Quality and usefulness of crowdsourced precipitation in Burundi: Current status and ways forward

by

Aliane Nahimana



A thesis presented for the degree of
Master of Science in Civil Engineering

Student number: 4790138

Committee members:
Dr. M.A. Schleiss (Daily supervisor)
Dr. Stefan Ligtenberg (co-supervisor)
Dr. ir. J.A.E. ten Veldhuis
Prof. dr. ir. N.C. van de Giesen

TU Delft Delft University of Technology

Preface

What were the odds for me to find a thesis project about Burundian farmers in the Netherlands? Almost none. Working on this particular thesis topic with a direct impact on smallholder farmers was a blissful and inspiring journey. But it has also been a roller-coaster between pain and gain. 1 month into the thesis, the COVID-19 crisis roared on one side but home welcomed me on the other side.

I am very grateful to Marc for all the meetings held and where I was invited to critically think about the results, making sure that I do not drown in the python codes and probably most importantly, try to write the report with the reader in mind. Stefan always reminded me f the importance and impact of this work but also ensured that it stays scientific enough for the university. Few months before my thesis, Nick introduced me to the TAHMO project as well as struggles of potato smallholder farmers in Rwanda. This opportunity later helped me better prepare the field visits. Marie Claire asked the questions that I am happy to have heard before my defense. I also acknowledge WeatherImpact for welcoming me in the team and showing me other parts of the project such as the seasonal forecast. AUXFIN BURUNDI has facilitated all my visits in the different provinces, making sure I visit all the stations I needed and talked to farmers with different backgrounds. Working with the WaterWatch team during the field visits and presentations at FABI and IGEBU helped me share the project with other (future) Burundian scientists and engineers. Whenever I visited farmers I told them that I have come to learn more than to teach. And I have learned quite a lot! Thanks for all the farmers who shared their joy and struggles of crowd-sourcing rainfall measurements and using the weather forecast for daily activities.

Since the beginning, I have been sharing experiences, career aspirations, etc with my classmates. Please, do not change and keep me posted again if you see any interesting position with "Africa" and "water/rain". I would also like to show gratitude to my family. For the 6 months I have been home, they made sure it was quiet whenever I have a Zoom meeting. When working to meet a deadline, they were the ones to remind me to nourish my body.

I am especially grateful for my fiancé Moses who has supported me since the start of my engineering studies up to now. You have seen my ups and downs and I could not dream of a greater support and a bigger sense of overcoming challenges.
Last but the greatest, I bless the LORD as he provided during the whole period, as he put on my path divine connections such as the women in BSF in the Hague and Trinity International Church. He also amazingly created miraculous ways through what we usually call coincidence. In your Majesty, thanks for being mindful of me!

Abstract

In this thesis, the quality and usefulness of crowdsourced precipitation measurements from 01/01/2019 to 31/01/2020 of 63 stations held by farmers in Burundi are investigated. Common sources of errors in the measurements are identified following field visits and discussions with farmers. Two filters for identifying and removing faulty zeros and suspiciously high precipitation amounts based on the measurements of neighbour stations are proposed. These filters successfully distinguish extreme rainfall events from outliers and show that within 13.5km radius, 90% of zeros corresponds to a median of 0mm/day. However, all these results are strongly affected by the number of neighbour stations. The high density crowdsourced precipitation data are also compared against the lower resolution Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) data set. Results show that although the range and the average over the whole period are almost equal for both data sets, there can be large temporal and spatial mismatches for the zero rainfall values and the extremes. Moreover, the inter-stations rainfall variability for the crowdsourced data is 4 times smaller than the stations-CHIRPS variability.

The agreement between the 0-day the European Centre for Medium-Range Weather Forecasts (ECMWF) forecast and the CHIRPS was higher than between CHIRPS and the crowdsourced rainfall data. This could be explained by the fact that both the CHIRPS dataset and the ECMWF model make use of the same satellite radiation data, making them more likely to be correlated with each other. The crowdsourced precipitation observations are used to verify ECMWF forecats delivered to farmers in Burundi by WeatherImpact. Results show that only the 0 and 1 day lead time forecasts of ECMWF are more skillful than the climatology-based forecasts and that the skill of the forecast is comparable to what is observed in the tropics. Regardless the lead time, most ECMWF forecasts tend to be better at discriminating no-rain from rain events than the climatology-based forecast.

Contents

1 Introduction **4**
- 1.1 Background . . . 4
- 1.2 Motivation . . . 6
- 1.3 Research questions . . . 8

2 Data **10**
- 2.1 Crowd-sourced precipitation measurements . . . 10
- 2.2 ECMWF precipitation forecast data . . . 11
- 2.3 Climatology data . . . 15
- 2.4 CHIRPS data . . . 15

3 Methodology **18**
- 3.1 Quality control methods . . . 18
  - 3.1.1 Radius of influence . . . 18
  - 3.1.2 Filter design . . . 18
- 3.2 Forecast verification methods . . . 19
- 3.3 Climatology-based forecast . . . 23

4 Results **25**
- 4.1 Identification of sources of errors . . . 25
  - 4.1.1 Wrong geographical coordinates . . . 25
  - 4.1.2 Station exposure . . . 26
  - 4.1.3 Rain gauge overflow . . . 28
  - 4.1.4 Reading and reporting . . . 28
  - 4.1.5 Change in stations' unique ID or locations . . . 29
  - 4.1.6 The internet connectivity . . . 31
- 4.2 Automatic quality control . . . 31
  - 4.2.1 Suspicious zeros . . . 32
  - 4.2.2 Suspicious high precipitation amounts . . . 32
  - 4.2.3 Sensitivity analysis of the filters . . . 33
- 4.3 Comparison with CHIRPS . . . 34
  - 4.3.1 Agreement with CHIRPS . . . 34
  - 4.3.2 Relationship between CHIRPS and crowd-sourced observations . . . 36
  - 4.3.3 CHIRPS sub-grid rainfall variability . . . 38
- 4.4 ECMWF forecast verification . . . 40

4.4.1 Use of CHIRPS for ECMWF forecast verification . . . 40
4.4.2 Forecast accuracy . . . 43
4.4.3 Forecast relevance for farmers . . . 44
4.4.4 Forecast skill . . . 47
4.4.5 Discrimination between rain and no-rain events . . . 50

5 Recommendations **53**

6 Conclusion **55**
.1 Farmer's instructions for crowd-sourcing rainfall . . . 58
.1.1 Visual instructions . . . 58
.1.2 Written instructions . . . 59
.2 Questionnaire: Field visit in Burundi . . . 60
.2.1 Questions to WeatherImpact . . . 60
.2.2 Questions to Auxfin . . . 60
.2.3 Questions to farmers . . . 60

References **62**

Abbreviations

CHIRP Climate Hazards Infrared Precipitation with Stations

ECMWF European Center for Medium-Range Weather Forecasts

GADM Global Administrative Areas Database

IGEBU Institut Géographique du Burundi

ROC Relative Operating Characteristics

SEEPS Stable Equitable Error in Probability Space

TIR CCD Thermal Infrared- Cold Cloud Duration

WMO World Meteorological Organisation

1

Introduction

1.1 Background

Burundi is a small country (27'834 km^2) in Eastern Africa. It lies in the tropics where it is crossed by the East African Rift from South to North (Figure 1.2). Both its geographical location and its landscape subject it to convective rainfall highly influenced by the orography. The national meteorological center, IGEBU (Institut Géographique du Burundi), is in charge of the collection and dissemination of national water resources data as well as the climate monitoring[1]. As mentioned in the Synthesis Report of the IPCC Fifth Assessment Report AR5[2], climate change risks in Africa result from both the continent's high exposure and the population vulnerability. About 95% of the agriculture in subsaharan Africa is rain-fed (Wani, Rockström, & Oweis, 2009), making it particularly vulnerable to climate change and changes in precipitation patterns.

To support the farmers in tackling these risks, the GAP4All (Good Agricultural Practices for All) project (2018-2021) is funded through the G4AW (Geodata for Agriculture and Water) programme by the Netherlands Ministry of Foreign Affairs and executed by the Netherlands Space Office (NSO) [3]. In Burundi,the project's objective is to reach and help 100'000 smallholder farmers make information-driven decisions. These farmers can then get access to the needed information through the AgriCoach app on their tablets.

As most farmers have a low education level, they were initially unfamiliar with the way precipitation is measured and what typical values they could expect (i.e. in mm/day). The initial 66 manual rain gauges that were distributed to the farmers not only increased the number of measurements points but also triggered the curiosity of the participants and provided them with the opportunity to learn more about the science behind precipitation measurement and forecasting. Furthermore, AUXFIN Burundi, one of the local partners, has put in place a hierarchical network to increase interactions with all different groups from the small-holder farmer to the national project supervisor (Figure 1.1). 50 neighbour households are gathered in a single group G50 with which AUXFIN will collaborate in activities such as teaching good agricultural practices or purchasing good quality fertilisers. The Master Activator (M.A), the Super Activator (S.A) and the Key Ac-

[1] https://www.adaptation-undp.org/partners/geographic-institute-burundi-igebu-ministry-water-environment-land-management-and-urban

[2] https://www.ipcc.ch/report/ar5/syr/

[3] https://g4aw.spaceoffice.nl/en/g4aw-projects/g4aw-projects/5/gap4a.html

tivator (K.A.) respectively oversee the farming and financial activities over a commune, a zone, a group of ±3 collines . This structure allows AUXFIN to better coordinate their activities at the national level.

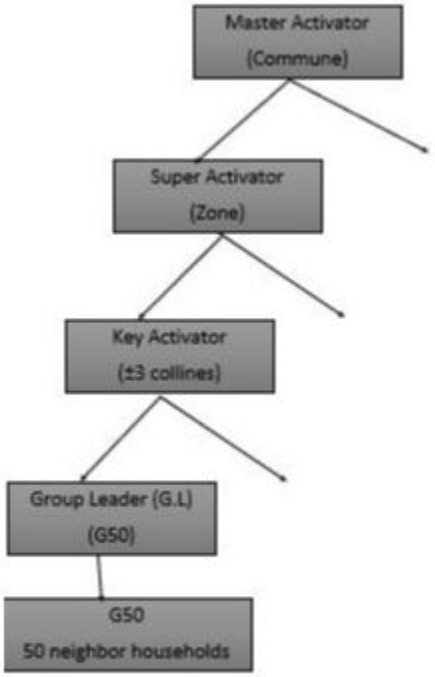


Figure 1.1: AUXFIN pyramidal network

As a result of the GAP4All project, farmers are now using the weather forecasts for planning important farming activities such as planting and harvesting.Thanks to these forecasts, they also manage the farm's labour force by reducing costs related to hired workers who could not work because of heavy rainfall. Before getting access to the information, farmers were very vulnerable to dry spells during the planting season. At the end of the day, these information-driven decisions save therefore the farmer's hard-earned money.

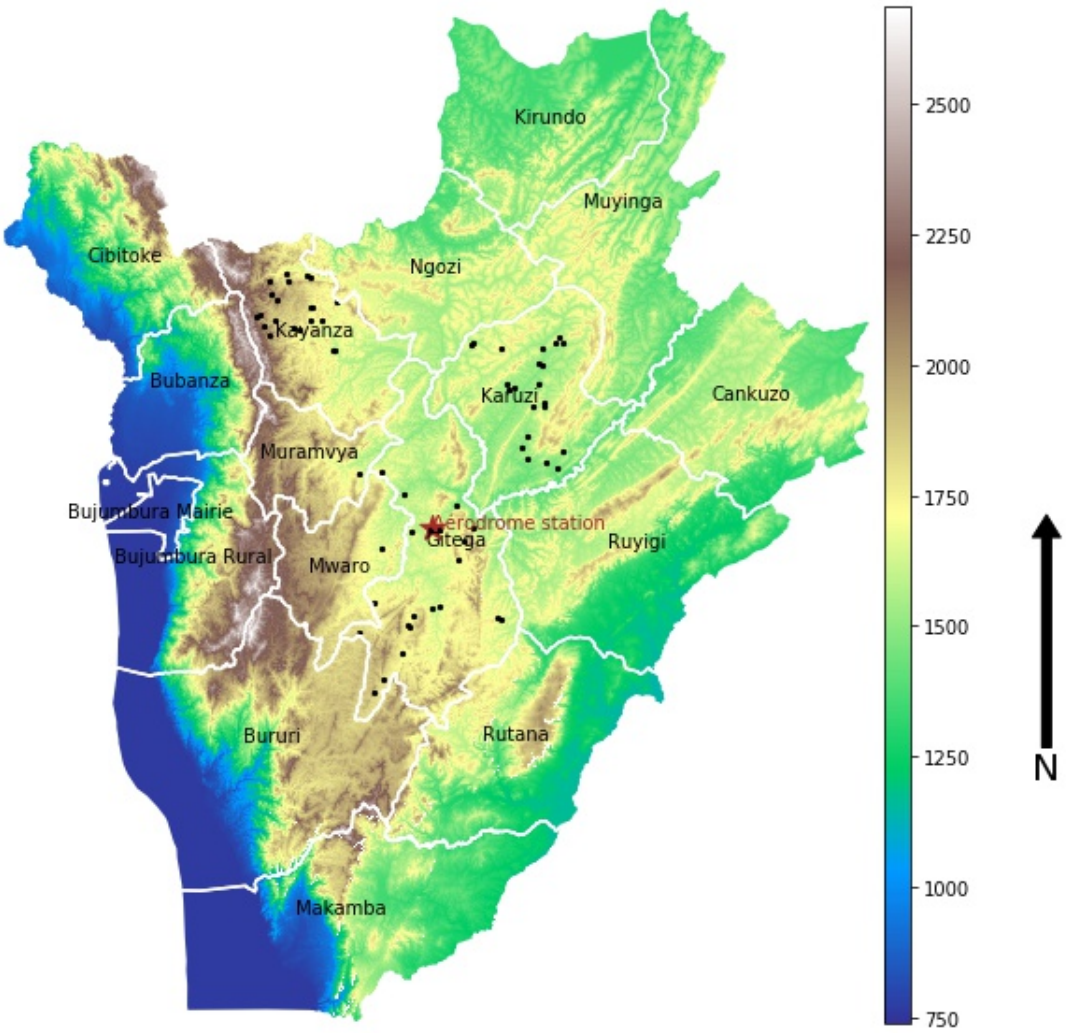


Figure 1.2: Digital Elevation Model of Burundi in meters. Source: The Shuttle Radar Topography Mission (SRTM) for digital elevation data and Global Administrative Areas (GADM) for the the administrative units boundaries. The black dots represent the reported farmers' gauge locations.

1.2 Motivation

In order to make accurate and reliable weather forecast, climate data availability and in-situ observations are crucial. However, these constitute a global challenge.For instance, the total number of rain gauges in the world does not even cover half of a football field (Kidd et al., 2017).In Africa, the amount of available data is lower than average. Burundi is unfortunately no exception to this. Moreover, the few data that are available are often not openly or easily shared and thus not very accessible (Dinku, 2019). A recent report estimates that less than half of the existing data sets are submitted to regional WMO associations in Africa (Figure 1.3).

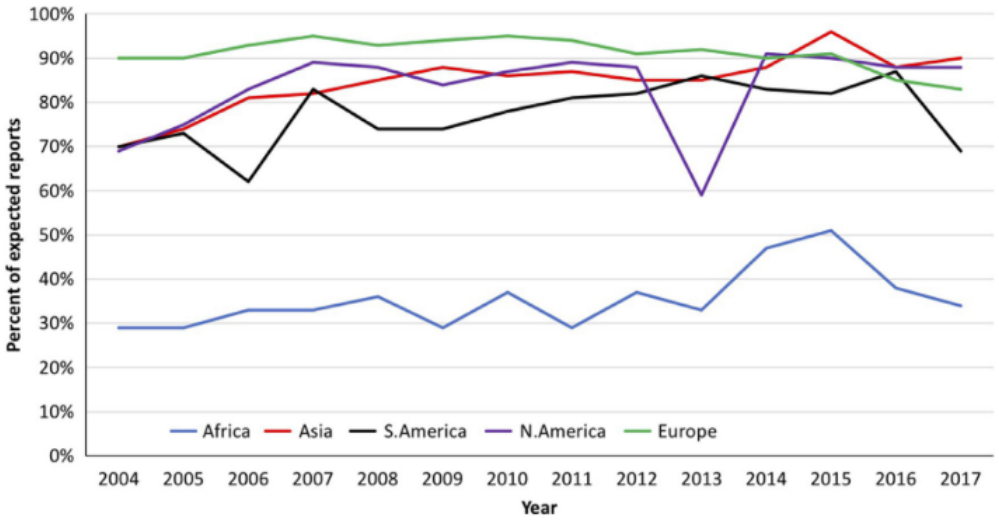


Figure 1.3: Submitted precipitation data sets to WMO [4]

One of the reasons of the reluctance to share climate data is that sometimes governments sell the intellectual property of the data to commercial companies (Overpeck, Meehl, Bony, & Easterling, 2011). Other challenges include lack of funding for performing maintenance, data sparsity and declining station density over time, especially in rural areas (Dinku et al., 2018). Moreover, socio-political and financial problems often make it difficult to access remote areas and uneven terrain. Data quality is also a challenge knowing that some institutions lack both the tools and the capacity to carry the quality control (Dinku, 2019). This is very unfortunate knowing that climate data are crucial for decision making in agriculture.

In recent years, several initiatives to tackle the issue of climate data availability and accessibility have emerged. One example is the ENACTS ("Enhancing National Climate Services") initiative launched by the International Research Institute for Climate and Society (IRI) of Columbia University supported by many National Meteorological Centers in Africa. Other efforts are aimed at improving the interpolation of existing data, such as the Global Precipitation Climatology Center (GPCC)[5], satellite-based products including the Tropical Applications of Meteorology using SATellite data and ground-based observations (TAMSAT)[6] and the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)[7]. However, satellite-based products suffer from many errors and biases and their quality heavily depends on the number of existing ground stations available for quality control and bias correction. Recently, the Trans-African HydroMeteorological Observatory (TAHMO) has started to expand the existing network of ground stations by installing a large number of automatic cost-effective hydro-meterological stations with the goal to achieve a density of 1 station per 1000 km^2 (van de Giesen, Hut, & Selker, 2014). Variables such as air temperature, relative humidity, pressure, wind speed and direction, and rainfall are measured. The network also includes stations in remote areas where regular monitoring and maintenance of the instruments is difficult.

[5]https://www.dwd.de/EN/ourservices/gpcc/gpcc.html
[6]https://www.tamsat.org.uk/
[7]https://www.chc.ucsb.edu/data/chirps

Another possibility to fill up the temporal and spatial gaps in professional weather observations is to rely on crowd-sourced precipitation measurements (Kidd et al., 2017). One good example of this is the Community Collaborative Rain, Hail and Snow Network (CoCoRaHS) [8] which is the largest daily precipitation provider in the United States. Other well-known examples include Weather Underground [9] and the Met Office Weather Observation Website (WOW) [10]. In developing countries already struggling with standard rainfall measurements collection, crowd-sourced precipitation measurements can greatly increase the number of available data. However, many obstacles related to quality control, storage and dissemination of the data still need to be overcome. In Nepal, citizens have been monitoring precipitation through the Smartphone4Water project (S4W) since 2017 using low-cost gauges made from soda bottles (Davids et al., 2019). To keep the volunteers motivated and engaged, personal messages, certificates of involvement and even payments were awarded. To encourage not only a larger quantity of the crowd-sourced data but also their quality, some assign badges in recognition of the great expertise whereas others such as the WOW prefer to use a star rating system in recognition of the quality of the instrument and the weather station exposure.

While crowd-sourcing is a good solution to increase data quantity, it does not solve the various quality issues. In fact, due to the lack of measurements standards, maintenance protocols and limited technical skills of the volunteers, many additional errors are introduced compared to standard measurements (Muller et al., 2015). Hence, before their use, crowd-sourced observations must undergo rigorous quality control. This can be achieved by validation or by verification (Tweddle, Robinson, Pocock, & Roy, 2012). Validation checks that the observations satisfy specific requirements (e.g., no negative rainfall values) whereas verification aims at comparing the data against each other or against a reference truth to detect outliers. This quality control also depends on the network stations and measurements. Recently, DeVos, Leijnse, Overeem, and Uijlenhoet (2019) developed a quality control methodology of crowd-sourced rainfall measurements with Personal Weather Stations (PWS) in the Netherlands. The particularity of this quality control mechanism is the absence of auxiliary or metadata data. Faulty zeros and high precipitation influx were filtered out based on the measurements of stations within 10 km radius as well as previous time interval rainfall values.

