Delft Measures Rain

A quality assessment of precipitation measurements from personal weather stations

Marchien Boonstra

MSc thesis Water Management

Graduation committee:

Dr. ir. Arjan Droste (Chair)	Water Resources
Dr. ir. Marie-Claire ten Veldhuis	Water Resources
Dr. Marc Schleiss	Geoscience and Remote Sensing

February 2024 Delft University of Technology, Faculty of Civil Engineering and Geosciences

Acknowledgements

I would like to express my gratitude to Arjan Droste, chair of the graduation committee for his supervision and infectious enthusiasm on the subject. I also thank the graduation committee members Marie-Claire ten Veldhuis and Marc Schleiss, for their expert input and feedback during the thesis research.

From the Delft Meet project I want to thank Marit Bogert and Sandra de Vries for their help with contacting participants, setting up the survey, supplying data and information from the project, and keeping the project running in general. Also thanks to Ronald Vendelmans for helping me try to understand the WOW API.

I could not have set up the experiment without Marijn Leeuwenberg, Jean-Paul de Garde, Floor Pino and Rob Mackenzie at The Green Village. Thank you for helping me with the manual and administrative work of installing the weather stations, obtaining reference data and your interest and support for my thesis project.

I would like to thank Lotte de Vos and Aart Overeem at KNMI for helping me understand the QC methods, finding reference data and updating me on current and relevant research on the topic of personal weather stations.

For help with the report writing, I want to thank Maurits Ertsen and Juliette Weitenberg for their writing classes. I would like to thank Diana Arias, Marijtje van Duijn and Marinda Vink for their feedback on the text of the draft thesis.

Thanks to everyone at the "Afstudeerhok" for your support and especially to Carolina Nanni for supplying the materials for my experiment.

Last but not least, I want to thank all the volunteers from the Delft Meet project for their time and energy, without whom there would be no DMR data.

Abstract

Personal weather station (PWS) networks have the potential to supply precipitation data at high spatial and temporal resolution for urban hydrological modeling. Past research has shown promising results on the quality of PWS data, for example from Netatmo gauges, but studies on other PWS brands are limited. This thesis assesses the quality of precipitation measurements from the Alecto WS-5500 personal weather station. During a controlled experimental setup in an urban environment, the Alecto was found to overestimate rainfall due to incomplete emptying of the tipping bucket. Correcting this mechanical error by a 10 percent reduction factor lowered the relative bias to 0.00 or 0.06, when comparing the station to official KNMI gauge or KNMI gauge-adjusted radar, respectively. Correlations were high between stations with non-faulty setups, but at the 5 minute resolution, correlations were substantially lowered by sampling errors caused during the data transfer to PWS data platforms. A quality control method from Vos et al. (2019) was adapted and applied to data from a citizen science project in Delft, the Netherlands, which had a 12-month period of measurements for 20 stations, and a 3-month period of measurements from 40 stations. The filtering of faulty zero measurements was improved by applying the filter on hourly accumulations, and the bias correction was stabilized. The variation over individual PWSs, however, remained high due to setup differences. The complex installation process for citizens and issues with software and data accessibility are limiting factors and warrant further research to improve the usability of PWS data for urban hydrological applications.

Contents

List of Acronyms 6

1 Introduction 7

2 Measuring precipitation with personal weather stations 9
- 2.1 Personal weather stations 9
  - 2.1.1 History of PWS platforms 9
  - 2.1.2 Quality issues with Personal Weather Station data 9
  - 2.1.3 Quality Control of Personal Weather Station data 10
  - 2.1.4 Tipping bucket rain gauges and associated errors 10
  - 2.1.5 Project PWS: Alecto WS-5500 12
- 2.2 Experimental site: Green Village 13
  - 2.2.1 Reference PWS: Davis Vantage Pro 2 13
  - 2.2.2 Ruisdael site: reference instruments 14
- 2.3 Citizen science project: Delft Measures Rain 14
  - 2.3.1 Delft Measures Rain: project history 14
  - 2.3.2 Delft Measures Rain: project ambitions 15

3 Methodology 16
- 3.1 The Green Village: experimental setup 16
  - 3.1.1 The Green Village setup: objectives 16
  - 3.1.2 The Green Village setup: lay-out 17
  - 3.1.3 The Green Village setup: Alecto station installation process 18
  - 3.1.4 The Green Village setup: access to Alecto data 19
  - 3.1.5 The Green Village setup: installation difficulties and observed errors after installation 20
  - 3.1.6 The Green Village setup: processing data from WOW 21
  - 3.1.7 The Green Village setup: tipping volume experiments 21
- 3.2 Quality control of DMR 2021-2023 22
  - 3.2.1 Data availability 22
  - 3.2.2 Reference data set and validation 23
  - 3.2.3 QC de Vos: Filters and parameters 24
  - 3.2.4 QC DMR: filters and parameter updates 26
  - 3.2.5 QC DMR addition: constant mechanical bias correction 27
- 3.3 DMR participants 28

