



# Exploratory analysis of citizen observations of hourly precipitation over Scandinavia

Cristian Lussana<sup>1</sup>, Emma Baietti<sup>2</sup>, Line Båserud<sup>1</sup>, Thomas Nils Nipen<sup>1</sup>, and Ivar Ambjørn Seierstad<sup>1</sup>

<sup>1</sup>The Norwegian Meteorological Institute, Oslo, Norway

<sup>2</sup>Università di Bologna, Bologna, Italy

**Correspondence:** Cristian Lussana (cristianl@met.no)

Received: 19 December 2022 – Revised: 17 April 2023 – Accepted: 19 May 2023 – Published: 31 May 2023

**Abstract.** We present a comparison between Netatmo hourly precipitation amounts and observations of the same quantity from weather stations managed by national meteorological services, the latter used as reference values. The empirical distributions of the crowdsourced observations in the surroundings of reference stations are used to assess accuracy and precision of crowdsourced data. We found that reference values are typically within the distribution of the crowdsourced data. However, as the amount of precipitation increases, the spread of the crowdsourced distribution increases and the reference values are more and more frequently found towards the right tail of the distribution. These results indicate that accuracy and precision of crowdsourced data change as precipitation increases. We have studied the sensitivity of our results to the size of the neighbourhood chosen around the reference stations and we show that by aggregating the values over those neighbourhoods, crowdsourced data can be trusted in determining precipitation occurrence. We have assessed the variability of precipitation within small neighbourhoods (of radius 1, 3 and 5 km) and we provide estimates on the basis of the precipitation amounts. Our study quantifies the variability of hourly precipitation over small regions, of the size of the so-called “unresolved spatial scales” in limited area models, based on three years of data collected at several places in Scandinavia.

## 1 Introduction

Observational networks of private weather stations managed by citizens, also known as crowdsourced data, are quickly becoming an important source of opportunistic information on weather that is hard to ignore for national meteorological services. Crowdsourced data opens up the possibility of continuously monitoring precipitation by means of in-situ measurements at km-scale resolutions over much wider regions than using only traditional stations. The use of private weather stations in spatial analysis has also been investigated by several research teams in Europe (e.g. Bárdossy et al., 2021; Mandement and Caumont, 2020; de Vos et al., 2017). All studies emphasize the issues encountered when using crowdsourced data, such as noisy measurements, the difficulties in assessing their representativeness and the ever-present problem of quality control (Alerskans et al., 2022). For instance, specific quality control techniques have been developed for precipi-

tation crowdsourced data by de Vos et al. (2019) and Båserud et al. (2020).

In Norway, the Netatmo network outnumbers the network of traditional stations by a factor of around 50 (Nipen et al., 2020). This information alone provides us with sufficient motivation to further investigate the characteristics of these measurements and, in previous works, we have combined crowdsourced hourly totals, weather radars and NWP model output to improve locally the reconstruction of precipitation (Lussana et al., 2021).

To the best of our knowledge, Netatmo’s rain gauges have not been part of metrological intercomparison studies with other types of rain gauges, such as those conducted by the World Meteorological Organization (WMO) (Colli et al., 2013). Then, it is not possible for us to refer to the scientific literature to obtain specific algorithms to post-process their measurements. Our study is not a metrological intercomparison either. Instead, we adopt the point of view of end users of

observational databases available today within national meteorological services. Crowdsourced data enriches the complexity of those databases. A conscious use of these new data source passes through the characterization of their uncertainty, which is often application-dependent. In many cases, applications such as verification or post-processing of numerical model output make use of reference observations from traditional weather stations, which are operated by national meteorological services according to the high standards described by WMO (2021d). In particular, we focus on Scandinavia and the reference stations considered are those managed by the Norwegian Meteorological Institute (MET Norway), the Swedish Meteorological and Hydrological Institute (SMHI) and the Finnish Meteorological Institute (FMI). The reference dataset is then a composite of different instruments and possibly different technologies to measure precipitation. Nonetheless, for our purpose all these stations are regarded to be of the same high quality, since for our applications they represent the unknown true value of hourly precipitation with satisfactory accuracy and precision.

We aim at assessing the uncertainty of Netatmo's hourly precipitation totals with respect to reference observations from traditional weather stations (i.e. the traditional "target" dataset for several meteorological applications). The main research questions we address are the following: if we consider a reference station and its observations, what is the distribution of the nearby crowdsourced observations? How does this distribution vary with the distance from the reference station?

The confidence we can have in crowdsourced observations originates from the redundancy of the observational network. If we move a little further along this line of thinking, we can ask ourselves if by aggregating crowdsourced observations together we can increase our confidence in the observed value. In fact, this is another of the research questions that will be addressed in this article.

Finally, we use the crowdsourced dense measurement network, which has been operational for several years, for studying the variability of precipitation over short distances. One way or another, numerous applications either require or benefit from information on spatial variability. Not only spatial quality checks require it (Lussana et al., 2010) but also spatial analysis procedures benefit from a realistic specification of observation representativeness errors (Uboldi et al., 2008; Soci et al., 2016), which is a concept strictly linked to the local spatial variability of a field. Stochastic precipitation ensembles that reconstruct daily precipitation, such as that proposed by Frei and Isotta (2019), may also use information on local variability for verification purposes, for instance. Other important applications are those related to the interpretation of numerical model output, either from numerical weather prediction (NWP) or reanalysis models. The local spatial variability can quantify the effects of spatial scales unresolved by the models.

The article is organized as follows. Section 2 describes the data used. Section 3 presents the results obtained. Finally,

in our Conclusions, we highlight the main outcomes of our work.

## 2 Data

WMO provides guidelines, such as WMO (2021d), defining expected accuracy, precision and uncertainty of precipitation measurements, as evaluated against ground-level reference gauges. The amount measured by commonly used gauges at national meteorological services may be less than the actual precipitation reaching the ground by up to 30 % or more. The main source of uncertainty is the error due to "systematic wind field deformation above the gauge orifice: typically 2 % to 10 % for rain and 10 % to 50 % for snow" (WMO, 2021d). There are several other sources of uncertainty, such as the "systematic mechanical and sampling errors, and dynamic effects errors (i.e. systematic delay due to instrument response time): typically 5 % to 15 % for rainfall intensity, or even more in high-rate events". The target uncertainty of precipitation measurements for hydrological purposes (i.e. on a 6 h reporting interval for hydrological forecasting) is reported in WMO (2021c), where it is set to:  $\pm 2$  mm below 40 mm,  $\pm 5$  % above 40 mm.

The Netatmo rain gauges are tipping buckets and their specification can be found online, <https://www.netatmo.com/el-gr/weather/weatherstation/specifications> (last access: 11 April 2023). The size of the bucket is 13 cm  $\times$  13 cm  $\times$  11 cm, the record frequency is 5 min, the resolution is 0.1 mm and the reported accuracy is 1 mm h<sup>-1</sup>, within the measurement range of precipitation rate of 0.2 to 150 mm h<sup>-1</sup>.

It is worth remarking that, although the nominal record frequency of Netatmo's stations is 5 min, it is often seen in the data that the updates are not synchronized between the stations (de Vos et al., 2018). Specifically, Netatmo's stations report hourly precipitation totals approximately every 5 min. For each station, we simply use the hourly value that is closest to the end of each hour, which for us is the timestamp that identifies the hour.

As stated in the Introduction, we use traditional stations, managed by national meteorological services, as our reference stations. Specifically, we have considered only stations having a WMO Integrated Global Observing System identification number (WMO, 2021b, a). The data sources are: MET Norway for Norway; SMHI for Sweden and FMI for Finland. We have used the open-access application programming interfaces of these institutions (see the Section on Data availability).

Netatmo rain gauges are not equipped with heating devices, therefore their measurements are more likely to be affected by gross-measurement errors for temperatures close to zero or negative. In those situations, the typical behavior of a Netatmo's station is that it often does not measure precipitation, even when we know there is precipitation. Fur-

thermore, not all Netatmo rain gauges stations are equipped with temperature sensors. Therefore, in this study, to filter out crowdsourced observations stuck to 0 mm because of cold weather, we require that: (i) our reference dataset includes only quality-controlled hourly precipitation totals registered when the temperature measurement at the same (WMO-compliant) station was higher than 2 °C; (ii) if the reference value is greater than 1 mm, then we exclude from our statistical analysis all those crowdsourced values less than 0.1 mm. This last check is a simple but effective way to decrease the statistical weight in our analysis of those cases where occurrence of precipitation is not measured because of gross-measurement errors, independently of the temperature. Apart from the filtering just mentioned, both reference and crowdsourced data have not been post-processed in any way, for instance they have not been corrected for the wind-induced undercatch.

We have decided not to process crowdsourced data with an automatic quality control procedure because this will inevitably remove some of the highest values, for instance. In general, by applying a quality control procedure our analysis would be affected by our subjective choices on setting up the procedure, which necessarily have to be linked to a specific application. Instead, we decided to show the potential of the data as they are made available to the public by Netatmo. In this way, the results provided are useful for a wide range of applications, including that of designing quality control procedures. The working assumptions allowing us to use the raw crowdsourced data are: (i) most of the crowdsourced data are representative of the atmospheric state, which we have verified by inspecting data samples and through their use in pilot studies (Lussana et al., 2021); (ii) we have to use statistical estimators that are resistant to possible outliers, such as the percentiles and we avoid using extremes (i.e. the 1st or the 99th percentiles).

For tipping-bucket rain gauges, WMO (2021d) recommends to apply intensity-dependent corrections “to improve the accuracy of the intensity measurements and to overcome the underestimation of intensity for high rainfall rates and the overestimation of intensity for low rainfall rates, both of which are typical in non-corrected tipping-bucket gauges”. In the paper by Lanza and Stagi (2008), the bias introduced by water losses during the tipping movement of the bucket is quantified on average as 10 %–15 % at rain intensities higher than 200 mm h<sup>-1</sup>. Examples of correction algorithms applied to tipping-buckets rain gauges are described by Habib et al. (2001), Lanza and Stagi (2009) and Stagnaro et al. (2016). By addressing the research questions reported in the Introduction, we believe we can go one step further in finding a way to adjust Netatmo precipitation data such that they become better comparable with measurements from traditional stations.