1.3 Research questions

The main objective of this study is to assess the quality and usefulness of crowd-sourced precipitation measurements in Burundi. Clear recommendations for how to improve data quality and reliability within the framework of the future, planned expansion of the rain gauge network will also be formulated. To this end, an automated quality control system that flags suspicious precipitation observations needs to be developed. To assess the added value of high-density rain gauge networks, we will perform detailed comparisons with satellite retrievals and daily forecasts from numerical weather models. The latter can be useful to better understand natural variability in precipitation over time and within regions as well as to reduce biases and measurement/forecast uncertainties.

[8] www.cocorahs.org
[9] www.wunderground.com
[10] http://wow.metoffice.gov.uk

Throughout this research, following questions will be answered:

1. What are common errors in daily crowd-sourced precipitation observations in Burundi and what can be done to improve the quality of the data in the future?

2. How can we filter out suspicious or erroneous crowd-sourced daily precipitation observations? And how reliable/good are these filters?

3. How well do crowd-sourced precipitation observations in Burundi compare to lower resolution precipitation data sets for the region such as CHIRPS?

4. How well do crowd-sourced precipitation measurements and CHIRPS agree with the ECMWF weather forecasts over this region? And how skillful is the ECMWF forecast compared to the climatology forecast?

To answer the above questions, the data are first analyzed for inconsistencies, such as missing data and wrong coordinates. Then, the quality of the measurement sites and day to day operation of the gauges by the farmers is assessed through field visits and interviews. Based on these findings, an automated quality control method to filter out abnormally high/low rainfall values is proposed. Then, the quality-controlled data are compared against the CHIRPS and the daily weather forecast delivered to the farmers to assess the practical value and provide recommendations for the future.

2

Data

2.1 Crowd-sourced precipitation measurements

Daily rainfall is collected with the help of manual rain gauges as shown in Figure 2.1. The rain gauge consists of a plastic bucket of a conical shape and a perforated removable plastic lid. Users can read a minimum of 1 mm of rainfall with 1 mm increment for up to 20 mm of rainfall and 5 mm from 20 mm to 70 mm of rainfall. In total, 66 rain gauges were distributed to the farmers (Figure 2.3). The result is a dense network with 0.5-8 km distance between adjacent stations, and all the stations are within $\approx 100km$ radius. Note that 3 rain gauges will not be considered for this analysis due to missing geographical coordinates and change in stations' names. For this study, only the precipitations measurements covering the period form January 1st, 2019 to January 31st, 2020 are considered leading to a total of 20342 measurements which is about 80% of expected reports.



Figure 2.1: Rainmeter[1]

Everyday at 7 a.m , farmers are expected to report the rainfall measurements via the UMVA platform. The UMVA platform was initially designed for money transfer but has since then been adjusted for "rainfall measurements transfer" as shown in Figure 2.2. The farmer has to select his group name in the *Account* drop-down whereas the Recipient is fixed to "WeatherImpact". To report a rainfall event, the farmer enters the measurement in mm in the *Amount* field, then clicks *Send* button. When there is no rain to report, the farmer directly clicks the *No Rain* button.

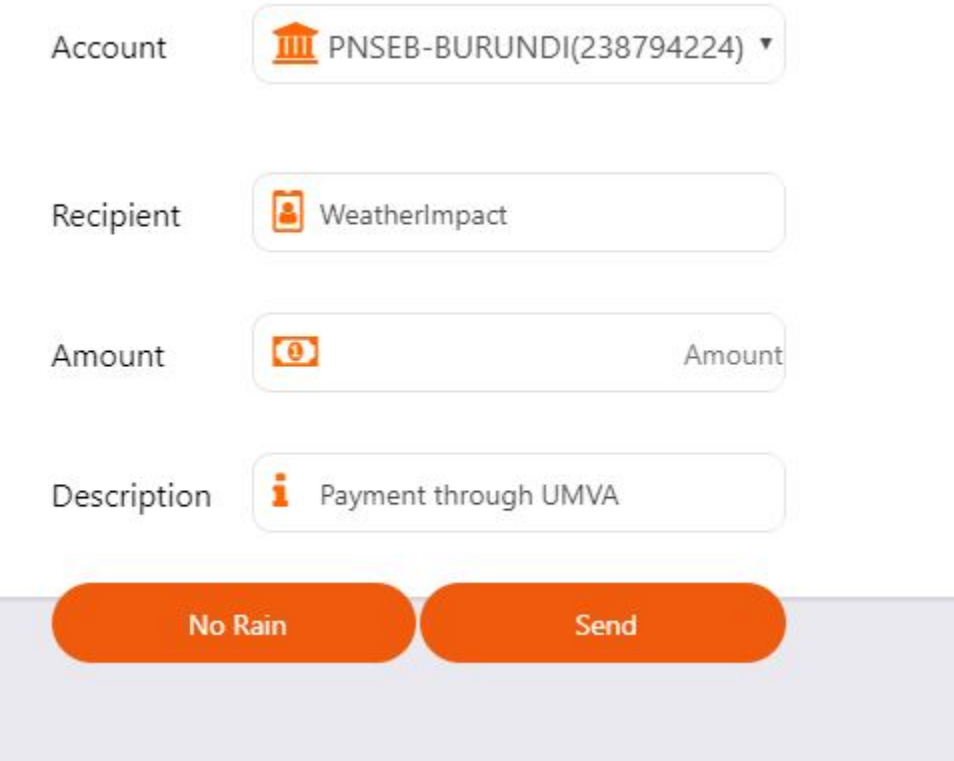


Figure 2.2: UMVA Platform

2.2 ECMWF precipitation forecast data

WeatherImpact is in charge of providing the meteorological information namely the weather and seasonal forecast to farmers. The weather forecast is generated from the ensemble forecast from the European Centre for Medium-Range Weather Forecasts (ECMWF) model (Stellingwerf, 2019). Among the various meteorological variables, temperature and rainfall forecasts are provided. Other meteorological variables such as relative humidity, pressure, wind speed as well as other precipitation forecast types such as hail are omitted. The ECMWF Ensemble Prediction System consists of 51 ensemble members with a resolution of 9×9 km. The model runs each day at 12 UTC and gives an output every 6 hours for 240 hours. A forecast day being defined as 00UTC-00UTC, the first and last hours are discarded resulting in a full 9 days forecast. These forecasts are visualised using symbols that can be easily interpreted by farmers (Figure 2.4).

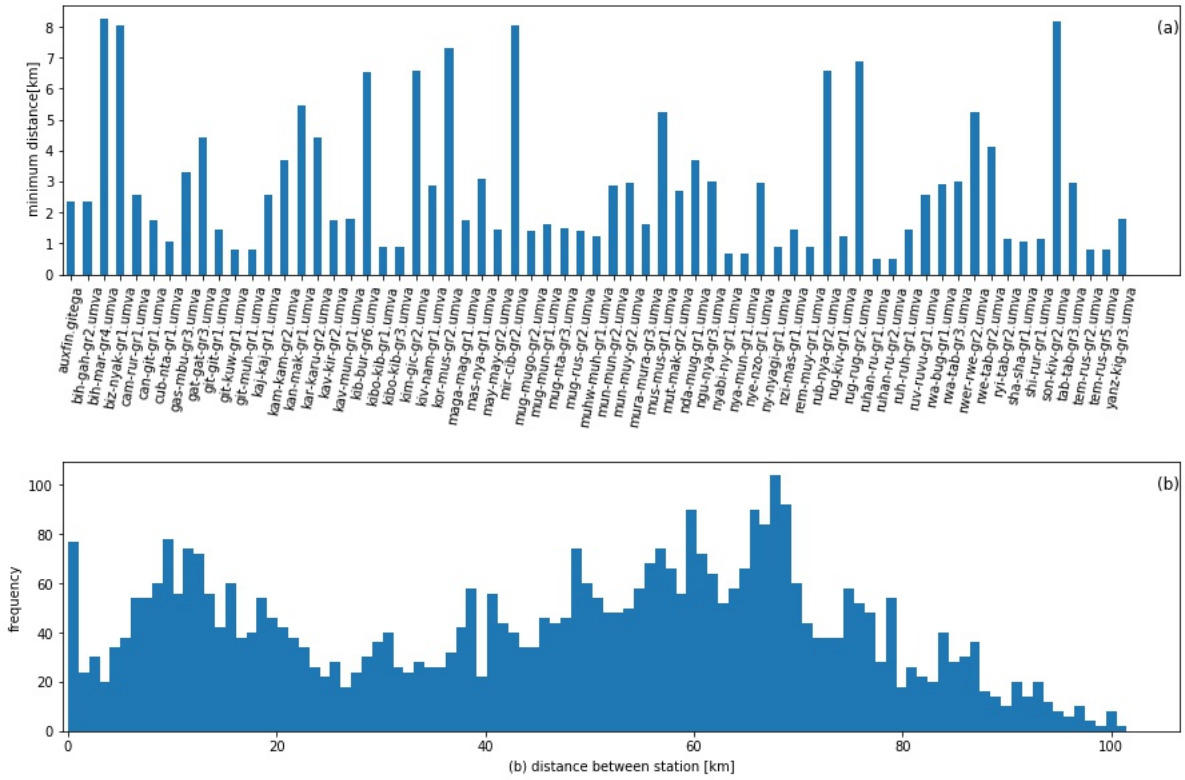


Figure 2.3: (a) The distance from the nearest neighbour station (b) The frequency of distance between the stations. Also known as a distogram

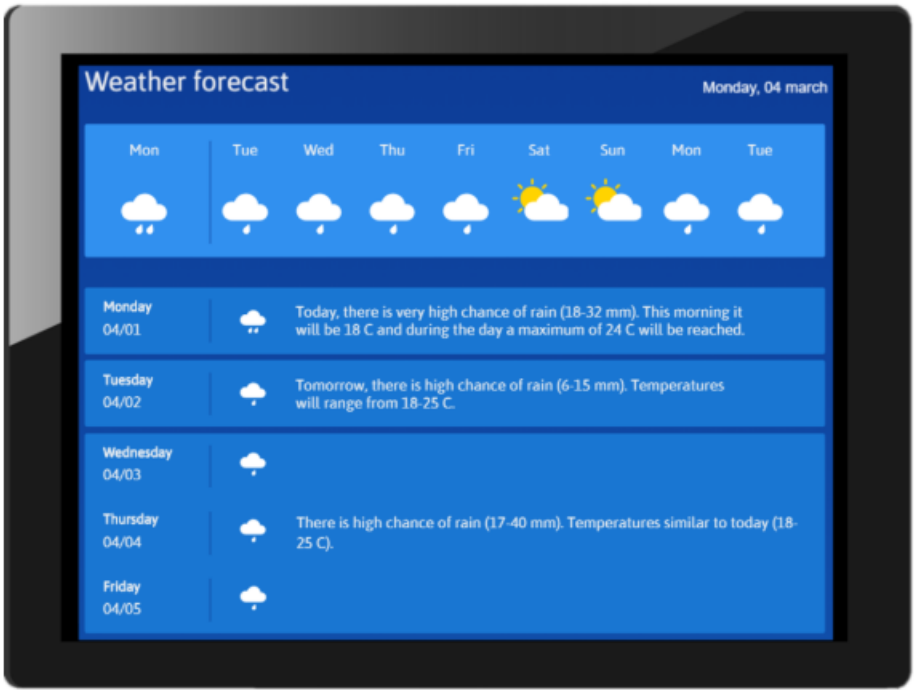


Figure 2.4: Weather Forecast in AgriCoach App[2]

The ECMWF defines the total forecasted precipitation as the sum of the large-scale and convective precipitation. [3]. To provide forecasts closer to the current observations, ECMWF model uses the short-range forecast and previous observations to make a better estimate of the initial condition of the atmosphere.[4]. Observations assimilated in the model are classified as direct/ground-based observations or indirect/satellite-based observations (ECMWF, n.d.). In 2007, 98% of the available observations were from satellites (Persson & Grazzini, 2007). This fraction probably increases when it comes to Africa where there has been a decrease in direct measurements such as from aircrafts or ground stations as discussed in Chapter 1.

The ECMWF model output is classified based on the percentage of the ensemble members exceeding a specific threshold. WeatherImpact uses 2 mm as the dry-wet threshold. As a result of this classification, five probability classes are obtained: no rain ($\leq 15\%$), low (15-40%), medium(40-65%), high (65-95%) and very high ($\geq 95\%$). The current approach is to retain the class with the highest probability among the five and use it for the forecast delivery. Later on, symbols are associated to these categories to help the farmer better understand the information.
For this study, the period of the forecast ranging from January 1st, 2019 to January 31st, 2020 was chosen to match the measurement period of the gauges. Figure 1.2 shows the landscape of the area of interest and the location of the rain gauges whereas Figure 2.5 shows the forecast grids location with respect to the location of the rain gauges.

[3]ECMWFDetails

[4]`https://www.ecmwf.int/en/about/media-centre/news/2020/fact-sheet-earth-system-data-assimilation`

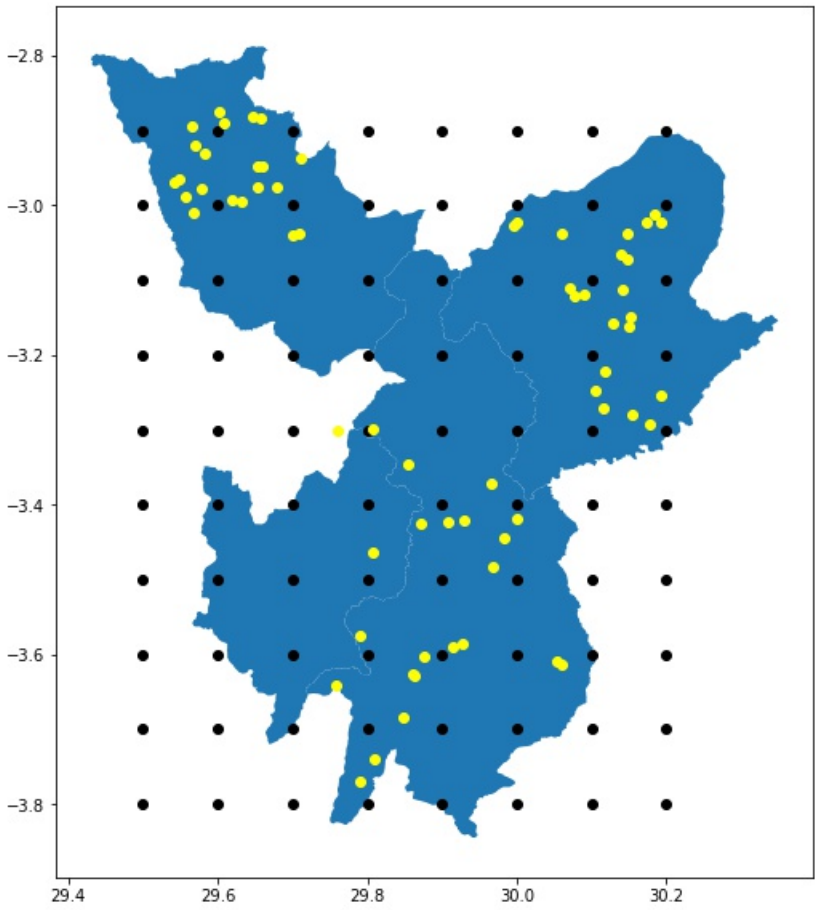


Figure 2.5: Rain gauge stations (yellow) and forecast grid (black)

2.3 Climatology data

In addition to the crowd-sourced and forecast precipitation data, long-term precipitation data are used as an indicator of the climatology of the region. IGEBU has been recording daily rainfall at the aerodrome of Gitega (Figure 1.2) since 1964. Unlike many of the other stations, the recordings did not stop during the civil war in 1993 and therefore the missing data account for only 0.31%. For this study, observation data from February 1964 to December 2009 are used.

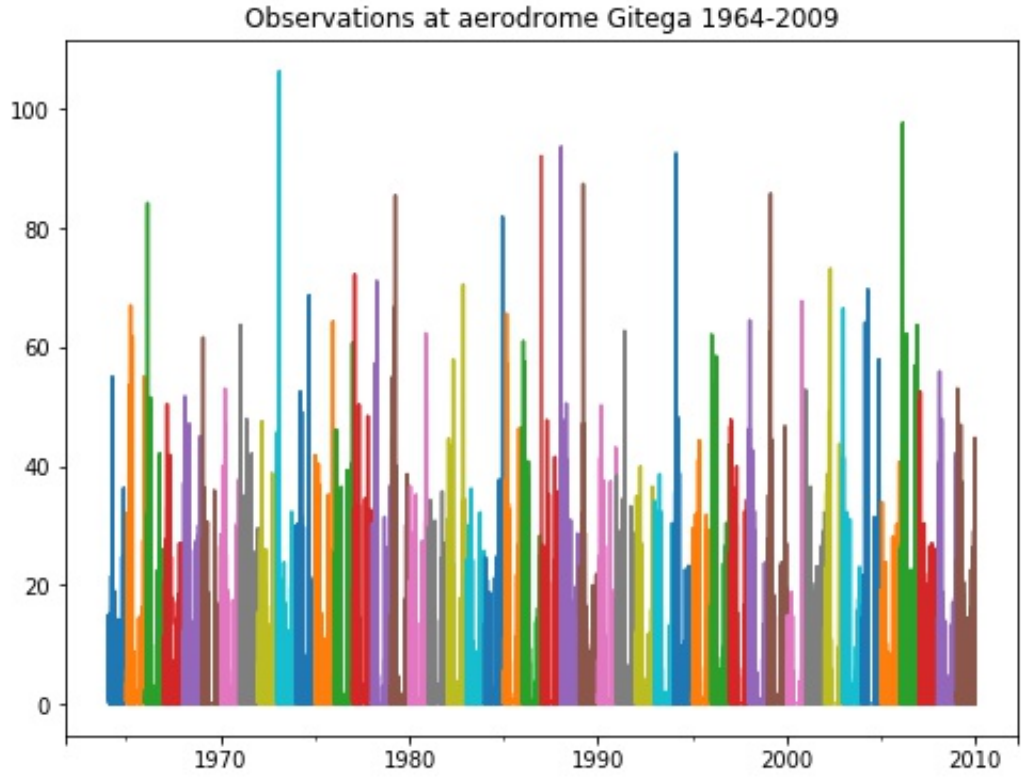


Figure 2.6: Rainfall measurements at Gitega aerodrome from February 1964 to December 2009

2.4 CHIRPS data

CHIRPS stands for the Climate Hazards Infrared Precipitation with Stations (Funk et al., 2015). It covers the earth from 50°N- 50°S with a resolution of 0.05°. The data set is available from 1981. CHIRPS has been designed with agriculture monitoring in mind. It has been developed by the United States Geological Survey(USGS) and the University of California, Santa Barbara. It builds on the CHIRP and gauge observations. CHIRP calibrates the climatological mean from (CHCclim) with statistics of historical thermal infrared cold cloud duration (TIR CCD). When stations are available CHIRPS use them for bias correcting CHIRP. Based on the description of Funk et al. (2015)of the CHIRP and CHIRPS data sets, a summary of how they were produced is presented in Figure 2.7 and 2.8.

