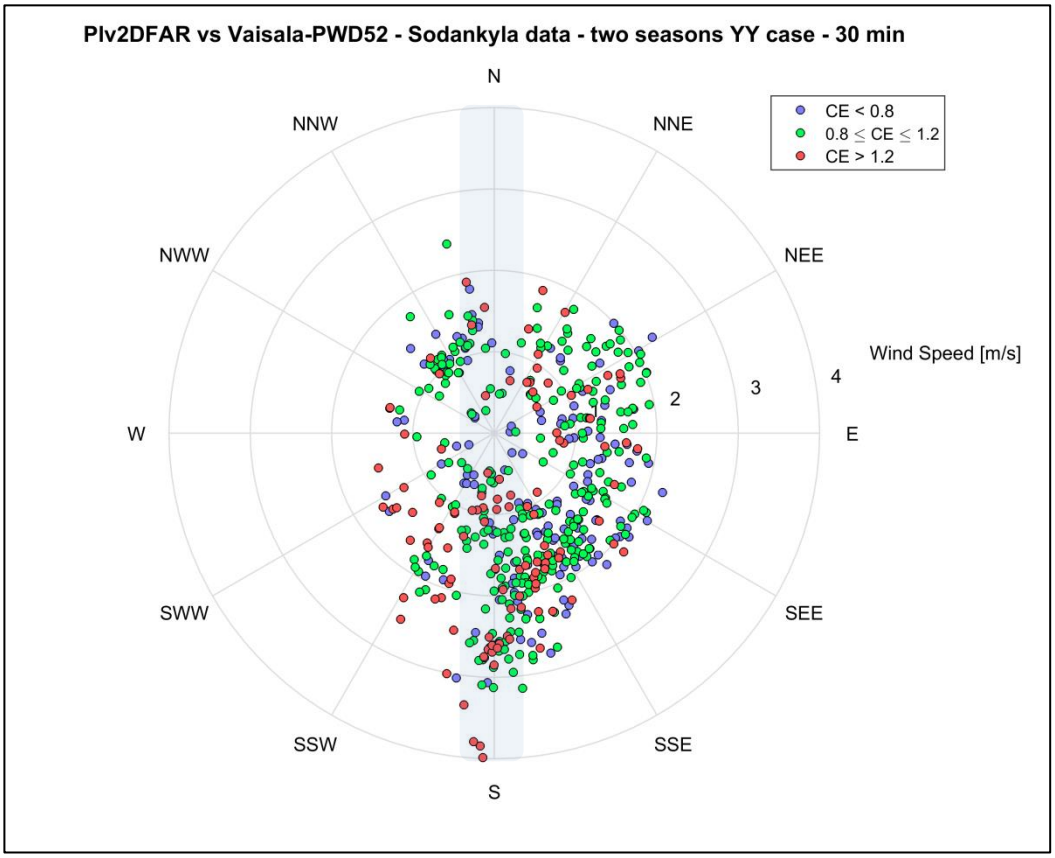


Figure 10: Precipitation events (YY cases) as function of wind speed and direction, and binned by catch efficiency (CE). The grey zone indicates the SUT orientation.

6.3.2. Yes-No and No-Yes cases

Events when the site reference and SUT do not agree on the occurrence of precipitation includes two categories of cases (Section 4.1.1): (1) when the field reference reported a precipitation event, while the SUT did not (Yes-No cases, 'YN'), and (2) when the field reference did not report a precipitation event while the SUT did (No-Yes cases, 'NY').

Histograms illustrating field reference and SUT reports and associated site conditions for all NY cases (no YN cases were reported for the PWD52) during the test period are provided in Figure 11.

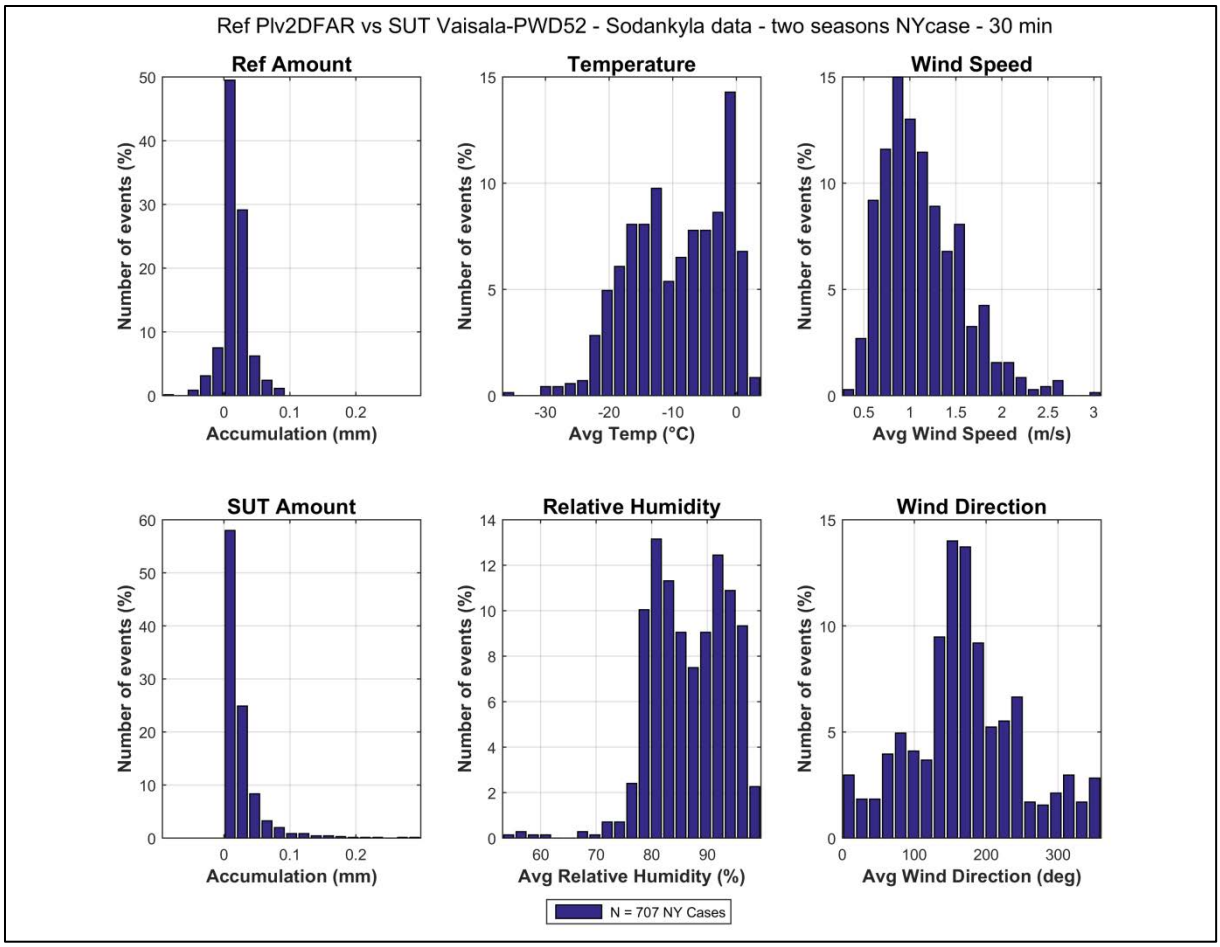


Figure 11: Histograms of SUT PWD52 and field reference accumulations (left column), along with distributions of mean temperature, mean relative humidity, mean wind speed, and mean wind direction for all NY cases (number indicated in the legend) of the 30 min intervals, as reported by Case 1 of Table 5.

7. Interpretation of results

7.1. Operating conditions

Sodankylä is a site characterized by very low wind speeds, as shown in Figure 3, with a majority of precipitation events occurring below 3 m/s. Regarding temperatures, Sodankylä recorded most of the precipitation events at temperature between -10 and 0°C, with extreme by -20°C. The site shows fairly humid climate conditions as expected from a Northern Boreal climate site.

7.2. Reliability in detecting precipitation

The Vaisala PWD52 was efficient in detecting precipitation when the reference reported precipitation, as presented in Table 6, with a POD of more than 99%, looking at the two different threshold cases. When no threshold is applied to the SUT (Case 1), the high FAR of 57% and bias of 235% are a direct consequence of the large number of NY events (see Section 7.5 for more details).
However, it can be noted that the FAR score is reduced to less than 5%, and the bias and HSS scores improved, when a threshold of 0.1 mm over 30 min events is applied to the SUT (Case 2), indicating that the majority of the NY events of Case 1 are for very low accumulations. This could be the consequence of a difference in sensitivity between the weighing gauge used as the reference and the non-catchment type SUT, related to their different measurement principle.

7.3. Performance of SUT during no-precipitation events

The PWD52 has a stable output signal, showing almost no noise during no-precipitation events (see Figure 4 and Table 7). Note that the no-precipitation study is performed using only the reference precipitation detector data (here a DRD11A for the first season and a Parsivel^2 for the second season) with the criteria ensuring that 30 min of 'no-precipitation' was recorded. For these cases, the PWD52 indicates 0 mm for more than 93% of the time, indicating a good agreement with the reference precipitation detector on the absence of precipitation. Most of the remaining cases (where the reference precipitation detector doesn't detect any precipitation while the SUT does) correspond to very low accumulations reported by the SUT (less than 0.1 mm/30 min) and can be related to the NY events mentioned above and analyzed in Section 7.5.

7.4. Performance of SUT during precipitation events

When both the reference and the PWD52 report precipitation (YY cases), the overall results show a very similar behavior of the sensor for all precipitation types.

The RMSE in Table 9 shows almost identical values for rain, mixed and snow precipitations, with 0.14 ± 0.01 mm. It could be noted from Figure 6, that the rain events with accumulation greater than 0.5 mm show a slight overcatch.

Despite this minor overcatch for some rain events, the PWD52 has a slight tendency to undercatch, as represented by Figure 7 andFigure 8 and Table 10, with a mean catch efficiency over the two seasons of 0.98, dropping to 0.95 for snow. It has to be reminded, however, that the number of snow events is significantly higher than that for rain events (few during winter seasons), and have hence a stronger impact over the global assessment.

Overall, the PWD52 seems to be a reliable instrument to account for the total accumulation over a longer period (e.g. one season), with a mean catch ratio very close to the 1-line, as shown in Figure 8 or Figure 9 (bottom). The relatively high scatter visible in Figure 9 (top), with catch ratio varying randomly from 0.3 to 2, on event based statistics (typically 30 min interval), occurring at all wind speeds (up to 4 m/s, maximum wind speed in Sodankylä), tends to show that the PWD52 is less reliable to derive solid and mixed precipitation accumulation over near real time periods.

Even if better in deriving accumulation for longer time periods, the PWD52, as other non-catchment type instruments, doesn't have an assurance of the continuity in the measurements, though, which is critical to long term data collection. If power is off or signal transmission is interrupted, no data is recorded and there is no possibility to know what has fallen during this time, thus affecting the long term data reports of the PWD52.

The results in Figure 9 indicate no clear dependency of the catch efficiency with increasing wind speed, for all precipitation types. The performance of the sensor for higher wind speed cannot be assessed, since Sodankylä is a site with low wind speed (up to 4 m/s), and the SUT was not tested in other environmental conditions.

The wind rose (see Figure 10) shows no dependency of the catch efficiency with wind direction. The sensor was installed parallel to the prevailing wind direction during precipitation events. An assessment with a sensor perpendicular to the main flow has not been examined, though, since no such configuration was tested.

7.5. Assessment of Yes-No, No-Yes events

The assessment of YN, NY events completes the picture of the performance of the SUT. As it can be seen in Table 5 and Figure 11, when no threshold is applied to the SUT (Case 1), the PWD52 shows a large amount of NY cases, but no YN cases were reported, the latter indicating the good agreement of the sensor to detect precipitation whenever the reference reported precipitation. When a threshold is applied to the SUT (Case 2) - here a threshold of 0.1 mm/30 min – the number of NY events is significantly reduced. Therefore, applying a threshold enables to ensure the good agreement of the PWD52 with the reference, in all cases (YN, NY, YY, NN), and prevents the user to deal with NY events of unknown origin (either real or artefacts).

Looking more closely at Case 1 and according to Table 5, more than 700 NY events occurred during the two seasons of the experiment. It could be useful to remind that these NY events are characterized, over the 30 minutes of an event, by an accumulation from the reference below the defined threshold of 0.1 mm, the reference precipitation detector having recorded 0 min of precipitation, and the SUT indicating more than 0 mm (see Section 4.1.1). This very high number of NY events appears to happen for very low accumulations (mostly less than 0.1 mm/30 min) reported by the PWD52 and for low temperatures and wind speeds (see Figure 11). Moreover more in-depth analysis showed that the NY events occurred primarily in the second season (621 for season 2 compared to 86 for season 1), where the precipitation detector being part of the reference was changed from the DRD11A capacitive sensor by a Parsivel^2 (see Table 4). Figure 12 shows some of the NY events of season 2 over the accumulation curve of the R2 reference weighing gauge, together with the precipitation detection measured by the Parsivel^2, the DRD11A and the PWD52. It can be noted that this light precipitation was well detected by the DRD11A and the PWD52, but nothing was

measured by the Parsivel^2. This led to the idea that the NY results could change if the choice of the reference precipitation detector was different for the second season.

To better address these differences in precipitation sensitivity and to give a sense of the results dependency on the selected reference precipitation detector, a second study has been made using the DRD11A as the reference precipitation detector for the second season as well (instead of the Parsivel^2). The same number of YY and YN events was found, but the number of NY events was reduced to 240 events (compared to 707 with the Parsivel^2), leading to a FAR of 31.5% (against 57.4% with the Parsivel^2), a bias of 146% and an HSS of 80.3%. Even if capacitive precipitation detectors, such as the DRD11A, are less sensitive than laser-based precipitation detectors, like the Parsivel^2 (see Section 4.1.3.5.2 of the SPICE Final Report), the DRD11A seems to better match the PWD52 results in terms of precipitation detection (as it could be seen in Figure 12). A similar graph to Figure 11, but this time using the DRD11A as the reference precipitation detector for both seasons, is given in Figure 13 for the PWD52 NY events. This figure shows a wider distribution over the wind speed range, but a distribution of temperatures closer to 0°C. It is still possible that these remaining 204 NY events are real light events that were not detected by the DRD11A (which measured 30 min of no precipitation for all of them), but this remains uncertain and no other parameters or photos are available to confirm or infirm this statement.

To conclude, the PWD52 NY events of either study (with DRD11A at season 1 and Parsivel^2 at season 2 or only DRD11A for both seasons) reflect therefore rather a reliable higher sensitivity of the PWD52 compared to the reference system. Indeed, on one hand the reference precipitation detector could have a lower sensitivity than the PWD52, on the other hand the reference weighing gauge has a different physical principle of measurement implying an expected lower sensitivity than the PWD52.

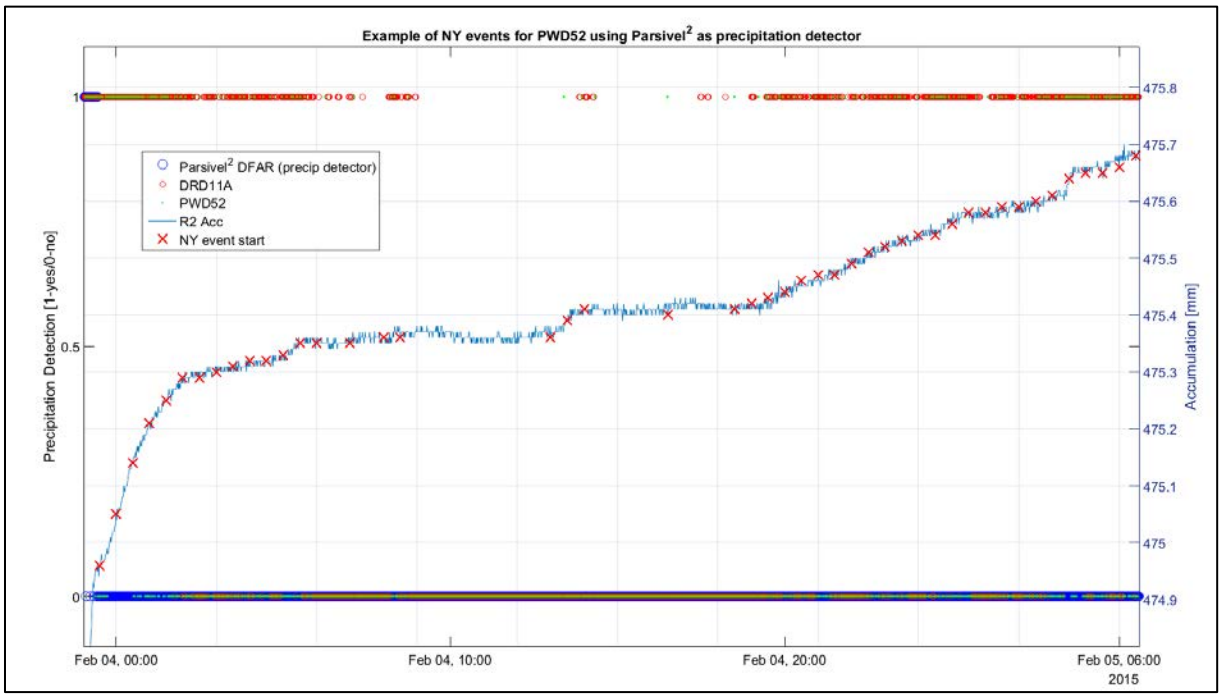


Figure 12: Example of NY events for the SUT PWD52 using the Parsivel^2 as the reference precipitation detector for the second season.

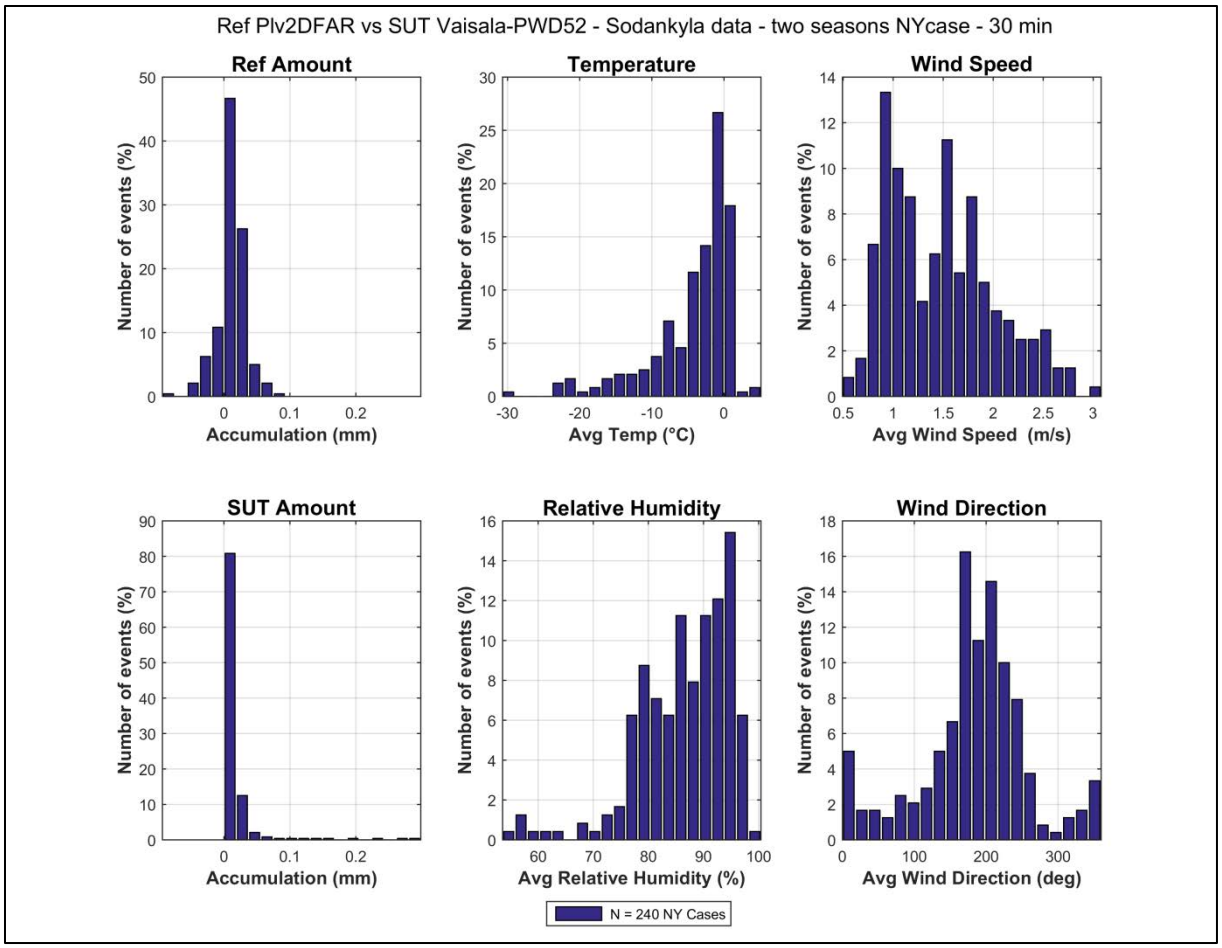


Figure 13: Histograms of NY events similar to results of Figure 11, but using a DRD11A as the reference precipitation detector (instead of the Parsivel^2) for the second season.

7.6. Threshold selection

The threshold to be set for the PWD52 in order to report precipitation adequately over a 30 min interval (3 STD, according to the methodology defined in Section 3.6.1.3.2 of the SPICE Final Report) corresponds to 0.03 mm (3 x 0.01 mm, from Table 7). Additionally, as shown in the contingency tables (Table 5), a threshold of 0.1 mm/30 min enables to get rid of events whose origin is not confirmed or well established.

8. Operational considerations

The overall experience with the PWD52 at Sodankylä was positive. The sensor appeared to be robust operationally with almost no data breaks due to the instrument (there was one major data gap of $1^{1/2}$ month in December 2013, but this was due to an issue with the data collection program).

During the experiment, some snow accumulation has been observed on the instrument device and mounting parts (see Figure 14). This is expected to happen on sites with calm weather conditions, as the Sodankylä site. According to the manufacturer, the probability that blowing snow from these accumulations could be counted as precipitation is very unlikely, though, because it would require that the precipitation detector 'RAINCAP' surface and the optical measurement volume, both part of the PWD52, are hit by the snow particles at the same time and with sufficient strength.

Figure 14 : Snow accumulation on PWD52 device, 2015-02-15, 11:02 UTC.

8.1. Maintenance

The PWD52 requires almost no maintenance. According to the User Manual, the sensor's lenses and the RAINCAP sensing surface should be cleaned every six months. Before using it operationally, though, the User Manual indicates also that an individual calibration on site may be needed to derive correctly the liquid water content of precipitation.

9. Performance considerations

- When no threshold is applied to the SUT accumulation (Case 1), a high number of false alarms events (NY) were reported by the PWD52 when comparing with the reference. These seem to be mainly due to the higher sensitivity of the SUT compared to the reference, but it remains that it has not been completely understood if some are real or artefacts. Applying a threshold (of 0.1 mm/30 min for instance in Case 2) to the PWD52 accumulation output, though, significantly reduces the FAR and ensures to derive reliable precipitation events.
- High scatter of CE on an event basis (30 min), but good mean CE over the seasons, makes the PWD52 more suitable for deriving solid precipitation amount over long periods of time, giving the condition that the sensor operates continuously.

SPICE Instrument Performance Report
Campbell Scientific CS725

1) Technical Specifications

Physical principle: SWE measurement based on the attenuation of terrestrial gamma radiation from the surface

Measurement Area: Coverage beam 60° from centre, approx. 7m diameter footprint (39m^2) at 2m height

Measurement Range: 0 to 600mm

Measurement Accuracy: ±15mm from 0-300mm, 15% from 300-600mm

Resolution: 1mm, 24 hour integration

Link to manual: https://s.campbellsci.com/documents/ca/manuals/cs725_man.pdf

Figure 1: Photo of the CS725 SWE sensor

2) SPICE Test Configuration

Test Sites: Caribou Creek (Canada) and Sodankylä (Finland)

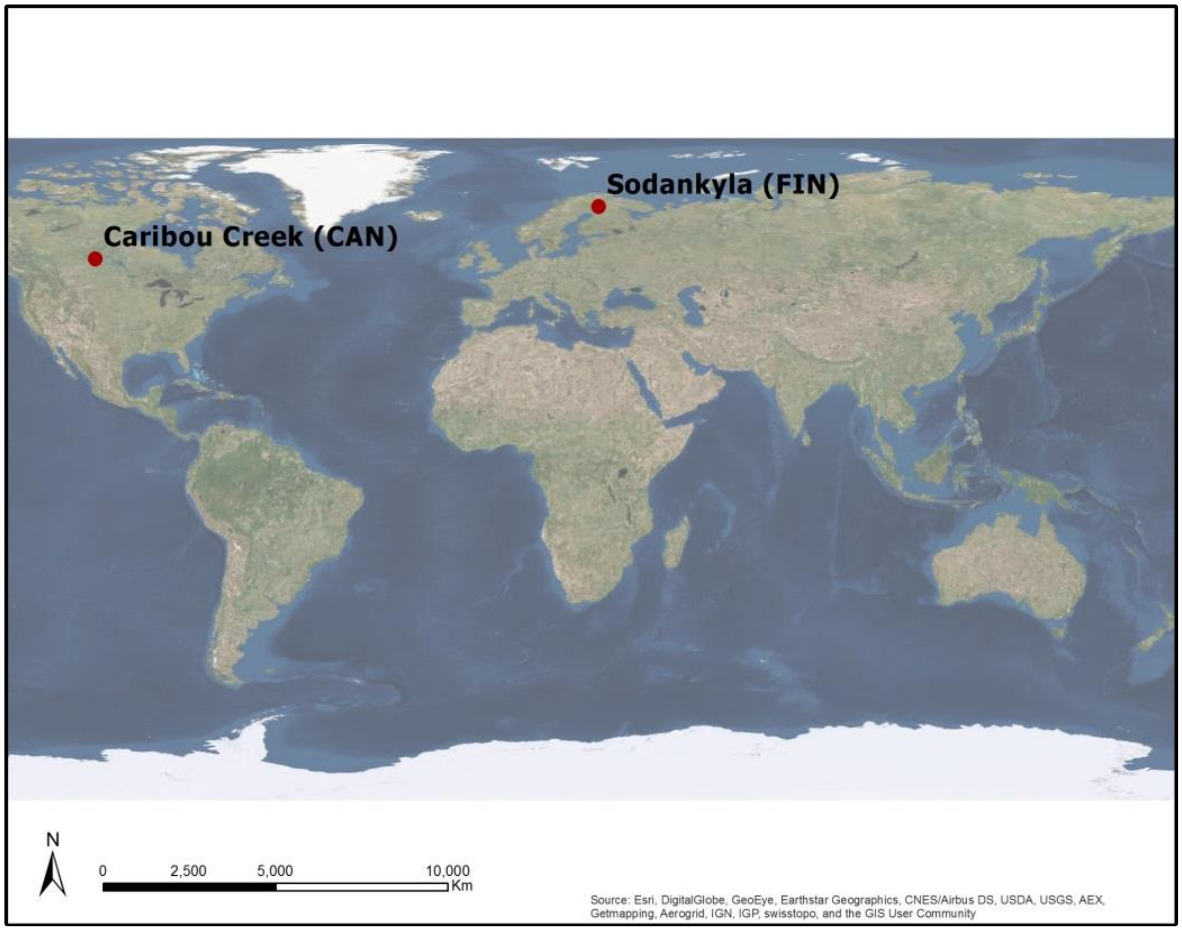


Figure 2: Location of the sensors under test

2.1) Site Specifications

Table 1: Site specific instrument and installation details

	Sodankylä	Caribou Creek
Date of Installation	11-Oct-2013	17-Oct-2013
Distance to Ground	1.9/2.3*m	2.0m
Serial Number(s)	1040	1051
Installation Location	40:62/40:72*	C2

*Sodankylä instrument moved 2014-09-18 from 40:62 to 40:72

2.2) Site Photos

Figure 3: Sodankylä

Figure 4: Caribou Creek

2.3) Instrument Footprint Diagram

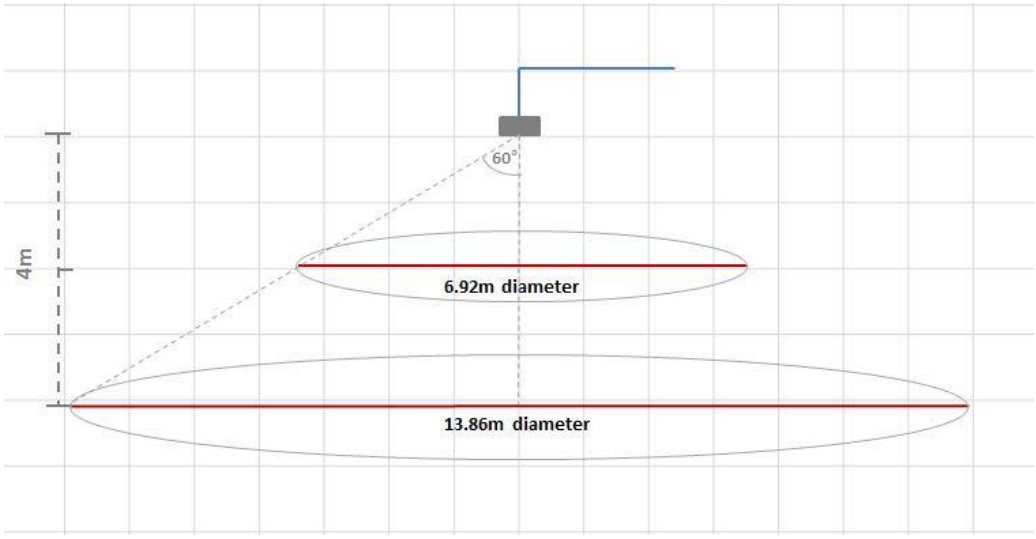


Figure 5: Conceptual diagram of instrument field of view

2.4) Environment Conditions During SPICE

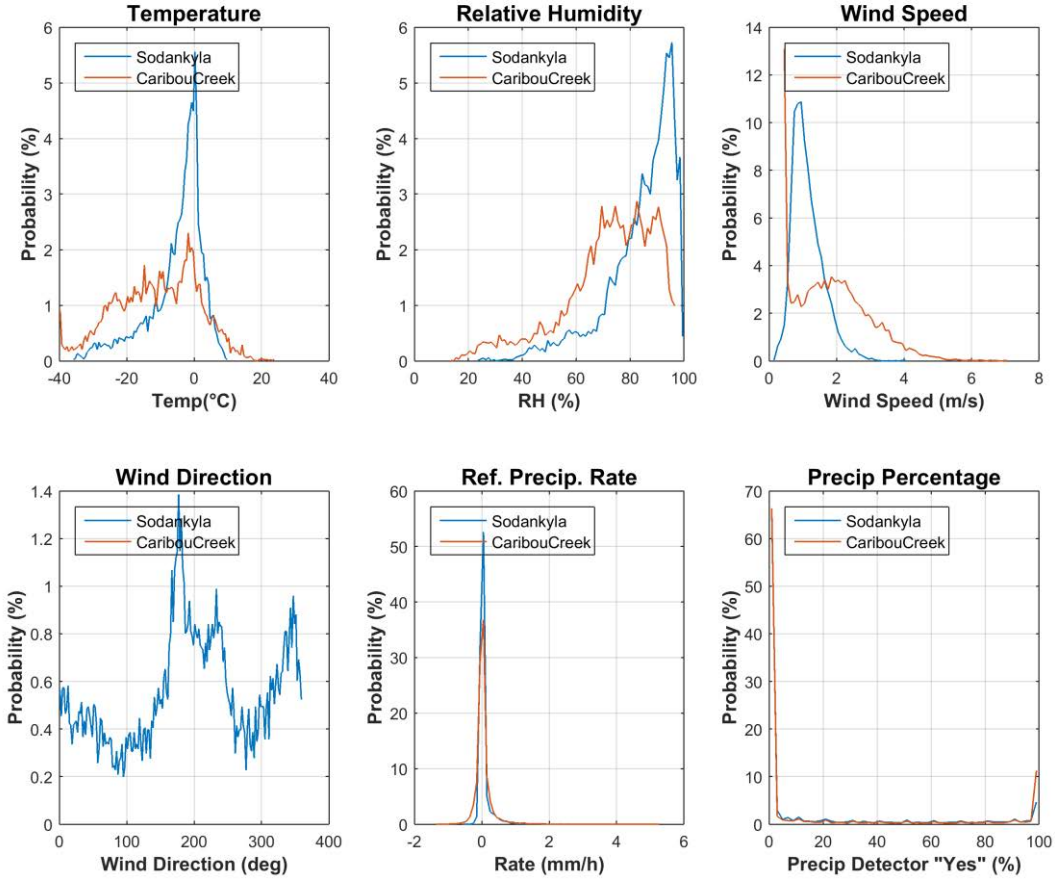


Figure 6: Summary of aggregated environmental conditions on the SPICE sites that operated a CS725 for the entire duration of formal tests.

3) Data Output

Table 2 : Parameters measured or output by the instrument

Measured Parameters	Units	Sodankylä	Caribou Creek
K Counts	Total	x	x
TL Counts	Total	x	x
SWE K	mm	x	x
SWE TL	mm	x	x
SWE K/TL Ratio	-	x	x
K Uncorrected	Total	x	
Crystal Temp Min	Deg C	x	x
Crystal Temp Max	Deg C	x	x
Statistics	-	x	
Histogram Blocks	Total	x	x
Displacement of K Peak	Bins	x	x
Instrument Power	Volts	x	x

Table 3: Measurement and data acquisition parameters

Measurement and Processing	Units	Sodankylä	Caribou Creek
Data Sampling Rate	[Hours]	Continuous	Continuous
Data Acquisition Interval	[Hours]	6	6
Date Processing		24 hour integration	24 hour integration
Communication Protocol		RS-232	RS-232

4) Quality Control Information

Missing data occurred when data were not recorded by the data logger. This may or may not be related to the function of the instrument. Suspicious data were identified as an outlier and were either removed automatically via pre-set range and jump filtering, or identified visually as an outlier and removed manually. Range and jump filters were set based on reasonable physical limits for the individual sites. Suspicious data may be a result of a malfunctioning instrument or from meteorological impacts on the instrument's measurement capability. Table 4 provides a seasonal breakdown of the quality control metrics by site. More commentary on individual instruments, where required, is included below. Only data classified as "Good" were used for the intercomparisons.

Table 4: Seasonal breakdown of data QC metrics

SEASON 2013-2014	Sodankylä	Caribou Creek
Collection Period	Oct-June	Oct-June
Good	99.2%	95.1%
Missing	0.8%	4.9%
Suspicious	0%	0%

SEASON 2014-2015	Sodankylä	Caribou Creek
Collection Period	Oct-June	Oct-April
Good	98.0%	99.9%
Missing	2.0%	0.1%
Suspicious	0%	0%

5) Evaluation of the Ability to Perform Over a Range of Operating Conditions

The instruments at Sodankylä and at Caribou Creek were compared against manual observations of Snow Water Equivalent at each site. The frequency of the manual measurement was approximately every two weeks. The manual measurement at Sodankylä was obtained at a single sample location (50:66) at a distance of about 12m from the centre of the sensor's footprint. The manual measurements at Caribou Creek consisted of a 5 point snow course combined with a sample within the footprint of the sensor, approximately 2m from the centre. The Caribou Creek intercomparison below used the SWE sample inside the sensor footprint and not the snow course average. The Sodankylä intercomparison between the sensor output and the manual SWE measurements is shown in Figure 7 with the time series for both seasons shown in Figure 8. The Caribou Creek intercomparison is shown in Figure 9 with the time series for both seasons shown in Figure 610. Summary statistics for both sites and years are shown in Section 5.2c.

5.1) Performance against the reference

5.1a Sodankylä

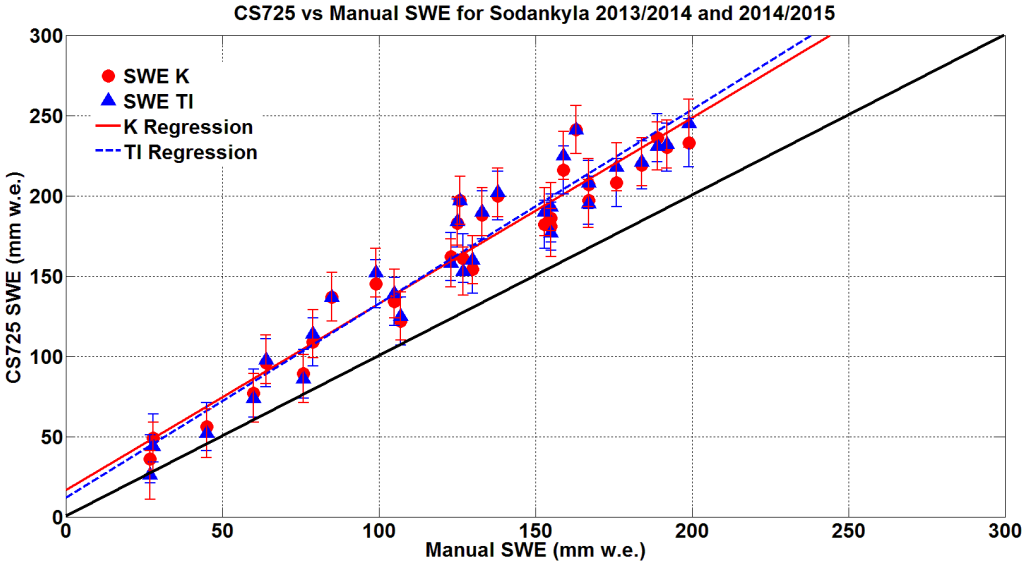


Figure 7: Sodankylä CS725 vs. manual SWE. Sensor Potassium output in red and Thalium output in blue. Error bars represent the manufacturers specified measurement accuracy

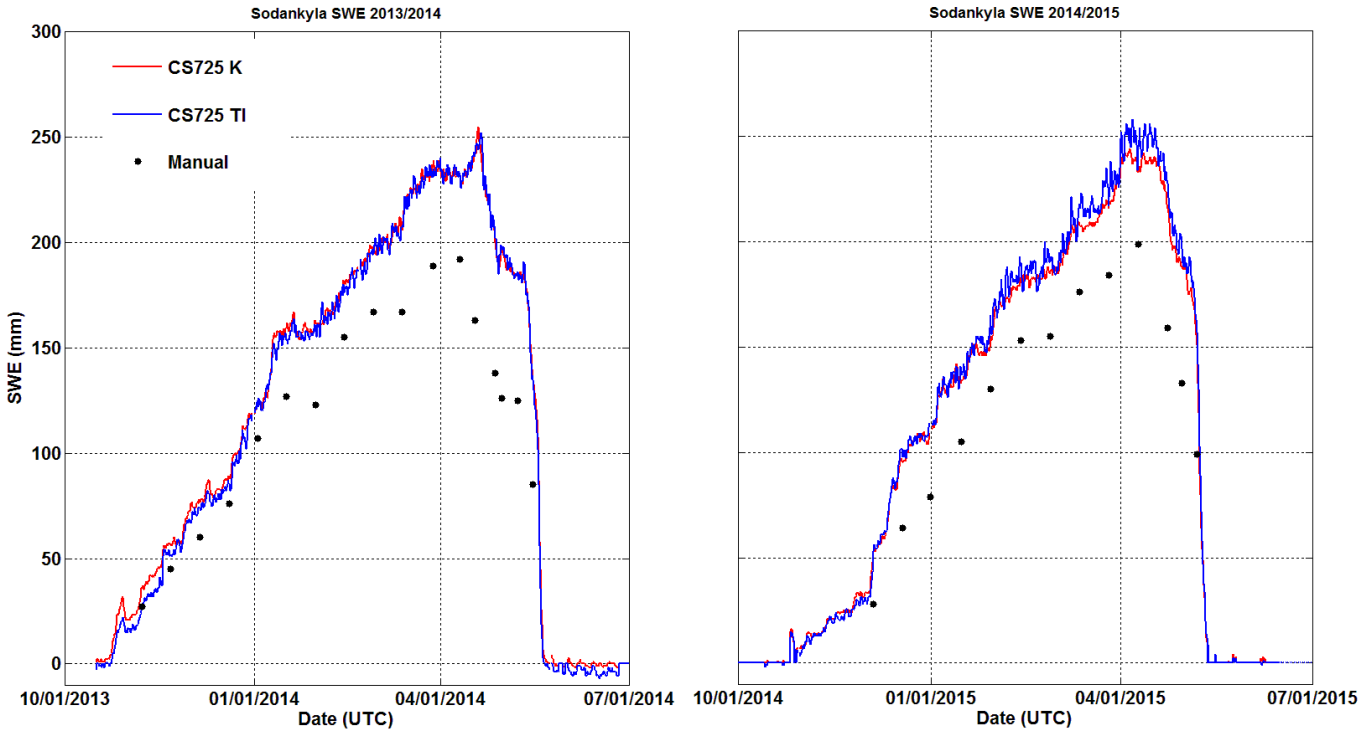


Figure 8: Sodankylä CS725 and manual SWE time series for 2013/2014 (left) and 2014/2015 (right). Sensor Potassium output in red and Thalium output in blue.

5.1b Caribou Creek

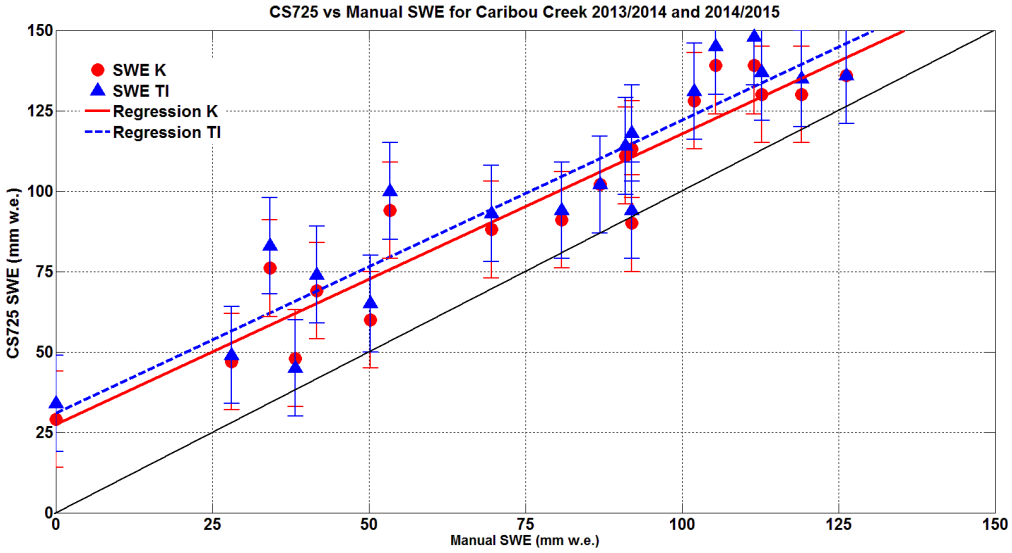


Figure 9: Caribou Creek CS725 vs. manual SWE. Sensor Potassium output in red and Thalium output in blue. Error bars represent the manufacturers specified measurement accuracy

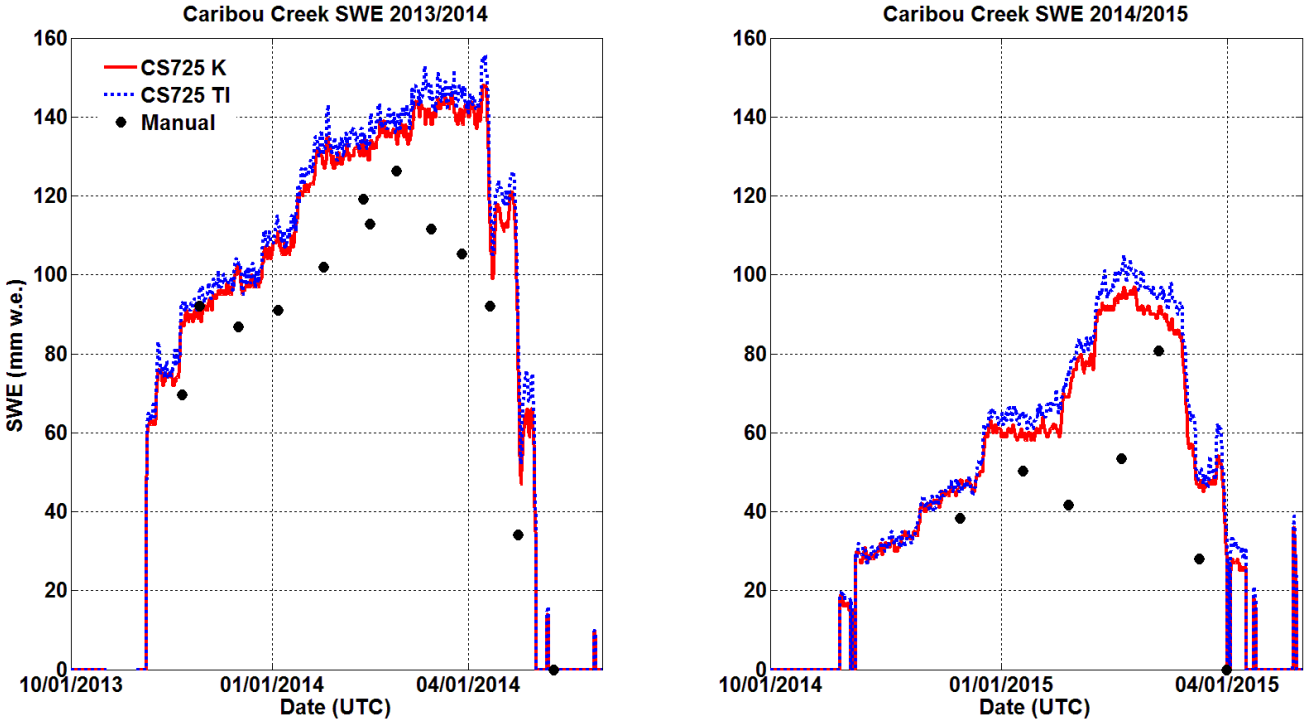


Figure 10: Caribou Creek CS725 and manual SWE time series for 2013/2014 (left) and 2014/2015 (right). Sensor Potassium output in red and Thalium output in blue

5.1c Intercomparison Statistics Summary

Table 5: Summary statistics for the CS725 at Sodankylä (from Smith et al. 2016)

Season	Slope K(Tl)	Intercept K(Tl) mm w.e.	r^2 K(Tl)	RMSE K(Tl) mm w.e.	Mean Relative Bias (K)	n
2013/2014	1.24(1.27)	8.77(3.17)	0.92(0.92)	43.0(42.2)	30.1%	17
2013/2014 (pre-melt)	1.24(1.28)	0.0123(-6.63)	0.97(0.97)	35.6(33.9)	24.6%	13
2014/2015	1.06(1.13)	26.9(24.2)	0.96(0.96)	36.6(42.2)	30.9%	13
2014/2015 (pre-melt)	1.05(1.12)	23.3(20.2)	0.99(0.99)	30.0(35.7)	28.1%	10
Combined	1.16(1.21)	16.8(11.9)	0.92(0.92)	40.3(42.2)	30.4%	30

Table 6: Summary statistics for the CS725 at Caribou Creek (from Smith et al. 2016)

Season	β K(Tl)	ε K(Tl) mm w.e.	r^2 K(Tl)	RMSE K(Tl) mm w.e.	Mean Relative Bias (K)	n
2013/2014	0.783(0.764)	40.6(46.9)	0.78(0.72)	22.8(27.5)	22.2%	12
2013/2014 (pre-melt)	0.982(0.997)	17.7(20.2)	0.79(0.75)	18.0(22.2)	15.4%	9
2014/2015	0.849(0.849)	27.1(30.4)	0.77(0.71)	23.6(27.4)	63.0%	7
2014/2015 (pre-melt)	1.12(1.31)	-8.38(-14.5)	0.55(0.60)	25.4(29.5)	42.4%	4
Combined	0.904(0.911)	27.5(31.0)	0.90(0.87)	23.1(27.4)	34.6%	19

5.2) Factors that influence instrument performance

The instrument performance is influenced by the amount of Potassium or Thallium present in the soil. These levels need to be high enough for the instrument to be functional. Pre-installation testing for these levels using the CS725 sensor is required to confirm functionality. The presence of trees or buildings in proximity to the sensor can be an additional source of gamma radiation that is not coming from the surface directly under the sensor and can influence the measurement. This is mitigated by the use of a collimator to eliminate non-soil gamma radiation sources. The instrument requires a gravimetric measurement of soil moisture within the instrument footprint as a calibration prior to the accumulation of snow.

5.3) Performance Considerations

The SPICE data analysis team suspected that the infiltration of liquid water draining from the snowpack and infiltrating into frozen (or non-frozen) sandy soils during the snow accumulation period resulted in a positive bias when comparing the instrument to the manual measurement of SWE. This is explained further in other sections of the SPICE final report.

Spatial variability in SWE between the location of the reference measurements and the sensor at Sodankylä and at Caribou Creek added uncertainty to the intercomparison. Snow course information at Caribou Creek showed that spatial variability of SWE at that site was substantial and it is expected that this variability also transfered through the field of view of the sensor, indicating that caution is required when comparing measurements taken at different spatial scales. From the 5-point snow course measurements at Caribou Creek, the stake used as the intercomparison reference can deviate from the site mean by as much as 30 mm in 2013/2014 and by as much as 40 mm in 2014/2015. This variability is discussed in other section of the SPICE final report.

The manual sampling of SWE using a snow tube, such as the ESC-30, generally has a low measurement bias (Goodison et al., 1987) but errors tend to occur when sampling more complex snow packs with ice layers (Powell, 1987). Depending on the skill of the observer and the condition of the sampler, snow will tend to escape out of the bottom of the tube while cutting through the layers of ice in the pack. This would result in an underestimate in SWE, perhaps as high as 5-10%, and could explain some of the overestimation shown by the sensor. Also, sampling a melting snowpack is prone to error, especially if the sample is being bagged and weighed as it was at Caribou Creek. When a sample from a melting snowpack is transferred from the tube to a bag for weighing, it is inevitable that some snow and water remains in the tube. This results in an underestimate which could perhaps be as high as 5%.

5.4) Maintenance

No maintenance was noted by the site managers for either season. The CS725 at Sodankylä was moved to a different location at the site after the first season to distance the instrument from buried cables.

6) Lessons Learned

The CS725 sensor, based on testing at Sodankylä and Caribou Creek, appeared to be a robust and reliable instrument. While the instrument is relatively expensive and is heavy when using the collimator, it required little to no maintenance. The instrument installed relatively easily above ground (away from animals and interference), had low power consumption and operated on 12VDC, sensor readings are not impacted by disturbance or the instrument infrastructure, and the footprint is dependent on installation height (and is not a point measurement).

Overestimation of SWE by the sensor following periods of melt could potentially impact data quality but may only be significant for sandy soils, which are prominent at both test sites. This is summarized in another section of the SPICE final report and in more detail in Smith et al. (2016).

References

Powell, D.: Observations on consistency and reliability of field data in snow survey measurements, paper presented at 55th Annual Meeting Western Snow Conference, Vancouver, B. C., 1987, 69‑77, 1987.

Goodison, B.E ,Glynn, J.E. ,Harvey, K.D.,Slater, J.E.: Snow surveying in Canada: A perspective, Canadian Water Resources Journal, 12(2), 27-42, DOI: 10.4296/cwrj1202027, 1987.

SPICE Instrument Performance Report
Campbell Scientific - SR50A/ATH

1) Technical Specifications

Physical principle: determines distance to a target by sending ultrasonic pulses, listening for an echo, and using the speed of sound to determine distance
Measurement Area: 30° beam angle from centre
Measurement Time: < 1sec
Measurement Range: 0.5 to 10 m
Measurement Accuracy: ± 1 cm or 0.4% of distance to target
Resolution: 0.25 mm
Link to manual: https://s.campbellsci.com/documents/us/manuals/sr50a.pdf

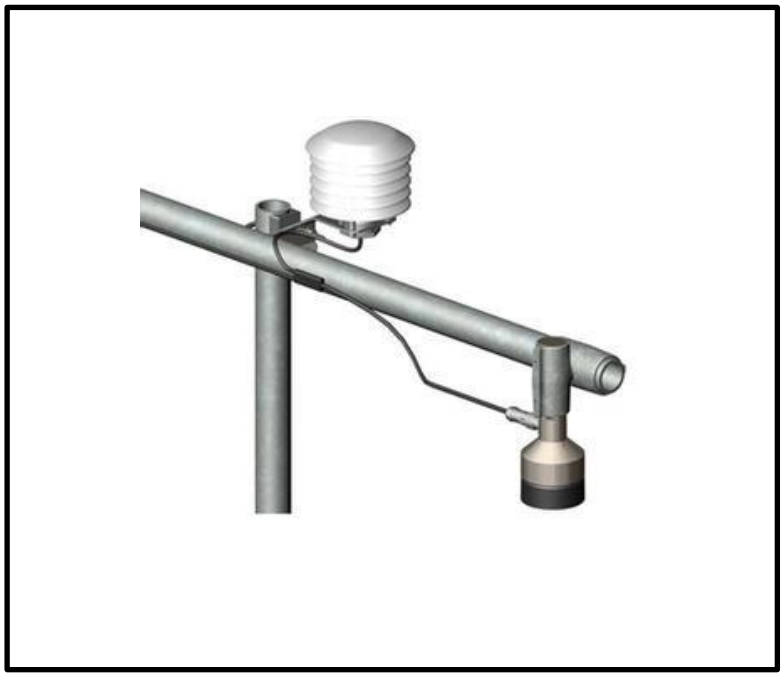


Figure1: Photo of the SR50ATH snow depth sensor

2) SPICE Test Configuration

Test Sites: Sodankylä (Finland), Col de Porte (France), and CARE (Canada)

The SR50ATH, as provided by the manufacturer, was tested at both Sodankylä and Col de Porte. The SR50A, as provided by the site host, was tested at CARE. One of the sensors at Sodankylä was mounted on a horizontal mounting beam (Figure 3) while one sensor was mounted on a heated angled beam (Figure 4).

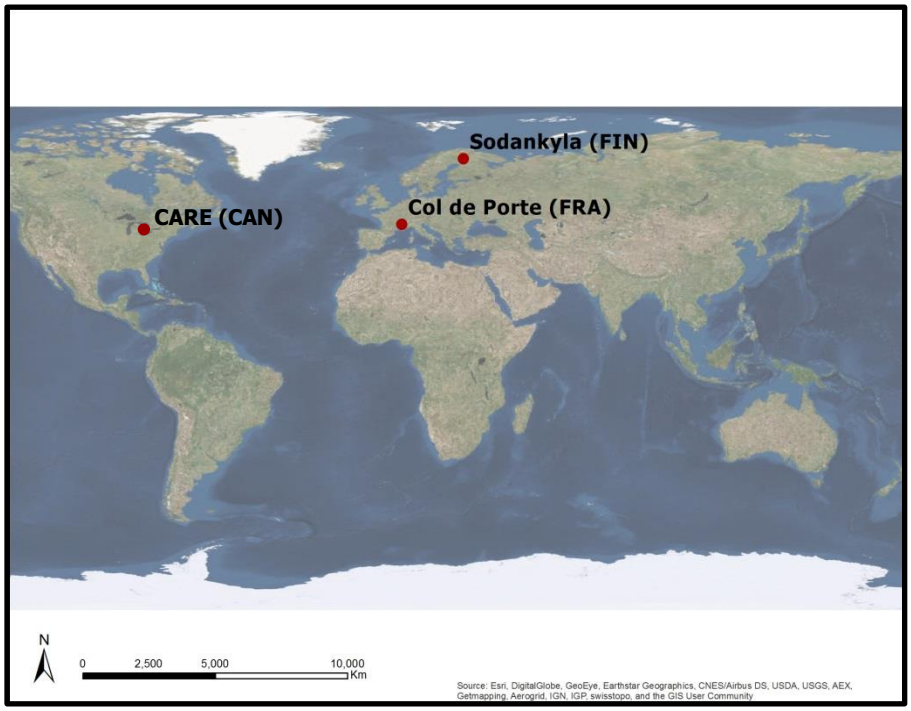


Figure 2: Location of the sensors under test

2.1) Site Specifications

Table 1: Site specific instrument and installation details

	Sodankylä	Col de Porte	CARE
Model Under Test	SR50ATH x 2	SR50ATH x 1	SR50A x 3
Date of Installation	2013-10-04, 2012-12-21	2014-01-15	2013-10-23
Serial Number(s)	4710; 4711	5327	2471;5203;5206
Installation Location	60:62, 70:52	SPICE Beam	12A; 11A; 20
Distance to Ground	2.0m, 2.0m	4.0m	1.67m;1.75m;1.70m
Target	Green artificial grass mats, 2m x 2.5m	Natural grass, mown	1.2mx1.2m grey textured plastic
Height of Temperature	2.05m, 2.05m	4.0m	1.5m
Instrument Heating	Yes, Yes	Yes	No
Mounting Beam	Horizontal, not heated; 45°, heated	Horizontal, not heated	Horizontal, not heated

2.2) Site Photos

Figure 3: Sodankylä 60:62, horizontal mounting beam

Figure 4: Sodankylä 70:52, angled mounting beam

Figure 5: Col de Porte, sensor shown inside blue square

Figure 6: CARE pedestal 12A, sensor and target inside red square

2.3) Instrument Footprint Diagram

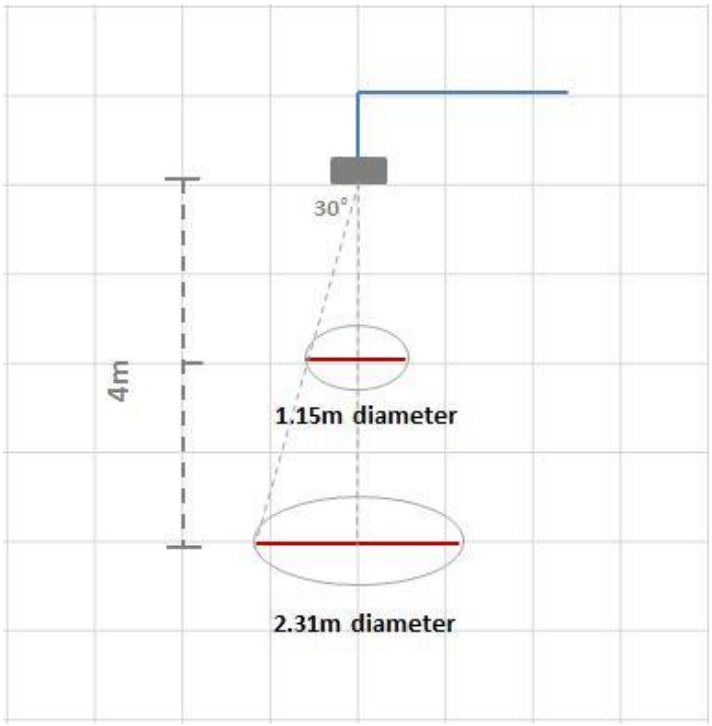


Figure 7: Conceptual diagram of instrument field of view

2.4) Environment Conditions During SPICE

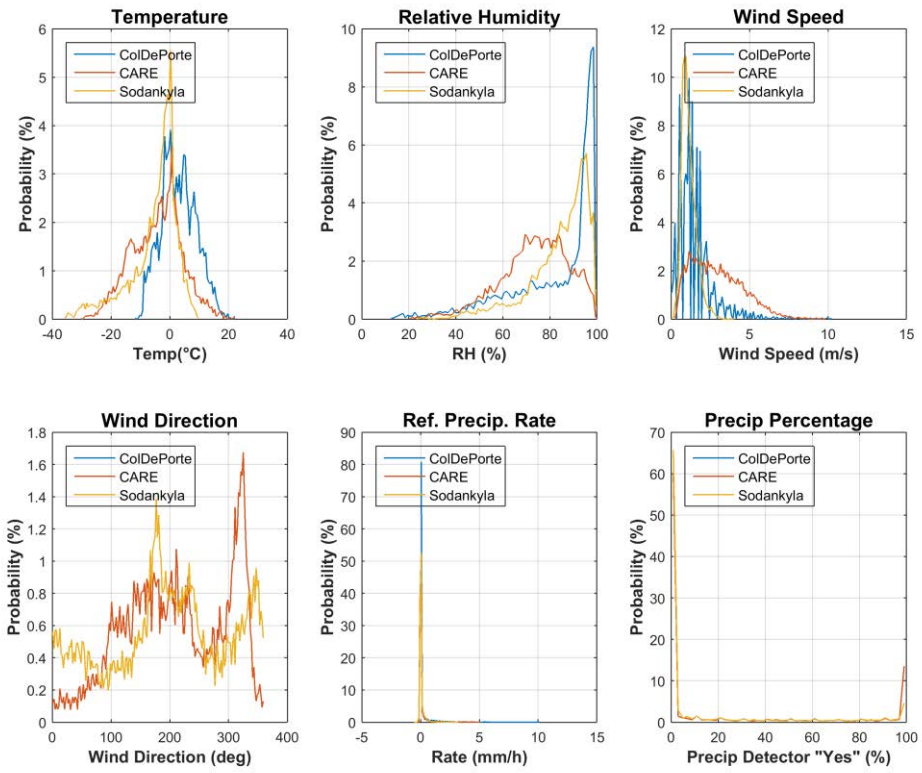


Figure 8: Summary of aggregated environmental conditions on the SPICE sites that operated a SR50A/TH for the entire duration of formal tests.