4 Results 29
- 4.1 The Green Village: experimental setup 29
  - 4.1.1 Tipping volume experiments 29
  - 4.1.2 Comparison of statistics for Alecto stations 31
  - 4.1.3 Comparison of Alecto 1 with reference data 35
- 4.2 Quality control of DMR 2021-2023 37
  - 4.2.1 Validation metrics of DMR 22-23 QC methods 37
  - 4.2.2 Faulty zeros analysis 39
  - 4.2.3 High influx analysis 40
  - 4.2.4 Bias correction analysis 40
  - 4.2.5 DMR 23 42

4.2.6 Subsets of installation years . . . 43
4.3 DMR participants survey and home visits . . . 44
4.3.1 Participants Survey . . . 44
4.3.2 Home visits to participants . . . 45
4.3.3 Participants and data quality . . . 46

5 Discussion and Outlook **47**
5.1 The Green Village: experimental setup . . . 47
5.1.1 Quantifying setup differences . . . 47
5.1.2 Tipping volume experiments . . . 47
5.1.3 Alecto vs. Netatmo . . . 48
5.1.4 Reference stations: Davis VP2 and MRR . . . 48
5.2 Quality control performance . . . 48
5.2.1 FZ filter . . . 49
5.2.2 HI filter . . . 49
5.2.3 Bias Correction . . . 49
5.2.4 QC Alecto vs QC Netatmo . . . 49
5.3 PWS data quality and hydrological applications . . . 50

6 Conclusions **51**

A Methodology appendix **56**
A.1 The Green Village set up: materials . . . 56
A.2 The Green Village set up: Tipping experiment A . . . 57
A.3 Quality control of DMR 2021-2023 : coding flow charts . . . 58
A.4 Participants' survey . . . 64

B Results appendix **66**
B.1 The Green Village set up: tipping volume experiments . . . 66
B.2 Quality control of DMR . . . 73

List of Acronyms

Alecto Alecto WS-5500 (weather station)

API Application Programming Interface

CAL Calibration data

CV Coefficient of variation of errors

DMR Delft Meet Regen (citizen science project)

DMR 22-23 Data set of stations ranging from November 2022 until October 2023

DMR 23 Data set of stations ranging from September 2023 until November 2023

DVP2 Davis Vantage Pro 2

KNMI Royal Netherlands Meteorological Institute

Met Office British Meteorological Service

PWS Personal Weather Station

r Pearson Correlation

RMSE Root Mean Square Error

TBR Tipping bucket rain gauge

TGV The Green Village

QC Quality Control

WMO World Meteorological Organization

WOW Weather Observations Website

VAL Validation data

1 Introduction

When one thinks of the danger of flood events, this is often associated with coastal and fluvial flooding, especially in the Netherlands. Pluvial flooding has received less attention in the past, both by the academic community (Rosenzweig et al., 2018) as well as the public (Rözer et al., 2016), even though it leads to significant damage and casualties. Pluvial flooding will pose an even larger threat in the near future, due to ongoing urbanization on the one hand, and climate change on the other (Kundzewicz et al., 2014; Praskievicz & Chang, 2009). The majority of the world's population will live in urban areas and occurrence of extreme rainfall events will increase. The need for accurate and high resolution data of extreme precipitation events is therefore higher than ever.

Hydrologists and meteorologists are looking to understand more about the climate and rainfall dynamics in the urban environment, and want to improve the modeling and quantification of the impact of precipitation events. Past research has emphasized the need for high resolution data to catch the rainfall dynamics on the urban scale (Schilling, 1991; Berne et al., 2004; Ochoa-Rodriguez et al., 2015; Cristiano et al., 2019), but most rain gauge networks do not meet the criteria of high resolution data. Even in countries with advanced meteorological institutions, observations in cities are scarce. Weather stations are often placed in rural areas, to obtain more reliable measurements.

Thanks to developing technology, it has become easier and cheaper for weather enthusiasts and hobbyists to purchase automatic personal weather stations (PWS) or amateur weather stations, which often collect meteorological data at high temporal resolutions, also referred to as crowd-sourced data. This has led to the growth of PWS networks and data platforms, such as Weather Underground (Weather Underground, 2023), Netatmo (Netatmo, 2023) and Weather Observations Website (Office, 2024). The use of PWSs as a data source is sometimes met with skepticism and concerns regarding the quality of the observations (Bell et al., 2015), as the measurements and set-up are often not validated by professionals. However, in the past years, researchers have proved the value of using crowdsourced data, for example in the quantification of the Urban Heat Island effect (Fenner et al., 2017; Meier et al., 2017), as well as for wind measurements (Droste et al., 2020), and hydrological applications (De Vos et al., 2017). To overcome the quality issues, several quality control methods have been developed, for example the quality control method for PWS precipitation data developed by Vos et al. (2019), which is based on geostatistical analysis.