Although few of the stations used in CHIRPS were found in Burundi in the 80s, CHIRPS does not integrate any station measurements from Burundi at the moment as seen in Figure 2.9. Knowing

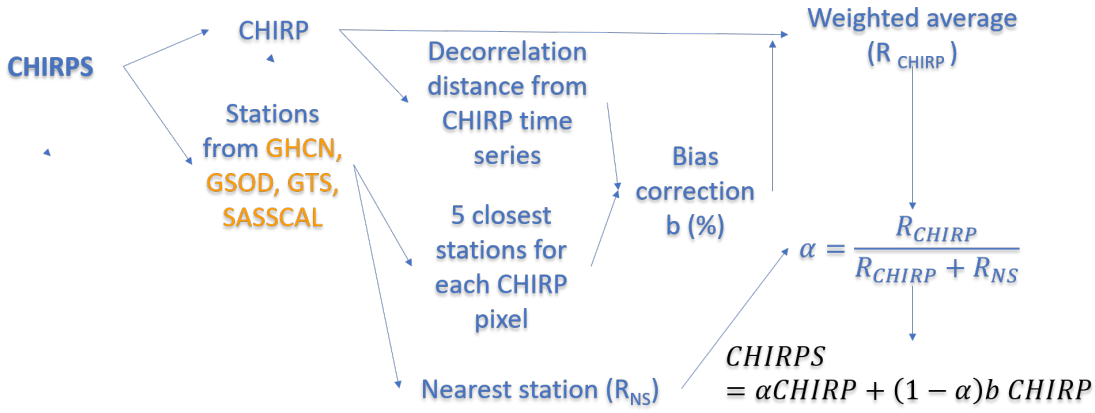


Figure 2.7: Production of CHIRPS data set

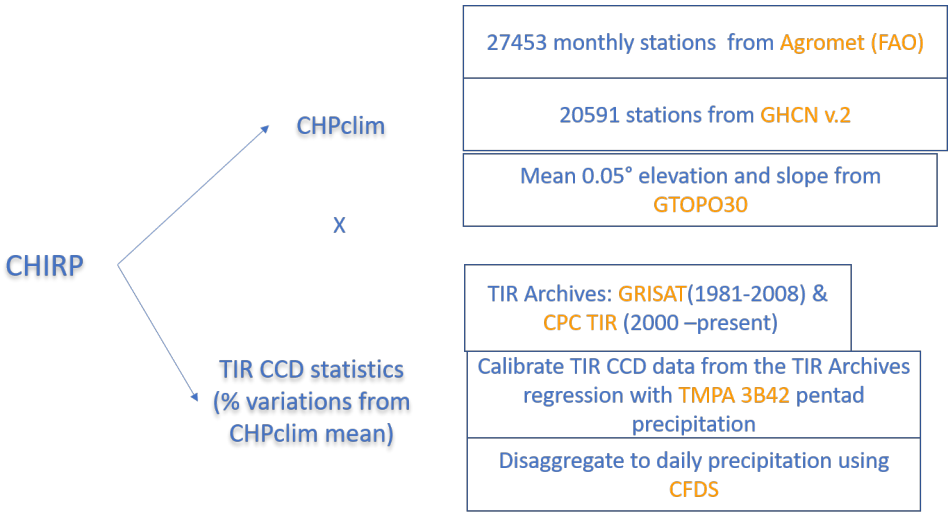
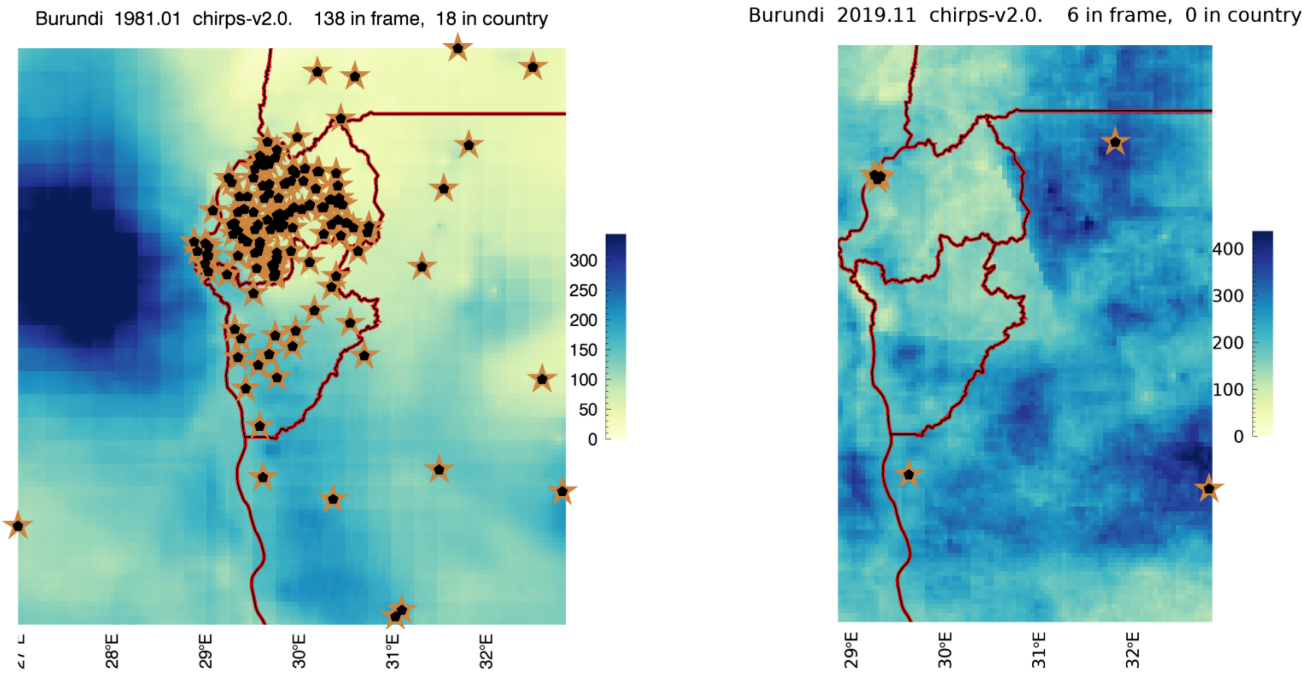


Figure 2.8: Production of CHIRP data set

that the stations are used for bias correction (Figure 2.8), the absence of these ground measurements result in precipitation observations directly correlated to TIR CCD statistics. In this particular case, CHIRPS reflect therefore more the cloud observations rather than precipitation observations.



(a) 01/1981

(b) 11/2019

Figure 2.9: rain gauges in CHIRPS data set

3

Methodology

3.1 Quality control methods

Sources of irregularities and discontinuities observed during the site visits are first described. Then, unrealistic precipitation measurements are flagged based on neighbour stations within a pre-defined radius of influence depending on whether it is a dry or wet day.

3.1.1 Radius of influence

The convective type of precipitation in the tropics implies that unusually high precipitation values compared with neighbour stations are not necessarily erroneous. Rainfall intensities can greatly vary over a scale of few hundred of meters only, making it difficult to apply spatial filters at small temporal scales (Goodrich, Faurès, Woolhiser, Lane, & Sorooshian, 1995). However, the daily scale of the precipitation measurements helps smooth out rainfall spatial variability for very close rain gauges. The World Meteorological Organisation (2014b) recommends at least 1 precipitation recording station for 575km^2 in hilly areas. Therefore, for this study, all measurements recorded at neighboring stations within an area of 575 km^2 i.e. a radius of 13.5km from the target gauge are considered for our quality control. Thanks to the high density of the rain gauge network (Figure 2.3), this results in a number of neighbour stations ranging from 2 to 20.

3.1.2 Filter design

Thresholds above which a measurement is considered suspicious are derived from histograms (Figure 3.1 & Figure 3.2.

For the quality control purpose, the rain and no-events are analysed separately.For the no-rain events, the median of the surrounding station measurements is calculated and compared with the reference station measurements(Figure 3.3). The difference should not exceed a statistical defined threshold. This will be referred as "the occurrence filter" in the following sections.

Given the nature of convective and orographic rainfall in this region, the median of surrounding stations should not be used for rain events as DeVos et al. (2019) used it for the Netherlands because the country landscape is flat. In this area, rainfall can vary within hundreds of meters without being necessarily an error. For this reason , an alternative filter based on the minimum difference and minimum ratio between rainfall at the station and a station within the radius of influence. The minimum difference is calculated by taking the minimum of the absolute difference between the

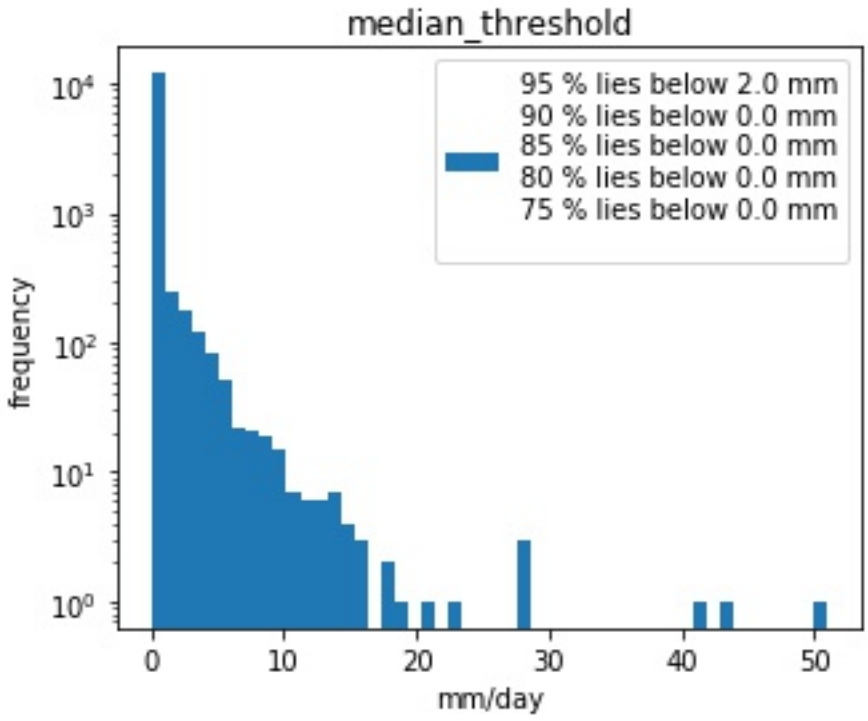


Figure 3.1: Occurrence filter: distribution of the **median** values

reference station and the neighbour stations measurements.The minimum ratio is the ratio of this minimum difference by the observed value at the reference stations. This helps put this difference between 2 stations in perspective with the magnitude of the rainfall value. This filter will hereafter be referred as "the intensity filter" (Figure 3.4).

For the occurrence filter, the threshold will be fixed to 95% percentile i.e. 2mm/days. Similarly, for the intensity filter, 95% percentile threshold corresponds to the minimum difference and minimum ratio of respectively 6mm/ day and 1.

3.2 Forecast verification methods

The objective of forecast verification is to assess the agreement of the forecast with respect to the crowd-sourced precipitation measurements. In fact, farmers may change unnecessarily their field work plan due to a forecast's false alarm or field work may be interrupted because of an unexpected rainfall resulting in money loss. Therefore, based on the forecast user's needs, three important characteristics of the forecast will be analysed:

1. Accuracy of the forecast: A forecast is defined as accurate when the observed value falls within the forecast range.

2. Skill of the forecast using the Relative Operating Characteristics (ROC) curve (Stanski, Wilson, & Burrows, 1989) and SEEPS score (Rodwell, Richardson, & Hewson, 2010).

3. Discrimination between rain and no-rain events using discrimination diagrams (Wilks, 2019)

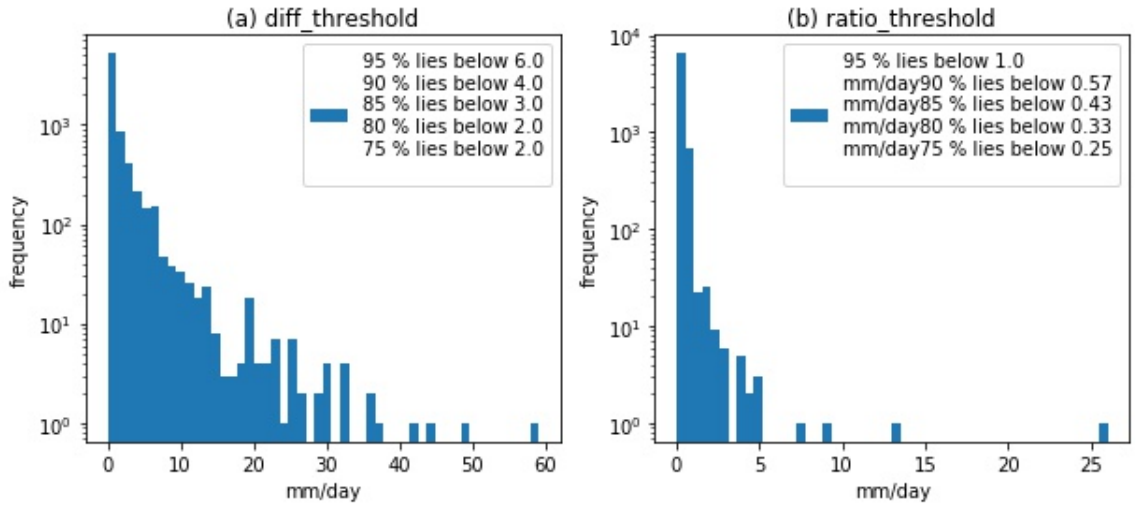


Figure 3.2: Intensity filter: distribution of (a) **minimum difference** (b) **minimum ratio** values

Given that the precipitation forecast is a 9 × 9 km grid box average, it cannot and should not be directly compared to the crowd-sourced point observations. To make a fair comparison between the 2 data sets, one should calculate the average of the point observations in a grid box [1]. However, this requires very dense rain gauge networks and is sensitive to the sample size. In our case, the maximum number of gauges within a grid box is 6 and only 2 out of 24 grid boxes contain the maximum number of 6 stations. Among these two, we selected the grid box with the highest number of days without any missing observations (i.e., 259 daily observations in 2019). The evaluation focuses on a 1 year period in 2019.

[1] https://apps.ecmwf.int/codes/grib/param-db?id=228

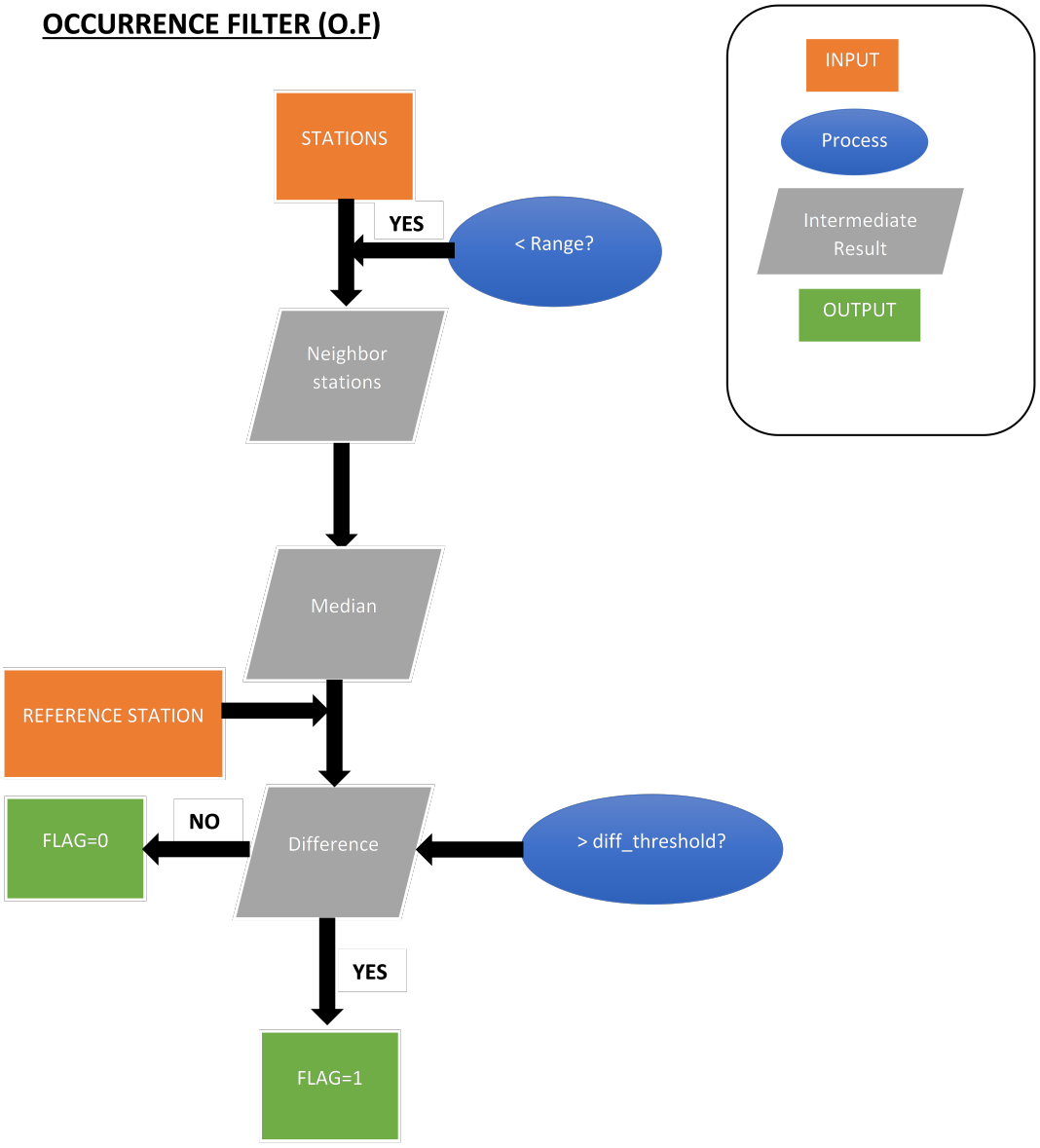


Figure 3.3: Flowchart of occurrence filter design

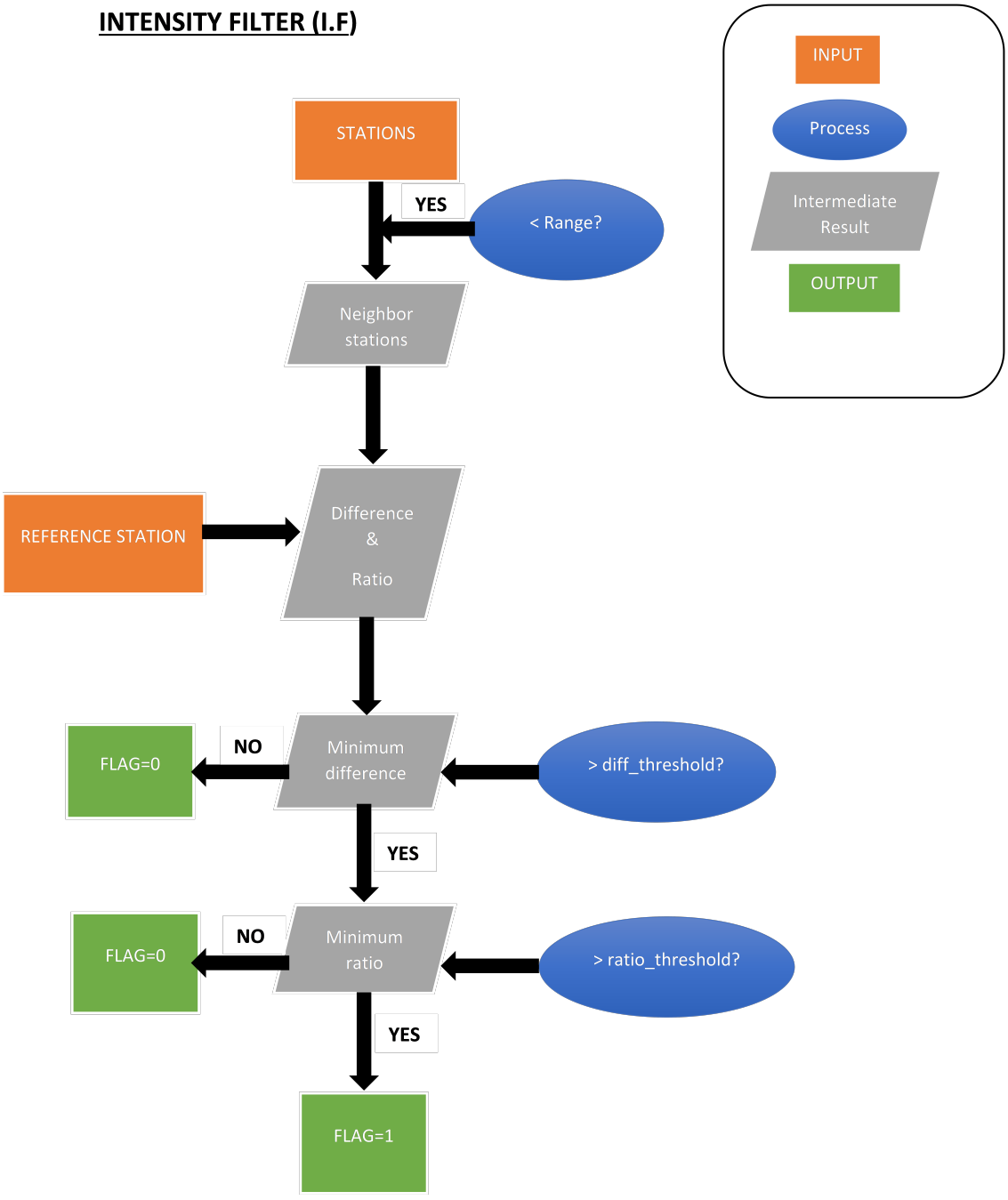


Figure 3.4: Flowchart of intensity filter design

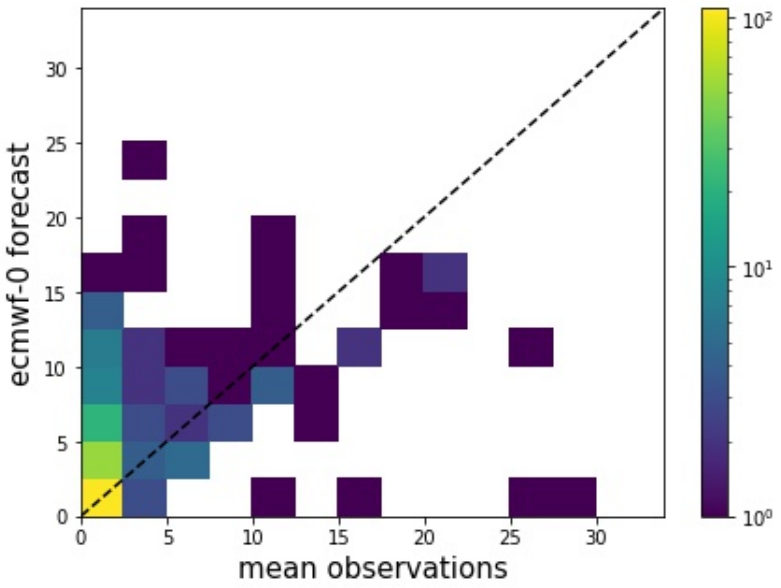


Figure 3.5: 2D histogram of average precipitation values (in mm/day) from the 6 crowd-sourced gauges (X-axis) within the forecast grid box centered at (30.1°E, -3.1°S) vs the median of the ensemble forecast (Y-axis) for this grid box. The model is run at about 00 UTC whereas the observations are reported around 05 UTC, resulting in a lead-time of about 5 hours.

