RESEARCH PROGRAM ON
Climate Change, Agriculture and Food Security

Workshop report: Training program on ENACTS climate time series, data library and maprooms, Kigali, Rwanda

December 2015

Tufa Dinku and John del Corral

USAID FROM THE AMERICAN PEOPLE

International Research Institute for Climate and Society
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Rwanda Meteorology Agency

Training program on ENACTS climate time series, data library and maprooms
Kigali, Rwanda, December 2015

Workshop Report

CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS)

Tufa Dinku
John del Corral

Correct citation:
Dinku T, del Corral J. 2016. Training program on ENACTS climate time series, data library and maprooms, Kigali, Rwanda, December 2015. CCAFS Workshop Report. CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS). Copenhagen, Denmark. Available online at: www.ccafs.cgiar.org

CCAFS Workshop Reports aim to disseminate interim climate change, agriculture and food security research and practices and stimulate feedback from the scientific community.

Published by the CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS).

CCAFS is a strategic partnership of the CGIAR and the Earth System Science Partnership (ESSP). CGIAR is a global research partnership for a food secure future. The program is supported by the Canadian International Development Agency (CIDA), the Danish International Development Agency (DANIDA), the European Union (EU), and the CGIAR Fund, with technical support from the International Fund for Agricultural Development (IFAD).

Contact:
CCAFS Coordinating Unit - Faculty of Science, Department of Plant and Environmental Sciences, University of Copenhagen, Rolighedsvej 21, DK-1958 Frederiksberg C, Denmark. Tel: +45 35331046; Email: ccafs@cgiar.org

Photo: Tufa Dinku

Abstract

Meteo Rwanda, in collaboration with the International Research Institute for Climate and Society (IRI), has implemented the first phase of the ENACTS (Enhancing National Climate Services) initiative. The ENACTS initiative brings climate knowledge into national decision-making by improving availability, access and use of climate information. Meteo Rwanda staff has received a number of trainings on the different aspects of generating the datasets and developing climate information products. However, due to the recent reorganization at Meteo Rwanda, as well as updates to the tools used to generate historical data and information products, it was necessary to revise the training and update climate data and information products. The current activities, which are part of CCAFS-USAID Climate Services for Agriculture project, had two major components: (1) data quality control and generating updated climate datasets; and (2) integrating the newly generated datasets into Meteo Rwanda's maprooms. Sixteen Meteo Rwanda staff members received six days of training on data quality control and generating updated climate datasets, and/or 4.5 days of training on developing and maintaining ENACTS maprooms

Keywords

Quality Control, Merging, Satellite, Reanalysis.

About the authors

Tufa Dinku is a Research Scientist at the International Research Institute for Climate and Society (IRI), at Columbia University, focused on evaluating and improving climate datasets. Contact: tufa@iri.columbia.edu

John del Coral is a Senior Staff Associate at the International Research Institute for Climate and Society (IRI), at Columbia University, focused on database management, GIS and semantic technology. Contact: jdcorral@iri.columbia.edu

Acknowledgements

This report is an output of USAID's Rwanda Climate Services for Agriculture project, and was made possible through support provided by the Rwanda Mission, U.S. Agency for International Development. The opinions expressed herein are those of the authors, and do not necessarily reflect the views of the U.S. Agency for International Development.

Contents

Introduction 8

Activity 1: Data quality control and generating updated climate data sets 9

Training 9

Work on generating data 10

Activity 2: Integrating the newly generated datasets into Meteo Rwanda's maprooms 20

Conclusion 21

Appendix 1. Training Program 22

Appendix 2. Participant List 24

Acronyms

CDT	Climate Data Tools
CHIRP	Climate Hazard Group Infrared Precipitation
ENACTS	Enhancing National Climate Services
IRI	International Research Institute for Climate and Society
JRA55	Japan 55 Years Reanalysis
QGIS	Quantum Geographic Information System
USAID	United States Agency for International Development

Introduction

The Rwanda Meteorological Agency (Meteo Rwanda), in collaboration with the International Research Institute for Climate and Society (IRI) has embarked on a unique multi-faceted initiative called Enhancing National Climate Services (ENACTS). The initiative aims to bring climate knowledge into national decision-making by improving availability, access and use of climate information. Availability of climate data is improved by combining quality-controlled data from the national observation network, which is very sparse over many parts of the country, with satellite estimates for rainfall and elevation maps and reanalysis products for temperature. These new data sets have been used to develop some information products that are made available through Meteo's web page. The Rwanda Climate Services for Agriculture (CSA) project, funded by USAID and led by CCAFS, builds on and incorporates ENACTS as part of its Outcome 3: Climate Information Provision.

Through an earlier project, Meteo Rwanda staff received several trainings by IRI experts on the different aspects of generating the data sets and developing the products. However, due to the recent, and ongoing, reorganization at Meteo Rwanda, many people have left and some new staffs have been hired. Thus, Meteo Rwanda has asked for refresher training for old staff as well as training new staff. There have also been some new developments in methodology and tools. Thus, this training introduced all participants to these new methods and tools. One of the new tools is quality control at daily time scale, which will enable Meteo Rwanda to make correction to the data in their database. This also prepared the grounds for generating data at daily time scale, which is required for new products planned for Rwanda's agriculture sector.

This training prepared Meteo Rwanda for the development of new products and services planned within the CSA project. The training activities had two major components:

1. Data quality control and generating updated climate datasets, and.
2. Integrating the newly generated datasets into Meteo Rwanda's maprooms.

The first training was conducted at two different time: 30 November to 3 December 2015 and 27 to 29 January 2016. Twelve participants received training on data quality control and generating updated climate datasets. The second training took place 14 to 18 December 2015. An overlapping set of twelve participants were trained on developing and maintaining ENACTS maprooms. Daily quality check was performed on temperature as part of the training. Quality control of daily rainfall data was done outside the training due time constraints. Some serious quality issues were observed. The quality-controlled data were used to update dekadal (10-day) rainfall and temperature time series. The updated rainfall time series goes from 1981 to 2015 while the temperature time series covers the period 1961 to 2014.

Activity 1: Data quality control and generating updated climate data sets

This was conducted during 30 November to 3 December 2015 and 27 to 29 January 2016. It included a training component, and actual guided work on generating datasets.