3) Data Output

Table 2: Parameters measured or output by the instrument

Measured Parameters	Units	Sodankylä	Col de Porte	CARE
Corrected Snow Depth	cm	x	x	x
Measurement Quality	-	x	x	x
Sensor Status	-	x		
Temperature	Deg C	x	x	

Table 3: Measurement and data acquisition parameters

Measurement and Processing	Units	Sodankylä	Col de Porte	CARE
Data Sampling Rate	[Minutes]	1	1	0.5
Data Acquisition Interval	[Minutes]	1	1	0.5
Data Processing		Sample	Sample	1-min Average
Communication Protocol		RS232	SDI-12	SDI-12

4) Quality Control Information

Missing data occured when data were not recorded by the data logger. This may or may not have been related to the function of the instrument. Suspicious data were identified as outliers and were either removed automatically via pre-set range and jump filtering, or identified visually and removed manually. Range and jump filters are set based on reasonable physical limits for the individual sites. Suspicious data may be a result of a malfunctioning instrument, or from meteorological impacts on the instrument's measurement capability. Table 4 provides a seasonal breakdown of the quality control metrics for each site. More commentary on individual instruments, where required, is included below. Only data classified as "Good" were used for the intercomparisons.

Table 4: Seasonal breakdown of data QC metrics

SEASON 2013-2014	**Sodankylä**	**Col de Porte**	**CARE**
Collection Period	Oct 2013 – June 2014	Jan 2014 – April 2014	Nov 2013 – April 2014
Good	97.0%	97.8%	98.3%
Missing	2.9%	2.0%	1.6%
Suspicious	0.1%	0.1%	0.1%

SEASON 2014-2015	**Sodankylä**	**Col de Porte**	**CARE**
Collection Period	Oct 2014 – June 2015	Nov 2014 – April 2015	Nov 2014 – April 2015
Good	99.0%	99.7%	97.6%
Missing	0.9%	0.0%	2.3%
Suspicious	0.1%	0.3%	0.1%

5) Evaluation of the Ability to Perform Over a Range of Operating Conditions

There are several options for comparing sensors under test to a reference. Manual snow depth measurements were performed in the Intercomparison Field at all three of these SPICE sites and are used as a measurement reference. Also, these sites hosted several automated snow depth sensors that when combined as an average, also served as a reference. Intercomparisons with both the manual and the automated references are shown below for each site.

6.1) Performance against the reference

6.1a Sodankylä

At Sodankylä, the manual reference consisted of a daily photograph observation of four graduated snow stakes distributed in the Intercomparison Field. The intercomparison below (Figure 9) shows the instrument plotted against the average of all stakes (red) and against Stake 3 which was closest to the instrument (blue). The seasonal time series for this intercomparison are plotted in Figures 10 through 13. The automated reference is an average of 1-minute snow depth data obtained from 6 instruments distributed in the intercomparison field. These six instruments include the two Campbell Scientific SR50ATH sensors, two Sommer USH-8 sensors, one Jenoptik/Lufft SHM30 and one Felix SL300. The SR50ATs are plotted against this reference in Figure 14.

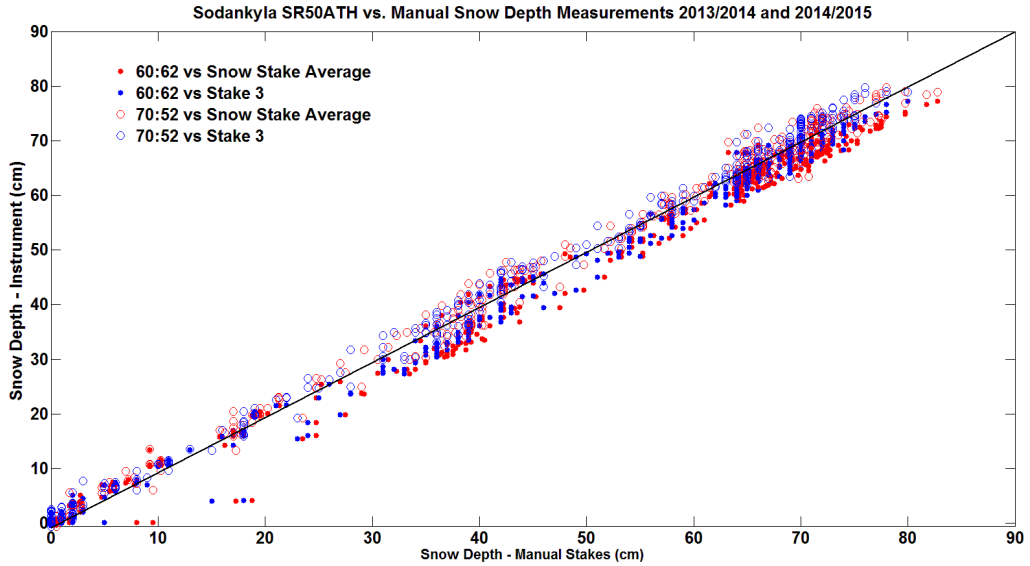


Figure 9: SR50ATH at Sodankylä compared with the manual refererence which is either the average of the four snow stakes (red) or the closest snow stake to the sensor (blue). The intercomparison statistics for the four stake average are shown in the tables in Section 5.1d.

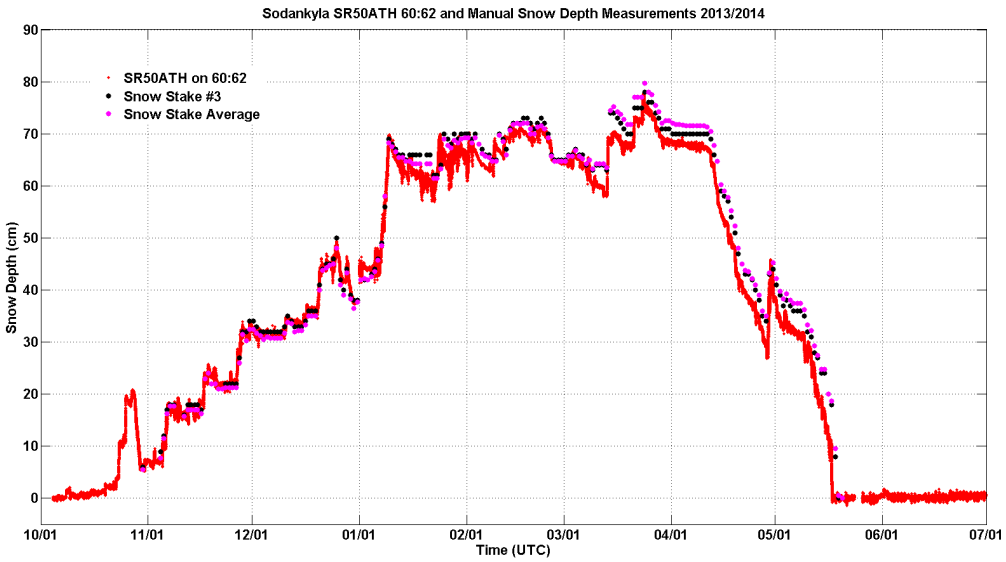


Figure 10: Time series of the SR50ATH on pedestal 60:62 and manual snow depths at Sodankylä for 2013/2014 including both the four snow stake average (magenta) and the snow stake closest to the sensor (black).

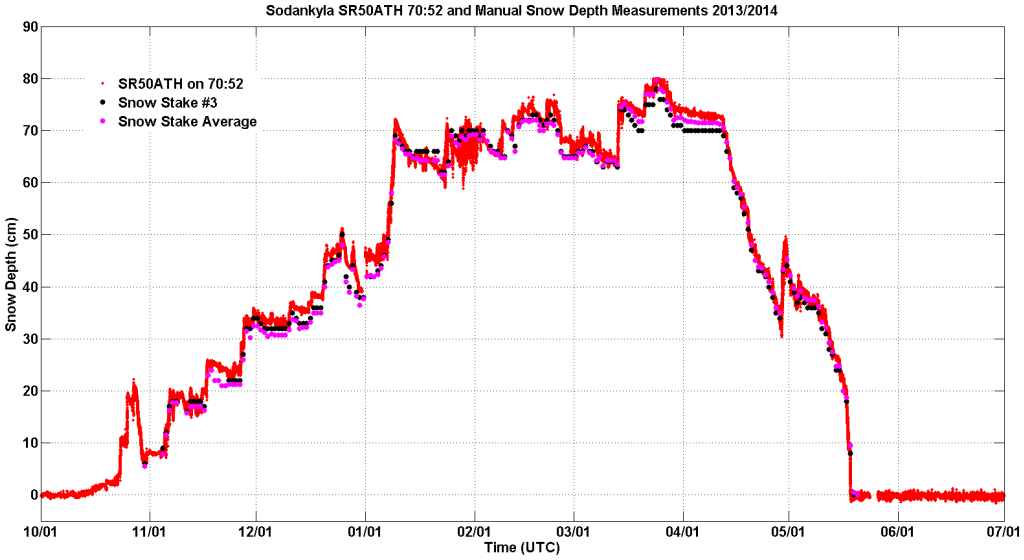


Figure 11: Time series of the SR50ATH on pedestal 70:52 and manual snow depths at Sodankylä for 2013/2014 including both the four snow stake average (magenta) and the snow stake closest to the sensor (black).

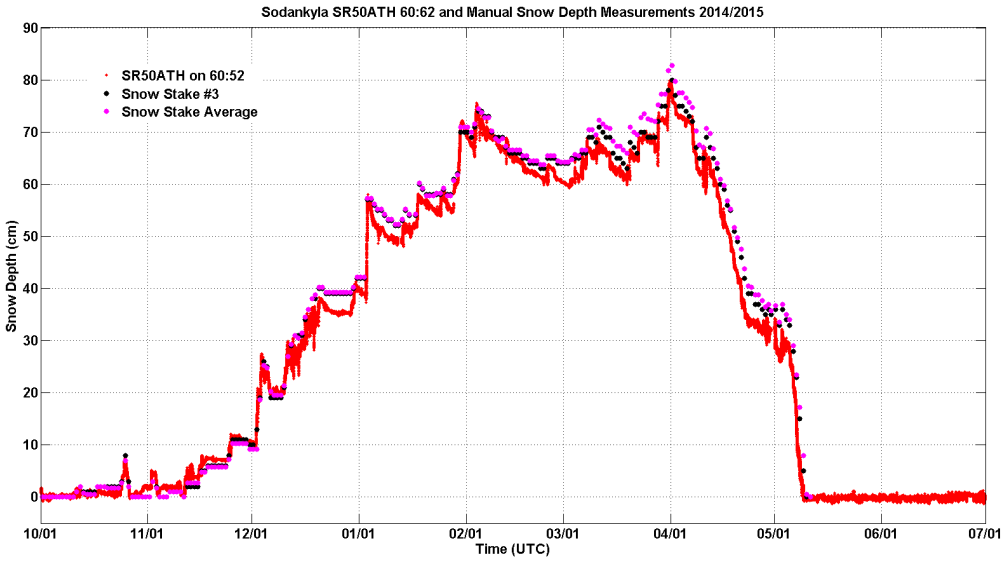


Figure 12: Time series of the SR50ATH on pedestal 60:62 and manual snow depths at Sodankylä for 2014/2015 including both the four snow stake average (magenta) and the snow stake closest to the sensor (black).

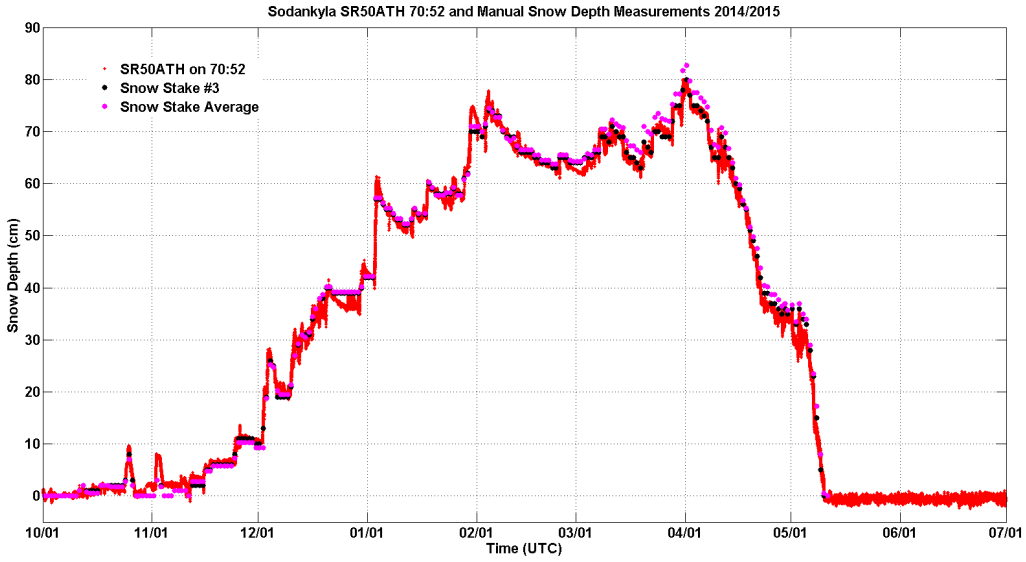


Figure 13: Time series of the SR50ATH on pedestal 70:52 and manual snow depths at Sodankylä for 2014/2015 including both the four snow stake average (magenta) and the snow stake closest to the sensor (black).

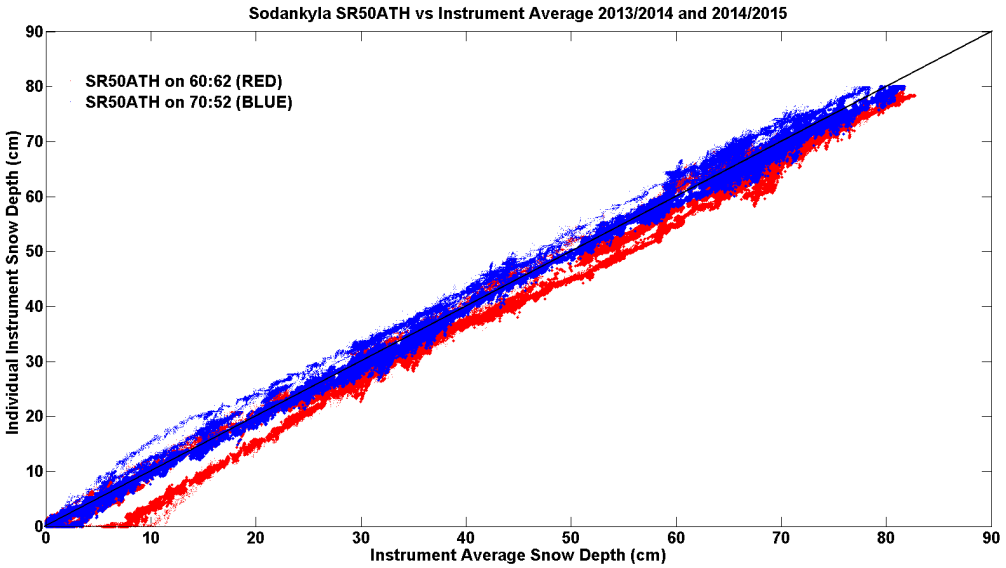


Figure 14: The SR50ATH compared with the average of five sensors at Sodankylä. The intercomparison statistics for the six sensor average are shown in the tables in Section 5.1d.

6.1b Col de Porte

At Col de Porte, the manual measurement consisted of a visual measurement of 3 snow stakes on a weekly basis. Of the 3 stakes, two are closer to the automated sensors than the other. For the manual intercomparison, the average of all stakes is plotted as red and the average of the two closest stakes (North and South) is plotted as blue in Figure 15. The corresponding time series are shown in Figures 16 and 17 for the 2013/2014 and 2014/2015 seasons respectively. The automated reference was an average of 1-minute snow depth data obtained from five automated sensors measuring roughly the same target under the SPICE beam. These sensors were a Campbell Scientific SR50ATH, a Campbell Scientific SR50A, a Jenoptik/Lufft SHM30, a Dimetix FLS-CH 10 and an Apical Technologies sensor. This is shown in Figure 18.

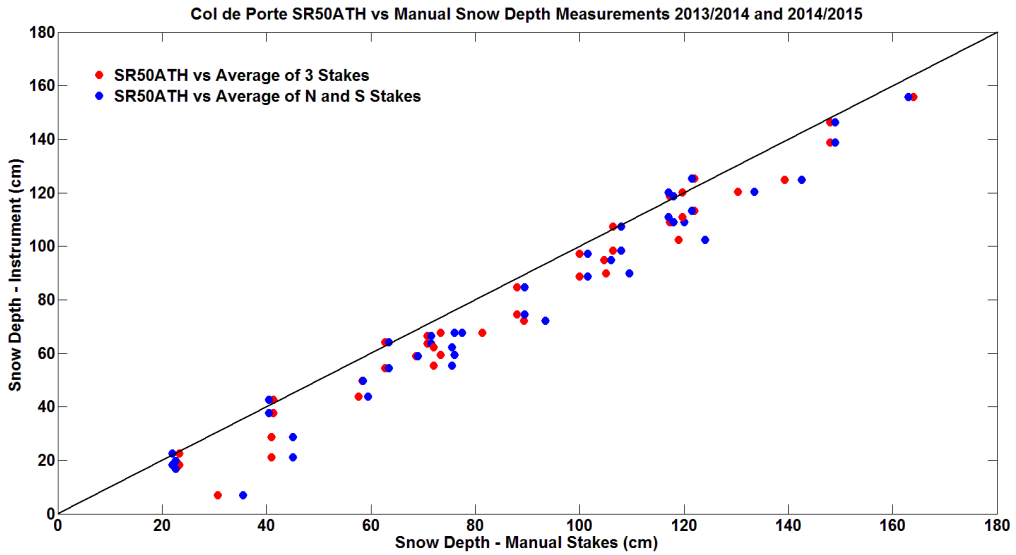


Figure 15: SR50ATH at Col de Porte compared with the manual refererence which is either the average of all three snow stakes (red) or the two closest snow stakes to the sensor (blue). The intercomparison statistics for the two (closest) snow stake average are shown in the tables in Section 5.1d

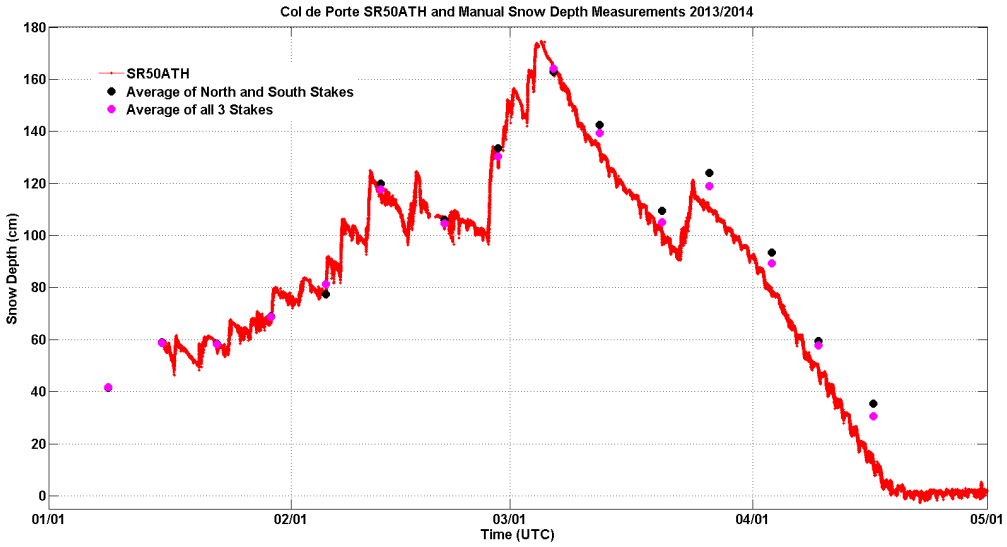


Figure 16: Time series of the SR50ATH and the manual snow depths at Col de Porte for 2013/2014 including both the three snow stake average (magenta) and the average of the two snow stake closest to the sensor (black).

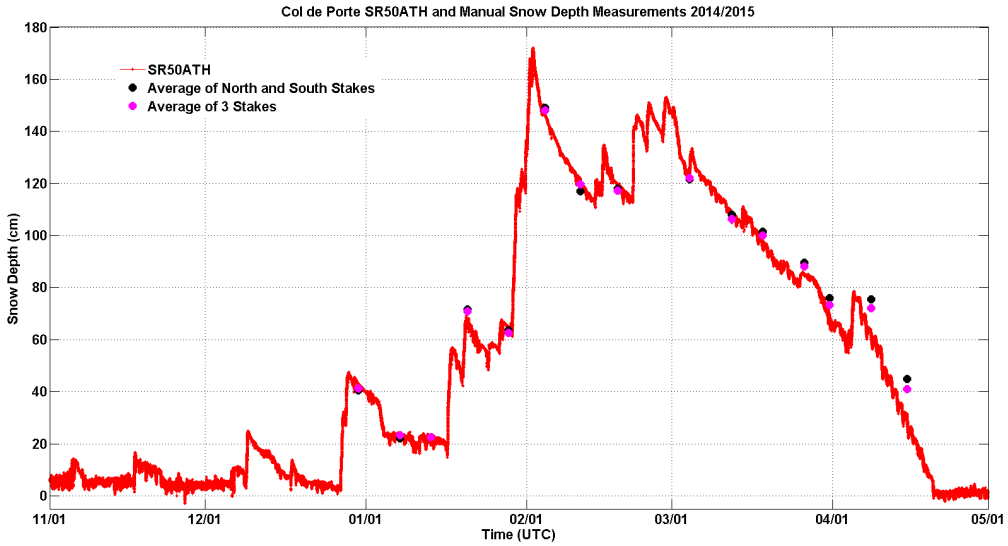


Figure 17: Time series of the SR50ATH and the manual snow depths at Col de Porte for 2014/2015 including both the three snow stake average (magenta) and the average of the two snow stake closest to the sensor (black).

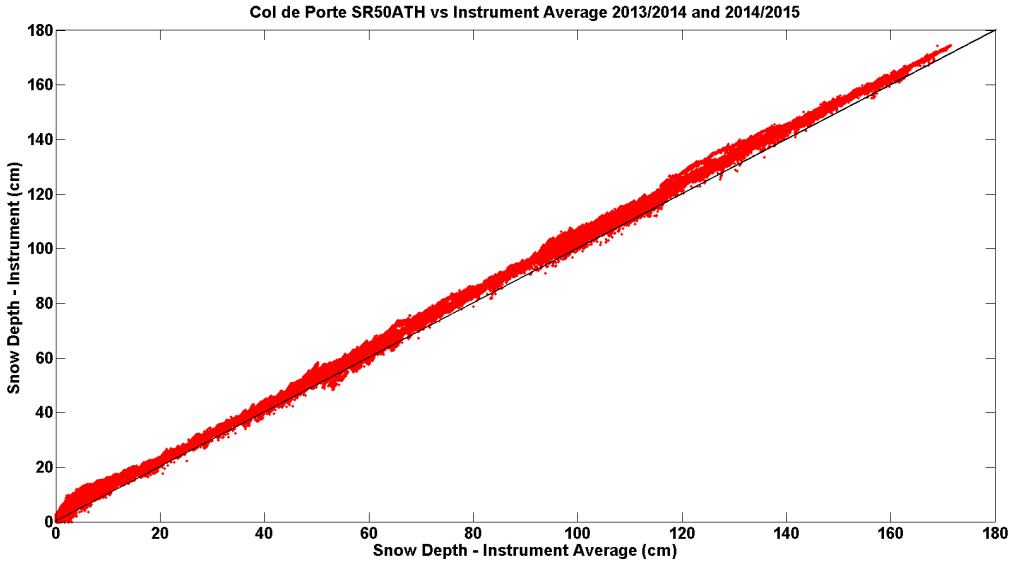


Figure 18: The SR50ATH compared with the average of five sensors at Col de Porte. The intercomparison statistics for the five sensor average are shown in the tables in Section 5.1d.

6.1c CARE

At CARE, each of the three pedestals (12A, 11A, and 20) included an SR50A that was aimed at a 1.2m x 1.2m plastic target under the sensor. Each target had a permanent snow stake at each corner that was observed visually each day. The average of all four stakes at each target comprised the manual reference for each of the three pedestals. This intercomparison is shown in Figure 19, with the time series shown in Figures 20 (2013/2014) and 21 (2014/2015). The automated reference consisted of the four automated instruments at each of the three pedestals averaged for each minute. These instruments were the Campbell Scientific SR50A, a Jenoptik/Lufft SHM30, a Sommer USH-8 and a Felix SL300. Each pedestal, therefore, has an independent automated reference. This intercomparison, separated by pedestal, is shown in Figure 22.

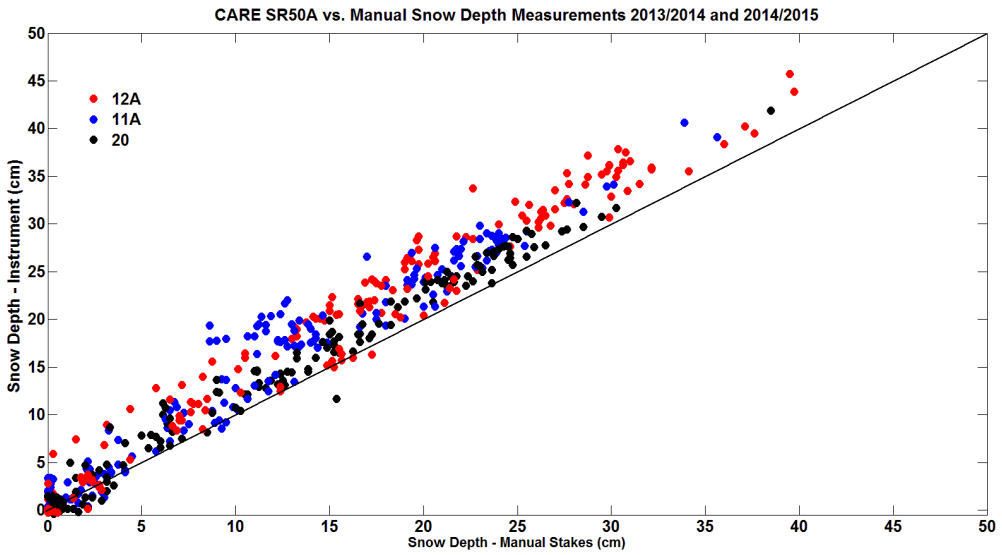


Figure 19: SR50A at CARE compared with the manual refererence which is the average of four snow stakes at the corner of the targets 12A (red), 11A (blue) and 20 (black). The intercomparison statistics are shown in the tables in Section 5.1d.

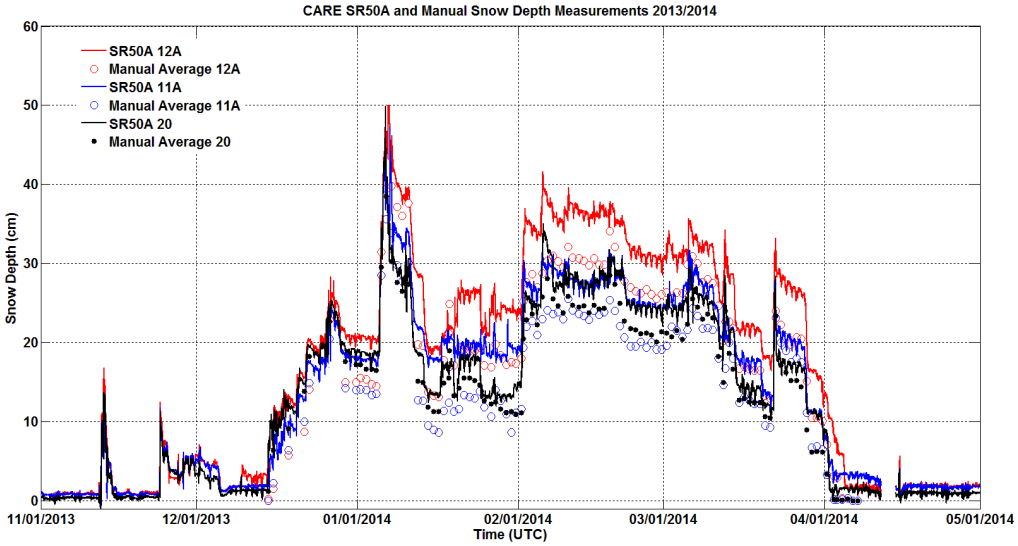


Figure 20: The time series for the SR50As on pedestals 12A (red), 11A (blue) and 20 (black) and their corresponding average manual snow depths (circles of the same colour) at CARE for 2013/2014.

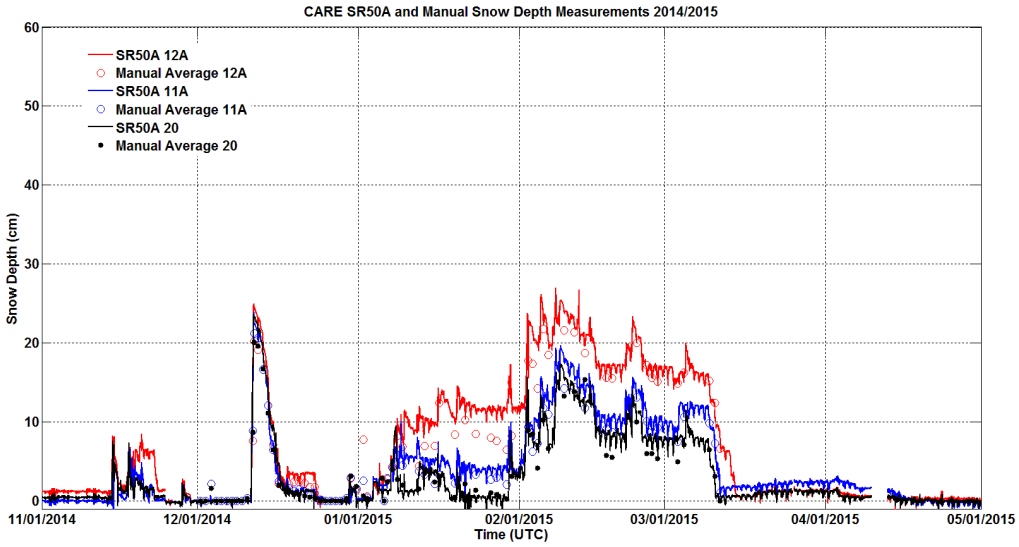


Figure 21: The time series for the SR50As on pedestals 12A (red), 11A (blue) and 20 (black) and their corresponding average manual snow depths (circles of the same colour) at CARE for 2014/2015.

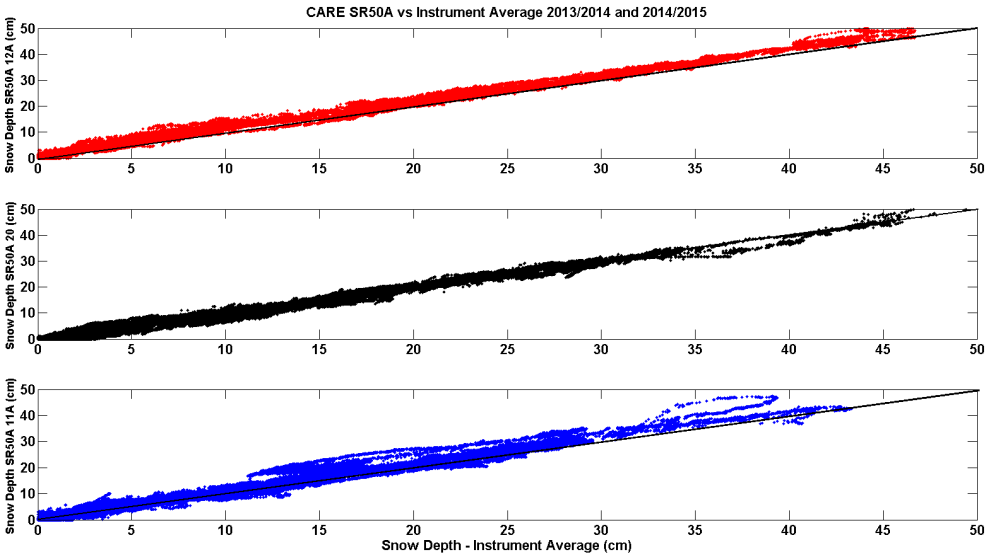


Figure 22: The three SR50As compared with the average of four sensors at pedestals 12A (top), 20 (middle) and 11A (bottom) at CARE. The intercomparison statistics for each of the pedestal intercomparisons are shown in the tables in Section 5.1d.

5.1d Intercomparison Statistics Summary

Table 4: Summary statistics for SR50ATH on pedestal 60:62 at Sodankylä

Sodankylä 60:62	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.98	0.96	-0.14	3.4	195
	2014-2015	0.99	0.94	0.02	3.6	221
Automated Reference	2013-2014	0.99	0.97	0.49	2.1	380768
	2014-2015	0.99	0.96	-0.30	2.3	388456

Table 5: Summary statistics for SR50ATH on pedestal 70:52 at Sodankylä

Sodankylä 70:52	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.99	1.00	1.83	2.3	193
	2014-2015	0.99	0.97	0.60	2.1	221
Automated Reference	2013-2014	0.99	1.03	0.92	2.3	378715
	2014-2015	1.00	0.99	-0.18	0.81	389284

Table 6: Summary statistics for the SR50ATH at Col de Porte

Col de Porte	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.97	1.07	-13.48	9.8	13
	2014-2015	0.98	1.04	-6.20	6.4	15
Automated Reference	2013-2014	1.00	1.02	2.00	3.9	148468
	2014-2015	1.00	1.01	2.97	4.0	260028

Table 7: Summary statistics for SR50A on pedestal 12A at CARE

CARE 12A	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.96	0.99	5.49	5.5	107
	2014-2015	0.95	1.10	0.67	2.4	70
Automated Reference	2013-2014	0.99	1.04	0.71	1.6	256304
	2014-2015	0.99	1.04	0.32	1.1	254022

Table 8: Summary statistics for SR50A on pedestal 20 at CARE

CARE 20	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.97	1.01	2.13	2.6	107
	2014-2015	0.95	1.13	-0.02	1.5	70
Automated Reference	2013-2014	0.99	0.99	0.08	0.9	256282
	2014-2015	0.97	0.93	-0.49	1.1	254773

Table 9: Summary statistics for SR50A on pedestal 11A at CARE

CARE 11A	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.93	0.97	5.20	5.2	106
	2014-2015	0.97	1.10	0.19	1.4	70
Automated Reference	2013-2014	0.98	1.07	0.78	2.3	256147
	2014-2015	0.96	1.00	0.18	1.0	254667

5.2) Factors that influence instrument performance

At Sodankylä, the snow depth instrumentation was distributed in the NE quadrant of the Intercomparison Field, with the average distance between instruments at approximately 16 m. The average distance between the SR50ATH at 60:62 and the 4 snow stakes was 23 m, with the closest stake (stake 3) being 7 m from the instrument. The average distance between the SR50ATH at 70:52 and the 4 snow stakes was 25 m, with the closest stake (stake 3) being 7 m from the instrument. Even though the snow depth in the Intercomparison Field was relatively uniform, there was still some variability in depth across 7 m, and certainly across 23 m. This likely accounted for much of the scatter shown in the reference intercomparison for Sodankylä (Figure 9). It was also noted that at this site, differential melt occurred between the targets under the sensors and the snow stakes, and also amongst the targets (depending on their individual exposures). The effects of this differential melt can be seen in both the manual and automated reference plots for Sodankylä, especially with lower snow depths during mid- to late-melt. This scatter likely cannot be attributed to sensor bias.

The variability in snow depth at the Col de Porte intercomparison site was much greater than at Sodankylä. This is quite evident in the manual reference plot (Figure 15). However, all of the instruments used in the automated reference were located on the same mounting structure, and target the same area beneath the instruments which resulted in much less scatter. There was nevertheless still some variability across the target area which resulted in the SR50ATH measuring slightly higher snow depths than the instrument average through much of the intercomparison period. Further analysis of the variability of snow depth below the mounting structure was assessed using an automated rugged laser scanner (Picard et al., The Cryosphere 2016), which demonstrated that the mounting structure itself had no impact on the snow conditions below, and that most of the variation between sensors was due to the natural variability of snow depth rather than instrument uncertainty.

The design and installation of the manual snow stakes at the corner of each target at the CARE site was implemented to measure the snow depth as close as possible to the sensor, without being within the sensor's FOV. Unfortunately, the manual snow stakes created a mounding effect, increasing the depth of the snow in the middle of the target as compared to the corners of the target. This is quite evident in Figures 19, 20 and 21, which suggest that the sensor overestimated snow depth compared to the manual measurement. Another factor to consider when interpreting the intercomparison at CARE is the potential for a change in the reference distance between the target and the sensor over the course of the intercomparison season. This is called "zero snow depth drift" and is discussed in the main body of the SPICE final report. The levels of the targets at CARE relative to the sensor were susceptible to change due to settling after installation (under the weight of the snow pack) or from frost heave. The change in this position could only be assessed at the end of the season when the targets were bare, but it changes in the zero snow depth distance can be

continuous and undetectable during the season, and therefore the impact on the snow depth intercomparison was difficult to assess. However, we estimated that the impact would not exceed ± 2 cm during the season. This error is generally not a concern at the other sites where either natural targets are used or the artificial targets were not prone to heaving or settling.

5.2) Performance Considerations

The site manager at Col de Porte reported that the sensor occasionally produced false echoes, predominately during snowfall over the natural grass target. Performance may be impacted by errors in the temperature measurement used for correcting the sonic signal for the speed of sound in air. There are potentially large errors associated with measuring temperature using a naturally aspirated radiation shield. This is discussed further in the main body of the SPICE final report. No other performance considerations were noted by the site managers during the intercomparison.

Although not evident during the SPICE intercomparison, site managers have noted that the instrument is affected by ground cover or disturbances to the footprint; measurements are occasionally incorrect or missing, especially when snow falls over a grass surface. This is not necessarily a weakness of this instrument but highlights the importance of the surface target and/or surface preparation prior to the first snowfall. Because this sensor measures the distance of the closest solid object within its measurement footprint rather than the average of the entire footprint, vegetation or other such obstacles may cause spurious measurements when the measurement footprint is not entirely buried in snow. In order to measure small amounts of snow on the ground, great care must therefore be taken in the choice and maintenance of the target. The performance of various targets is discussed further in the main body of the SPICE report.

5.3) Maintenance

One issue with signal noise was noted in October 2013 for the instruments at Sodankylä. This was rectified by adjusting the way the instrument was grounded after consultation with the manufacturer. A similar problem was noted by the Col de Porte site manager in October 2013 and was rectified after consultation with the manufacturer.

6) Lessons Learned

The SR50ATH is a compact sensor with low power consumption (except for when the heater is used). It is easily installed and calibrated and operates reliably. The heating works well, but increases the power consumption considerably (3 W).

SPICE Instrument Performance Report
Dimetix FLS-CH 10

1) Technical Specifications

Physical principle: Laser ranger (620-690 nm)
Measurement Area: 4 mm at 5 m height
Measurement Range: 0.05 to ~65 m (500 m using reflective foil
Measurement Accuracy: ± 1 mm
Measurement reproducibility: ± 0.3 mm
Measurement Resolution: 0.1 mm
Operating Range: -40 °C to +50 °C
Link to manual:
http://www.dimetix.com/downloads/Manuals/DLS_FLS_C_TechnicalManual_V502_en.pdf

Figure 1: Photo of FLS-CH 10 snow depth sensor

2) SPICE Test Configuration

Test Site: Col de Porte (France)

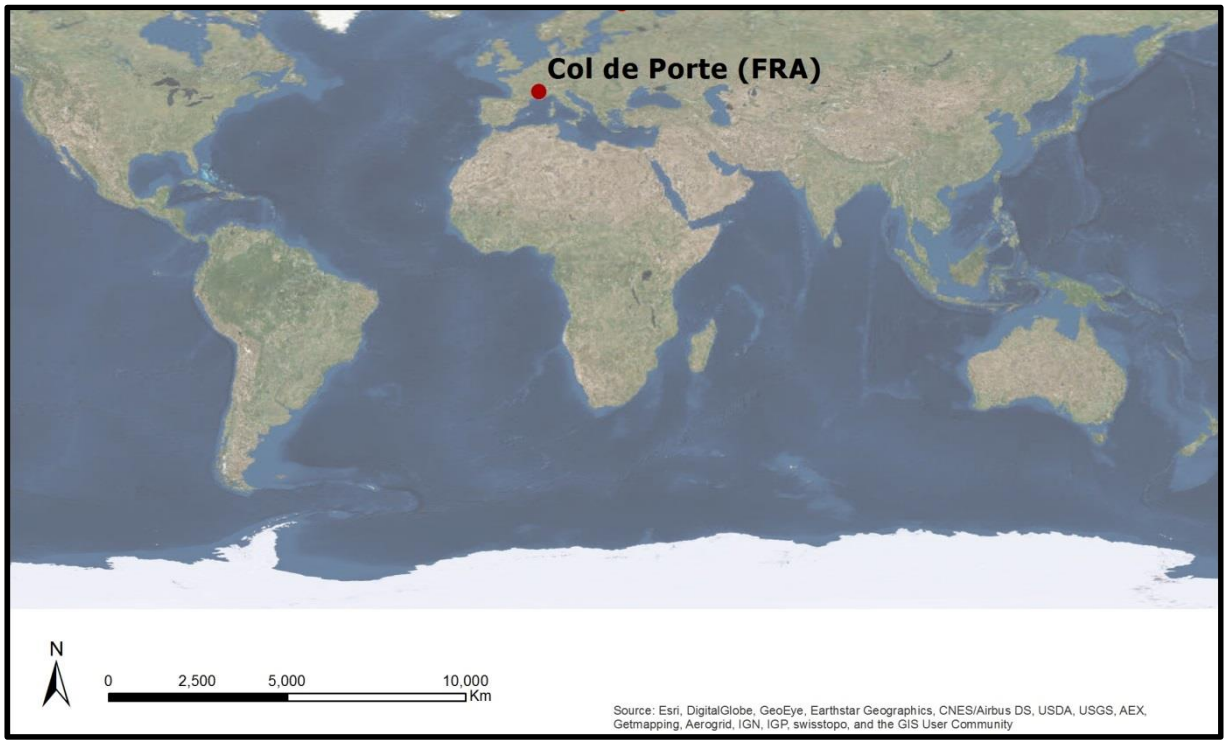


Figure 2: Location of the sensors under test

2.1) Site Specifications

Table 1: Site specific instrument and installation detail

	Col de Porte
Date of Installation	Jan. 15, 2014
Serial Number(s)	22260036
Installation Location	SPICE Beam
Firmware	0040 (measuring module) 0501 (interface)
Distance to Ground	4m
Target	Natural (mown) grass
Angle of Instrument	20°

2.2) Site Photos

Figure 3: Dimetix FLS-CH 10 on the SPICE beam at Col de Porte

2.3) Instrument Field of View

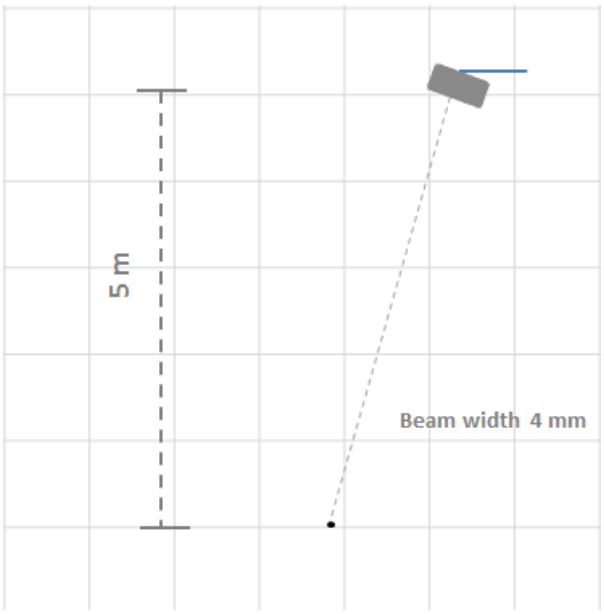


Figure 4: Conceptual diagram of instrument field of view

2.4) Environment Conditions During SPICE

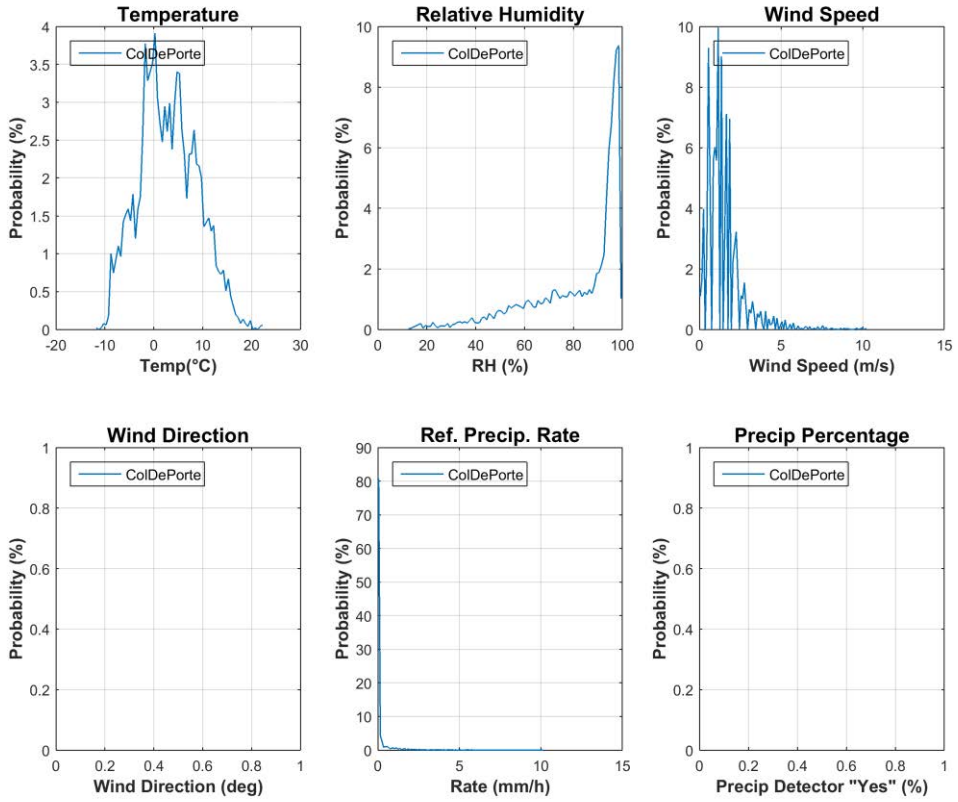


Figure 5: Summary of aggregated environmental conditions on the Col de Porte SPICE site during operation of the FLS-CH 10.

3) Data Output

Table 2: Parameters measured or output by the instrument

Measured Parameters	Units	Col de Porte
Snow Depth	m	x
Signal Strength	-	
Temperature	Deg C	
Error Code	-	

Table 3: Measurement and data acquisition parameters

Measurement and Processing	Units	Col de Porte
Data Sampling Rate	[Minutes]	1
Data Acquisition Interval	[Minutes]	1
Date Processing		Sample
Communication Protocol		RS232
Heating		Yes

4) Quality Control Information

Missing data occurred when data were not recorded by the data logger. This may or may not be related to the function of the instrument. Suspicious data was identified as an outlier and are either removed automatically via pre-set range and jump filtering, or identified visually as an outlier and removed manually. Range and jump filters were set based on reasonable physical limits for the individual sites. Suspicious data may be a result of a malfunctioning instrument or from meteorological impacts on the instrument's measurement capability. Table 4 provides a seasonal breakdown of the quality control metrics. More commentary on individual instruments, where required, is included below. Only data classified as "Good" were used for the intercomparisons.

Table 4: Seasonal breakdown of data QC metrics

SEASON 2013-2014	Col de Porte
Collection Period	Jan 2014-April 2014
Good	97.2%
Missing	2.8%
Erroneous	0.0%

SEASON 2014-2015	Col de Porte
Collection Period	Nov 2014-April 2015
Good	99.8%
Missing	0.0%
Erroneous	0.2%

5) Evaluation of the Ability to Perform Over a Range of Operating Conditions

There are several options for comparing this instrument to a reference. Manual snow depth measurements were performed inside the Intercomparison Field at Col de Porte and these can be used as a measurement reference. Also, Col de Porte hosted more than three automated snow depth sensors that when combined as an average, served as a reference. The instruments used in the automated reference at this site are the Campbell Scientific SR50ATH, a Campbell Scientific SR50A, a Jenoptik/Lufft SHM30, a Dimetix FLS-CH 10 and an Apical Technologies sensor. Both intercomparisons to the manual and automated reference (where available) are shown below.

At Col de Porte, the manual measurement consisted of a visual measurement of three snow stakes observed on a weekly basis. Of the three stakes, two were closer to the automated sensors than the other. For the manual intercomparison, the average of all stakes was plotted as red and the average of the two closest stakes were plotted as blue in Figure 6. The time series of this intercomparison are shown in Figures 7 (2013/2014) and Figure 8 (2014/2015). The automated reference is an average of 1-minute snow depth data obtained from five automated sensors measuring roughly the same target under the SPICE beam and plotted in Figure 9.

5.1) Performance against the reference

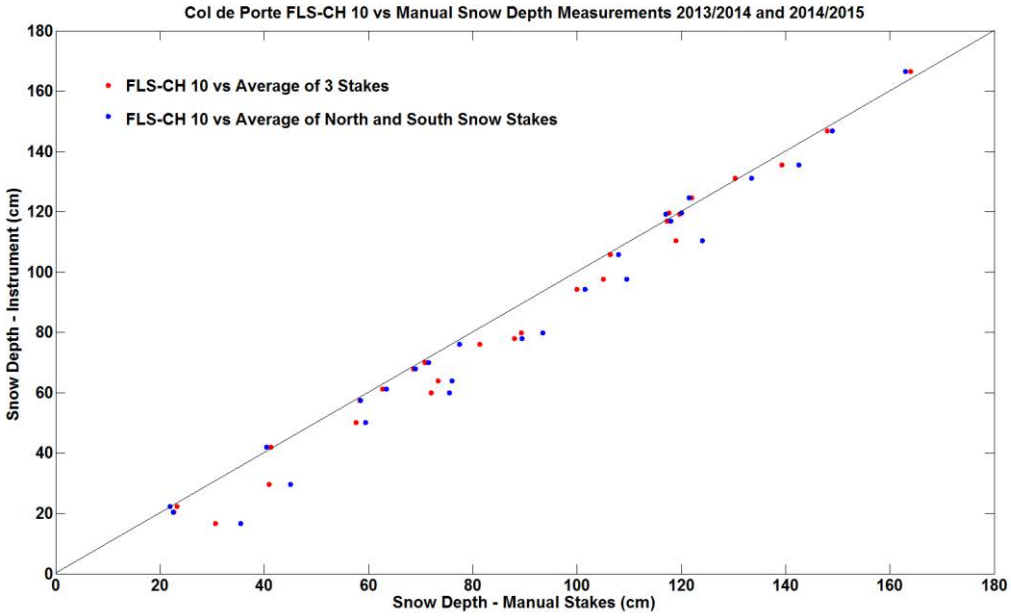


Figure 6: FLS-CH 10 at Col de Porte compared with the manual refererence which is either the average of three snow stakes (red) or the average of the two closest snow stakes to the sensor (blue). The intercomparison statistics for the two closest stake average are shown in Table 5.

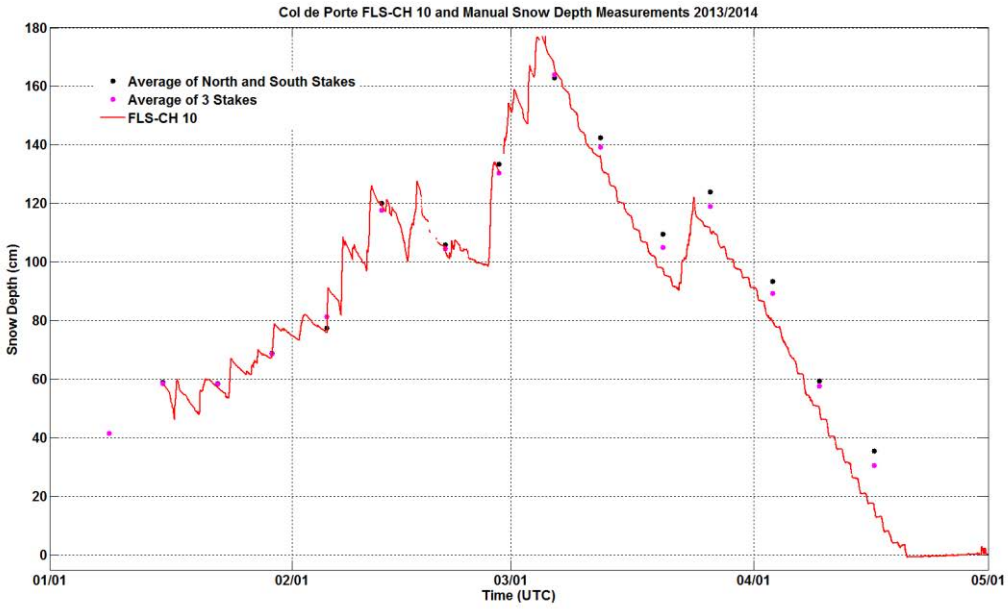


Figure 7: Time series of the FLS-CH 10 and manual snow depths at Col de Porte for 2013/2014 including both the three snow stake average (magenta) and the average of the two snow stake closest to the sensor (black).

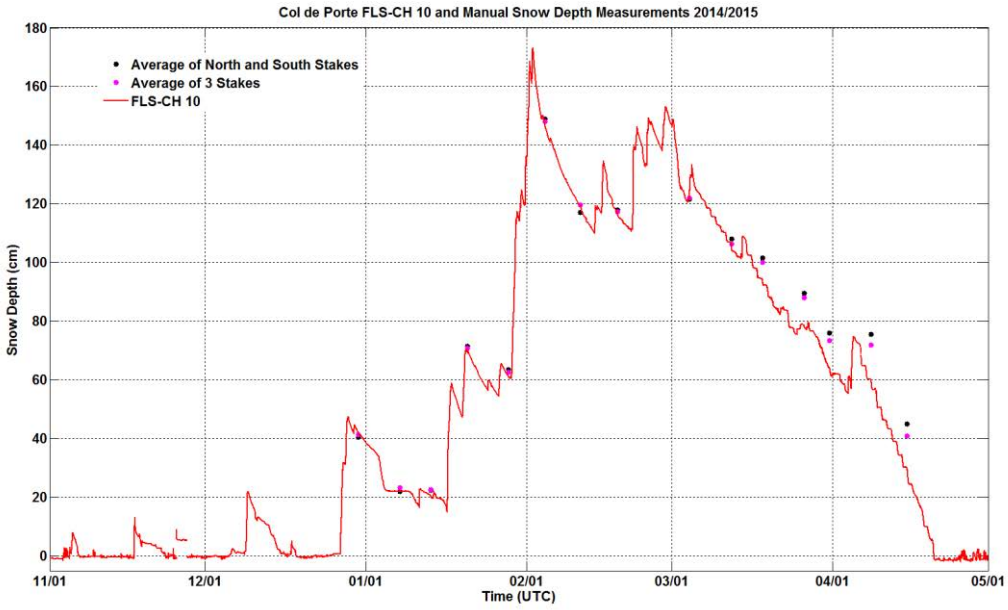


Figure 8: Time series of the FLS-CH 10 and manual snow depths at Col de Porte for 2014/2015 including both the three snow stake average (magenta) and the average of the two snow stake closest to the sensor (black).

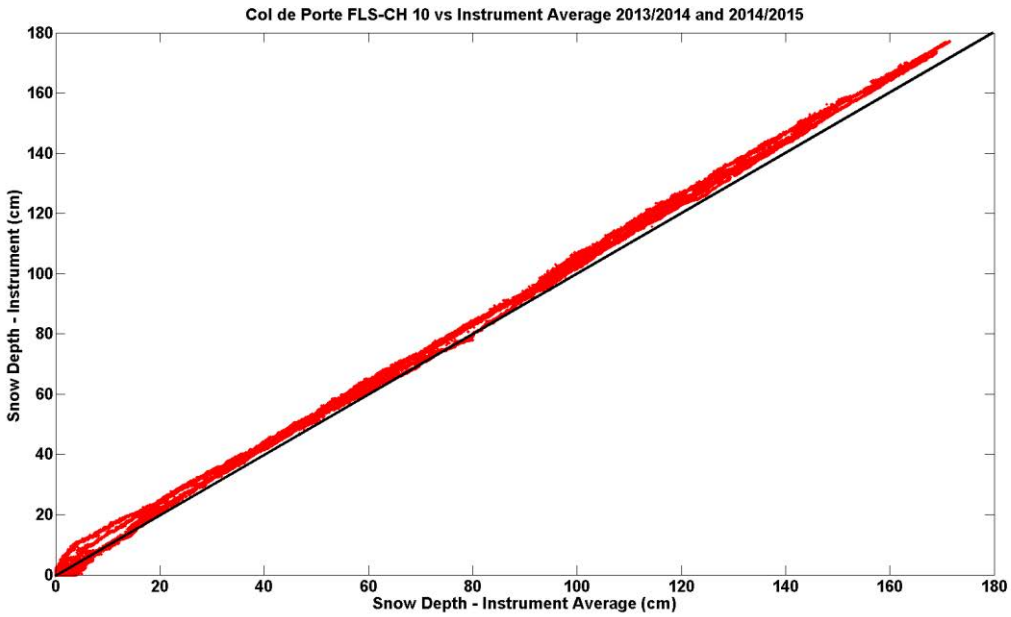


Figure 9: The FLS-CH 10 compared with the average of five sensors at Col de Port. The intercomparison statistics for the five sensor average are shown in the Table 5.

Table 5: Intercomparison statistics summary for the FLS-CH 10 at Col de Porte

Col de Porte	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	.98	1.07	-12.9	9.2	15
	2014-2015	.98	1.03	-6.5	7.5	15
Automated Reference	2013-2014	1.00	1.03	1.37	4.2	172800
	2014-2015	1.00	1.03	-0.74	2.5	260640

5.2) Factors that influence instrument performance

Although the variability in snow depth across the Col de Porte site was quite high, all of the instruments used in the automated reference (Figure 9) were located on the same mounting structure and measured roughly the same target area beneath the instruments. This resulted in a relatively strong correlation between the sensor measurements and the reference. Further analysis of the variability of snow depth below the mounting structure was assessed using an automated rugged laser scanner (Picard et al., The Cryosphere 2016) which demonstrates that the mounting structure itself had no impact on the snow conditions below, and that most of the variations between sensors were due to natural variability of snow depth rather than instrumental uncertainty.

5.3) Performance Considerations

There are no Performance considerations to be included here.

5.4) Maintenance

No maintenance items are listed by the site manager during the intercomparison

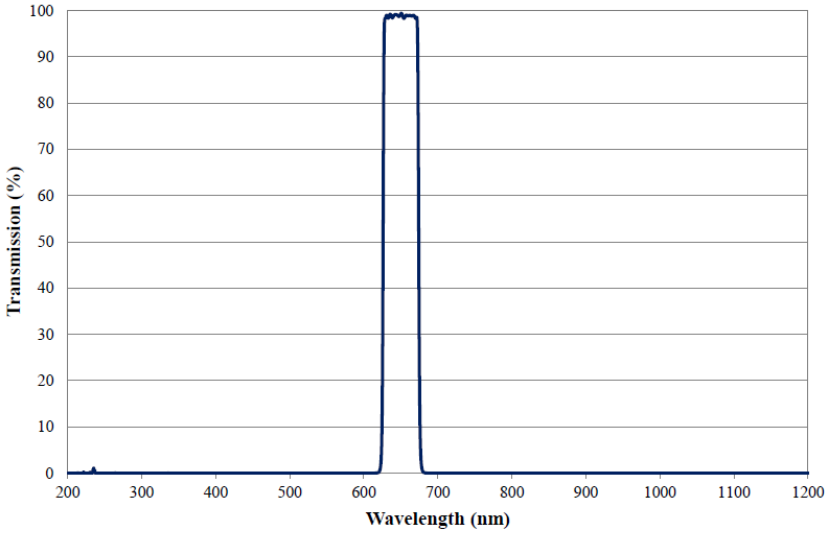
6) Lessons Learned

From Jan-April 2014, the instrument was installed and operated as provided by the manufacturer. However, on 25 November 2014, a bandpass filter was installed to reduce the amount of sunlight collected by the sensor. The specifications of the bandpass filter are available on line at http://www.edmundoptics.com/optics/optical-filters/bandpass-filters/hard-coated-od4-50nm-bandpass-filters/84786/ (link accessed on 4 July 2016) and are briefly provided below in Table 7.1. A similar filter was installed on the Rugged LaserScan device (RLS) at Col de Porte during the season 2014-2015 (see Picard et al., The Cryosphere, 2016).