The Relative Operation Characteristics curve (ROC) shows the hit rate (percentage of the recorded wet days that were rightly forecasted as wet days) against the false alarm rate (percentage of the dry days that were forecasted as wet days). The probabilities of having a rain day are classified in 10 bins with 0-10%, 10-20% and etc... up to 90-100%. For each probability class, the hit rate and the false alarm rate are calculated. The results for all the probability classes and all forecast lead times can be visualized in a single plot known as the ROC curve.

The discrimination diagram 4.25 shows the ability of the model to distinguish wet and dry days. This diagram is plotted for different lead times and compared against the climatology. It shows the full joint distribution of the forecasts and the observations. Given that it rains or not, the likelihood for each class of the forecast probability is plotted. The distance between the 2 distributions shows the ability of the forecast to discriminate between the rain/ no rain event. For a perfect forecast, there is no overlap between the 2 distributions .

3.3 Climatology-based forecast

To assess the skill of ECMWF forecast, a reference forecast is needed. For this purpose, a climatology-based forecast is developed from the climatology data at the aerodrome in Gitega (see Section 2.3). We assume here that a uniform climatology applies over the whole region of interest and that all observed values are equally probable. A moving average with a window of $\pm$ 5 observation days is applied to remove noise (Figure 3.6). The result is then grouped by similar days and months but different years and the average value of each group is calculated. The output is made of daily precipitations values for each specific day of the year. Thereafter, in order to calculate

the probability of a rain event for a specific date, similar dates for the same year with a window of ± 5 days are grouped with a maximum size of 45 (years) x 11(days) elements. The observed precipitation on those days are compared with the wet/dry threshold of 0.2mm/day (WMO, 2014b) and the probability of a rain event is calculated from the number of days exceeding the threshold. The final result is a rainfall amount and probability for each day of the year.

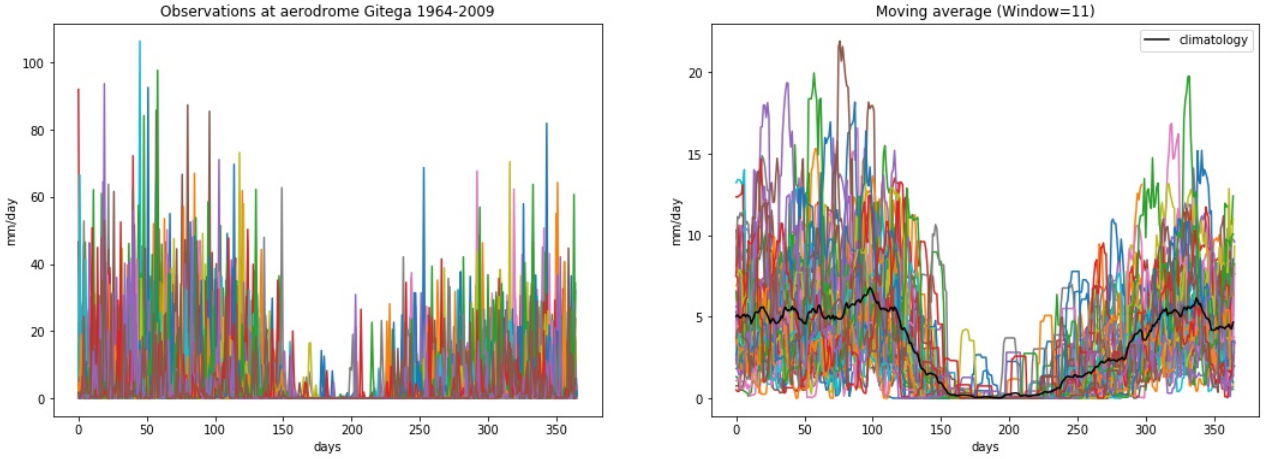


Figure 3.6: Left figure: Raw rainfall measurements for different years(1964-2009) at a specific day .Right figure: $\pm 5 days$ window moving average.The black line shows the value of the climatology-based forecast for each day of the year

4

Results

4.1 Identification of sources of errors

The collection and reporting of precipitation measurements are prone to irregularities that can lead to errors in the dataset. To get a better overview of the situation, field visits to the rain gauges stations were undertaken from 11-13 March, 2020 as well as on March 27^{th}, 2020. A total of 10 farmers and 8 stations were visited in 2 different provinces. The following section describes irregularities observed when farmers are measuring precipitation .

4.1.1 Wrong geographical coordinates

During the rain gauge installation, the geographical coordinates of the rain gauge stations were retrieved with the help of a GPS tracker. However, often, the initial coordinates had to be updated several times as they were not matching the actual position. For example, in some cases, coordinates of neighbouring countries were reported. Nevertheless, even the most updated coordinates still contained errors. In particular, one wrong coordinate corresponding to the `muh-gah-gr1.umva` station was found. Consequently, this station was removed from subsequent analyses due to the uncertainty surrounding its location. The time series shows that the removal of the station resulted in valuable information loss as shown in Figure 4.1.

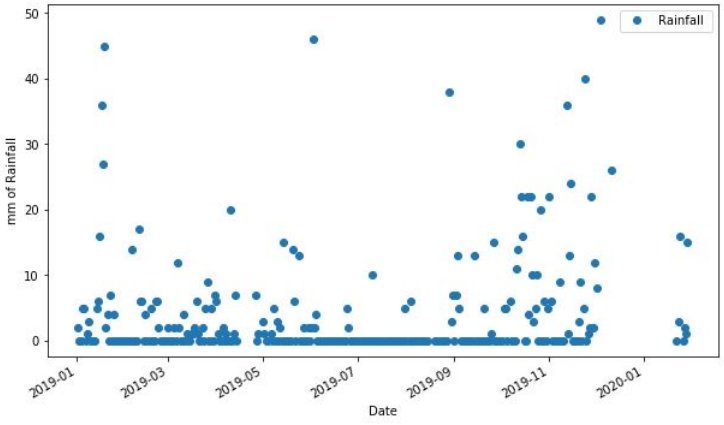


Figure 4.1: `muh-gah-gr1.umva` station time series

In addition to removing stations with obviously wrong coordinates, all the coordinates were checked against the Global Administrative Areas (GADM) dataset [1]. Province, commune, colline and locality names were compared for the stations geographical coordinates and classified into different categories as shown in Figure 4.2. For more than 50% of the stations, the Province, Commune, Colline and Locality names matched for both data sets. Although 5 out of 65 stations had different Province or Commune names according to the GADM dataset (Table 4.1, they were reported in the neighbour Province/ Commune. Therefore, this discrepancy could be explained by errors in drawing administrative boundaries in either of the data sets.

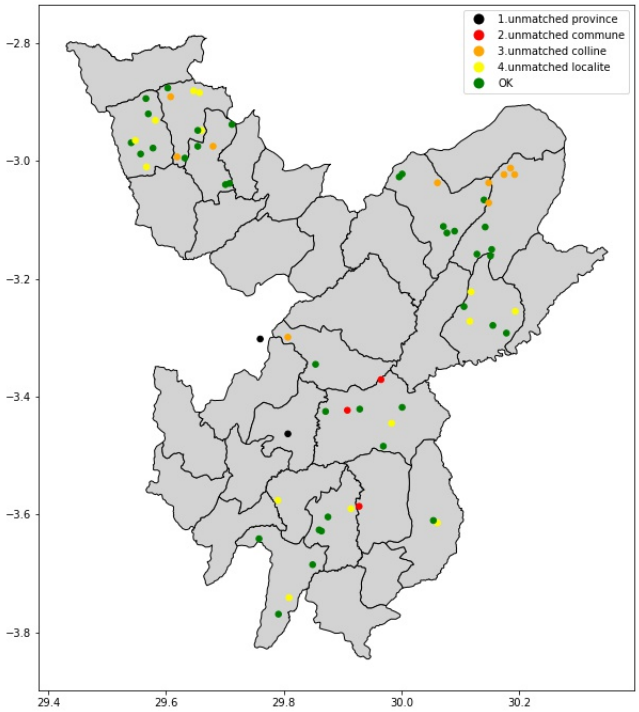


Figure 4.2: Geographical coordinates check

Table 4.1: Number of stations with matching administrative names with GADM data set

	Provinces	Commune	Colline	Locality
Number of stations	63	60	50	36

4.1.2 Station exposure

The stations were not necessarily fully exposed as recommended by WMO (2014b) which recommends that objects should not be within a distance of twice their height above the gauge. This prevents shielding and mitigates the bias due to wind effects. WMO also recommends that in case

[1] https://gadm.org/data.html

there is some dense vegetation around the gauge, the latter should be regularly clipped at equal height than that of the gauge.

Figures 4.3a & 4.3b show that although the stations do not stand directly under a tree, they often have many obstacles in their vicinity especially tree branches. This is because the farmers had to compromise optimal exposure with security, in order to prevent theft or vandalism.



(a) `auxfin.burundi` station(10/03/2020)



(b) `son-kiv-gr2` station(26/03/2020) /2020) /

Figure 4.3: Station (a) is installed at AUXFIN headquarters in Bujumbura whereas station (b) is in a farmer's field in Gitega. None of the station is fully exposed as branches of trees taller than the station pole extend above the gauge openings

It is also worth noting that the poles on which the gauges are installed are not standard and neither are their heights. Therefore, some stations are a few centimeters above the ground whereas others are about a person's height (Figure 4.4). This may lead to additional discrepancies in the collected rainfall data as 2 rain gauges at the same location but different heights do not necessarily report the same measurements (Rodda & Dixon, 2012).



(a) `auxfin.gitega` station(12/03/2020)



(b) `mun-mun-gr2` station(11/03/2020)

Figure 4.4: Both pictures are taken in Gitega province. The difference in collected rainfall can be influenced by this observed difference in pole heights and not only the different locations.

4.1.3 Rain gauge overflow

The rain gauges delivered to farmers have a capacity of 70 mm. Although, it does not occur very often, it has happened at least 12 times that a gauge overflows due to a heavy rainfall event. Sometimes even, this maximum limit was reached by 2 different gauges on the same day.

In case of overflow, farmers simply report the maximum precipitation they could read on the gauge i.e. 70 mm. For this reason, very extreme precipitation observations cannot be accurately measured though these are of greater concern for the farmers as they usually cause a lot of damages in the fields like floods, landslides, etc.

4.1.4 Reading and reporting

Although these types of rain gauges are quite simple to read, potential measurement errors can be ascribed to inclined gauges or difficult to read values due to dirt or the growth of vegetation and other organisms inside the gauge (see Figure 4.5). This is especially the case for light precipitation i.e. when the readings are taken at the gauge base.
While reporting the measurement, it can happen that the user makes a wrong click or erroneously enters a value. Unfortunately, these can not be rectified later. Another concern is the reporting time. Although farmers report to take the measurements early morning at 7 a.m -8 a.m, the reporting time in the AgriCoach app corresponds mostly to afternoon and evening time. A further investigation has shown that this delay is linked to the internet connectivity issues as discussed in Section 4.1.6.



Figure 4.5: An unclean gauge makes the reading of the measurements very difficult, especially for light rain .Picture taken on 11/03/2020 in Kayanza.

4.1.5 Change in stations' unique ID or locations

Unique IDs have been attributed to every single station. Those IDs are made from the first letters of the *Colline, Locality* and *Group* names. However, collected data show that some stations have changed their names probably due to an updated group name as the colline and locality remains the same. The error was probably the result of checking and updating the coordinates. Figure 4.6 shows the nyak-rut-gr1 & nyak-rut-gr2 stations rainfall data. The station `nyak-rut-gr1` of coordinates (-3.445, 29.9832) was changed to station `nyak-rut-gr2` with coordinates (-3.445, 29.983).

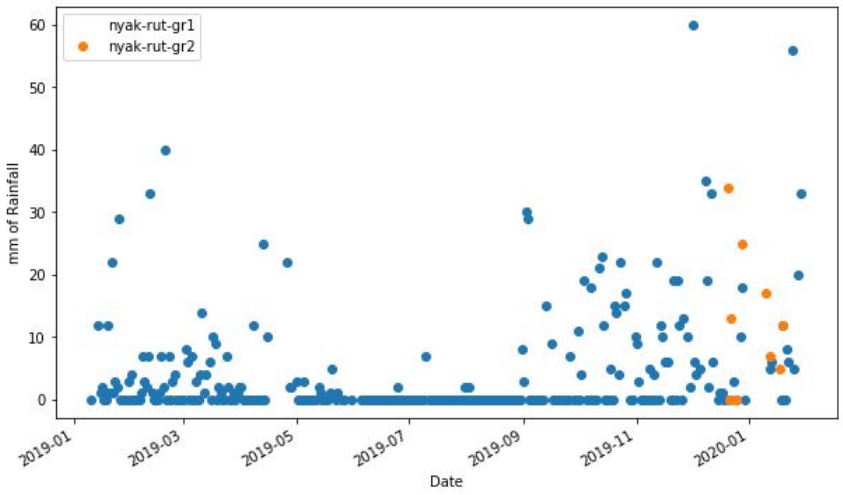


Figure 4.6: `nyak-rut-gr1` & `nyak-rut-gr2` overlapping time series

However not all the changes can be captured by the stations' time series. In fact, during the visit in Gitega, the rain gauge manager also known as the "Rain master" of the station `kibo-kib-gr3` reported to have shifted the station to about 100 m of distance due to wood-boring insects that were threatening the gauge pole. As a result of this station shift, the station was within few tens of meter from another station `kibo-kib-gr1`. While this proximity is great for the quality control, it also meant that the measurements of this station stopped being reported for a while (Figure 4.7). This might be the consequence of the additional load on a single person to measure daily precipitations from 2 different stations or the fact that the user thought that it was not necessary to report the values of two stations very close to each other. This is especially challenging knowing that the locality is suffering heavily from a low internet connectivity and thus, the capacity to report smoothly the measurements.

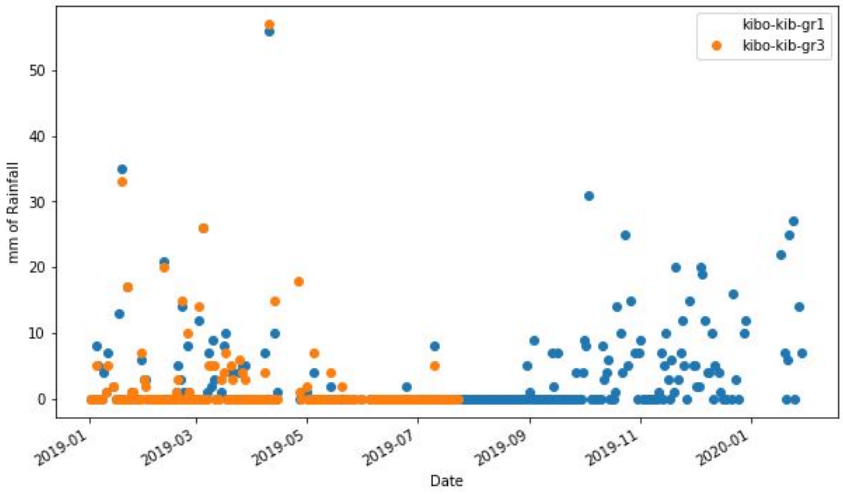


Figure 4.7: kibo-kib-gr3 interrupted time series

4.1.6 The internet connectivity

One of the main challenges linked to the discontinuity in the reported measurements is the internet connectivity. The tablets are equipped with dual-sim cards. Regardless, internet seems to be almost totally absent in some areas and periods of time. As a consequence, people have to move to a nearer hilly or more urban location to transfer the measurements. For these reasons, some measurements made are never reported and thus, lost forever. Additionally, farmers who are not regular smartphones users receive data bundles each month for the various tablet applications. This ensures that all farmers, whether they can afford it or not, are able to use the internet on a daily basis. However, irregularities in activating the bundles for the farmers' network may result in the whole system's shutdown. This scenario was observed shortly in April 2019 and for a much longer period in November 2019 - January 2020 (Figure 4.8).

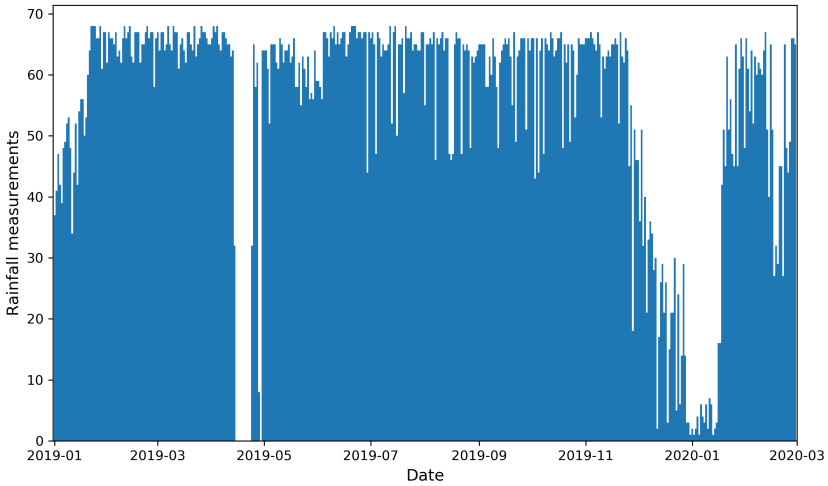


Figure 4.8: Overall missing data gaps

In the absence of internet connectivity, most of Group Leaders (G.L), try to report the measurements via phone calls and/or whatsapp. Hence, they submit the measurements to the Key activators (K.A.), Super Activators(S.A) or Master Activators(M.A) so that they report the measurements in the system on their behalf. In this case, the issue is that a single person is responsible of reporting many measurements gathered via phone call, whatsapp or even face to face meeting. When the person forgets to enter the measurements in the system, measurements from several stations or entire neighbourhoods or regions will be missing.

4.2 Automatic quality control

The aim of the following quality control system is to flag suspicious zeros and high precipitation measurements as a result of previously identified sources of errors.

4.2.1 Suspicious zeros

Figure 4.9 shows the results of applying the occurrence filter. We can see that on that day within the radius of influence, all 20 stations recorded rain. The median of those measurements is 28.5 mm/day is much greater than the 4 mm/ day defined threshold in Section 3.1.2. Therefore, it is suspicious that no rain fell at that location for that specific day and hence this reported measurement is flagged. In total, 1.96% of the zero measurements are flagged.

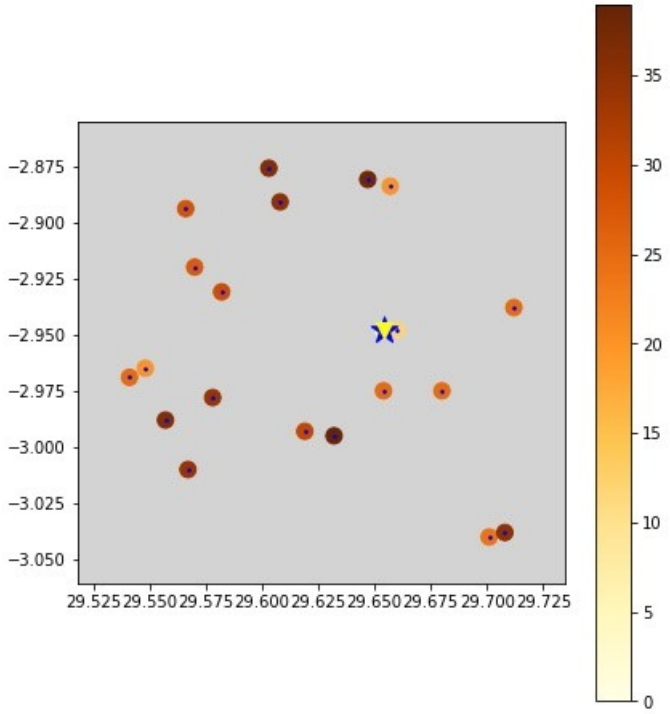


Figure 4.9: Filtered zero precipitation measurement at station `nya-mun-gr1.umva` on October 14^{th}, 2019. The right figure shows all stations measurements for that day and the left figure zooms in stations within the radius of influence. The yellow marker shows the flagged station whereas the blue marker shows the station measurement which is quality controlled. A blue dot with no halo shows a station that did not report any measurement.