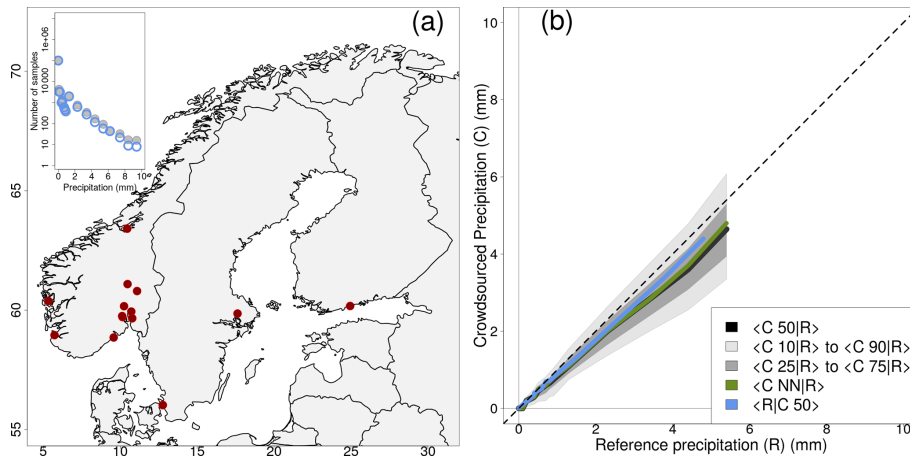
### 3 Results

The number of Netatmo rain gauges available in Scandinavia is increasing constantly over the recent years. For this reason, we consider only the most recent years in our analysis. The results presented are based on hourly measurements from 1 September 2019 to 1 November 2022.

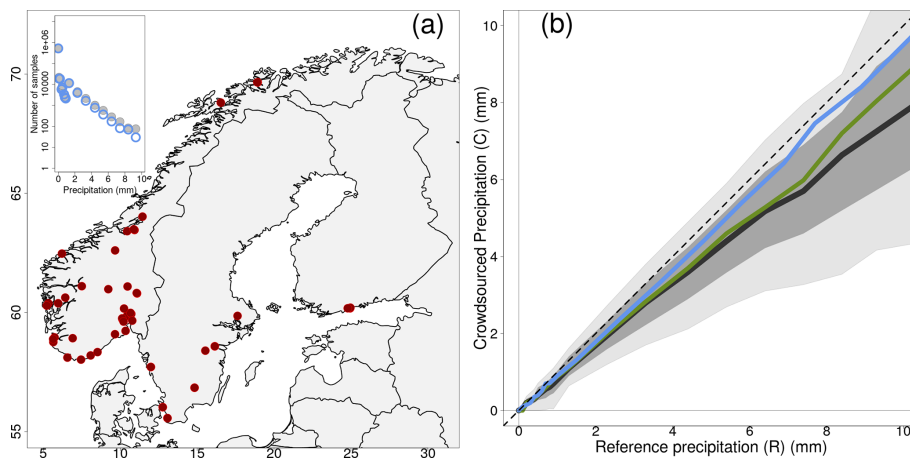
Our investigation is on the quantification of the empirical distribution of crowdsourced data over small circular regions and its comparison with reference observations. The circular regions chosen have radii (indicated by  $r$  in the figures and in the text) of 1, 3 and 5 km. By considering three regions, we can evaluate the sensitivity of the outcomes to the size of the neighbourhood. We consider these regions as small ones because they are representative of meteorological phenomena not well resolved by current numerical models (Frogner et al., 2019; Haakenstad and Øyvind Breivik, 2022). Besides, for atmospheric processes, there is a connection between the aggregation time and the spatial scales represented (Orlanski, 1975; Thunis and Bornstein, 1996). Specifically, when considering hourly precipitation totals, the smaller scales which make sense to study are within the Meso- $\gamma$  scale (i.e. characterized by a minimum horizontal length scale of the order of 1 km).

The maps in Figs. 1a–3a show the spatial distribution of reference stations with enough crowdsourced data within neighbourhoods of 1, 3 and 5 km, respectively. As reported in the captions of the figures, the numbers of stations used are: 15, 51 and 81; for the radii of 1 km (Fig. 1), 3 km (Fig. 2) and 5 km (Fig. 3), respectively. Note that when considering neighbourhoods of 1 km we set a threshold of at least 5 crowdsourced observations, while for 5 and 10 km we raise that threshold to 10. We impose a minimum number of crowdsourced observations such that we can have confidence in the statistics obtained. As expected, the spatial distributions on the maps show that by increasing the neighbourhood size, we get more samples. However, for all three configurations, the reference stations considered are located in densely populated areas and often in the bigger cities. This is not surprising, given that for this type of opportunistic data sources we expect a higher redundancy where most people live.

The procedure used to collect the samples for the study is the following. Given the reference stations in Figs. 1–3, we collect one “sample” for each station every hour. Each sample is a collection of the following values (or records): the reference observed value  $R$ ; the observed value of the nearest crowdsourced rain gauge  $C_{NN}$ ; percentiles from the distribution of the crowdsourced observations, such as: the 10th  $C_{10}$ , the 25th  $C_{25}$ , the median  $C_{50}$ , the 75th  $C_{75}$  and the 90th  $C_{90}$ . We are considering percentiles because they provide more robust (i.e. less dependent on prior assumptions on probability distribution functions that precipitation should follow) and resistant (i.e. less influenced by outliers) estimates (Lanzante, 1996). Then, aggregated statistics of each record over all samples are calculated with several different mathemati-



**Figure 1.** Empirical distribution of crowdsourced hourly precipitation totals ( $C$ ) as a function of reference observations ( $R$ ) based on measurements from 1 September 2019 to 1 November 2022. The crowdsourced stations used lie within circular regions of radius equal to 1 km from the reference stations and there must be at least 5 crowdsourced observations simultaneously available. Panel (a) shows the location of the 15 reference stations and the inset on the top right shows the number of observations as a function of the precipitation classes (see Sect. 3). Panel (b) shows aggregated statistics (i.e. median over all samples) on  $C$  conditional to  $R$ , such as: the crowdsourced observation nearest to the reference (green, NN in the legend stands for nearest neighbour); the median of all observations within the circular region (black); the IQR (i.e. interquartile range, dark gray shaded area); the IDR (i.e. interdecile range, light gray shaded area).



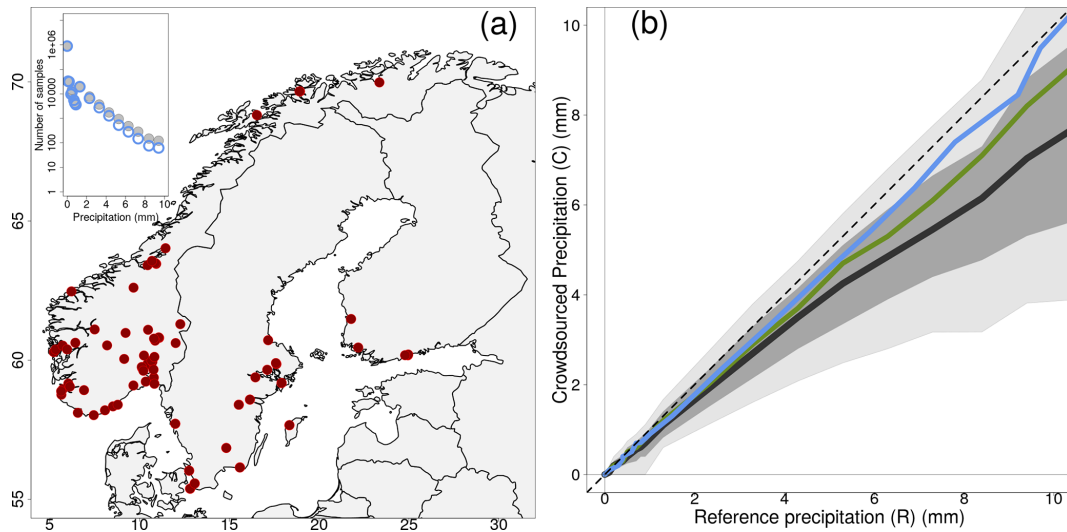
**Figure 2.** Empirical distribution of crowdsourced hourly precipitation totals ( $C$ ) as a function of reference observations ( $R$ ) based on measurements from 1 September 2019 to 1 November 2022. The crowdsourced stations used lie within a circular region of radius equal to 3 km from the reference stations. We have used only those cases when at least 10 crowdsourced observations were simultaneously available. Panel (a) shows the location of the 51 reference stations. The layout is similar to Fig. 1.

cal operators, depending on the specific elaboration or score we want to compute. The aggregated statistics will be indicated with the symbol  $\langle \dots \rangle$  (e.g.  $\langle R \rangle$  indicates the aggregated statistics of the reference observed values over all samples).

Precipitation measurement uncertainties follow a multiplicative error model (Tian et al., 2013), as a consequence our assessment takes into account that observation uncertainty increases with the amount. This leads us to define a number of precipitation classes for hourly precipitation amounts, which we will use to stratify the input samples

and, consequently, the outcomes of our study. The classes with respect to the generic record  $X$  (either one of  $R$  or  $C_{50}$ , as we will see in the following) are (units are mm):  $0 \leq X < 0.1$ ,  $0.1 \leq X < 0.2$ , ...,  $0.9 \leq X < 1$ ,  $1 \leq X < 2$ ,  $2 \leq X < 3$ , ...,  $9 \leq X < 10$ . The whole list of classes is reported in the first column of Tables 1–3.