The Netherlands has a relatively dense PWS network, especially in urban areas. Many of the PWSs are connected to commercial platforms Wundermap and Netatmo. The British meteorological service, or Met Office, founded the open data platform 'Weather Observations Website' (WOW), where participants can choose to upload and share their PWS data.

One of the areas where the PWS network is particularly dense, is the city of Delft in the Netherlands. It counts approximately 50 stations connected to WOW, as volunteers were stimulated to connect a PWS as part of the citizen science project of *Delft Meet Regen* (DMR). The PWSs within the DMR project are all of the same brand, Alecto WS-5500, which measures precipitation with a tipping bucket mechanism. The project is run in Delft, and managed by a TU Delft-based team, which gives the possibility of contact with participants and availability of meta data. The project therefore offers an excellent opportunity for research on the quality of precipitation data collected by PWS networks, and its potential use, for example as input for urban runoff models, or to study precipitation variability in the urban climate.

This thesis adds to the efforts currently made to prove the scientific value of PWS networks, specifically for hydrological applications. To do so, research on the quality of the data is needed. The main objective is to assess the quality of precipitation measurements from the Alecto WS-5500 personal weather station, in the context of the *Delft Meet Regen* citizen science project. This objective is approached from three different research questions, dividing the research into three parts:

1. How accurate and reliable are the precipitation observations of the Alecto WS-5500, and is it possible to quantify setup differences?
2. To what extent can the quality of the DMR data set be improved by applying a QC method, and if so, how can the QC method be optimized for the Alecto WS-5500 data?
3. How does the monitoring and maintenance of owners of PWS influence data quality?

To answer the first research question, an experiment setup with 8 Alecto stations is created at TU Delft's local field lab 'The Green Village'. To answer the second question, the possibility of improving data quality is explored by applying and modifying an existing quality control method, developed by Vos et al. (2019). The data set comes from the DMR project, containing measurements from the participants in the period 2021-2023.

These first two research questions are answered by means of quantitative analyses, using statistical parameters and reference measurements. Local gauge data and radar data from the Royal Netherlands Meteorological Institute (KNMI) are used. In contrast, the third research question is answered by a qualitative assessment of the DMR project participants' experience with the installation and maintenance of their personal weather station. This is achieved by both a users survey and notes from home visits made at participants' stations.

The thesis is structured as follows. First, chapter 2 provides more information about PWSs. It briefly discusses the history, previous studies on PWS data and data quality concerns. The mechanism of tipping bucket gauges and associated errors are explained, as well as the details of the Alecto. The second part of the chapter provides context of the DMR project, describing both its history and future ambitions. Next, chapter 3 explains the methodology. It is split into three sub chapters, each explaining the methods for the three sub objectives. Section 3.1 describes the set-up of the experiment at the Green Village, 3.2 explains the application of quality control on the DMR data set. Finally, the interaction with the participants is briefly discussed in 3.3. Results of these three parts will be presented in 4, which leads to the discussion chapter 5. The thesis concludes with chapter 6, where recommendations are made for the Delft Meet project and further research on crowdsourced data in general.

2 Measuring precipitation with personal weather stations

In this chapter, background information is provided on both personal weather stations (2.1) and the DMR project (2.3). Sections 2.1.1, 2.1.2 and 2.1.3 discuss the development of PWSs and previous research on precipitation measurements with PWS. Hereafter, 2.1.4 provides an overview of research on errors associated specifically with tipping bucket rain gauges. The PWSs used in the research, the Alecto WS-5500 (2.1.5) and Davis Vantage Pro 2 (2.2.1) are discussed separately. Since the research is conducted in the context of the Delft Meet Regen project, 2.3.1 and 2.3.2 explain how the project was founded and expanded over the past years.

2.1 Personal weather stations

The weather has always been a popular subject of conversation. This is why there are many weather enthusiasts or 'hobby meteorologists' who like to monitor the weather with simple rain gauges or anemometers. With the widespread development of consumer electronics and wireless technology, many manufacturers have developed amateur weather monitor devices at reasonable prices. These devices perform basic measurements such as temperature, humidity, wind and precipitation and are often referred to as personal or private weather stations (PWS) as they are purchased by individuals. The UN's World Meteorological Organization (WMO) describes these PWSs as "low-cost automatic weather stations" (WMO, 2018a), as they have a low power consumption and transmit data automatically in (near) real time. The WMO defines low cost as $100 -$7000, which is relatively low compared to advanced weather stations and other observation instruments. The price of the PWS varies with the its complexity and quality, simple all-in-one PWSs with classic anemometers are for example cheaper than modular PWSs with sonic anemometers.