4.2.2 Suspicious high precipitation amounts

Figure 4.10 shows the application of the intensity filter. Within the radius of influence, 13 stations have reported 0 to 2mm whereas 13mm have been reported at the station of interest. Therefore, the minimum difference (11 mm) and ratio (0.15) between this station and neighbour stations' measurements is respectively above 9mm/day and below 0.5 according to the intensity filter (section 3.1.2). This high amount of rainfall at that station is therefore suspicious with respect to what has been observed in the neighbourhood. In total, 1.93% of the rain events are flagged.

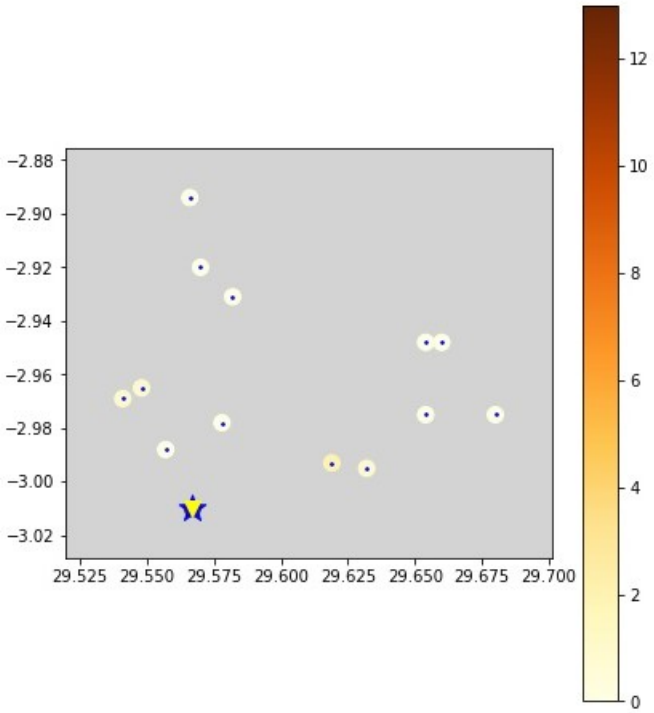


Figure 4.10: Filtered precipitation measurements at station `mut-mak-gr2.umva` on May 31^{th}, 2019. 13 mm of precipitation was recorded at this station.The yellow marker shows the flagged station and the blue marker shows the station which is quality controlled.A blue dot with no halo shows a station that did not report any measurement.

4.2.3 Sensitivity analysis of the filters

The filters are sensitive to the number of neighbour stations. Figure 4.11 shows that stations with lots of neighbors have smaller minimum differences on average. This makes them less likely to exceed the critical threshold and . Moreover, the distribution over the minimum difference classes becomes less spread with increasing number of stations, thus increasing the confidence of interval when filtering out measurements. For example, The figure shows that increasing the number of stations in the neighborhood from 3 to 12 results in an increase of the probability of having a minimum difference of 0 mm from 0.2 to 0.65. Therefore, flagged measurements with many neighbouring stations are more suspicious than those having only a few neighbours.

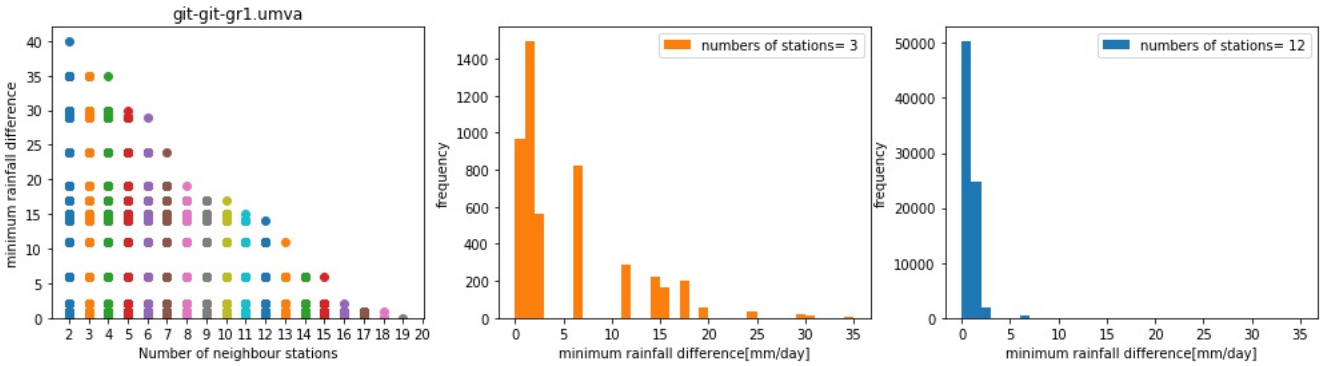


Figure 4.11: Station `git-git-gr1.umva`. Left figure. Decrease in minimum difference with increasing number of stations. Center and Right figure. Change of minimum rainfall difference frequency with increasing number of stations.

4.3 Comparison with CHIRPS

The previous sections dealt with an internal assessment of the crowd-sourced precipitation data set. However, in order to understand its contribution to existing precipitation data sets and in the absence of a true ground precipitation data set , CHIRPS and ECMWF precipitation forecast are used. This section compares the crowd-sourced rainfall observations to the CHIRPS observed rainfall whereas the next section deals with the precipitation forecast.

4.3.1 Agreement with CHIRPS

As far as satellite products like CHIRPS are concerned, we want to know whether they reflect direct rainfall measurements. Figure 4.12 shows the number of agreements between CHIRPS and all stations measurements from January 2019 to January 2020. Both data sets are grouped into bins of 5mm/day except the first bin which contain zeros for the crowd-sourced observations and 0 to 0.5mm for CHIRPS. During this period, there is almost no bias between the CHIRPS and crowd-sourced observations values. Their respective average over the whole year is respectively 3.616mm/day and 3.628 mm/day. This is in line with previous research and the good performance of CHIRPS at annual time scale (Rivera, Hinrichs, & Marianetti, 2019). Interestingly, CHIRPS and citizen gauges report similar ranges of daily rainfall i.e 0-80 mm and 0-70mm respectively. The difference in the maximum limit can be ascribed to the insufficient size of the rain gauges. This differs much from the forecast range (0 to about 35mm). Even so the extremes of the 2 data sets do not take place simultaneously.

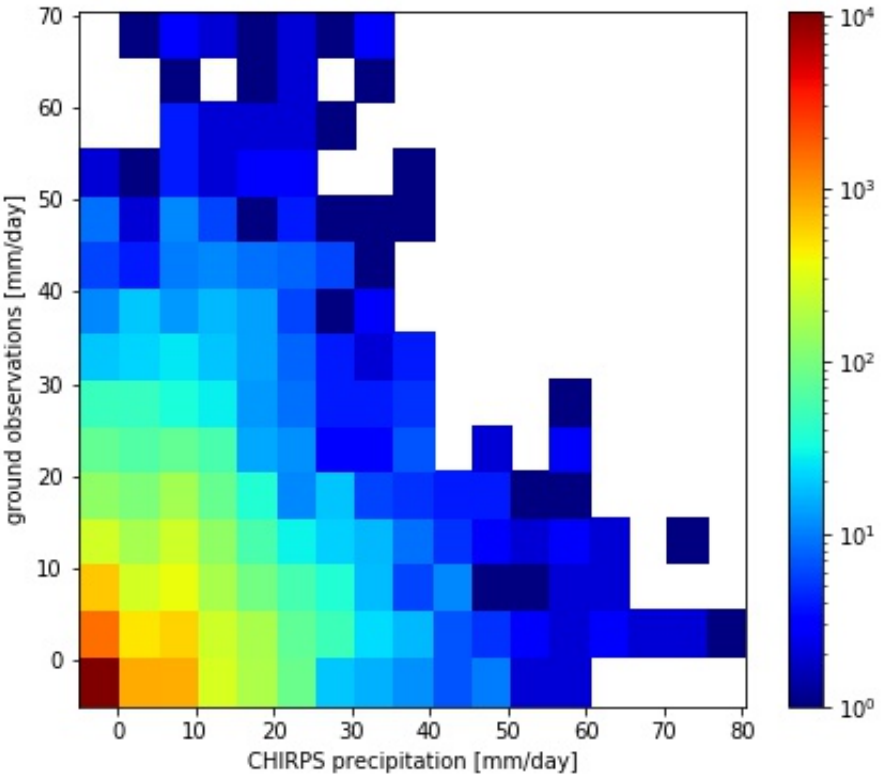


Figure 4.12: Absolute agreements of CHIRPS and rain meters observations.

Figure 4.13 shows one day with CHIRPS extreme and another day with rainmeters extreme measurements. The right figure shows higher rainfall values by CHIRPS whereas on the left figure, there seems to be a mismatch of the spatial locations between CHIRPS and farmers' gauge measurements. However, to make any useful conclusion a larger data set and longer record of extremes is needed.

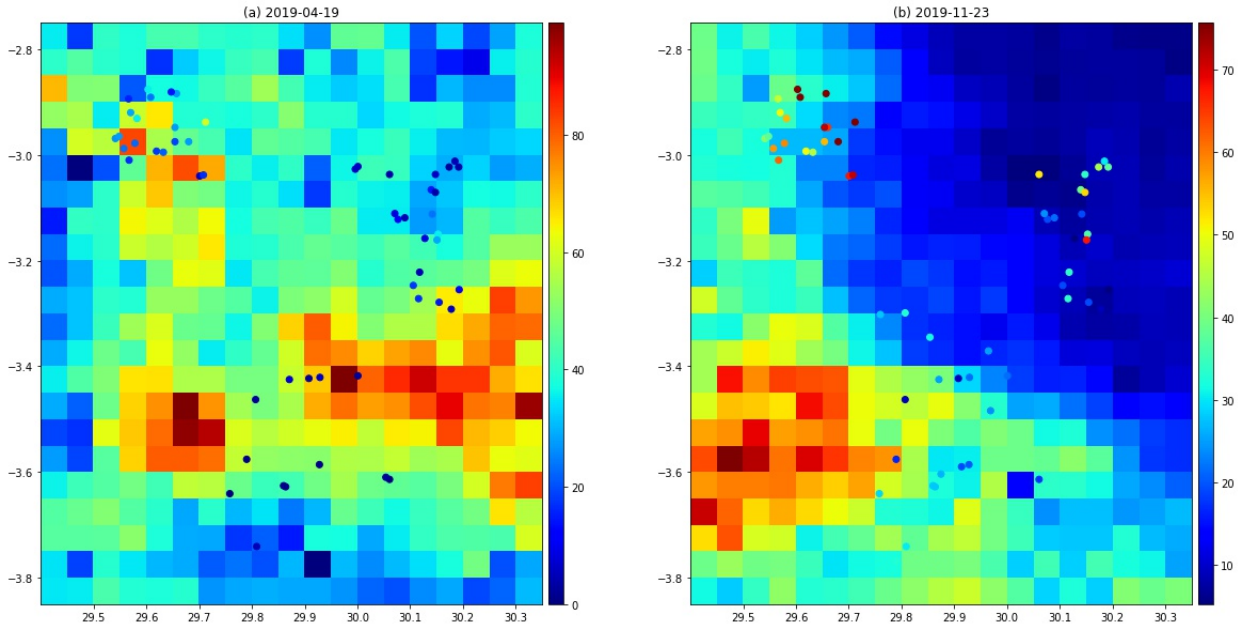


Figure 4.13: Right side figure shows the CHIRPS extremes whereas left figure shows the citizen gauge extreme measurements in mm/day.

The histograms in Figure 4.14 shows comparable frequency of zeros. Nevertheless, those zeros events do not always match as shown in Figure 4.12. Contrary to the absolute agreements between CHIRPS and citizen gauge measurements, the CHIRPS original distribution is taken into account in Figure 4.14. As a result, we observe a tendency of the CHIRPS observations to record higher rainfall amount than crowd-sourced gauges observations. It is noteworthy that in his paper about the development of the CHIRPS dataset, Funk et al.(2015) acknowledged that CHIRPS dataset may overestimate rainfall relative to ground observations due to the use of Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA)-3B42 for calibration instead of ground measurements . Therefore, the crowd-sourced observations present the same behavior as what would be expected from conventional ground observations in this context. However, there is still a need to compare these crowd-sourced measurements to trustworthy rainfall measurements to quantify the potential overestimate of the satellite data set CHIRPS with respect to ground observations.

4.3.2 Relationship between CHIRPS and crowd-sourced observations

To establish whether there is a correlation between the 2 data sets, only CHIRPS and crowd-sourced observations below 30 mm/day are considered because 98% of the data lie within this interval. CHIRPS data are grouped in classes of 1mm/day interval and the median and mean of the crowd-sourced measurements are calculated for each class. Figure 4.15 shows that the mean of the crowd-sourced observations is higher than the 50 percentile of the same observations for each CHIRPS class. This implies that the distribution of the crowd-sourced observations for each CHIRPS class is positively skewed. In the association with Figure 4.12, the dispersion of the zero-rain events over almost all CHIRPS classes help us understand the causes of this positive skewness. Hence, there is no agreement between the observed and the CHIRPS zero rain events .Therefore,

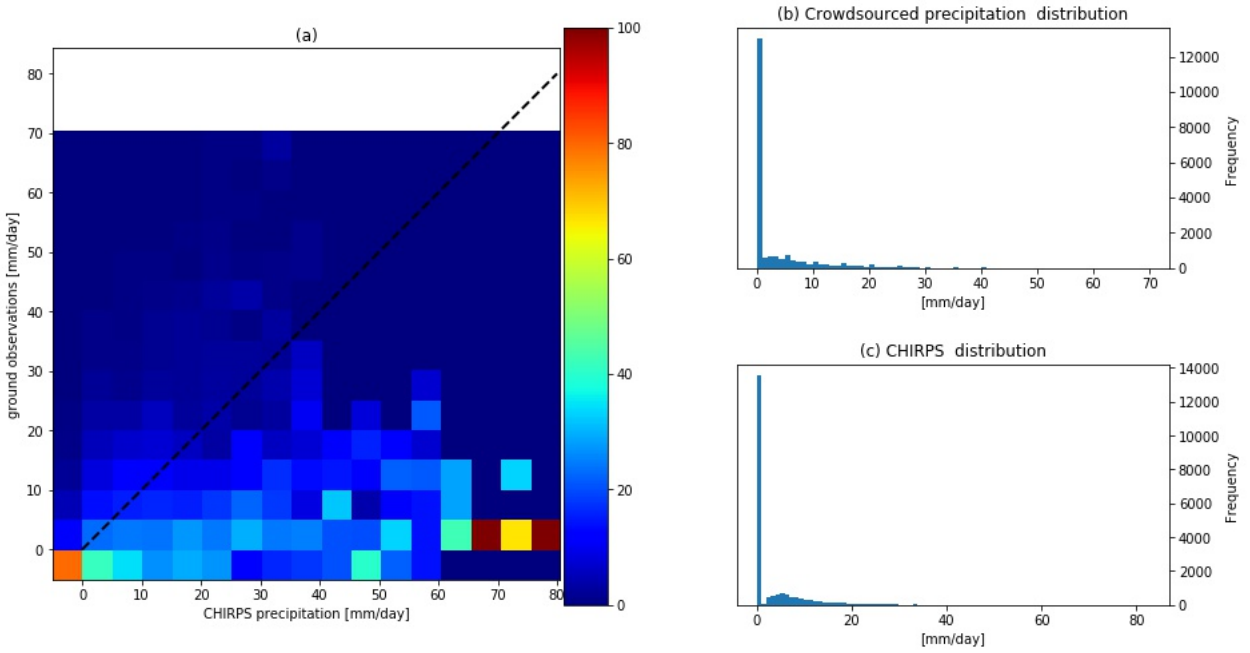


Figure 4.14: Right hand side: CHIRPS-conditioned agreement. Left hand side: Distribution of crowd-sourced rainfall measurements and CHIRPS value with 1mm/day interval.

despite the rather high correlation neither the mean nor the more robust median cannot correctly represent the data set as they are both affected by the high frequency of zeros in the crowd-sourced measurements and thus a clear relationship among the 2 data sets cannot be established.

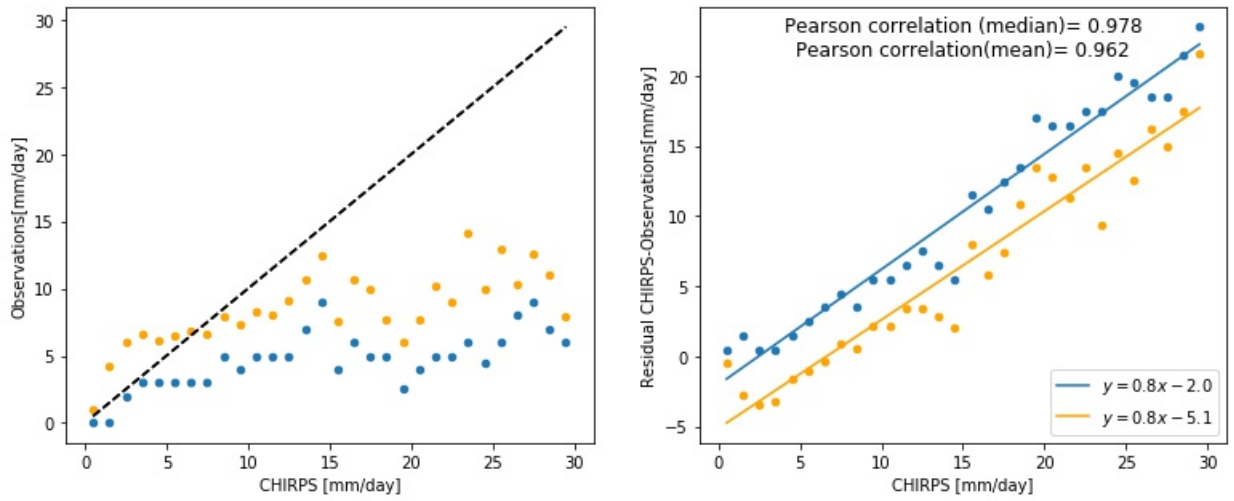


Figure 4.15: . The right figure plots directly the mean(orange) and the median (blue) crowd-sourced precipitation observations against the CHIRPS value. The left figure the difference of the 2 data sets in function of CHIRPS

4.3.3 CHIRPS sub-grid rainfall variability

The high density of the farmer's rain gauges makes it possible to have more than one station within a CHIRPS grid cell, hereafter referred as CHIRPS pixel. To assess the stations' reported rainfall variability within the CHIRPS pixels and how it compares to CHIRPS pixel value, 5 clusters of 3 stations are identified. Figure 4.16 depicts the dispersion of the stations' measurement relative to their mean and CHIRPS pixel value. The stations' measurements are closer both to CHIRPS mean value and to their own average from June to August than for the rest of the year. This period corresponds to the dry season which is characterised by many consecutive dry days observed. Thus, most of the stations measure no rainfall, and CHIRPS correctly report less precipitation during that period. 60% of the time, the stations show no deviation from their mean but when comparing to CHIRPS mean value this is the case for slightly less than 50%.

We also observe that outside the dry season, the station's observations are closer to their own mean than to CHIRPS pixel value. In fact, the stations' measurements deviate up to 30mm/day and 50 mm/day from respectively their mean and CHIRPS value. The 95 percentile standard deviation from their mean and CHIRPS mean value amounts respectively to 4.5mm/day and only 20 mm/day . These results corroborate that rainfall observations at one station is closer to another station' observations within a 5×5 km grid cell than to CHIRPS pixel value . Therefore, in general, using CHIRPS in this region at a daily scale and at a spatial scale smaller than its resolution would result in up to 20mm/day of standard deviation i.e. almost 5 times than the sub-grid stations' observations standard deviation.