Considering the climatology of hourly precipitation in Scandinavia, most of the samples should refer to situations of no- or light precipitation (from 0.1 to 2 mm). Then, the number of samples will decrease for increasing precipitation amounts. The exact number of samples in each class is re-



**Figure 3.** Empirical distribution of crowdsourced hourly precipitation totals ( $C$ ) as a function of reference observations ( $R$ ) based on measurements from 1 September 2019 to 1 November 2022. The crowdsourced stations used lie within circular regions of radius equal to 5 km from the reference stations. We have used only those cases when at least 10 crowdsourced observations were simultaneously available. Panel (a) shows the location of the 81 reference stations. The layout is similar to Fig. 1.

**Table 1.** Statistics of the empirical distribution of crowdsourced observations conditional to classes of reference precipitation  $R$  based on measurements from 1 September 2019 to 1 November 2022. The crowdsourced stations used lie within circular regions of radius equal to 1 km from the reference stations and there must be at least 5 crowdsourced observations simultaneously available. The data shown in the table have been used to obtain the graph in Fig. 1b. The first column reports the definitions of the precipitation classes. The second column is the number of samples in a class  $n$ , besides  $n^*$  shown in brackets is the number of samples when the classes are defined with respect to  $C_{50}$  (e.g.  $\dots \leq C_{50}$ )  $\dots$ ). For the symbols in the remaining three columns see Sect. 3. Note that in the column  $\langle C_{NN}|R \rangle$ , the relative difference between  $\langle C_{NN}|R \rangle$  and  $\langle R \rangle$  is reported in brackets (only when  $|\langle C_{NN}|R \rangle - \langle R \rangle| > 0.1$  mm). In the last column, the 5-tuple is the set of percentiles and the closest to  $\langle R \rangle$  is shown in bold (more than one bold value is admissible for ties).

$R$ class (mm)	$n$ ( $n^*$ )	$\langle R \rangle$ (mm)	$\langle C_{NN} R \rangle$ (mm)	$\langle (C_{10}, C_{25}, C_{50}, C_{75}, C_{90}) R \rangle$ (mm)
$0 \leq R < 0.1$	97 835 (99 739)	0.00	0.00 (–)	<b>(0.00, 0.00, 0.00, 0.00, 0.00)</b>
$0.1 \leq R < 0.2$	4321 (3297)	0.10	0.00 (–)	(0.00, 0.00, 0.00, <b>0.10</b> , <b>0.10</b> )
$0.2 \leq R < 0.3$	3433 (3186)	0.20	0.20 (–)	(0.04, 0.10, <b>0.20</b> , 0.28, 0.30)
$0.4 \leq R < 0.5$	1080 (1000)	0.40	0.30 (–)	(0.11, 0.20, 0.30, <b>0.40</b> , 0.51)
$0.5 \leq R < 0.6$	1552 (1173)	0.50	0.50 (–)	(0.18, 0.30, <b>0.45</b> , 0.58, 0.66)
$0.6 \leq R < 0.7$	591 (596)	0.70	0.60 (–)	(0.20, 0.40, 0.60, <b>0.70</b> , 0.82)
$0.8 \leq R < 0.9$	532 (458)	0.80	0.70 (–)	(0.30, 0.50, 0.66, <b>0.80</b> , 0.93)
$0.9 \leq R < 1$	443 (383)	0.90	0.80 (–)	(0.32, 0.58, 0.75, <b>0.90</b> , 1.04)
$1 \leq R < 2$	2104 (1982)	1.30	1.10 (–15 %)	(0.75, 0.91, 1.10, <b>1.31</b> , 1.55)
$2 \leq R < 3$	764 (626)	2.30	2.00 (–13 %)	(1.37, 1.65, 2.00, <b>2.27</b> , 2.63)
$3 \leq R < 4$	345 (274)	3.40	2.83 (–17 %)	(2.00, 2.40, 2.83, <b>3.23</b> , 3.76)
$4 \leq R < 5$	169 (118)	4.40	3.70 (–16 %)	(2.55, 3.08, 3.60, <b>4.14</b> , 4.78)
$5 \leq R < 6$	93 (58)	5.40	4.80 (–11 %)	(3.35, 3.94, 4.65, <b>5.30</b> , 6.08)
$6 \leq R < 7$	46 (44)	6.20	5.03 (–19 %)	(3.86, 4.58, 5.68, <b>6.60</b> , 7.60)
$7 \leq R < 8$	33 (22)	7.40	5.80 (–22 %)	(4.46, 5.23, 6.40, <b>7.53</b> , 8.82)
$8 \leq R < 9$	17 (9)	8.40	8.10 (–4 %)	(3.09, 5.10, 6.70, <b>7.93</b> , 10.12)
$9 \leq R < 10$	16 (8)	9.40	8.40 (–11 %)	(4.85, 6.11, 7.22, 8.56, <b>10.20</b> )
$10 \leq R < 20$	23 (12)	12.20	12.02 (–1 %)	(6.44, 7.90, 10.15, <b>11.63</b> , 13.52)



**Table 2.** Statistics of the empirical distribution of crowdsourced observations conditional to classes of reference precipitation  $R$  based on measurements from 1 September 2019 to 1 November 2022. The crowdsourced stations used lie within circular regions of radius equal to 3 km from the reference stations and there must be at least 10 crowdsourced observations simultaneously available. The data shown in the table have been used to obtain the graph in Fig. 2b. The layout is similar to Table 1.

$R$ class (mm)	$n$ ( $n^*$ )	$\langle R \rangle$ (mm)	$\langle C_{NN} R \rangle$ (mm)	$\langle (C_{10}, C_{25}, C_{50}, C_{75}, C_{90}) R \rangle$ (mm)
$0 \leq R < 0.1$	514 457 (521 327)	0.00	0.00 (–)	<b>(0.00, 0.00, 0.00, 0.00, 0.00)</b>
$0.1 \leq R < 0.2$	21 790 (18 948)	0.10	0.00 (–)	(0.00, 0.00, 0.05, <b>0.10</b> , 0.18)
$0.2 \leq R < 0.3$	19 414 (18 599)	0.20	0.20 (–)	(0.00, 0.10, <b>0.20</b> , 0.3, 0.37)
$0.4 \leq R < 0.5$	5830 (5830)	0.40	0.30 (–)	(0.00, 0.20, 0.30, <b>0.40</b> , 0.55)
$0.5 \leq R < 0.6$	8828 (7005)	0.50	0.40 (–)	(0.00, 0.28, <b>0.40</b> , 0.60, 0.71)
$0.6 \leq R < 0.7$	3357 (3364)	0.70	0.60 (–)	(0.02, 0.33, 0.56, <b>0.71</b> , 0.90)
$0.8 \leq R < 0.9$	2954 (2381)	0.80	0.70 (–)	(0.05, 0.40, 0.61, <b>0.81</b> , 1.00)
$0.9 \leq R < 1$	2531 (2137)	0.90	0.80 (–)	(0.04, 0.45, 0.70, <b>0.90</b> , 1.10)
$1 \leq R < 2$	12 212 (11 493)	1.30	1.11 (–15 %)	(0.64, 0.90, 1.10, <b>1.40</b> , 1.62)
$2 \leq R < 3$	4515 (3972)	2.30	2.00 (–13 %)	(1.20, 1.60, 1.95, <b>2.33</b> , 2.73)
$3 \leq R < 4$	2148 (1655)	3.30	2.83 (–14 %)	(1.70, 2.23, 2.80, <b>3.31</b> , 3.80)
$4 \leq R < 5$	1018 (772)	4.40	3.70 (–16 %)	(2.12, 2.90, 3.60, <b>4.24</b> , 4.90)
$5 \leq R < 6$	575 (367)	5.40	4.60 (–15 %)	(2.69, 3.59, 4.40, <b>5.20</b> , 5.90)
$6 \leq R < 7$	282 (172)	6.40	5.23 (–18 %)	(3.11, 4.22, 5.18, <b>6.20</b> , 7.02)
$7 \leq R < 8$	166 (82)	7.40	5.98 (–19 %)	(3.27, 4.60, 5.70, <b>6.90</b> , 7.98)
$8 \leq R < 9$	84 (71)	8.40	7.20 (–14 %)	(3.54, 5.20, 6.64, <b>7.75</b> , <b>8.74</b> )
$9 \leq R < 10$	76 (30)	9.30	8.04 (–14 %)	(4.17, 5.73, 7.25, <b>8.94</b> , 10.72)
$10 \leq R < 20$	135 (75)	11.6	10.1 (–13 %)	(4.57, 7.10, 8.84, <b>10.75</b> , 13.04)

**Table 3.** Statistics of the empirical distribution of crowdsourced observations conditional to classes of reference precipitation  $R$  based on measurements from 1 September 2019 to 1 November 2022. The crowdsourced stations used lie within circular regions of radius equal to 5 km from the reference stations and there must be at least 10 crowdsourced observations simultaneously available. The data shown in the table have been used to obtain the graph in Fig. 3b. The layout is similar to Table 1.