Training

The training was a revision of previous ENACTS training. However, most of the trainees were totally new to subjects and lack background in meteorology. The training included both theoretical background and practical exercises on Meteo Rwanda's actual data. The topics covered in the training (Appendix 1) included the following:

Quality control of station data. Data quality control is a critical component of ENACTS. Thus, significant amount of time is spent on this aspect of the training. Trainees were first be introduced to the need for quality control of climate data, different types and sources of error with examples and different approaches to quality control of climate data. Then trainees then performed actual quality control of temperature data using IRI's Climate Data Tools (CDT).

Satellite rainfall estimation. An overview of satellite remote sensing, different satellites and sensor types, and the use of the different sensors for rainfall estimation was presented and discussed.

Climate Reanalysis Data. Climate reanalysis data are used for interpolating temperature data. Thus, a brief introduction was provided on the concept of climate reanalysis products.

Interpolation of climate data. The trainees were introduced to different interpolations methods, and their strengths and weakness. Then they explored different interpolation methods and the different factors that may affect interpolated values using their data and the CDT tool.

Combining satellite data with station measurements. The trainees were exposed to some merging techniques and then were shown how using auxiliary information, such as topography for temperature or satellite rainfall estimates for rainfall, could improve the quality of the interpolated values.

Work on generating data

Quality control

Most of this was done mostly outside the training time because there was no enough time to do it as part of the training. However, the trainees did do some practical exercises on data quality control and merging station data with proxies.

The quality control for rainfall has two parts: (1) detecting false zeros; and (2) identifying extremely high rainfall values. The first case happens when given station reports significantly higher number of zero values compared to the surrounding stations. This occurs mainly when missing observations are reported as blanks and the data entry staff interpret the blanks as zeros. The IRI Climate Data Tools (CDT) compares the given station with surrounding stations to identify the problem. An example of the output from CDT is given in Table 1. For instance, the first station in the table (ID = 20101600) reported all zero values (val.stn=100%) for Nov 1988. However, a station just 11km away reported only 45% zeros for same month. The average for stations in the area is 44%, while the maximum is 53%. Thus, station 20101600 seems to have false zeros, which need to be checked.

The other quality heck procedure is to identify extreme values. This is done both by comparing each observation of each station with historical records for a specific month (temporal check) as well as comparing the station value with surrounding stations for the same day (spatial check). Two examples of the temporal check are presented in Figures 1 and

2 below. The red bars indicate are what CDT identified as suspect values. These values need to be checked with the original paper data and fixed. Figure 3 is an example of the spatial check where a station observation is compared with values from the surrounding stations. The suspect value is shown in read. In Fig. 3, the value of the station is 248 mm while the maximum from the neighbouring stations is 63mm. Note that these are daily values. Thus, the 248 mm for that station is very unlikely and need to be checked.

Table 1: A sample output of zero-check.

ID	YY	MM	val.stn	nrst.val	nrst.stn	avg.nbrs	max.nbrs
20101600	1988	11	100	45	11	44	53
20107500	1984	1	100	32	3	48	58
20306600	1992	11	100	30	6	31	40
20306600	1992	12	100	45	6	45	61
20513500	1984	6	100	97	8	55	97
20513500	1985	7	100	0	17	0	0
20513500	1986	7	100	19	17	19	19
20601500	1981	9	100	53	6	62	70
20602100	1997	3	100	45	4	48	52
20604200	1981	12	100	52	5	55	61
20703200	1990	6	100	50	14	50	50

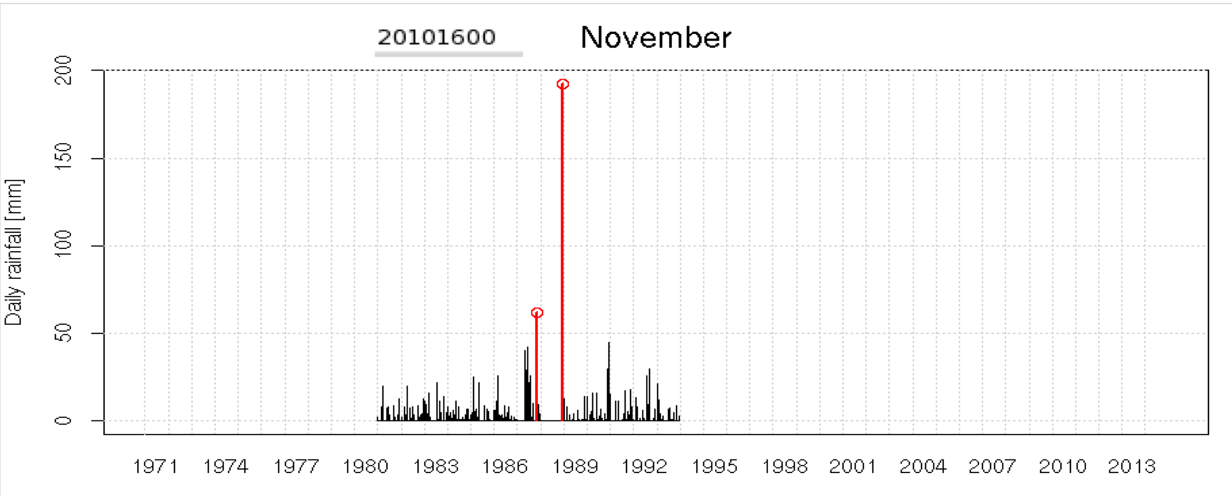

Figure 1: A sample output of the temporal quality check procedure. The extreme values suspected by the quality control procedure (CDT) are shown in red. The number at the top is station ID.