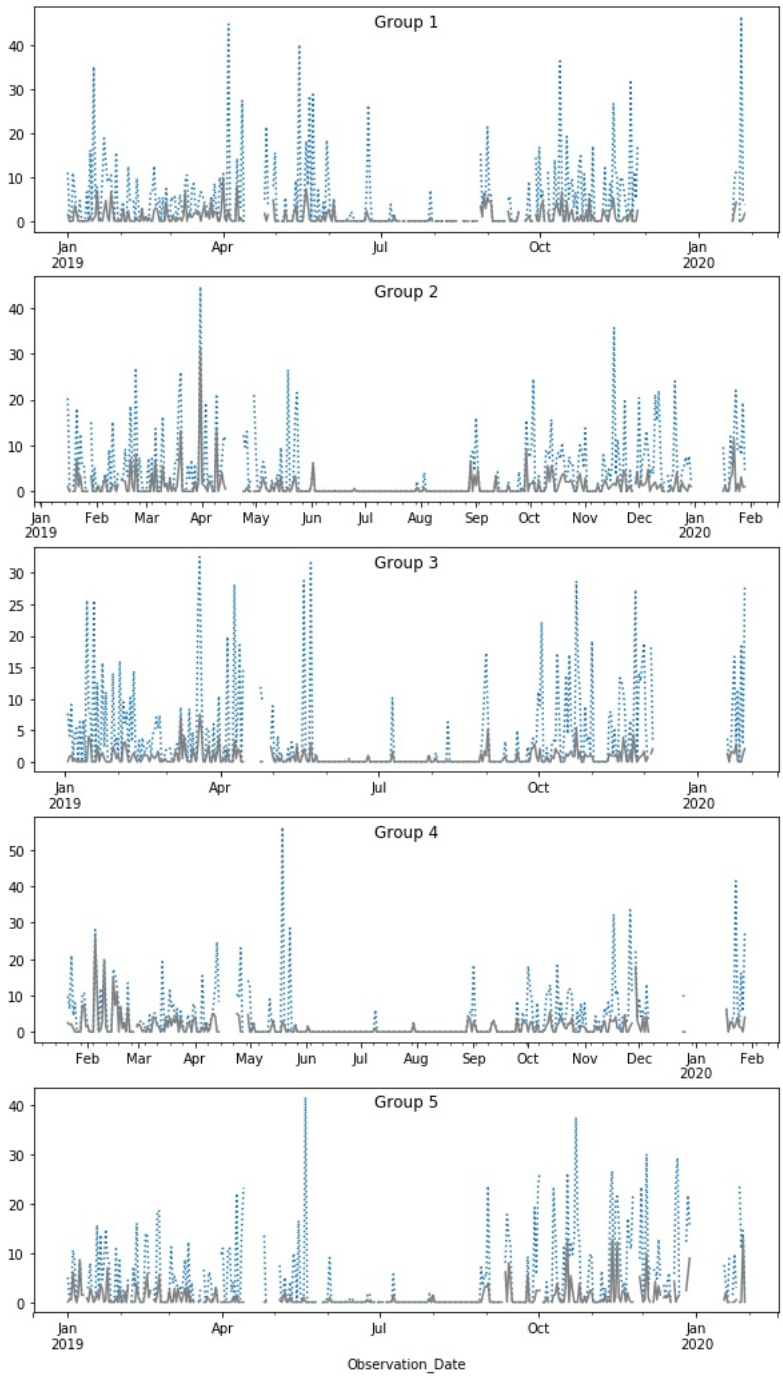


Figure 4.16: Standard deviation of the sub-grid stations relative their mean (black continuous line) and relative to CHIRPS grid value (blue dotted line

4.4 ECMWF forecast verification

This section discusses the verification of the medium-range precipitation forecast using the crowd-sourced rainfall measurements. According to the WMO (2014a), the goodness of a forecast can be assessed both by verification which is observations-based and by value which is user-based and evaluates the increase in economic benefits, risk reduction or other benefits. Besides, verification can also be used as a part of quality assurance like for the evaluation of the timeliness and the relevance of the forecast thus contributing to the potential forecast improvement. Among the three main aspects (i.e. consistency, quality and value) that describe a "good" forecast (Murphy, 1993), only the first two will be evaluated. The value of the forecast being linked to the economic benefits of the forecast users requires a deeper socio-economical study and therefore is beyond the scope of this research. The accuracy, discrimination and skill of the weather forecast are studied as attributes of a quality forecast whereas the farmers' observations are compared to the daily delivered forecast to assess the consistency of the latter.

4.4.1 Use of CHIRPS for ECMWF forecast verification

Due to sparse spatial coverage of local rain gauges and data accessibility challenges discussed in the Introduction, WeatherImpact decided to use CHIRPS for operational monitoring. Other satellites products such as IMERG, TAMSAT and CMORPH were left out due to the following reasons (Weather Impact BV, 2017):

- A long lag time
- Absence of historical archive
- High cost or very low spatial coverage

Before using the newly acquired crowd-sourced precipitation data set, CHIRPS is used to verify ECMWF precipitation forecast. This will help to assess the added value of using the crowd-sourced data instead of CHIRPS. Figure 4.17 displays the agreement between CHIRPS precipitation value and ECMWF rainfall forecast for 0 day lead time. We see that there is a strong correlation between the 2 values and that they cluster around the line of a perfect forecast. This is quite surprising knowing that CHIRPS is only a sattelite product and hence does not always reflect the reality at the ground as discussed in the previous section. It is therefore likely that such connection between the forecast and CHIRPS values is the result of common data sources such as satellite products. This section dissects the satellite sources of each data set to verify any overlap between their respective assimilated data .

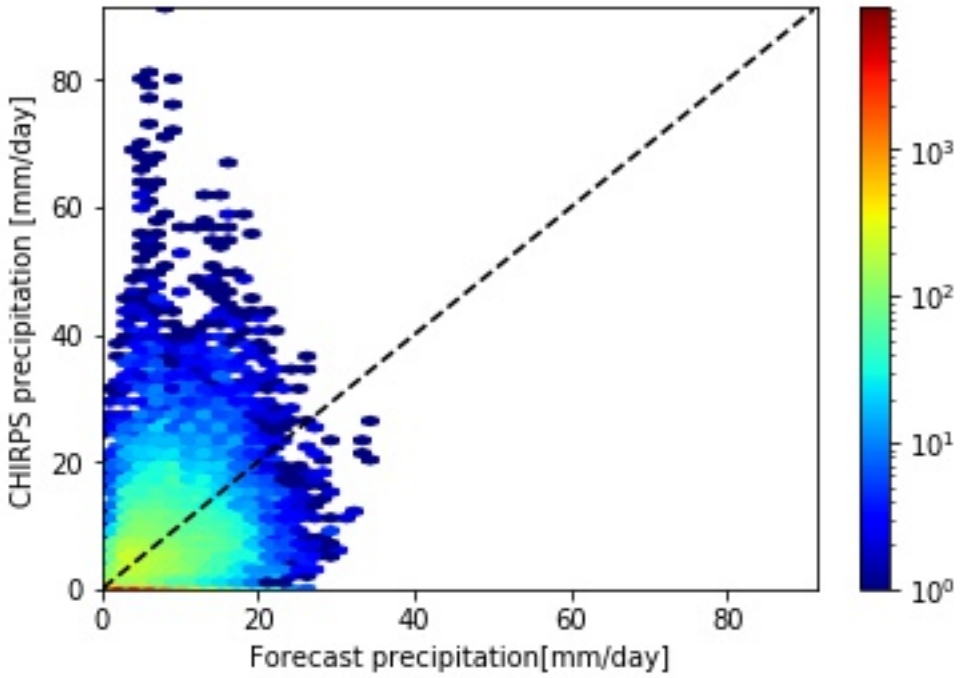


Figure 4.17: Agreement between CHIRPS and ECMWF 0 day lead time forecast. The dashed black line represents a perfect forecast with respect to CHIRPS value.

As shown in Figure 2.7 and 2.8 one of the satellite product used in the production of CHIRPS is Climate Prediction Center Thermal Infrared data(CPC TIR). This product merges satellite data from all geostationary satellites namely :Geostationary Operational Environmental Satellite system (GOES) operated by the National Oceanic and Atmospheric Administration (NOAA), Japan's Geostationary Meteorological Satellite (GMS) and Meteosat operated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT)) thus fulfilling a global coverage (60°N-60°S (Janowiak, Joyce, & Yarosh, 2001)).

The classification of the assimilated observations at ECMWF is shown in Figure 4.18. Infrared (IR) and microwave(MW) sensors dominate other assimilated data sources. IR sensors being mostly on geostationary satellites have a good temporal resolution, but the IR imagery is less directly linked to precipitation. On the other side, MW sensors thanks to their wavelength comparable to rain droplet and ice particle give a better estimate of rainfall. However, they are mounted on low earth orbiting satellites and thus have a much longer revisiting time(Hong et al., 2019).

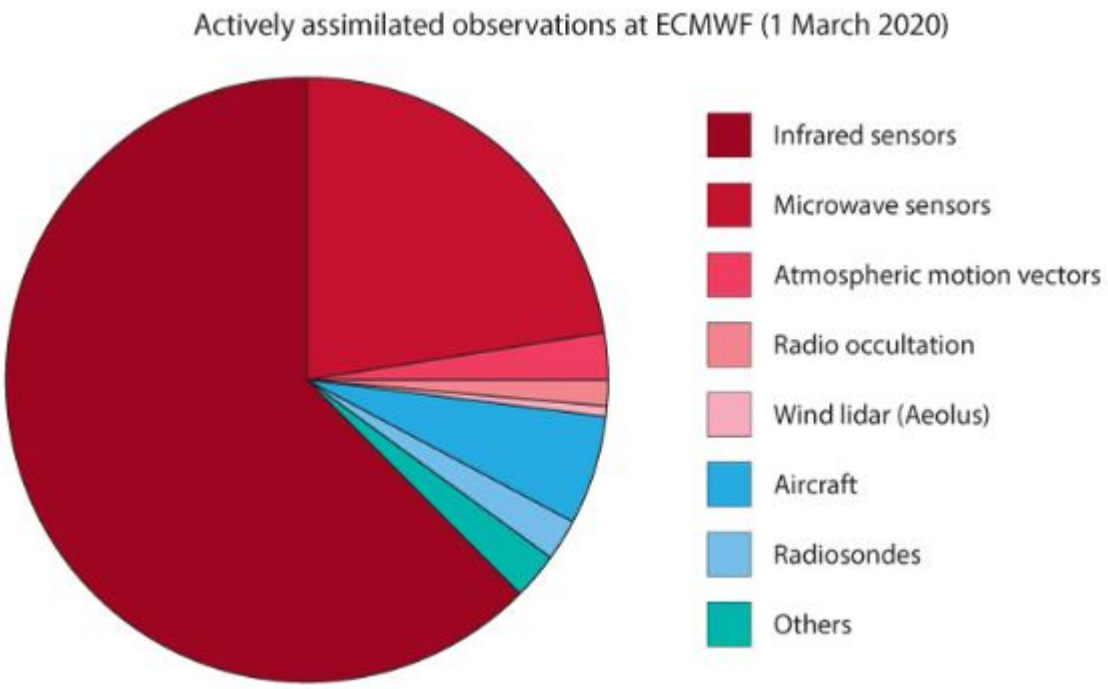


Figure 4.18: Assimilated observations at ECMWF [2]

Among the satellites monitored by ECMWF (Figure 4.19), we recognize GOES and Meteosat that help collect the earth's radiation information. ECMWF uses these satellite product to retrieve and assimilate in the model water vapour as well as wind information from sequences of imageries hence playing a key role in weather forecasting. [3][4].

[3]https://www.ecmwf.int/en/about/media-centre/news/2015/meteosat-weather-satellite-launched

[4]https://www.ecmwf.int/en/about/media-centre/news/2018/noaas-goes-s-launching-1-march-help-improve-weather-forecasts

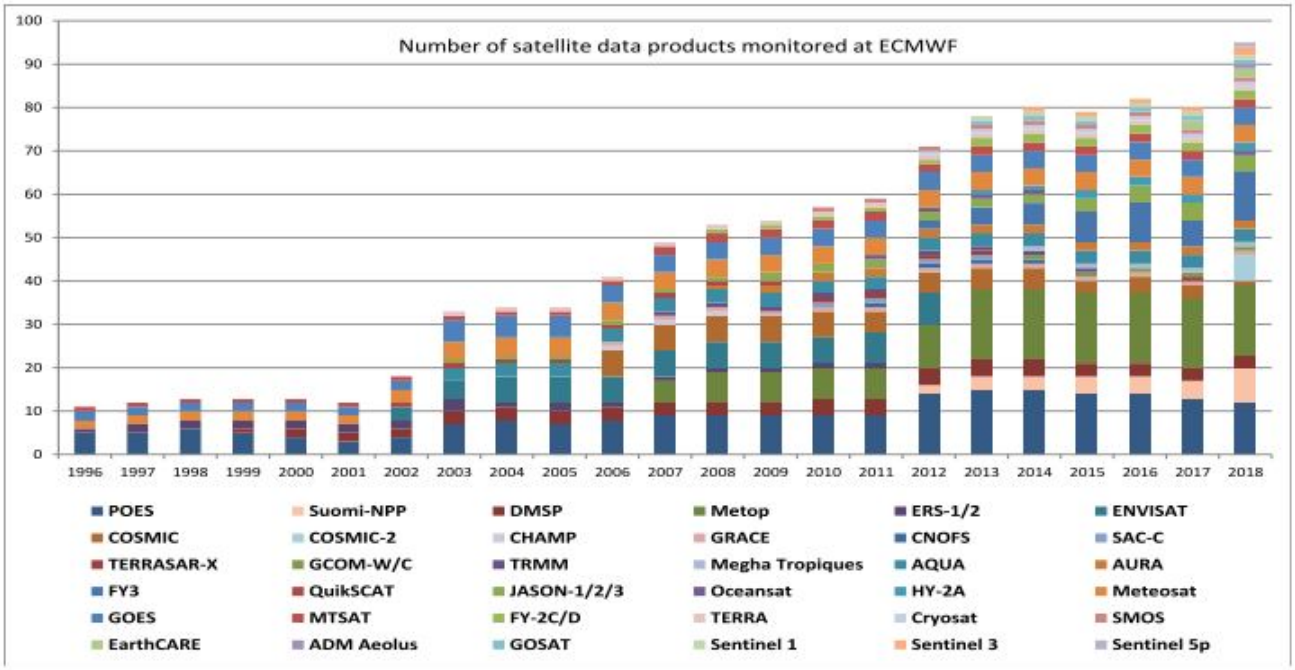


Figure 4.19: Some of the satellite data products monitored at ECMWF (Rabier, 2018)

As a result, the Meteosat and the GOES data sets are assimilated by ECMWF on one side and generates the TIR CCD data used in the CHIRPS dataset on the other side. This clear linkage between the forecast data and CHIRPS observations tends to explain the strong association between the 2 data sets as shown in Figure 4.17.

4.4.2 Forecast accuracy

The accuracy of the forecast is defined as the level of agreement of the forecast with the observations [5]. In this section crowd-sourced precipitation is averaged over the forecast grid-box area in order to be compared to the 0 day lead time precipitation forecast. Figure 3.5 shows that during dry and slightly rainy days, 50% of the forecast members overestimate the amount of precipitation in the grid box . In fact, the mean of the observed and forecasted rainfall amounts are 2.43 mm/day and 4.12 mm/day respectively. However, these values increase to 5.52 mm/day and 5.13 mm/day for the observations and forecast respectively when we omit the zeros in both data sets. The bias is therefore mostly caused by a different frequency of zeros between forecasts and observations, with zeros being almost 3 times more frequent in the observations ($\approx$ 60%) than in the forecasts ($\approx$ 20%).

On the other side, the number of days with heavy rainfall up to 30 mm/day (averaged over the grid box of 9 km $\times$ 9 km) is highly underestimated by at least 50% of the ensemble members (50percentile).

Whereas Figure 3.5 compares the intensity of the forecasted precipitation to the crowd-sourced

[5]https://www.cawcr.gov.au/projects/verification/

precipitation measurements, Figure 4.20 shows the temporal evolution of the precipitation forecast and the average of the crowd-sourced precipitation in 2019. In fact, farmers take the measurements to verify the forecast and at the same occasion get more familiar with the rainfall measurements so that they understand better the forecast. On the Agricoach app, farmers read the expected occurrence of rainfall, as well as the range of the forecasted rainfall amount which corresponds to the 30-90 percentile of the ensemble forecast. Hence in Figure 4.20, as long as the average of the crowd-sourced data falls within the forecasted rainfall range, the forecast can be considered as accurate. However, accuracy greatly varies over time. For example, in the month of May, 27 out of 31 mean observations lie below the forecast range. Around day 150 is a transition period between the wet and dry seasons. During this period, the daily average precipitation measurements exceeded 2 mm on only 3 days. However, the model predicted 9 days. Also, note that many of the largest differences between the predicted and observed rainfall amounts were recorded between day 75 and day 100. For the same period, the average precipitation measurements exceeded the maximum predicted rainfall 5 times (out of 16).

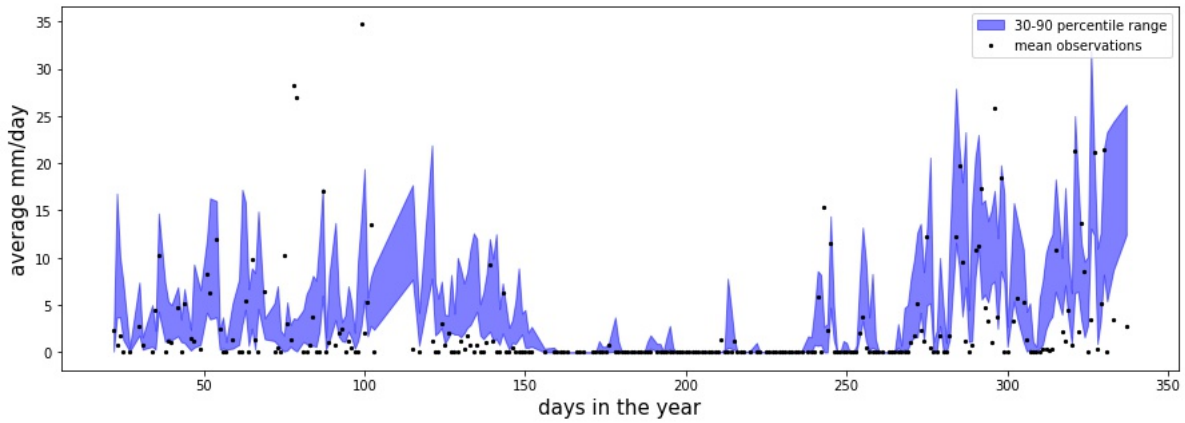


Figure 4.20: The average precipitation observations and the 0 day lead time forecast precipitation range for 2019

4.4.3 Forecast relevance for farmers

The farmers' use of the forecast depends on the farming activities. In Burundi, 3 agricultural seasons are considered: Season A (September-January), Season B (February-May), and Season C (June-August). Season C corresponds to the dry season and therefore concerns owners of marshes or farmers who have access to irrigation. The remaining 2 seasons correspond to a rain-fed agriculture and makes thus weather forecast very relevant. This weather forecast is made of grid box-average rainfall forecast whereas farmers record point rainfall measurements in their backyard. In the following, the effect of the rainfall variability within the grid box on the agreement between the 0 day lead time forecast and the crowd-sourced measurements is analysed.

Table 4.2 shows that during the rainy seasons A & B, respectively about half and a third of the farmer's measurements are above the minimum predicted rainfall as shown by the "min" column. In more than half of these cases, no rain was measured at the station. However, these results vary not only with seasons but also with the stations. For example, in season B for the same grid box the `kar-karu-gr2.umva` only exceeds the minimum predicted forecast 23.8% of the time while

		A		B		C	
		min	**max**	**min**	**max**	**min**	**max**
1	**cub-nta-gr1.umva**	50.5	90.5	31.1	85.8	83	96.6
2	**kar-karu-gr2.umva**	46.5	89.9	23.8	88.6	82.8	94.3
3	**mug-mugo-gr2.umva**	47.2	92.5	41.5	84.9	83.1	96.6
4	**mug-nta-gr3.umva**	45.4	90.7	32	85.4	84.1	95.5
5	**mug-rus-gr2.umva**	46.2	90.6	34	86.8	83.5	97.8
6	**sha-sha-gr1.umva**	40.2	95.1	31.7	83.2	83.1	95.5

Table 4.2: Percentage of station measurements above the 30percentile of the ensemble forecast(min) and below the the 90percentile of the ensemble forecast(max) for season A (01/09/2019-31/01/2020),season B (01/02/2019-31/05/2019) and season C (01/06/2019-31/08/2019)

`mug-mugo-gr2.umva` exceeds it 41.5% of the time. Hence, more than 75% of the time, less or no rain may fall at `kar-karu-gr2.umva` station when a positive minimum rainfall is predicted in the grid box area. However, this improves in Season C where at least 80% of all station data fall within the forecast range. In fact, if we compare these results to the ones in Figure 4.21, we can see that even during the rainy period, it is quite uncommon that all of the stations measure rain when rain is forecasted. Compared to the minimum predicted rainfall, the condition of the maximum predicted rainfall is more satisfied. More than 80% of each station's measurements lie below the 90percentile of the ensemble forecast. In general, it is more common for the stations in the grid box to agree during the dry season than during the wet season.