Table 6: Specification of a 650nm CWL, 25mm Dia. Hard Coated OD 4 50nm Bandpass Filter, reference #84-786 by Edmund Optics used on the FLS-CH 10 from 25 November 2014.

Diameter (mm): 25.0	Optical Density OD: ≥4.0
Diameter Tolerance (mm): +0.0/-0.1	Blocking Wavelength Range (nm): 200 - 1200
Center Wavelength CWL (nm): 650	Surface Quality: 80-50
Center Wavelength CWL Tolerance (nm): ±5	Typical Light Sources: 635-670nm Laser; 660nm LED
Coating: Hard Coated	Substrate: Fused Silica
Full Width-Half Max FWHM (nm): 50	Type: Bandpass Filter
Full Width-Half Max FWHM Tolerance (nm): ±5	Manufacturer: Edmund Optics
Mount Thickness (mm): 5.0	RoHS Compliant
Transmission (%): ≥90	

650nm Hard Coated Broadband Bandpass Interference Filter: 50nm FWHM
OD >4.0 Coating Performance
FOR REFERENCE ONLY



Edmund optics | worldwide
www.edmundoptics.com

SPICE Instrument Performance Report
Felix Technology - SL300

1) Technical Specifications

Physical principle: Ultrasonic pules to measure the distance from the sensor to a target
Measurement Area: 15° beam angle from centre
Measurement Range: 0.45-6.10m
Measurement Accuracy: ±0.1% over entire range
Link to manual:
http://www.wmo.int/pages/prog/www/IMOP/intercomparisons/SPICE/Manuals/Felix_SL-300_v1-1-2.pdf

Figure 1: Photo of the SL300 sensor under test

2) SPICE Test Configuration

Test Sites: Sodankylä (FIN), CARE (CAN)

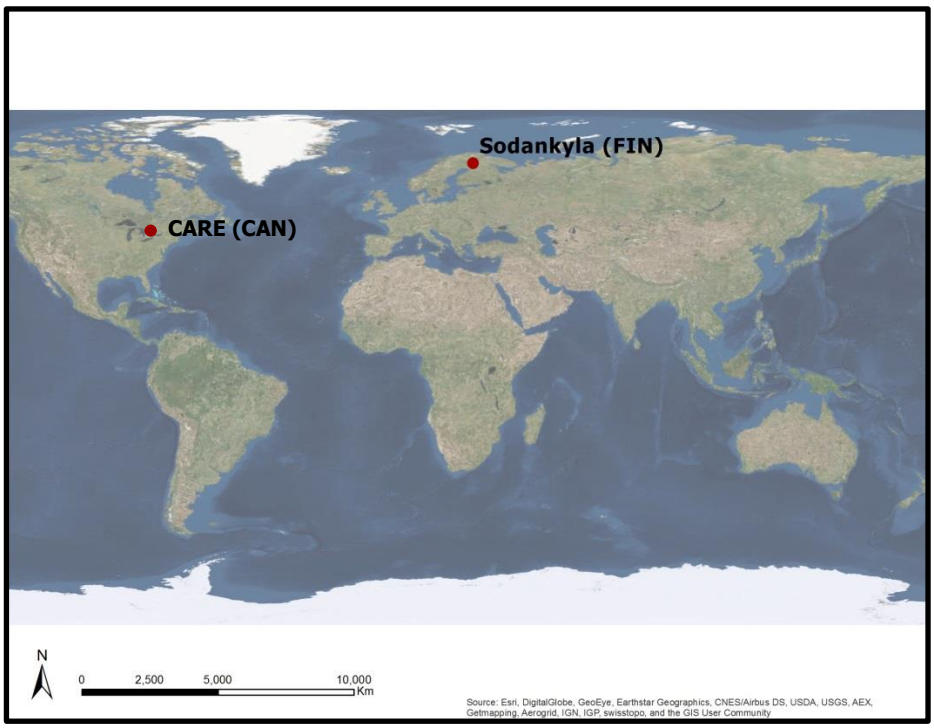


Figure 2: Location of the sensors under test

2.1) Site Specifications

Table 1: Site specific instrument and installation details

	Sodankylä	CARE
Date of Installation	2012-11-15	2013-10-23
Serial Number(s)	**Figure 1.1:** Location of the sensors under test 20120193	120190, 20120191, 20120192
Installation Location	60:32	20,11A,12A
Distance to Ground	2.0 m	2.06 m, 1.92 m, 2.07 m
Target	Green artificial grass mats 2m x 2.5m	Textured Plastic, 1.2mx1.2m
Height of Temperature	Integrated	Integrated
Instrument Heating	No	No
Mounting Beam	Horizontal	Horizontal

2.2) Site Photos

Figure 2: Sodankylä

Figure 3: CARE (Pedestal 11A), sensor and target indicated by red

2.3) Instrument Footprint Diagram

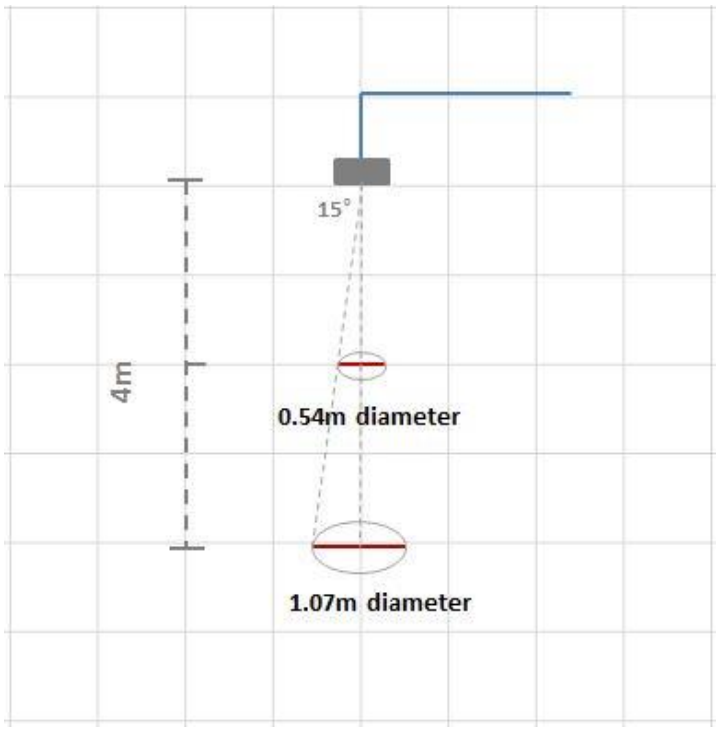


Figure 4: Conceptual diagram of instrument field of view

2.4) Environment Conditions During SPICE

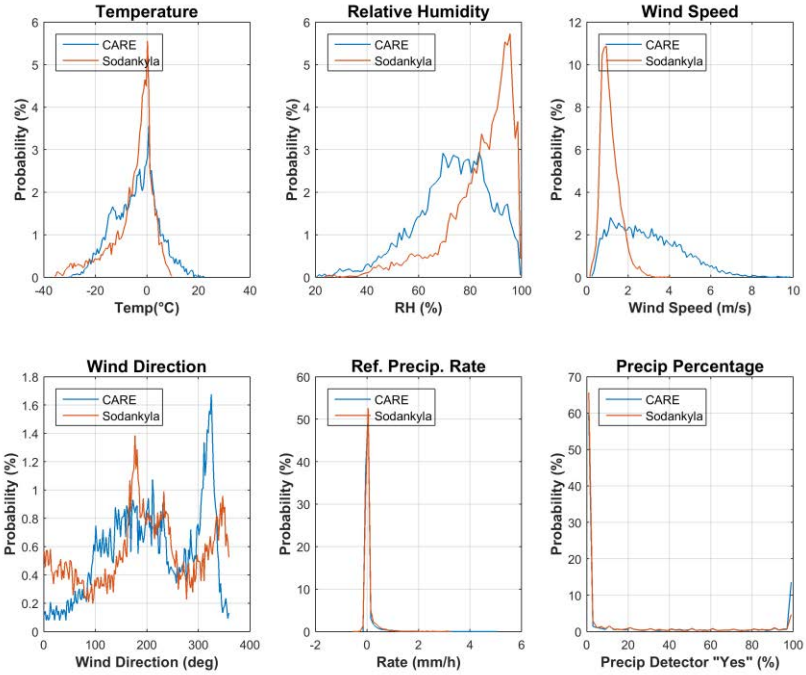


Figure 5: Summary of aggregated environmental conditions on the Sodankylä and CARE SPICE sites during operation of the SL300.

3) Data Output

Table 2: Parameters measured or output by the instrument

Measured Parameters	Units	Sodankylä	CARE
Temperature	Deg C	x	
Travel Time	ms	x	
Corrected Snow Depth	mm	x	x
Measured Distance to Target	mm	x	

Table 3: Measurement and data acquisition parameters

Measurement and Processing	Units	Sodankylä	CARE
Data Sampling Rate	[Seconds]	0.45	30
Data Acquisition Interval	[Minutes]	1	1
Date Processing		Averaged	Averaged
Communication Protocol		RS-485	SDI-12

4) Quality Control Information

Missing data occurred when data were not recorded by the data logger. This may or may not be related to the function of the instrument. Suspicious data were identified as outliers and were either removed automatically via pre-set range and jump filtering, or identified visually as an outlier and removed manually. Range and jump filters were set based on reasonable physical limits for the individual sites. Suspicious data may be a result of a malfunctioning instrument or from meteorological impacts on the instrument's measurement capability. Table 4 provides a seasonal breakdown of the quality control metrics by site. More commentary on individual instruments, where required, is included below. Only data classified as "Good" were used for the intercomparisons.

Table 3: Seasonal breakdown of data QC metrics

SEASON 2013-2014	Sodankylä	CARE
	Oct 2013 – June 2014	Nov 2013 - April 2014
Good	98.0%	98.3%
Missing	1.9%	1.6%
Suspicious	0.1%	0.1%

SEASON 2014-2015	Sodankylä	CARE
	Oct 2014 – June 2015*	Nov 2014 – April 2015
Good	69.8%	97.3%
Missing	2.5%	2.6%
Suspicious	27.7%	0.1%

*one of the two instruments tested at Sodankylä during the 2014/2015 season failed mid-season due to an electronics issue.

The sensor at Sodankylä began malfunctioning 9 December 2014 outputting large negative values with intermittent good values, resulting in a large percentage of “Suspicious” values flagged by the quality control process. It is difficult to pinpoint the cause of this intermittently bad data but the site manager believes that it is weather related. The instrument was returned to the manufacturer and the diagnosis was a malfunctioning electronics board. This was the only SL300 that failed during testing.

5) Field Intercomparison Sensor Performance

There were several options for comparing sensors under test to a reference. Manual snow depth measurements were performed in the Intercomparison Field at these two SPICE sites and were used as a measurement reference. Also, these sites hosted several automated snow depth sensors that when combined as an average, served as a reference. Both intercomparisons are shown below in Figures 6 through 13.

5.1) Performance against the reference

6.1a Sodankylä

At Sodankylä, the manual reference consisted of a daily photograph observation of four graduated snow stakes distributed in the Intercomparison Field. The intercomparison below (Figure 6) shows the instrument plotted against the average of all stakes (red) and against Stake 4 which was closest to the instrument (blue). The time series of this intercomparison is shown in Figure 7 (2013/2014) and Figure 8 (2014/2015). The automated reference was an average of 1-minute snow depth data obtained from 6 instruments (two Campbell Scientific SR50ATH sensors, two Sommer USH-8 sensors, one Felix SL300 sensor and one Jenoptik/Lufft SHM30 sensor) distributed in the Intercomparison Field. The SL300 is plotted against this reference in Figure 9.

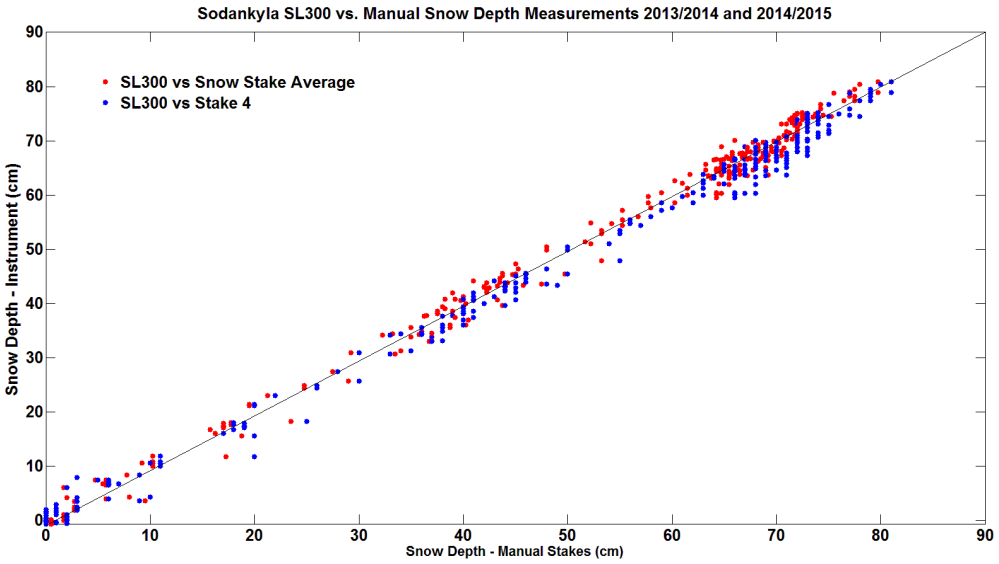


Figure 6: SL300 at Sodankylä compared with the manual refererence which is either the average of the four snow stakes (red) or the closest snow stake to the sensor (blue). The intercomparison statistics for the four stake average are shown in the tables in Section 5.1c.

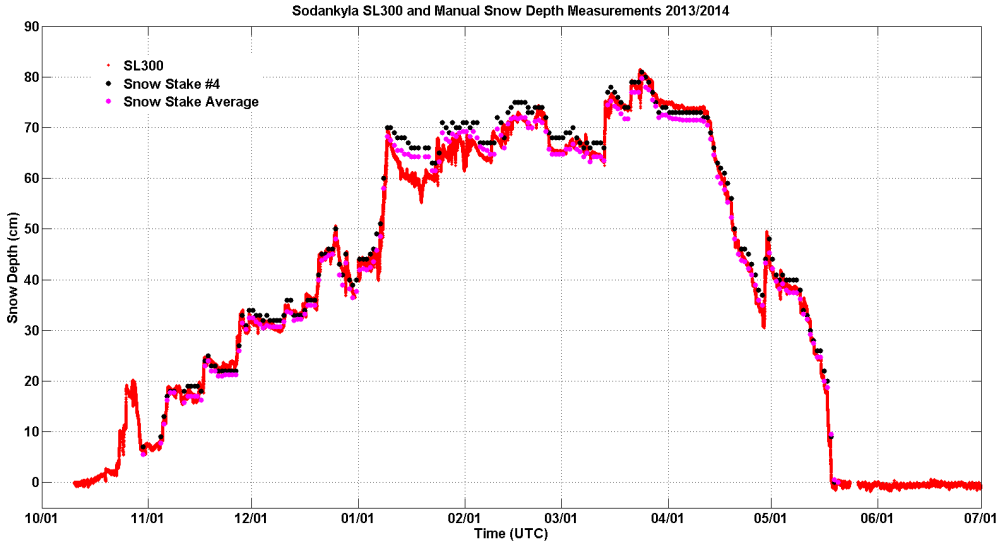


Figure 7: Time series of the SL300 and manual snow depths at Sodankylä for 2013/2014 including both the four snow stake average (magenta) and the snow stake closest to the sensor (black).

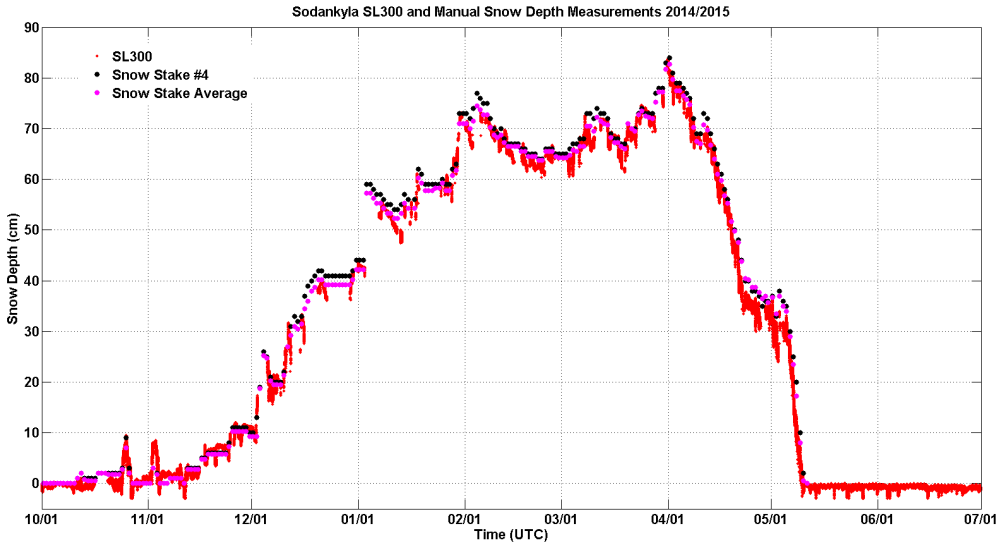


Figure 8: Time series of the SL300 and manual snow depths at Sodankylä for 2014/2015 including both the four snow stake average (magenta) and the snow stake closest to the sensor (black).

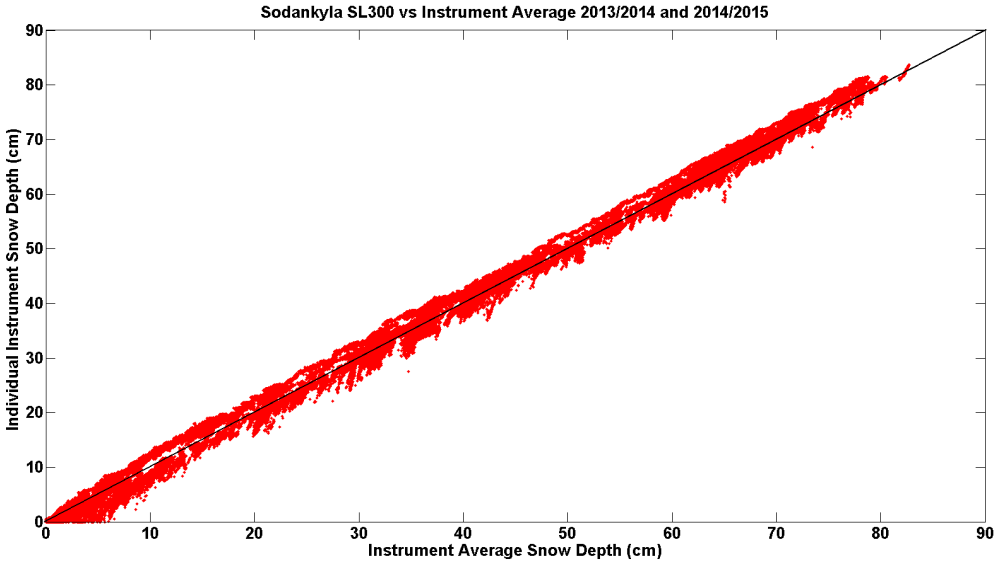


Figure 9: The SL300 compared with the average of six sensors at Sodankylä. The intercomparison statistics for the five sensor average are shown in the tables in Section 5.1c.

6.1b CARE

At CARE, the manual reference consisted of an average of daily visual observations of four snow stakes that were mounted at each of the corners of the target (shown in Figure 3.2). The intercomparison of the sensors on the three pedestals (12A, 20, and 11A) with the manual average for each of the targets is shown in Figure 10. The corresponding time series are shown in Figure 11 (2013/2014) and Figure 12 (2014/2015) with pedestal 12A shown in red, 11A shown in blue, and 20 shown in black. The automated reference consisted of the average of four sensors (Campbell Scientific SR50A, Jenoptik/Lufft SHM30, Sommer USH-8 and Felix SL300) at each of the three pedestals. This intercomparison is shown in Figure 13 for each of the pedestals.

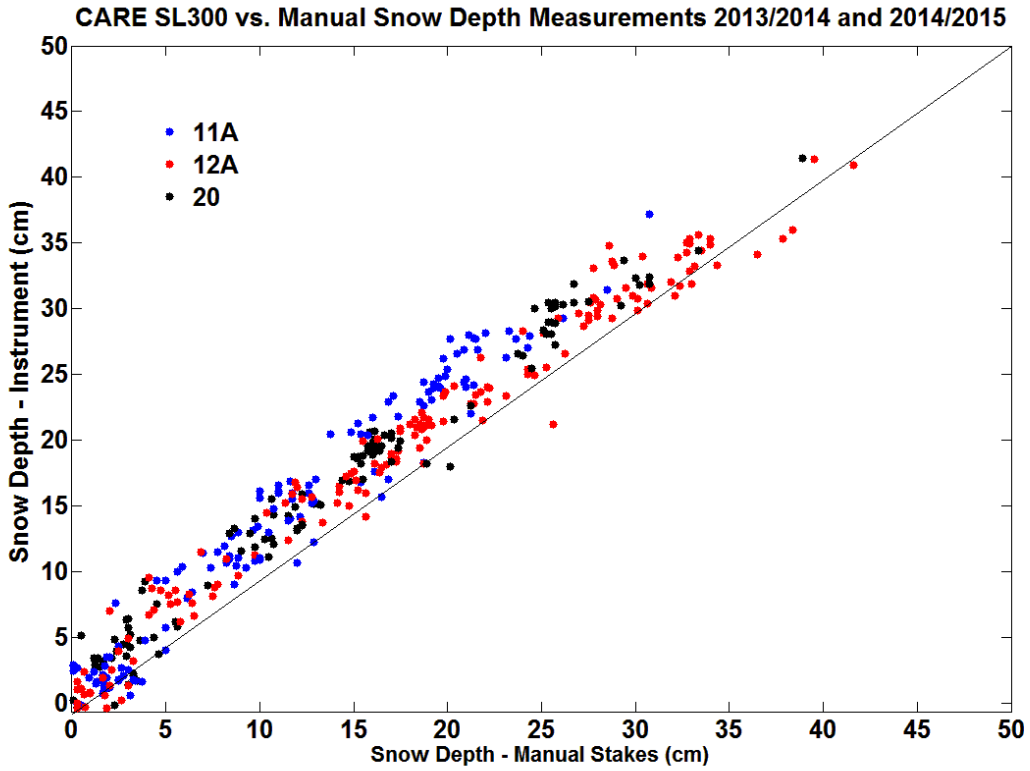


Figure 10: SL300 at CARE compared with the manual refererence which is the average of four snow stakes at the corner of the targets 12A (red), 11A (blue) and 20 (black). The intercomparison statistics are shown in the tables in Section 5.1c.

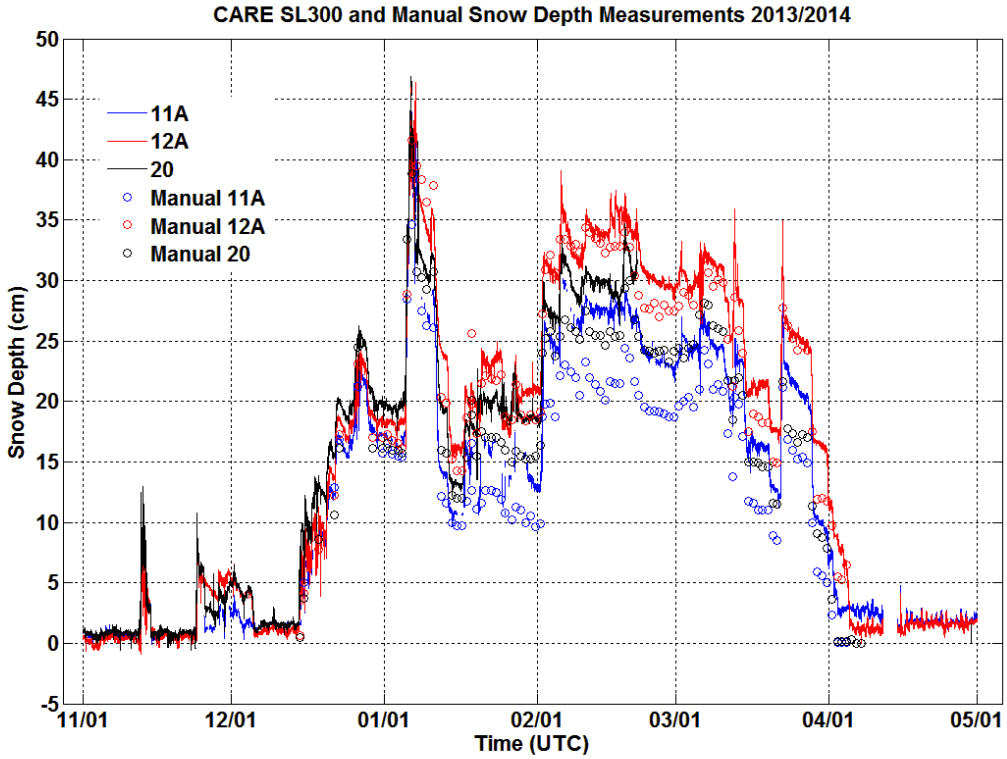


Figure 11: The time series for the SL300s on pedestals 12A (red), 11A (blue) and 20 (black) and their corresponding average manual snow depths (circles of the same colour) at CARE for 2013/2014.

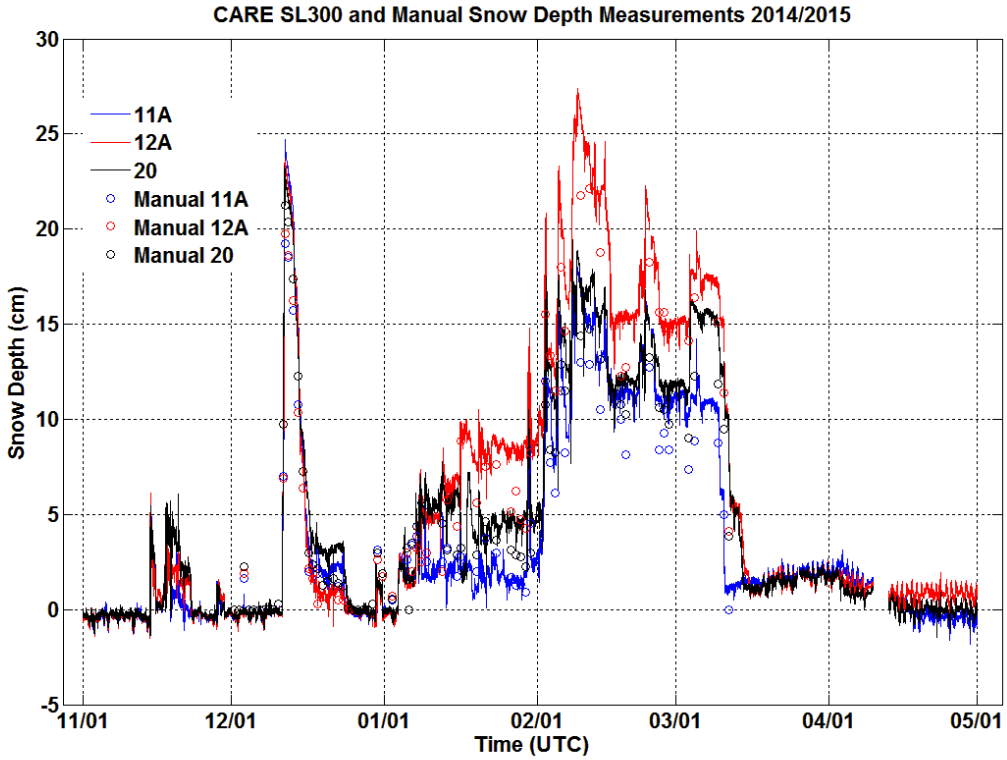


Figure 12: The time series for the SL300s on pedestals 12A (red), 11A (blue) and 20 (black) and their corresponding average manual snow depths (circles of the same colour) at CARE for 2014/2015.

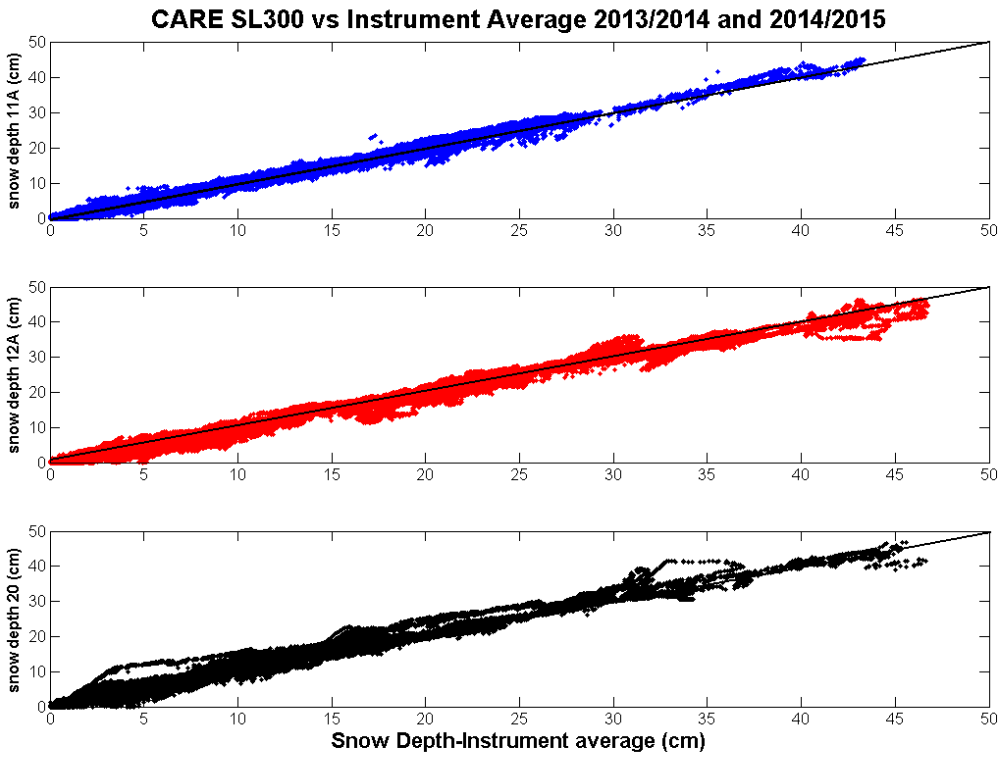


Figure 13: The three SL300s compared with the average of four sensors at pedestals 12A (top), 20 (middle) and 11A (bottom) at CARE. The intercomparison statistics for each of the pedestal intercomparisons are shown in the tables in Section 5.1c.

5.1c Intercomparison Statistics Summary

Table 4: Summary statistics for the SL300 at Sodankylä

Sodankylä	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.99	1.00	0.56	1.9	193
	2014-2015	0.99	0.99	-0.14	1.8	140
Automated Reference	2013-2014	1.00	1.02	0.31	1.6	371672
	2014-2015	1.00	1.01	-0.72	1.1	274383

Table 5: Summary statistics for the SL300 on pedestal 12A at CARE

CARE 12A	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.97	0.95	2.82	2.3	107
	2014-2015	0.96	1.14	0.29	2.2	70
Automated Reference	2013-2014	0.99	0.98	0.17	0.9	255489
	2014-2015	0.98	0.98	-0.35	1.2	254090

Table 6: Summary statistics for the SL300 on pedestal 11A at CARE

CARE 11A	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.94	1.07	2.85	4.4	77
	2014-2015	0.96	1.13	-.20	2.0	70
Automated Reference	2013-2014	0.99	1.02	0.02	0.8	181013
	2014-2015	0.99	1.04	-0.48	0.6	254253

Table 7: Summary statistics for the SL300 on pedestal 20 at CARE

CARE 20	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.96	0.96	3.80	3.4	60
	2014-2015	0.95	1.14	0.52	2.0	70
Automated Reference	2013-2014	0.99	1.03	0.58	1.5	160447
	2014-2015	0.94	1.16	0.16	1.8	254612

5.2) Factors that influence instrument performance

At Sodankylä, the snow depth instrumentation was distributed in the NE quadrant of the Intercomparison Field with the average distance between instruments of approximately 16m. The average distance between the SL300 at pedestal 60:32 and the 4 snow stakes was 27.5 m with the closest stake (Stake 4) being 7 m from the instrument. Even though the snow depth in the Intercomparison Field was relatively uniform, there was still variability in depth across 7 m and certainly across 27 m. This is outlined in another section of the SPICE final report. This spatial variability likely accounted for much of the scatter shown in reference intercomparison for Sodankylä and cannot be attributed to sensor bias.

It was reported by the Sodankylä site manager that the instrument intermittently underperformed during certain weather conditions. This report needs to be substantiated and documented during future analysis.

The design and installation of the manual snow stakes at the corner of each target at the CARE site was implemented to measure the snow depth as close as possible to the sensor without being within the sensor's FOV. Unfortunately, the manual snow stakes created a mounding effect, increasing the depth of the snow in the middle of the target as compared to the corners of the target. This is quite evident in Figures 10, 11 and 12 which suggests that the sensor overestimated snow depth as compared to the manual measurement. Another factor to consider when interpreting the intercomparison at CARE is the potential for a change in the reference distance between the target and the sensor over the course of the intercomparison season. This is called "zero snow depth drift" and is discussed in the main body of the SPICE final report. The levels of the targets at CARE relative to the sensor were susceptible to change due to settling after installation (under the weight of the snow pack) or from frost heave. The change in this position can only be assessed at the end of the season when the targets are bare but it is recognized that changes in the zero snow depth distance can be continuous and undetectable during the season and therefore the impact on the snow depth intercomparison was difficult to assess. However, it is estimated that the impact would not exceed ± 2 cm during the season. This error is generally not a concern at the other sites where either natural targets are used or the artificial targets are not prone to heaving or settling.

5.3) Performance Considerations

The Sodankylä site manager noted that the instrument tended to be sensitive to temperature changes and measurements were often impacted by adverse weather conditions, especially after December 2014. This is reflected in the high percentage of suspicious data identified during data quality control. The instrument did not fail completely and provided intermittently good data points as shown in Figure 6-9. The manufacturer later noted that this was a technical issue with an electronics board. CARE did not have the same issue so the performance limitation in 2014/2015 at Sodankylä can be attributed to either the partial failure of that sensor or on adverse climatic conditions experienced at Sodankylä and not at CARE.

5.4) Maintenance

The original instrument sent to Sodankylä in 2013 was damaged during shipping and needed to be replaced.

6) Lessons Learned

One of the SL300 tested at Sodankylä experienced a malfunction during the 2014/2015 SPICE measurement period that created blocks of missing or suspicious data. According to the site manager quality assurance reports, the instrument measurements and the reporting of suspicious values appeared to be sensitive to temperature changes and measurements during certain weather conditions appear to have accuracy issues. This appears to only be an issue at Sodankylä and no issues were reported at CARE. It was also noted by the site managers at Sodankylä that the top of the instrument tends to accumulate snow during snowfall events with low wind speeds. This was investigated further and although the snow eventually drops (or is manually cleared) off the instrument and falls onto the target area, there is no evidence to support the hypothesis that this impacted the snow depth measurements.

SPICE Instrument Performance Report Lufft/Jenoptik - SHM30

1) Technical Specifications

Physical principle: Laser ranger (650 nm) measuring distance from phase information
Measurement Area: < 11 mm at 10 m distance
Measurement Range: 0 to 10 m
Measurement Accuracy: < ± 5 mm (95% statistical spread)
Measurement reproducibility: ≤ 0.5 mm
Measurement Resolution: 0.1 mm
Operating Range: -40 °C to +50 °C
Link to manual:
http://www.lufft.com/dateianzeige.php?Dateiname=/download/manual/SHM30_V1_e.pdf

Figure 1: Photo of the SHM30 snow depth sensor

2) SPICE Test Configuration

Test Sites: CARE (Canada), Col de Porte (France), Sodankylä (Finland), Weissfluhjoch (Switzerland)

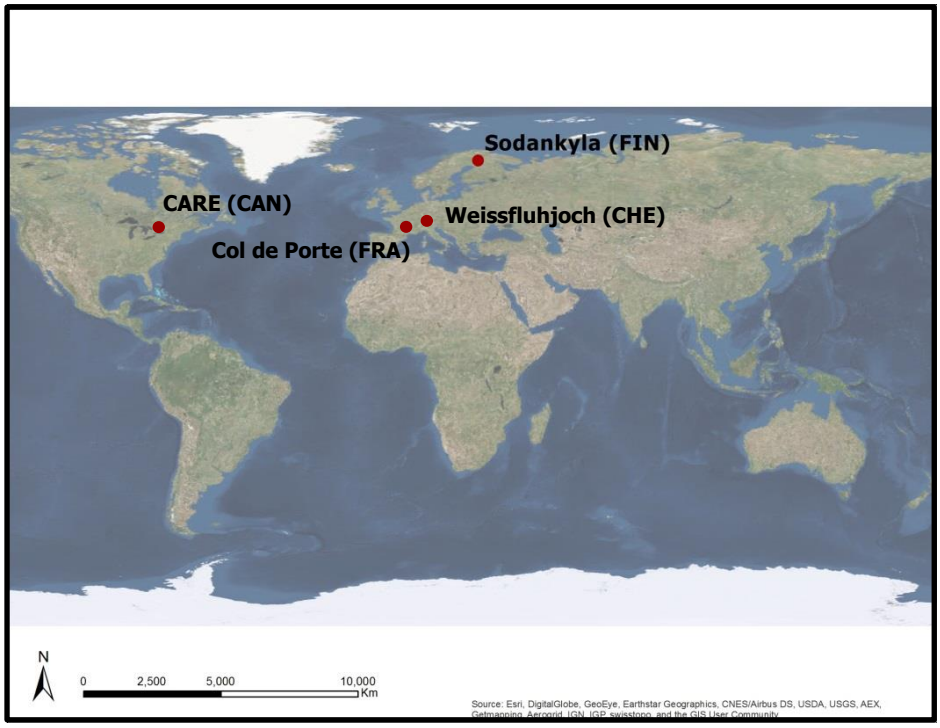


Figure 2: Locations of the sensors under test

2.1) Site Specifications

Table 1: Site specific instrument and installation details

	Sodankylä	Col de Porte	CARE	Weissfluhjoch
Date of Installation	Oct. 20, 2013	Jan. 15, 2014	Nov. 1, 2013	Oct. 1, 2013
Serial Number(s)	121421	120739	121883,130874,130875	91103
Installation Location	70:42	SPICE Beam	12A, 11A,20	18
Firmware	9.07	9.06	9.08	9.04
Distance to Ground	2m	4m	1.70,1.75,1.72m	5m
Target	Artificial Grass Mat 2m x 2.5m	Natural (mown) grass	Textured Plastic, 1.2mx1.2m	Natural rock and grass
Angle of Instrument	18.2°/28°(after 29/8/2014)	20°	30°, 30°, 30°	12.5°

2.2) Site Photos

Figure 3: Sodankylä

Figure 4: Col de Porte

Figure 5: CARE

Figure 6: Weissfluhjoch

2.3) Instrument Footprint Diagram

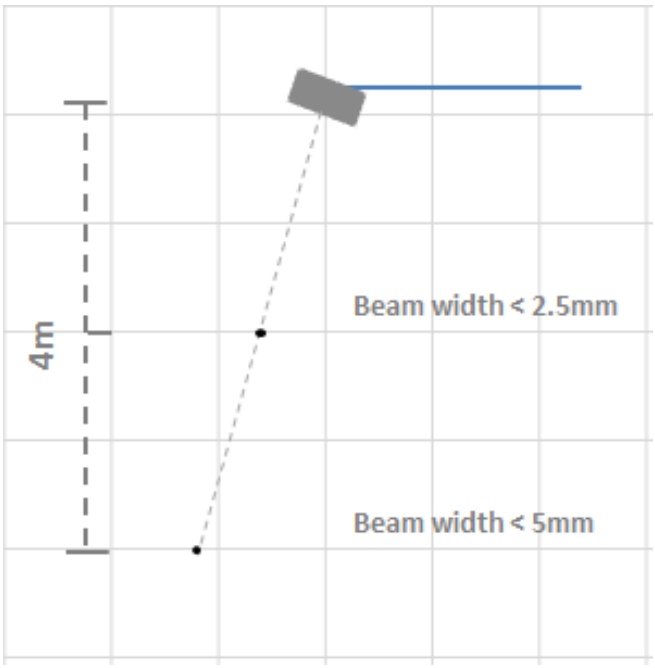


Figure 7: Conceptual diagram of instrument field of view

2.4) Environment Conditions During SPICE

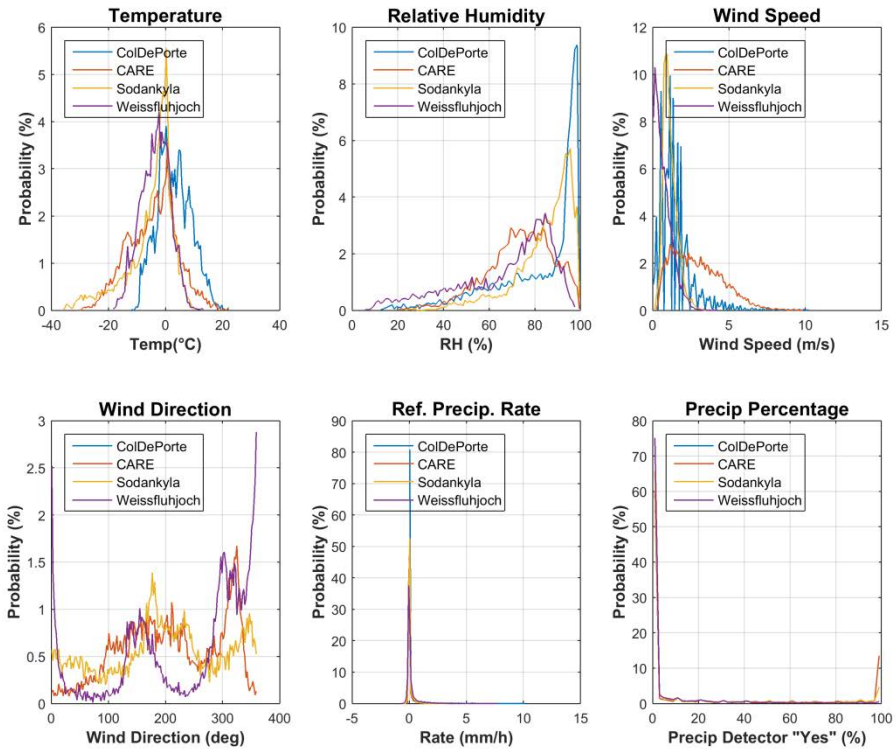


Figure 8: Summary of aggregated environmental conditions on the SPICE sites that operated a SHM30 for the entire duration of formal tests.

3) Data Output

Table 2: Parameters measured or output by the instrument

Measured Parameters	Units	Sodankylä	Col de Porte	CARE	Weissfluhjoch
Snow Depth	m	X	x	x	x
Signal Strength	-	X		x	
Temperature	Deg C	X			x
Error Code	-	X		x	x

Table 3: Measurement and data acquisition parameters

Measurement and Processing	Units	Sodankylä	Col de Porte	CARE	Weissfluhjoch
Data Sampling Rate	[Minutes]	1	1	0.5	5
Data Acquisition Interval	[Minutes]	1	1	0.5	5
Date Processing		Sample	Sample	Averaged	Sample
Communication Protocol		RS-232	RS-232	RS-232	RS-232
Heating		Yes	No	Yes	No

4) Quality Control Information

4.1) Data Quality and Availability

Missing data occurred when data were not recorded by the data logger. This may or may not be related to the function of the instrument. Suspicious data were identified as an outlier and are either removed automatically via pre-set range and jump filtering, or identified visually as an outlier and removed manually. Range and jump filters were set based on reasonable physical limits for the individual sites. Suspicious data may be a result of a malfunctioning instrument or from meteorological impacts on the instrument's measurement capability. Table 4 provides a seasonal breakdown of the data quality metrics for each site. More commentary on individual instruments, where required, is included below. Only data classified as "Good" were used for the intercomparisons.

Table 4: Seasonal breakdown of data QC metrics

SEASON 2013-2014	Sodankylä	Col de Porte	CARE	Weissfluhjoch
Collection Period	Oct 2013-June 2014*	Jan 2014-April 2014	Nov 2013-April 2014**	Oct 2013-July 2014
Good	89.0%	98.0%	71.7%	99.3%
Missing	1.8%	2.0%	1.6%	0.2%
Suspicious	9.2%	0.0%	26.7%	0.5%

* sensor went into error mode at Sodankylä following snow melt and remained in this state due to reduced monitoring at the site. It returned to a functional state after onsite re-set.

**a data logger issue, not a sensor issue, at CARE resulted in the reporting of out-of-range measurements

SEASON 2014-2015	Sodankylä	Col de Porte	CARE	Weissfluhjoch
Collection Period	Oct 2014-June 2015	Nov 2014-April 2015	Nov 2014-April 2015	Oct 2014-July 2015
Good	98.9%	98.9%	95.0%	95.9%
Missing	1.1%	0%	2.3%	2.9%
Suspicious	0%	1.1%	2.6%	1.2%

Quality metric statistics for CARE and Sodankylä were skewed by a relatively large proportion of data that are flagged as Suspicious by the automated QC procedures for the 2013/2014 season.

For Sodankylä, a large portion of this erroneous data occurred after 2014-06-01 and after the end of snowmelt which makes the missing data inconsequential for analysis. Investigation has shown that the suspicious data were a result of the sensor going into an error state as a result of an electrical storm at the site. During this period, the sensor was on and functioning but reporting out-of-range data while in an error state. The data reported by the sensor was interpreted as erroneous by the QC procedures and was flagged as Suspicious. When the site manager identified the problem and reset the sensor, it then continued to function normally. Also, a large portion of the missing data at Sodankylä in the 2013/2014 season occurred in late May and can be attributed to a site power outage also caused by lightning.

For CARE, the suspicious data occurred at the beginning of the 2013/2014 season from 2013-11-01 through 2013-12-18 and can also be attributed to a data logging issue resulting in an "out of range" value and flagged as Suspicious by the QC procedures. This suspicious data were not a result of a sensor malfunction.

5) Evaluation of the Ability to Perform Over a Range of Operating Conditions

There are several options for comparing sensors under test to a reference. Manual snow depth measurements were performed in the Intercomparison Field at all three of these

SPICE sites and were used as a measurement reference. Also, many of these sites hosted several automated snow depth sensors that when combined as an average, served as a reference. The automated references are detailed further in the following sections. Both intercomparisons are shown below in Figures 6.1 through 6.15. The summary statistics for these intercomparisons are shown in Section 6.1e.

5.1) Performance against the reference

5.1a Sodankylä

At Sodankylä, the manual reference consisted of a daily photograph observation of four graduated snow stakes distributed in the Intercomparison Field. The intercomparison below (Figure 8) shows the instrument plotted against the average of all stakes (red) and against Stake 4 which was closest to the instrument (blue). The seasonal time series for this intercomparison are shown in Figures 9 (2013/2014) and Figure 10 (2014/2015). The automated reference was an average of 1-minute snow depth data obtained from six instruments, distributed in the Intercomparison Field. These instruments were two Campbell Scientific SR50ATH sensors, two Sommer USH-8 sensors, a Felix SL300 sensor and the Jenoptik/Lufft SHM30 sensor. The SHM30 is plotted against this reference in Figure 11.

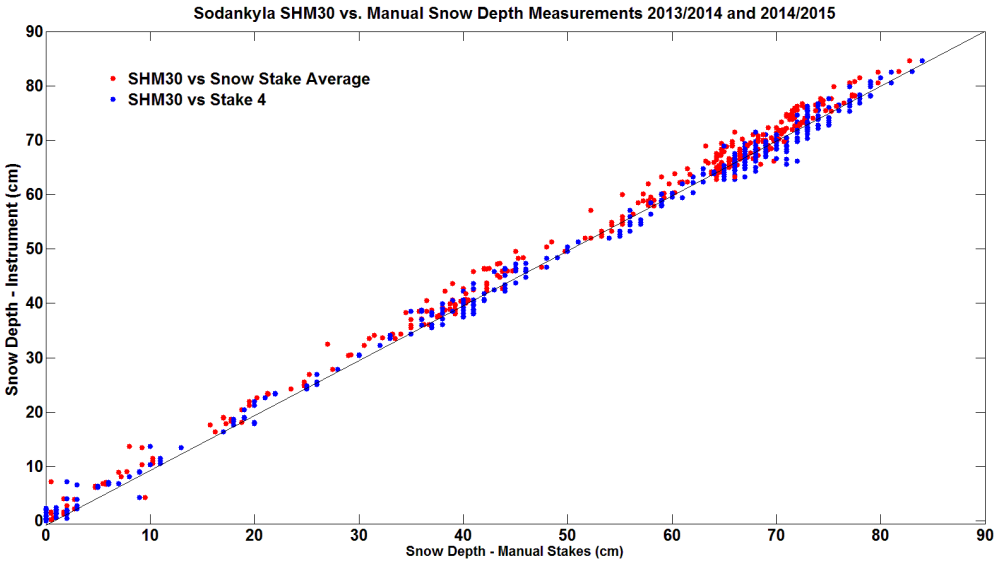


Figure 8: SHM30 at Sodankylä compared with the manual refererence which is either the average of the four snow stakes (red) or the closest snow stake to the sensor (blue). The intercomparison statistics for the four stake average are shown in the tables in Section 5.1e.

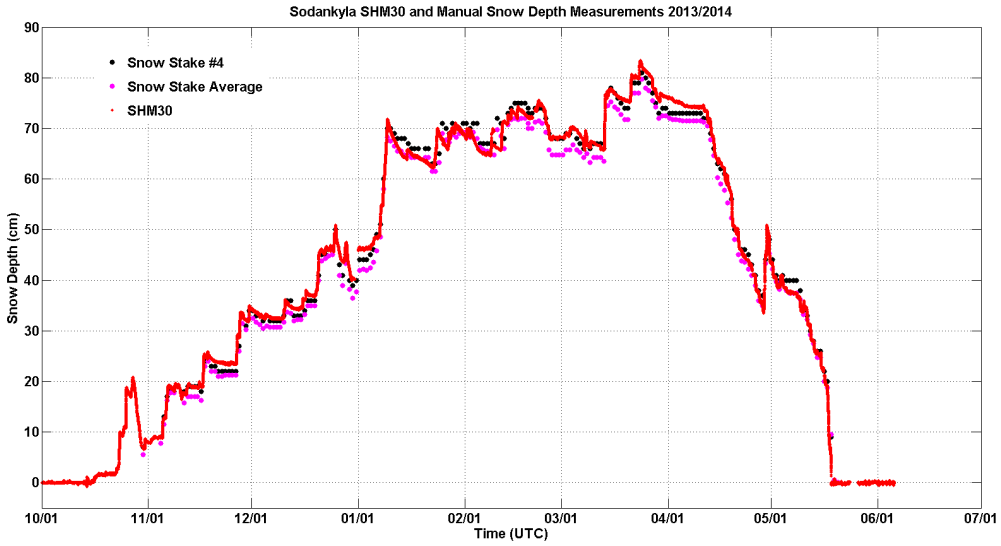


Figure 9: Time series of the SHM30 and manual snow depths at Sodankylä for 2013/2014 including both the four snow stake average (magenta) and the snow stake closest to the sensor (black).

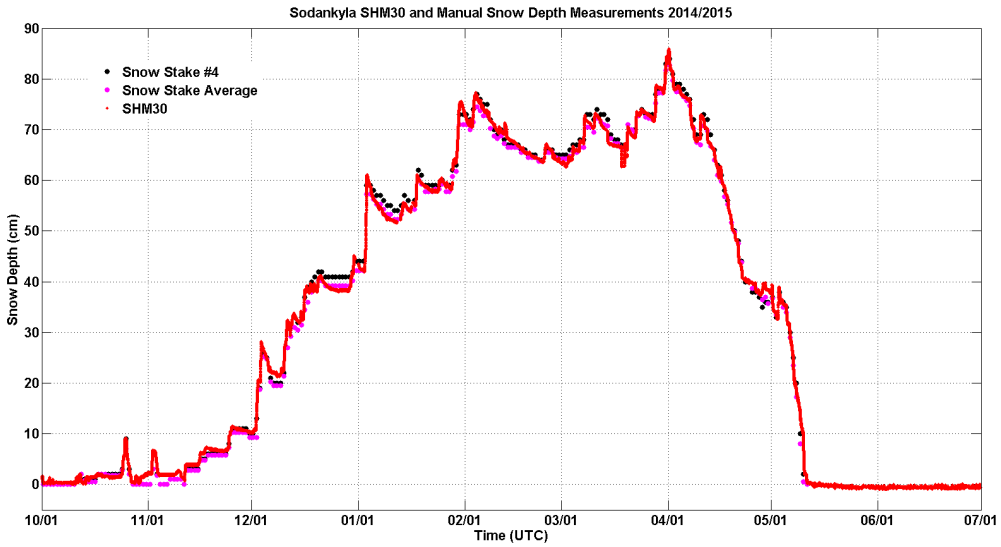


Figure 10: Time series of the SHM30 and manual snow depths at Sodankylä for 2014/2015 including both the four snow stake average (magenta) and the snow stake closest to the sensor (black).

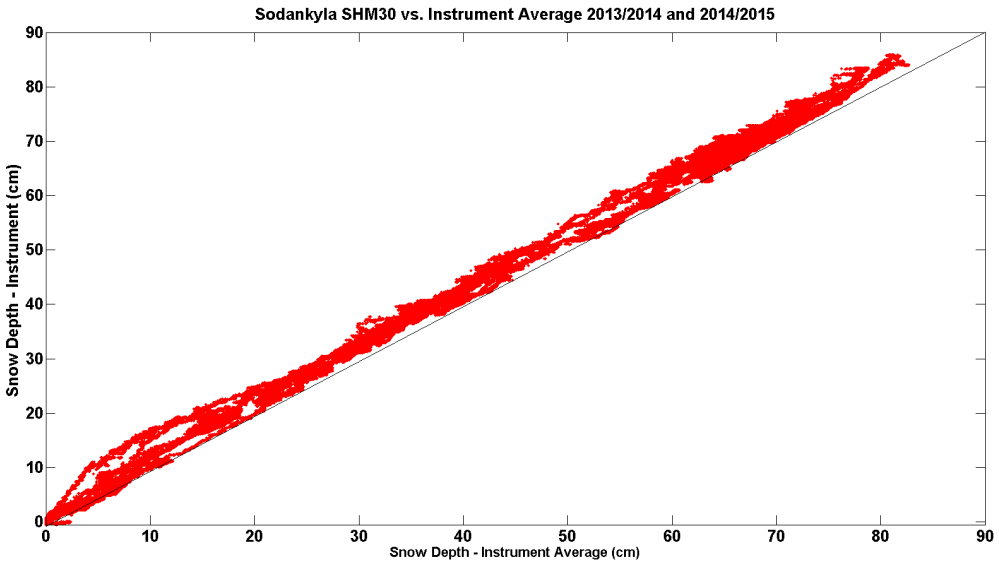


Figure 11: The SHM30 compared with the average of six sensors at Sodankylä. The intercomparison statistics for the five sensor average are shown in the tables in Section 5.1e.

5.1b Col de Porte

At Col de Porte, the manual measurement consisted of a visual measurement of three snow stakes observed on a weekly basis. Of the three stakes, two were closer to the automated sensors than the other. For the manual intercomparison, the averages of all stakes are plotted as red and the average of the two closest stakes are plotted as blue in Figure 12. The time series of this intercomparison are shown in Figures 13 (2013/2014) and Figure 14 (2014/2015). The automated reference was an average of 1-minute snow depth data obtained from five automated sensors measuring roughly the same target under the SPICE beam. These instruments were a Campbell Scientific SR50ATH, a Campbell Scientific SR50A, a Jenoptik/Lufft SHM30, a Dimetix FLS-CH 10 and an Apical Technologies sensor. This is plotted in Figure 15.

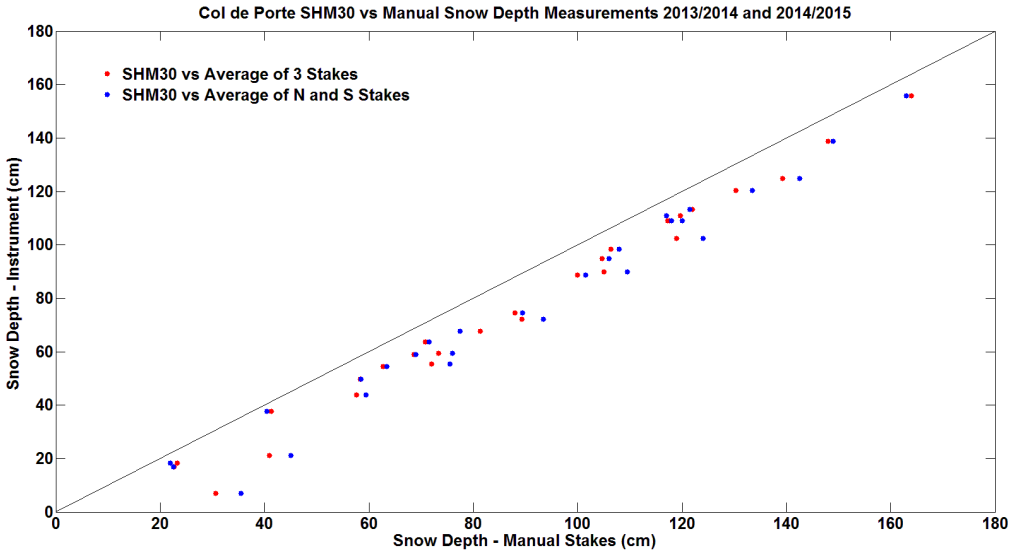


Figure 12: SHM30 at Col de Porte compared with the manual refererence which is either the average of three snow stakes (red) or the average of the two closest snow stakes to the sensor (blue). The intercomparison statistics for the two closest stake average are shown in the tables in Section 5.1e.

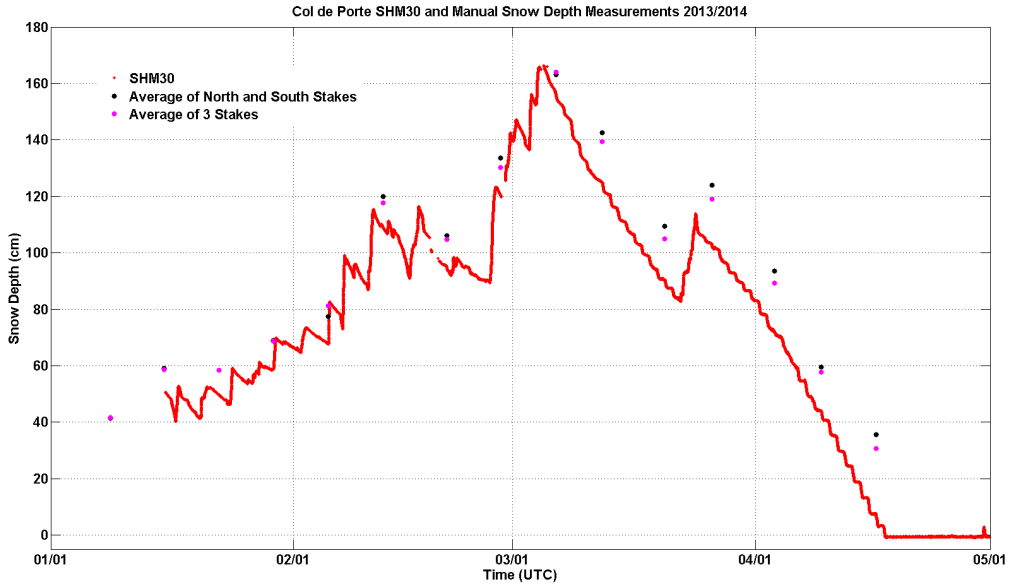


Figure 13: Time series of the SHM30 and manual snow depths at Col de Porte for 2013/2014 including both the three snow stake average (magenta) and the average of the two snow stake closest to the sensor (black).