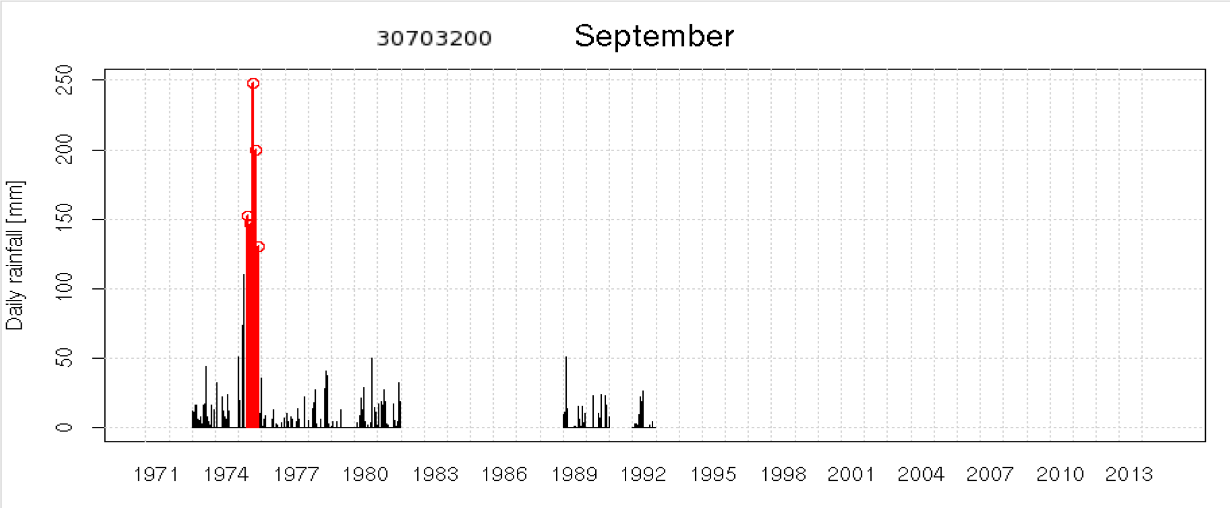

Figure 2: Same as Fig. 1, but for another station.

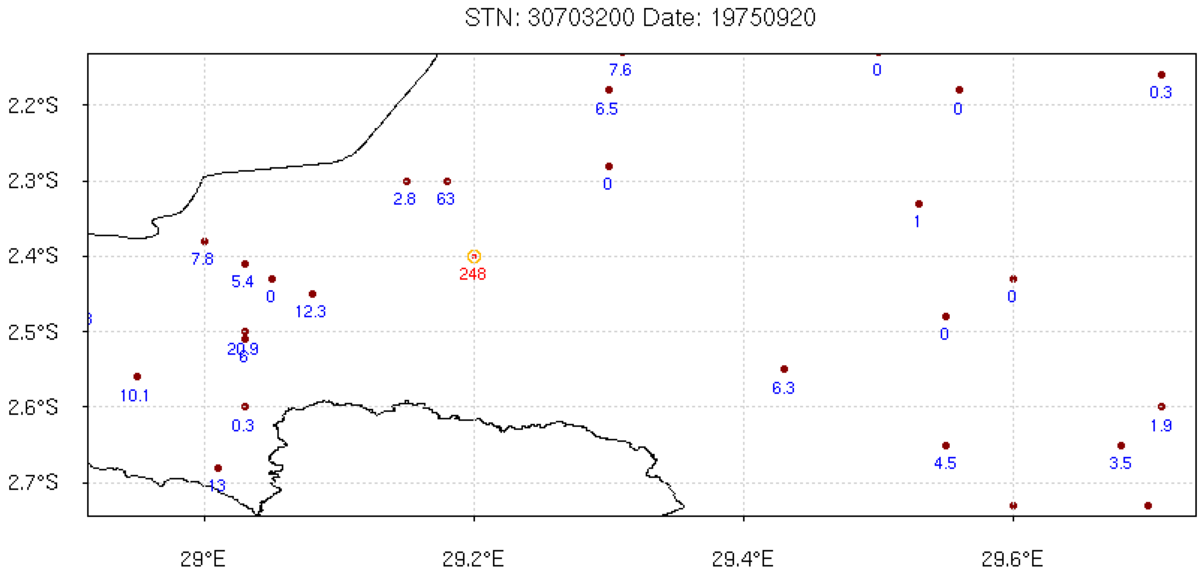

Figure 3: Result of spatial quality check. The suspect value is given in red.

The quality control for temperature is different from rainfall in that there is no zero-check. However, the extreme values could be either low or high. Both temporal and spatial checks have been performed. Examples are given below in Figures 4 through 7. Figures 4 and 5 are for minimum temperature while figures 6 and 7 are for maximum temperature. Figure 4 shows both high and low extreme values flagged by the quality check procedure. In this case, particularly for the low values case, a number of observations have been flagged as suspect. On the other hand, Fig. 5 shows one very unlikely extreme value (> 120^{o}C). This could most probably be a data entry problem.

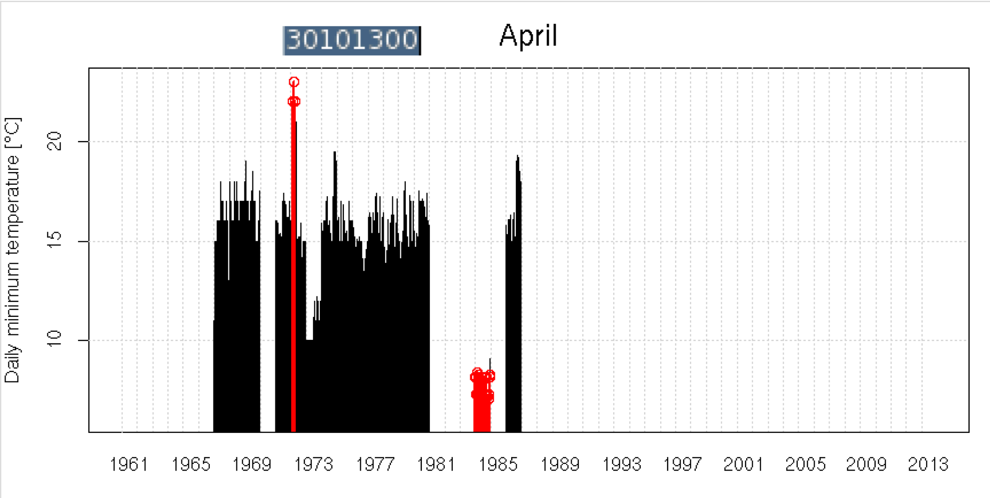

Figure 4: Minimum temperature time series for the month April. The red color shows extreme values flagged by CDT as suspect observations.