Given the spatial variability of rainfall within the grid box induce that the perfect forecast range correspond to the range of the station measurements. During the whole year, for 59 days this aspect of the forecast is satisfied. In contrast, a forecast is less useful to the farmers if none of the stations' observations fall within the forecast range. Whereas the observed rainfall is often lower than the forecast range, the observed rainfall is higher than the forecast range only 7 times in total: 1 in season A, 5 in season B and 1 in season C .

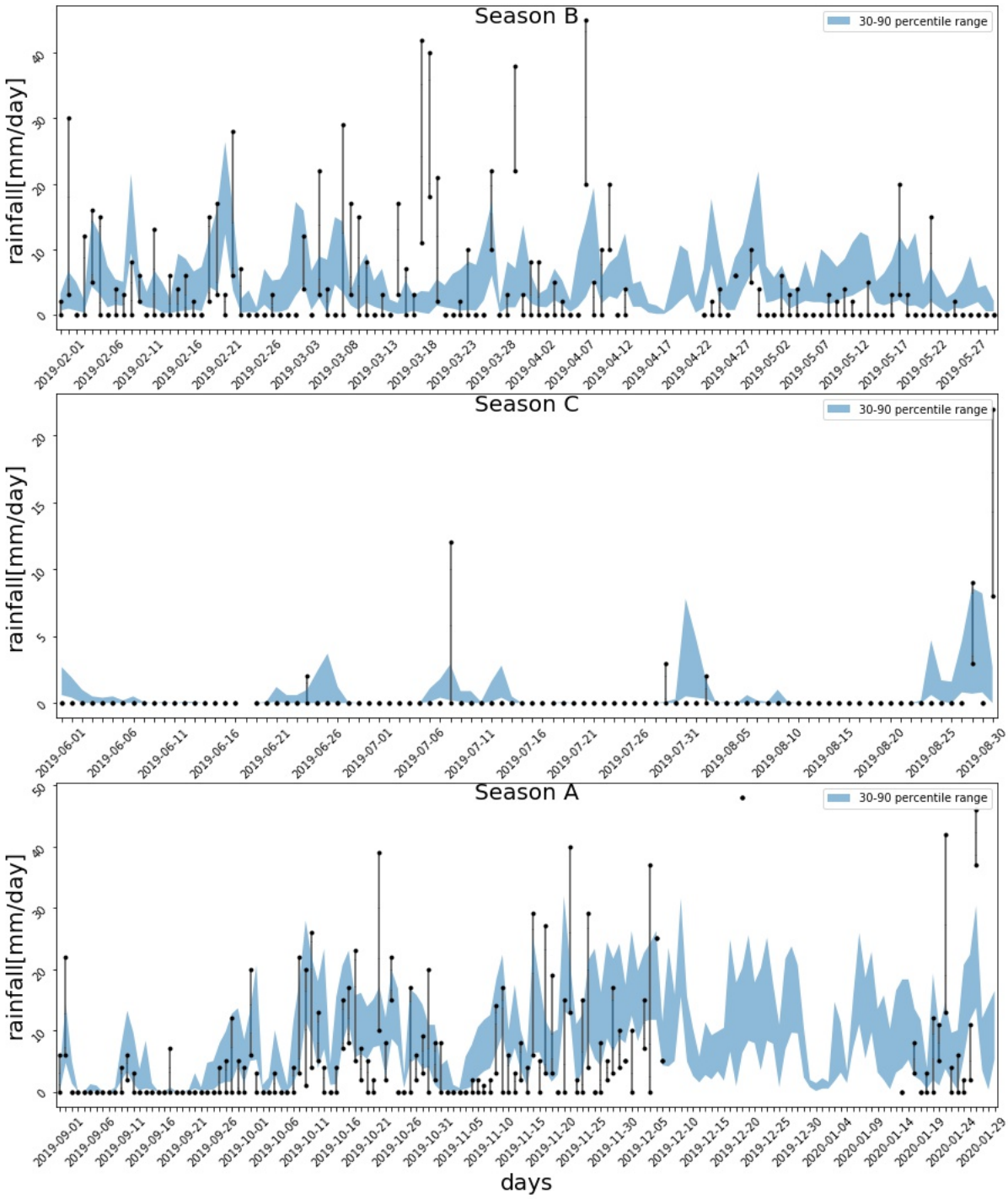


Figure 4.21: Grid box rainfall range (black bars) and the 0 day lead time forecast range (shaded blue area)

4.4.4 Forecast skill

The previous sections used only the crowd-sourced precipitation to assess the accuracy of the forecast. However, one of the most important information about a forecast is how good it performs compared to another alternative forecast. The skill of a forecast is therefore defined as "the relative accuracy of a set of forecasts, with respect to some set of standard reference forecasts" (Wilks, 2019). The reference forecasts used here are derived from the 1964-2008 Gitega climatology (see Section 3.3) . A skillful forecast is therefore the one whose accuracy is better than that of a forecast based on the climatology.

Relative Operating Characteristics (ROC) Curve

The Relative Operating Characteristics (ROC) curve (Figure 4.22) by plotting the hit rate (percentage of forecasted events that were correct) vs the false alarm rate (percentage of forecasted events that were incorrect). For higher probability threshold, the ECMWF forecast has a significant increase in the hit rate for the 0 and 1 day lead time, and thus an increase in the forecast skill. Therefore, when false alarms are very costly, for example at the start of the season,one should be cautious when forecasting rain with a probability below ≈ 0.8 .Also only forecasts at those 2 lead times are more skillful than the climatology-based forecast. .

The Area Under the Curve (AUC) is calculated in Table 4.3. When looking at the increase in AUC from one lead time to another, again only for lead times of 0 and 1 day we have a significantly improved skill (> 0.1 AUC increase). When Stanski et al. (1989) introduced the signal detection theory in forecast verification a general minimum threshold for useful forecast was suggested as 0.7. In our case, forecast for all lead times can be rounded to 0.7; however, the AUC for the climatology is 0.73 (Table 4.3).Since it exceeds 0.7, we will consider as useful a forecast with AUC greater than 0.73. For forecast with 6 to 8 days lead times, we observe a slight increase in the forecast skill with the lead time.Knowing that only forecast with 0 and 1 day lead times are more skillful than the climatology, this behaviour at the higher lead times is possibly explained by the forecast tendency toward a climatology-based forecast.

These results suggest that only forecast with 0 and 1 day lead time should be considered as skillful.

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**climatology**
Area under ROC	0.8	0.77	0.73	0.72	0.73	0.7	0.69	0.7	0.71	0.73

Table 4.3: Area under the ROC curves for different lead times and for the climatological forecast

Stable Equitable Error in Probability Space (SEEPS) score

The Stable Equitable Error in Probability Space is an equitable score less sensitive to precipitation spatial distribution as it includes information about the climatology of the region (Rodwell et al., 2010). It is an error-oriented score ranging from 0 (a perfect forecast) to 1 (unskillful forecast). These limits are valid for a sufficient long record of data (Rodwell et al., 2010). Figure 4.23 shows that (1-SEEPS) vary for the different stations. Along with score values beyond the above range, this

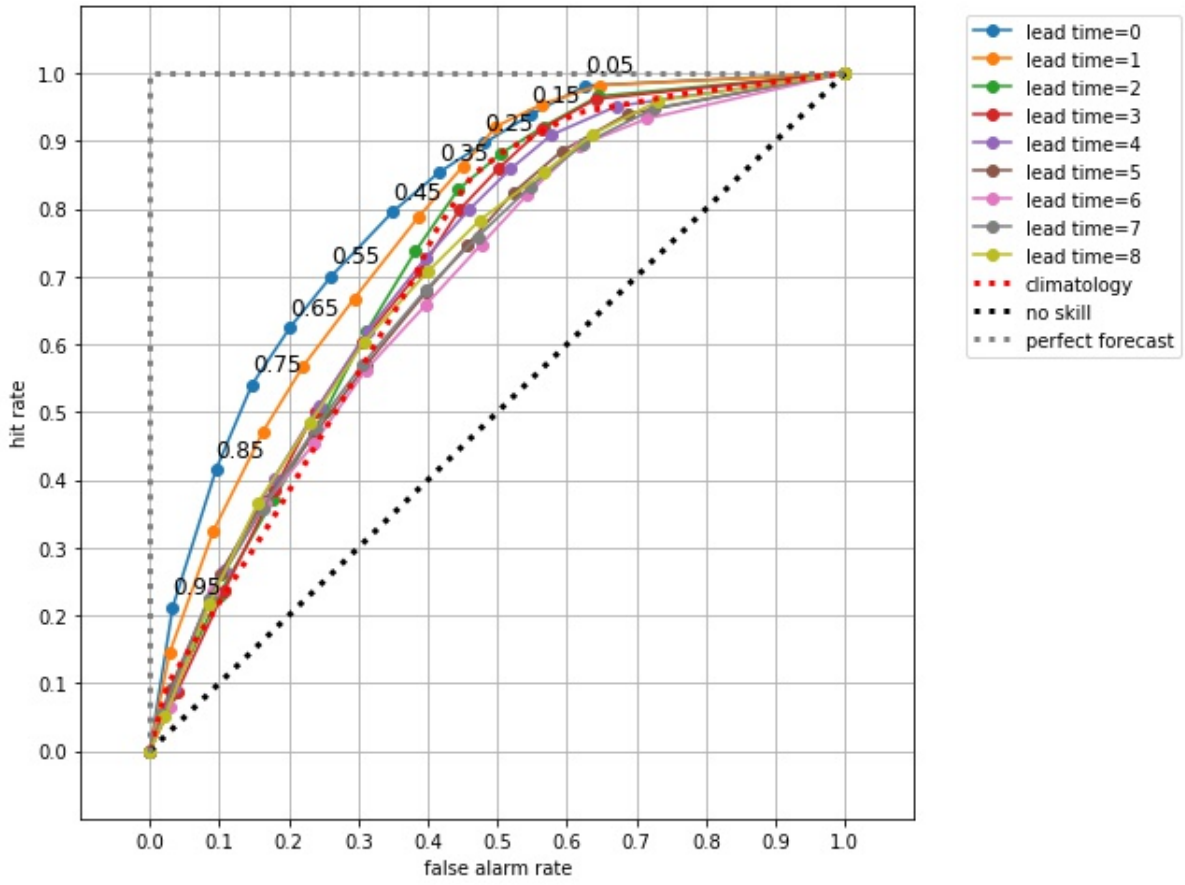


Figure 4.22: ROC curve. The big dots with labeled numbers represent the probability of having a rainy day

could be explained by the requirement of a long record of observations which can only be achieved by aggregating all the different stations. In his paper, Haiden et al.(2012) used 4500 global stations among which 1200 were in the tropics. The resulting 1-SEEPS score is shown in Figure 4.24. We see that the SEEPS score is much lower in the tropics than on global level. This can be explained by the difficulty in forecasting convective rainfall. Convection-driven precipitation also reduces the forecast skill in the Extra Tropics mostly during summer (Haiden et al., 2012) but affects the forecast in the Tropics during the whole year. The 1-SEEPS value at the Tropics are quite comparable to what we observe in Burundi using our crowd-sourced precipitation observations. Another difference between 1-SEEPS at the global vs at the tropical level is the shape of the decrease of the SEEPS score with the lead time. The lead time of the forecast is from 24hrs to 144hrs. At the tropics, we see a steeper decrease in the score for 24hrs to 48hrs lead times. Using the crowd-sourced precipitation measurements we also observe a sharp decrease in the score from 0 to 2 days lead time.

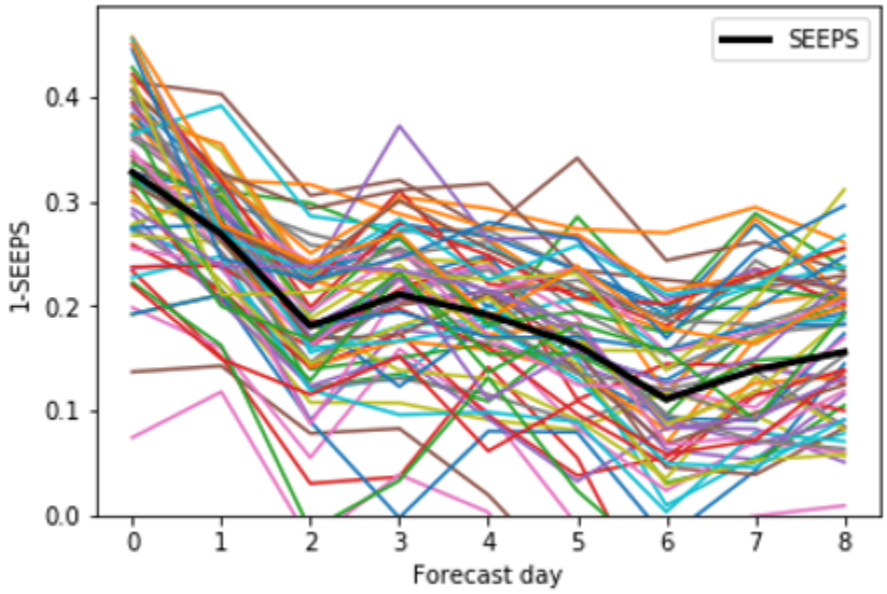
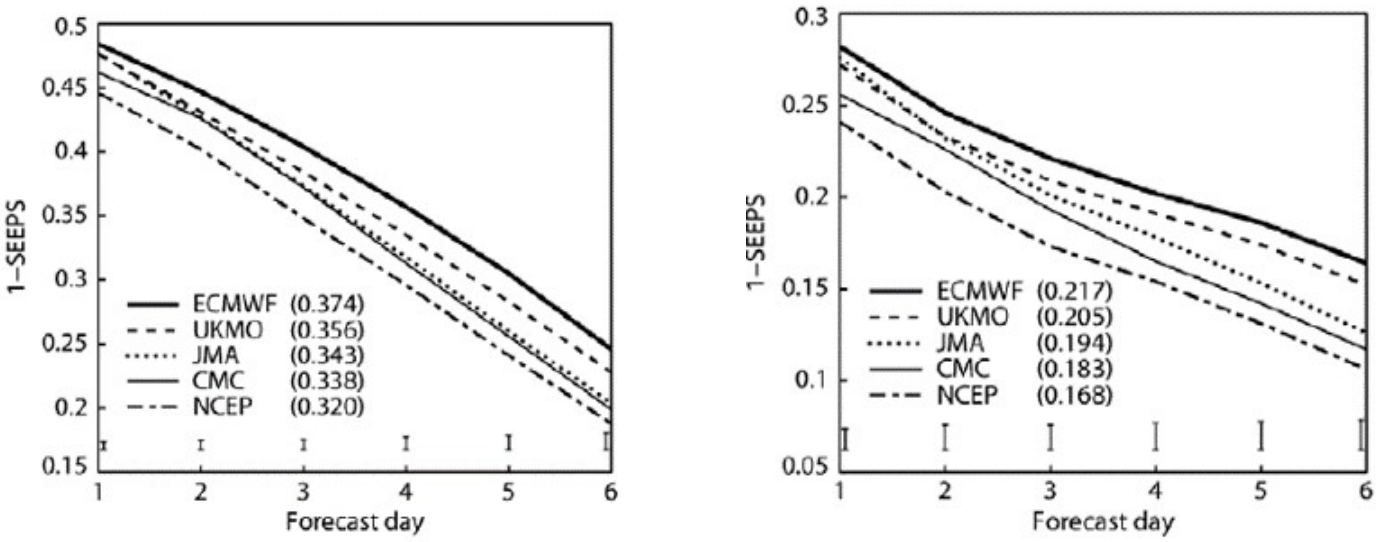


Figure 4.23: (1-SEEPS) score for different lead times (X axis) and different stations under consideration (different colours). The black line shows the average score for the entire crowd-sourced precipitation data set for all stations.



(a) 1-SEEPS score globally

(b) 1-SEEPS score at the Tropics

Figure 4.24: 1-SEEPS value globally (a) and in the tropics (b)

Both the ROC curve and SEEPS score are important metrics of the forecast skill. The ROC curve allows the comparison of 2 different forecast whereas the SEEPS score help to compare the skill for different climatological regions. Therefore, those two scores complete each other.

4.4.5 Discrimination between rain and no-rain events

The discrimination diagram (Figure 4.25 show that the no-rain events have a higher likelihood with a probability of 0.637 (sum of all orange bars) against 0.363 (sum of all blue bars) for rain events. A perfect forecast will have a 0% probability when there is no rain, and a 100% probability when rain is observed. At a lead time of 0 day, 39.2% of no rain events were forecasted with a probability within 0-10% whereas this is only 27.5% of rain events for a probability of 90-100%. This proves that it is more difficult to forecast a rainy day than a dry day with extreme probabilities. Sharpness being defined as the ability of the forecast to predict extreme probabilities, there is no 90% or 100% forecast based on the climatology; therefore, although the climatology-based forecast has some skill, it has no sharpness. The closer to 0% and 100% the probability of a forecast is, the more meaningful for the user. This is because a probability of around 50% distributes the probability equally between the rain/ no rain alternatives and this is of little help for decision making.

Table 4.4 shows a jump in the discrimination distance from day 1 to day 2 lead time when forecasting wet days. This reveals that from 2 days lead time onwards, it becomes more difficult to discriminate a rain event from no- rain events. Similarly to the forecast skill, there is a slight increase in the ability of the forecast to discriminate wet days. On the other side, the climatology performs worse when it comes to discriminating rain days from dry days. This could be explained by the dry days in the dry season which are easier to forecast than the wet days in rainy season.

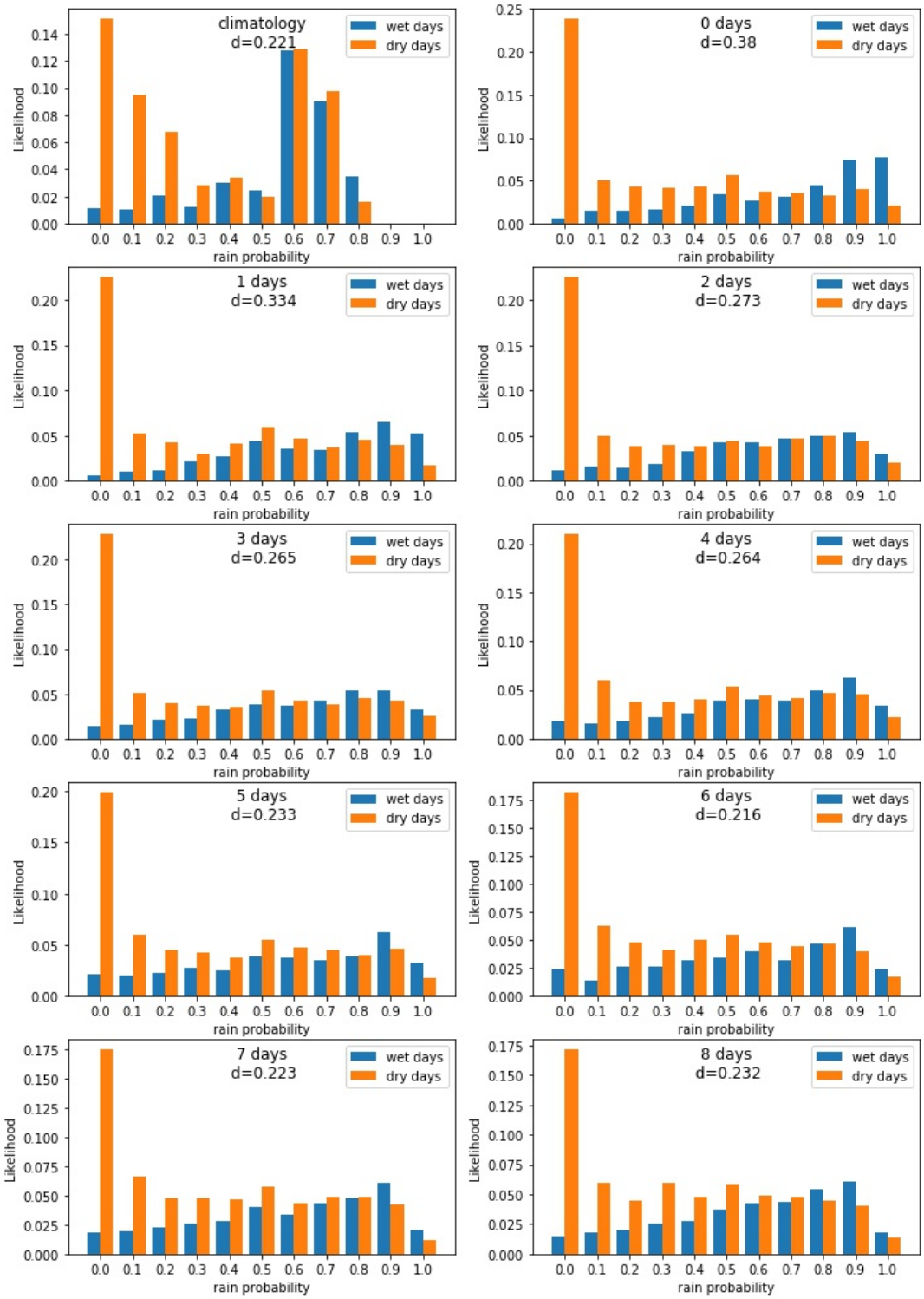


Figure 4.25: Discrimination diagram for the climatological forecast and different lead times. d is the distance between the mean probability for the rainy days(blue) and the mean probability for the dry days (orange). In the case of a perfect forecast, the distance d is equal to 1.

	climatology	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
wet days	0.45	0.3	0.33	0.39	0.39	0.39	0.42	0.43	0.42	0.41
dry days	0.33	0.32	0.33	0.34	0.34	0.35	0.35	0.36	0.36	0.36

Table 4.4: Mean distance from the perfect forecast mean. This is negatively oriented , 0 representing the distance for a perfect forecast

5

Recommendations

As discussed in the introduction, one of the challenges of citizen science is the quality of the data. In addition to automatic data filters, metadata plays a crucial role during the quality control by giving precious information about the location, collection and storage of the data. For example, in this study, one station had to be removed due to wrong geographical coordinates, and many others had partially inconsistent information about their location (e.g., commune, colline and locality). Also of concern is the suboptimal location and installation of some of the stations, which adds errors and uncertainties. As a result, some gauges probably act more like rain gutters rather than rain gauges.