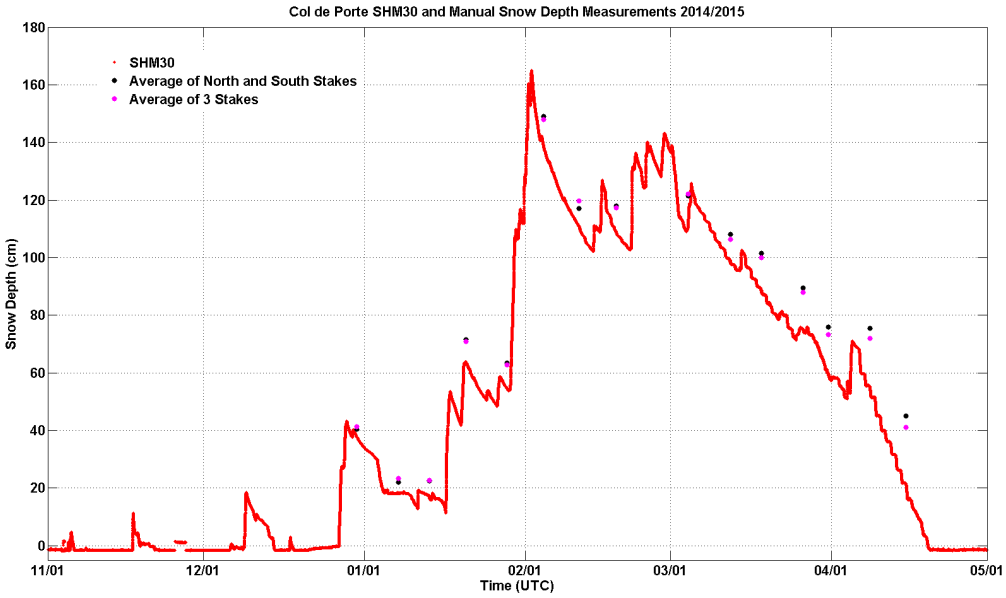


Figure 14: Time series of the SHM30 and manual snow depths at Col de Porte for 2014/2015 including both the three snow stake average (magenta) and the average of the two snow stake closest to the sensor (black).

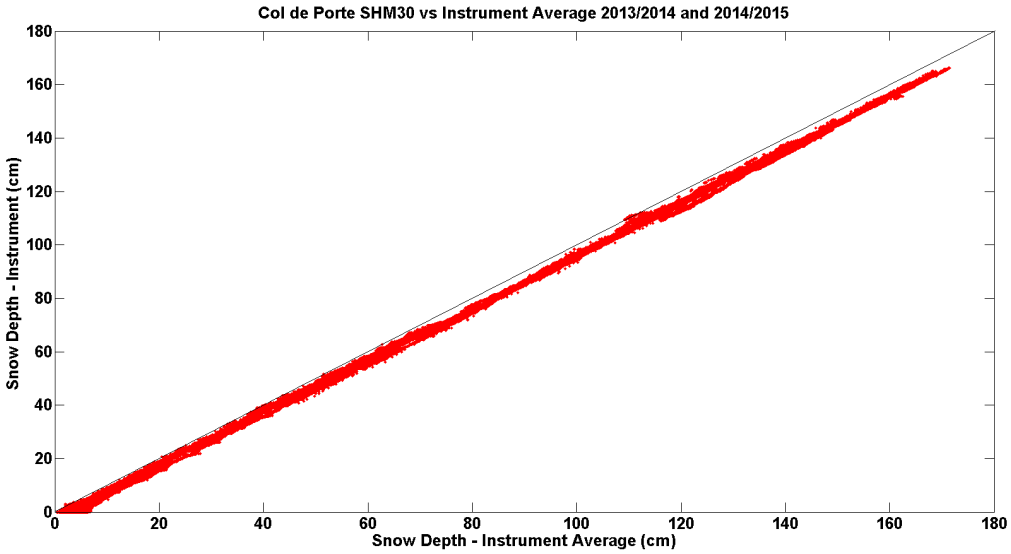


Figure 15: The SHM30 compared with the average of five sensors at Col de Port. The intercomparison statistics for the five sensor average are shown in the tables in Section 5.1e.

5.1c CARE

At CARE, the manual reference consisted of an average of daily visual observations of four snow stakes that were mounted at each of the corners of the target (shown in Figure 5). An intercomparison of the sensors on the three pedestals (12A, 20, and 11A) along with the manual average for each of the targets is shown in Figure 16. The corresponding seasonal time series are shown in Figures 17 (2013/2014) and Figure 18 (2014/2015) with pedestal 12A shown in red, 11A shown in blue, and 20 shown in black. The automated reference consisted of the average of four sensors (Campbell Scientific SR50A, Jenoptik/Lufft SHM30, Felix SL300 and Sommer USH-8) at each of the three pedestals. This intercomparison is shown in Figure 19 for each of the pedestals.

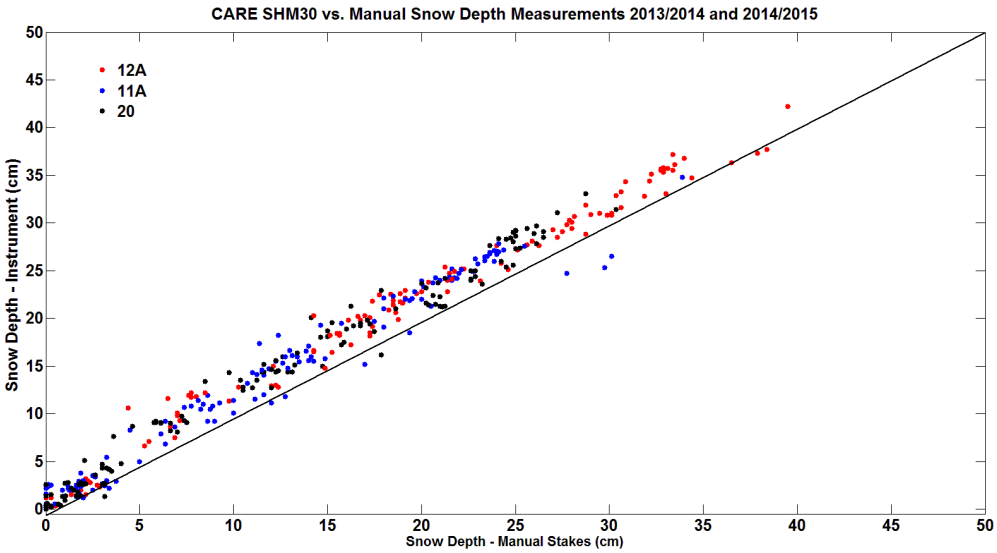


Figure 16: SHM30s at CARE on pedestals 12A (red), 11A (blue), and 20 (black) compared with the manual refererence which is the average of four snow stakes at the corner of the target under the sensor for each of the three pedestals. The intercomparison statistics for the 4-corner averages are shown in the tables in Section 5.1e.

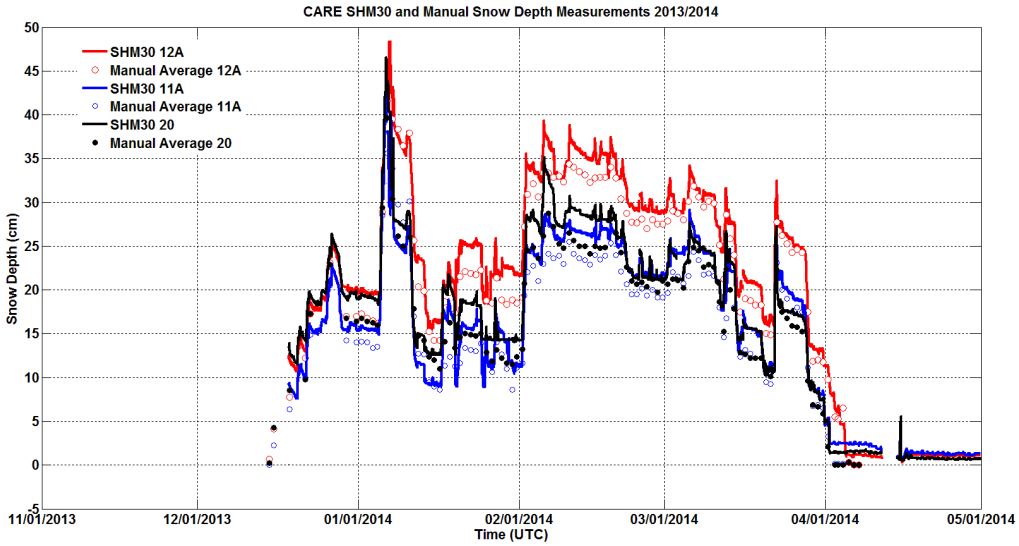


Figure 17: The time series for the SHM30s on pedestals 12A (red), 11A (blue) and 20 (black) and their corresponding average manual snow depths (open circles of the same colour) at CARE for 2013/2014.

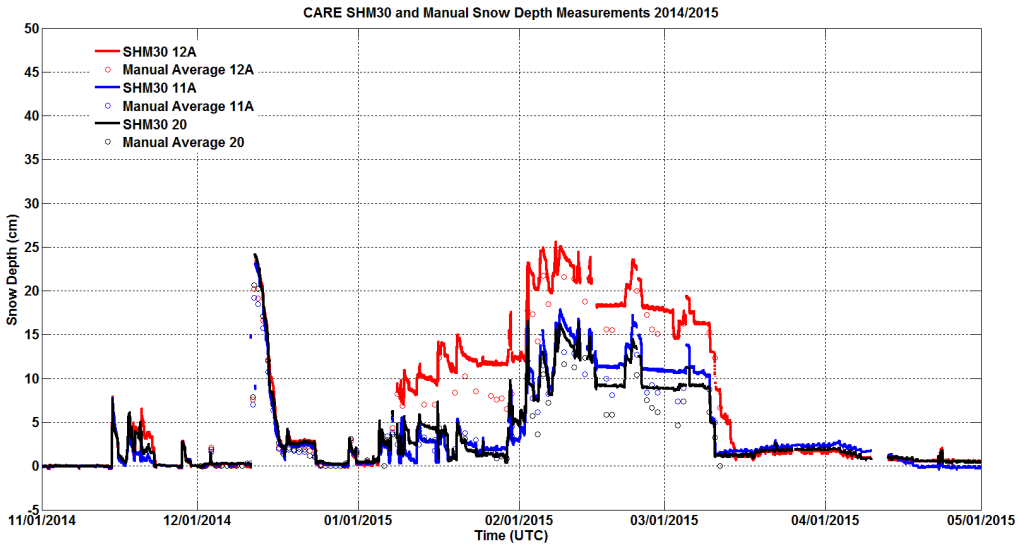


Figure 18: The time series for the SHM30s on pedestals 12A (red), 11A (blue) and 20 (black) and their corresponding average manual snow depths (open circles of the same colour) at CARE for 2014/2015.

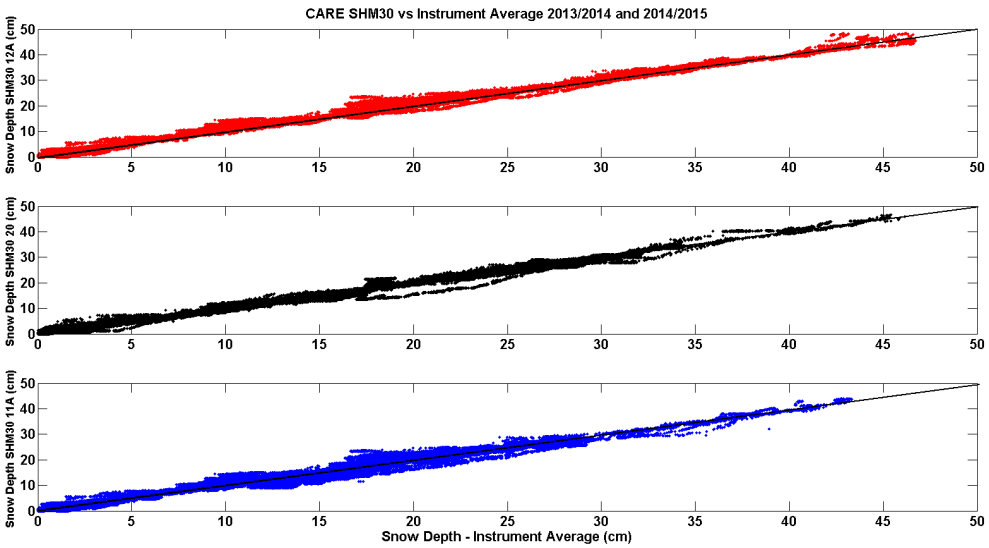


Figure 19: The three SHM30s compared with the average of four sensors at pedestals 12A (top), 20 (middle) and 11A (bottom) at CARE. The intercomparison statistics for each of the pedestal intercomparisons are shown in the tables in Section 5.1e.

6.1d Weissfluhjoch

At Weissfluhjoch, the manual measurement was a visual observation of a graduated snow stake located about 17 m from the SHM30. This measurement was made once per day at approximately 08:00 UTC. Since there were only two automated measurements at this site, no automated reference is available for intercomparison. The manual intercomparison is shown in Figure 20 and the time series for each season are shown in Figure 21 (2013/2014) and Figure 22 (2014/2015).

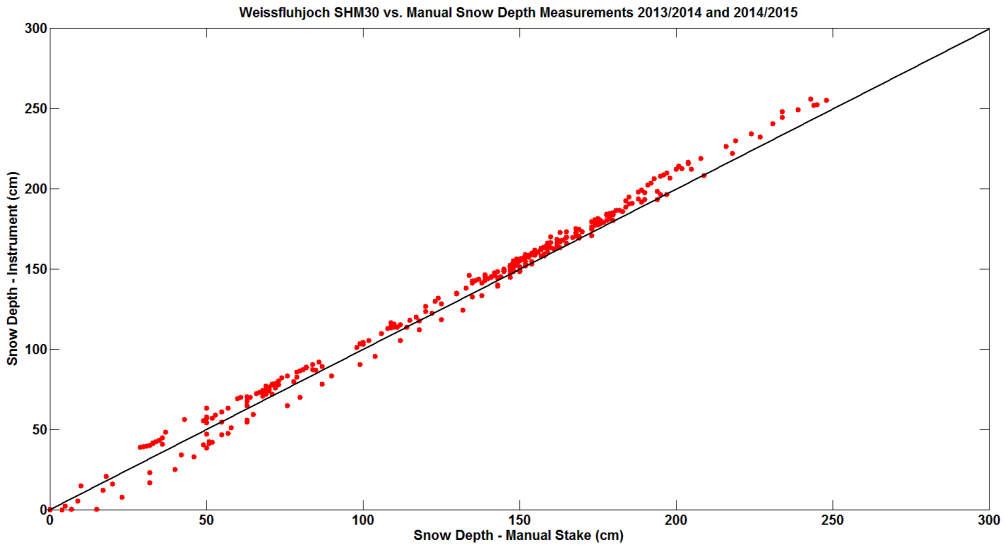


Figure 20: SHM30 at Weissfluhjoch compared with the manual refererence snow stake. The intercomparison statistics are shown in the tables in Section 5.1e.

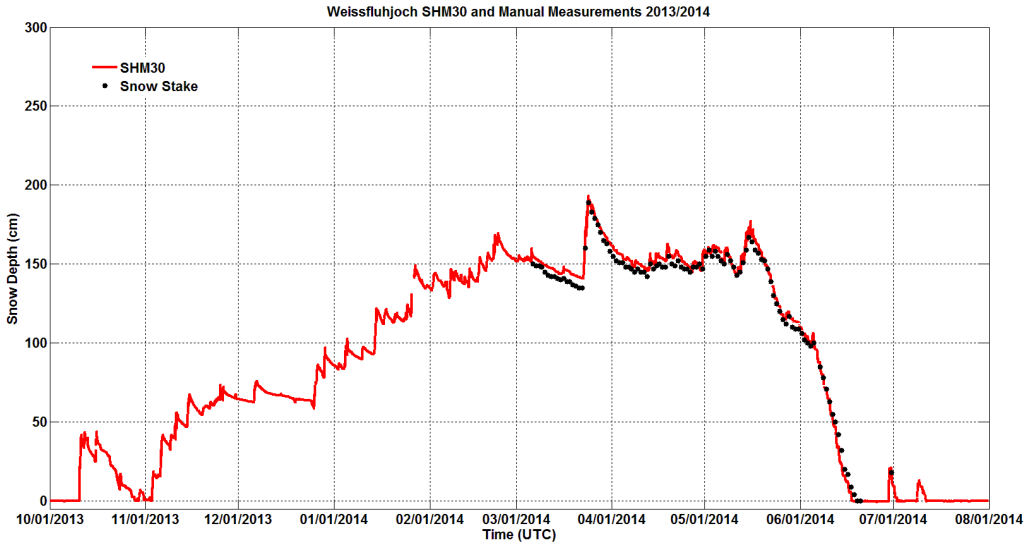


Figure 21: Time series of the SHM30 and manual snow depths at Weissfluhjoch for 2013/2014.

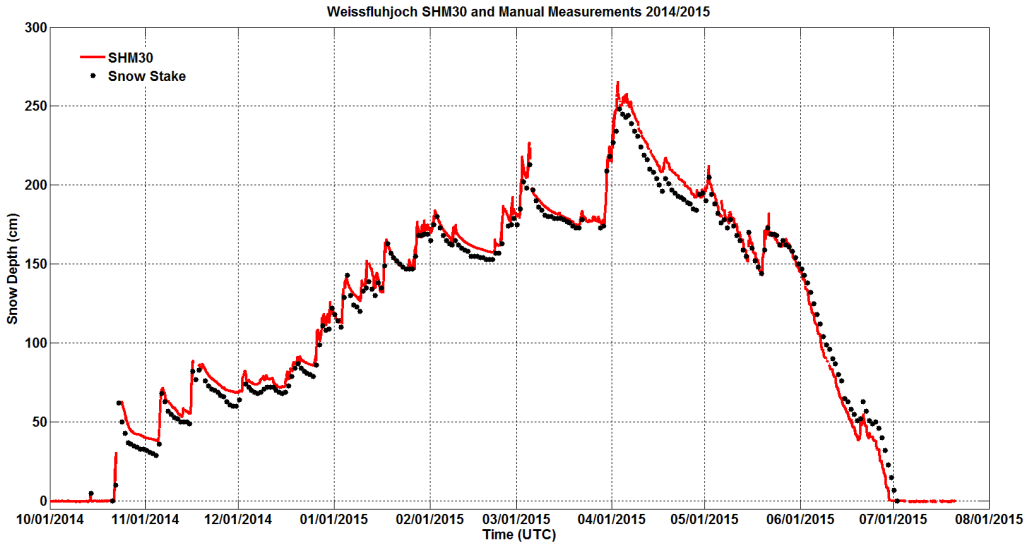


Figure 22: Time series of the SHM30 and manual snow depths at Weissfluhjoch for 2014/2015.

5.1e Intercomparison Statistics Summary

Table 5: Summary statistics for SHM30 at Sodankylä

Sodankylä	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.99	1.02	0.78	2.7	166
	2014-2015	1.00	0.99	1.12	1.5	219
Automated Reference	2013-2014	1.00	1.04	0.91	2.9	349757
	2014-2015	1.00	1.02	0.23	1.5	388749

Table 6.: Summary statistics for SHM30 at Col de Porte

Col de Porte	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.98	1.05	-19.48	16.2	13
	2014-2015	0.98	0.99	-9.79	12.2	15
Automated Reference	2013-2014	1.00	0.98	-2.40	4.3	148588
	2014-2015	1.00	0.99	-2.97	3.8	260590

Table 7: Summary statistics for SHM30 on pedestal 12A at CARE

CARE 12A	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.98	0.98	2.59	2.4	78
	2014-2015	0.97	1.14	0.63	2.4	63
Automated Reference	2013-2014	0.99	1.01	-0.19	0.7	186243
	2014-2015	0.99	1.07	0.15	1.1	245562

Table 8: Summary statistics for SHM30 on pedestal 20 at CARE

CARE 20	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.96	1.02	2.15	2.8	78
	2014-2015	0.96	1.19	0.43	1.7	64
Automated Reference	2013-2014	0.99	0.99	-0.21	0.9	187043
	2014-2015	0.98	0.95	0.08	0.6	248380

Table 9: Summary statistics for SHM30 on pedestal 11A at CARE

CARE 11A	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.94	0.97	2.81	2.8	78
	2014-2015	0.97	1.19	0.24	1.7	64
Automated Reference	2013-2014	0.99	1.00	-0.68	1.2	187147
	2014-2015	0.99	1.00	-0.08	0.6	248277

Table 10: Summary statistics for SHM30 at Weissfluhjoch

Weissfluhjoch	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Graduated Stake	2013-2014	1.00	1.05	-2.83	4.6	107
	2014-2015	0.99	1.03	-0.51	6.9	237

5.2) Factors that influence instrument performance

At Sodankylä, the snow depth instrumentation was distributed in the NE quadrant of the Intercomparison Field with the average distance between instruments at approximately 16 m. The average distance between the SHM30 and the four snow stakes was almost 27 m with the closest stake being 7 m from the instrument. Even though the snow depth in the Intercomparison Field was relatively uniform, variability in depth was still expected across 7 m and certainly across 27 m and is documented in the main body of the SPICE final report. This probably accounted for much of the scatter shown in the reference intercomparison for Sodankylä. It was also noted at this site that differential melt occurred amongst the targets under the sensors and the snow stakes (depending on their individual exposures). The effects of this differential melt can be seen in both the manual and automated reference plots for Sodankylä at lower snow depth amounts. This scatter most likely cannot be attributed to sensor bias.

The variability in snow depth at the Col de Porte intercomparison site was much greater than at Sodankylä. This is quite evident in the manual reference plot (Figure 6.5). However, all of the instruments used in the automated reference are located on the same mounting structure and target the same area beneath the instruments resulting in much less scatter. There appeared to be some variability across the target area resulting in the SHM30 measuring slightly lower snow depths than the instrument average. Further analysis of the variability of snow depth below the mounting structure was assessed using an automated rugged laser scanner (Picard et al., The Cryosphere 2016) which demonstrated that the mounting structure itself has no impact on the snow conditions below, and that most of the variations between sensors was due to natural variability of snow depth rather than instrumental uncertainty.

The design and installation of the manual snow stakes at the corner of each target at the CARE site was implemented to measure the snow depth as close as possible to the sensor without being within the sensor's FOV. Unfortunately, the manual snow stakes created a mounding effect, increasing the depth of the snow in the middle of the target as compared to the corners of the target. This is quite evident in Figures 19, 20 and 21 which suggests that the sensor is overestimating snow depth as compared to the manual measurement. Another factor to consider when interpreting the intercomparison at CARE is the potential for a change in the reference distance between the target and the sensor over the course of the intercomparison season. This is called "zero snow depth drift" and is discussed in the main body of the SPICE final report. The levels of the targets at CARE relative to the sensor were susceptible to change due to settling after installation (under the weight of the snow pack) or from frost heave. The change in this position can only be assessed at the end of the season when the targets are bare but it is recognized that changes in the zero snow depth distance can be continuous and undetectable during the season and therefore the impact on the snow depth intercomparison is difficult to assess. However, it was estimated that the

impact would not exceed ± 2 cm during the season. This error is generally not a concern at the other sites where either natural targets are used or the artificial targets are not prone to heaving or settling.

5.3) Performance Considerations

Some analysis was performed using SPICE data to look at the SHM30's capabilities to determine the presence of snow in the target using the single strength output area even if the snow depth is very small. This is documented in the main body of the final report. It is shown that the response of the signal strength output to the presence of snow cover on the target area was strongly dependent on the optical properties of the target and that the response was much more distinct on artificial and natural turf than it was on a grey plastic target. It should also be noted here that although the firmware version (from 9.05 and higher) does not impact the snow depth measurement, it does impact the signal strength output.

It was also noted that the sensor heating created some noise in the signal strength output. No other performance considerations are noted for this instrument during the intercomparison.

5.4) Maintenance

A logger programming change was initiated at CARE on 2013-12-18 that remedied a data issue with the three SHM30s installed at this site.

The sensor at Sodankylä required a reset after the 2014-06-01 electrical storm before it would continue to function properly.

No other maintenance issues are noted by the site managers during the intercomparison

6) Lessons Learned

The SHM30s, as tested during the SPICE intercomparison, were found to be a reliable instrument with few operational problems or breaks in the data. The instrument appeared to be an accurate snow depth sensor producing very little noise and few incorrect measurements. Negative aspects about using the instrument include its larger size and weight. Also, the instrument casing tends to gather snow in high snow/low wind conditions which could potentially be rectified by external heating. The instrument requires a substantial amount of power to function, especially when the heater is active, and may therefore be less suited to more remote sites where AC power is not available.

SPICE Instrument Performance Report
Sommer Messtechnik –USH-8

1) Technical Specifications

Physical principle: Ultrasonic snow depth measurement
(Frequency 50 kHz; beam width 12°)
Measurement Area: 0.6 m diameter at 2m height
Measurement Range: 0 to 8 m
Measurement Accuracy: 0.1% of full scale
Measurement Resolution: 1 mm
Operating Temperature: -40°C to +60°C
Firmware: V1.80r00
Link to manual:
http://www.hydrologicalusa.com/images/uploads/User_manual_USH-8_V01_00_V01_58_ENG_www.pdf

Figure 1: Photo of the USH-8

2) SPICE Test Configuration

Test Sites: CARE (Canada), Sodankylä (Finland)

The USH-8, as provided by the manufacturer, was tested at Sodankylä in two different configurations, one with a horizontal boom (Figure 3) and the other with an angled boom (Figure 4). The sensors at CARE were provided by the site host where three sensors were installed on three instrument pedestals in close proximity to each other.

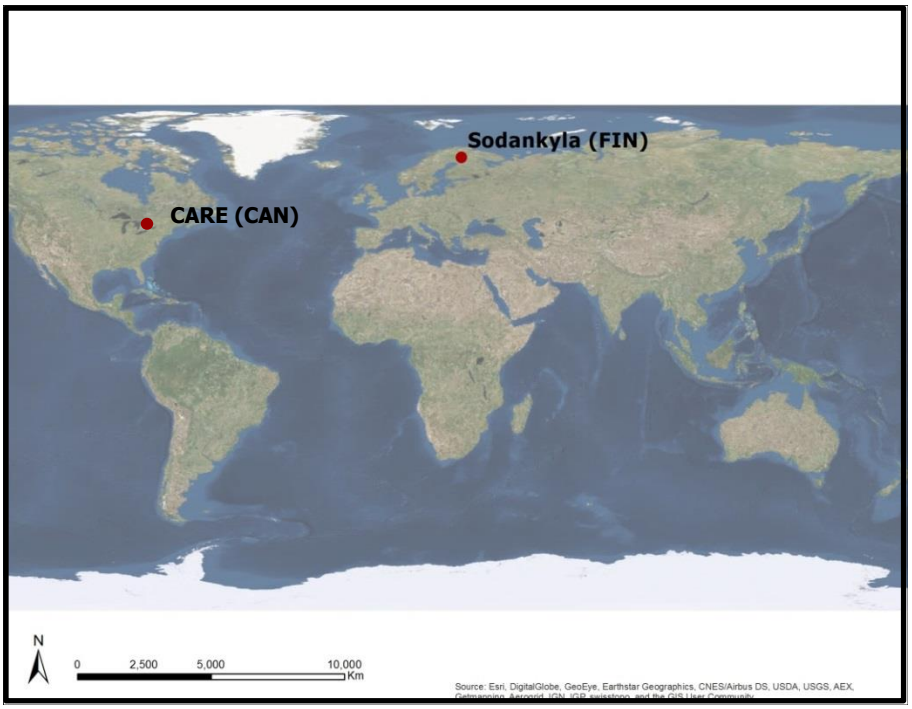


Figure 2: Location of the sensors under test

2.1) Site Specifications

Table 1: Site specific instrument and installation details

	Sodankylä	CARE
Date of Installation	2012-12-19; 2013-10-10	2013-11-01
Serial Number(s)	32120820; 32120816	32120822; 32120823; 32120810
Installation Location	60:42; 70:32	12A; 20; 11A
Distance to Ground	205 cm	1.84m; 1.76m; 1.86m
Target	Green artificial grass mat 2m x 2.5m	Textured Plastic, 1.2mx1.2m
Height of Temperature	Integrated	Integrated
Instrument Heating	Automated de-icing of ultrasonic membrane; Automated de-icing of ultrasonic membrane AND heated boom	Automated de-icing of ultrasonic membrane
Mounting Beam	Horizontal boom; Boom angled 45°	Horizontal boom

2.2) Site Photos

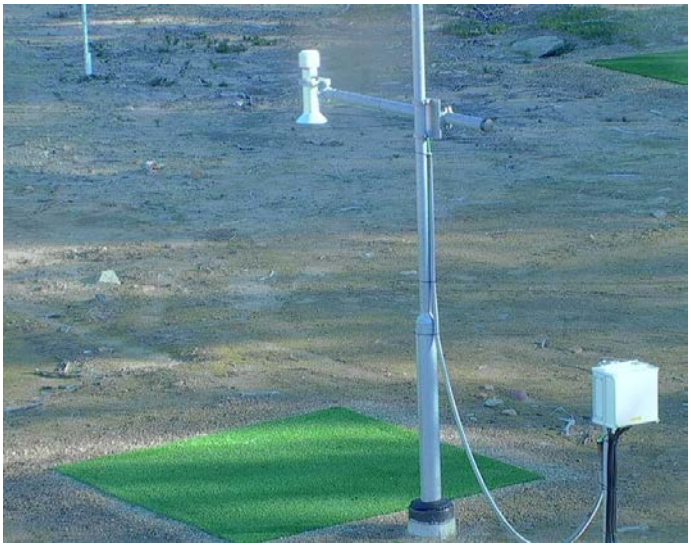


Figure 3: Sodankylä, horizontal mounting beam on 60:42

Figure 4: Sodankylä, angled mounting beam on 70:32

Figure 5: CARE, pedestal 12A. Instrument and target inside red square.

2.3) Instrument Footprint Diagram

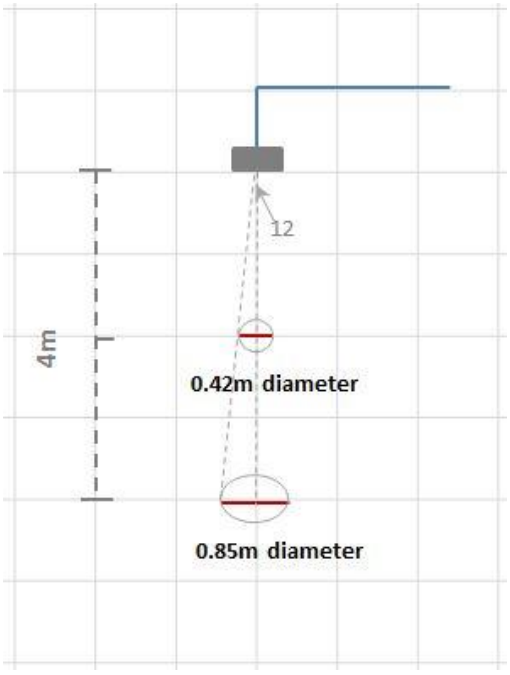


Figure 6: Conceptual diagram of instrument field of view

2.4) Environment Conditions During SPICE

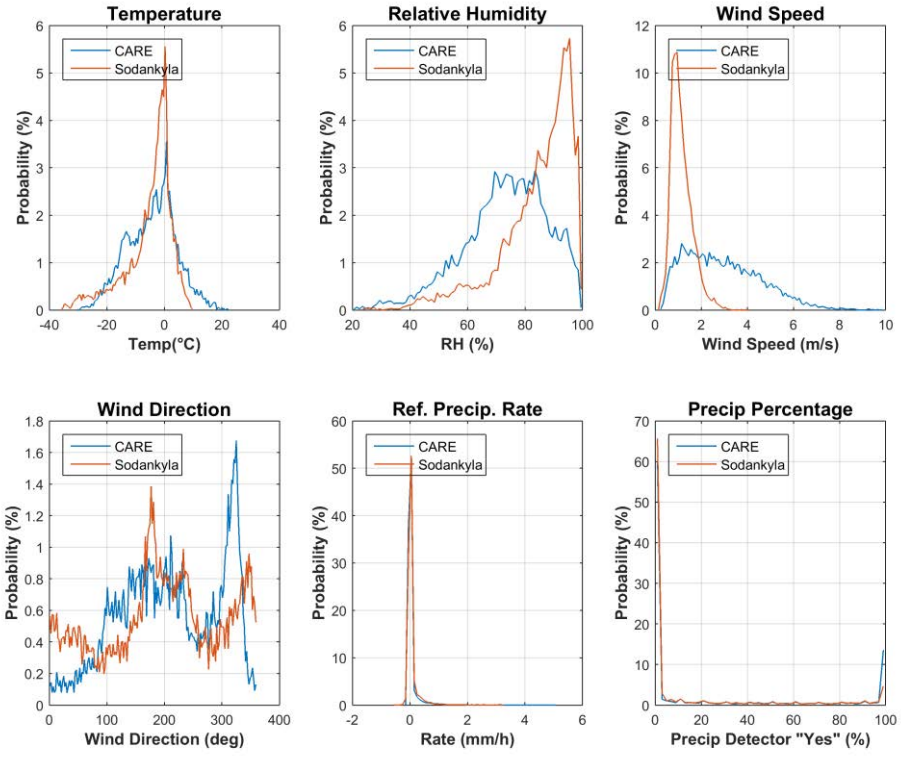


Figure 7: Summary of aggregated environmental conditions on the SPICE sites that operated a USH-8 for the entire duration of formal tests.

3) Data Output

Table 2: Parameters measured or output by the instrument

Measured Parameters	Units	Sodankylä	CARE
Corrected Snow Depth	mm	x	x
Sensor Status	-	x	x
Uncorrected Snow Depth	mm	x	-
Temperature	Deg C	x	-

Table 3: Measurement and data acquisition parameters

Measurement and Processing	Units	Sodankylä	CARE
Data Sampling Rate	[Minutes]	1	0.5
Data Acquisition Interval	[Minutes]	1	0.5
Date Processing		Sample	1-minute Average
Communication Protocol		RS-232	RS-232

4) Quality Control Information

Missing data occurred when data were not recorded by the data logger. This may or may not be related to the function of the instrument. Suspicious data were identified as outliers and are either removed automatically via pre-set range and jump filtering, or identified visually as an outlier and removed manually. Range and jump filters were set based on reasonable physical limits for the individual sites. Suspicious data may be a result of a malfunctioning instrument or from meteorological impacts on the instrument's measurement capability. Table 4 provides a seasonal breakdown of the quality control metrics by site. More commentary on individual instruments, where required, is included below. Only data classified as "Good" were used for the intercomparisons.

Table 4: Seasonal breakdown of data QC metrics

SEASON 2013-2014	Sodankylä	CARE
Collection Period	Oct 2013-June 2014	Nov 2013-April 2014
Good	98.3%	98.3%
Missing	1.7%	1.6%
Suspicious	0%	0.1%

SEASON 2014-2015	Sodankylä	CARE
Collection Period	Oct 2014 – June 2015	Nov 2014-April 2015
Good	99.2%	97.3%
Missing	0.8%	2.6%
Suspicious	0%	0.1%

5) Evaluation of the Ability to Perform Over a Range of Operating Conditions

There were several options for comparing these instruments to a reference. Both sites performed manual measurements in the Intercomparison Field that can be used as a measurement reference. Also, both sites hosted several automated snow depth sensors that when combined as an average, serves as a reference. Both intercomparisons are shown below. The summary statistics for the intercomparisons is shown in Section 6.1c.

5.1) Performance against the reference

6.1a Sodankylä

At Sodankylä, the manual reference consisted of a daily photograph observation of four graduated snow stakes distributed in the Intercomparison Field. The intercomparison below (Figures 8) shows the instrument plotted against the average of all stakes (red) and against Stake 4 which is closest to the instrument (blue). The time series of this intercomparison is shown in Figure 9 (2013/2014) and Figure 10 (2014/2015). The automated reference was an average of 1-minute snow depth data obtained from 6 instruments distributed in the intercomparison field. These instruments were the two Sommer USH-8s, two Campbell Scientific SR50ATHs, one Jenoptik/Lufft SHM30 and one Felix SL300. The USH-8s are plotted against this reference in Figure 11.

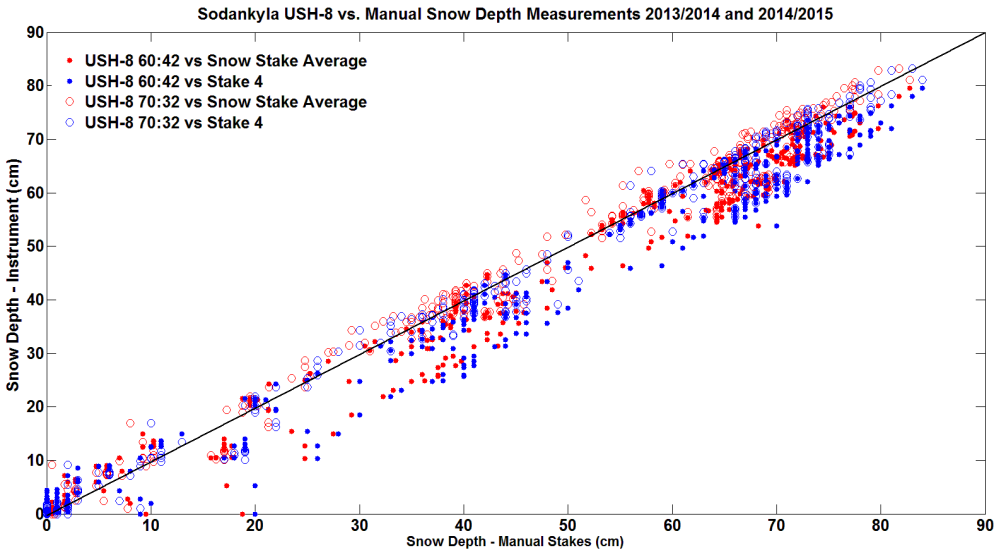


Figure 8: USH-8 at Sodankylä compared with the manual refererence which is either the average of the four snow stakes (red) or the closest snow stake to the sensor (blue). The intercomparison statistics for the four stake average are shown in the tables in Section 5.1c.

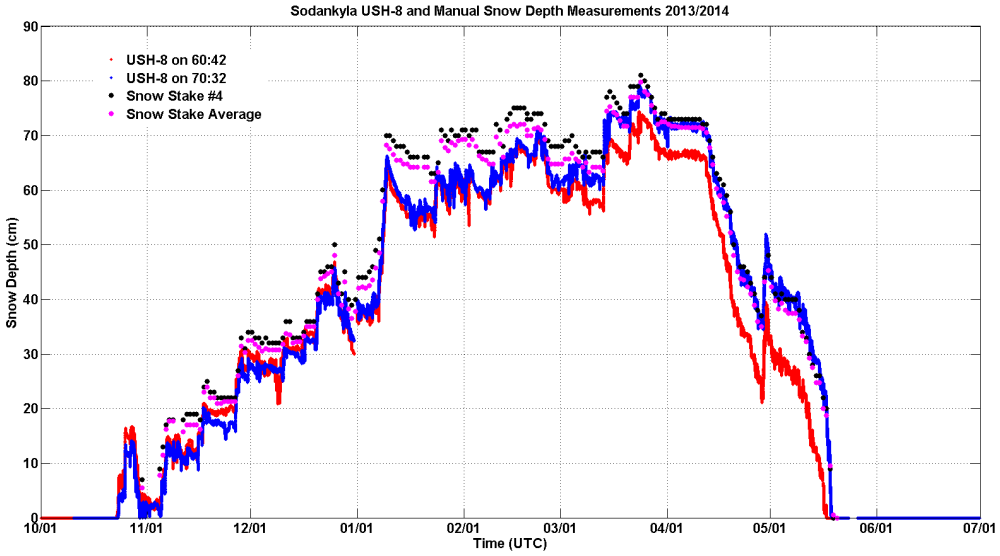


Figure 9: Time series of the USH-8s on pedestal 60:42 (red line) and pedestal 70:32 (blue line) and manual snow depths at Sodankylä for 2013/2014 including both the four snow stake average (magenta markers) and the snow stake closest to the sensor (black markers).

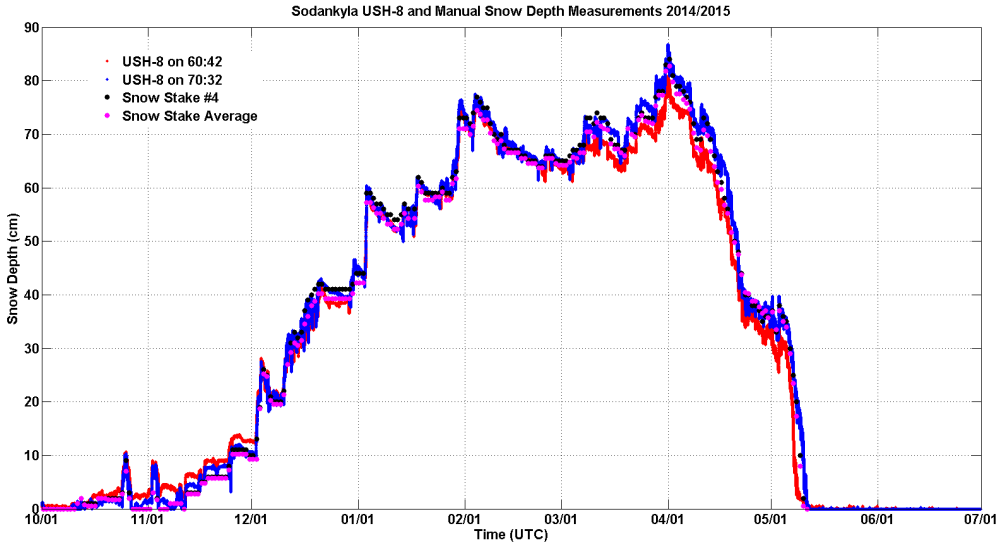


Figure 10: Time series of the USH-8s on pedestal 60:42 (red line) and pedestal 70:32 (blue line) and manual snow depths at Sodankylä for 2014/2015 including both the four snow stake average (magenta markers) and the snow stake closest to the sensor (black markers).

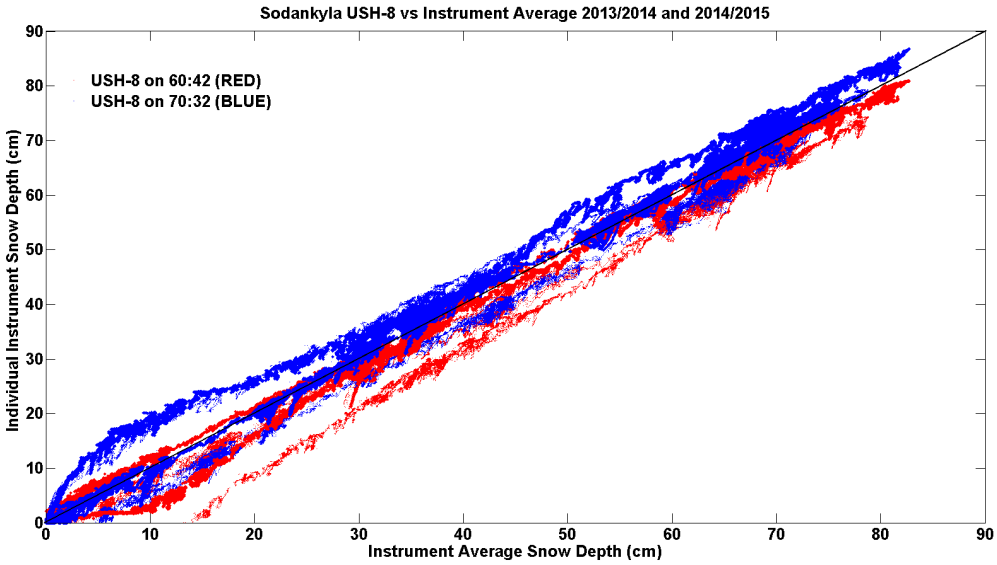


Figure 11: The USH-8s on pedestals 60:42 (red) and 70:32 (blue) compared with the average of six sensors at Sodankylä. The intercomparison statistics for the five sensor average are shown in the tables in Section 5.1c.

6.1b CARE

At CARE, each of the three pedestals (12A, 11A, and 20) included a USH-8 aimed at a 1.2m x 1.2m plastic target under the sensor. Each target had a permanent snow stake at each corner which was observed visually each day. The average of all four stakes at the corner of each target comprised the manual reference for each of the three pedestals. This intercomparison is shown in Figure 12 with the time series shown in Figure 13 (2013/2014) and Figure 14 (2014/2015). The automated reference consisted of the four automated instruments at each of the three pedestals averaged for each minute. These instruments were the Sommer USH-8, the Campbell Scientific SR50A, the Jenoptik/Lufft SHM30 and the Felix SL300. Each pedestal, therefore, had an independent automated reference. This intercomparison, separated by pedestal, is shown in Figure 15.

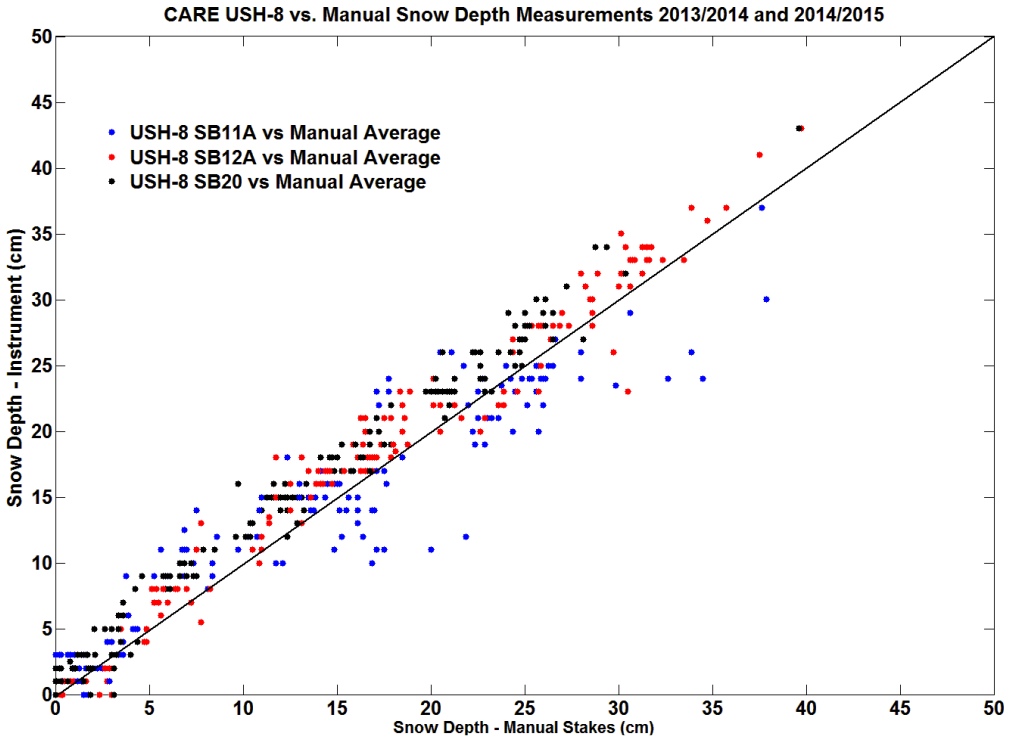


Figure 12: USH-8 at CARE compared with the manual refererence which the average of the four snow stakes. The intercomparison statistics for the four stake average are shown in the tables in Section 5.1c.

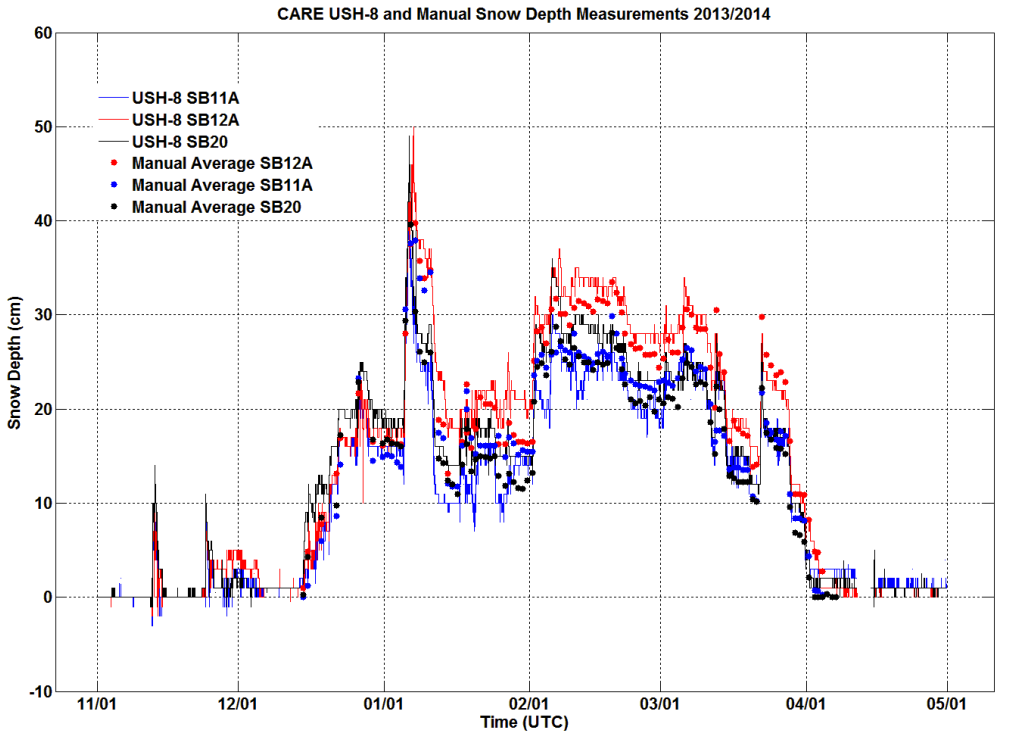


Figure 13: Time series of the USH-8s on pedestal 11A (blue line) ,pedestal 12A (red line), and pedestal 20 (black line) and manual snow depths at CARE for 2013/2014 including the four snow stake average.

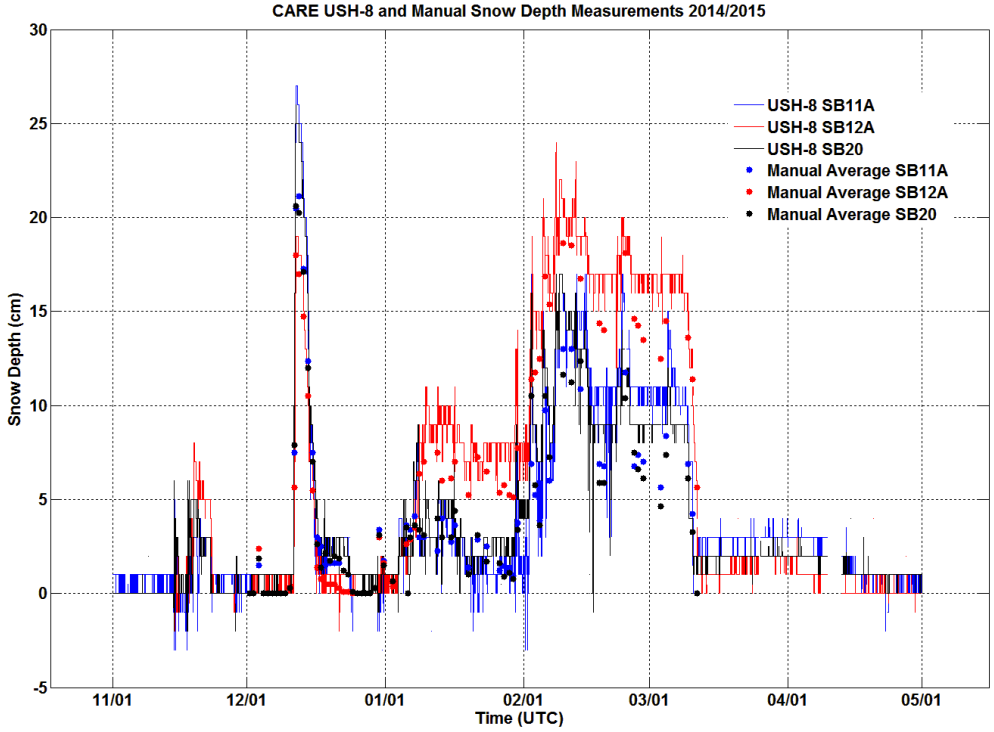


Figure 14: Time series of the USH-8s on pedestal 11A (blue line) ,pedestal 12A (red line), and pedestal 20 (black line) and manual snow depths at CARE for 2014/2015 including the four snow stake average.

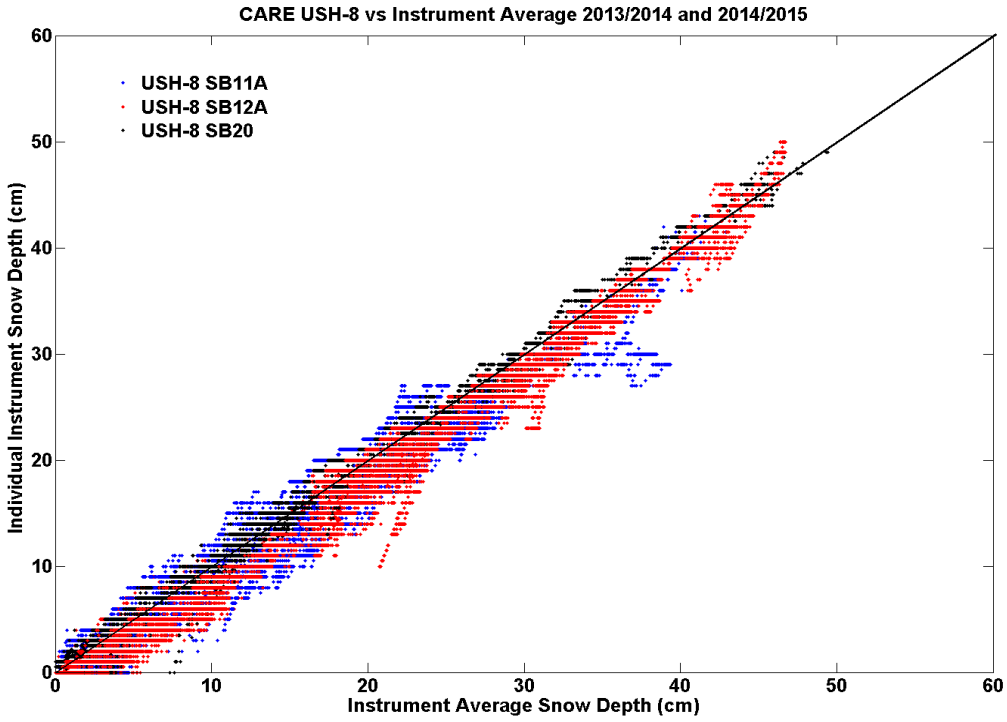


Figure 15: The USH-8s 11A (blue), 12A (red), and 20 (black) compared with the average of 4 sensors at CARE. The intercomparison statistics for the five sensor average are shown in the tables in Section 5.1c.

5.1c Intercomparison Statistics Summary

Table 5: Summary statistics for the USH-8 on pedestal 60:42 at Sodankylä

Sodankylä 60:42	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.98	0.97	-4.4	6.6	195
	2014-2015	0.99	0.96	1.5	2.6	221
Automated Reference	2013-2014	0.99	0.94	-1.4	4.4	386302
	2014-2015	0.99	0.99	0.63	1.3	390085

Table 6: Summary statistics for the USH-8 on pedestal 70:52 at Sodankylä

Sodankylä 70:52	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.98	1.00	-2.3	4.0	195
	2014-2015	0.99	1.01	0.85	2.1	221
Automated Reference	2013-2014	0.99	0.99	-1.00	3.2	372744
	2014-2015	0.99	1.03	0.26	2.2	390081

Table 7: Summary statistics for the USH-8 on pedestal 12A at CARE

CARE 12A	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.95	1.02	1.09	2.5	107
	2014-2015	0.97	1.11	0.42	1.8	69
Automated Reference	2013-2014	0.99	0.97	-0.78	1.6	256165
	2014-2015	0.98	0.92	-0.13	1.3	251695

Table 8: Summary statistics for the USH-8 on pedestal 20 at CARE

CARE 20	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.98	1.00	2.3	2.6	107
	2014-2015	0.95	1.19	0.61	2.1	70
Automated Reference	2013-2014	0.99	1.00	-0.31	0.72	256201
	2014-2015	0.98	0.95	0.26	0.72	254669

Table 9: Summary statistics for the USH-8 on pedestal 11A at CARE

CARE 11A	Season	r^2	Slope	Intercept (cm)	RMSE (cm)	n
Manual Stake Average	2013-2014	0.86	0.80	2.7	3.3	106
	2014-2015	0.93	1.25	0.5	2.6	69
Automated Reference	2013-2014	0.99	0.94	-0.5	1.8	256186
	2014-2015	0.96	0.96	0.4	1.0	254403

5.2) Factors that influence instrument performance

At Sodankylä, the snow depth instrumentation was distributed in the NE quadrant of the Intercomparison Field with the average distance between instruments at approximately 16m. The average distance between the USH-8 at 60:42 and the four snow stakes was 22.5 m with the closest stake (Stake 4) being 7 m from the instrument. The average distance between the USH-8 at 70:32 and the four snow stakes was 31 m with the closest stake (Stake 4) being 7 m from the instrument. Even though the snow depth in the Intercomparison Field was relatively uniform, there was still variability in depth across 7 m and certainly across 31 m. This could account for much of the scatter shown in reference intercomparison for Sodankylä (Figures 8 and 11). It was also noted at this site that differential melt occurs between the targets under the sensors and the snow stakes and even amongst the targets (depending on their individual exposures). The effects of this differential melt can be seen in both the manual and automated reference plots, especially at lower snow depth amounts during mid- to late-melt. Figure 11 shows a bi-modal relationship between the USH-8s and the other instruments as the different pedestals accumulate and melt snow at different rates. This means that source of the scatter shown in the above figures is likely not due to sensor bias.

The design and installation of the manual snow stakes at the corner of each target at the CARE site was implemented to measure the snow depth as close as possible to the sensor without being within the sensor's FOV. Unfortunately, the manual snow stakes created a mounding effect, increasing the depth of the snow in the middle of the target as compared to the corners of the target. This is quite evident in Figures 12, 13 and 14 which suggests that the sensor overestimated snow depth as compared to the manual measurement. Another factor to consider when interpreting the intercomparison at CARE is the potential for a change in the reference distance between the target and the sensor over the course of the intercomparison season. This is called "zero snow depth drift" and is discussed in the main body of the SPICE final report. The levels of the targets at CARE relative to the sensor were susceptible to change due to settling after installation (under the weight of the snow pack) or from frost heave. The change in this position can only be assessed at the end of the

season when the targets are bare but it is recognized that changes in the zero snow depth distance can be continuous and undetectable during the season and therefore the impact on the snow depth intercomparison is difficult to assess. However, it is estimated that the impact would not exceed ± 2 cm during the season. This error is generally not a concern at the other sites where either natural targets are used or the artificial targets are not prone to heaving or settling.

5.3) Performance Considerations

CARE experienced an issue with a sensor parameter which had the sensor reporting snow depth to the nearest cm rather than the nearest mm. This explains the appearance of the sensor data plotted in Figures 12 through 15 and may have a small impact on sensor performance. Other than this, no limitations to instrument performance were noted by the site managers during the intercomparison.

5.4) Maintenance

The Sodankylä site manager occasionally reported that snow would collect on the top of the instrument and instrument mount due to the lack of wind at the Sodankylä site. When this occurred, the instrument was left for a 24hour period to self-clear prior to having it cleared manually. There were no reports stating that this snow collection actually impacted the instrument's capability to make a measurement but there was some concern that the target surface beneath the sensor was impacted from falling snow when this snow was cleared from the sensor. Potential impacts are discussed elsewhere in the SPICE final report.

6) Lessons Learned

The USH-8, as tested, appeared to be a reliable instrument and experienced no data breaks during the testing period. The site manager at Sodankylä reported that the instrument was easy to install and calibrate. However, reports from Sodankylä noted that the instrument, which is equipped with a cone like shield placed at the bottom of the instrument around the signal transmitter/receiver, tended to collect snow during low wind snowfall events. This snow eventually dropped off onto the surface target although no evidence of this dropping snow could be discerned in the instrument measurements. This has been documented to occur about 16 times during the 2013/2014 season at Sodankylä. Sensor 70:32 was installed on a boom extending out at a 45° angle to test if this change would reduce the amount of snow collecting on the instrument and the mounting infrastructure. Photographic evidence suggested that there appeared to be a small reduction in the amount of snow collecting on the instrument cone.