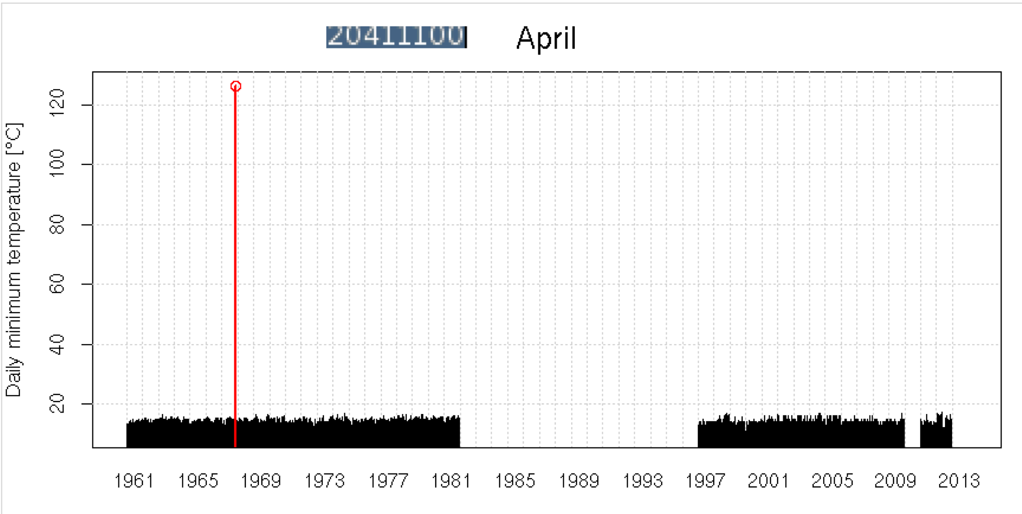

Figure 5: Minimum temperature time series for station 20411100, for the month April.

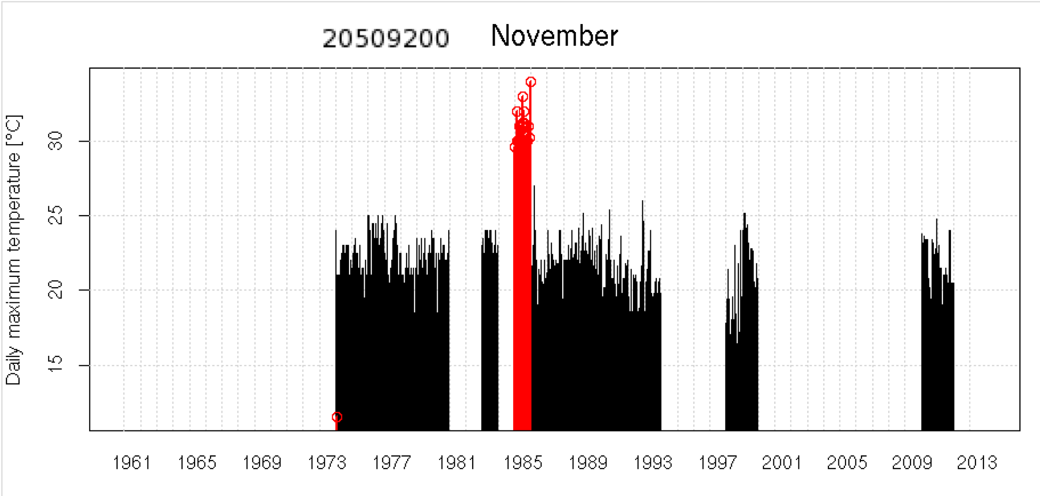

Figure 6. Maximum temperature time series for station 20509200 in the month of November.

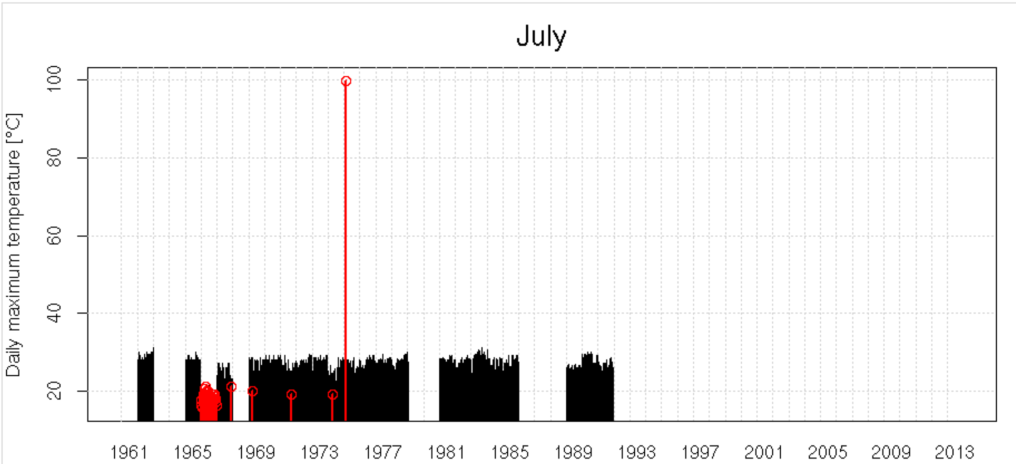

Figure 7: Maximum temperature time series for the month of July.

Sometimes it might not be to detect observation errors just by using objective methods. An example is given in Figure 8. Here the values are not extreme; thus, the quality-check procedure did not flag any values (no red bars). However, the constant values of 15°C shown in the figure do not look like reasonable observations. It is hard to imagine that temperature values will be constant for so many days. These values must be wrong. These kinds of errors could only be found by looking at the data station-by-station and month-by-month as shown here.