$R$ class (mm)	$n$ ( $n^*$ )	$\langle R \rangle$ (mm)	$\langle C_{NN} R \rangle$ (mm)	$\langle (C_{10}, C_{25}, C_{50}, C_{75}, C_{90}) R \rangle$ (mm)
$0 \leq R < 0.1$	882 111 (891 622)	0.00	0.00 (–)	<b>(0.00, 0.00, 0.00, 0.00, 0.00)</b>
$0.1 \leq R < 0.2$	34 588 (32 724)	0.10	0.10 (–)	(0.00, 0.00, 0.05, <b>0.10</b> , 0.20)
$0.2 \leq R < 0.3$	32 718 (31 332)	0.20	0.20 (–)	(0.00, 0.1, <b>0.2</b> , 0.30, 0.40)
$0.4 \leq R < 0.5$	10 054 (9741)	0.40	0.30 (–)	(0.00, 0.20, 0.30, <b>0.43</b> , 0.60)
$0.5 \leq R < 0.6$	14 792 (11 691)	0.50	0.50 (–)	(0.00, 0.25, <b>0.40</b> , 0.60, 0.74)
$0.6 \leq R < 0.7$	5537 (5515)	0.70	0.60 (–)	(0.00, 0.30, 0.56, <b>0.73</b> , 0.90)
$0.8 \leq R < 0.9$	4996 (4179)	0.80	0.70 (–)	(0.00, 0.40, 0.61, <b>0.83</b> , 1.01)
$0.9 \leq R < 1$	4153 (3649)	0.90	0.80 (–)	(0.00, 0.40, 0.70, <b>0.90</b> , 1.11)
$1 \leq R < 2$	2049 (19 552)	1.30	1.20 (–)	(0.60, 0.83, 1.10, <b>1.40</b> , 1.67)
$2 \leq R < 3$	7781 (6613)	2.30	2.00 (–13 %)	(1.11, 1.50, 1.90, <b>2.32</b> , 2.73)
$3 \leq R < 4$	3656 (2749)	3.30	2.90 (–12 %)	(1.61, 2.15, 2.70, <b>3.30</b> , 3.80)
$4 \leq R < 5$	1748 (1273)	4.30	3.70 (–14 %)	(2.08, 2.80, 3.50, <b>4.15</b> , 4.75)
$5 \leq R < 6$	940 (529)	5.30	4.70 (–11 %)	(2.48, 3.35, 4.25, <b>5.10</b> , 5.80)
$6 \leq R < 7$	485 (282)	6.30	5.30 (–16 %)	(2.80, 3.90, 4.85, <b>5.90</b> , 6.80)
$7 \leq R < 8$	286 (148)	7.30	6.10 (–16 %)	(3.17, 4.40, 5.45, 6.65, <b>7.76</b> )
$8 \leq R < 9$	149 (76)	8.40	7.10 (–15 %)	(3.17, 4.78, 6.15, 7.30, <b>8.79</b> )
$9 \leq R < 10$	122 (62)	9.40	8.20 (–13 %)	(3.82, 5.31, 7.04, <b>8.81</b> , 10.43)
$10 \leq R < 20$	222 (105)	11.60	10.10 (–13 %)	(3.98, 6.00, 8.44, 10.49, <b>12.50</b> )

ported in Tables 1–3 (second columns) and it is shown in the upper-right insets of Figs. 1a–3a (the  $y$ -axis has a logarithmic scale). Note that when  $R$  is used to distinguish between the classes (i.e.  $X = R$  in the definition of classes above), this corresponds to  $n$  in the tables and the gray dots in the figures. Alternatively, when  $X = C_{50}$ , this corresponds to  $n^*$  in the tables and blue dots in the figures. It is worth remarking that not all results reported in the tables are shown in the figures. In particular, in the figures, we show values between 0 and 10 mm and we require at least 60 samples ( $n > 60$ ) within a class in order to show the aggregated statistics. The choice is based on the fact that in the figures we do not want to compare with each other values characterized by rather different uncertainties. In the tables, the choice is left to the readers.

In Tables 1–3, for the class  $0 \leq X < 0.1$  (i.e. first row in the tables)  $n^*$  is always higher than  $n$ . Then, for classes where the maximum precipitation amount is below 1 mm, we still have some cases when  $n^*$  exceeds  $n$ . This never happens for classes referring to amounts higher than 1 mm and the deviations become greater as the amount increases. This mismatch between crowdsourced and reference precipitation in the frequency of occurrence of several classes, especially those with more intense precipitation, will be further investigated in Sect. 3.1.

### 3.1 Comparison of crowdsourced data against traditional observations

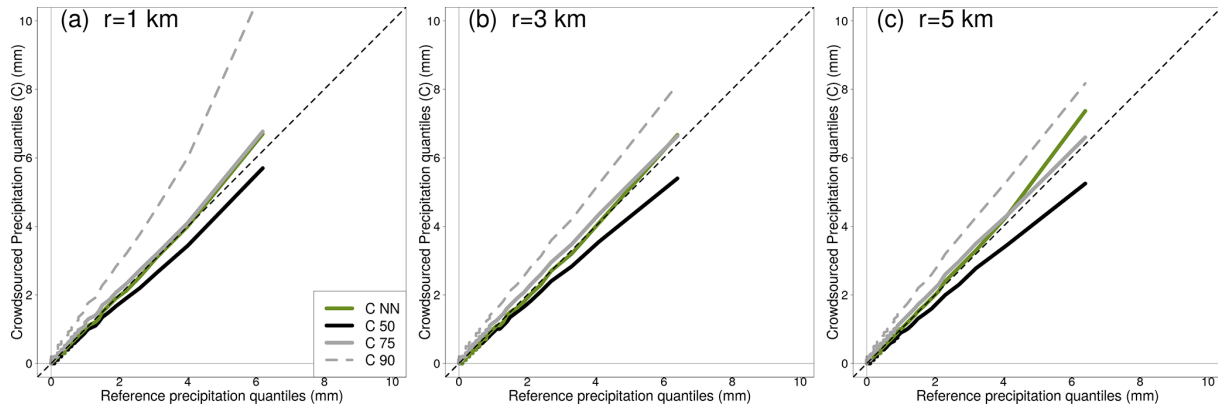
The distributions of crowdsourced precipitation conditional to (classes of) reference precipitation amounts have been computed and they are reported in Tables 1–3. Besides, in Figs. 1b–3b, the black lines and the gray shaded regions show an estimate of the distribution of crowdsourced precipitation conditional to reference precipitation amounts (i.e.  $R$  is now a continuous range of values, instead of a set of discretized classes). Figures show “estimates” obtained from the data in the tables, in the sense that we begin our elaboration by classifying our samples with respect to  $R$ . Then, for each of the classes, the aggregated statistics over all samples for every record is obtained using the median as the aggregation operator. This procedure is indicated with the following notation e.g.  $\langle C_{50}|R \rangle$  that stands for: the median over all samples (i.e.  $\langle \dots \rangle$ ; remember one sample corresponds to one hour) of the median of the crowdsourced observations ( $C_{50}$ ) within a circular region surrounding a reference station, when we select only those samples belonging to a specific precipitation class ( $R$ ). The aggregated statistics are then reported in Tables 1–3, in the third ( $\langle C_{NN}|R \rangle$ ) and fourth columns (the 5-tuple  $\langle C_x|R \rangle$ , with  $x = \{10, 25, 50, 75, 90\}$ ). In the second column,  $\langle R \rangle$  is the median over all samples of all  $R$  values within a class. From the tables, we build the figures. Let us consider the thick black line in Fig. 1, which show  $\langle C_{50}|R \rangle$  when  $r = 1$  km, the line is obtained by joining together the pairs of points ( $\langle R \rangle$ ,  $\langle C_{50}|R \rangle$ ) in Table 1. Similar procedures apply for all the other lines and for the

shaded regions. The light gray region spans the area between  $\langle C_{10}|R \rangle$  and  $\langle C_{90}|R \rangle$  (i.e. the interdecile range or IDR in brief). The dark gray region spans the area between  $\langle C_{25}|R \rangle$  and  $\langle C_{75}|R \rangle$  (i.e. the interquartile range or IQR). The two shaded areas give an indication of the variability expected on crowdsourced data given a reference precipitation amount; we will explore these aspects in more detail in Sect. 3.2. It is however worth noting that the IDR and the IQR are rather symmetric around  $\langle C_{50}|R \rangle$ . The graphs of Figs. 1b–3b show something more: the green line is  $\langle C_{NN}|R \rangle$  and the blue line is  $\langle R|C_{50} \rangle$ . The latter is an aggregated statistics conditional to the crowdsourced precipitation amounts, which is defined as the median of  $R$  over all samples within a circular region surrounding a reference station, when we select only those samples with  $C_{50}$  belonging to a specific precipitation class. Sometimes, the lines in Fig. 1 do not cover the whole range of reference precipitation values because for higher amounts we do not have enough samples (see Sect. 3).

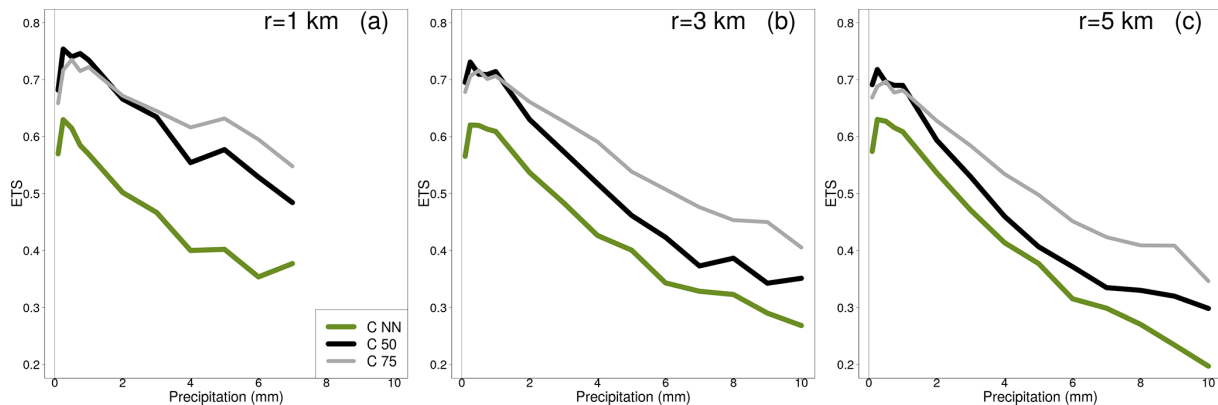
The results in Tables 1–3 show that  $\langle R \rangle$  is always included in the IDR of the crowdsourced observations and often it is within the IQR. However, we can notice a sort of drift in the positioning of  $\langle R \rangle$  within the distributions. In the first lines,  $\langle R \rangle$  stands close to the median  $\langle C_{50}|R \rangle$ , then it gradually moves towards the higher percentiles. For instance,  $\langle R \rangle$  is in between  $\langle C_{50}|R \rangle$  and  $\langle C_{75}|R \rangle$  up until the classes:  $8 \leq R < 9$  ( $r = 1$  km);  $7 \leq R < 8$  ( $r = 3$  km);  $6 \leq R < 7$  ( $r = 5$  km). Then, for even higher amounts,  $\langle R \rangle$  falls often between  $\langle C_{75}|R \rangle$  and  $\langle C_{90}|R \rangle$ . The “drifting” of  $R$  within the crowdsourced distributions is also shown in Figs. 1–3 (i.e. the gradual increase in the deviation between the dashed and the thick black lines, as measured using the gray shaded areas as references). In the ideal situation of the reference and crowdsourced precipitation being random variables having the same probability density function, one should expect the line of  $\langle C_{50}|R \rangle$  to lie on the identity line (the black dashed line) and the gray regions to be centered on that line too. This is not a bad assumption for classes of light precipitation (vast majority of the cases) but it becomes an increasingly less good approximation as the amount increases.