SPICE Instrument Performance Report
Sommer Messtechnik – SSG1000

1) Technical Specifications

Physical principle: load cell measures the weight of the snow on the measurement platform
Scale Surface Area: 2.8 m x 2.4 m
Measurement Area: 1.2 m x 0.80 m
Measurement Range: 0 to 1000 mm Snow Water Equivalent (SWE)
Measurement Accuracy: 0.3% of full range
Resolution: 1 mm SWE, 0.1 kg/m^2
Link to manual:
http://www.hydrologicalusa.com/images/uploads/Manual_SSG_V1.4_EN-2_.pdf

Figure 1: Photo of the SSG1000

2) SPICE Test Configuration

Test Site: Sodankylä (Finland)

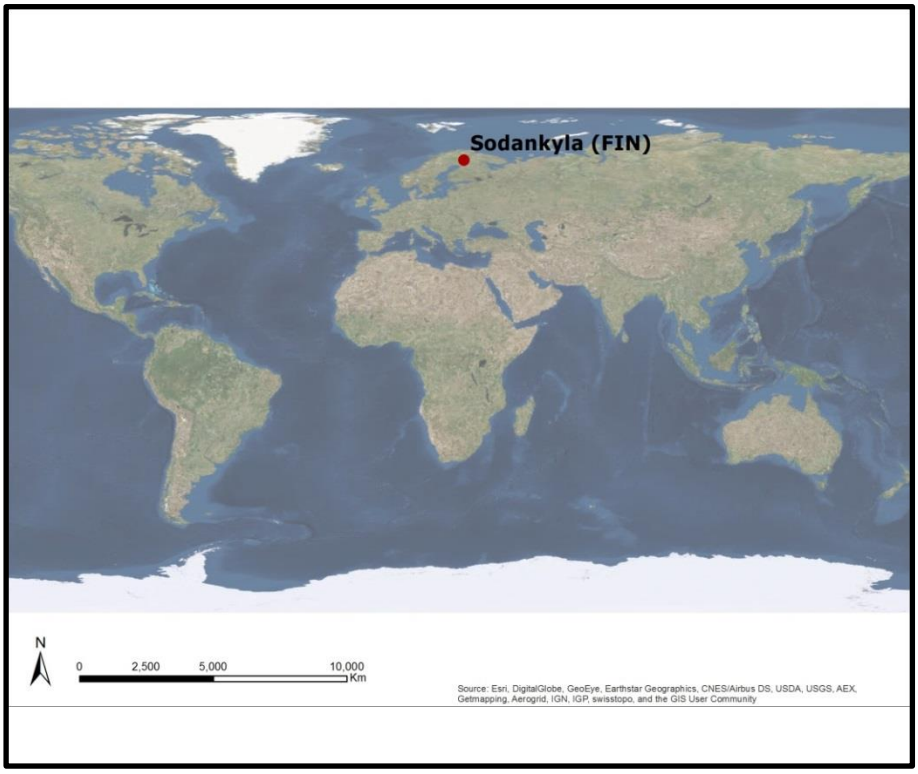


Figure 2: Location of sensors under test

2.1) Site Specifications

Table 1: Site specific instrument and installation details

	Sodankylä
Date of Installation	17-Oct-2013
Serial Number(s)	Not Available
Installation Location	60:52

2.2) Site Photos

Figure 3: Sommer Messtechnik SSG1000 snow scale at the Sodankylä SPICE intercomparison site.

2.3) Instrument Footprint Diagram

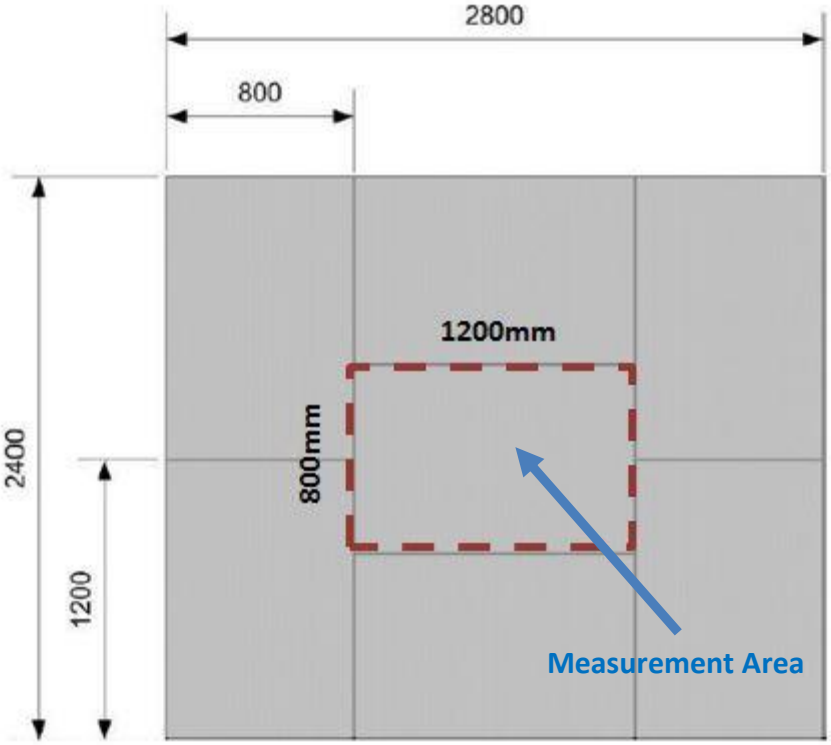


Figure 4: Conceptual diagram of instrument dimensions and measurement area

2.4) Environment Conditions During SPICE

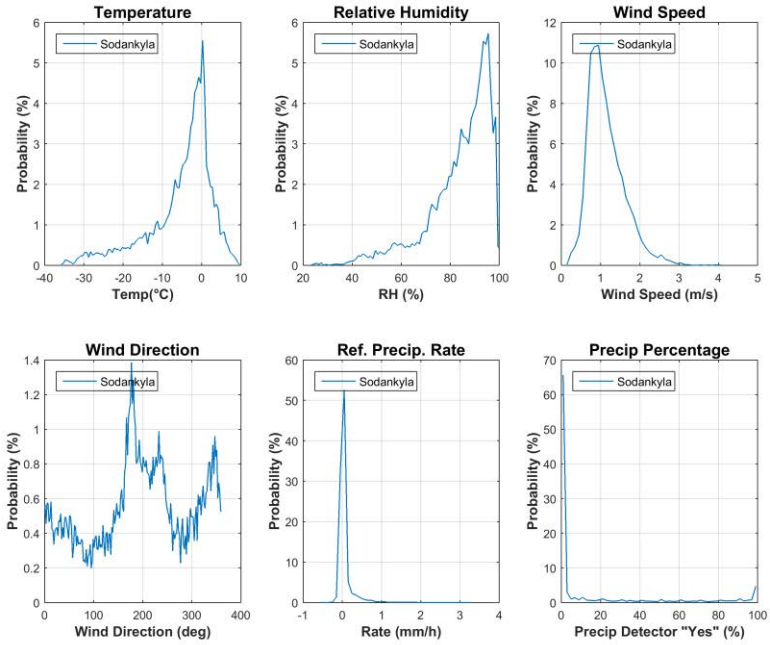


Figure 5: Summary of aggregated environmental conditions on the SPICE sites that operated a SSG1000 for the entire duration of formal tests.

3) Data Output

Table 2: Parameters measured or output by the instrument

Measured Parameters	Units	Sodankylä
Snow Water Equivalent	mm	x

Table 3: Measurement and data acquisition parameters

Measurement and Processing	Units	Sodankylä
Data Sampling Rate	[Minutes]	1
Data Acquisition Interval	[Minutes]	1
Date Processing		Sample
Communication Protocol	Analog	

4) Quality Control information

Missing data occurred when data were not recorded by the data logger. This may or may not be related to the function of the instrument. Suspicious data were identified as outliers and were either removed automatically via pre-set range and jump filtering, or identified visually as an outlier and removed manually. Range and jump filters were set based on reasonable physical limits for the individual sites. Suspicious data may have been a result of a malfunctioning instrument or from meteorological impacts on the instrument's measurement capability. Table 4 provides a seasonal breakdown of the quality control metrics. More commentary on individual instruments, where required, is included below. Only data classified as "Good" were used for the intercomparisons.

Table 4: Seasonal breakdown of data QC metrics

SEASON 2013-2014	Sodankylä
Collection Period	Oct-June
Good	82.6%
Missing	8.4%
Suspicious	9.0%

SEASON 2014-2015	Sodankylä
Collection Period	Oct-June
Good	66.7%
Missing	2.3%
Suspicious	31.0%

The quality metrics breakdown for the SSG1000 shows a relatively high percentage of "Suspicious" data, especially for the 2014/2015 season. The reason for this is a malfunction of the instrument towards the end of the intercomparison seasons. For both years, the malfunction was attributed to water damage of the electronics which resulted in the sensor reporting spurious values leading up to a complete instrument failure. This is discussed further in Sections 5.3.

5) Evaluation of the Ability to Perform Over a Range of Operating Conditions

The instrument at Sodankylä was compared against manual observations of Snow Water Equivalent. The frequency of the manual measurement was approximately every two weeks and was obtained at a single sample location (50:66) at a distance of about 16 m from the centre of the sensor's footprint.

5.1) Performance against the reference

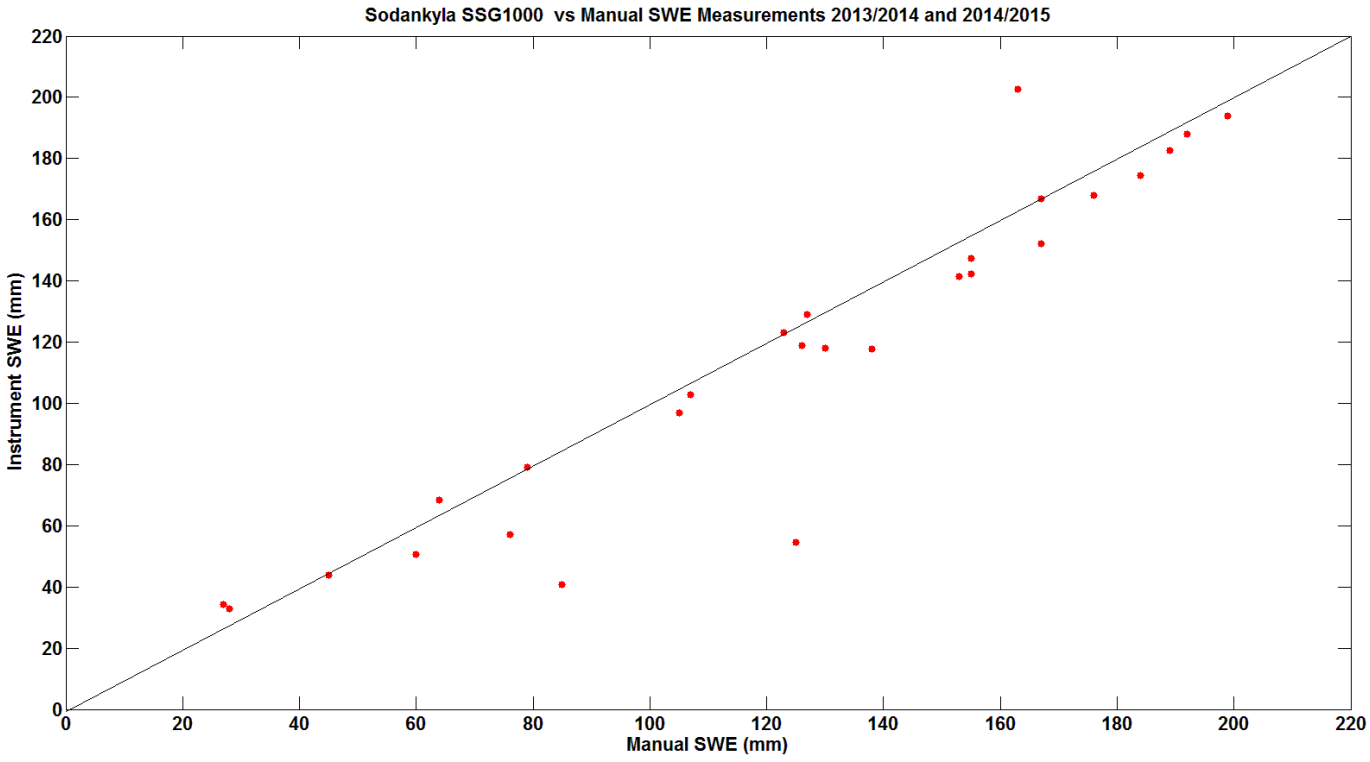


Figure 6: SSG1000 vs Manual SWE for 2013/2014 and 2014/2015

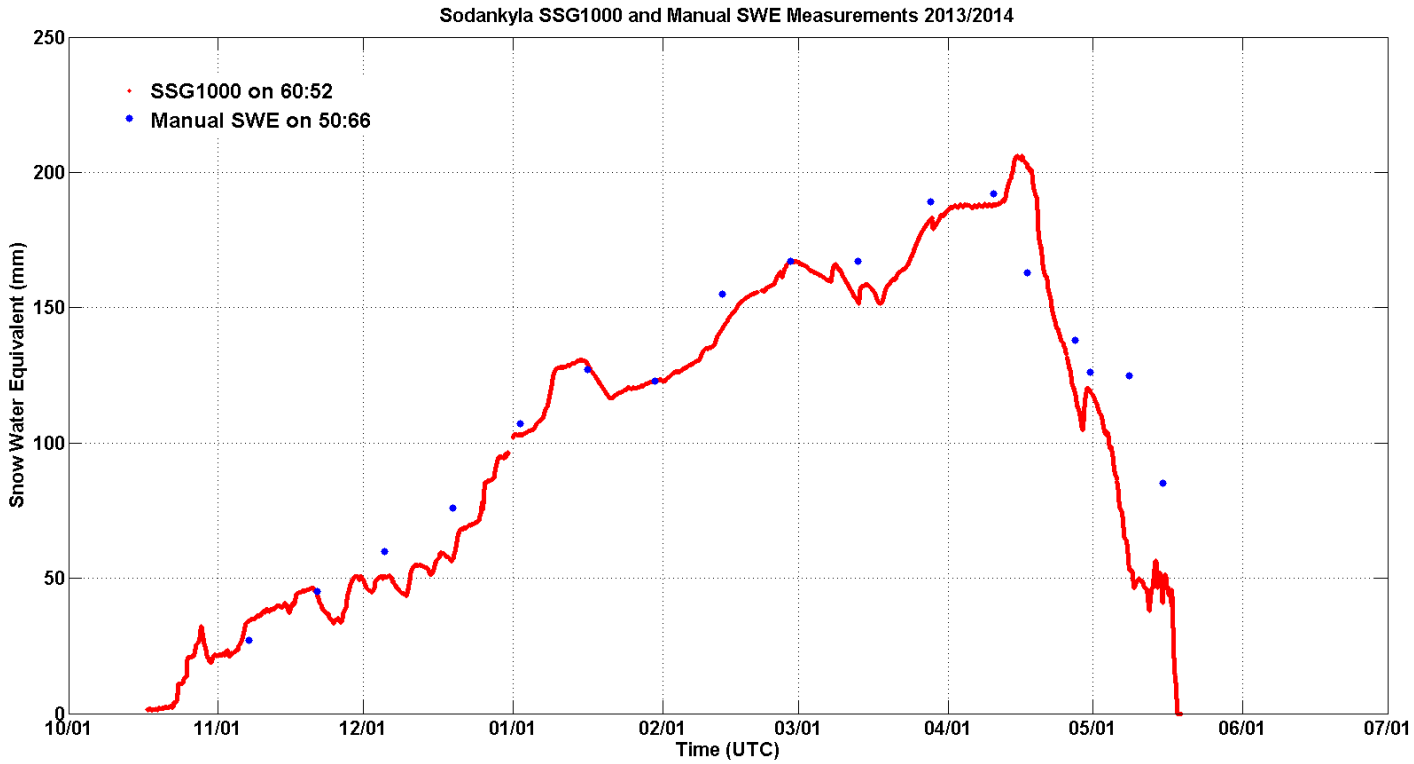


Figure 7: SSG1000 and Manual SWE time series for 2013/2014

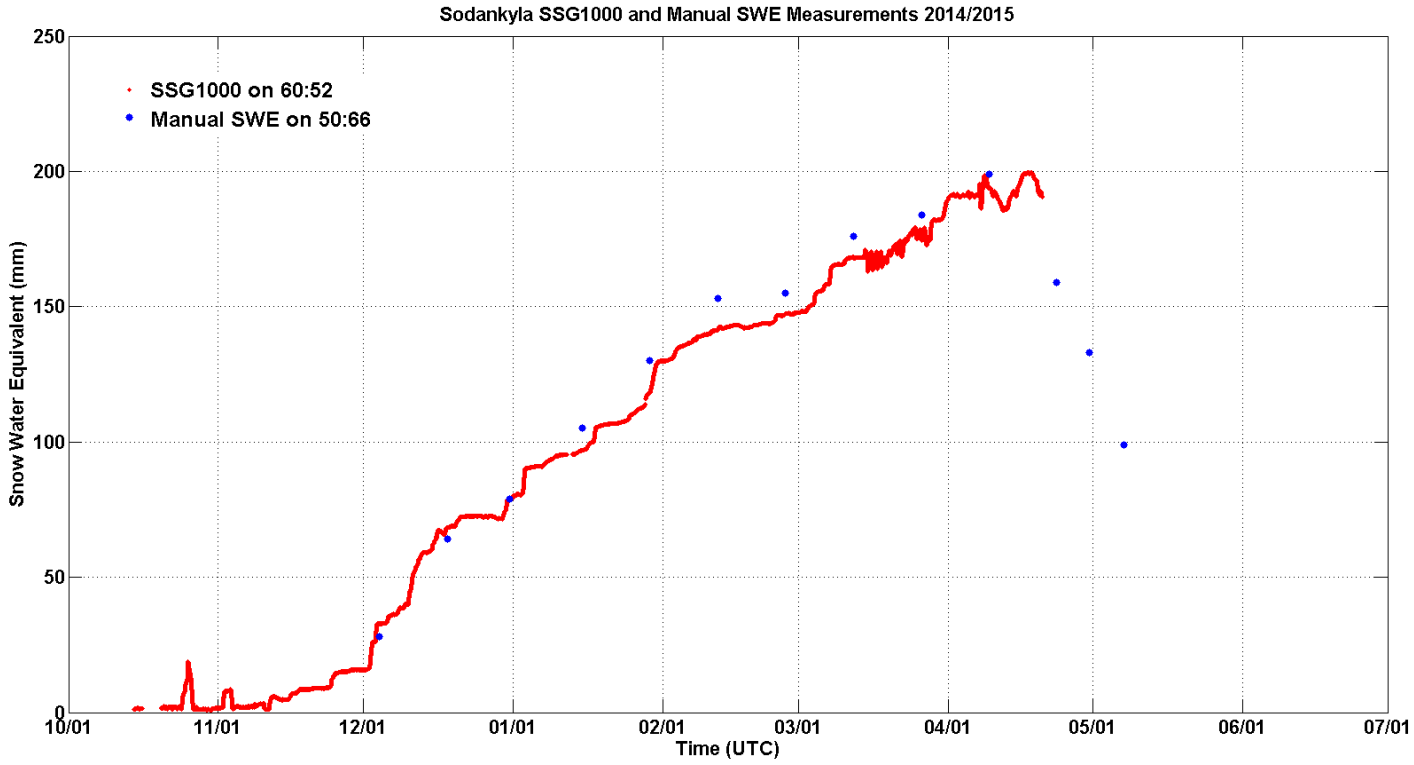


Figure 8: SSG1000 and Manual SWE time series for 2014/2015

Table 5: Summary statistics for the SSG1000 at Sodankylä

Season	*Slope*	Intercept mm w.e.	r^2	RMSE mm w.e.	Mean Relative Bias (K)	n
2013/2014	*1.05*	-15.5	0.84	24.2	-15.1%	17
2014/2015	*0.92*	5.5	0.99	7.9	-2.3%	10
Combined	*0.99*	-7.3	0.88	19.8	-10.8%	27

5.2) Factors that influence instrument performance

There were possible "bridging" occurrences in the 2013/2014 season. Bridging occurs when snow builds up across the measurement platform but is not entirely in contact with the measuring surface. The result is the sensor underestimating the true snow pack. These occurrences could be the cause of the outliers shown in Figure 6 and Figure 7. Bridging can occur when warmer temperatures are followed by decreasing temperatures resulting in a crusty snow layer forming near the top of the snow pack. This potentially creates a "bridge" such that the full weight of the overlying snow pack is no longer transferred to the weighing mechanism, resulting in an underestimation of SWE. Users should be aware of this potential error when doing data quality control.

Another environmental factor that could influence performance of the sensor is settling or frost heave of the sensor itself. Depending on the substrate and installation, the sensor could either settle deeper into the soil during the accumulation season or be pushed out of the soil by frost heave. Potentially, both could occur during an accumulation season. Either situation could change the force that the snow pack is applying on the weighing mechanism and thereby influence the accuracy of the SWE measurements. The occurrence of this issue was not noted during SPICE and no recommendation can be made on how to correct resulting errors. It is suggested that users examine the sensor platform at the beginning and end of each accumulation period, noting any change in the platform's height relative to the surface of the soil. These errors can be minimized or eliminated by following the manufacturer's recommendations to install the sensor on solid ground and/or use wood or concrete supports under the sensor platform.

5.3) Performance considerations

The intercomparison of the sensor with the manual reference was also influenced by potential errors in the manual reference measurement. Errors in the manual measurements using a snow tube tend to occur when sampling a complex snow pack with ice layers (Powell, 1987). Depending on the skill of the observer and the condition of the sampler, snow will tend to escape out of the bottom of the tube while cutting through the layers of ice in the pack. This would result in an underestimate in SWE, perhaps as high as 5-10%.

5.4) Maintenance

The instrument had to be returned to the manufacturer at the end of the 2013/2014 season (after the SPICE observation period) due to a failure related to water damage. The instrument failed again early spring of the 2014/2015 season for the same reason, and was returned to the manufacturer for repair. Re-installation of the instrument after the end of the SPICE intercomparison periods involved re-locating the electronics to a higher location to prevent water infiltration and damage. This adaptation seems to have corrected the issue for SPICE. Since the end of the SPICE field intercomparison, the manufacturer has also made other changes to make the sensor electronics less susceptible to water damage.

6) Lessons Learned

The SSG1000 sensor provided an estimate of snow water equivalent at a very high temporal resolution that is very closely correlated with manual measurements. There appeared to be very few cases of over- or under-estimation of SWE. Two outliers in Figure 6 show the SSG1000 substantially underestimating SWE, but these SWE estimates are from late in the 2013/2014 melt season and are likely a result of sensor “bridging” but could also be exacerbated by differential melting in the Intercomparison Field.

The instrument design tested during SPICE is susceptible to water damage of the electronics during snow melt and it is highly recommended that the electronics be re-located to a higher location where they will be less exposed to melt water.

References

Powell, D.: Observations on consistency and reliability of field data in snow survey measurements, paper presented at 55th Annual Meeting Western Snow Conference, Vancouver, B. C., 1987, 69‑77, 1987.

Antifreeze and oil mixtures used in Geonor/Pluvio² reference gauges at SPICE sites

Maximum capacity of various bucket sizes

600mm: 12 l max capacity
1000 mm: 20 l max capacity
1500 mm: 30 l max capacity

Antifreeze

	Temp [°C]	Total volume [l] / % of total gauge capacity	Ethylene glycol	Methanol	Propylene glycol	Ethanol with additives (methylated spirits)	Powercool DC 924-PXL Based on propylene glycole plus coolant	Water
Geonor manual Alternative 1 (600mm)	-35 -30 -25 -20 -15 -10 -5	6.0 l (50%) 5.6 l (47%) 5.0 l (42%) 4.2 l (35%) 3.5 l (29%) 2.6 l (22%) 1.5 l (12.5%)	2.4 l 2.3 l 2.0 l 1.7 l 1.4 l 1.0 l 0.6 l	3.6 l 3.3 l 3.0 l 2.5 l 2.1 l 1.6 l 0.9 l				
Geonor manual Alternative 2 (600mm)	-51 -46 -40 -34 -29 -23 -20 -18 -12	7.2 l (60%) 7.0 l (58 %) 6.6 l (55 %) 6.2 l (52 %) 5.8 l (48 %) 5.16 l (43 %) 4.8 l (40 %) 4.3 l (36 %) 3.6 l (30%)		4.32 l 4.2 l 3.96 l 3.72 l 3.48 l 3.096 l 2.88 l 2.58 l 2.16 l	2.88 l 2.8 l 2.64 l 2.48 l 2.32 l 2.064 l 1.92 l 1.72 l 1.44 l			

	-7 -3	2.4 l (20 %) 1.2 l (10 %)		1.44 l 0.72 l	0.96 l 0.48 l			
Canada Geonor 600 mm	-10 -5	3 l (25%) 2 l (17%)		1.8 l (60%) 1.2 l (60%)	1.2 l (40%) 0.8 l (40%)			
New Zealand Geonor 1000mm		3.62 l (18 %)			50%	50%		
Norway Geonor 600mm	-25 -10	5 l (42%) 2.5 l (21%)	2 l (40%) 1 l (40%)	3 l (60%) 1.5 l (60%)				
Norway Geonor 1000m	-25	8.3 l (42%)	3.3 l (40%)	5 l (60%)				
USA Geonor 600mm Boulder Testsite	-34 -29 -23 	6.2 l (52%) 5.8 l (48%) 5.16 l (43%) 4.8 l (40%)		3.72 l (60%) 3.48 l (60%) 3.096 l (60%) 2.88 l(60%)	2.48 l (40%) 2.32 l (40%) 2.064 l (40%) 1.92 l (40%)			

Pluvio^2 OTT manual (200cm^2, 1500mm)	-34 to -6 -34 to -10 -34 to -15	5.5 l (18%) 8.5 l (28%) 11 l (37%)					5 l 7.5l 10 l	0.5 l 1 l 1 l
Switzerland Pluvio^2 Ott	-40 to -12	10 l (33%)			7.5 l			2.5 l
Finland Pluvio^2 Ott	> -58	5 l Meltium = Kalium formate liquid (49-51% concentration)						

Table 1: Antifreeze mixture for Geonor and Pluvio^2 gauges, as recommended from the manufacturers and as used in different SPICE sites.

Oil

	Volume	Type	Temp-range [°C]
Geonor Manual (600mm, 1000mm)	0.4 l	Esso (Exxon) UNIVISJ13	
Canada	0.5 l	5W30 weight motor oil	-15 to -20
New Zealand	”a little”	BP Enerpar M002 (viscosity 15) or M006 (viscosity 68) Safe for consumption if in contact with food	
Norway (600mm,1000mm)	0.4 l	Statoil Hydraway 15LT	
USA – Marshall	0.47 l	Super Hydraulic Oil	
Switzerland	0.4 l	Isopar	

Table 2: Oil type and volume used in different SPICE sites to prevent evaporation from the reference gauges Geonor 600mm and Pluvio^2.

Issues encountered with the PowerCool recommended for use with the OTT Pluvio^2, at Col de Porte, during WMO-SPICE

Yves Lejeune, Jean-Michel Panel, Samuel Morin

Durring the SPICE fiele campaigns an OTT Pluvio^2 400 cm² supplied by Météo-France was installed at the Col de Porte field site. The site has a long-term experience in the use of alternative weighing-gauges instruments, in particular GEONOR T200 1B since 1993. Two main types of challenging situations were observed with the PowerCool during the year 2013-2014, leading to the use of the GEONOR-recommended antifreeze mixture for the year 2014-2015 in the Pluvio2.

1. Use of PowerCool together with oil.

In order to prevent evaporation, the staff at Col de Porte regularly uses Univis oil on top of the antifreeze mixture used. This was originally done on top of the PowerCool mixture. On 31 January 2014, with 90 % of the Pluvio^2 capacity reached, it was decided to empty the bucket at around 11:00 UTC. During the operation, it was observed that during the previous days an ice ring had formed within the bucket, and that there were snow and ice aggregates floating in the bucket. The formation of these features must have occurred between 27 January and 31 January, during which the air temperature never exceeded 0 °C, was -3.6 °C on average with a minimum of -6.4 °C and a maximum of -0.2 °C. Cumulative snow precipitation was on the order of 20 kg/m during this time period. After having emptied the bucket from its liquid part, pictures of the ice blocks attached to the bucket sides and at its bottom were taken, see below :

It seems that the problem encountered could be due to the use of the Esso (Exxon) Univis J13 oil in addition to the PowerCool mix. Although with the antifreeze mix recommended by GEONOR (methanol and ethylene glycol) the oil has never frozen in the weighing gauges at Col de Porte, it appears that the miscibility between the Powercool, water and oil was not appropriate leading to oil freezing (reddish ice features, due to the red color of the oil).

Since this event, oil was never added anymore to the PowerCool mixture. It is not recommended to use PowerCool together with anti-evaporation oil, based on this experience.

Figure 1 : Picture of the Pluvio^2 bucket after the emptying of the bucket, showing an ice ring (containing some oil hence the red color) on 31 January 2014.

2. Frozen anti-freeze

On 19 February and 27 February 2014, a mixture ice/snow/water was observed inside the bucket.

19 February : In this case, snowfalls occurred on 16 February (27 kg/m), and from 17 February 18:00 UTC to 18 February 2:00 UTC mean air temperature was -1.2 °C with -1.5 °C/-0.2 °C minima/maxima, respectively, then rose from 18 February at 2:00 UTC to 19 February at 15:00 UTC with mean air temperature of 1.8 °C with 0.0 °C/6.0 °C minima/maxima. After such a warm period, it would be anticipated that no solid water could still be observed in the bucket, given that it was originally filled with 7.5 L PowerCool and 3 L water, which is supposed to ensure freezing protection down to -34 °C for a 60 % filling rate and -14 % for a 80 % filling rate. At the time of the observations, the filling rate was 58 % only.

27 February : Similar to 19 February we observed a mixture of snow, ice and water in the bucket, as well as a thin ice ring attached to the bucket sides. From 26 February 6:00 UTC to 27 February 9:00 UTC, mean air temperature was -1.2 °C with minima/maxima of -4.1 °C and -0.8 °C and 36 kg/m precipitation was recorded. The bucket filling rate was 65 %, under which conditions it was totally unexpected to observe snow and ice in the bucket.

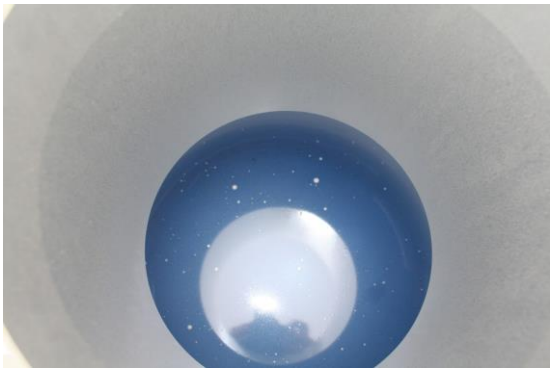


Figure 2a (left) : Frozen bucket on 19 February 2014 at Col de Porte Figure 2b (right) : Frozen bucket on 17 February 2014 at Col de Porte

06.2014 – 08.2014

OTT Pluvio^2: Cold chamber tests to investigate different types of oil and anti-freeze

Table of contents

Table of figures

1. Introduction

For the accurate measurement of solid precipitation in rain gauges, these are usually filled with an initial charge of anti-freeze and oil. The anti-freeze mixture has the function of preventing the freezing of the bucket content, whereas the oil prevents the evaporation of the volatile anti-freeze mixtures. The freezing of the bucket content or formation of slush within the bucket may both lead to inaccurate measurements as the contents “drift” around in the bucket, or cause for difficulties when emptying the gauges. Also, slush or ice drifting on the surface of the mixture and penetrating the oil layer may be lost to evaporation and sublimation processes.

The current mixtures used within the Swiss meteorological network are a Propylene Glycol mixture without oil in the Pluvio2 gauges (since the Accumulated NRT parameter in the Pluvio2 gauge compensates for evaporation losses) and a calcium chloride with water mixture topped by a layer of Vaseline in the totalizer rain gauges.

Within the context of SPICE (Solid Precipitation Inter Comparison Experiment) the accurate measurement of solid precipitation and thus the choice in type of oil and anti-freeze mixture used in the precipitation gauges is important. For reasons of comparability with other types of rain gauges, the Pluvio2 gauges were also filled with this mixture. During the 2012-2013 measurement season, the precipitation gauges were filled with a mixture of Propylene Glycol (PG) and linseed oil. The oil was chosen for its environmental friendly nature which would allow the emptying of the gauges on the terrain (the PG also being relatively harmless). However, the feedback from MétéoSwiss technicians indicated two problems with this mixture:

- The contents of one of the gauges appeared to be frozen when the bucket was emptied.
- Over the season, the oil formed a solid layer in the gauges (Figure 2, Appendix A).

Also within the context of SPICE, extensive studies have been conducted by Jeffery Hoover (et al. 2014) on the properties of different types of oil and anti-freeze mixtures. His recommendations included an anti-freeze mixture of PG and Methanol with the oil Isopar M.

In order to better understand what happens to the different oil and anti-freeze mixtures when subjected to low temperatures and to evaluate which mixture to use for the 2013-2014 SPICE season, the above mentioned mixtures were subjected to tests in a cold chamber and subsequently subjected to qualitative tests.

2. Methods

A schematic of the experimental setup, the quantities used for the anti-freeze and oil mixtures and the types of mixtures tested are given in Figure 1. A detailed log of the tests and temperature settings can be found in the Appendix B.

The PG/linseed oil mixture for example, was first tested without any addition of water (except for the standard 2,5 L of water which is added to the PG in order to saturate it, thus preventing the hygroscopic effect of the PG (attracting water molecules from its environment). The mixture was than subjected cold temperatures in steps of -5 °C (from 0 °C to -25 °C, the lowest temperature to be reasonably expected over a longer period on the SPICE site (Weissfluhjoch, Davos, Switzerland)). At every temperature step, the mixture was subjected to qualitative tests. These tests included:

- visual observations of the bucket contents (pictures).
- stirring the contents with a probe.
- dripping water droplets onto the surface to see if and how fast these pass through the oil layer.
- shedding frost onto the surface to see if and how fast this passes through the oil layer.

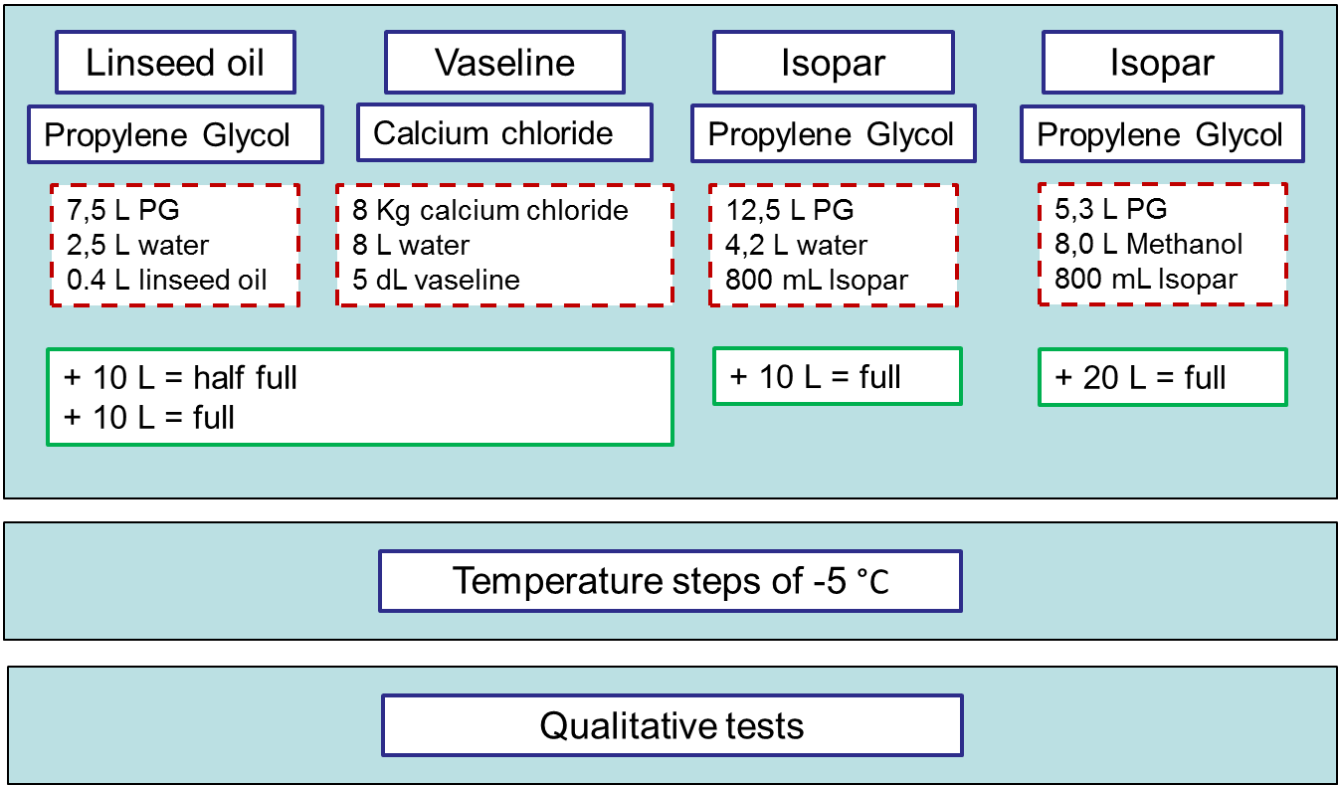


Figure 1: Experimental setup.

These steps were then repeated with additional quantities of water in order to investigate at what dilution and temperature the anti-freeze mixture would become less effective as well

as what consequences this might have for the functioning of the oil layer. For the before last PG/Isopar mixture 10 litres of water was added only once, because the quantities of PG in the initial mixture were much higher (double dose) so that the bucket was already full with the first addition of water. For the last PG/Isopar mixture, 20 litres of water were added at once due to time restraints.

During one of the tests, saucers with the three different types of oil (linseed, Vaseline and Isopar) were put in the cold chamber and subjected to the same temperature steps in order to better distinguish how the quality of the oils is affected by lower temperatures.

3. Results

The first tests were conducted on the PG/linseed oil mixture, both the water droplets and the frost remained captured in the oil layer starting at temperatures of 0.4 °C (Figure 3, Appendix A). Even after some time, the frost remained on the oil and would not traverse. At temperatures of -15 °C the water droplets also remained stuck on the oil surface, forming puddles of water. The oil seemed more viscous and dense. After letting the bucket contents warm up to ambient temperatures, the frost and water would again pass through the oil. In the "empty" bucket situation, the mixture did not freeze. However, for the bucket half full and full situations the bucket content became "slushy" or even frozen (Figure 4 and Figure 5, Appendix A) (layer) at chamber temperatures of -15 °C to -20 °C (Table 1).

Mixture	**Bucket status**	**Slush temperature (bucket / cold chamber)**	**Freezing temperature (bucket /cold chamber)**
PG	Empty	-	-
PG	Half full (+10 L)	17.7 °C / -20 °C	-
PG	Full (+20 L)	-9 °C / -15 °C	-16.32 °C / -20 °C
Calcium Chloride	Empty	-	-
Calcium Chloride	Half full (+10 L)	-	-4.08 °C / -15 °C
Calcium Chloride	Full (+20 L)	-	-3.6 °C / -5 °C
PG+	Empty	-	-
PG+	Full (+10 L)	-	-10.8 °C / -10 °C [1]
PG/Methanol/Isopar	Full (+20 L)	-	-

Table 1: freezing temperatures for different types of anti-freeze mixtures and dilutions.

The calcium chloride mixture was prepared by mixing 8 kg of solid calcium chloride in 8 Litres of water. The mixture was stirred during several minutes and the calcium chloride was completely dissolved in the water. The temperature of the bucket then rose to 55 °C as an exothermic reaction took place. However, after one night in the cold chamber at 0 °C, solid salt crystals had deposited on the bottom of the bucket (Figure 6, Appendix A). The bucket was once again left at ambient temperatures, and after a day, the salt crystals had disappeared from the bottom of the bucket. At -10 °C however, salt crystals were once again

[1] A small layer of ice had already formed after six days and nights (app. 144 hours) in the cold chamber at -5 °C.

deposited on the bottom of the Pluvio bucket. Where the quality of the Vaseline is concerned, the frost has difficulty (takes a few hours) going through the Vaseline layer starting at bucket temperatures of -2.2 °C and chamber temperatures of -15 °C. At bucket temperatures of -15.6 °C (cold chamber at -15 °C) water droplets seem to have more difficulty passing through the layer and the frost remains deposited on the layer (Figure 7, Appendix A). The Vaseline has also become so viscous that it can easily be pushed to the sides, leaving part of the surface of the anti-freeze uncovered (Figure 8, Appendix A). Where freezing of the bucket contents is concerned, this seems to happen more slowly than for the PG mixtures. For the half full bucket, no slush formation was observed, though a small layer of ice could be observed at the bucket temperature of 0 °C and the chamber temperature of -15 °C, which then thickened and became a solid layer of ice over the weekend. For the full bucket, the cold chamber was left for a much longer period at -5 °C. As a result the bucket content transformed into a layer of ice in 24 hours at this temperature (bucket temperature at -3.6 °C).

A final remark concerning the Calcium Chloride mixture, is that due to its high density, the Pluvio bucket becomes very heavy so that for emptying the buckets on the terrain this mixture is very impractical.

Name	Use	Advantages	Disadvantages
Linseed oil	Was used during the SPICE 2012-2013 season	Eco-friendly	- High density at low temperatures such that precipitation cannot traverse oil layer. - Formation of impermeable layer. - Layer does not uniformly cover the surface.
Vaseline	Used in totalizer rain gauges on Swiss meteorological network.	Familiar procedure, products are present.	- High density/viscosity at low temperatures, though water droplets still seem to pass through. - Layer does not uniformly cover the surface. - Lengthy on-site preparation. - Heavy.
Isopar M	Recommended by J. Hoover.	- Maintains low density at low temperatures. - Water and frost traverse easily.	- No known supplier in Europe. - Not eco-friendly, and some small safety precautions.

Table 2: overview results for different types of oil.

The second Propylene Glycol mixture was prepared using a higher concentration of PG and with the Isopar M oil on top. During the experiment three saucers with linseed oil, Vaseline and Isopar M were also put in the cold chamber in order to have a better view of the effects of cold temperatures on the oil types (Figure 9, Appendix A). Overall, the linseed oil became very dense and viscous, water droplets as well as frost traversed the oil difficultly and

remained mostly captured in the oil or remained on top of it (Figure 10, Appendix A). The Vaseline was less dense than the linseed oil, typically, water droplets would traverse the Vaseline in three steps: first they remained on the surface, then they would agglomerate to form a larger water bubble and finally this larger bubble would have enough mass to pass through the oil layer (Figure 10, Appendix A). The Isopar M oil remained liquid for all temperatures and both water and frost passed through rapidly (Figure 12, Appendix A). This was also the case for the Isopar oil in the Pluvio bucket, though sometimes it seemed to become “misty” or traces could be observed when stirring the liquid with the probe. An overview of the advantages and disadvantages of different types of oils is given in Table 2.

The Anti-freeze mixture did not form any slush or ice during the empty bucket run. For the run with an additional 10 Litres of water in the bucket (bucket full), the temperature of the bucket contents was descended to -5 °C by leaving the cold chamber at this temperature for a long time instead of setting the cold chamber at a very low temperature in order to descend the bucket contents’ temperature more rapidly. A small layer of ice started forming after six days and nights at -5 °C. A thick layer of ice had formed after approximately another five days and four nights at -10 °C.

The last mixture that was tested contained 40% of Propylene Glycol which was mixed with 60% of Methanol, the Isopar M oil was then added to the bucket. Due to time restraints 20 Litres of water were added directly such as to test the highest dilution factor. The cold chamber temperature was set to -25 °C. Though the Isopar layer became “misty” in colour, both the oil and the bucket contents were still liquid in consistency after five days and nights at this temperature.

4. Discussion

The purpose of the adding initial charge of anti-freeze mixtures to gauges is to prevent the formation of slush and ice which could lead to errors in the measurement or problems when emptying the buckets. The purpose of the oil cover is to prevent the evaporation of the volatile anti-freeze mixtures as well as the precipitation in the bucket (especially for gauges which do not compensate for this). Different oil and anti-freeze mixtures were evaluated in this context, as well as for their practical use on the terrain.

Propylene Glycol (PG) is currently used in the Swiss meteorological network. Since the OTT Pluvio2 gauges in the network compensate for evaporation losses, no oil is added to this mixture. Propylene Glycol is used as an additive in food and is thus considered harmless enough to be disposed of on the terrain. This is especially advantageous for stations which are difficult to access as it means that the bucket contents (for the Pluvio2 approximately 30 Litres) do not have to be transported for elimination. However, Propylene Glycol is denser than water and as a consequence the precipitation which falls into the bucket does not mix well with the anti-freeze. The result, as was observed in the cold chamber experiment, is

that the water forms a layer on top of the PG and becomes a frozen layer. This problem was first addressed by McSaveney (1979) who advised mixing the PG with Methanol in order to reduce its density. It should be emphasized that during the experiment in the cold chamber the 10 Litres of water were added at once (at ambient temperatures), whereas in practice precipitation would fall in smaller quantities over a much longer time period, thus perhaps permitting better mixing with the anti-freeze. Also, the problem of frozen bucket contents on the terrain has been reported only once until now, though the formation of slush seems to occur more often. Admittedly, the gauges are not emptied often in the middle of the winter season; as such, the contents may already have melted by the time maintenance occurs.

The Calcium Chloride mixture is used in the Swiss meteorological network in the totalizer rain gauges. During the experiments in the cold chamber salt crystals regularly deposited on the bottom of the bucket when temperatures were lowered to around 0 °C. This deposition of salt, leading to lower concentrations of salt in the anti-freeze mixture, could also explain why the bucket contents froze at lower temperatures. It is unclear whether the deposition of salt was due to inadequate prior mixing of the salt water mixture or whether it was due to the fact that water becomes saturated for this quantity of salt at lower temperatures. For our experiment, the mixture was stirred during several minutes and this seemed to suffice for ambient temperatures as the salt was completely dissolved. MétéoSwiss technicians have indicated that they usually mix for a much longer time (approximately 15 minutes). On the other hand, the totalizer rain gauges are specifically used for areas which are difficult to access and are emptied once a year and not in wintertime. As such, if salt crystals are formed over the winter, these may already have been dissolved by the time the gauge is emptied. Even so, there is one report of the formation of salt crystals on the terrain. Finally, it should also be added that the totalizer rain gauges are more solid in structure and do not have to be lifted for emptying. Due to its high density, the salt/water mixture is much heavier than other types of anti-freeze which makes its use in other rain gauges, which do sometimes have to be lifted, much more impractical.

As mentioned earlier, McSaveney (1979) already recommended mixing Propylene Glycol with Methanol in order to diminish its density so that the anti-freeze mixes more efficiently with water. This was also recommended by J. Hoover (2014) as a result of his research in the context of SPICE. During the experiments in the cold chamber this mixture remained liquid in consistency at the highest dilution and while being exposed to a long period of very low temperatures. As such, this is the anti-freeze mixture which responds best to the requirements. However, Methanol is more volatile than PG and even though the Methanol is mixed with PG this anti-freeze mixture probably requires an oil layer to prevent evaporation of the Methanol (whereas this is not necessarily required for a pure PG mixture). Furthermore, Methanol is more hazardous than PG, requiring more strict safety precautions for those handling it, as well as that the mixture has to be recovered and transported for professional chemical elimination. Finally, the addition of Methanol makes the anti-freeze

mixture more expensive. The advantages, disadvantages and the approximate relative costs of the different combinations of anti-freeze and oil are summarised in Table 3.

Mixture	Advantages	Disadvantages	Relative price per Pluvio (1500 mm)[2]
PG/Linseed oil	- relatively cheap - practical on the terrain - Eco-friendly	- Formation of ice blocks and slush - formation of impermeable layer	100 %
Calcium Chloride / Vaseline	- cheap - low maintenance	- impractical on terrain due to weight - formation of ice - formation of impermeable layer - not eco-friendly	8 %
PG+/Isopar M	- Isopar M remains liquid at low temperatures	- freezing issues are not resolved - relatively expensive - emptying of gauges will be required more often - Isopar is less eco-friendly	226 %
PG/Methanol/Isopar M	- allows for precise measurements - mixture remains liquid at low temperatures	- expensive - mixture must be retrieved when emptying gauges - mixture is not eco-friendly	325 %

Table 3: the "price to pay" for different anti-freeze and oil combinations.

5. Conclusion

The purpose of this analysis was to assess the quality of different mixtures of anti-freeze and oil at low temperatures. The choice of a mixture has a direct influence on the measurement precision which is particularly important in the context of SPICE. It was found that the Propylene Glycol anti-freeze and linseed oil mixture, though being the most practical and eco-friendly option, cannot prevent the formation of ice and slush in the bucket. This is due to the high density of the PG and cannot be resolved with a higher concentration of anti-freeze. What is more, the linseed became too dense at low temperatures and degraded over the course of the season to form an impermeable layer. The Calcium Chloride Vaseline mixture had the same problems as the PG/linseed mixture and was evaluated too impractical for use in non-totalizer precipitation gauges. The mixture which responded to the requirements is a Propylene Glycol/Methanol mixture covered by Isopar M oil. It has

[2] Prices are approximate calculations.

been decided that this mixture will be used in the precipitation gauges for the upcoming SPICE winter season.

6. References

Hoover, J. et al. (2014) SPICE-5 Sodaniklä meeting, Antifreeze and oil assessment. *J:\pay\CIMO\SPICE\SPICE_Problems\5-3(3)_SPICE-5_S3C_Antifreeze&Oil.pdf.*

Hoover, J. et al. (2014) Precipitation gauge oil quantity for SWX/RCS Network. *J:\pay\CIMO\SPICE\SPICE_Problems\ EC Isopar M Precipitation Gauge Oil Quantity Recommendations June 2 2014.pdf.*

McSaveney, M.J.(1979) An effective antifreeze for storage raingauges. *J:\pay\CIMO\SPICE\SPICE_Problems\ JoHNZ_1979_v18_1_McSaveney.pdf* .

7. Appendix A

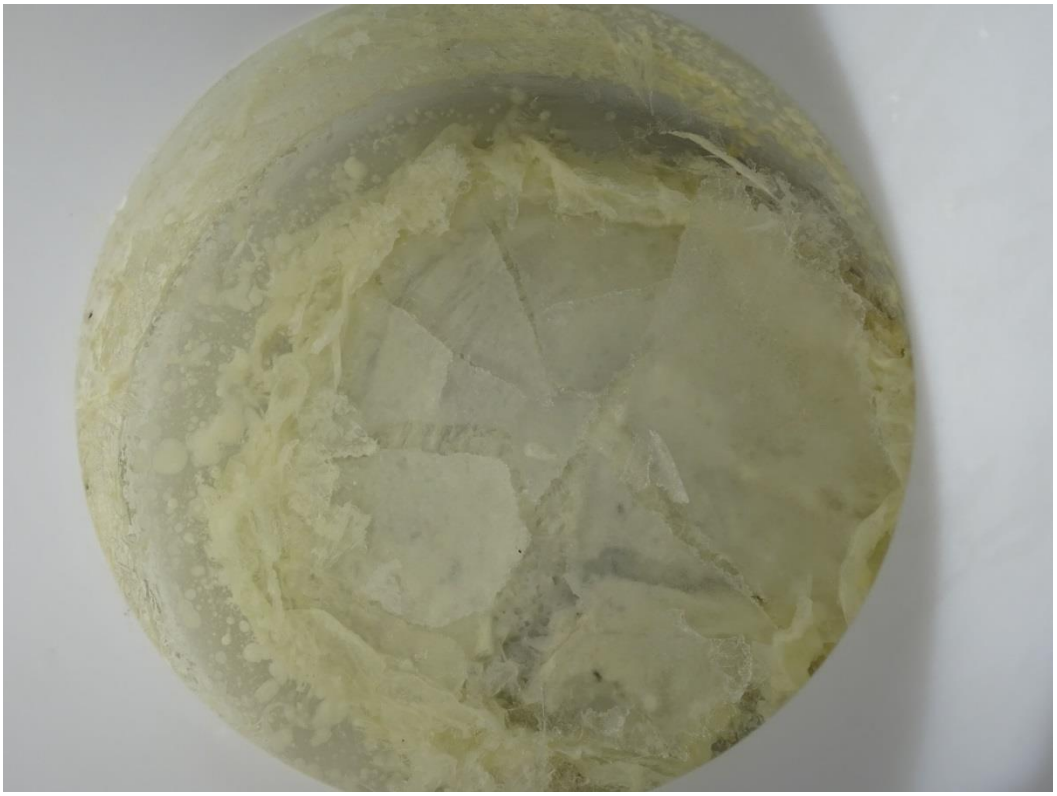


Figure 2: bucket contents after a summer on the Swiss SPICE measurement site at the Weissfluhjoch, Davos.

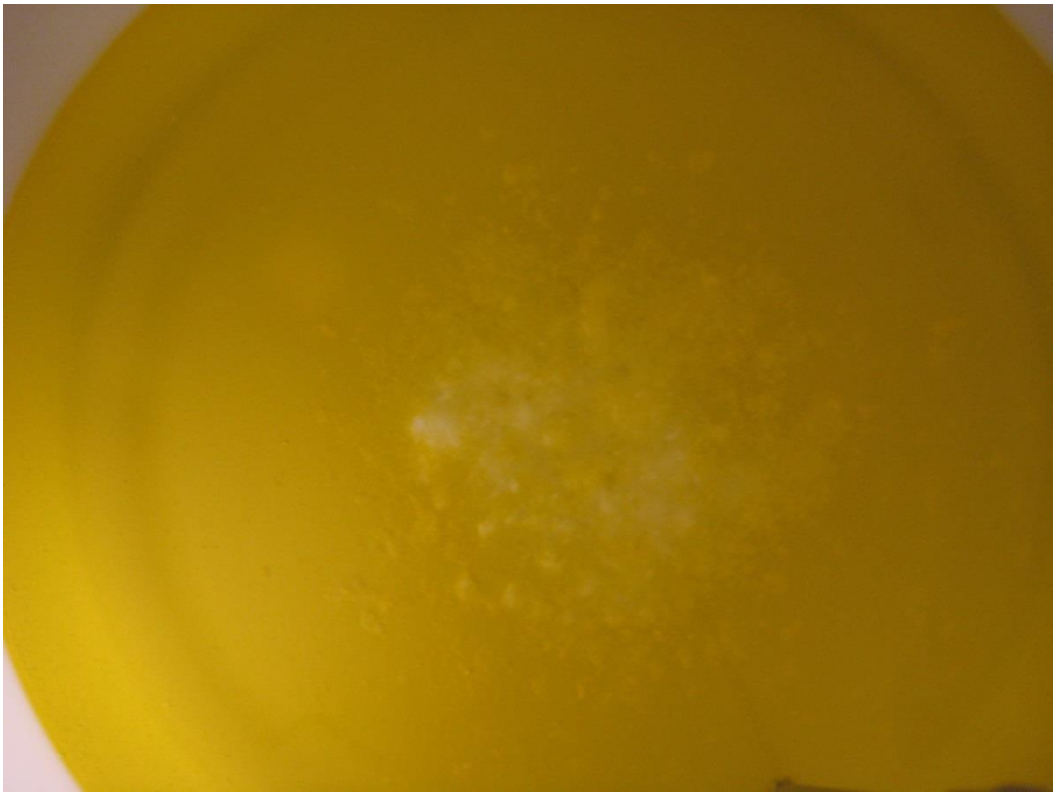


Figure 3: frost and water droplets remain on the surface of the linseed oil.

Figure 4: frozen bucket contents.

Figure 5: Frozen layer with liquid layer underneath.

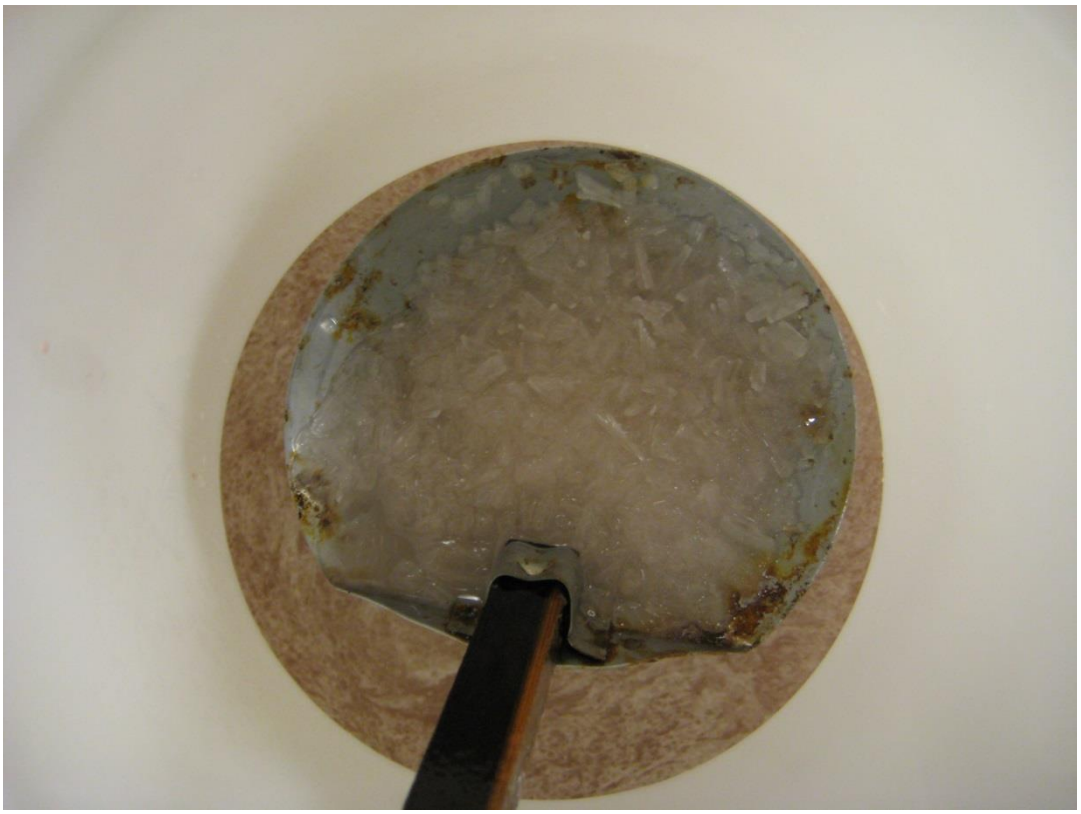


Figure 6: solid salt crystals on the bottom of the bucket.

Figure 7: frost remains on the Vaseline layer.

Figure 8: Vaseline layer is easily displaced.

Figure 9: saucers with from left to right: linseed oil, Vaseline and Isopar M.

Figure 10: water and frost captured in linseed oil.

Figure 11: water droplets agglomerate in Vaseline.

Figure 12: frost passing through Isopar layer (left), and frost has disappeared in anti-freeze (right).

8. Appendix B

Date	Heure	Tbucket (°C)	Tset (°C)	Comments
03.06.2014	13.15		0	Start cold chamber with 7.5 L of Propylene Glycol, 2.5 L water, 0.4 L of linseed oil.
	13.46		-5.0	No observations
	15.45		-7.0	No observations *(photo 4032)*
	16.20		-9.0	No observations *(photo 4033)*
	16.53		-11.0	No observations *(photo 4034)*
	17.25		-13.0	No observations *(photo 4035)*
	17.43		-15.0	No observations *(photo 4036)*
04.06.2014	8.45		-17.0	Pluvio spent the entire night in the cold chamber at -15°C. Small air bubbles have appeared in the oil and the oil seems more viscous? *(photo 4037)*. Test with water droplets: the water droplets seem to remain on the surface of the oil without passing through *(photo 4038, 4039)*.
	10.44		-19.0	Water droplets remain frozen on top of the oil layer *(photo 4041)*.
	11.29		-19.0	Water droplets are frozen. Oil was tested and is still liquid in consistency.
	15.40		5.0	Pluvio is warmed up to see whether the water will pass through the oil.
		-0.3	10.0	
			12.0	Because bucket content warms up very slowly the temperature was set to be higher.