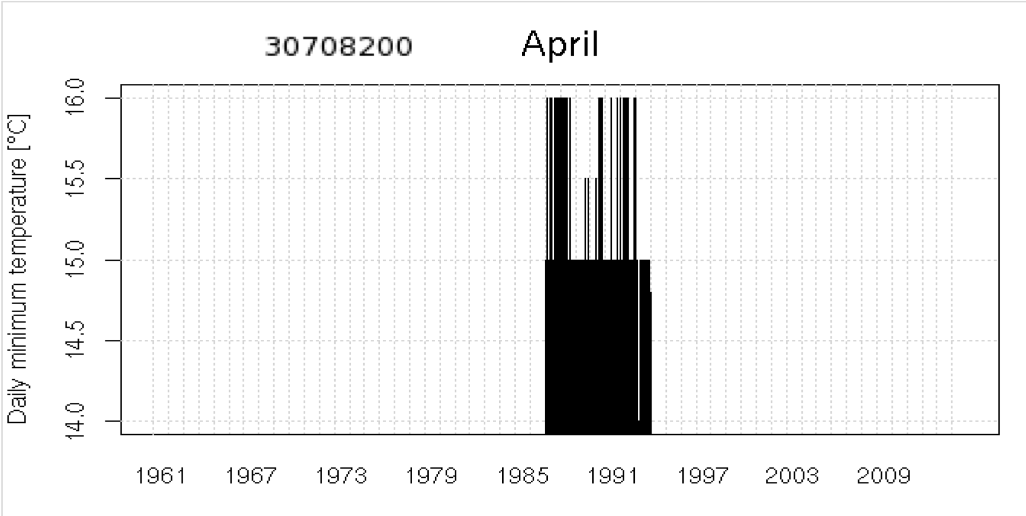

Figure 8: Suspect observations of constant value (15°C). These values have not been flagged as suspects, but do look right.

Generating updated climate data sets

The approach adopted for generating rainfall time series involves the following steps:

1. Use the data from 1981 to 1993 to calculate climatological adjustments factors for each dekad (dekads 1 to 36). Figure 9 shows stations used for this purpose
2. Interpolate the adjustment factors to the required grid points; and
3. Apply the adjustment factors to all satellite time series from 1981 to present;
4. Remove mean dekadal bias (averaged over the whole country) for each year using available stations;
5. Merge the output from the pervious step with available station data for each of dekad of each year. Stations used for this step are shown in Fig. 9.

The last two steps results in two different rainfall time series. The time series from step 4 have relatively more homogenous time series except around 1994 (Fig 10). These are used for the Climate Analysis maproom. They can also be used for other climate analysis activities. The data from step 5 is relatively the most accurate as it combines station observation for each dekad. The problem with this data is that the number of stations used varies from year-to-years, with almost no station data for part of 1994 and few stations for many years after that (Fig. 10). These data are good for applications that do need to compare one year with another.

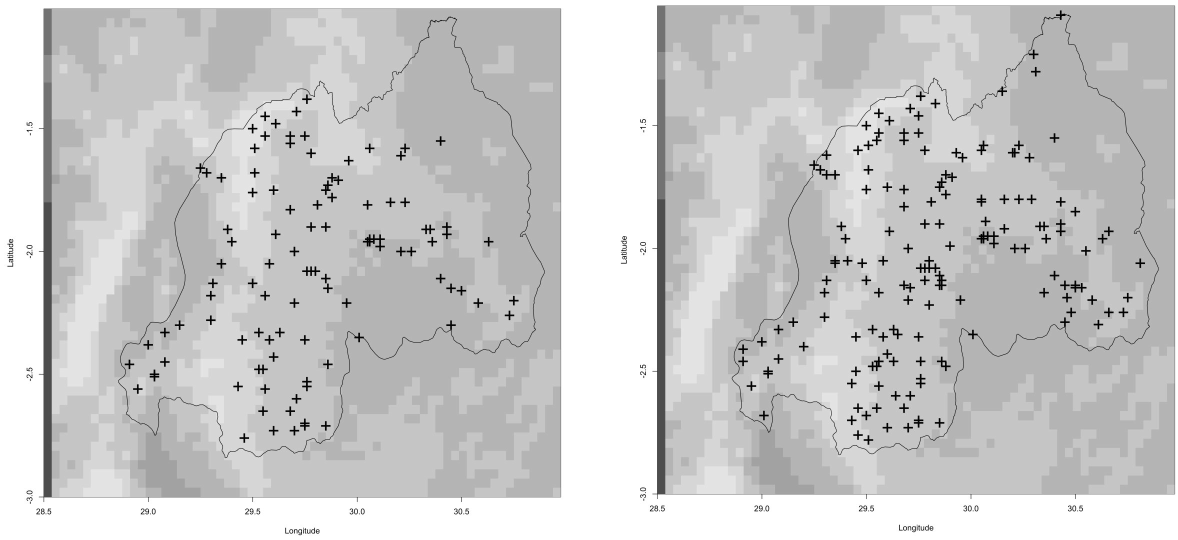

Figure 9: Distribution of stations used for bias removal (left) and combining with satellite estimates dekad-by-dekad (right).

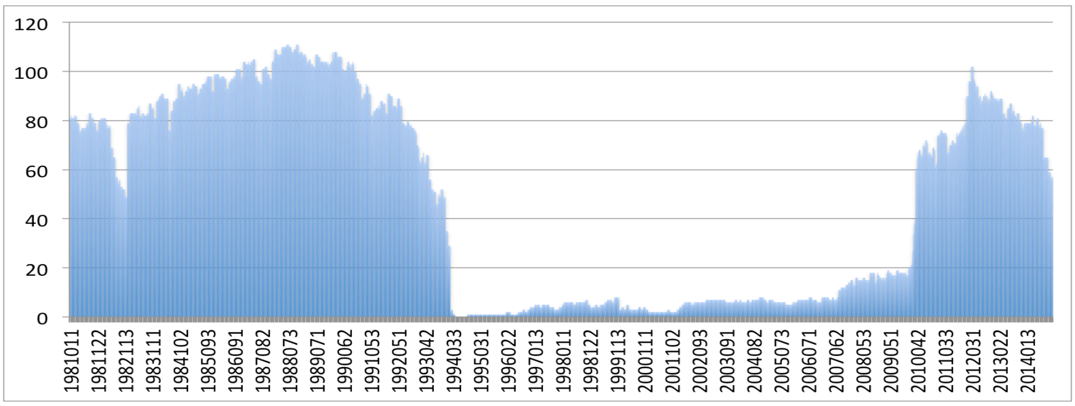

Figure 10: Number of stations with data for each year-month-dekad from 1981 to 2014.

Figure 11 below compares station data with satellite and combined station-satellite products. The combined product is an output from step 5 above.