The first recommendation is therefore to improve the quality of the crowd-sourced rainfall measurements by making sure that gauges are properly installed, meet WMO standards and that their geographical coordinates are correct. As simple as it may sound, collecting pictures for each station would contribute to the quality assurance. It would help identify badly located stations and identify problematic setups. It could also be used to assign weights during interpolation or during the quality control process.

Another big problem of crowd-sourced data is the presence of gaps in the measurement records. Our results show that the latter are often due to technical issues. In this context, the second recommendation is to set up an automated alarm system that sends a notification whenever the number of reporting gauges falls below 90% (i.e. less than 60 out of the 66 stations). This would allow to quickly detect problems such as the lack of data bundles, and would allow to solve connectivity issues within a few days. The goal should be to have a maximum of 1-2 missing days in a given month and to avoid long gaps of consecutive missing measurements. For installation, the calculation of monthly mean precipitations for a specific station require less than 11 missing values in a given month, and less than 5 missing values when those are for consecutive days (WMO, 2017).

Another problem identified during this study is that the forecasts provided to the farmers are made for the time interval from 00 UTC to 00 UTC whereas the crowd-sourced observations roughly run from 7 a.m to 7 a.m (local time). This makes it impossible to directly compare the daily forecasts and observations and limits the possibilities for validation. In fact, with a local time equivalent to UTC+2, an observation day is defined from 05 UTC to 05 UTC. Hence, any precipitation that would be forecasted between 00 UTC and 05 UTC will be ascribed to the previous observation day.

Knowing that the model output is produced every 6 hours, the third recommendation would be to shift the time intervals for the forecasts from 00UTC-00UTC to 06UTC-06UTC to enhance the overlap of the forecast period and the observation period from 19 hours to 23 hours a day.

Besides technical issues, another crucial aspect is to make sure farmers stay engaged for as long as possible. In his study about increasing public engagement with the help of citizen science, Martin (2017) shows that people engage better and longer when the specific citizen science project matches their own interest. He stresses that this is especially important for groups of people with lower education level, who expect to receive knowledge from others but do not esteem themselves knowledgeable enough to make any useful contributions. Therefore, simply asking farmers to make rainfall measurements is not enough to achieve a high engagement. But measuring precipitation for forecast verification purposes can make a huge difference as most farmers feel that their measurements serve a clear purpose and will report the measurement daily and for free. One strength of the GAP4All project is that many partners operate locally from Burundi rather than from abroad which helps partly bridge the cultural and educational gap. In fact, the AUXFIN pyramidal network makes it possible to establish a direct communication with the volunteering farmers almost on daily basis. Following a discussion with a particular farmer who was proudly wearing an old hat from UMVA (the money and "rain transfer" platform), the distribution of appreciation gifts or certificates does give the farmers a feeling of belonging to the large family of crowd-sourcers. This would lead to a potential long-term engagement of the farmers beyond the hype.

Finally, the quality of crowd-sourced data should eventually be achieved by preventing errors (quality assurance) rather than fixing them (quality control). For this reason, the current use of tablets for reporting the measurements can also inspire their potential use for reading the measurements through automatic image analysis in the future. By automating both the reading and reporting of the measurements, sources of errors will then be reduced. Similarly, the replacement of the current gauges by 2 buckets rain gauges as the one used in CocaRaHS footnotewww.cocorahs.org for example. This will prevent the current gauge overflow and give a better understanding of the extremes range and spatial variability.

6

Conclusion

This thesis was undertaken to assess the quality and to evaluate the usefulness of the crowd-sourced precipitation measurements in Burundi. For this purpose, different research questions have been defined and below are the conclusions of the study:

- *What are common errors in daily crowd-sourced precipitation observations in Burundi and what can be done to improve the quality of the data in the future?*

Field visits and discussion with farmers revealed that the quality of the crowd-sourced rainfall data was hindered by : wrong geographical coordinates, interrupted internet connectivity, insufficient rain gauge size, informal change of location and suboptimal station exposure. Careful and supervised installation was therefore highly recommended to meet the WMO standards wherever possible. In order to detect and correct errors in the measurements, there is a need to reduce the missing gaps by reporting them via an alarm system whenever they exceed a 10% threshold. To keep a satisfactory long-term engagement , the distribution of appreciation gifts or certificates will show the farmer that his contribution is very valuable and therefore will motivate him to be more careful and consistent in his measurements.

- *How can we filter out suspicious or erroneous crowd-sourced daily precipitation observations? And how reliable/good are these filters?*

As part of an automated quality control mechanism, suspicious faulty zeros were filtered out based on the median value of the stations whereas high precipitation amount were filtered out based on the minimum difference and ratio value between the reference station and neighbour stations within 13.5km radius. Thresholds were defined as 95percentile limit of these parameter values for the observation period . 4.3% and 1% of respectively no-rain and rain events were flagged summing up to a total of 2.9% of the whole data set. The lower number of flagged rain events is explained by the fact that when comparing one station to another, higher differences in rainfall corresponded mostly to extreme rainfall events. Therefore, only about 15% of rainfall above the minimum difference threshold also exceeded the minimum ratio threshold. Thus, this minimum ratio threshold prevents the common observed extreme rainfall events for being flagged as unrealistic values. A stronger coupling than that of the extremes is also observed for the zeros as 95% correspond to a median of 2mm/day or less of the neighbour stations . The main weakness of both these filters is their sensitivity to the number of available neighbour stations . For optimal

performance, the thresholds should probably be updated to local network characteristics or when more rain gauges will be deployed and longer data records will be available. This would increase the performance of the filters, especially for stations having only 2-3 stations at the moment.

- *How well do crowd-sourced precipitation observations in Burundi compare to lower resolution precipitation data sets for the region such as CHIRPS?*

The results show that although the range and the average over the whole period are almost equal for both data sets , CHIRPS reports higher values than the crowd-sourced observations at daily timescale. Also, the zeros and extremes do not necessarily match temporally nor spatially and for consequently, no clear relationship between the 2 data sets could be derived at the daily time scale. This is in agreement with previous research on CHIRPS that suggests that comparisons with in-situ observations may only be suitable for larger timescales of about 1 month. Furthermore, the sub-grid rainfall variability of the stations was less than the variability between CHIRPS and stations values. In comparison with the sub-grid average of stations rainfall measurements, CHIRPS value has a deviation of up to 20 mm/day whereas a rain gauge has a deviation of only $\approx$ 5 mm/day. Therefore, although the crowd-sourced measurements add little information to rainfall data for monthly and larger timescales, more sub-grid rain gauges will help assess the significant rainfall variability at resolution lower than that of CHIRPS. This finding is thus of great encouragement for the present expansion of the rain gauge network.

- *How well do crowd-sourced precipitation measurements and CHIRPS agree with the ECMWF weather forecasts over this region?*

The comparison of the crowd-sourced measurements with ECMWF forecast show a zero-dominance for the crowd-sourced data 3 times higher than that of the forecast. Regarding CHIRPS data, the results show a stronger correlation between CHIRPS and the 0 day ECMWF forecast than between the ECMWF forecast and the crowd-sourced data. Knowing that the 0 day lead time forecast is obtained using the model assimilated data, this correlation is likely due to the fact that both the CHIRPS and the ECMWF forecast model make use of the same satellite radiation data. Using CHIRPS for the purpose of forecast verification may therefore be misleading due to the overestimated forecast accuracy.

- *How skillful is the ECMWF forecast compared to the climatology forecast?*

The crowd-sourced precipitation observations were primarily reported in order to verify the ECMWF forecast delivered to farmers in Burundi by WeatherImpact. The ROC curve and AUC show that only the 0 and 1 day lead time forecasts were more skillful than the climatology-based forecast. Knowing that the AgriCoach app is shared by different farmers and that there were well known internet connectivity issues, these findings mean that every farmer should check the weather forecast either the day itself or the following day but not later. The forecast for the remaining days were almost as skillful as the climatology but with a higher ability to discriminate between no rain and rain events. The SEEPS score corresponded to what has been previously observed in the tropics showing that the forecast accuracy is poorer in the tropics than in the rest of the world.

However, to calculate this score for individual stations, a longer period of observations is needed.

Because of the issue of the sparsity, availability and accessibility of conventional rainfall measurements, it was not possible to assess the "true" ground rainfall observations. A natural progression of the work is to directly compare the crowd-sourced to the conventional ground rainfall measurements . The trending installation of automated stations all over the country by IGEBU will probably help overcome data collection in remote areas, decreasing the cost and increasing both their quantity and quality.

Appendix

.1 Farmer's instructions for crowd-sourcing rainfall

.1.1 Visual instructions

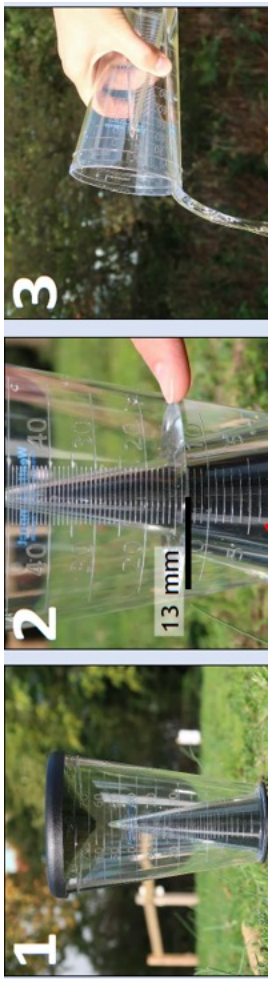


Figure 1: This is a visual used to demonstrate the 3 important steps:installing, measuring and cleaning the rain gauge

.1.2 Written instructions

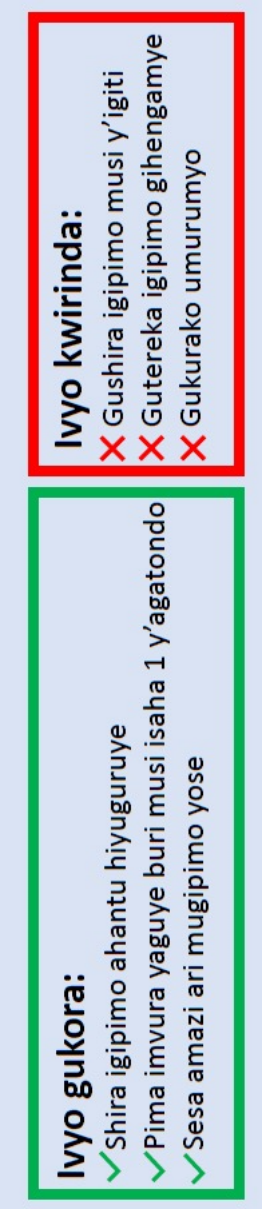


Figure 2: Written instructions in Kirundi. The green points show the dos whereas the red points, the dont's. In summary, farmers are invited to install the stations away from the trees, not to incline the gauge and to make sure to pour all water after taking the measurement.

.2 Questionnaire: Field visit in Burundi

.2.1 Questions to WeatherImpact

1. What kind of decisions do you want that require these meteorological data.
    - What type of information do you need and at what scale do you need them?
    - Do these decisions require data from the past (observation) or potential future scenarios(forecast)?

.2.2 Questions to Auxfin

Auxfin is the institution in charge of the rain meter distribution along other financial services to the group of farmers.

1. Management: Who is responsible of what for this rain gauges network?
2. Metadata collection: What happens when stations information or location change? How do you notify that?
3. Network problem: How is the known network problem usually handled when reporting the rainfall measurements?
4. Coordinates errors: How have were the geographic coordinates retrieved and checked?
5. Reporting errors: What happens when an error is noticed? Can it be deleted/modified? If yes, how do you keep track of modifications?
6. Reporting observations: Do you record the measurement time of these precipitation observations?
7. Project outcome:
    - What improvements have you seen based on these precipitation observations?
    - What are your suggestions for future improvement?

.2.3 Questions to farmers

1. Farmers' engagement (Essential for understanding and avoiding missing data for example):
    - Have you learned anything since you have been measuring the rainfall?
    - Do you know what this information is useful for?
    - What additional information would you like to receive?
2. Reporting challenges
    - What time exactly do you take the measurements? Do you record every time the measurement time?
    - How do you report no-rain days? precipitation above 70mm?

- What has been most difficult since you have started measuring precipitation?
- What kind of circumstances have made you stop reporting the precipitation observations in the past?

References

Davids, J. C., Devkota, N., Pandey, A., Prajapati, R., Ertis, B. A., Rutten, M. M., ... van de Giesen, N. (2019, 3). Soda Bottle Science—Citizen Science Monsoon Precipitation Monitoring in Nepal. *Frontiers in Earth Science*, *7*, 46. Retrieved from `https://www.frontiersin.org/article/10.3389/feart.2019.00046/full` doi: 10.3389/feart.2019.00046

DeVos, L. W., Leijnse, H., Overeem, A., & Uijlenhoet, R. (2019). Quality Control for Crowdsourced Personal Weather Stations to Enable Operational Rainfall Monitoring. *Geophysical Research Letters*, *46*(15), 8820–8829. doi: 10.1029/2019GL083731

Dinku, T. (2019). Challenges with availability and quality of climate data in Africa. Retrieved from `https://doi.org/10.1016/B978-0-12-815998-9.00007-5` doi: 10.1016/B978-0-12-815998-9.00007-5

Dinku, T., Thomson, M. C., Cousin, R., del Corral, J., Ceccato, P., Hansen, J., & Connor, S. J. (2018, 10). Enhancing National Climate Services (ENACTS) for development in Africa. *Climate and Development*, *10*(7), 664–672. Retrieved from `https://www.tandfonline.com/doi/full/10.1080/17565529.2017.1405784` doi: 10.1080/17565529.2017.1405784

ECMWF. (n.d.). *2.4 Atmospheric Model Data Sources - Forecast User Guide.* Retrieved from `https://confluence.ecmwf.int/display/FUG/2.4+Atmospheric+Model+Data+Sources`

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... Michaelsen, J. (2015, 12). The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes. *Scientific Data*, *2*(1), 1–21. doi: 10.1038/sdata.2015.66

Goodrich, D. C., Faurès, J.-M., Woolhiser, D. A., Lane, L. J., & Sorooshian, S. (1995). *Measurement and analysis of small-scale convective storm rainfall variability* (Vol. 173; Tech. Rep.).

Haiden, T., Rodwell, M. J., Richardson, D. S., Okagaki, A., Robinson, T., & Hewson, T. (2012). *Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score* (Tech. Rep.). Retrieved from `http://www.ecmwf.int/publications/`

Hong, Y., Tang, G., Ma, Y., Huang, Q., Han, Z., Zeng, Z., ... Guo, X. (2019). Remote Sensing Precipitation: Sensors, Retrievals, Validations, and Applications. In *Observation and measurement of ecohydrological processes* (p. 2). Retrieved from `https://doi.org/10.1007/978-3-662-47871-4_4-2` doi: 10.1007/978-3-662-47871-4{_}4-2

Janowiak, J. E., Joyce, R. J., & Yarosh, Y. (2001, 2). *A Real-Time Global Half-Hourly Pixel-Resolution Infrared Dataset and Its Applications* (Vol. 82; Tech. Rep. No. 2). Retrieved from `http://journals.ametsoc.org/bams/article-pdf/82/2/205/3867309/1520-0477` doi: 10.1175/1520-0477(2001)082⟨0205:ARTGHH⟩2.3.CO;2

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017, 1). So, How Much of the Earth's Surface is Covered by Rain Gauges? *Bulletin of the American Meteorological Society*, *98*(1), 69–78. Retrieved from `http://journals`

.ametsoc.org/doi/10.1175/BAMS-D-14-00283.1 doi: 10.1175/BAMS-D-14-00283.1

Martin, V. Y. (2017). Citizen Science as a Means for Increasing Public Engagement in Science: Presumption or Possibility? *Science Communication*, *39*(2), 142–168. Retrieved from https://doi.org/10.1177/1075547017696165 doi: 10.1177/1075547017696165

Muller, C., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., ... Leigh, R. (2015, 9). Crowdsourcing for climate and atmospheric sciences: current status and future potential. *International Journal of Climatology*, *35*(11), 3185–3203. Retrieved from http://doi.wiley.com/10.1002/joc.4210 doi: 10.1002/joc.4210

Murphy, A. H. (1993, 6). *What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting* (Vol. 8; Tech. Rep. No. 2). Retrieved from http://journals.ametsoc.org/waf/article-pdf/8/2/281/4650824/1520-0434 doi: 10.1175/1520-0434(1993)008⟨0281:WIAGFA⟩2.0.CO;2

Overpeck, J. T., Meehl, G. A., Bony, S., & Easterling, D. R. (2011). Climate Data Challenges in the 21st Century. Retrieved from www.ngdc.noaa.gov/wdc/usa/paleo.html. doi: 10.1126/science.1197869

Persson, A., & Grazzini, F. (2007). User Guide to ECMWF forecast products. *Ecmwf*(March), 153.

Rabier, F. (2018, 1). *Satellite Data Assimilation at ECMWF.* Austin: American Meteorological Society.

Rivera, J. A., Hinrichs, S., & Marianetti, G. (2019). Using CHIRPS Dataset to Assess Wet and Dry Conditions along the Semiarid Central-Western Argentina. *Advances in Meteorology*, *2019*. doi: 10.1155/2019/8413964

Rodda, J. C., & Dixon, H. (2012, 5). Rainfall measurement revisited. *Weather*, *67*(5), 131–136. Retrieved from http://doi.wiley.com/10.1002/wea.875 doi: 10.1002/wea.875

Rodwell, M. J., Richardson, D. S., & Hewson, T. D. (2010). *A New Equitable Score Suitable for Verifying Precipitation in NWP A New Equitable Score Suitable for Verifying Precipitation in NWP* (Tech. Rep.). Retrieved from http://www.ecmwf.int/publications/

Stanski, H. R., Wilson, L. J., & Burrows, W. R. (1989). COMMON VERIFICATION METHODS. In *Survey of common verification methods in meteorology.*

Stellingwerf, S. (2019). *Agricultural applications using ECMWF ensemble forecasts in Africa* (Tech. Rep.).

Tweddle, J. C., Robinson, L. D., Pocock, M., & Roy, H. E. (2012). *Guide to citizen science: developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK.* Natural History Museum and NERC Centre for Ecology & Hydrology for UK-EOF. Retrieved from www.ukeof.org.uk

van de Giesen, N., Hut, R., & Selker, J. (2014, 7). The Trans-African Hydro-Meteorological Observatory (TAHMO). *Wiley Interdisciplinary Reviews: Water*, *1*(4), 341–348. Retrieved from http://doi.wiley.com/10.1002/wat2.1034 doi: 10.1002/wat2.1034

Wani, S. P., Rockström, J., & Oweis, T. (2009). *Rainfed Agriculture: Unlocking the Potential.* Retrieved from www.cabi.org

Weather Impact BV. (2017). *Weather Monitoring in East Africa Examples from CropMon and CommonSense G4AW* (Tech. Rep.). Retrieved from www.g4aw.spaceoffice.nl

Wilks, D. S. (2019). Forecast Verification. In *Statistical methods in the atmospheric sciences* (pp. 369–483). doi: 10.1016/b978-0-12-815823-4.00009-2

WMO. (2014a). *Forecast Verification for the African Severe Weather Forecasting Demonstration*

Projects (No. 1132).
WMO. (2014b). *WMO Guide to meteorological instruments and methods of observation* (Tech. Rep.). World Meteorological Organisation.
WMO. (2017). *WMO Guidelines on the Calculation of Climate Normals* (Tech. Rep.).