Date	Heure	Tbucket (°C)	Tset (°C)	Comments
	18.15	5		The Pluvio was left at ambient temperature for the night with the objective of letting the water droplets pass through the oil.
05.06.2014	08.21	17.5	Ambient	Still small water and air bubbles in the oil; oil was mixed to make these disappear *(photo 4043)*.
	08.55	17.5	Ambient	Water droplets test: big droplets (accumulation of smaller droplets at the same spot) disappear almost instantaneously, small droplets remain on the surface or slightly within the oil *(photo 4044, 4045, 4046)*.
	09.50-10.00	17.5		Water droplets test was repeated, test with frost from freezer: both pass through the oil *(video1)*.
	10.28	17.5	0.0	Surface of the oil has become smooth again *(photo 4048).*
	11.45	12.0	-10.0	To make the temperature of the bucket content diminish more rapidly.
	12.36	7.7	-15.0	
	14.00	-0.4		Larger bubbles pass through the oil, smaller remain on top/within the oil *(photo 4049, 4050) (video 2)*. Frost does not melt, remains within the oil (not on surface). The bucket content still seems liquid in consistency.
	15.25-15.30	-5.2		Larger bubbles pass through the oil but remain on top of the oil *(photo 4052, 4053)*. The frost remains on top of the oil *(video 3)*.
	15.40	-5.0		The frost remains captured within the oil without passing through it or melting *(4055, 4056)*. The content seems more viscous or dense.
	16.25	-7.0		An additional thermometer was used to check whether the temperature probe in the Pluvio bucket was not influenced by the air temperature. The thermometer indicated -6 °C and the probe in the bucket indicated -6.9 °C.
	18.14	-9.6		A large quantity of frost was poured into the bucket: part of it seems to remain captured in the oil and part of it (white in the picture) seems to remain on the surface *(photo 4057 4058, 4059)*. The bucket content seems to be more dense.
06.06.2014	9.30-9.45	-15.5	-30.0	The frost from the test performed the previous day at -10 °C still seems captured within the oil and is still in a solid form *(photo 4064, 4065)*. Water droplets remain on the oil surface, even the bigger collections or puddles of droplets. A large quantity of frost was poured into the bucket but did not pass through the oil. The oil seems somewhat viscous (no real change in density but oil droplets remains suspended on a finger or probe for a long time without falling).

Date	Heure	Tbucket (°C)	Tset (°C)	Comments
	10.30	-20		Frost remains on top of the oil and forms small heaps of frost *(photo 4066, 4067)*. Oil seems more viscous.
	12.28	-24.5		Frost drifts in small blocks of ice on the surface of the oil. The oil is more viscous and denser. In order to make the frost melt, the temperature of the cold chamber was set to 30 °C.
	15.55	7.7		Frost melted.
				END
06.06.2014	18.25			Start program for test cases, bucket "empty."
10.06.2014	18.40			Start program for test cases, bucket "empty."
13.06.2014	18.08			Start program for test cases, bucket "half full."
16.06.2014	13.46	14.7	-15.0	Content Pluvio : 1229.14 mm (10 liters have been added to simulate the situation "bucket half full"). Surface : small bubbles after having mixed the oil. After the previous test, there were "holes" in the oil layer (parts where the anti-freeze had been uncovered) *(photo 4068, 4069).*
	13.49/50		-15.0	Accessed cold chamber.
	15.37/38	8.8	-30.0	Accessed cold chamber.
	16.48	1.4		
	17.19	-1.3		Frost passes through the oil *(photo 4071),* the oil seems more viscous and dense. No openings in the oil.
	18.01	-4.4	-10.0	Frost seems to stay on the surface of the oil, creating small white spots *(photo 4072)*. Temperature set at -10.0 °C for the night.
17.06.2014	08.07	-9.8		Frost stays on the surface of the oil and forms large white spots *(photo 4073)*. The oil seems to be more viscous. A darker trace of oil sticks to the thermometer probe *(photo 4074, 4075)*. Machine was stopped for a visit.
	13.29	-5.4	-30	Temperature set to diminish the temperature of the bucket content to -15 °C.
	17.34	-14	-20	The properties of the oil surface have changed *(photo 4076, 4077, 4078, 4079, 4080)*.

Date	Heure	Tbucket (°C)	Tset (°C)	Comments
18.06.2014	08.24	-17.7	-20	The contents of the bucket have transformed into a block of slush, three different layers can be distinguished *(4081, 4083)*.
	08.54 – 10.55	-17.7	-20	Regular visits to the cold chamber.
	10.55	-17.7	Ambient	The bucket is exposed to ambient temperature in order to melt its contents.
	11.43	-14.4	Ambient	The bucket contents are still slushy.
	13.36	-7.3	Ambient	The bucket contents are a mixture of slush and water *(photo 4085)*.
	14.56	-5.7	Ambient	Slush at the interior of the bucket and liquid on the sides.
	17.16	-1.3	Ambient	Small icebergs of slush are floating in the otherwise liquid content. The icebergs stick out of the oil surface *(photo 4086)*.
19.06.2014	08.45	14	Ambient	The bucket content is back to its initial state, all ice has melted.
	09.21			Start program for test case, scale without bucket.
20.06.2014	13.00			End program for test case, scale without a bucket. Bucket put back on scale.
20.06.2014	13.10			Start program for test cases, bucket “half full.” Observation of small “holes” in the oil layer (parts where the anti-freeze had been uncovered) *(photo 4092)*.
24.06.2014	09.00 – 09.15	Ambient	Ambient	10 liters of water have been added to the Pluvio bucket to simulate the situation “bucket full” and the oil has been left to reconstitute itself.
	13.06	14	-30	Small bubbles of air in the oil *(photo 4093)*. The oil layer is easily discernible from the exterior of the bucket *(photo 4094)*.
	15.53	0.6		The position of thermometer in the bucket was changed since it gave a wrong reading of the temperature of the bucket contents.
	18.33/43	-8.3	-10	An impermeable layer has formed under the oil layer. Perhaps this is the start of the freezing process. No ice or slush can be observed yet, but the layer seems dense and cohesive. The layer can be broken when turning the bucket contents with a probe *(photo 4095 and 4096)*. Temperature was set for the night.

Date	Heure	Tbucket (°C)	Tset (°C)	Comments
25.06.2014	07.58	-9	-15	A larger layer of slush can be observed.
	15.45			Visit with the visitors from OTT, disturbance of the Pluvio.
	16.55	-11.4		Very dense slush, the oil layer seems to have frozen.
	18.21	-14.6	-20	A very thick ice layer makes it difficult to move a probe in the bucket. Underneath this layer is a layer with slush and liquid. The oil layer has frozen *(photo 4100)*. The temperature was set for the night.
26.06.2014	08.24	-16.32	-20	The bucket has spent the night at -20 °C, the content has been transformed into a block of ice : perhaps underneath this layer there is still a layer of non-frozen anti-freeze but the frozen layer on top is too solid to verify this *(photo 4101 and 4102)*.
	09.33	-16.4	-30	*(photo 4103).*
	13.29	-20.4		*(photo 4104).*
	16.45	-22.5		*(photo 4105)*.
	18.29	-23		*(photo 4106)*.
				END
27.06.2014	18.00			Start program for test cases, bucket “full.” There is still a block of ice in the bucket.
02.07.2014	10.38			Observation of “holes” in the oil layer (parts where the anti-freeze had been uncovered) *(photo 4107)*.
04.07.2014	16.14			Start program for test cases, bucket “full.”
07.07.2014				Liquid-like stains in the oil layer which deform when stirred *(photo 4109)*.
10.07.2014	15.00 – 15.30			The bucket was filled with 8 L of water, 8 Kg of salt (calcium chloride) and 5 dl of Vaseline, which is the mixture used within the MeteoSwiss network for totalizers. The mixture of water and salt was stirred for about 5 minutes so that the salt was completely dissolved. Due to the exothermic reaction, the contents warmed up to 55 °C. The contents were left to cool down for a while *(photo 4111, 4111(a), 4111(b), 4112 4113(a), 4113(b), 4113(c), 4113)*.
	± 18.00		0	Temperature is set for the night.

Date	Heure	Tbucket (°C)	Tset (°C)	Comments
11.07.2014	15.21	3.6	Ambient	The bucket spent the night in the cold chamber at 0 °C. The salt has deposited on the bottom of the bucket forming a solid block of salt crystals. It was attempted to detach and stir the contents again, without any results. It was decided to let the contents warm up in order to try and dissolve these at a higher temperature *(photo 4114)*.
14.07.2014	08.57	20.36		The Pluvio has spent the weekend in ambient temperatures. The ice crystals have been completely dissolved. Perhaps the Vaseline has been mixed somewhat with the other layers *(photo 4115)*.
	09.02	20.36	-10	
	13.28	3.6	-15	There are once again salt crystals on the bottom of the bucket.
	15.23	1.5	-15	
	17.29	-2.2	-15	An important, thick layer of salt on the bottom of the bucket. Water droplets go through the Vaseline layer, frost seems to stay floating in the layer. The bucket contents are still liquid in consistency.
	18.00		-5	Temperature set for the night.
15.07.2014	08.09	-4.8	-15	Water droplets go through the Vaseline layer, frost seems to stay a bit on the surface or traverses slowly.
	11.16 – 12.00	-10.6	-15	The frost of 08.09 hours has gone through the Vaseline layer, the layer seems somewhat milky white and more viscous *(photo 4116(a)))*. Water droplets go through the layer of Vaseline, frost seems to stay on the surface somewhat *(photo 4116(b), 4116(c))*.
	13.30		-25	
	15.47	-15.6	-15	The layer of Vaseline is dense and viscous and can be pushed away to the sides. Water droplets do not seem to go through the layer of Vaseline. Temperature was set for the night.
16.07.2014	08.52	-15.3	-15	The layer of Vaseline is viscous but still covers the entire surface. It reconstitutes itself easily (after having been pushed to the sides it tends to replace itself such as to cover the entire surface again) *(photo 4116, 4117)*. Water droplets go through the layer *(photo 4117)*, frost does not immediately go through *(4118).*
	10.29	-15.3	Ambient	The frost still has not passed through the layer of Vaseline *(photo 4119, 4121, 4122)*.
	13.19	-3.7	Ambient	The frost has gone through the layer of Vaseline and the colour of the layer is starting to change *(photo 4123)*.

Date	Heure	Tbucket (°C)	Tset (°C)	Comments
17.07.2014	09.32	17.2	Ambient	The layer of Vaseline is more transparent and liquid in consistency *(photo 4124)*. There are still salt crystals on the bottom of the bucket.
	09.32	17.2	Ambient	10 L of water have been added to the bucket in order to simulate the case bucket " half full."
	13.26	19.9	-15	The layer of Vaseline has become transparent, *(photo 4125)* there are still salt crystals on the bottom of the bucket.
18.07.2014	09.08	0.7	-15	The bucket content is liquid in consistency, the layer of salt crystals on the bottom is hard *(photo 4126)*.
	09.08			Started logging data
	12:00	0	-15	Small layer of ice has formed under the layer of Vaseline *(photo 4127)*.
	14.06	-4.08	-15	The layer of ice becomes thicker, the bucket content seems to be cooling down more slowly than during previous tests.
21.07.2014	09.17	-10.2		The Pluvio has spent the weekend in the cold chamber at -15 °C, a thick layer of solid ice has formed.
	12.15 – 12.30		-15	Frequent visits to Cold chamber.
	12.30		Ambient	Exposure of Pluvio to ambient temperatures in order to melt the ice in the bucket.
22.07.2014	09.19	11.45	Ambient	The Pluvio has spent the night at ambient temperatures. There is still a thick layer of salt crystals on the bottom of the bucket.
	09.40	11.14	-5	10 L of water have been added to the Pluvio bucket in order to simulate the case bucket "full."
	15.23	6.4		The bucket contents are liquid however the layer of Vaseline may have become a bit more dense.
23.07.2014	09.17	-2.5	-5	The Pluvio has spent the night at -5 °C. A layer of ice has formed which is still easy to break through *(photo 4129)*. The layer of Vaseline is still liquid in consistency. There is also still an important layer of liquid under the ice layer.
	13.24	-2.9	-5	The same *(photo 4130)*.
	17.54	-3.1	-5	Same, but the ice layer has become thicker *(photo 4131, 4132)*.

Date	Heure	Tbucket (°C)	Tset (°C)	Comments
24.07.2014	09.16	-3.6	-5	The Pluvio has spent the night in the cold chamber. There is now a thick layer of ice in the bucket and it is now impossible to pull the thermometer probe out of the ice. The ice layer around the thermometer moves up when trying to pull the probe out of the bucket *(photo 4133)*.
25.05.2014		-3.9	Ambient	
28.07.2014	09.59	20.8		The Pluvio has been left at ambient temperatures for the weekend. There seem to be less salt crystals on the bottom and the layer seems less solid. The layer of Vaseline has again changed in quality, it now seems that there are small "bags" of oxygen caught in the layer *(photo 4134)*.
30.07.2014	11.12	Ambient	-5	Start cold chamber with 12,6 L of Propylene Glycol, 4,2 L of water and 800 mL of Isopar M oil. Additionally, three saucers with a layer of linseed oil, Vaseline and Isopar M oil have been put in the cold chamber *(photo 4135, 4136, 4137, 4138)*.
	18.25		0	Temperature set for the night
31.07.2014	08.21 – 08.34	-0.2	-5	Pluvio spent the night at 0 °C, no changes can be observed in the bucket content *(photo4139)*. For the saucer with linseed oil, the water droplets remain somewhat captured in the oil layer *(photo 4140, 4144)*, for the saucer with Vaseline, the water droplets remain first captured in the Vaseline layer and then proceed to the bottom of the saucer *(photo 4141, 4145)*, for the saucer with Isopar M oil, the water droplets immediately go through the layer to the bottom of the saucer *(photo 4142, 4146; the Isopar layer on the side is too thin so that the water droplets still point out of the surface)*. The water droplets immediately disappear in the Pluvio bucket *(photo 4143)*. The frost remains somewhat caught in the linseed oil *(photo 4148, 4151)*, it traverses gradually the Vaseline *(photo 4149, 4152)*, and it goes through the Isopar oil *(photo 4150, 4135)*. This is also the case for the Pluvio *(photo 4147, 4154)*.
06.08.2014	09.37 – 09.50	-5.3	-10	Pluvio spent several days and nights at -5 °C, the bucket content is still liquid in consistency *(photo 4155)*. The water droplets from the previous tests seem to have frozen on the bottoms of the saucers *(photo 4156, 4157, 4158)*. For the Pluvio, the water droplets traverse almost immediately the Isopar layer *(photo 4159)*. For the linseed oil, the droplets remain in suspension in the oil *(photo 4160)*, for the Vaseline, the water droplets traverse the layer and spread on the bottom of the saucer *(photo 4161)*, for the Isopar oil, the water traverses the surface and remains on the bottom of the saucer *(photo 4162)*. For the linseed oil, the frost seems to traverse / remain somewhat in suspension in the oil *(photo 4163, 4167)*, for the Vaseline, the frost traverses the surface partially (quantity too important?) *(photo 4164, 4168)*, for the Isopar the frost also traverses partially due to the important quantities *(photo 4165, 4169)*. In the Pluvio, the frost traverses completely and melts rapidly in the anti-freeze *(photo 4166, 4170, 4171)*.

Date	Heure	Tbucket (°C)	Tset (°C)	Comments
12.08.2014	14.58 – 15.11	-9.8	-15	Pluvio spent several days and nights at -10 °C, the bucket content still seems liquid in consistency, though a trace can be observed in the Isopar for where the probe has been moved around. For the saucer with linseed oil, the water droplets seem to remain on the surface of the oil *(photo 4172)*, the oil seems very viscous and dense. For the Vaseline, the droplets remain on the surface at first and then traverse the layer rapidly *(photo 4174).* For the Isopar, water droplets traverse the layer at once *(photo 4175)*, and in the Pluvio the droplets traverse the Isopar layer instantaneously as well. The frost seems to remain on the surface of the linseed oil *(photo 4176)*, it is caught in the layer of Vaseline *(photo 4177)* and it traverses the layer of Isopar *(photo 4178)*. In the Pluvio as well, the frost traverses the layer of Isopar, remains between this layer and the anti-freeze for a while, and then slowly disappears in the anti-freeze / water mixture *(photo 4179)*.
18.08.2014	08.27	-8.2	-20	Pluvio spent several days and nights at -15 °C, the bucket contents still seem liquid in consistency. For the linseed oil, the water droplets remain on the surface *(photo 4180, 4184)*. For the Vaseline, the water droplets remain on the surface for a few seconds, then aggregate and pass through the layer *(photo 4181, 4185)*. For the Isopar, the water droplets pass through the oil layer immediately *(photo 4182, 4186 – droplets can be observed on the side of the saucer because the Isopar layer is too shallow here)*. For the Pluvio, this happened too fast to make a picture with the water droplets *(photo 4183)*. The frost remains on the surface of the linseed oil and is somewhat absorbed into the layer *(photo 4187)*. For the Vaseline, the frost seems to remain caught in the layer *(photo 4188)*. For the Isopar, the frost traverses the layer, though the quantity of solid material present in the saucer prevents some of it from passing through *(photo 4189)*. The phenomenon can be better observed in the Pluvio; the frost traverses the Isopar layer and disappears as it reaches the antifreeze *(photo 4190, 4191)*.
25.08.2014	11.54 – 12.30	-4.5 to -2.7	Ambient	The cold chamber has probably malfunctioned over the last period as the temperature within the bucket seems to have increased instead of decreased. Also, the interior of the cold chamber did not seem as cold as it should have been. The linseed oil is no longer as viscous and dense in consistency as it was during the other tests. The water droplets still stay on the surface of the oil *(photo 4232)*. For the Vaseline and the Isopar oil the results were the same as for the previous tests: the droplets remain on the surface at first, then traverse for the Vaseline and traverse directly for the Isopar *(photo 4233, 4234, 4235)*. Where the frost is concerned, it remains stuck within the linseed oil *(photo 4236, 4241)*. For the Vaseline and Isopar the effects are hard to evaluate due to the quantity of frozen water and frost already present in the saucers *(photo 4237 and 4242, 4238 and 4243),* however, the frost traverses the Isopar in the Pluvio bucket and disappears in the anti-freeze *(photo 4239, 4240)*. The results of this experiment have to be put into the context of a cold chamber temperature well above -20 °C. 10 Liters of water have been added to the Pluvio bucket which is now almost full.

Date	Heure	Tbucket (°C)	Tset (°C)	Comments
26.08.2014	08.35	20.2	-5	A thermometer has been added to the cold chamber in order to check its temperature.
	14.20	3.2		The interior of the cold chamber is indeed at -5 °C.
28.08.2014	08.59	-5.9		The interior of the cold chamber is still at -5.5 °C.
01.09.2014	08.30 – 09.07	-5.6	-10	The pluvio has spent several days and nights at -5 °C. The thermometer in interior of the cold chamber indicates -4.5 °C, which means that the cold chamber has managed to stay at this temperature. A small ice layer has formed on the top of the water layer, under the Isopar layer *(photo 4244)*. Water droplets pass through directly, though frost passes through the Isopar layer and then remains stuck on the frozen water layer *(photo 4245)*.
	11.46	-7.1		The frost did not pass through or reach the anti-freeze layer. It seems to be stuck on top of the ice layer *(photo 4246)*. The thermometer in interior of the cold chamber indicates -9.5 °C.
02.09.2014	13.30 14.10 14.20	-9.15	-10	The frost did not pass through to the anti-freeze and seems stuck between the Isopar and the ice layer *(photo 4247)*.
05.09.2014	16.35	-10.8	-10	The bucket contents have transformed into a block of ice. The cold chamber has been stopped.
11.09.2014	18.22	31.0	-25	The Pluvio bucket has been filled with 5.3 L of PG, 8.0 L of Methanol and 800 mL of Isopar. 20 liters of water have been added and the temperature was set to -25°C.
12.09.2014	08.44	-19	-25	The bucket contents are still liquid, though under the Isopar layer the anti-freeze has become a misty in colour.
16.09.2014	16:00	-25	-25	The bucket contents have spent several days and nights at -25 °C and are still liquid in consistency.

SPICE Calibration and Configuration Recommendations for the GEONOR Precipitation Gauge

Prepared by Jeffery Hoover. Developed from:
I. Roy Rasmussen's GEONOR site visit summary
II. GEONOR calibration demonstration by Ivar Fredriksen, Tuesday October 16, 2012, Brussels, Belgium
III. GEONOR T-200B user manual

This document concerns the recommendations for the calibration and configuration of the GEONOR reference instrument. It is described through the four following steps.

1. Laboratory calibration of the GEONOR vibrating wire transducers

SPICE Recommendation: The independent calibration of the individual GEONOR vibrating wire transducers is not recommended for site managers.

GEONOR's experience is that the factory calibration of the vibrating wire transducer does not change significantly over time. This is supported by DiBiagio, Rasmussen, and previous studies within the CRN group. Vibrating wire transducers that are found to be out of calibration should be returned to GEONOR for factory recalibration.

2. Empty bucket field calibration of the GEONOR precipitation gauge

SPICE Recommendation: Field calibration of the GEONOR gauge should be performed using an empty bucket as per section 5.4 in the GEONOR user manual. Prior to the field calibration the gauge should be installed and leveled according to the GEONOR user manual. If any of the empty bucket transducer frequencies (f0) differ by more than 10 Hz from the GEONOR calibration values, a new A' coefficient must be calculated using the equation provided in section 5.4. This field calibration is to be completed at the beginning of each SPICE measurement period with the old and new transducer coefficients documented.

3. Field check of the GEONOR precipitation gauge

SPICE Recommendation: A field check shall be performed after the empty bucket field calibration as a check of the gauge function and calibration values (wires hanging correctly, interference with bucket, calibration coefficients etc.). The field check is performed by adding a 1.5 kg of water to the bucket, corresponding to 75 mm of precipitation. Each observed transducer frequency must be within 0.5 % of the calibration frequency for acceptance. If the transducer frequency is outside this range it should be returned to GEONOR for recalibration. This field check is to be completed at the beginning and end of each SPICE measurement period with each transducer frequency recorded for reference. The water measurement procedure and scale features (make, model, resolution, repeatability, linearity, calibration identification, and calibration expiry date) should be recorded as well.

If possible, additional observations at 3.0 kg, 4.5 kg, 6.0 kg, 7.5 kg, 9.0 kg, 10.5 kg, and 12.0 kg masses are recommended for reference.

4. Recommended GEONOR mounting configuration

SPICE Recommendation: In order to minimize noise due to mechanical vibrations on the GEONOR gauge, a configuration with the Alter shield isolated from the gauge is recommended. In this configuration the GEONOR gauge is mounted to the GEONOR pedestal (see section 2 in the GEONOR manual), while the Alter shield is independently mounted to the ground or DFIR. Examples of these Alter shield configurations are shown in Figures 1 and 2 for reference.

Figure 1: Single Alter Shield Mounting in DFIR, R2(G,SA) Reference, CARE, Canada

Figure 2: Single Alter Shield Mounting, R3(G,SA) Reference, CARE, Canada

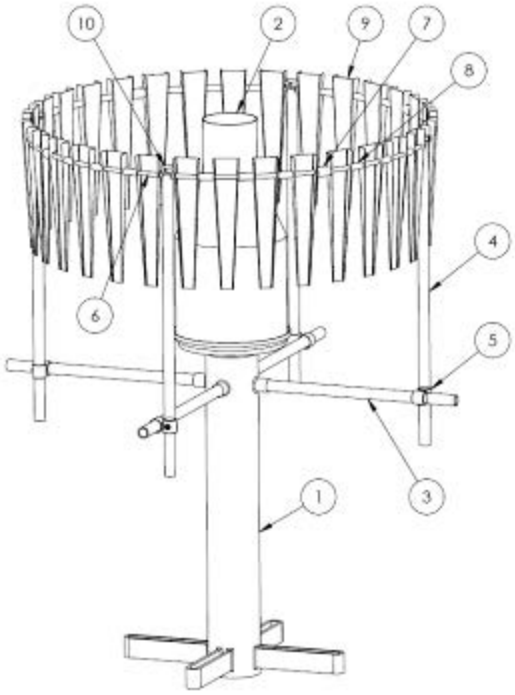
Specifications for the WMO-SPICE Single Alter Shield Configuration

Prepared by Craig Smith, Jeffery Hoover, John Kochendorfer, Rodica Nitu

The single Alter configuration that will be used during the WMO-SPICE is a modified configuration of the design distributed by Geonor, originally designed by the Norwegian Meteorological Institute (Circa 1985).

This shield will be used for the gauges which are part of the reference systems, i.e. the gauge in the Double Fence Intercomparison Reference, the R2 reference, and the gauge in a single Alter, part of the R3 reference.

The single Alter shield configuration will consist of a single ring of "blades", also known as slats or fins, mounted on a ring of 1230 mm diameter (or approximately 4 feet), centered about the gauge. The height of the blades will be positioned at a height of 20mm above the orifice of the gauge.



ITEM	PART NAME	QTY.
1	PEDESTAL	1
2	GEONOR T200B 600MM	1
3	HORIZONTAL SUPPORT BAR	4
4	VERTICAL SUPPORT BAR	4
5	FITTING	4
6	SUPPORT ROD	4
7	COUPLING	4
8	SHIELD SPACER	24
9	SHIELD ELEMENT	32
10	END CAP	4

Figure 1: Alter shield: general view and components

Note: although shown connected to the post of the gauge, it is strongly recommended that the shield is mounted independently of the posts of the gauge.

Blades (shield elements)

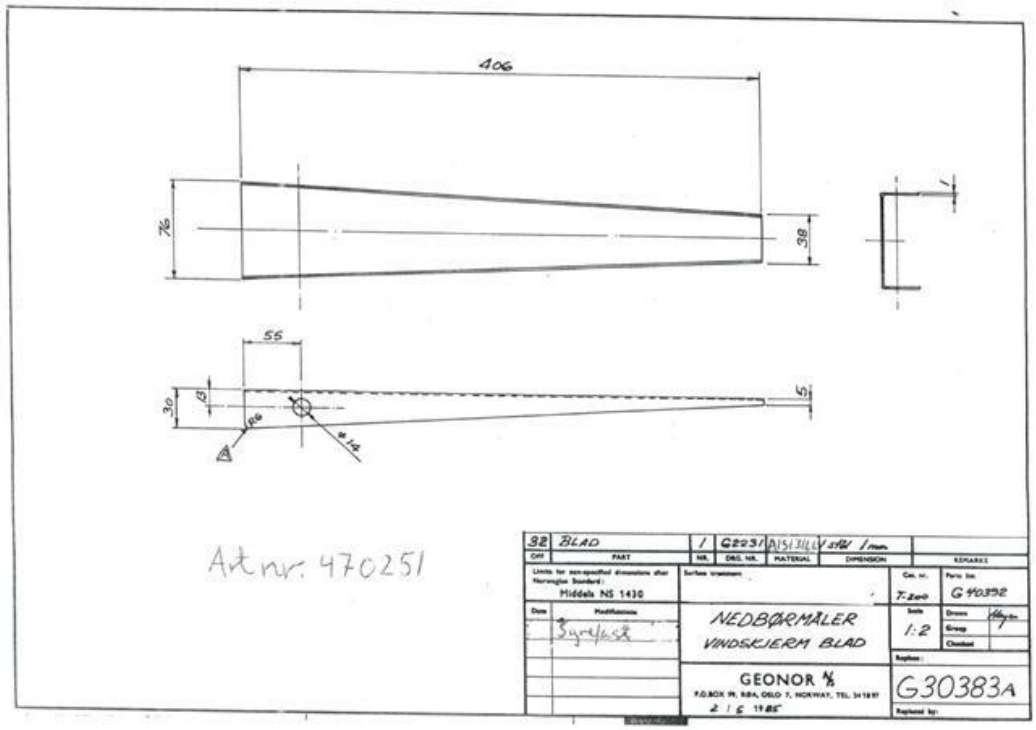


Figure 2: Blade drawing (courtesy of Geonor)

Material: stainless steel AISI 316L, 1 mm thick.

Number of blades/shield: 32

The blades should be able to swing freely on the shield ring.

The blade specifications are shown in Figure 2. A 3-dimensional view of the blade is included in Figure 3. The tapered blade will be 76mm wide at the top and 38mm wide at the bottom with a total length of 406 mm (16 inches). The blade will be boxed (open at the top and bottom) with the open side of the box facing towards the outside of the ring.

The edges of the boxed blade are also tapered from 30 mm at the top and decreasing down to 5mm at the bottom. The blades will be supported by a shield ring passing through a 14mm diameter hole located 55mm from the top of the blade and 13 mm from its inside edge.

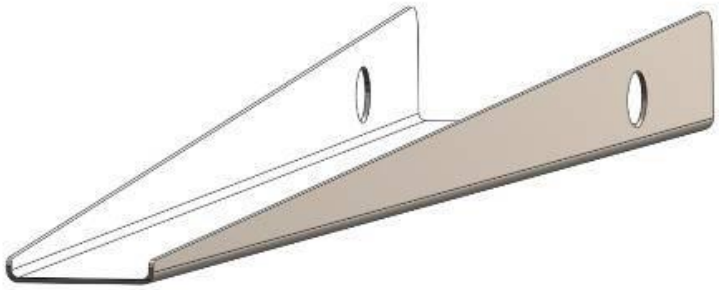


Figure 3: Blade (shield element)

Spacers

The blades are separated by spacers, with a nominal length of 40 mm.

Number of spacers per shield: 24

Note: The Geonor spacer was measured to be 44 mm long with an inner diameter of 16.18 mm and an outer diameter of 20 mm.

Figure 4: Spacer

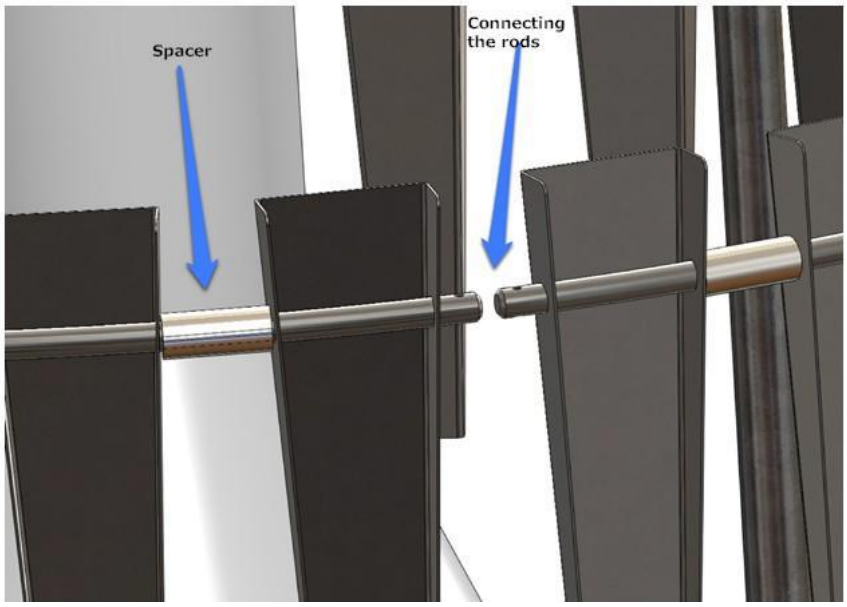


Figure 5: Shield spacers and rod connection: details

Shield Ring

The ring is manufactured from 4 stainless steel rods of 13 mm (1/2”) diameter.

The diameter of the ring, when installed is 1230 mm.

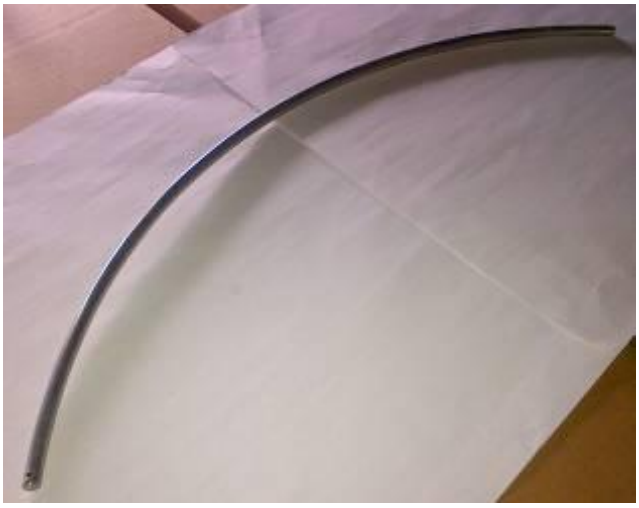


Figure 6: Support rod (1/4 shield ring)

The rods are connected with 4 couplers:

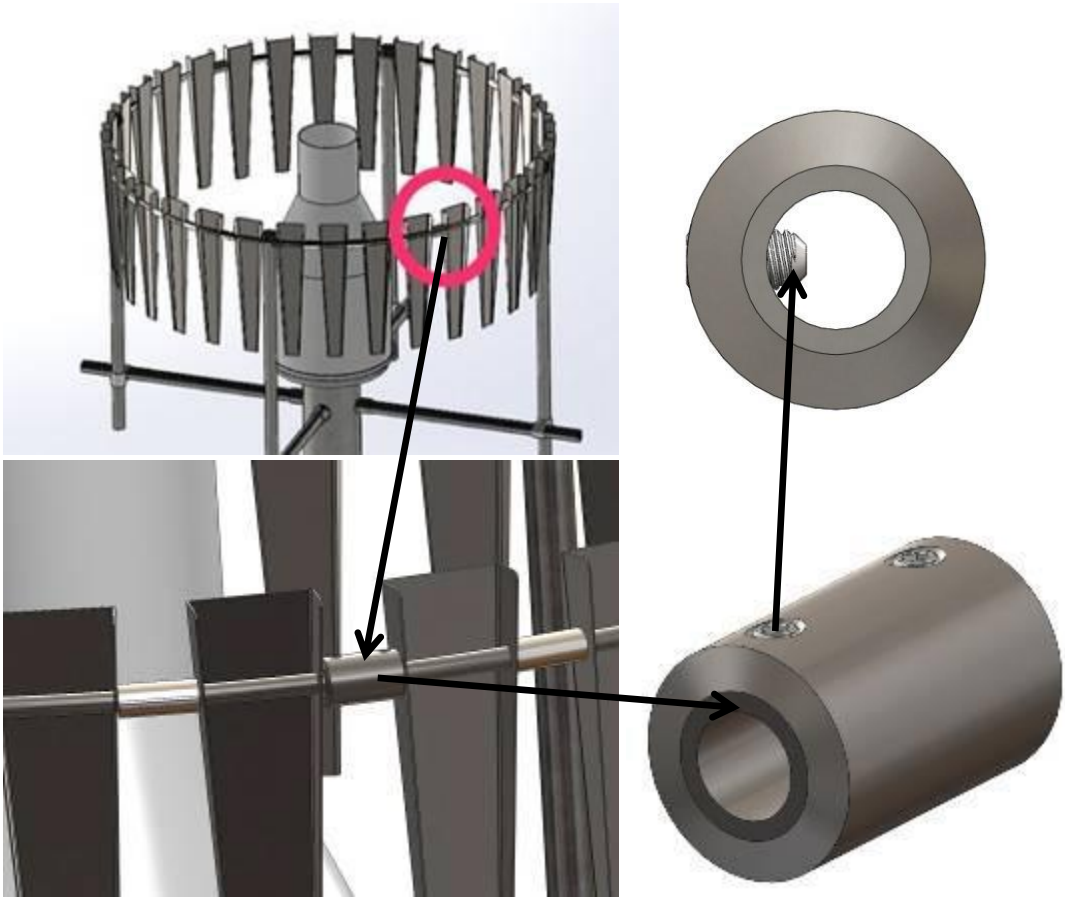


Figure 7: Coupling assembly

Shield Posts

The shield is mounted on 4 vertical posts, equally spaced. It is strongly recommended that the shield is mounted independently from the gauge to reduce the impact of wind vibration on the instrument. Generally, 4 (or more) vertical support pipes are used with the support pipes fitting between the blades (replacing a spacer at that point on the ring).

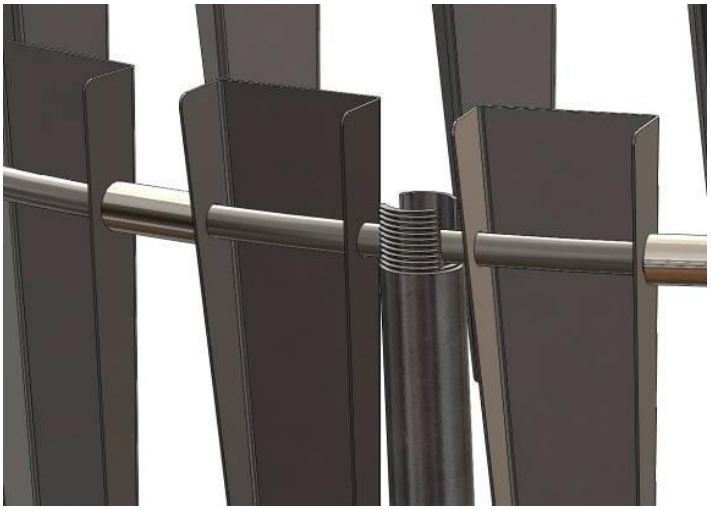


Figure 8: Support post and rod mounting - view 1

Figure 9: Support post and rod mounting - view 2

Note: The installation solutions presented in this document are for illustration purposes; the user may choose to assemble the shield differently, based on current practice, experience, and availability.

Manual observation procedures at CARE

Michael Earle and Phil Raczynski
Observing Systems and Engineering, Meteorological Service of Canada, Environment Canada

Version	Date	Notes
1	Jan. 4, 2013	Prepared by Michael Earle (OSE)
2	Jan. 8, 2013	Incorporated feedback from Phil Raczynski (OSE)
3	Feb. 5, 2013	Updated frequency of manual precipitation observations as decided upon for WMO-SPICE; included description of manual observation procedures as conducted during previous WMO Solid Precipitation Measurement Intercomparison
3A	Apr. 9, 2014	Minor corrections
3B	May 28, 2014	Updated link to Environment Canada's Manual of Surface Observations (MANOBS)

1. Overview

The following is an overview of procedures for manual measurements of precipitation accumulation and snow depth at the Observing Systems and Engineering (OSE) field measurement site at the Environment Canada (EC) Centre for Atmospheric Research Experiments (CARE) in Egbert, Ontario, Canada. This overview is based on descriptions of procedures outlined in the World Meteorological Organization (WMO) Solid Precipitation Intercomparison Experiment (SPICE) Proof of Performance (PoP) report for CARE prepared by Amal Samanter (OSE), the WMO SPICE document on measurements of snow on the ground prepared by Barry Goodison (WMO), as well as e-mail correspondence with the observer at CARE, Phil Raczynski (OSE). The equipment and methodology for measuring and recording precipitation accumulation and snow depth are outlined in Sections 2 and 3, respectively. For continuity, pertinent details of manual precipitation observations during the previous WMO Solid Precipitation Measurement Intercomparison are included in Section 2.

2. Precipitation accumulation

Manual observations of precipitation accumulation at CARE employ a Tretyakov gauge within a double-fence intercomparison reference (DFIR) shield; the former is comprised of a cylindrical collector inside a wind shield, and the latter is comprised of concentric octagonal wooden fences. The on-site observer is provided with two collectors, one of which is positioned on a base within the DFIR shield as part of the Tretyakov gauge. For additional details regarding the design and construction of the Tretyakov gauge and DFIR shield, please refer to the final report from the WMO Solid Precipitation Measurement Intercomparison (Goodison et al., 1998).

The measurements of accumulated precipitation amount are obtained through the following sequence of events, typically once daily, with one or more additional measurements to be obtained on days with precipitation (either during or following precipitation events, depending on event duration and/or observer availability):

- The empty ('clean') collector is weighed to the nearest 0.1 g on a scale in one of the sheltered shacks on-site and covered
- The clean collector is carried to the base in the DFIR shield, currently housing the 'weathered' collector, which has been open to the air and has collected any incident precipitation that has fallen within it
- The time is recorded
- If it is not precipitating, the weathered collector is left uncovered to avoid disturbing any snow/ice above the rim; if it is precipitating, the cover from the clean collector is transferred to the weathered collector, trying not to disturb any snow/ice above the rim
- In the same sequence, the weathered collector is replaced with the clean collector, now uncovered, on the base in the DFIR shield
- The weathered collector is taken inside the shack housing the weighing scale
- Any precipitation adhering to the outside of the weathered collector is removed
- In the event that there is frozen precipitation on the collector above the rim, a short waiting period is provided for melting. Any melt not falling into the collector is wiped from the outside of the container with a cloth, along with any condensation that has formed.
- The collector is weighed – again, to the nearest 0.1 g – and the specific collector used is recorded, along with the prevailing weather conditions

- Following the measurement, all precipitation is evacuated from the collector, the insides are cleaned with a cloth, and the collector is then left upside-down in the shack to dry until the time of the next measurement

All measurements and times are recorded on a designated sheet (hard copy), along with the details of the collector, observer name, and prevailing weather conditions. Weather conditions are recorded using the format outlined in Environment Canada's Manual of Surface Observations, MANOBS:

http://www.ec.gc.ca/manobs/

The precipitation measurements and associated details should also be input into an online database via a web interface, to provide users access to the most recent observations and also to provide a secondary backup of measurements.

Using the procedure outlined above, the accumulated precipitation amount in mm is determined from the difference between the weight of the weathered collector (combined weight of collector and any accumulated precipitation) and the weight of the empty (clean, dry) collector. The procedure during the first WMO Solid Precipitation Measurement Intercomparison was similar to that described above, but the accumulated precipitation amount was determined from the volume of precipitation in the collector. The specifics of this procedure are detailed in the final report (Goodison et al., 1998). A brief description of the volume-based approach and associated corrections is provided below, along with a discussion of key distinctions between this approach and the current, weight-based, approach.

During the previous intercomparison, the clean and weathered collectors were exchanged using a similar configuration[1] and procedure as the current study; however, when the (covered) weather collector was taken into a sheltered area, any solid precipitation was melted, and the collected precipitation was drained into a measuring glass. The precipitation amount was then recorded in mm. To account for any precipitation adhering to the inside of the collector ('wetting losses'), the collector was weighed immediately following draining, and then again later when the collector was dry. The difference in grams was recorded, averaged over a series of measurements (typically 40) to establish a wetting loss constant for a given collector, and then converted to a water depth in mm. In addition, to account for evaporation losses, the collector and accumulated precipitation were weighed immediately after a snowfall event and at a later time. The difference between the two weightings was converted to an evaporation loss in mm.

The wetting loss correction is not necessary when manual measurements are obtained based on weight, rather than volume, as no precipitation is drained from the collector when taking measurements. Hence, where the weight-based approach is applied during WMO-SPICE (e.g. at CARE), wetting loss corrections are not required. The evaporation loss correction may not be

[1] Canadian sites also used the Nipher shielded snow gauge, a manual gauge similar in concept and design to the Tretyakov gauge, but with a solid, inverted bell-shaped shield. At the time of the previous intercomparison, the Nipher gauge was the national standard for measuring snowfall amount.

possible during WMO-SPICE, given the logistical difficulty of having an observer on-hand to make a measurement immediately following a precipitation event. An additional consideration when manual measurements are based on both weight and volume is the potential contribution of frost on the collector. At CARE, any occurrences of frost that may influence the measurements are noted by the observer on-site and reported along with the measurement.

3. Snow depth measurements

Manual observations of snow depth at CARE employ a network of 30 graduated snow stakes placed at various locations on the site for the purpose of assessing spatial variability. The placement of stakes generally corresponds with the locations of gauge bases on-site, with three or four stakes located at different locations around each base. It is important that stakes are placed outside the field of view of automatic snow depth sensors (Campbell Scientific SR50A, Jenoptik, GMON) to avoid disturbing the snow in areas interrogated by these gauges, and hence producing erroneous snow depth readings. Stakes are presently located just outside the field of view (FOV) of each automatic snow depth sensor location at CARE for comparison with the automatic measurements (three or four stakes outside the FOV of each SR50A; two stakes outside the FOV of all other automatic sensors).

Daily snow depth measurements are obtained as follows:

- The start time is recorded
- The snow depth is measured at each snow stake to the nearest 0.5 cm, typically starting at the southeast corner of the site
- The end time is recorded

All measurements are recorded on a designated sheet (hard copy), along with the observer's name and any notes. This information should also be input into an online database via a web interface.

A method for identifying the seasonal snow depth reference on glacier ice

Antonella Senese, University of Milano, Department of Environmental Science and Policy

The measurement of precipitation and snow depth on glacial surfaces is very important for monitoring and understanding the mass balance of glaciers. Seasonal trends and variability are also obviously very important. One of the difficulties of measuring snow depth on the surface of a glacier is establishing a point of reference as a zero snow depth due to the dynamic nature of that surface. When measuring snow depth on a glacier with an automated sensor, it could be very difficult to determine the transition from ablation to accumulation.

The SPICE site on the Forni Glacier, pictured in Figure 1, uses a technique employing albedo to establish the transition from seasonal snow and glacial melt to the beginning of the seasonal accumulation of snow in the fall and winter.

Figure 1: The SPICE site on the Forni glacier.

The correct assessment of the first snowfall on the glacial surface can be defined by analyzing albedo data (Figure 2). In general, the surface of the ice, without snow, usually has an albedo value lower than 0.35. At the onset of snow accumulation, with the addition of new snow on the surface, the albedo experiences a step jump due to the difference in albedo between surface ice and new snow. Figure 3 shows that at Forni in the fall of 2014, the first snowfall occurred on 14 October (marked by the green line in Figure 2) and the snow accumulation season starts from 22 October (marked by the red line in Figure 2). Establishing this point of reference allows for the calculation of the distance to the target at the beginning of the accumulation season which then allows for the accurate determination of snow depth throughout the accumulation season. The result is the seasonal snow depth time series for 2014/2015 as shown in Figure 3.

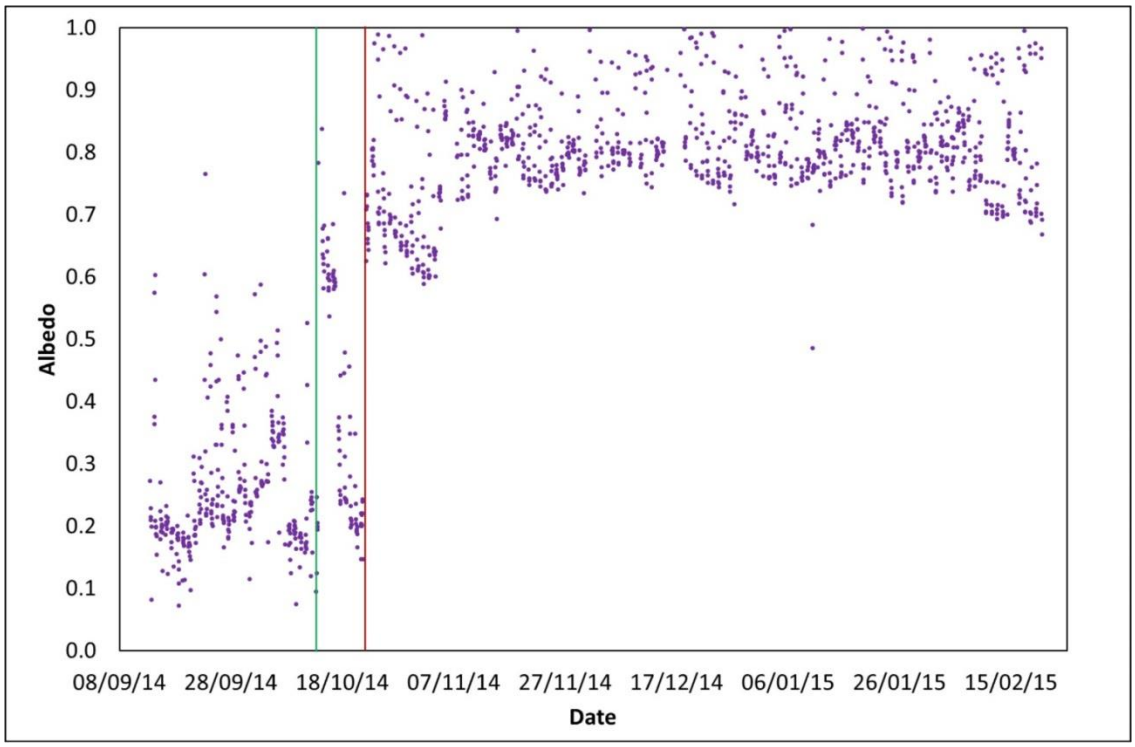


Figure 2: Albedo values estimated from solar radiation data acquired by the net radiometer during winter 2014/2015 at the Forni glacier SPICE site. The green line shows the first snowfall (14 Oct 2014) and the red one the beginning of the snow accumulation season (22 Oct 2014).

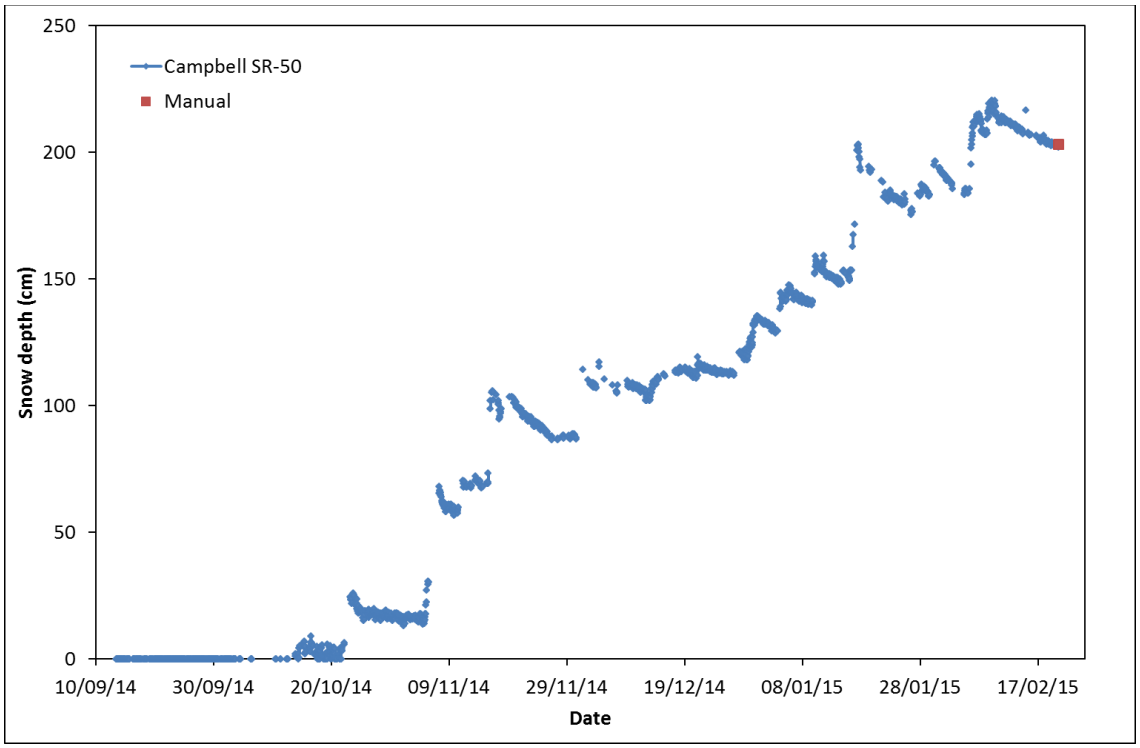


Figure 3: Snow depth values measured by an SR50 during winter 2014/2015 at Forni glacier.

Investigation of the impact of crystal type and particle size distribution on the measurement of snow using different gauge-shield configurations

Authors: Matteo Colli, Luca Lanza, Roy Rasmussen, Julie Thériault

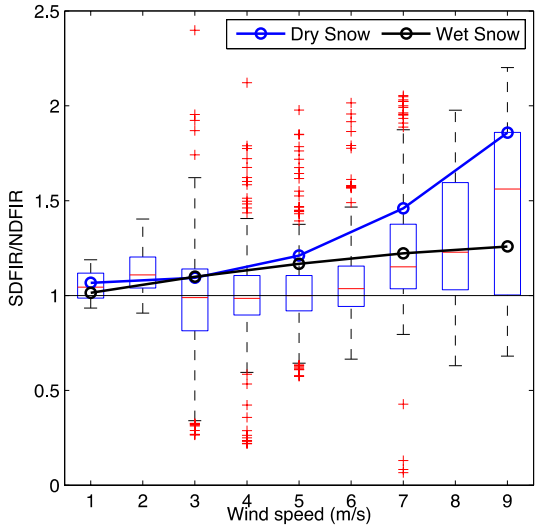
Weighing snow gauge collection efficiency decreases with increasing wind speed. This is mainly due to the perturbation of the flow in the vicinity of the gauge impacting the trajectory of snow particles. In addition to this, high scatter is present in the data even after a wind correction is made. The following section studies the factors impacting the scatter in the data after wind correction using a combination of numerical simulations and field measurements. The numerical approach uses a combination of CFD and a Lagrangian tracking model to predict the snow particle trajectory near the gauge (Thériault et al., 2012). First, we used known characteristics of different types of snow such as wet and dry snow, graupel, rimed aggregates and densely rimed snow. Second, the collection efficiency is calculated for a given snow particle based on the area associated with the number of snowflakes falling in the gauge of a given diameter (Area[D]) with respect to the area of the orifice of the gauge (Area gauge) . Finally, the collection efficiency of a size distribution is calculated, in which the size distribution is represented by an inverse exponential:

$$CE = \frac{\int_0^\infty Area(D)\exp(-\lambda D)\,dD}{\int_0^\infty Area\exp(-\lambda D)\,dD}$$

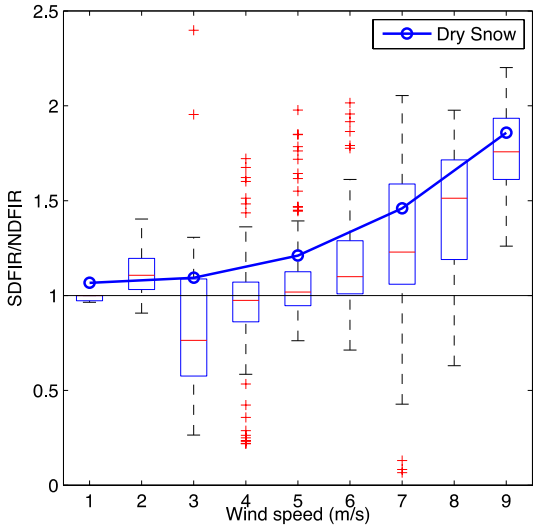
where λ is the slope parameter of the size distribution. The numerator represents the amount of snow collected by the gauge-shield configuration that we are studying and the denominator would be the amount of snow measured by the reference gauge-shield configuration (DFAR). It is clear from this equation that the collection efficiency depends on the characteristics of the precipitation because it will influence the trajectory and, in turn, the factor Area(D). It also depends on the slope of the size distribution and the accuracy of the reference gauge.

To verify the accuracy of the reference gauge-shield configuration, numerical simulations have been conducted (Thériault et al., 2015). The results show that this gauge-shield configuration does not always collect 100% of the precipitation. Because the DFAR is octagonal, two DFAR orientations have been studied. Our findings showed two mechanisms impacting the snow trajectories. First, when the flow field direction is onto a vertex of the DFAR, the airflow produced by the DFAR converges near the top of the gauge. Second, when the flow field direction is onto a flat side of the DFAR, the airflow produced by the DFAR will create a much stronger updraft upstream of the side. These mechanisms influenced the overall collection efficiency of the DFAR. The numerical results have been compared to measurement from the Marshall Test sites where two DFARs are oriented differently with respect to the north (Figure 1). Using the temperature threshold, the gauge measurement samples have been divided by crystal types: dry and wet snow. It is assumed that dry snow occurs at temperature <-4C° and wet snow at temperatures >-4C°. The results showed that the flow field further impacts dry snow collection efficiency because it follows the converging or diverging streamlines depending on the wind speed and direction. Therefore, the collection of a gauge placed in a DFAR depends on the wind speed and direction as well as the snowflake types.

(a)



(b)



(c)

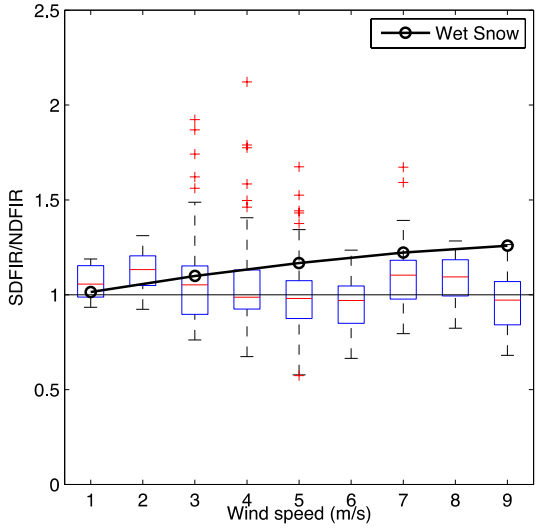


Figure 1: The collection of the Double fence inter-comparison reference (DFIR) ratio based on 2 orientations with respect to the wind direction. The SDFIR has the vertices oriented northward and the NDFIR is oriented 22.5° east with respect from the north. The box plot represents the measurements and the solid lines are obtained with the numerical simulations. (a) all snow, (b) dry snow and (c) wet snow. Dry snow is associated with snowfall at temperature < -4C and wet snow with temperature >-4C. [Figure adapted from Thériault et al. (2015) published by © American Meteorological Society]

In parallel, other studies has been conducted using an unshielded gauge and a gauge placed in a single-Alter shield to determine the impact of a time-varying flow on the trajectory of the snow particles (Colli et al., 2016a). The analysis was carried out using a combination of CFD simulations assuming time-invariant flow based on the Reynolds Averaged Navier-Stokes equations (RANS) and a time-variant flow based on the Large Eddy Simulations (LES) model. A comparison between the RANS and LES modelled airflows highlighted a general under-estimation of gauge generated turbulence by the former model just above the gauge orifice rim. The LES revealed that the intensity and the spatial extension of such turbulent region has a dependence on wind speed that was not detected by using a RANS approach (Figure 2). As a result, the CE from the LES approach was slightly lower than that derived from the RANS model (Colli et al. 2016b).

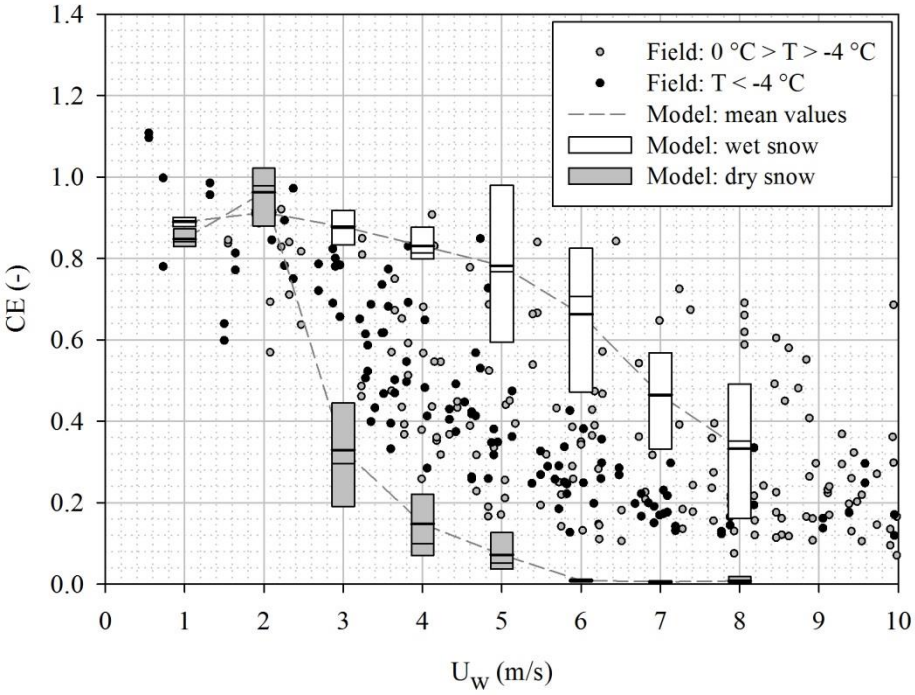


Figure 2: Comparison between in-field SA shielded GEONOR (grey and black dots) and CFD LES modeled (dashed curves with boxplots) CE vs. the undisturbed horizontal wind speed Uw. Field data are classified based on the environmental temperature T separating precipitation occurring under 0°C > T > -4°C from T < -4°C. [Figure adapted from Colli et al. (2015) published by © American Meteorological Society]

The time-dependent simulations showed that the propagation of the turbulent structures, produced by the aero-dynamic response of the upwind SA blades, has an impact on the turbulent kinetic energy realized above the gauge collector and on the particle trajectories.