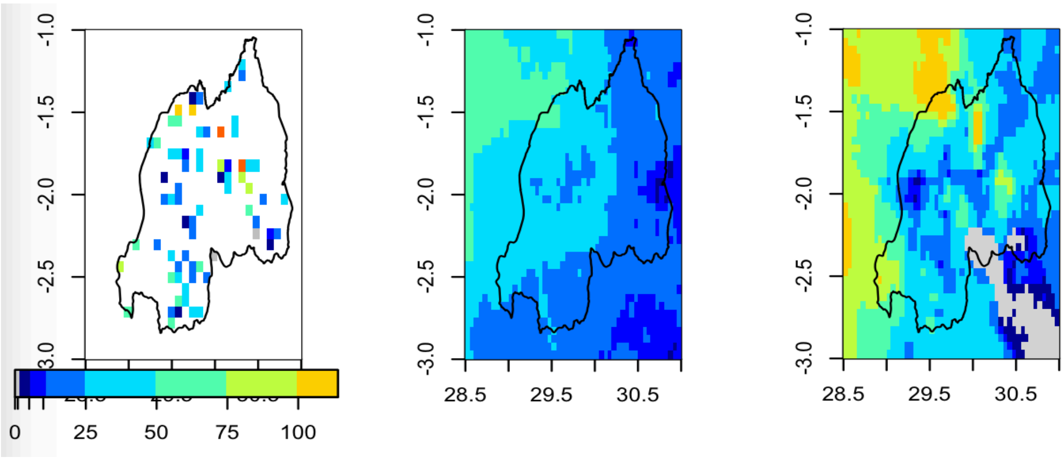

Figure 11: Comparisons of station observation (left), satellite estimate (center) and combined product (right) for a given dekad.

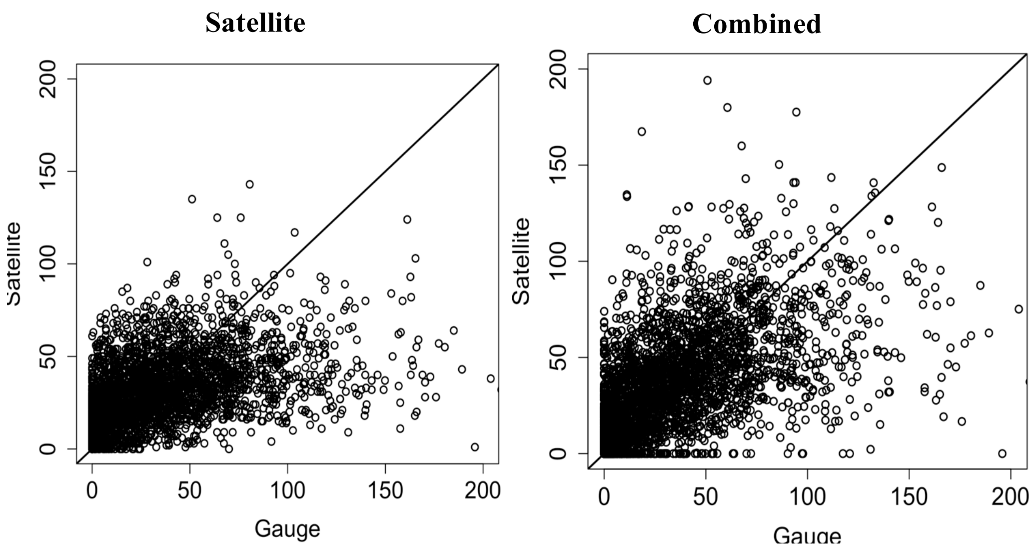

Figure 12: Comparison of satellite estimate (left) and combined-station product (right) with station observations.

There are no satellite temperature estimates going back 30 years. Thus, reanalysis data are used as a proxy. Reanalysis products are climate data generated by systematically combining climate observations (analyses) with climate model forecasts using data assimilation schemes and climate models. For this work, we use the Japanese 55-year Reanalysis, or JRA55 (http://jra.kishou.go.jp/JRA-55/index_en.html#about). This product has a coarse spatial resolution of about 50km. Thus, the reanalysis data are downscaled to 5km spatial resolution using station observations and elevation maps. The following steps are used to reconstruct the temperature time series:

1. Downscale reanalysis data from 50km to 5km;
2. Use the data from 1981 to 1990 to calculate adjustments factors for each dekad;
3. Interpolate the adjustment factors; and
4. Apply the adjustment factors to all downscaled reanalysis data from 1981 to 2014.
5. Merge the adjusted reanalysis from previous step with station measurements from every dekad of every year.

As in the case of rainfall, step 4 produces a more homogenous data with less accuracy, while step 5 results in relatively more accurate products with less homogeneity. The first products can be used for other climate analysis that requires comparing data from different periods. The second product can be used for applications that may not involve comparison of data from different periods. This is because the data from different periods, particularly periods with significantly different number of stations, will have different qualities.

Figure 13 compares station measurements of maximum temperature with downscaled reanalysis products and combined station-reanalysis data for a particular dekad. For maximum temperature, even the downscaled reanalysis data is very close to station observation. This could also be observed from the scatter plots in Fig. 14, which compare both reanalysis and the combined products with station measurements.

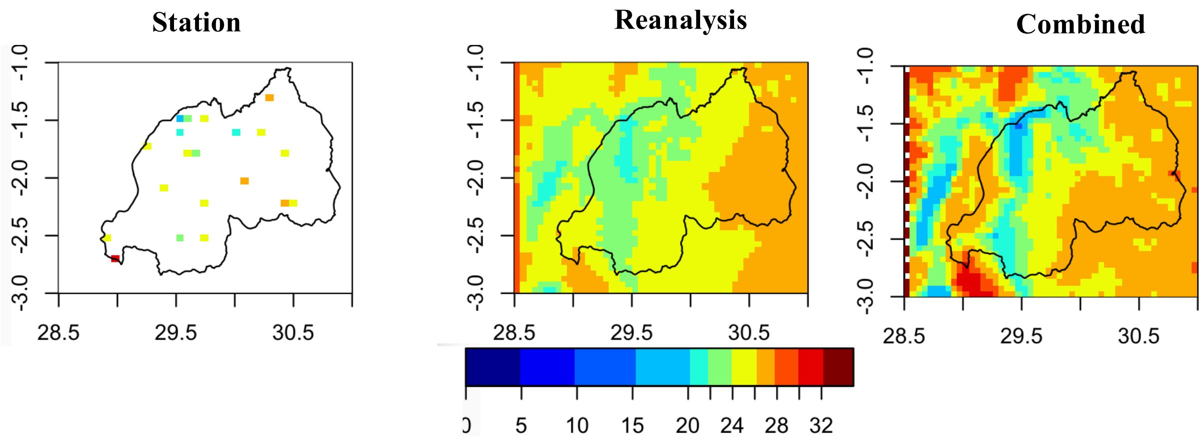

Figure 13: Comparison of station observation (left), reanalysis (center) and combined product for a given dekad.