The blue lines in Figs. 1–3, on the other hand, show that  $\langle R|C_{50} \rangle$  stay close to the identity line even for the highest precipitation classes. This means that we can rely on  $C_{50}$  as an estimator of  $R$  across the range of precipitation amounts. However, for the higher values there is a systematic underestimation in the frequency of occurrence of those events, as pointed out in Sect. 3 when discussing on the differences between  $n$  and  $n^*$  in Tables 1–3.

Since the aggregated statistics in the last column of Tables 1–3 involve regionalization of point values into values representative of an area, then part of the underestimation may be due to the smoothing inherent in the regionalization processes (or “conditional bias”; Wilks, 2019). It is then interesting to consider  $\langle C_{NN}|R \rangle$  (fourth column in the tables or the green lines in the figures), since in this case we are



**Figure 4.** Quantile–quantile (Q–Q) plots comparing reference and crowdsourced hourly precipitation. The datasets are the same used for Figs. 1–3 and the three panels refer to the three neighbourhood sizes  $r = 1, 3, 5$  km. The crowdsourced variables are listed in the legend, with reference to Sect. 3 for the meaning of the symbols.



**Figure 5.** Equitable threat score (ETS) comparing reference and crowdsourced hourly precipitation. The datasets are the same used for Figs. 1–3 and the three panels refer to the three neighbourhood sizes reported ( $r$  equal to 1, 3 and 5 km). The green line shows the ETS for the crowdsourced observation nearest to the reference. The black and the gray lines show the ETS for selected percentiles of the crowdsourced observations within the neighbourhood.

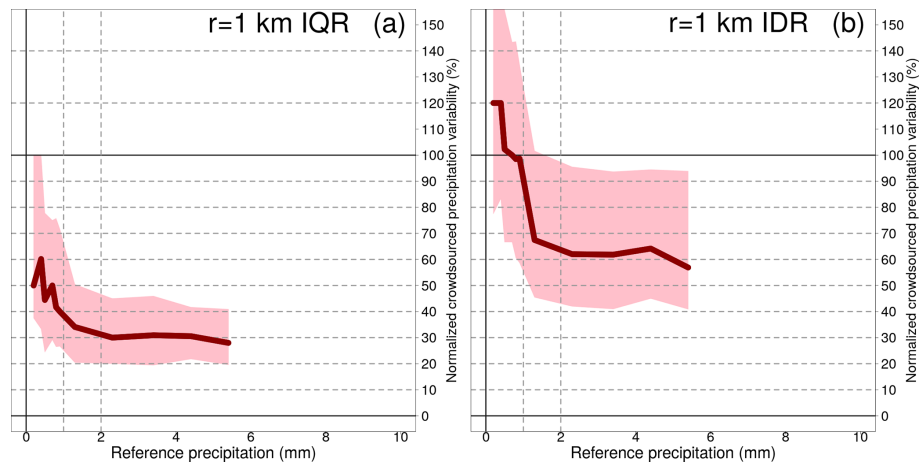
comparing point values against point values and we should not expect conditional biases. The tables show that  $\langle R \rangle$  and  $\langle C_{NN}|R \rangle$  do have very similar values (between  $\pm 0.1$  mm) for precipitation classes below: 1 mm ( $r = 1$  and  $r = 3$  km); 2 mm ( $r = 5$  km). Then,  $\langle C_{NN}|R \rangle$  underestimates  $\langle R \rangle$  and the relative differences are within the range of values:  $-22\%$  and  $-1\%$  ( $r = 1$  km);  $-19\%$  and  $-13\%$  ( $r = 3$  km);  $-16\%$  and  $-11\%$  ( $r = 5$  km).

The empirical distributions of crowdsourced and reference hourly precipitation observations are compared in the quantile–quantile (Q–Q) plots shown in Fig. 4. The thin-dashed black lines mark the identity lines  $y = x$ , which is where the points would lie if the two distributions were similar. We point out that: (i) the Q–Q plots for both  $C_{75}$  and  $C_{NN}$  stay close enough to the identity line; (ii) the graphs in the three panels are rather similar among each other, although a slight worsening of the agreement can be noticed as the distance increases. It is worth remarking that the crowdsourced

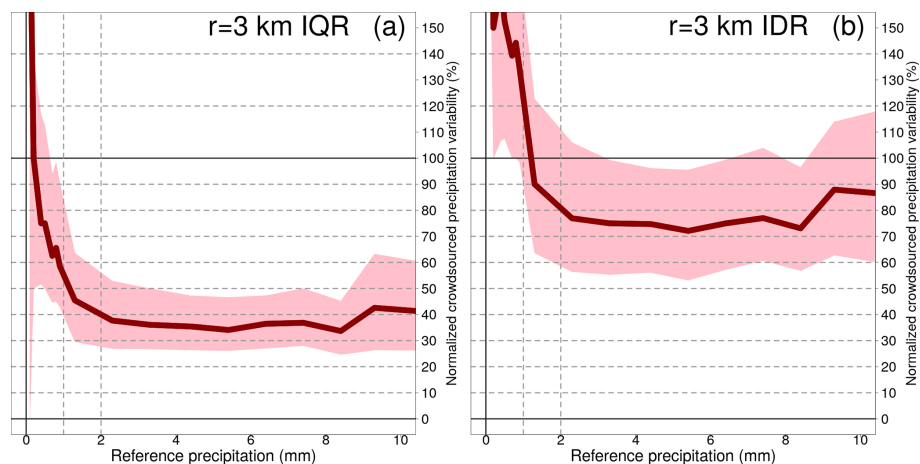
data have not been quality controlled, then the higher quantiles are most likely affected by outliers (e.g.  $C_{90}$  in Fig. 4a). Looking at the figure, it is possible to state that a Q–Q mapping procedure (Wilks, 2019) might be a good way to deal with some of the underestimation issues we have reported above.

The last result we present in this section focuses on the performances of crowdsourced observations in distinguishing between precipitation yes/no cases and, more in general, on the agreement between crowdsourced and reference precipitation being simultaneously above predefined thresholds. In Fig. 5, the equitable threat score (ETS, Jolliffe and Stephenson, 2012) of the crowdsourced precipitation is shown for the three neighbourhoods used in our study. Given a threshold of precipitation (on the  $x$ -axis), the ETS measures the fraction of crowdsourced observations that were correctly predicting an amount above that threshold, adjusted for the hits associated with random chance. A “hit” is defined as “both the





**Figure 6.** Spatial variability of crowdsourced hourly precipitation as a function of reference precipitation amounts, within circular regions of 1 km radius. IQR stands for interquartile range, while IDR stands for interdecile range. The dataset used, the reference precipitation classes and the constraints imposed are the same as for Fig. 1. In panel (a), The tick red lines show the median of the IQRs (IDRs in panel b) within the precipitation classes. In panel a, The shaded pink regions delimit the area between the 25th and the 75th percentiles of the IQRs (IDRs in panel b).



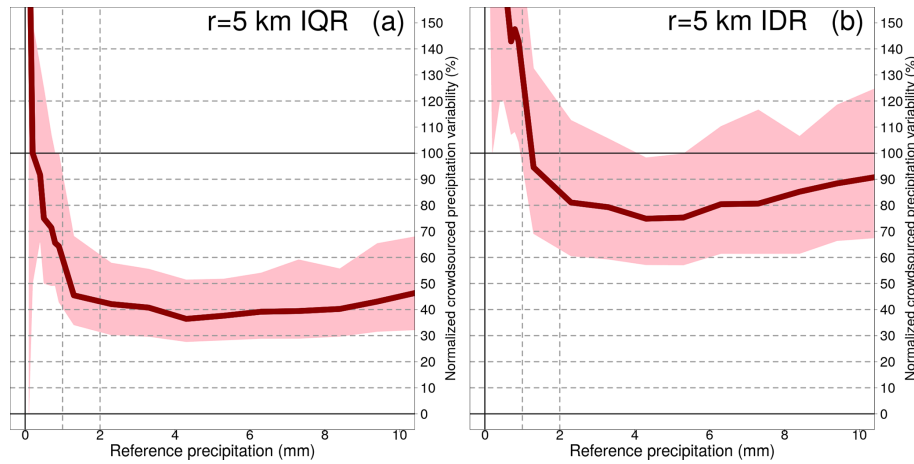
**Figure 7.** Spatial variability of crowdsourced hourly precipitation as a function of reference precipitation amounts, within circular regions of 3 km radius. The layout is similar to Fig. 6 and the dataset used is the same as that used in Fig. 2.

crowdsourced observation (or derived statistics) and the reference are greater than the threshold”. Because of the differences between  $n$  and  $n^*$  (see Sect. 3), we should expect an increase of “misses” when the amount increases (i.e. “the crowdsourced observation is below the threshold, while the reference is above”). The ETS graphs show that we can have good confidence in the ability of crowdsourced data in distinguishing between precipitation yes/no events. As expected, performance deteriorates with increasing precipitation. The aggregation of crowdsourced data over small regions yields somewhat better and more stable results, as can be seen in the graphs by comparing the results for the nearest neighbours (green line) and those for the aggregated statistics (black and – especially – the gray line).