The time-dependent CE estimates provided by Colli et al. (2016b) are appreciably lower than existing numerical simulation results obtained by using RANS models. Both shielded and unshielded gauge were tested. Figure 3 shows the trajectory of the same type and size of snow in the vicinity of the gauge and gauge-shield configuration. Noticeable difference between the CE for dry and wet snow crystals demonstrated the importance of the physical parameterization of hydrometeors (terminal velocity and mass) for both the shield and unshielded gauge.

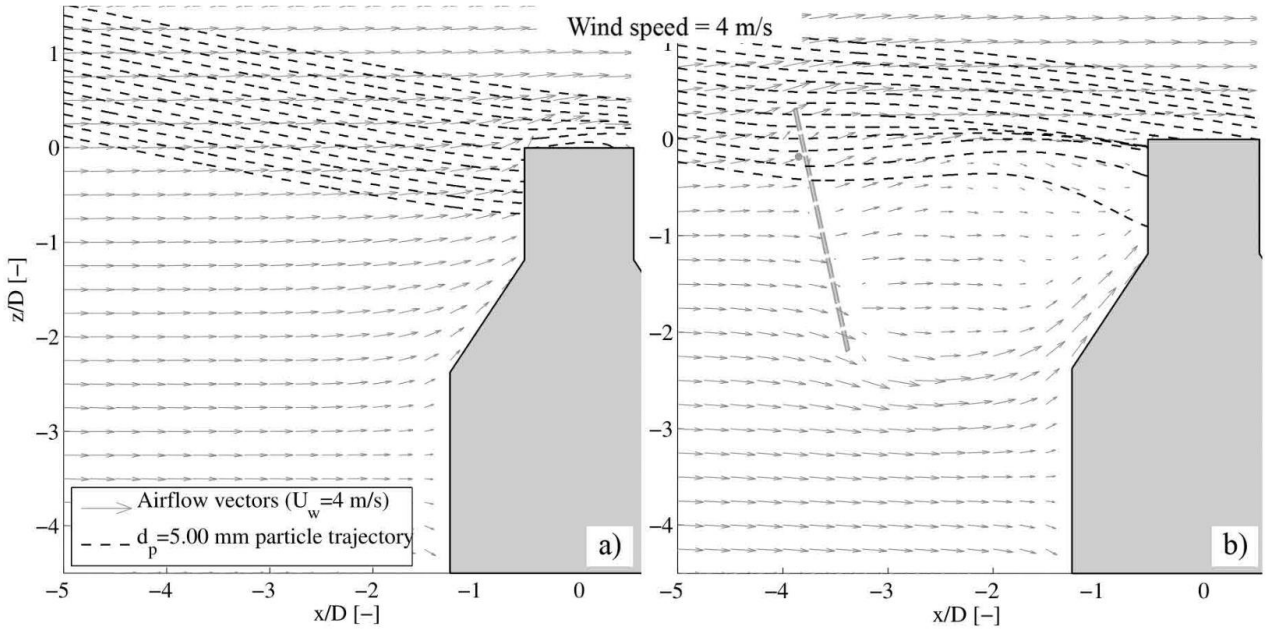


Figure 3: Unshielded (panel a) and single Alter shielded (panel b) GEONOR 600 mm gauge. Sample dry snow (dp=5 mm) particle trajectories (black lines) and vector plot (grey lines) on a vertical plane as computed by the Reynolds Averaged Navier-Stokes (RANS) airflow and the Lagrangian tracking model at a horizontal wind speed equal to 4 m/s. The spatial coordinates x and z are normalized with the gauge collector diameter D. [Figure adapted from Colli et al. (2015) published by © American Meteorological Society]

A first step towards addressing this issue was to study the formation of the drag coefficient used in the Lagrangian model. The dynamic formulation of the drag coefficient adopted by Colli et al. (2015) yielded significantly improved numerical model estimates of snow collection efficiencies by increasing the theoretical collection efficiency leading a better comparison with observations (Figure 4).

It was also shown that the slope of the impacting size distribution influenced the collection efficiency due to larger particles having a higher collection efficiency (Thériault et al. 2012). Rimed particles were also shown to have a large impact on the observed collection efficiency.

These studies show the importance of accounting for microphysical factors on the snow gauge collection efficiency. In particular, it is important to account for the type of snow and its size distribution to explain the scatter in the data at a given wind speed.

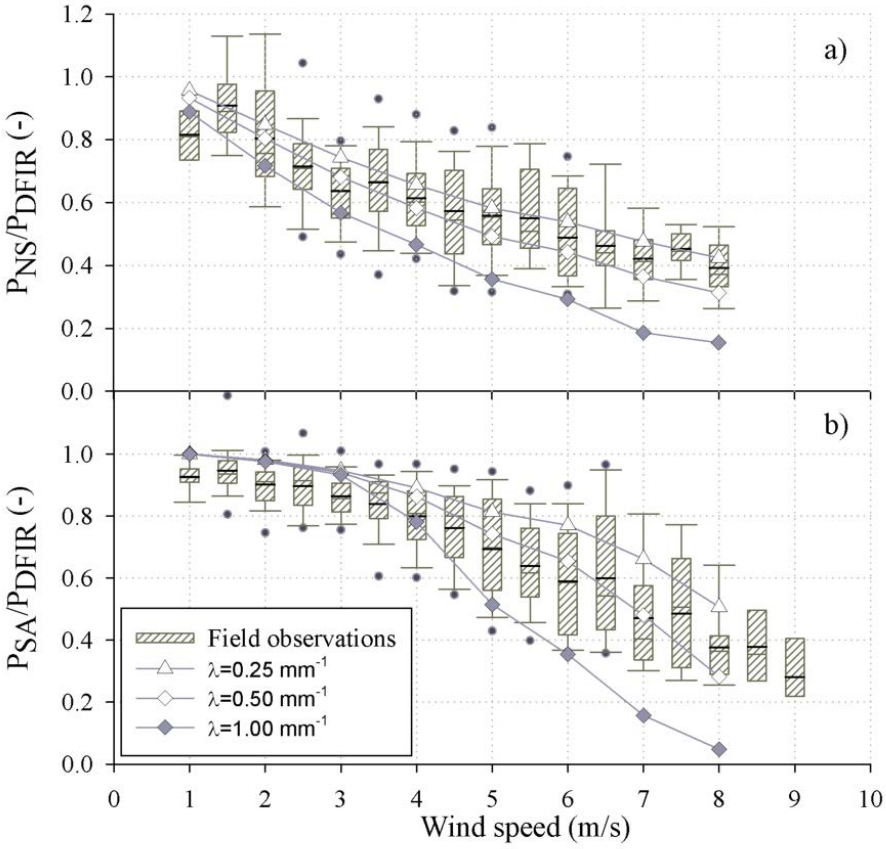


Figure 4: Collection efficiency vs. the horizontal wind speed for an unshielded (a) and a single Alter shielded (b) GEONOR computed by tracking the snowflake trajectories with an improved formulation of the particles drag coefficient (Colli et al., 2015). Three different particle size distributions for snow are simulated according to the slope parameter λ (mm−1). The Marshall unshielded gauge data are shown in the background. [Figure adapted from Colli et al. (2015) published by © American Meteorological Society]

SPICE site report: Rikubetsu, Japan

Naohiko Hirasawa[1],

Hiroyuki Konishi[2],

Koichi Nishimura[3],

Christophe Genthon[4]

and

Project group[5] of Japan Meteorological Agency (JMA)

[1]hira.n@nipr.ac.jp, National Institute of Polar Research,

10-3, Midoricho, Tachikawa, Tokyo 190-8518, Japan

[2]Osaka-kyoiku University, [3]Nagoya University,

[4]Institut des Géosciences de l'Environnement, Université de Grenoble

[5]JMA project group: K. Bessho, Y. Tahara, K. Honda, A. Umehara

ABSTRACT

This paper describes the snowfall measurement at the Rikubetsu measurement-site for SPICE in two winters of 2013/14 and 2014/15 for SPICE (Solid Precipitation Inter-Comparison Experiment), organized by WMO. Rikubetsu (43.5°N, 143.8°E, 217 m a.s.l.) is located in the northern part of Japan, which represents one of the coldest areas in Japan such as around -30 °C of daily-minimum temperature in winter. At Rikubetsu, dry snow particles fall more frequently and the snow accumulation is smaller, around 300 mm for December to March, than other snow areas in Japan. The frame of the observation is constructed of DFIR (Double Fence Inter-comparison Reference for the SPICE project) using a weighing gauge, Geonor T-200B, and nine instruments for test measurement of snowfall together with instruments for surface meteorological elements. For the test measurements, three kinds of disdrometer (LPM, Parsivel and PWS), a snow particle counter (SPC), two types of tipping gauge (RT3 and RT4, conducted by Japan Meteorological Agency: JMA) and a ceilometer have been installed. The main results are as follows:

- As for the total amount of snowfall, those measured by the disdrometers gave the largest, the DFIR-Geonor followed them, and the RT4 and RT3 gave the smallest, which are 50-60% of DFIR-Geonor.
- The Geonor showed false diurnal variation synchronized with ambient air temperature. The range of the variation is approximately 0.5 mm for 20 °C.
- The disdrometers were affected by strong wind, in many cases of which the snowfall rete was erroneously enhanced. 4) When snow does not fall at all, the LPM frequently showed erroneous counts in a region of smaller particle size (< approximately 1 mm) and higher velocity (> approximately 1 m/s). 5) RT4 improves the total amount of snowfall by around 20% compared with RT3 due to evaporation loss.

1. INTRODUCTION

Moisture circulation in the atmosphere varies in association with variation in the climatic system. Precipitation is not only a passive element for the climatic change but also plays a key role in the climatic change. Namely, precipitation may change wetness condition of surface layers, which kicks subsequent processes differently, e.g. sensible and latent heat flux between surface and lower atmosphere, formation of synoptic-scale circulation pattern, or convective activity. In particular, change in snowfall amount, snowfall frequency, and length of period with snow cover drives ice-albedo feedback. Furthermore, snowfall is the main feeder of the polar ice-sheets, i.e. Greenland and Antarctica, and glaciers in mountain ranges of the earth. Their evolution totally relates to sea level change.

On estimation of present climatic state and of future climate under global warming, measurement of the snowfall amount has got to be important and it will be more and more toward the future. However, there is still serious difficulty in snowfall measurement more than rain. In 1990's, WMO conducted the solid precipitation measurement inter-comparison to determine the degree of inaccuracy in measurement and to develop correction methods. The final report edited by Goodison et al. [1998] shows the correction functions of daily snowfall amount for each tested instrument. The results compiled in the final report have been applied to precipitation for the Arctic Ocean [Yang, 1999a] and to precipitation for Greenland [Yang, 1999b]. Adam and Lettenmaier [2003] showed that terrestrial precipitation were under-estimated based upon the final report. Tian et al. [2007] showed that the correction of the catchment error results in increment in runoff of fresh water from the northern continent into the Arctic Ocean.

The most serious problem has been wind-induced reduction in catchment of snow particles into the gauge. When wind speed is more than 5 m/s, the measured snowfall rate gets to be less than half of true value. The ideal and perfect measurement avoided from the wind-induced reduction is considered as bush gauge, which is installed in a tree-free area in a forest. Practically, in order to suppress the wind-induced reduction, wind shields have been installed around the instruments.

The underestimation results also from evaporation loss, wetting loss, and miss-count of trace precipitation. Some instruments catch snow particles in their buckets, and evaporation continues to occur at the surface of liquid water which is made by melting of the sampled snow particles by heating. Wetting loss occurs by evaporation of snow particles at inner lateral surface of the instruments gauge. It may be considered as the first process of evaporation loss. Trace precipitation is one with too small falling rate and also the amount to measure it by instrument, which, however, can be verified by eyes. It also may be considered as the first process of evaporation loss. In order to suppress the evaporation loss, some instruments catch snow particles into antifreezing solution capped with thin oil layer, e.g. Geonor (weighing gauge) and RT4 (tipping gauge employed by Japan Meteorological Agency, JMA).

WMO conducted a project for snowfall measurement again, entitled "Solid Precipitation Inter-Comparison Experiment (SPICE)" [Qiu, 2012]. Rasmussen et al. [2012] summarized the present state of snowfall measurement which has advanced since the first intercomparison [Goodison et al., 1998]. While many kinds of wind shield have been tried to improve the wind-induced reduction in catchment of snow particles, no shield has yet performed at sufficiently high collection rate of snow

particles with the diameter of smaller than 4 m. On the other hand, the resolution of amount of snowfall has been improved such as 0.2 mm/hr for weighing gauge. Disdrometer, which measures the size and the fall velocity for each precipitation particle, has got to be expected not only to detect occurrence of weak-rate snowfall but also to measure the snowfall rate.

The goal of the authors is reliable measurement of snowfall in Antarctica, which severe conditions inhabit. One of the main snowfall processes emerges associated with synoptic-scale disturbance, when the snowfall occurs a great deal in strong wind more than 10 m/s [e.g. Hirasawa et al., 2000; Schlosser at al., 2010; Reijmer and Broeke, 2003]. Another main process is diamond dust, which occupies around 80% of time in the Antarctic interior [Kuhn et al., 1975; King and Turner, 1997]. The snowfall rate at the event with synoptic-scale disturbance is several 10 mm/day in the Antarctic coast and a few mm/day in the interior [e.g. Welker et al., 2014]. Although the snowfall rates under such disturbance at coastal stations drop in the suit range of capturing by recent instruments, the relatively strong wind speed gives some problems, which are the wind-induced reduction and contamination of blowing snow. As for the diamond dust, it is less than 1 mm/day in the interior [e.g. Hirasawa et al., 2013]. Walden et al. [2003] summarized, based upon observation at South Pole, that the size of diamond dust is generally smaller than a hundred micro meter. Thus, it is difficult to detect diamond dust in the Antarctic interior even by recently developed disdrometers because the detectable size of precipitation particle is generally larger than 125 micro meter. So, we need to employ such a particle counter that can sense particles smaller than a hundred micro meter, i.e. Snow Particle Counter (SPC) manufactured by Niigata Electronics Co.

Rikubetsu (43.5°N, 143.8°E, 217 m a.s.l.) is located in the northern part of Japan, which is one of the coldest areas in Japan with around -30°C of daily-minimum temperature in winter, as shown by Sorai et al. [2016]. Surface-based temperature inversion is established at such cold night. At Rikubetsu, dry snow particles fall more frequently and the snow accumulation is smaller, around 300 mm for December to March, than other snow areas in Japan. Therefore, the meteorological conditions and snow features of Rikubetsu are the most comparable to the Antarctica among Japan, and the authors decided to test the instruments for measuring snowfall there. The activity will contribute to SPICE and the efforts for SPICE must contribute to coming observation in Antarctica.

For the SPICE site, in the Rikubetsu site, a double fence inter-comparison reference (DFIR) with full-scale fences of 4 m and 12 m in diameters and a weighing gauge, GEONOR, was installed. In the test field, three types of disdrometer, two particle counters, two types of gauge which are the main systems of JMA, a ceilometer, and a snow depth sensor. The activity started at 2012/13 winter and continues through 2016/17 winter.

This paper documents the snowfall measurement and the relevant meteorological measurement at Rikubetsu for SPICE, summarizes the results of observation, and describes new findings in instrumental problems, characters of measured values, and future direction for improvement based upon the results of 2013/14 and 2014/15 winters.

2. OBSERVATION FIELD AND INSTRUMENTS

2.1 The observation field

Rikubetsu (43.5°N, 143.8°E, 217 m a.s.l.) is a small town in the central area of Hokkaido (Figure 1a), the northernmost island of Japan. As Rikubetsu is approximately 100 km far from the surrounded coasts, the daily minimum temperature in winter drops around -25 °C. Then, it is one of the coldest areas in Japan, as discussed by Sorai et al. [2016]. Figure 1b shows the landscape of the Rikubetsu site seen approximately toward the northeast. The prevailing wind is from approximately the north. There are two sets of double fence wind shield of the full frame-size, a field for testing instruments, and a hut for installing data logger, PCs, and facilities of power supply.

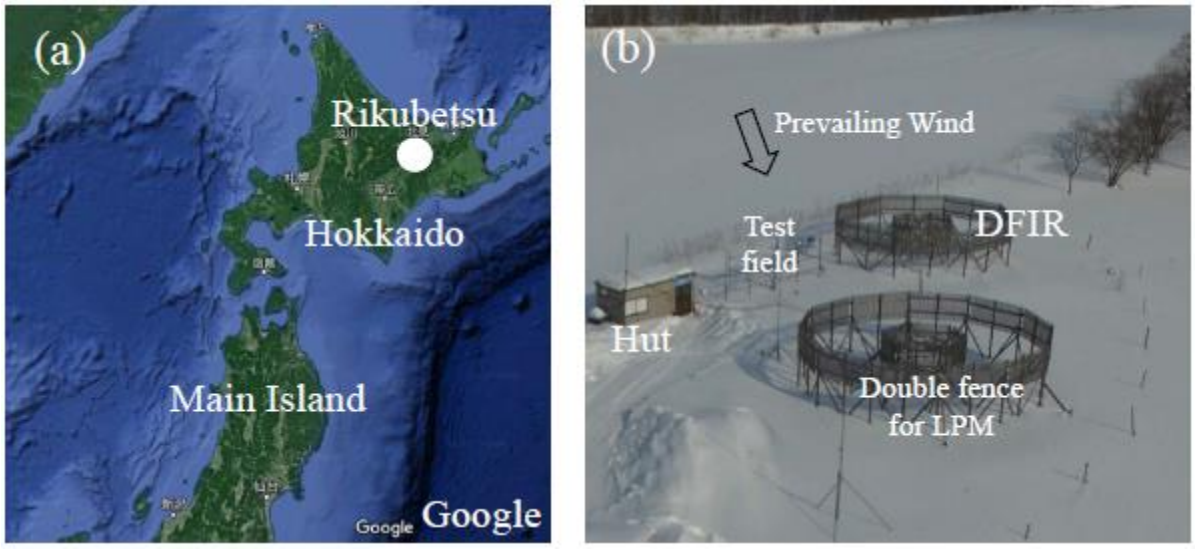


Figure 1 (a) The Rikubetsu site on a map of the northern part of Japan. (b) A view of the main area of the Rikubetsu site from a radio-controlled small plane.

2.2 Meteorological characters of the 2013/14 and 2014/15 winters

Figure 2 shows time series of surface air temperature, wind speed, and precipitation for the two winters (from December to March) at the national site in Rikubetsu operated with automatic weather station (AWS) by Japanese Meteorological Agency (JMA). The site locates to the southwest of the SPICE site with the distance of approximate 3 km. For both winters, the daily minimum temperature frequently drops to around -25 °C. It sometimes remains in range of temperature higher than -10 °C under southerly winds associated with synoptic-scale cyclones. In such days, relatively high wind speed and relatively large amount of precipitation are observed.

Figure 3 shows time series of surface air temperature of daily mean and accumulated snowfall amount from 1 December to 31 March for the recent six winters from 2011/12 to 2016/17. The sub-seasonal variation in air temperature in 2013/14 winter showed relatively higher temperature in December and relatively lower temperature in March than other winters. In 2014/15 winter, air temperature was relatively higher in the middle of January and from the middle of February through March.

The accumulated snowfall amount for the 2013/14 winter was approximately 100 mm, which drops in the standard range of the six winters. Although that for the 2014/15 winter progressed in the standard range till the end of February, it reached about 200 mm at the end of March, around twice the accumulation of the other winters.

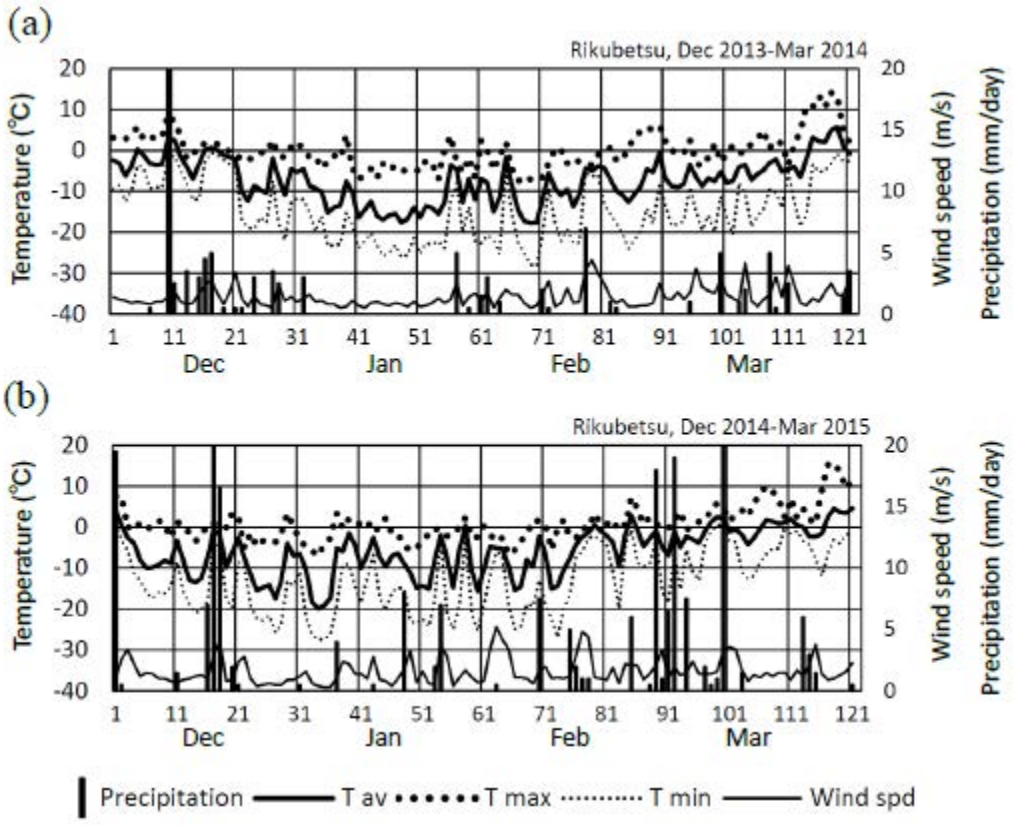


Figure 2 Time series of surface air temperature, wind speed, and precipitation for the two winters 2013/14 and 2014/15 (from December to February) at the national site in Rikubetsu operated with automatic weather station (AWS) by Japanese Meteorological Agency (JMA). Surface air temperature (SAT) is shown in three kinds of parameters of daily mean, daily minimum, and daily maximum. Wind speed is the daily mean and precipitation is the daily total.

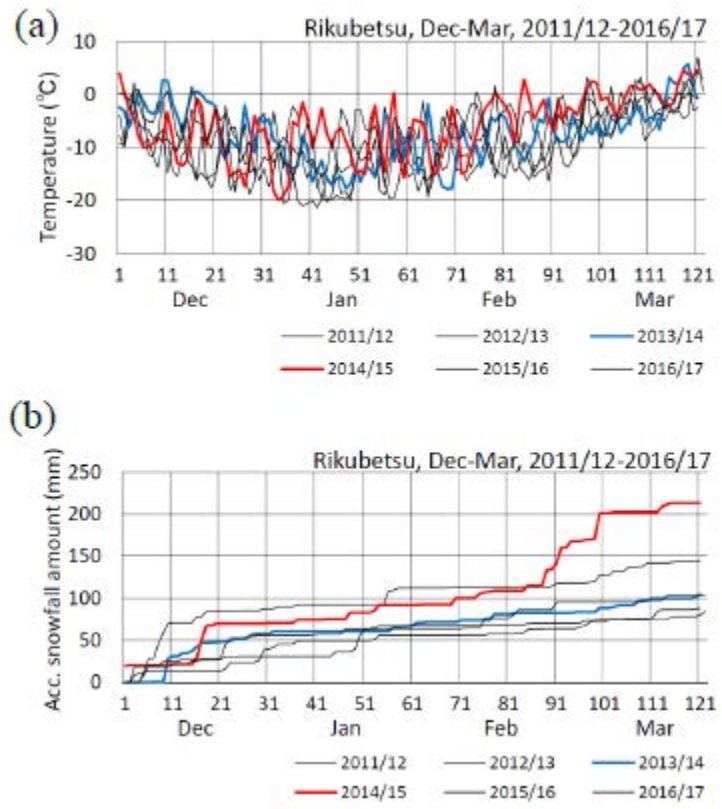


Figure 3 Inter-annual variability of the surface meteorology of Rikubetsu for the recent six years, including winters 2013/14 and 2014/15 for (a) surface air temperature of daily mean and (b) accumulated snowfall amount from 1 December to 31 March.

3. INSTRUMENTS

Table 1 shows specifications for each instrument installed at the Rikubetsu site. The information are derived from the catalogues indicated by the manufacturers. The details will be discussed in the following sections. Figure 4 shows deployment of instruments in the test field. The three kinds of disdrometers (LPM, Parsivel and PWS), two SPCs, gauge type instruments (RT3 and RT4), a ceilometer and meteorological sensors appear in the photo. Almost all the sensors, except for a meteorological sensor by Vaisala, measure at 3 m above the ground.

Table 1 Specifications for each instrument installed at the Rikubetsu site.

	Type Name of instrument (Manufacturer)	Output parameter(s) on precipitation	Way of measurement and measuring area	Major information on measurement scheme	Detectable minimum limit	Resolution for Geonor RT3 and RT4 / Error intensity for disdrometers
DFIR	Weighing gauge T-200B (Geonor)	(1) Accumulated amount of snowfall (For every minute)	Collection in gauge. 200 cm²	(1) Trap of snow particles in a bucket with non-freeze solution covered with oil layer on the surface. (2) Continuous weighing* of the bucket by tension meter hanging the bucket. * By conversion from vibration frequency of wire equipped to the tension meter to the weight (3) Installing in a double fence wind shield. (4) 600 mm capacity of the bucket. (5) Temperature drift of 0.001 %/°C	0.05 mm	Around 0.01 - 0.05 mm (Depends on the total weight at once)
Instrument for test	Tipping bucket gauge RT3 (Optical rain)	(1) Count of the tips (For every 10 minute)	Collection in gauge. *** mm in diameter	(1) Melting snow particles on surface of funnel by heating. (2) Running the melt water into the tipping bucket below.	0.2 mm	0.2 mm
Instrument for test	Tipping bucket gauge RT4 (Yokogawa)	Same as above	Collection in gauge. *** mm in diameter	(1) Trap of snow particles in a bucket with non-freeze solution covered with oil layer on the surface. (2) Running the non-freeze solution pouring* out of the bucket into the tipping bucket below. * Induced by increment in volume of the non-freeze solution by snow particles which fall into.	0.5 mm	0.5 mm
Instrument for test	Disdrometer LPM (Thies)	(1) Precipitation rate (mm/hr). (2) Particle size (22 classes). (3) Fall velocity (20 classes). (For every minute)	Horizontal laser sheet. 20 × 228 (mm²)	(1) Measuring each snow particle at falling through the laser sheet and the statistics for every minute. (2) The classes of particle size covers for 125 μm -8 mm and larger than 8 mm. (3) The classes of fall velocity covers for 0 - 20 m/s. (4) Installing in a double fence wind shield, same as DFIR.	0.005 mm/hr for 1 min (drizzle rain)	< 30 % for snow
Instrument for test	Disdrometer LPM (Thies)	Same as above	Same as above	Same as above without double fence wind shield	Same as above	Same as above
Instrument for test	Disdrometer Parsivel (Ott)	(1) Precipitation rate (mm/hr). (2) Particle size (32 classes). (3) Fall velocity (32 classes). (For every minute)	Horizontal laser sheet. 30 × 180 (mm²)	(1) Same as above. (2) The classes of particle size covers for 200 μm -25 mm. (3) The classes of fall velocity covers for 0 - 20 m/s.	0.001 mm/hr for 1 min (drizzle rain)	–
Instrument for test	Disdrometer PWS100 (Campbell Scientific)	(1) Precipitation rate (mm/hr). (2) Particle size (34 classes). (3) Fall velocity (34 classes). (For every minute)	Horizontal laser sheet. 4800 (mm²)	(1) Same as above. (2) The classes of particle size covers for 0 mm -2.56 mm and larger than 2.56 mm up to 7.44 mm. (3) The classes of fall velocity covers for 0 - 25.6 m/s and higher than 25.6 up to 74.4 m/s.	Rainfall Resolution 0.0001 mm	10 % for rain
Instrument for test	Particle counter SPC (Nagata electronics)	(1) Particle size (32 classes) (For every second)	Horizontal laser sheet. 25 × 2 (mm²)	(1) Same as above. (2) The classes of particle size covers for 40-500 μm.	–	–
Instrument for test	Particle counter SPC (Nagata electronics)	Same as above	Same as above but for vertical laser sheet	Same as above.	–	–
Instrument for test	Lidar Ceilometer CT25K (Vaisala)	(1) Backscatter coefficients (sr^{-1}·m^{-1}) (For every 15 seconds)	Vertical laser beam	(1) Estimation of the backscatter coefficient for every vertical layer of 30 m in thickness from the ground to 7500 m in height	–	–

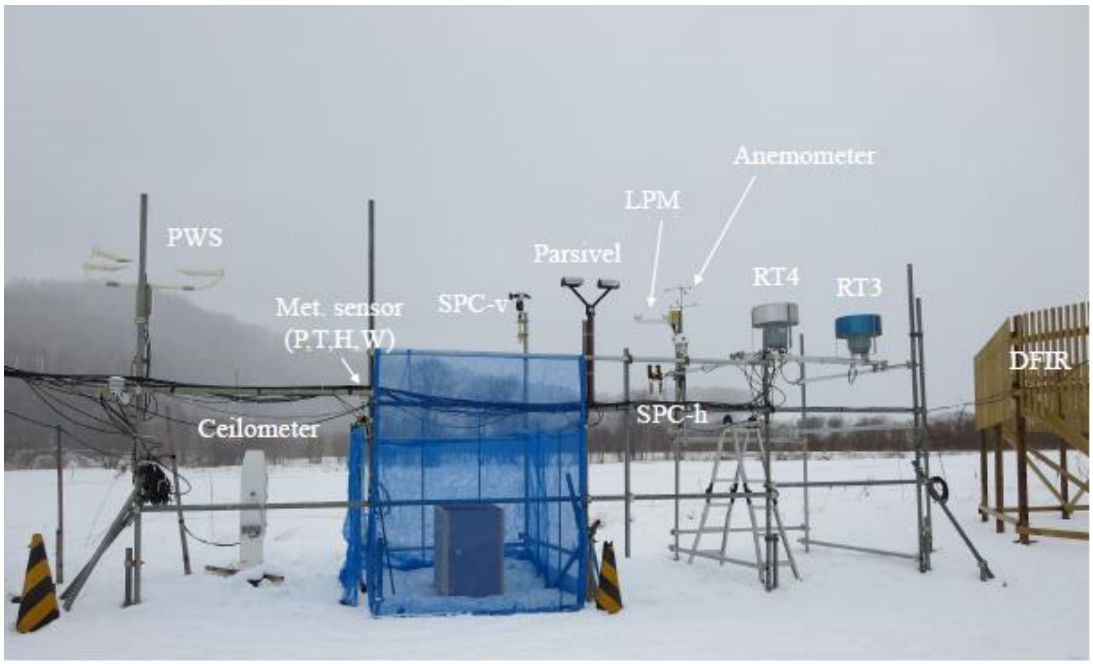


Figure 4 Deployment of the instruments in the test field.

3.1. Weighing gauge, Geonor, as the instrument for the DFIR

The measurement by a weighing gauge in a double fence wind shield is necessary for the SPICE sites because the data enable us to do inter-comparison between the SPICE sites. It is referred to as double fence inter-comparison reference (DFIR) in the project. Geonor-T200B (Figure 5) has been installed in the Rikubetsu site as the weighing gauge.

Geonor catches snow particles into bucket with anti-freeze solution, measures the weight automatically and continuously, and converts it into precipitation amount. The surface of anti-freeze solution is covered by thin oil layer to avoid evaporation from the bucket. Geonor actually measures frequency of vibration by a sensor of load cell, and it is converted into weight according to calibrated function. The unit count of the frequency provides the resolution of the snowfall amount. The resolution depends on the accumulated snowfall amount in the instrument, as shown in Figure 6. It is from approximately 0.02 mm at small amounts of accumulated snowfall to 0.05 mm at large amounts.

We would like to note that we should take care of the difference between snowfall amount (the unit is mm) and snowfall rate (in mm/hr). That is, an increment of 0.05 mm in snowfall amount for a day means only snowfall rate of 0.05 mm/day but it is not the evidence of snowfall rate of 0.0208 mm/hr (0.05 mm/24 hours). In other words, a longer time for estimation of snowfall amount gives a mean snowfall rate for the longer time period. The snowfall rate is sometimes smaller value than the resolution of snowfall amount but we cannot detect when or how the snowfall occurred in the period.

Figure 5 Geonor T200B in the double fence.

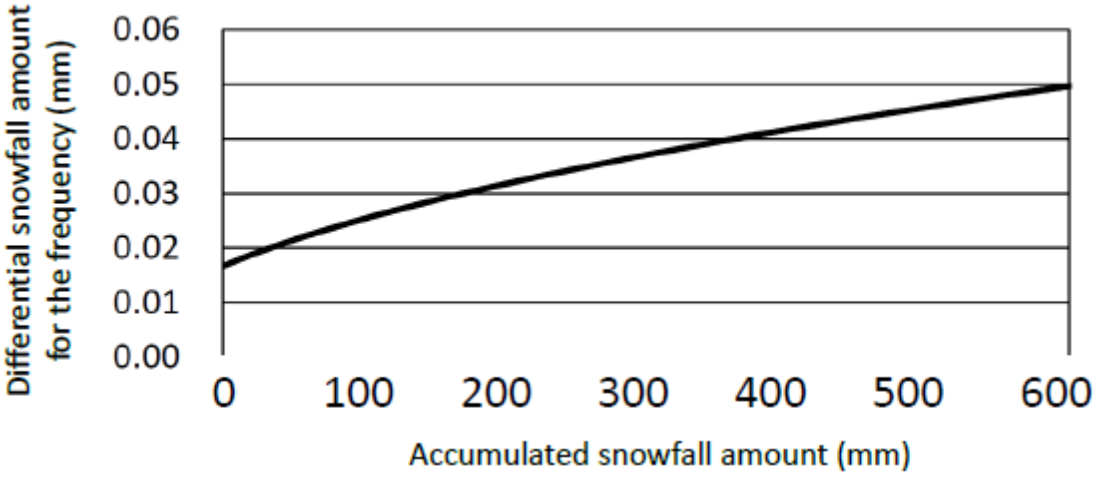


Figure 6 Resolution of the measurement of snowfall amount, depending on accumulated snowfall amount.

3.2 Disdrometers, LPM, Parsivel and WPS

Three kinds of disdrometers, LPM manufactured by Thies, Parsivel by OTT, and PWS by Campbell Scientific (Figure 7). The disdrometers evaluate snowfall rate based upon scattering ratio of laser with their original function. Each disdrometer has a table of classification for size and fall velocity of snow particles. The details are given in manufacture's catalogs.

Figure 7 Three kinds of disdrometers installed in the test field. (a) LPM (Thies), (b) Parsivel (OTT), and (c) WPS (Campbell Scientific).

3.3 Snow particle counter, SPC

Figure 8 is a photo of SPC. SPC measures only size of snow particles with the range of 40 micro meter to 500 micro meter, dividing into 32 classes. As it does not cover all the size of snow particles, we intend to detect very weak snowfall such as diamond dusts with it.

Figure 8 SPC installed in the test field.

3.4 Tipping buckets

Figure 9 shows tipping buckets, RT3 and RT4. JMA has employed both the types. Figure 10 illustrates schema of the measurement mechanism in RT3 and RT4 RT3 melts snow particles by heating the inner slope and pours it into the tipping bucket. RT4 catches snow particles into anti-freeze solution and then, the anti-freeze solution brims over into the tipping bucket. The surface of the anti-freeze solution is covered by thin oil layer to avoid evaporation loss. One tipping corresponds to 0.5 mm in water level for both gauges.

Figure 9 Tipping buckets, RT3 and RT4. The mechanism of measurement in each instrument is illustrated in Figure 10.

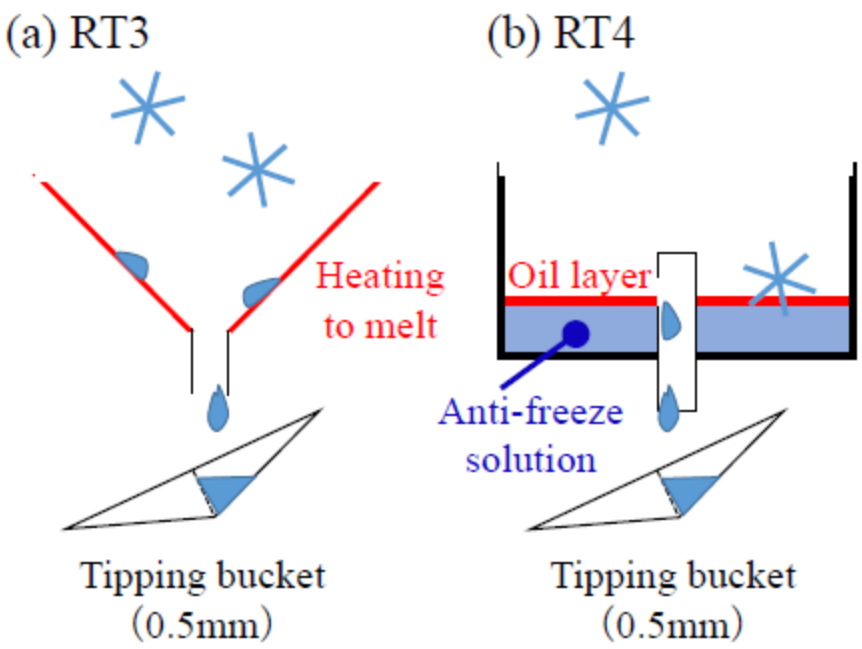


Figure 10 Illustration of measurement mechanism in (a) RT3 and (b) RT4.

3.5 Ceilometer

The ceilometer is manufactured by Vaisala. As the measurement is free from the catchment loss of snow particles and evaporation loss, the backscatters near surface layers are worth estimating snowfall rate.

3.6 Meteorological measurement

Multi types of meteorological observation were operated at the Rikubetsu site. Vaisala's weather station is employed to monitor ordinary surface meteorological elements, air temperature, pressure, humidity, wind speed and direction. A 3-D ultra-sound anemometer is employed to monitor the turbulent condition around the instruments in the test field. Moreover, another measurement for the surface meteorology, air temperature, pressure, wind speed and direction, is operated in the site by another activity, and the data are considered as a backup.

JMA operates official and routine observation at another site approximately 3 km away from the Rikubetsu site. This report contains the results of the preliminary analyses using JMA data, which had been processed before the meteorological data obtained at the Rikubetsu site had not been finished preparing to analyze.

4. RESULTS IN OBSERVATION FOR THE TWO WINTERS

Figure 11 shows each observed period by each instrument. Parsivel had finished the work after 14:00 JST, 23 February in the 2013/14 winter. The measurement of PWS, RT3 and RT4 started on 18 December 2013, and several short periods with troubles in the measurement are contained till the end of March 2015. LPM and LPM-DF observed extremely large snowfall rate for 7:00 JST, 2 February 2014 and for 14:00-15:00 JST, 17 December 2013 but no evidences were confirmed by other instruments. Thus, the data at those times were replaced with zero. As the two systems are perfectly independent each other, the signals may reflect some natural phenomena such as fog.

Figure 12 shows time series of accumulated snowfall amount observed by each instrument for the two winters, respectively. In this case, the data gap periods of PWS, RT3 and RT4 are interpolated with the Geonor (DFIR). Table 2 shows ratio of each total accumulated snowfall amount to that of Geonor at the end of March.

LPM measured the largest accumulated snowfall amount for the two winters, around 1.5 times larger than Geonor. LPM-DF measured the second largest, around 1.2 times larger than Geonor. Parsivel and WPS measured slightly smaller amount than Geonor in the 2013/14 winter and slightly larger amount than Geonor in the 2014/15 winter. The reason for the systematic difference in snowfall amount between LPM and the other disdrometers should be clarified to improve reliability in measurement.

All the disdrometers measured larger amounts than Geonor, and it is considered to reflect the difference in character of snow particles and/or meteorological features, e.g. wind speed, during snowfall between the two winters. To clarify the reason will contribute to improvement in measurement of snowfall amount.

RT3 and RT4 estimated the amount 30-50% lower than Geonor. The amount observed by RT3 is approximately 20% lower than by RT4. It will come from avoiding evaporation loss, wetting loss at RT4.

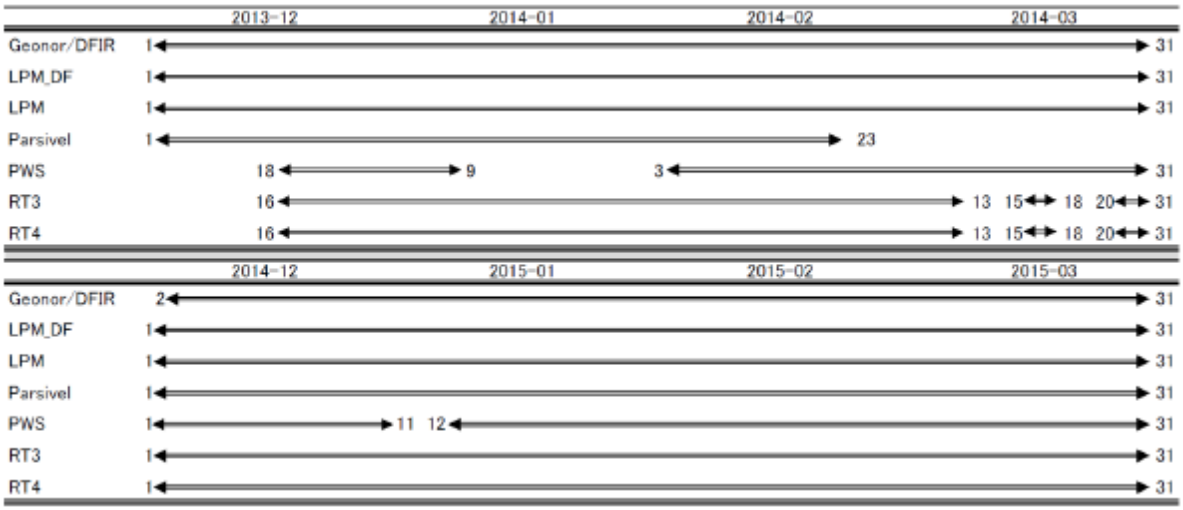


Figure 11 Each period of observation by each instrument.

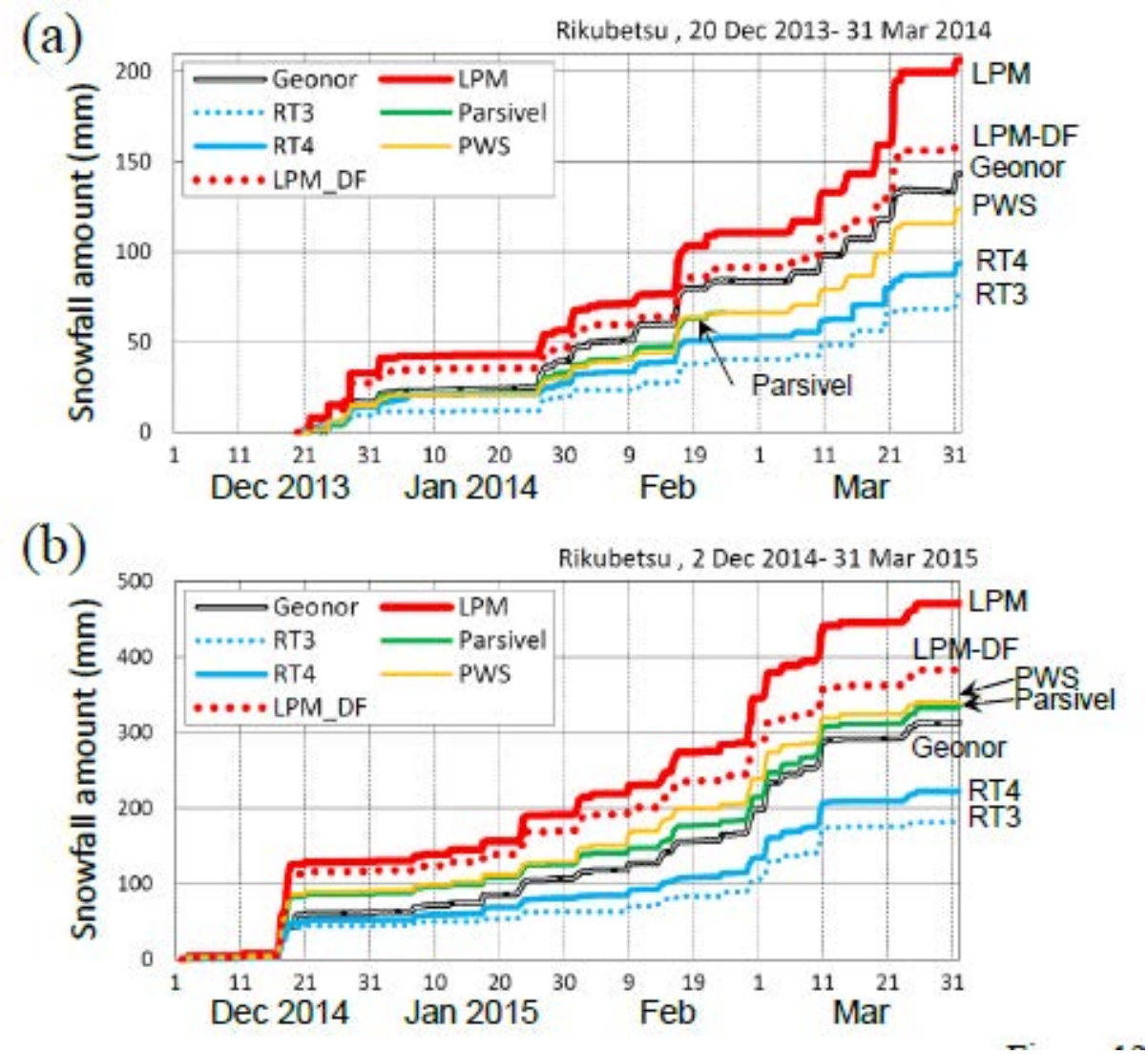


Figure 12 Time series of accumulated snowfall amount observed by each instrument for the two winters, respectively.

Table 2 Ratio of each total accumulated snowfall amount to that of Geonor (DFIR) at the end of March.

	20 Dec 2013 - 31 Mar 2014		2 Dec 2014 - 31 Mar 2015	
Insruments	Total amount (mm)	Ratio (/Geonor)	Total amount (mm)	Ratio (/Geonor)
Geonor (DFIR)	143.3	1.00	313.7	1.00
RT3 *1	75.9	0.53	182.5	0.58
RT4 *2	93.9	0.66	222.5	0.71
LPM_DF *3	161.4	1.13	384.9	1.23
LPM_DF *3 (Original Algorism *4)	121.0	0.84	247.7	0.79
LPM	206.4	1.44	472.4	1.51
LPM (Original Algorism *4)	166.6	1.16	347.6	1.11
Parsivel	(66.4/83.7)	(0.79)	335.0	1.07
Parsivel (Original Algorism *4)	(84.9/83.7)	(1.01)	422.1	1.35
PWS	132.8	0.93	341.2	1.09

*1 RT3: Heated tipping bucket
*2 RT4: Spill-typed tipping bucket
*3 LPM_DF: LPM installed in another double fence wind shield same as DFIR
*4 It is evaluated with size and vertical velocity disribution (see appendix).

5. THE CHARACTERS AND PROBLEMS IN DATA OBTAINED BY EACH MEASUREMENT

5.1 Geonor

5.1.1 Measurements of three load-cells

Figure 13 shows time series of measurement of snowfall rate by the three load-cells for the two winters. For the 2013/14 winter, two load-cells (S1 and S2) worked well and for the 2014/15 winter, one load-cell (S2) only worked well. One of the reasons of the trouble in measurement may come from remaining the pin, only locking off, at the load-cells. Because same kind of trouble in measurement has not occurred in and after 2015/16 winter, when the pins were taken off from the load-cells (evidences not shown here). For the 2013/14 winter, the average of measurements by S1 and S2 is adopted as the DFIR and for the 2014/15 winter, the measurement by S2 is adopted as the DFIR.

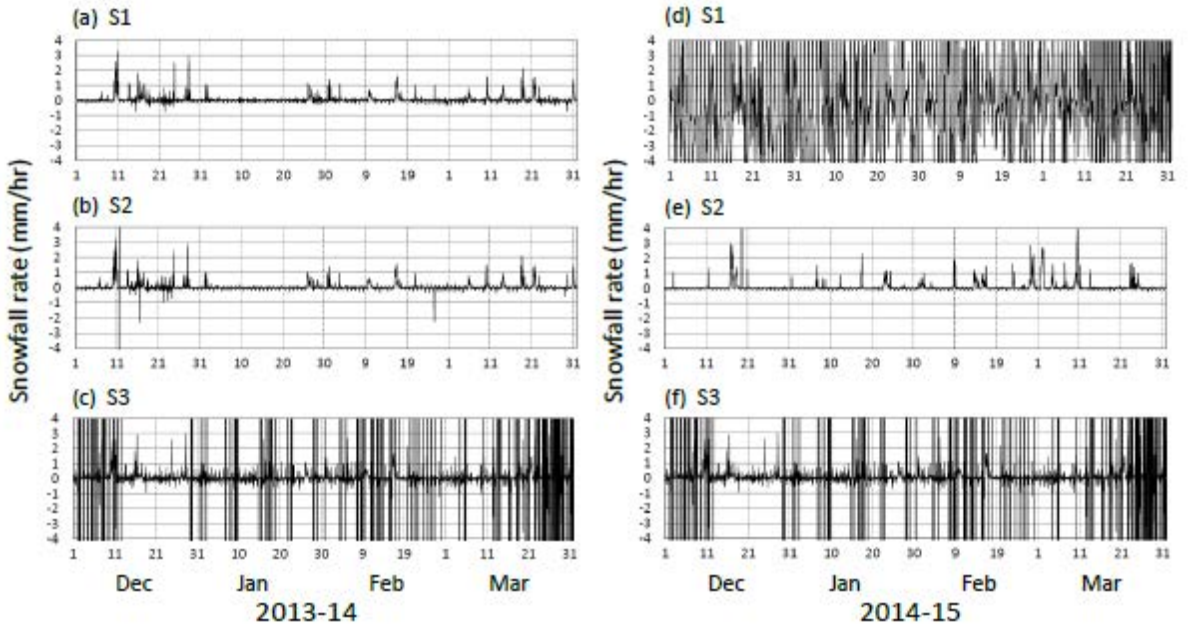


Figure 13 Time series of measurement of snowfall rate by the three load-cells for the two winters.

5.1.2 Wind-induced vibration in time series

Noise in measurement associated with wind were observed at Rikubetsu site. Although the influence of wind will clarified in later analysis, as the time-scale of fluctuation in frequency measured is from sub-second to a few second, running mean around for a minute is expected to be effective to many cases.

5.1.3 Temporal drift synchronized diurnal variation in air temperature

The measured values by Geonor drift in response to ambient air temperature. Figure 14a shows time series of each of the three load-cells for 5-26 January 2014 when there was almost no influence of disturbances. Each time series shows diurnal variation. We focus on S1 and S2 here because S3 has some errors as discussed in previous section. Time series in Figure 14b are of diurnal component of S1 and S2 and of air temperature. The measured values by S1 and S2 are higher in lower temperature and lower in higher temperature. This relationship is more clearly seen in the scatter diagrams of temperature and S1, S2 in Figure 15. The magnitude of the drift synchronized with the diurnal temperature variation is about 0.5 mm/20 °C. It is not a negligible value at Rikubetsu and in polar regions where the snowfall rate is low.

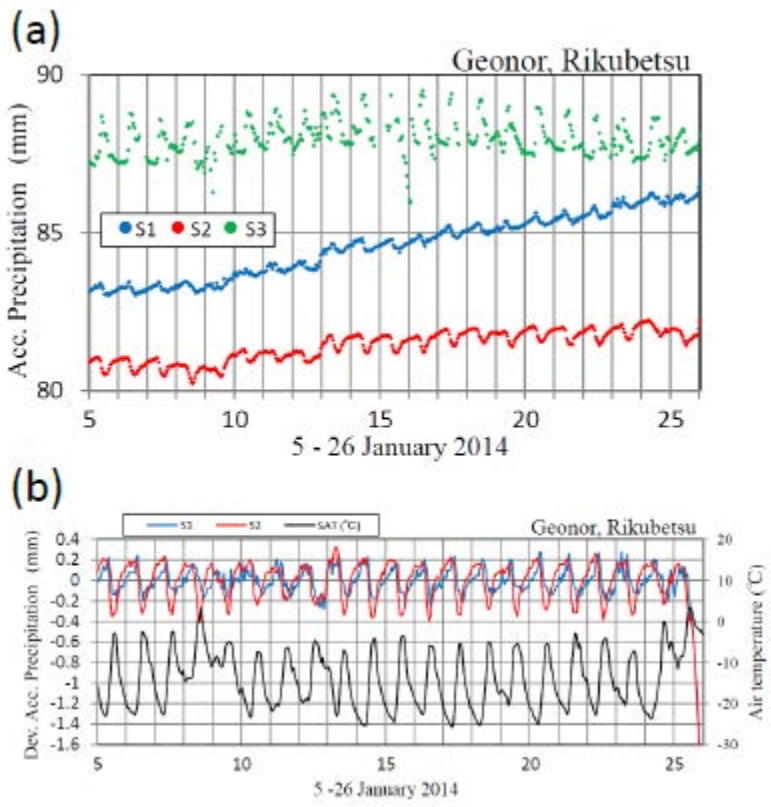


Figure 14 (a) Time series of each of the three load-cells for 5-26 January 2014. (b) Time series of diurnal component of S1(blue line) and S2 (red line) and of air temperature (black line).

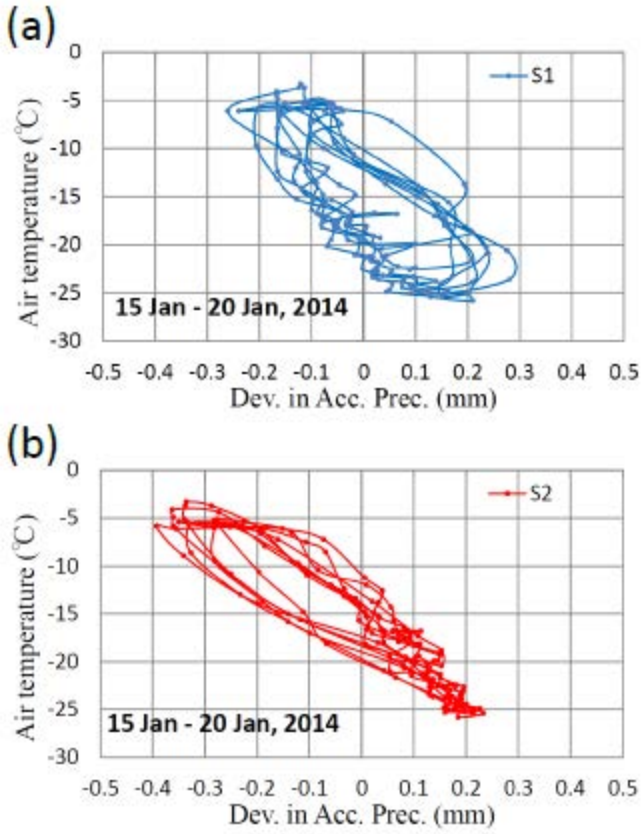


Figure 15 Scatter diagrams of ambient air temperature and diurnal component of (a) S1 and (b) S2.

5.2 Disdrometers

5.2.1 Caution in measurement

In some cases with no snowfall, LPM outputs existence of snow particles in the lowermost classes of size, mainly less than 1 mm in diameter, as shown in Figure 16. The false signals tend to emerge in low temperatures and also tend to disappear in true snowfall event. It is serious problem when we investigate very light snowfall or the features in snow particles for such areas that light snowfall event occurs frequently as Rikubetsu and polar region. Thus, it is desirable to find an algorism to pick up true snowfall from the measurement.

On the other hand, Parsivel does not output data for the two lowermost classes of size, i.e. 0-125 and 125-250 micro meter (see Appendix A). Therefore, Parsivel is detecting 250 micro meter to 26 mm particle size.

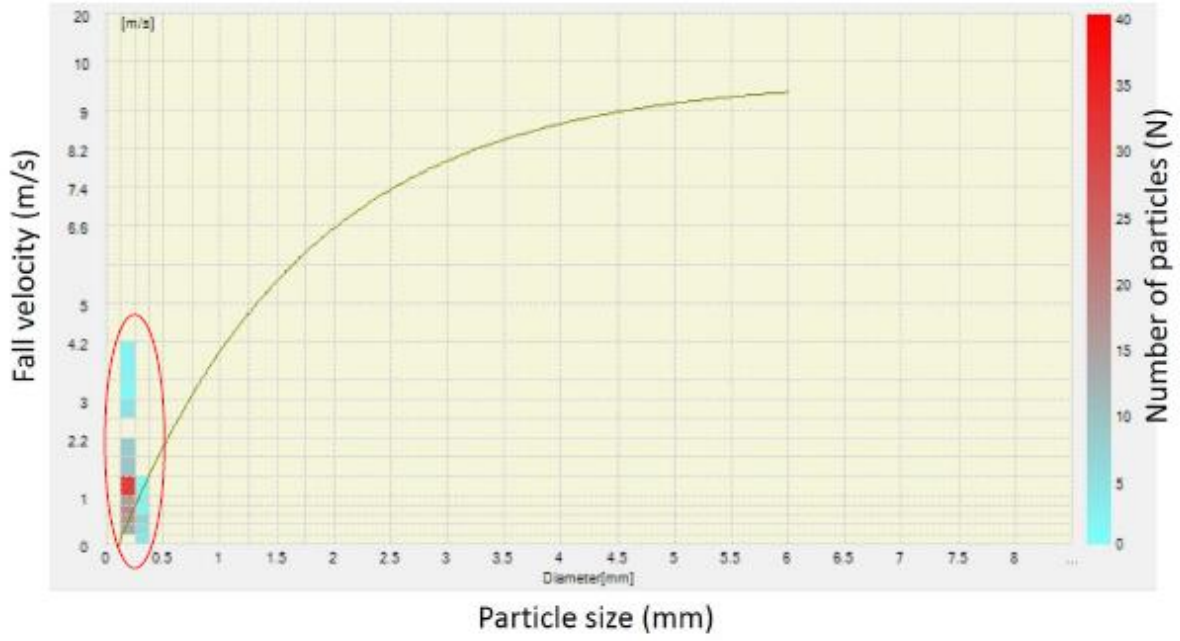


Figure 16 Sample of erroneous counts of snow particles in the lowermost particle sizes in measurement of PLM for a no snowfall case.

5.2.2 Comparison for hourly averaged snowfall rate

Figure 17 shows the comparison of Geonor and LPM, Parsivel, PWS for hourly averaged snowfall rate. Although the values of R2 are almost identical for LPM, Parsivel and PWS, the features of scatter are different from each other. Inclinations of linear fitting line with Parsivel and PWS are almost 1 but PWS shows more scattered data. LPM is overestimating compared to Geonor, and its lower limit value is close to Geonor.

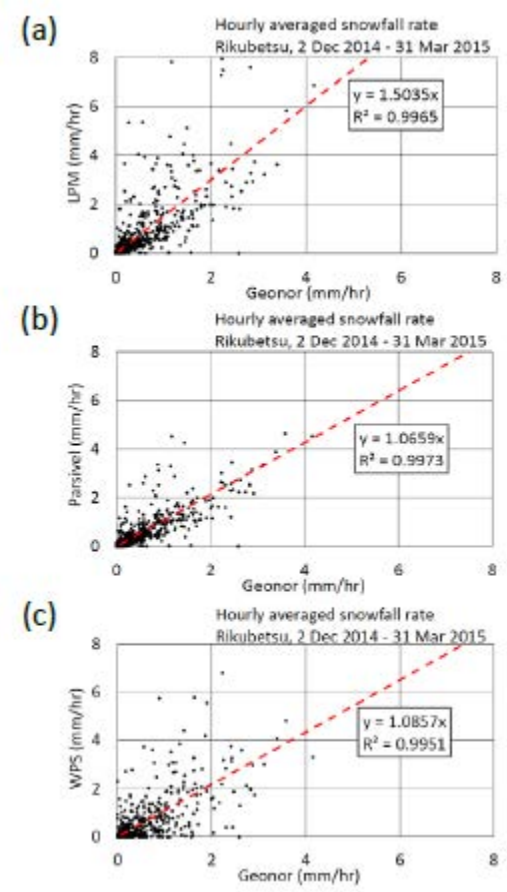


Figure 17 Scattered diagrams for hourly averaged snowfall rate of (a) LPM, (b) Parsivel and (c) PWS compared to Geonor.