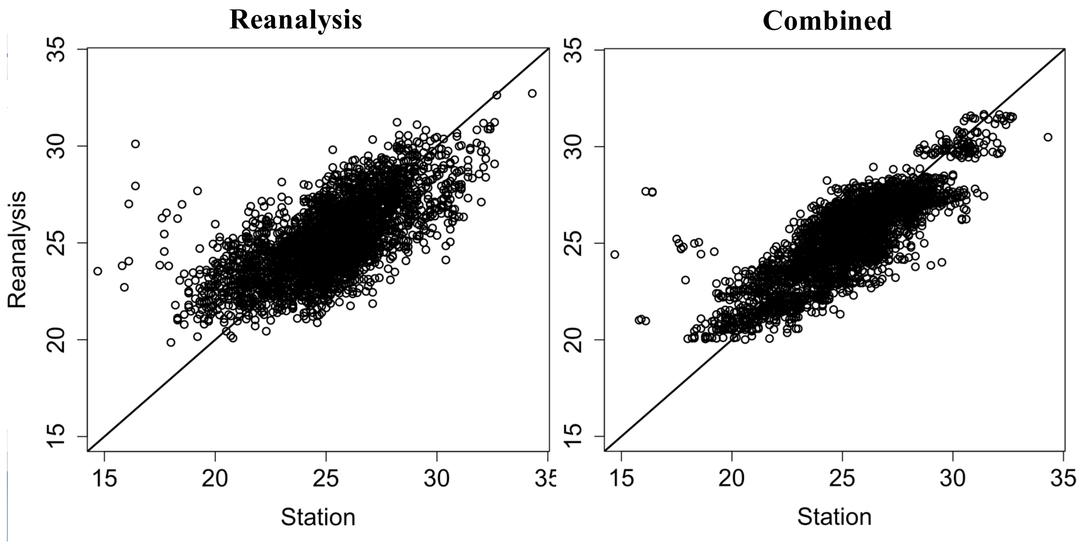

Figure 14: Comparison of downscaled reanalysis (left) and combined product (right) with station observations of maximum temperature

For minimum temperature (Fig. 15), the reanalysis product overestimates temperature values. This is also shown in the scatter plot of Fig. 16. Figure 16 also shows that the combined product is a significant improvement over the downscaled reanalysis data.

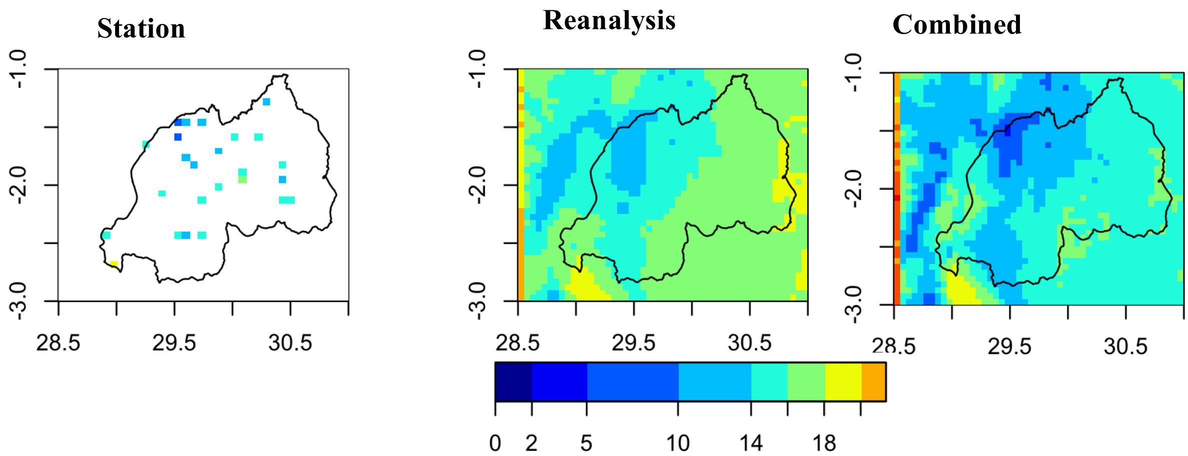

Figure 15: Comparison of station observation (left), reanalysis (center) and combined product of minimum temperature for a given dekad.

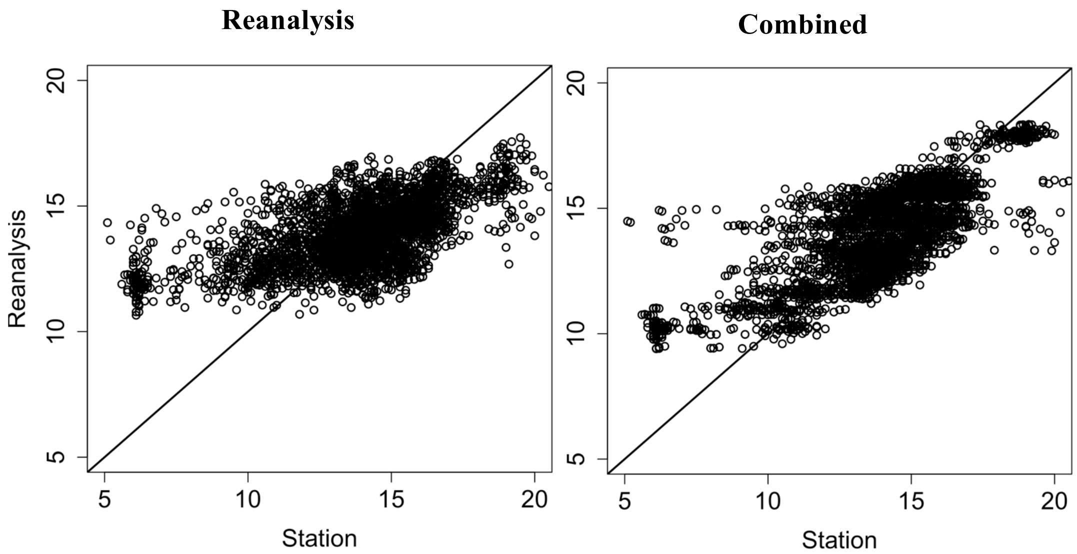

Figure 16: Comparison of downscaled reanalysis (left) and combined product (right) with station observations of minimum temperature.