### 3.2 Spatial variability of precipitation

The spatial variability of precipitation over neighbourhoods of 1, 3 and 5 km has been measured using the IDR (i.e.  $\langle(C_{90} - C_{10})|R\rangle$ ) and the IQR (i.e.  $\langle(C_{75} - C_{25})|R\rangle$ ). Given a neighbourhood, IDR gives an idea of the total range of values we should expect to find in the crowdsourced observations, extreme values included. IQR represents the typical (i.e. most likely) range of values. Variability is assumed to depend on precipitation intensity, then we will present our results using the same precipitation classes defined in Sect. 3.

The results are presented in Tables 4–6 and Figs. 6–8 for the three neighbourhoods. In both figures and tables, we have used the normalized crowdsourced variability (units %), where the normalization of the spread is meant with respect



**Figure 8.** Spatial variability of crowdsourced hourly precipitation as a function of reference precipitation amounts, within circular regions of 5 km radius. The layout is similar to Fig. 6 and the dataset used is the same as that used in Fig. 3.

**Table 4.** Spatial variability of crowdsourced observations conditional to classes of reference precipitation  $R$  based on measurements from 1 September 2019 to 1 November 2022. The crowdsourced stations used lie within circular regions of radius equal to 1 km from the reference stations and there must be at least 5 crowdsourced observations simultaneously available. The data shown in the table have been used to obtain the graph in Fig. 6. The first column reports the definitions of precipitation classes. For the notation in the remaining three columns, see Sect. 3. The second (fourth) column is equivalent to  $\langle C_{90} - C_{10} | R \rangle$  ( $\langle C_{75} - C_{25} | R \rangle$ ). The 3-tuple in these columns show the results with three different aggregation operators  $\langle \dots \rangle$ , which are over all samples in each class: (25th percentile, median (bold), 75th percentile). The third (fifth) column is the normalized IDR (IQR), which is defined as  $\langle C_{90} - C_{10} | R \rangle / \langle C_{50} | R \rangle$  ( $\langle C_{75} - C_{25} | R \rangle / \langle C_{50} | R \rangle$ ). The 3-tuple in these columns show the results with three different aggregation operators, as for the second and fourth columns, with the difference that the operator used for  $\langle C_{50} | R \rangle$  is always the median (i.e.  $\langle C_{50} | R \rangle$  is constant over each 3-tuple).

$R$ class (mm)	$\langle \text{IDR}   R \rangle$ (mm)	norm. $\langle \text{IDR}   R \rangle$ (%)	$\langle \text{IQR}   R \rangle$ (mm)	norm. $\langle \text{IQR}   R \rangle$ (%)
$0.1 \leq R < 0.2$	(0.00, <b>0.10</b> , 0.20)	(–, –, –)	(0.00, <b>0.07</b> , 0.10)	(–, –, –)
$0.2 \leq R < 0.3$	(0.16, <b>0.24</b> , 0.39)	(78, <b>120</b> , 195)	(0.07, <b>0.10</b> , 0.20)	(37, <b>50</b> , 100)
$0.4 \leq R < 0.5$	(0.25, <b>0.36</b> , 0.54)	(83, <b>120</b> , 181)	(0.10, <b>0.18</b> , 0.30)	(33, <b>60</b> , 100)
$0.5 \leq R < 0.6$	(0.30, <b>0.46</b> , 0.70)	(67, <b>102</b> , 156)	(0.11, <b>0.20</b> , 0.35)	(24, <b>44</b> , 78)
$0.6 \leq R < 0.7$	(0.40, <b>0.60</b> , 0.86)	(67, <b>100</b> , 143)	(0.17, <b>0.30</b> , 0.45)	(29, <b>50</b> , 75)
$0.8 \leq R < 0.9$	(0.40, <b>0.65</b> , 0.95)	(61, <b>98</b> , 144)	(0.17, <b>0.28</b> , 0.50)	(27, <b>42</b> , 76)
$0.9 \leq R < 1$	(0.44, <b>0.74</b> , 1.01)	(59, <b>99</b> , 135)	(0.20, <b>0.30</b> , 0.54)	(27, <b>40</b> , 72)
$1 \leq R < 2$	(0.50, <b>0.74</b> , 1.12)	(45, <b>67</b> , 102)	(0.22, <b>0.38</b> , 0.56)	(20, <b>34</b> , 50)
$2 \leq R < 3$	(0.84, <b>1.24</b> , 1.91)	(42, <b>62</b> , 96)	(0.40, <b>0.60</b> , 0.90)	(20, <b>30</b> , 45)
$3 \leq R < 4$	(1.16, <b>1.75</b> , 2.65)	(41, <b>62</b> , 94)	(0.55, <b>0.88</b> , 1.30)	(19, <b>31</b> , 46)
$4 \leq R < 5$	(1.62, <b>2.31</b> , 3.40)	(45, <b>64</b> , 94)	(0.78, <b>1.10</b> , 1.50)	(22, <b>31</b> , 42)
$5 \leq R < 6$	(1.90, <b>2.64</b> , 4.36)	(41, <b>57</b> , 94)	(0.91, <b>1.30</b> , 1.90)	(20, <b>28</b> , 41)
$6 \leq R < 7$	(2.22, <b>3.44</b> , 5.78)	(39, <b>61</b> , 102)	(1.10, <b>1.47</b> , 2.80)	(19, <b>26</b> , 49)
$7 \leq R < 8$	(2.65, <b>4.08</b> , 7.04)	(41, <b>64</b> , 110)	(1.25, <b>1.73</b> , 2.95)	(20, <b>27</b> , 46)
$8 \leq R < 9$	(2.95, <b>6.60</b> , 9.55)	(44, <b>99</b> , 143)	(1.82, <b>2.67</b> , 3.90)	(27, <b>40</b> , 58)
$9 \leq R < 10$	(2.67, <b>4.36</b> , 7.54)	(37, <b>60</b> , 104)	(1.01, <b>2.49</b> , 3.22)	(14, <b>34</b> , 45)
$10 \leq R < 20$	(4.50, <b>7.18</b> , 10.70)	(44, <b>71</b> , 105)	(2.44, <b>3.83</b> , 5.53)	(24, <b>38</b> , 54)

to the observed amount. We have used three different aggregation operators  $\langle \dots \rangle$ , which are: the 25th, the 50th (median) and the 75th percentiles of all samples within a class. Then, for instance, in Table 4, the 3-tuple in the first row of the second column ( $\langle \text{IDR} | R \rangle$ , units mm) is (0.00, 0.10, 0.20) and it means: 0.00 mm is the 25th percentile of all IDR samples within the class  $0.1 \leq R < 0.2$ ; 0.10 mm is the me-

dian of all IDR samples within the same class; 0.20 mm is the 75th percentile of all IDR samples within the same class. Since we are using the spread of the crowdsourced observation, the normalization is done on the basis of  $\langle C_{50} | R \rangle$  (and not of  $\langle R \rangle$ ). Then, the normalized IDR (IQR) is defined as  $\langle C_{90} - C_{10} | R \rangle / \langle C_{50} | R \rangle$  ( $\langle C_{75} - C_{25} | R \rangle / \langle C_{50} | R \rangle$ ). Note that, for the results presented in this section, the op-

**Table 5.** Spatial variability of crowdsourced observations conditional to classes of reference precipitation  $R$  based on measurements from 1 September 2019 to 1 November 2022. The crowdsourced stations used lie within circular regions of radius equal to 3 km from the reference stations and there must be at least 10 crowdsourced observations simultaneously available. The data shown in the table have been used to obtain the graph in Fig. 7. The table layout is similar to Table 4.

$R$ class (mm)	$\langle \text{IDR}   R \rangle$ (mm)	norm. $\langle \text{IDR}   R \rangle$ (%)	$\langle \text{IQR}   R \rangle$ (mm)	norm. $\langle \text{IQR}   R \rangle$ (%)
$0.1 \leq R < 0.2$	(0.10, <b>0.16</b> , 0.30)	(200, <b>320</b> , 600)	(0.00, <b>0.10</b> , 0.15)	(0, <b>200</b> , 300)
$0.2 \leq R < 0.3$	(0.20, <b>0.30</b> , 0.50)	(100, <b>150</b> , 250)	(0.10, <b>0.20</b> , 0.28)	(50, <b>100</b> , 138)
$0.4 \leq R < 0.5$	(0.32, <b>0.50</b> , 0.70)	(107, <b>167</b> , 233)	(0.16, <b>0.22</b> , 0.35)	(52, <b>75</b> , 117)
$0.5 \leq R < 0.6$	(0.43, <b>0.61</b> , 0.87)	(108, <b>152</b> , 217)	(0.20, <b>0.30</b> , 0.45)	(50, <b>75</b> , 112)
$0.6 \leq R < 0.7$	(0.56, <b>0.78</b> , 1.00)	(100, <b>139</b> , 179)	(0.25, <b>0.35</b> , 0.52)	(45, <b>62</b> , 94)
$0.8 \leq R < 0.9$	(0.61, <b>0.88</b> , 1.12)	(100, <b>144</b> , 184)	(0.28, <b>0.40</b> , 0.60)	(45, <b>66</b> , 98)
$0.9 \leq R < 1$	(0.69, <b>0.94</b> , 1.22)	(99, <b>134</b> , 174)	(0.30, <b>0.41</b> , 0.64)	(43, <b>59</b> , 91)
$1 \leq R < 2$	(0.7, <b>0.99</b> , 1.35)	(64, <b>90</b> , 123)	(0.32, <b>0.50</b> , 0.70)	(30, <b>45</b> , 64)
$2 \leq R < 3$	(1.10, <b>1.50</b> , 2.07)	(56, <b>77</b> , 106)	(0.52, <b>0.74</b> , 1.03)	(27, <b>38</b> , 53)
$3 \leq R < 4$	(1.55, <b>2.10</b> , 2.78)	(55, <b>75</b> , 99)	(0.75, <b>1.01</b> , 1.4)	(27, <b>36</b> , 50)
$4 \leq R < 5$	(2.02, <b>2.69</b> , 3.46)	(56, <b>75</b> , 96)	(0.95, <b>1.27</b> , 1.70)	(26, <b>35</b> , 47)
$5 \leq R < 6$	(2.34, <b>3.17</b> , 4.20)	(53, <b>72</b> , 95)	(1.15, <b>1.50</b> , 2.05)	(26, <b>34</b> , 47)
$6 \leq R < 7$	(2.96, <b>3.88</b> , 5.14)	(57, <b>75</b> , 99)	(1.40, <b>1.89</b> , 2.45)	(27, <b>36</b> , 47)
$7 \leq R < 8$	(3.46, <b>4.39</b> , 5.92)	(61, <b>77</b> , 104)	(1.60, <b>2.10</b> , 2.85)	(28, <b>37</b> , 50)
$8 \leq R < 9$	(3.77, <b>4.85</b> , 6.40)	(57, <b>73</b> , 96)	(1.64, <b>2.24</b> , 3.00)	(25, <b>34</b> , 45)
$9 \leq R < 10$	(4.55, <b>6.38</b> , 8.26)	(63, <b>88</b> , 114)	(1.91, <b>3.09</b> , 4.58)	(26, <b>43</b> , 63)
$10 \leq R < 20$	(5.07, <b>7.52</b> , 10.80)	(57, <b>85</b> , 122)	(2.31, <b>3.54</b> , 5.10)	(26, <b>40</b> , 58)