5.2.3 Wind effect to the relationship between Geonor and disdrometers

Figure 18 is identical to Figure 17 but for the plot expresses the wind speed, that is, the dots are blue for the wind speed of 2 m/s or less and red for more than 4 m/s. When the wind speed is 2 m/s or less for snowfall rate of 2 mm/hr or less, the measurement by LPM is the closest to Geonor. In that case, the measurement by Parsivel is underestimating in comparison with Geonor and PWS shows higher scatter.

At LPM and Parsivel, many cases of overestimation of snowfall rate are found for wind speed more than 4 m/s. The reason for the difference between disdrometers of LPM and Parsivel and Geonor get to be larger in stronger wind should be solved to derive accurate snowfall rate from disdrometers and also from weighing gauge such as Geonor.

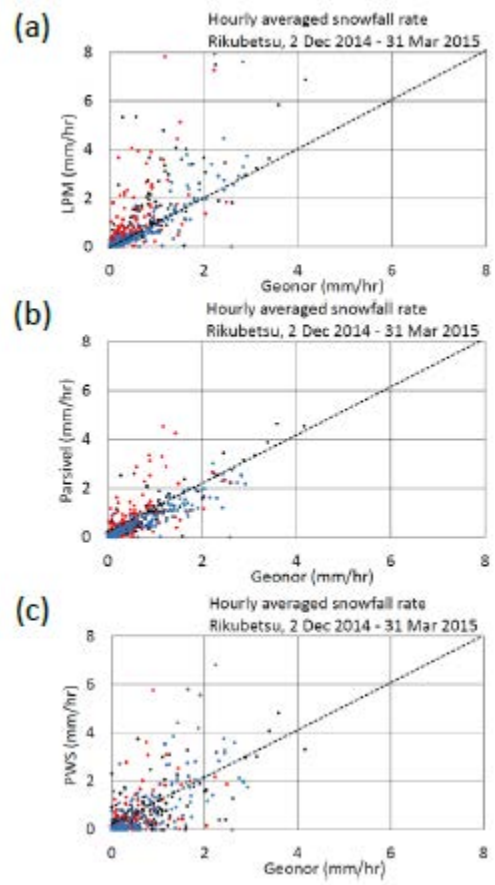


Figure 18 Same to Figure 17 but for the plot expresses the wind speed, that is, the dots are blue for the wind speed of 2 m/s or less and red for more than 4 m/s.

5.3 The tipping gauge, RT3 and RT4

The resolution of RT3 and 4 is 0.5 mm, which is coarse compared with other instruments. The first thing to check is in what extent this resolution can capture the hourly snowfall rate (mm/hr) at Rikubetsu. Figure 19 shows the frequency distribution of hourly snowfall rate and cumulative snowfall rate derived from Geonor. The snowfall rate at 0.5 mm/hr or less is about 86% and at 1 mm/hr or more is about 7%. For snowfall amount (not shown), snowfall rate of less than 0.5 mm/hr contributes about 32% of total snowfall amount at Rikubetsu. Similarly, snowfall rate of 1 mm/hr or more contributes to about 47%. However, there is temperature drift in Geonor, and current analysis does not remove this fake snowfall. This is conceivable as a cause that the frequency of 0.1 mm/hr or less increases in Geonor's measurement result.

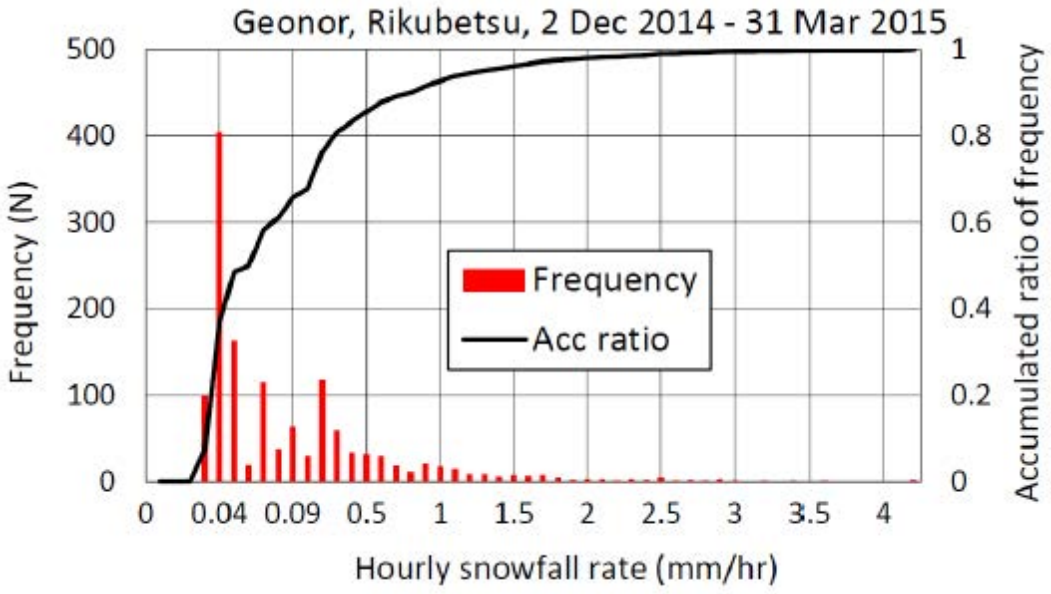


Figure 19 Frequency distribution of hourly snowfall rate (red bar) and cumulative snowfall rate (black line) derived from Geonor.

Regardless, at Rikubetsu, about 80% of the snowfall event has a snowfall rate of 0.5 mm/hr or less. In such events RT3 and 4 will be detected after accumulating snowfall for 1 hour or more. Therefore, we cannot directly compare the hourly data of RT3 and 4 with Geonor. Here, we compare cases where hourly snowfall rate of Geonor is 1 mm/hr or more.

Figure 20 shows the comparison between Geonor and RT3 with the wind speed divided into less than 2 m/s and the wind speed divided into 4 m/s or more. RT3 tends to be smaller than Geonor regardless of wind speed. The difference in the catchment rate of snow particles as a function of wind speed cannot be read clearly from this figure.

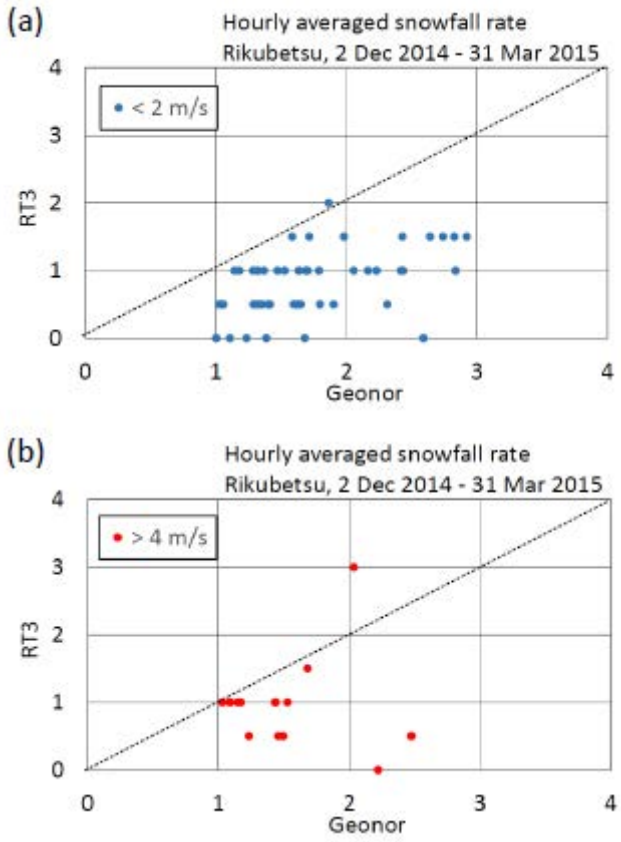


Figure 20 The comparison between Geonor and RT3 with (a) the wind speed divided into less than 2 m/s and (b) the wind speed divided into 4 m/s or more.

Figure 21 is the comparison between Geonor and RT4. RT4 tends to be smaller than the value of Geonor, as RT3. In comparison between Figure 20 and 21, the measurement with wind speed of 2 m/s or less by RT4 is closer to Geonor than RT3.

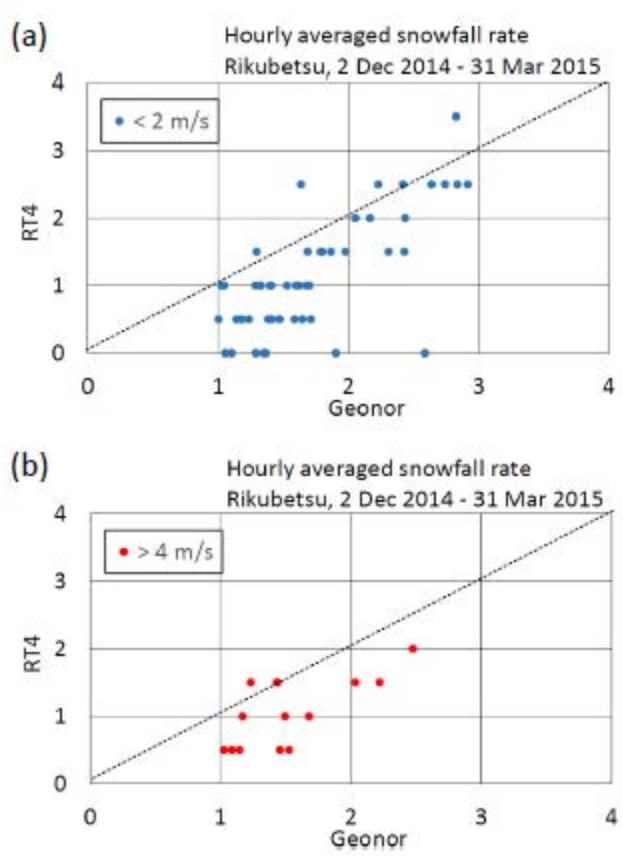


Figure 21 Same as Figure 20 but for RT4.

It is suggested here that it is not possible to capture the one hour resolution of snowfall rate at RT3 and RT 4 at Rikuzuki. It means, at Rikubetsu, to clarify the relationship between wind speed and catchment rate of snow particles of RT3 and 4 is difficult based on hourly data. To do it, it is necessary to target snowfall over longer interval. However, since the wind speed is not constant for a long time, a method of correction considering the fluctuation of the wind speed should be developed.

Acknowledgements

The authors are grateful to participate in SPICE's activities. We wish to express our special gratitude to Dr. Yves-Alain Roulet and Dr. Isabelle Rüedi who have waited for our manuscript to the end.

References

Adam and Lettenmaier, 2003: Adjustment of global gridded precipitation for systematic bias, J. Geophys. Res. Atmosphere, 108(D9), 4257, doi:10.1029/2002JD002499.

Goodison, B. E., P. Y. T. Louie and D. Yang, 1998: WMO solid precipitation measurement intercomparison. WMO Instruments and Observing Methods Rep. 67, WMO/TD-872, 212 pp.

Hirasawa, N., H. Nakamura and T. Yamanouchi, 2000: Abrupt changes in meteorological conditions observed at an inland Antarctic station in association with wintertime blocking formation, Geoph. Res., 27, 1911-1914.

Hirasawa, N. et al., 2013: The role of synoptic-scale features and advection in prolonged warming and generation of different forms of precipitation at Dome Fuji station, Antarctica, following a prominent blocking event. Geoph. Res., 118, 6916-6928.

King, J. C. and J. Turner, 1997: Antarctic Meteorology and Climatology. Cambridge University Press.

Kuhn, M. H., A. J. Reordan and I. A. Wagner, 1975: The climate of Plateau Station. Climate of the Antarctic, 255-267.

Qiu, Jane, 2012: Snow survey hopes for avalanche of data, Nature, 491, 312-313.

Rasmussen, R. and al., 2012: How well are we measuring snow? The NOAA/FAA/NCAR winter precipitation test bed. Bull. Amer. Meteor. Soc., 93, 811-829, doi:10.1175/BAMS-D-11-00052.1.

Reijmer C. H. and M. R. van den Broeke, 2003: Temporal and spatial variability of the surface mass balance in Dronning Maud Land, Antarctica, as derived from automatic weather stations, Journal of Glaciology 49, 512–520.

Schlosser, E., J. G. Powers, M. G. Duda, K. W. Manning, C. H. Reijmer and M. R. van den Broeke, 2010: An extreme precipitation event in Dronning Maud Land, Antarctica: a case study with the Antarctic Mesoscale Prediction System. Polar Res., 29, 330-344, doi:10.1111/ j.1751-8369.2010.00164.x.

Sorai, T., H. Hamada, T. Kameda, S. Takahashi, 2016: Rikubetsu, the coldest town in Japan : according to the meteorological data by Japan Meteorological Agency from 2007 to 2016 in winter seasons. Tenki, 63(11), 879-887.

Tian, X. et al., 2007: Effects of precipitation-bias corrections on surface hydrology over northern latitudes, J. Geophys. Res., 112, D14101.

Walden, V. P., S. G. Warren and E. Tuttle, 2003: Atmospheric ice crystals over the Antarctic plateau in winter, J. Appl. Meteor., 42, 1391-1405.

Welker, C., O. Martius, P. Froidevaux, C. H. Reijimer, and H. Fischer, 2014: A climatological analysis of high-precipitation events in Dronning Maud Land, Antarctica, and associated large-scale atmospheric conditions. J. Geophys. Res., 119, 11932-11954, doi:10.1002/2014JD022259.

Yang, D., 1999a: An improved precipitation climatology for the Arctic Ocean, Geoph. Res., 26, 1625-1628.

Yang, D., 1999b: Bias correction of daily precipitation measurements for Greenland, J. Geoph. Res., 104, 6171-6181.

Measurements at Joetsu (Japan) for the WMO/CIMO SPICE Project

Katsuya Yamashita, Sento Nakai, and Hiroki Motoyoshi

(Snow and Ice Research center, National research institute for earth science and disaster prevention)

1. SITE AND INSTRUMENTATION

The Joetsu site is located in the middle of Honshu main island, Joetsu city, Niigata prefecture, Japan (37°06’56’’N, 138°16’23’’E, 10 m asl). The site is about 6 km away from the coast of Sea of Japan, and is surrounded by paddy fields. Due to the winter monsoon wind from the Siberian air mass and the warm current in Sea of Japan, there is a large amount of precipitation at the site every year. Climate classification of the Joetsu site is Sea of Japan side climate. The area around the Joetsu site is designated in special heavy snowfall zones from the central government of Japan. According to the record from Takada weather station (37°06’24’’N, 138°14’48’’E, 13 m asl) where 2 km apart from the site, the climate normal (1981-2010) of annual mean temperature and precipitation is 13.6 °C and 2755.3 mm, respectively. Snow precipitates from November to April (Figure 1). The climate normal of mean precipitation and maximum snow depth between January and March are 875.3 mm and 122 cm, respectively. The wind speed is not so strong absolutely but is relatively strong in winter. The mean temperature of the coldest month is greater than the freezing point. The ratio of solid precipitation to the winter precipitation varies largely year by year. Thunderstorms and graupel are quite frequent in this season.

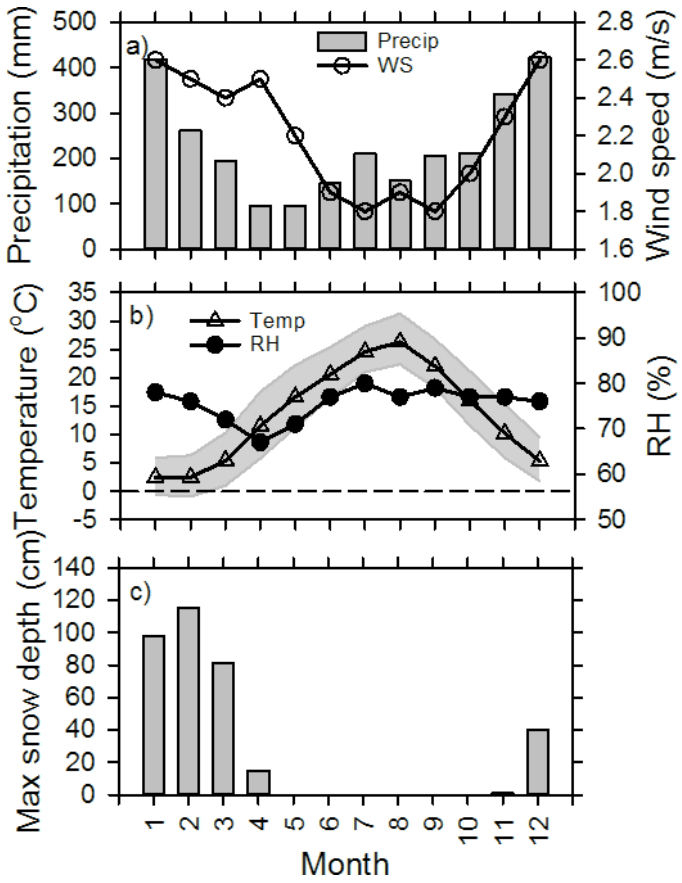


Figure 1 Seasonal variation of the climate normal of meteorological element obtained from Takada weather station managed by the Japan Meteorological Agency.

The Double Fence Intercomparison Reference (DFIR) used in this study is the revised one that was used in the previous WMO Solid Precipitation Measurement Intercomparison (Goodison et al. 1998) as the Hokuriku site (Ohno et al. 1998). The single alter Geonor rain gauge (hereafter DFIR-Geonor: Model: T-200B-MD-3-W) recommended by SPICE was installed at the center of the DFIR to continuously measure precipitation amount. The height of the inlet of the Geonor gauge was set to 3.5 m above the ground so as to match the top height of the inner fence of the DFIR. The wind speed and direction around the DFIR were measured by the windmill type anemometer set to 4 m above ground. Temperature and relative humidity were measured by ventilated thermometer (Model; C-HPT-10-JM, Climatec) and thermohygrometer (Model: HMP155, Vaisala) set to 2.5 m above the ground.

Six precipitation gauges (4 tipping bucket gauges, 1 melted-drop-count gauge, and 1 optical disdrometer) were also installed at Joetsu site to compare precipitation measured by DFIR-Geonor used as the reference instrument. The list is following,

1. Overflow type precipitation gauge with tipping bucket, 314 cm^2 orifice area (Model: RT-4, Yokogawa) with wind shield (hereafter RT-4WS). This gauge collects precipitation to the water reservoir installed in the orifice which is heated to 5 °C. The water equivalent to the

collected precipitation flows down to the measuring unit through the overflow drain gage. The evaporation from the reservoir is protected by an oil layer covering the water surface.

2. Overflow type precipitation gauge with tipping bucket, 314 cm^2 orifice area (Model: RT-4, Meisei) without wind shield (hereafter RT-4).
3. Warm-water precipitation gauge with tipping bucket, 314 cm^2 orifice area (Model: RT-3, Ogasawara) with wind shield (hereafter RT-3WS). This gauge has thick sidewall in which hot water is filled. Hot water is kept to 5 °C.
4. Warm water precipitation gauge with tipping bucket, 314 cm^2 orifice area (Model: RT-3, Ogasawara) without wind shield (hereafter RT-3) The JMA deploys this type at most of observatory which have a probability of solid precipitation in winter.
5. Tamura precipitation intensity meter (hereafter Tamura. Model: SR-2A, Sanyo) counts droplet by photoelectric device. The droplet are produced by the heated funnel (140cm^2 orifice area) which collects rain and snow. Wind shield was not installed.
6. The Laser Precipitation Monitor (hereafter LPM, Model:5.4110.01.000, Thies) measures size and fall speed for the different types of precipitation such as drizzle, rain, hail, snow, sleet. The precipitation type can be defined as a function of drop size and fall speed. A laser-optical beaming source produces a parallel laser beam (infrared 785 nm). A photo diode with a lens is situated on the receiver side in order to measure the optical intensity by transforming it into an electrical signal. When a precipitation particle falls through the laser beam (measuring area 2 cm x 22.8 cm = 45.6 cm^2) the receiving signal is reduced. The diameter of the particle is calculated from the amplitude of the reduction. Moreover, the fall speed of the particle is determined from the duration of the reduced signal.

JMA deploys RT-3 and RT-4 type gauges for observatories at which solid precipitation is possible in winter. These gauges have a resolution of 0.5 mm to measure the precipitation amount. The resolution of Tamura and LPM were 0.005 mm.

The heights of all gauge orifices were adjusted to 3.5 m above the ground. A compact weather sensor (Model: WS600-UMB, Lufft) was also installed near the non-reference gauges to measure the meteorological parameter around the gauge orifices. The heights of sensor were also adjusted to 3.5 m above the ground.

2. OBSERVATION

The observations during three winters (2013/14, 2014/15, and 2015/16) were conducted. The observation periods of each winter were from Jan. 2014 to Mar. 2014, from Nov. 2014 to May 2015, and from Nov. 2015 to May 2016, respectively. Figure 2 shows temperature and precipitation anomaly with the climate normal at Takada station for the observed winters. Temperature anomaly and precipitation amount in Dec. 2014 was lower about 2 °C and higher twice, respectively, than the climate normal. Temperature anomaly in 2015/16 winter was higher than the climate normal, and precipitation amount in Mar. 2016 was 60% of the climate normal.

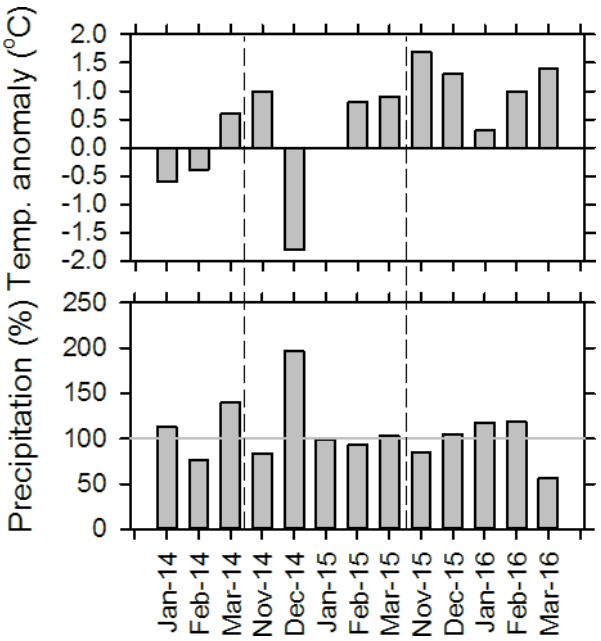


Figure 2 Temperature (top) and precipitation (bottom) anomaly with respect to normal period (1981-2010) at Takada station, the closest station managed by Japan Meteorological Agency to Joetsu site.

The data obtained from all precipitation instruments were recorded by the three data loggers (Model: CR1000, Campbell). The list of the parameters recorded by each loggers were shown in Table 1. A data acquisition interval is 1 minute.

Table 1 List of the parameters recorded by the loggers.

Logger ID	**Obtained parameter**
Logger 1	Wind direction(Young), wind speed (Young), Temperature(Vaisala), Humidity(Vaisala), and Disdrometer (LPM) data
Logger 2	Accumulated precipitation(Geonor)
Logger 3	Wind direction(Lufft), Wind speed(Lufft), Temperature(Lufft), Humidity(Lufft), Pressure(Lufft), Precipitation rate(Lufft、RT-3、RT-3WS、RT-4、RT-4WS、Tamura)

In this report, the results until Mar for each winters were analyzed. The quality controlled data downloaded from “SPICE NCAR website” were used. As the quality control by analysis team were conducted until 7 Mar 2016 at present, the data until that period was used.

3. OBSERVATION RESULTS

Figure 3 shows time series of sensor readings of DFIR-Geonor and summation of data classification flag from the beginning of the observation to the end of March. There were oscillation noise when the sensor reading of the Geonor with 1500 mm capacity exceeded about 1100 mm. In order to reduce the period having the oscillation noise, the precipitation water and antifreezing fluid in the container for collecting precipitation was drained once and twice a winter of 2014/15 and 2015/16, respectively. Analysis in this report were conducted by removing the data with the oscillation noise that corresponds to besides 30 of the summation flag (Table 2).

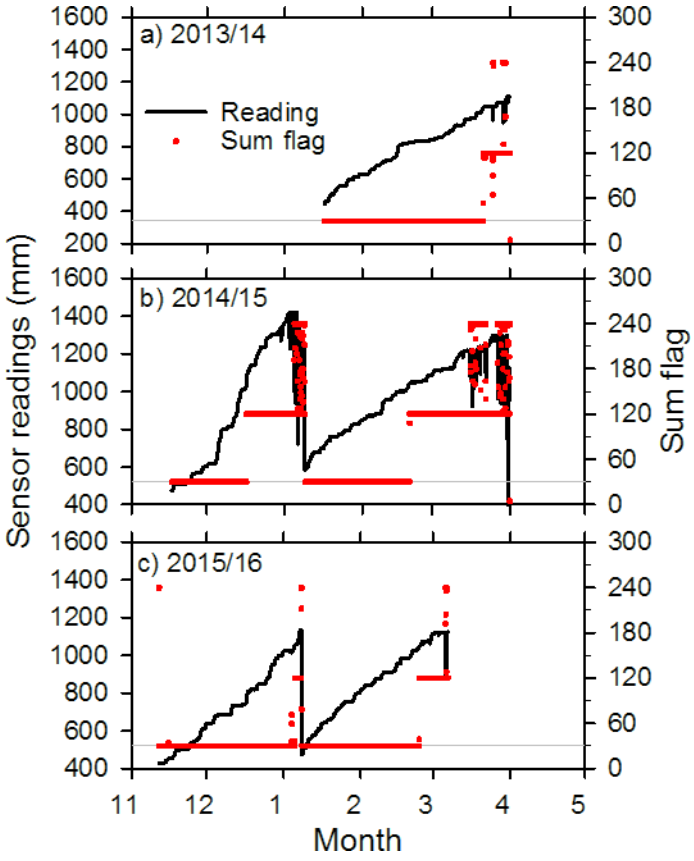


Figure 3 Time series of the sensor reading obtained by Geonor and summation of flag attached by the SPICE data analysis team. The sensor reading and summation of flag is a mean and an integrated value for 30 minutes. A flag was added to one minute interval data. The value 1 of the flag means good data that no issues detected and other values means inconsistent, suspect, erroneous, missing, site maintenance, etc. The value 30 of summation flag means good data.

Table 2 Exclusion period list for data analysis.

Winter ID	Start time (UTC)	End time (UTC)
2013/14Winter	2014/03/21 07:00	2014/03/31 23:59
2014/15Winter	2014/12/16 19:30	2015/01/09 04:30
2014/15Winter	2015/02/19 22:30	2015/03/31 23:59
2015/16Winter	2015/11/12 00:00	2015/11/12 02:00
2015/16Winter	2016/01/05 10:30	2016/01/08 06:30
2015/16Winter	2016/02/24 10:00	2016/03/31 23:59

Figure 4 shows the time series of accumulated precipitation from the beginning of observation to the end of March (except for period indicated in Table 1) for various precipitation gauges. The accumulated precipitations of the DFIR-Geonor in the end of March were 560.7 mm (2013/14), 910.3 mm (2014/15), and 1189.7 mm (2015/16).

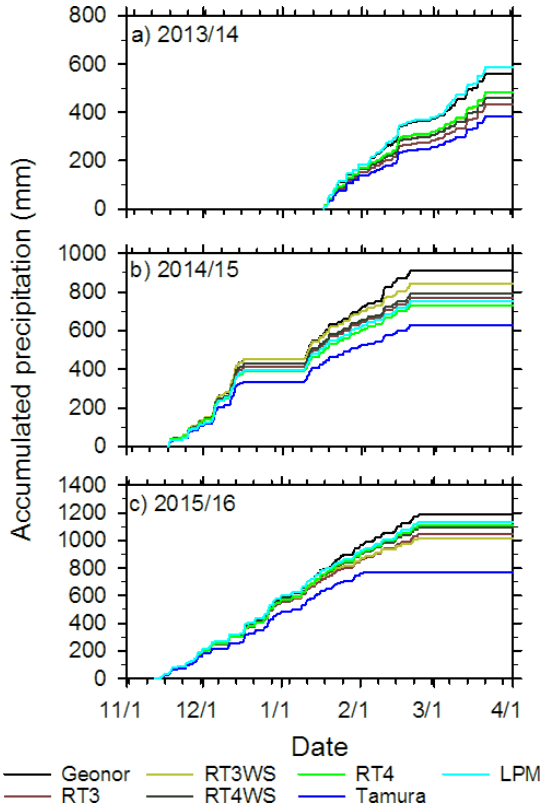


Figure 4 Accumulated precipitations during 2013/14 (a), 2014/15 (b), and 2015/16 (c) winter for a DFIR-Geonor, RT-3, RT-3WS, RT-4, RT-4WS, Tamura, and LPM.

Figure 5 shows the ratio of accumulated precipitation of various gauges to that of the DFIR-Geonor from the beginning of observation to the end of March. The ratios were smaller than 1 for gauges except for the LPM. It suggest that the gauges have collection loss. Especially, the ratios were small when the temperature were less than freezing point. It suggest that the collection loss was large while precipitating snow. The ratio of RT-4 in 2014/15 and RT-3 in 2015/16 were larger than RT-4WS and RT-3WS, respectively. As this results suggest that the collection loss of the gauges with wind shield is smaller than that without wind shield, it is contradictory results. The cause of the contradictory results is being investigated. As the result of the simple check, the contradictory result in 2014/15 and 2015/16 winter was appeared at the observation while all precipitation and raining, respectively. So, the data obtained by RT-4WS in 2014/15 winter while all precipitation and RT-3WS in 2015/16 winter while raining were removed from analysis after Chapter 5 in this report.

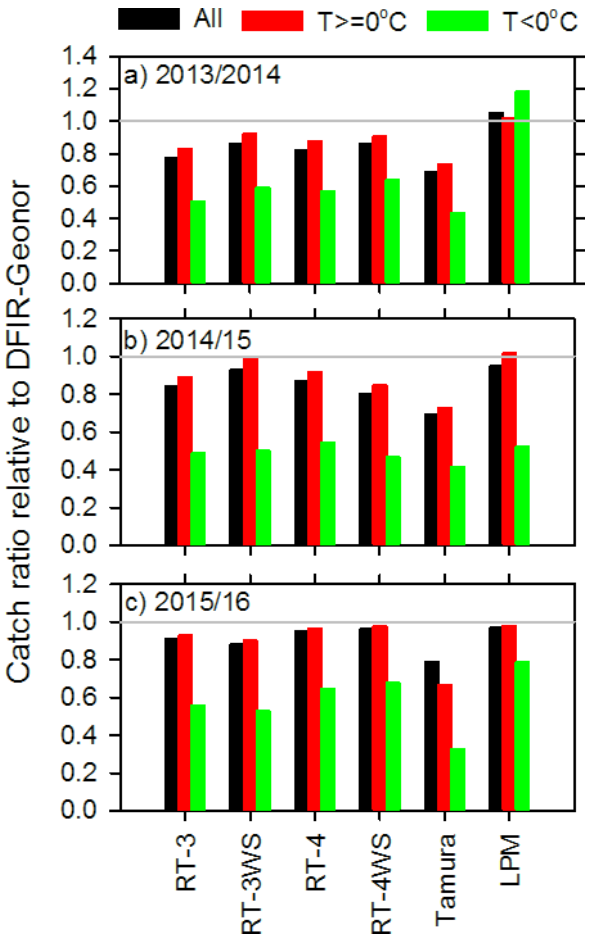


Figure 5 Catch ratio for various precipitation gauges relative to DFIR-Geonor classified for temperature (all, negative, and positive).

4. EVENT SELECTION

The algorithm of the event selection was referred to the documents "SPICE_Reference_Report_V0_1.docx". Figure 6 shows the event selection algorithm flow chart used in this paper. The time of the beginning and end of the event were decided in first step and second step, respectively, using one minute interval data. The accumulated precipitation and statistics (average, standard deviation, etc.) of meteorological parameter (temperature, wind speed, etc.) were calculated by using the data between start and end time of a precipitation event. About the size and fall velocity of precipitation particles obtained by the LPM, the averages of Center of Mass Flux (hereafter CMF), proposed by Ishizaka et al. (2013), for one minute interval were also calculated. CMF is the averages of the size and fall speed from all measured hydrometeors weighted by the mass flux estimated in precipitation particles during arbitrary time. It expresses the predominant type of hydrometeors of the precipitation particles quantitatively.

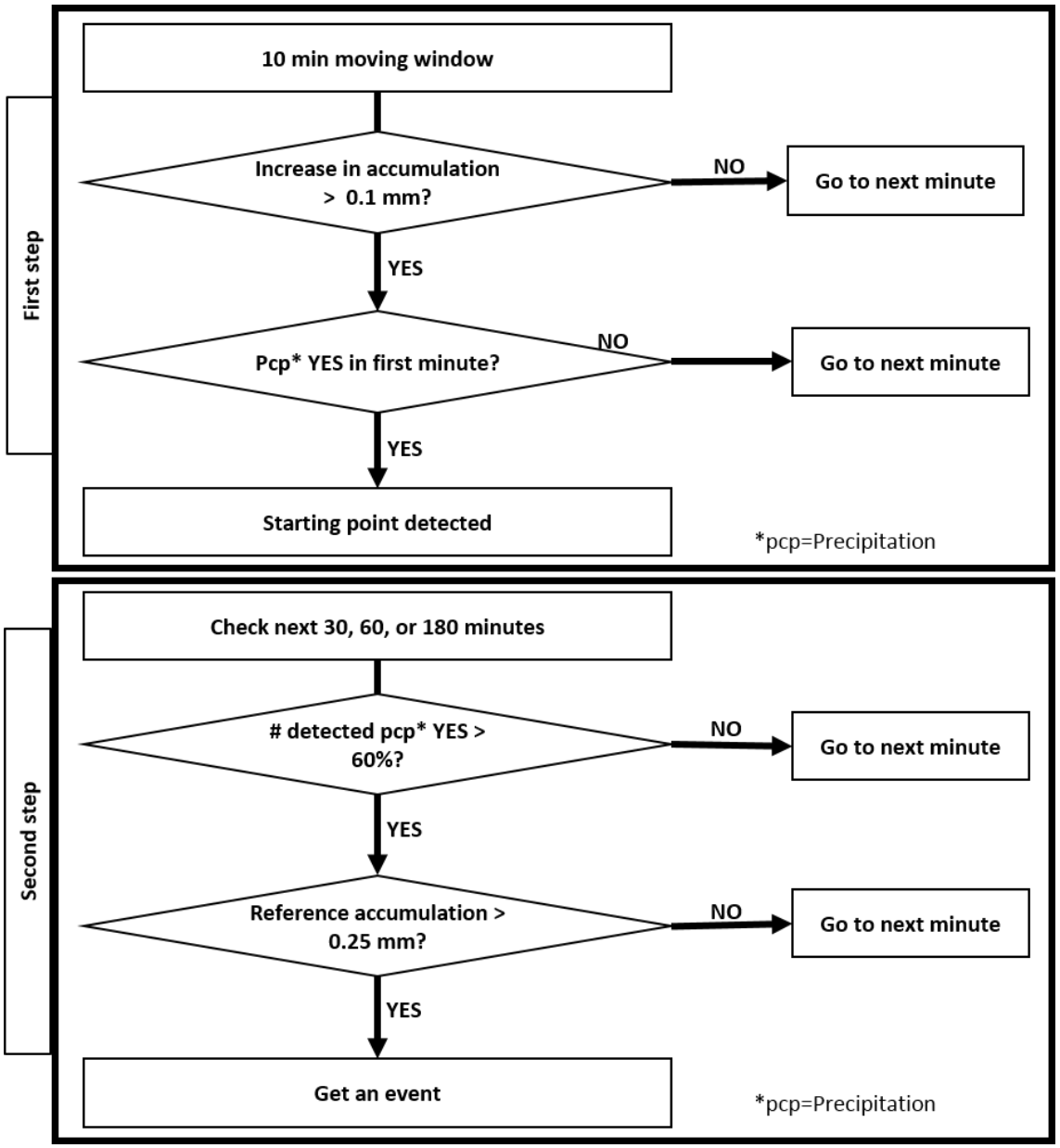


Figure 6 Event selection algorithm flow chart.

Figure 7 shows the histograms of accumulated precipitation for DFIR-Geonor, RT-3, and Tamura extracted by the event selection algorithm assumed a precipitation duration with 30 and 180 min. Although those assumed precipitation duration with 60 min were produced, not shown here. The fraction over accumulated precipitation amount 3 mm was 0.1 (30 min), 11 (60 min), and 96% (180 min). As the resolution of the tipping bucket precipitation gauges used by JMA is 0.5 mm, the data of precipitation duration 180 min were used to analyze and discuss the catch ratio.

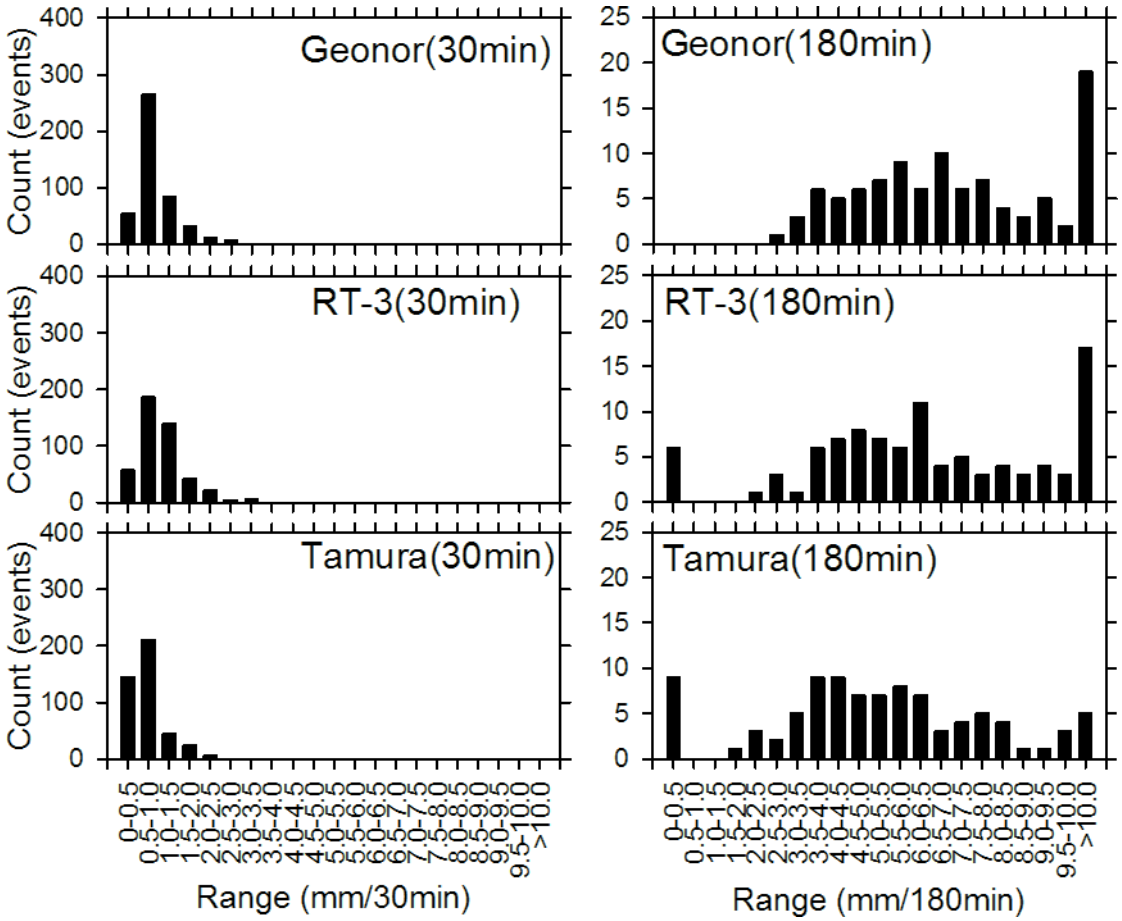


Figure 7 Histogram of the accumulated precipitation extracted by the event selection algorithm for 30 (left) and 180 (right) min.

Figure 8 shows time series of accumulated precipitation, temperature, and wind speed of the precipitation event. The ranges of each element are from 2.7 to 15.8 mm (accumulated precipitation), from -3.2 to 17.2 °C (temperature), and from 0.8 to 6.2 m/s (wind speed).

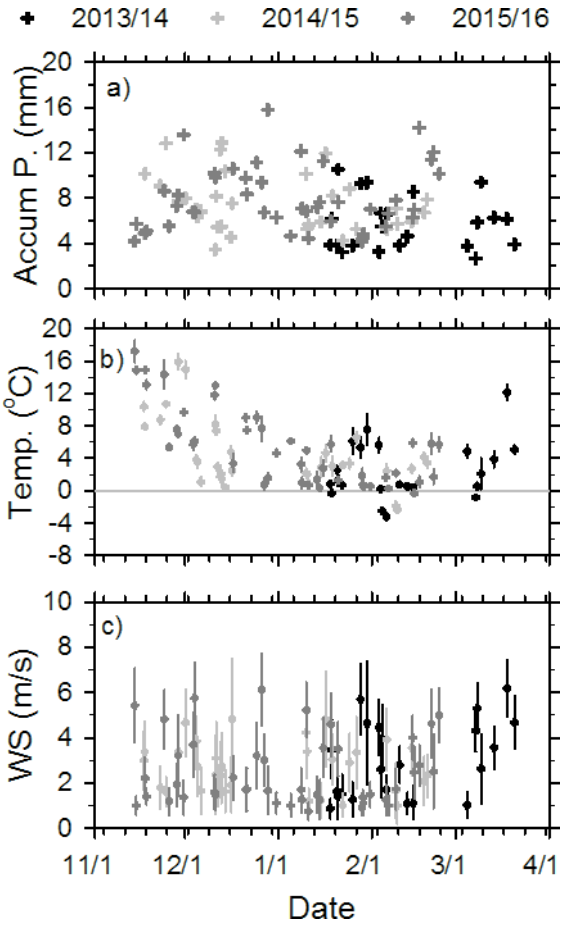


Figure 8 Accumulated precipitation (a), mean temperature (b), and mean wind speed (c) extracted by the event selection algorithm for 180 min.

Figure 9 shows counts of precipitation event for three winters. Those under freezing point were 5 (2013/14), 2 (2014/15), and 2 (2015/16).

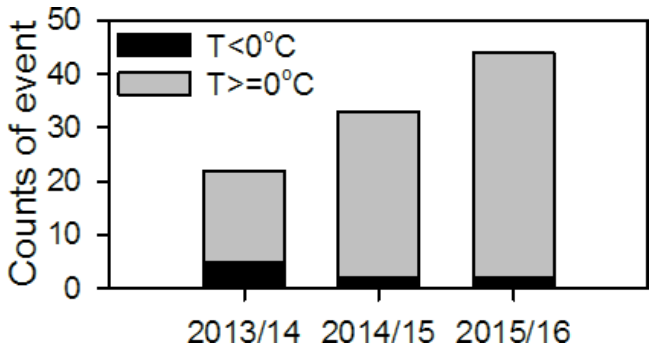


Figure 9 Counts of precipitation event for 2013/14, 2014/15, and 2015/16 winter.

Figure 10 shows histograms of accumulated precipitation, mean wind speed, and mean temperature extracted by the event selection algorithm for 180 min. Amounts above half of cumulative frequency were 6.5 mm (accumulated precipitation), 2 m/s (mean wind speed), and 3 °C (mean temperature).

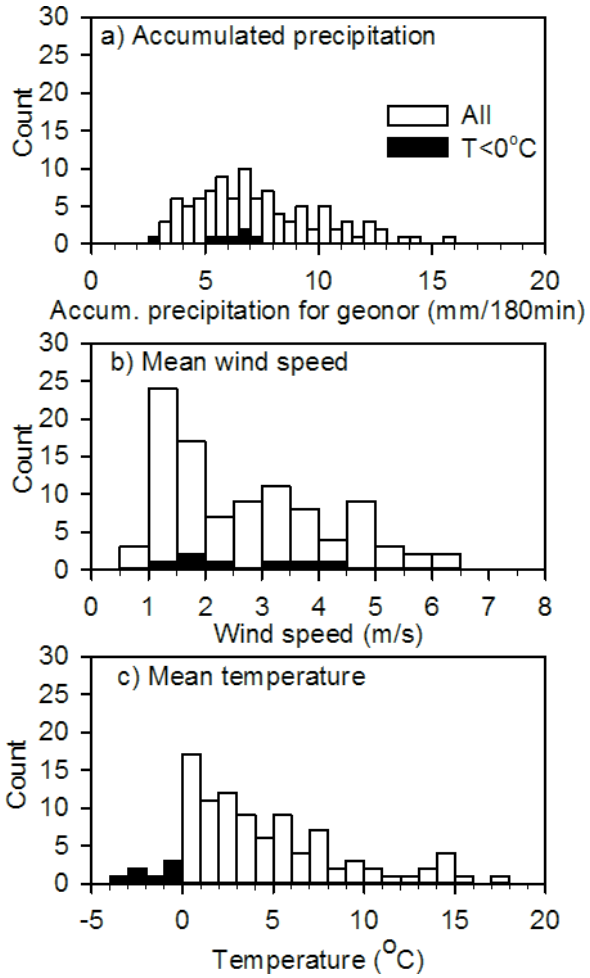


Figure 10 Histogram of accumulated precipitation (a), mean wind speed (b), and mean temperature (c) extracted by the event selection algorithm for 180 min.

5. CATCH RATIO DEPENDENCE TO WIND SPEED

Figure 11 shows the catch ratios versus mean wind speed for various precipitation gauges. All data extracted by event selection algorithm for 180 min were plotted. The mean wind speed measured by the Lufft compact weather sensor used in Figure 11, because the height of instrument above the ground is the same height as the inlet of the precipitation gauges. The curve lines shown in Figure 11 is the regression curve derived by Yokoyama et al. (2003, hereafter YK03). There was no clear relationship between the catch ratios of the LPM and mean wind speed. The catch ratios relative to DFIR-Geonor except for the LPM decreased with increasing of wind speed. This tendency is different in rain and snow like YK03. As the plots between 0 and 2 °C temperature distributed around rain and snow group, it is difficult to distinguish rain or snow only temperature.

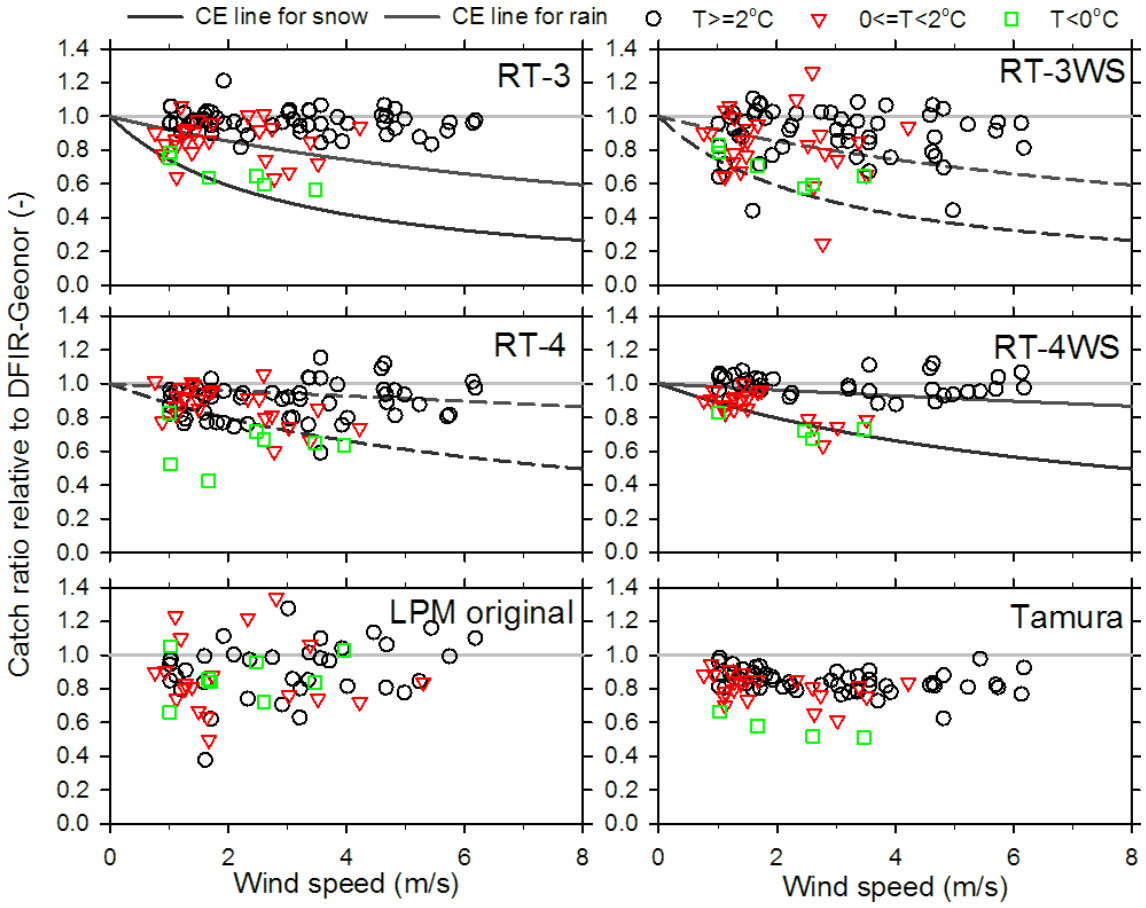


Figure 11 Catch ratio for various precipitation gauges relative to DFIR-Geonor classified for temperature (all, negative, and positive) versus mean wind speed.

Therefore, distinction of rain and snow was performed by using the CMF size and CMF fall velocity. Figure 12 shows the relation of CMFs in the size-fall velocity coordinates for the data extracted by the event selection. Each points express the mean size and fall velocity averaged the 1 minute CMS in the event duration. Rain, graupel, and snow were roughly separated by the location in the size-fall velocity coordinates.

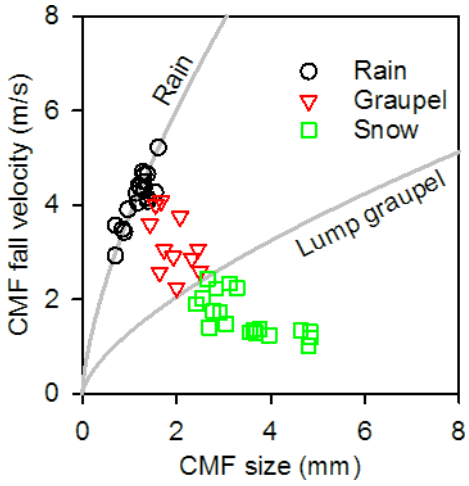


Figure 12 Relation of CMFs (Center of Mass Flux proposed by Ishizaka et al. (2013)) in the size-fall velocity coordinates for the data extracted by the event selection. Curve lines of the rain and lump graupel indicate the empirical curves reported by Atlas et al. (1977) and Locatelli and Hobbs (1974), respectively.

Figure 13 shows the catch ratio for various precipitation gauges relative to DFIR-Geonor classified for precipitation type (rain, graupel, and snow) versus mean wind speed. The data less than 3mm of the accumulated precipitation were removed. Types of precipitation is roughly identified from Figure 13. Following the previous work in YK03, the relationships between catch ratio (CR) and wind speed (U) was assumed to obey the following formula:

$$\mathrm{CR} = \frac{1}{1+mU} \quad (1)$$

where *m* is a parameter assumed to be gauge dependent. The relationship between catch ratio and wind speed and their regressed relationship are shown in Figure 13. The parameters *m*, counts of events, and root mean square error of the regression are tabulated in Table 3.

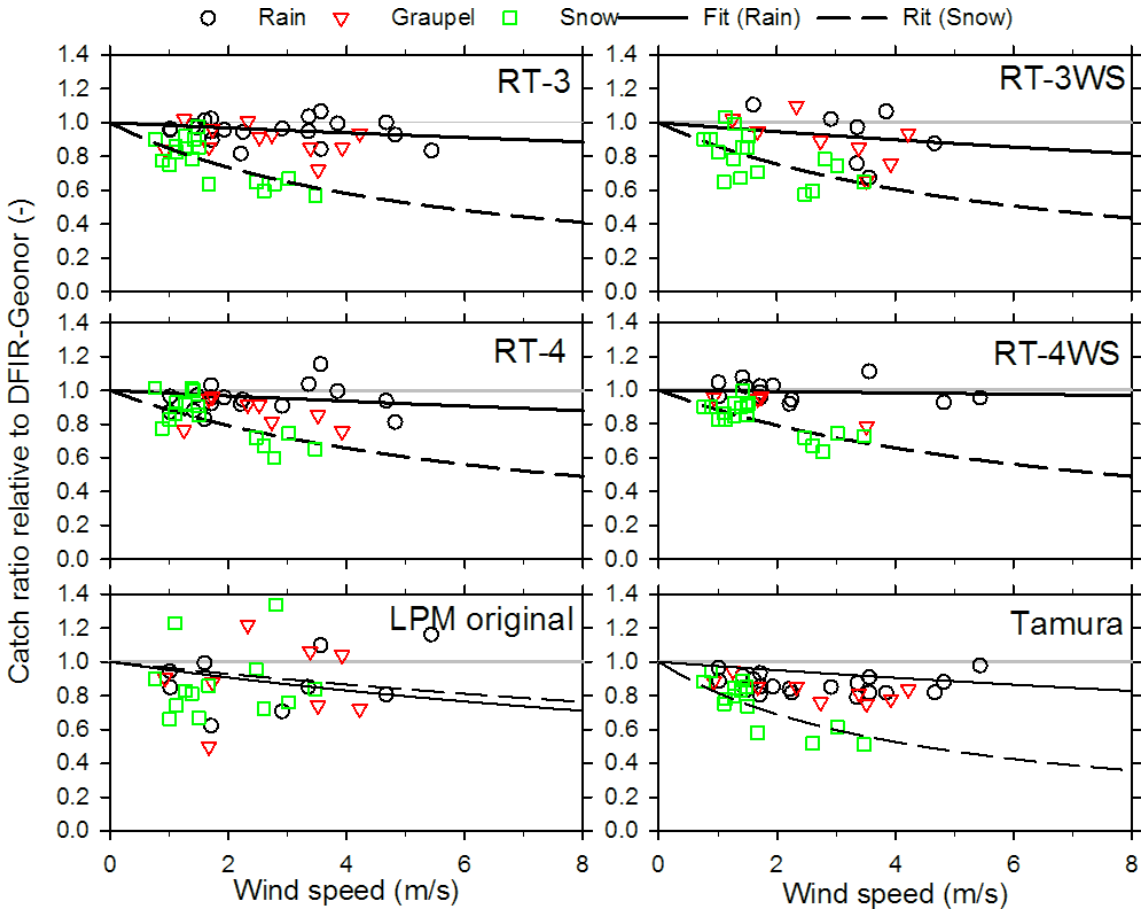


Figure 13 Same as Figure 11 except for precipitation type (rain, graupel, and snow).

Table 3 Result of the regression analysis for catch ratio using the data plotted in Figure 13. The parameter "m" in Eq. (1), number of events and root mean square error for gauge models and precipitation types are listed.

Precip. Type	Parameter	Gauge Type					
		RT-3	RT-3WS	RT-4	RT-4WS	LPM	Tamura
Snow	M	0.180 (0.346)	0.162	0.131	0.131 (0.128)	0.039	0.226
	Events	17 (33)	16	16	17 (43)	14	13
	RMSE	0.081 (0.164)	0.114	0.092	0.070 (0.107)	0.269	0.081
Rain	M	0.016 (0.086)	0.028	0.017	0.004 (0.019)	0.026	0.051
	Events	20 (22)	7	17	13 (23)	9	20
	RMSE	0.066 (0.112)	0.153	0.067	0.058 (0.067)	0.196	0.073

Figure 14 shows the regression curves using the parameter m in Table 2. The curve for the overflow type precipitation gauge with tipping bucket with wind shield (RT-4WS) for snow case is consistent with the report of YK03. On the other hand, that for warm-water precipitation gauges with tipping bucket without wind shield for rain and snow case is higher than the report of YK03. The cause of this difference is not clear at present. It probably become clear by a further analysis and observation.

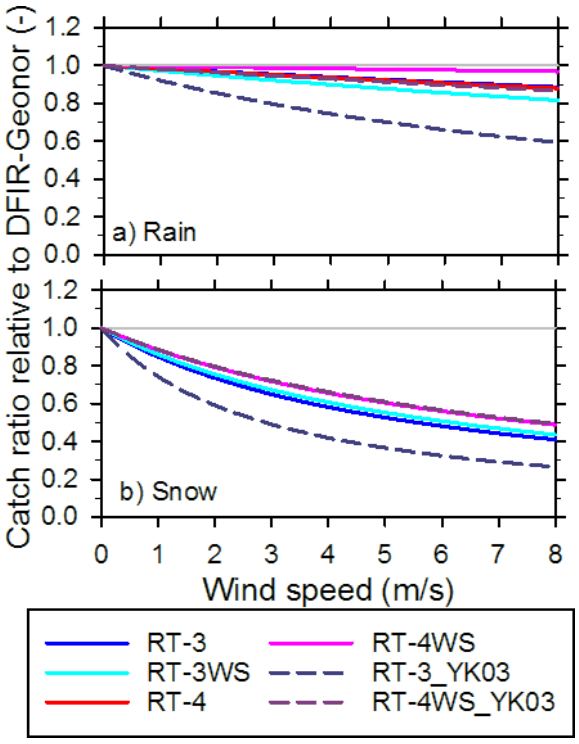


Figure 14 Catch ratio curves for various precipitation gauges using the results of the regression analysis versus mean wind speed. YK03 in legend means Yokoyama et al. (2003) that indicated the regression results for RT-3 and RT-4WS.

6. SUMMARY

From comparative observation among DFIR-Geonor (reference), warm-water precipitation gauge with tipping bucket (RT-3, RT-3WS), overflow type precipitation gauge with tipping bucket (RT-4, RT-4WS), Tamura precipitation intensity meter, and optical disdrometer (LPM) for three winters, the results present that the catch ratios of each precipitation gauges relative to DFIR-Geonor except for the LPM decreased with increasing of wind speed. As the wind speed dependence of the catch ratio is different depending on the precipitation types, the analysis which is classified by the precipitation type estimated by using the temperature or the relation of between size and fall velocity is needed. When the preliminary classification of precipitation type using CMF size and fall velocity was conducted, it was possible to classify into rain and snow. Using the results of the classification, the regression analysis of the relation between the catch ratio and the mean wind speed were

performed like YK03. The results of the regression analysis for the overflow type precipitation gauge with tipping bucket with wind shield (RT-4WS) is consistent with the report of YK03. On the other hand, that for warm-water precipitation gauges with tipping bucket without wind shield is higher than the report of YK03. It probably become clear by a further analysis and observation.

Although the catch ratio about the LPM is not described in this report, it is needed to analyze because it is useful instrument to measure property of precipitation particles automatically and in-situ. As accumulated precipitation and precipitation rate from the LPM is estimated by using size distribution and some assumptions, its amount is different depending on the assumptions. It is needed to analyze the influence of the assumptions for each precipitation types, because the assumptions is different depending on the precipitation type.

REFERENCES

Atlas, D., and C. W. Ulbrich, 1977: Path- and area-integrated rainfall measurement by microwave attenuation in the 1-3 cm band. J. Appl. Meteor., 16, 1322-1331.

Ishizaka, M., H. Motoyoshi, S. Nakai, T. Shiina, T. Kumakura, and K. Muramoto, 2013: A New Method for Identifying the Main Type of Solid Hydrometeors Contributing to Snowfall from Measured Size-Fall Speed Relationship. J. Meteor. Soc. Japan, 91, 747-762.

Locatelli, J. D., and P. V. Fobbs, 1974: Fall speed and mass of solid precipitation particles. J. Geophys. Res., 79, 2185-2197.

Yokoyama, K., H. Ohno, Y. Kominami, S. Inoue, and T. Kawakata, 2003: Performance of Japanese precipitation gauges in winter, Seppyo, 65, 303-316. [in Japanese with English summary.]

For more information, please contact:

World Meteorological Organization

7 bis, avenue de la Paix – P.O. Box 2300 – CH 1211 Geneva 2 – Switzerland

Communication and Public Affairs Office

Tel.: +41 (0) 22 730 83 14/15 – Fax: +41 (0) 22 730 81 71

E-mail: cpa@wmo.int

public.wmo.int