Activity 2: Integrating the newly generated datasets into Meteo Rwanda's maprooms

This was conducted from 14 to 18 December 2015. The first 4 training days were held at the Classic Hotel, Kigali with 10 participants from the Rwanda Meteorological Agency, 2 of which were female. The final day of training was held at the Rwanda Meteorological Agency, Kigali with 3 male participants. All training materials and training schedule were distributed to the participants through online access to a GoogleDrive folder.

Day 1: Presentation of Data Library and maproom components. Detailed demonstration and explanation of all the ENACTS maprooms in the Rwanda Data Library. Participants were able to browse the maprooms and try different functions in the maproom.

Day 2: Introduction to the ENACTS Data Catalog. Description of how to add and update datasets in the Rwanda Data Library. Demonstration of basic Linux commands. Demonstration of viewing and manipulating ENACTS datasets.

Day 3: Introduction to QGIS (open source geographic information system). Installation of QGIS on participants' laptops. Demonstration of digitizing a seasonal forecast for Rwanda using QGIS. Practice by participants of digitizing using QGIS. More detailed explanation of ENACTS maprooms.

Day 4: Demonstration of adding GIS datasets to ENACTS Data Catalog. Demonstration of building and modifying ENACTS maprooms. Answering of follow-up questions about QGIS.

Day 5: Demonstration of creating a backup disk for the Rwanda Data Library and maprooms. Review of backup and recovery procedures for the Rwanda Data Library. Updated ENACTS data was transferred to the Rwanda Data Library server.

Conclusion

The two training activities built Meteo Rwanda's capacity in different areas. The training on data further improved Meteo Rwanda's capacity in quality control of station data and generating merged climate time series. The generated and updated datasets also enable Meteo Rwanda to provide improved services.

The maproom training prepared for official launch of the ENACTS Maproom on the website of the Rwanda Meteorological Agency in 2016. It is recommended to monitor connectivity and functionality of Maproom server and monitor updates of the ENACTS Monitoring maproom. Meteo Rwanda has solid computer, network, and electricity infrastructure. It is poised to be able to provide consistent online climate services for many sectors in Rwanda. Further development is needed to engage stakeholders. This can be done with sector-focused maprooms and sector-focused training workshops. In-depth training on the maintenance and use of the ENACTS products and maprooms needs to be done. This will make the staff at Meteo Rwanda more confident and self-sufficient.

Appendix 1. Training Program

Activity 1. Generating historical climate time series
Monday, 30 November
Introduction to the training
Installation and testing of software
Introduction to the Climate Data Tools (CDT)
Introduction satellite rainfall estimation
Tuesday 1 December
Introduction to Climate Reanalysis products
Introduction to quality control of station data
Practical: QC of RR and TT data
Wednesday 3 December
Introduction to interpolation of climate variables
Introduction to merging data from different sources
Practical: merging data from different sources
Thursday 4 December
Updating data for the monitoring maproom
Extracting data for points, boxes, and administrative boundaries
Summary
Activity 2. Maproom development and exploitation
Monday 14 December
Overview of Maproom structure
Maproom and Data Library linkage to datasets and functions
Practical: Analysis of Dominant Climate trends in each region
Tuesday 15 December
Add updated Monitoring data to Data Library
Quality check Monitoring maproom with new data
Practical: Process most recent dekad for Monitoring maproom
Wednesday 16 December
Introduction to git repository of Maproom on BitBucket
Updating the local copy and repository copy of the Maproom
Practical: Set up BitBucket accounts for primary users

Thursday 17 December
Procedure for making backup copy of Data Library
Practical: Make backup of Data Library
Friday 18 December
Monitoring methods for Data Library and Maproom
How to debug problems and fix them
Practical: Open platform for questions and follow-up

Appendix 2. Participant List

Name	Title	Institute	Gender	Training	
				1	2
Jean Marie Niyitegeka	Forecasting Officer	Meteo-Rwanda	M	Y	Y
Valens Rwakaozyo	Data Processing	Meteo-Rwanda	M	Y	Y
Didace Musoni	Assistant Director General	Meteo-Rwanda	M	Y	Y
Serge Senyana	Meteo Co-Application Officer	Meteo-Rwanda	M	Y	Y
Blandine Mukamanea	Observations Processing Officer	Meteo-Rwanda	F	Y	Y
Vuguziga Floribert	Forecasting Officer	Meteo-Rwanda	M	Y	Y
Annah Muteteri		Meteo-Rwanda	F	Y	N
Felicien Nsabakunze		Meteo-Rwanda	M	Y	N
Janet Umuhoza		Meteo-Rwanda	F	Y	N
Mathieu Mugunga Mbati		Meteo-Rwanda	M	Y	N
Francois Nsengiyumva		Meteo-Rwanda	M	Y	N
Clarisse Mukazarukommo	Observations Officer	Meteo-Rwanda	F	N	Y
Jonah Kazora	Forecasting Officer	Meteo-Rwanda	M	N	Y
Joseph Hazabintwari	Observations Officer	Meteo-Rwanda	M	N	Y
Jean Paul Kalisa	Observations Officer	Meteo-Rwanda	M	N	Y
Brave Rukagumya	Observations Officer	Meteo-Rwanda	M	Y	Y
John del Coral	Trainer	IRI	M	N	Y
Tufa Dinku	Trainer	IRI	M	Y	N