**Table 6.** Spatial variability of crowdsourced observations conditional to classes of reference precipitation  $R$  based on measurements from 1 September 2019 to 1 November 2022. The crowdsourced stations used lie within circular regions of radius equal to 5 km from the reference stations and there must be at least 10 crowdsourced observations simultaneously available. The data shown in the table have been used to obtain the graph in Fig. 8. The table layout is similar to Table 4.

$R$ class (mm)	$\langle \text{IDR}   R \rangle$ (mm)	norm. $\langle \text{IDR}   R \rangle$ (%)	$\langle \text{IQR}   R \rangle$ (mm)	norm. $\langle \text{IQR}   R \rangle$ (%)
$0.1 \leq R < 0.2$	(0.10, <b>0.20</b> , 0.30)	(200, <b>400</b> , 600)	(0.00, <b>0.10</b> , 0.20)	(0, <b>200</b> , 400)
$0.2 \leq R < 0.3$	(0.20, <b>0.34</b> , 0.52)	(100, <b>170</b> , 260)	(0.10, <b>0.20</b> , 0.30)	(50, <b>100</b> , 150)
$0.4 \leq R < 0.5$	(0.36, <b>0.51</b> , 0.74)	(120, <b>170</b> , 247)	(0.20, <b>0.28</b> , 0.40)	(67, <b>92</b> , 133)
$0.5 \leq R < 0.6$	(0.48, <b>0.68</b> , 0.92)	(120, <b>170</b> , 230)	(0.20, <b>0.30</b> , 0.50)	(50, <b>75</b> , 125)
$0.6 \leq R < 0.7$	(0.60, <b>0.80</b> , 1.10)	(107, <b>143</b> , 196)	(0.28, <b>0.40</b> , 0.60)	(49, <b>71</b> , 107)
$0.8 \leq R < 0.9$	(0.66, <b>0.90</b> , 1.21)	(108, <b>148</b> , 198)	(0.30, <b>0.40</b> , 0.61)	(49, <b>66</b> , 100)
$0.9 \leq R < 1$	(0.73, <b>1.00</b> , 1.30)	(104, <b>143</b> , 186)	(0.30, <b>0.45</b> , 0.70)	(43, <b>64</b> , 100)
$1 \leq R < 2$	(0.76, <b>1.04</b> , 1.46)	(69, <b>95</b> , 133)	(0.38, <b>0.50</b> , 0.75)	(34, <b>45</b> , 68)
$2 \leq R < 3$	(1.15, <b>1.54</b> , 2.14)	(61, <b>81</b> , 113)	(0.58, <b>0.80</b> , 1.10)	(30, <b>42</b> , 58)
$3 \leq R < 4$	(1.60, <b>2.14</b> , 2.85)	(59, <b>79</b> , 106)	(0.80, <b>1.10</b> , 1.50)	(30, <b>41</b> , 56)
$4 \leq R < 5$	(2.00, <b>2.62</b> , 3.44)	(57, <b>75</b> , 98)	(0.97, <b>1.27</b> , 1.80)	(28, <b>36</b> , 51)
$5 \leq R < 6$	(2.42, <b>3.20</b> , 4.24)	(57, <b>75</b> , 100)	(1.20, <b>1.60</b> , 2.20)	(28, <b>38</b> , 52)
$6 \leq R < 7$	(2.98, <b>3.90</b> , 5.35)	(61, <b>80</b> , 110)	(1.40, <b>1.90</b> , 2.62)	(29, <b>39</b> , 54)
$7 \leq R < 8$	(3.35, <b>4.40</b> , 6.36)	(61, <b>81</b> , 117)	(1.57, <b>2.15</b> , 3.22)	(29, <b>39</b> , 59)
$8 \leq R < 9$	(3.78, <b>5.24</b> , 6.55)	(61, <b>85</b> , 107)	(1.82, <b>2.48</b> , 3.42)	(30, <b>40</b> , 56)
$9 \leq R < 10$	(4.67, <b>6.22</b> , 8.34)	(66, <b>88</b> , 119)	(2.22, <b>3.03</b> , 4.60)	(32, <b>43</b> , 65)
$10 \leq R < 20$	(5.80, <b>7.90</b> , 11.15)	(69, <b>94</b> , 132)	(2.78, <b>4.24</b> , 6.00)	(33, <b>50</b> , 71)

erator used in the aggregation of  $\langle C_{50} | \rangle$  is always the median over the samples (i.e. even when calculating the 25th percentile of e.g.  $\langle C_{90} - C_{10} | R \rangle / \langle C_{50} | R \rangle$ , we have used the median to obtain  $\langle C_{50} | R \rangle$ ). In the tables, the absolute values of the variability are reported (i.e.  $\langle \text{IDR} | R \rangle \equiv \langle C_{90} - C_{10} | R \rangle$  and  $\langle \text{IQR} | R \rangle \equiv \langle C_{75} - C_{25} | R \rangle$ ) but not on the figures.

The link between figures and tables is explained with an example. Let us consider Fig. 7a, the pink area shows the range of values delimited by the first and the third values of the 3-tuple in the fifth column of Table 5 (i.e. the range of values for the normalized  $\langle \text{IQR} | R \rangle$ , between the 25th and the 75th percentiles). The thick red line joins together the second

values of the 3-tuple in the fifth column of Table 5 (i.e. the medians of norm.  $(IQR|R)$ ). Then, the pink area is the variability we have found in the normalized crowdsourced variability. The thick red line is the “typical” value of the normalized crowdsourced variability.

A common feature of all the three figures is that the normalized spatial variability is very high for light precipitation (i.e. less than 1 mm), often even with values higher than 100 %. Then, the normalized variability stabilizes and reaches a plateau which remains fairly constant throughout the range of precipitation amounts. As expected, the variability increases as the area of the neighbourhood considered increases too.

For the IQRs, the plateaus, reached for values  $\geq 2$  mm, are (we consider the medians here): between 26 % and 40 % (1 km); between 34 % and 43 % (3 km); between 36 % and 50 % (5 km). For the IDRs, the plateaus, reached for values higher than 2 mm, are: between 57 % and 99 % (1 km); between 72 % and 88 % (3 km); between 75 % and 94 % (5 km).

#### 4 Conclusions

The relationship between the empirical distributions of Netatmo’s hourly precipitation totals conditional to reference precipitation has been investigated. We have found that the reference observations are always included in the envelope of the empirical distribution of crowdsourced data (i.e. between the 10th and the 90th percentiles). However, there are indications that for intense precipitation crowdsourced data may underestimate precipitation. This is inline with WMO guidelines, which recommend to correct measurements from tipping-buckets rain gauges and to adjust measurements of different rain gauges to make them more comparable.

The results obtained by comparing the empirical distributions of crowdsourced and reference precipitation suggest that it would probably be possible to use a quantile-quantile mapping procedure to adjust the crowdsourced observations toward the reference values.

It might also be beneficial to aggregate the crowdsourced data over small neighbourhoods, of the size of 1 to 5 km, instead of using the raw data. In this way, crowdsourced data performs better in distinguishing between precipitation yes/no events, for instance.

The investigation of the crowdsourced precipitation spatial variability shows that when comparing measurement from two points, even if not very far from each other (i.e. distance between 1 to 5 km), one should not be surprised to find values that are quite different from each other (up to 50 % of the mean hourly precipitation in the area). The variability is quantified in detail in the presented tables. The results are representative of the actual spatial variability of precipitation over small distances, as described in Sect. 3. However, part of the variability is certainly given by the not ideal siting exposure of Netatmo’s stations and, in this sense, the results

obtained can be considered as a maximum estimate of the variability.

**Data availability.** MET Norway station data are open and publicly accessible via <https://frost.met.no/index.html> (The Norwegian Meteorological Institute, 2023) or <https://seklima.met.no/> (The Norwegian centre for climate services, 2023). SMHI data used are open and publicly accessible at <https://opendata.smhi.se> (The Swedish Meteorological and Hydrological Institute, 1975). FMI data are open and publicly accessible data at <https://en.ilmatieteenlaitos.fi/open-data> (Finnish Meteorological Institute, 2023). Netatmo rain gauge data are available from Netatmo <https://www.netatmo.com/> (Netatmo, 2023). Restrictions apply to the availability of Netatmo data, which were used under license for this study.

**Author contributions.** CL, TNN, IAS and LB equally contributed to the conceptualization. CL and EB performed the formal analysis. TNN, IAS and LB contributed to the validation of results. CL has written the software for the formal analysis and prepared the manuscript with contributions from all co-authors.

**Competing interests.** The contact author has declared that none of the authors has any competing interests.

**Disclaimer.** Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Special issue statement.** This article is part of the special issue “EMS Annual Meeting: European Conference for Applied Meteorology and Climatology 2022”. It is a result of the EMS Annual Meeting: European Conference for Applied Meteorology and Climatology 2022, Bonn, Germany, 4–9 September 2022. The corresponding presentation was part of session OSA3.2: Spatial climatology.

**Financial support.** This research was partially supported by “EUMETNET – R&D Study A1.05 – QC for data from personal weather stations” and by the internal project “CONFIDENT” of the Norwegian Meteorological Institute.

**Review statement.** This paper was edited by Mojca Dolinar and reviewed by two anonymous referees.

## References

- Alerskans, E., Lussana, C., Nipen, T. N., and Seierstad, I. A.: Optimizing Spatial Quality Control for a Dense Network of Meteorological Stations, *J. Atmos. Oceanic Tech.*, 39, 973–984, <https://doi.org/10.1175/JTECH-D-21-0184.1>, 2022.
- Bárdossy, A., Seidel, J., and El Hachem, A.: The use of personal weather station observations to improve precipitation estimation and interpolation, *Hydrol. Earth Syst. Sci.*, 25, 583–601, <https://doi.org/10.5194/hess-25-583-2021>, 2021.
- Båserud, L., Lussana, C., Nipen, T. N., Seierstad, I. A., Oram, L., and Aspelién, T.: TITAN automatic spatial quality control of meteorological in-situ observations, *Adv. Sci. Res.*, 17, 153–163, <https://doi.org/10.5194/asr-17-153-2020>, 2020.
- Colli, M., Lanza, L., and Chan, P.: Co-located tipping-bucket and optical drop counter RI measurements and a simulated correction algorithm, *Atmos. Res.*, 119, 3–12, <https://doi.org/10.1016/j.atmosres.2011.07.018>, 2013.
- de Vos, L., Leijnse, H., Overeem, A., and Uijlenhoet, R.: The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam, *Hydrol. Earth Syst. Sci.*, 21, 765–777, <https://doi.org/10.5194/hess-21-765-2017>, 2017.
- de Vos, L. W., Raupach, T. H., Leijnse, H., Overeem, A., Berne, A., and Uijlenhoet, R.: High-Resolution Simulation Study Exploring the Potential of Radars, Crowdsourced Personal Weather Stations, and Commercial Microwave Links to Monitor Small-Scale Urban Rainfall, *Water Resour. Res.*, 54, 10293–10312, <https://doi.org/10.1029/2018WR023393>, 2018.
- de Vos, L. W., Leijnse, H., Overeem, A., and Uijlenhoet, R.: Quality Control for Crowdsourced Personal Weather Stations to Enable Operational Rainfall Monitoring, *Geophys. Res. Lett.*, 46, 8820–8829, <https://doi.org/10.1029/2019GL083731>, 2019.
- Finnish Meteorological Institute: The Finnish Meteorological Institute's open data, <https://en.ilmatieteenlaitos.fi/open-data> (last access: 30 May 2023), 2023.
- Frei, C. and Isotta, F. A.: Ensemble Spatial Precipitation Analysis from Rain-Gauge Data – Methodology and Application in the European Alps, *J. Geophys. Res.-Atmos.*, 124, 5757–5778, <https://doi.org/10.1029/2018JD030004>, 2019.
- Frogner, I.-L., Singleton, A. T., Kjøltzow, M. Ø., and Andrae, U.: Convection-permitting ensembles: Challenges related to their design and use, *Q. J. Roy. Meteorol. Soc.*, 145, 90–106, <https://doi.org/10.1002/qj.3525>, 2019.
- Haakenstad, H. and Øyvind Breivik: NORA3. Part II: Precipitation and Temperature Statistics in Complex Terrain Modeled with a Nonhydrostatic Model, *J. Appl. Meteorol. Clim.*, 61, 1549–1572, <https://doi.org/10.1175/JAMC-D-22-0005.1>, 2022.
- Habib, E., Krajewski, W. F., and Kruger, A.: Sampling Errors of Tipping-Bucket Rain Gauge Measurements, *J. Hydrol. Eng.*, 6, 159–166, [https://doi.org/10.1061/\(ASCE\)1084-0699\(2001\)6:2\(159\)](https://doi.org/10.1061/(ASCE)1084-0699(2001)6:2(159)), 2001.
- Jolliffe, I. T. and Stephenson, D. B.: *Forecast verification*, Wiley, Oxford, ISBN 978-0-470-66071-3, 2012.
- Lanza, L. G. and Stagi, L.: Certified accuracy of rainfall data as a standard requirement in scientific investigations, *Adv. Geosci.*, 16, 43–48, <https://doi.org/10.5194/adgeo-16-43-2008>, 2008.
- Lanza, L. G. and Stagi, L.: High resolution performance of catching type rain gauges from the laboratory phase of the WMO Field Intercomparison of Rain Intensity Gauges, *Atmos. Res.*, 94, 555–563, <https://doi.org/10.1016/j.atmosres.2009.04.012>, 2009.
- Lanzante, J. R.: Resistand, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data, *Int. J. Climatol.*, 16, 1197–1226, [https://doi.org/10.1002/\(SICI\)1097-0088\(199611\)16:11<1197::AID-JOC89>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-0088(199611)16:11<1197::AID-JOC89>3.0.CO;2-L), 1996.
- Lussana, C., Uboldi, F., and Salvati, M. R.: A spatial consistency test for surface observations from mesoscale meteorological networks, *Q. J. Roy. Meteorol. Soc.*, 136, 1075–1088, <https://doi.org/10.1002/qj.622>, 2010.
- Lussana, C., Nipen, T. N., Seierstad, I. A., and Elo, C. A.: Ensemble-based statistical interpolation with Gaussian anamorphosis for the spatial analysis of precipitation, *Nonl. Processes Geophys.*, 28, 61–91, <https://doi.org/10.5194/npg-28-61-2021>, 2021.
- Mandement, M. and Caumont, O.: Contribution of personal weather stations to the observation of deep-convection features near the ground, *Nat. Hazards Earth Syst. Sci.*, 20, 299–322, <https://doi.org/10.5194/nhess-20-299-2020>, 2020.
- Netatmo: The easy way to make your AC smart, <https://www.netatmo.com/> (last access: 30 May 2023), 2023.
- Nipen, T. N., Seierstad, I. A., Lussana, C., Kristiansen, J., and Hov, Ø.: Adopting Citizen Observations in Operational Weather Prediction, *B. Am. Meteorol. Soc.*, 101, E43–E57, <https://doi.org/10.1175/BAMS-D-18-0237.1>, 2020.
- Orlanski, I.: A rational subdivision of scales for atmospheric processes, *B. Am. Meteorol. Soc.*, 56, 527–530, 1975.
- Soci, C., Bazile, E., Besson, F., and Landelius, T.: High-resolution precipitation re-analysis system for climatological purposes, *Tellus A*, 68, 29879, <https://doi.org/10.3402/tellusa.v68.29879>, 2016.
- Stagnaro, M., Colli, M., Lanza, L. G., and Chan, P. W.: Performance of post-processing algorithms for rainfall intensity using measurements from tipping-bucket rain gauges, *Atmos. Meas. Tech.*, 9, 5699–5706, <https://doi.org/10.5194/amt-9-5699-2016>, 2016.
- The Norwegian centre for climate services: Observations and weather statistics, <https://seklima.met.no/> (last access: 30 May 2023), 2023.
- The Norwegian Meteorological Institute: The Norwegian Meteorological Institute's open data, <https://frost.met.no/index.html> (last access: 30 May 2023), 2023.
- The Swedish Meteorological and Hydrological Institute: The Swedish Meteorological and Hydrological Institute's open data, <https://opendata.smhi.se> (last access: 30 May 2023), 2023.
- Thunis, P. and Bornstein, R.: Hierarchy of mesoscale flow assumptions and equations, *J. Atmos. Sci.*, 53, 380–397, [https://doi.org/10.1175/1520-0469\(1996\)053<0380:HOMFAA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1996)053<0380:HOMFAA>2.0.CO;2), 1996.
- Tian, Y., Huffman, G. J., Adler, R. F., Tang, L., Sapiano, M., Maggioni, V., and Wu, H.: Modeling errors in daily precipitation measurements: Additive or multiplicative?, *Geophys. Res. Lett.*, 40, 2060–2065, <https://doi.org/10.1002/grl.50320>, 2013.
- Uboldi, F., Lussana, C., and Salvati, M.: Three-dimensional spatial interpolation of surface meteorological observations from high-resolution local networks, *Meteorol. Appl.*, 15, 331–345, <https://doi.org/10.1002/met.76>, 2008.



- Wilks, D. S.: Statistical methods in the atmospheric sciences, in: 4th Edn., Elsevier, <https://doi.org/10.1016/B978-0-12-815823-4.00009-2>, 2019.
- WMO: WMO-No.1160: Manual on the WMO Integrated Global Observing System: Annex VIII to the WMO Technical Regulations, Tech. Rep. 1160, World Meteorological Organization, ISBN 978-92-63-11160-9, 2021a.
- WMO: WMO-No.49\_volI: Technical Regulations, Volume I: General Meteorological Standards and Recommended Practices, Tech. Rep. 49, World Meteorological Organization, ISBN 978-92-63-10049-8, [https://library.wmo.int/doc\\_num.php?explnum\\_id=10955](https://library.wmo.int/doc_num.php?explnum_id=10955) (last access: 30 May 2023), 2021b.
- WMO: WMO-No.49\_volIII: Technical Regulations, Volume III: Hydrology, Tech. Rep. 49, World Meteorological Organization, ISBN 978-92-63-10049-8, [https://library.wmo.int/doc\\_num.php?explnum\\_id=11293](https://library.wmo.int/doc_num.php?explnum_id=11293) (last access: 30 May 2023), 2021c.
- WMO: WMO-No.8: WMO Guide to Meteorological Instruments and Methods of Observation, Tech. Rep. 8, World Meteorological Organization, ISBN 978-92-63-10008-5, 2021d.