2.1.1 History of PWS platforms

The commercial weather service Weather Underground was the first online service that used data from the personal weather station community to create a database and interactive station map, 'Wundermap' (Weather Underground, 2023). Approximately 10 years ago, PWS networks started to grow worldwide, and more data platforms were founded, like the Weather Observations Website (WOW) (KNMI, 2023a) and the commercial Netatmo (Netatmo, 2023). It became easy for weather enthusiasts to upload and share the data of their local stations. The crowdsourcing of weather data via these platforms offers a potential treasure of data for meteorologists and hydrologists, due to the combination of high resolution in the spatial as well as the temporal dimension.

2.1.2 Quality issues with Personal Weather Station data

Although PWS networks have a high potential, they also come with drawbacks, related to availability and quality of the data (Bell et al., 2015). Due to connectivity issues (WiFi, low battery), data gaps may occur. When data is not freely accessible, availability is limited. This is the case for Netatmo, who charges researchers for the use of datasets. On the other hand, data from WOW is available for use as long as the owner of the station has given permission. The participants of DMR are connected to WOW, and the DMR-PWS data is hence available through this platform.

The quality of data obtained from PWSs, however, is not monitored and cannot be guaranteed, because it is automatically uploaded and the stations are owned by individual citizens. Several types of errors are associated with personal weather stations, related to the instrument itself (e.g.

low sensor quality, deterioration over time) as well as the set-up of the instrument with regard to the surroundings (Bell et al., 2015; De Vos et al., 2017). More information on possible errors is given in 2.1.4. The precipitation measurements from a PWS network of Netatmo in Amsterdam were found to underestimate rainfall by 11.1% (Vos et al., 2019) compared to climatological gauge-adjusted radar data, most likely caused by the fact that the PWSs are not shielded from wind. (Pollock et al., 2018).

Netatmo rain gauges have been used in several studies, and the accuracy is reported to be 1 mm/hr by the manufacturer. De Vos et al. (2017) have tested an experimental set-up with three Netatmo stations, and concluded that the accuracy was high compared to a well-calibrated operational reference rain gauge used by KNMI. For the device used by DMR, the Alecto WS-5500, the manufacturer reports a precipitation measurement accuracy of '+/- 10%', however no further details are provided. This value probably means that measurements may deviate up to 10% from the calibrated value. As of yet, no literature is available on the quality of the Alecto.

2.1.3 Quality Control of Personal Weather Station data

Since the quality of PWSs is variable, the data can be improved by filtering and addressing the instrument bias. Several methods of quality control (QC) or quality assurance tests have been proposed in literature (Estavez et al., 2011; Bárdossy et al., 2021; Beele et al., 2022).

A QC method for the Netatmo stations in the urban area of Amsterdam, published by Vos et al. (2019), uses several filters based on geo-statistics. It was found that the filtered dataset of PWS measurements shows great improvement in the Pearson correlation (r), the relative bias, and the coefficient of variation of errors (CV), using the climatological gauge-adjusted radar product of the KNMI as reference data set.

Building on the previous efforts of building a QC method, a recent study by Overeem et al. (2023) found that the crowdsourced precipitation data from Netatmo stations can be used to improve pan-European radar precipitation products. After performing the QC, the average underestimation by the radar product was reduced from 28% to 3% in the PWS gauge-adjusted radar set, offering a much better real-time product than previously available. It should be noted that at lower temperatures, the Netatmo gauges were found to highly underestimate precipitation, which is explained by their inability to measure solid precipitation (e.g. snow). This limitation is also highlighted by WMO (2018a).

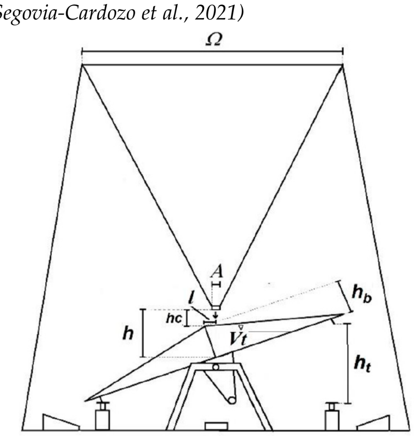
2.1.4 Tipping bucket rain gauges and associated errors

Both the Netatmo and the Alecto WS-5500 use a tipping bucket mechanism to record precipitation. The use of gauges with a tipping bucket mechanism is a popular method to collect precipitation data, not only by individuals but also by official meteorological institutes. Still, many errors are associated with the mechanism. Segovia-Cardozo et al. (2023) have collected research on tipping bucket gauges and written a review on these uncertainties and strategies to reduce errors. Below, the research on the topic relevant for this project is summarized.

The tipping bucket rain gauges (TBRs) that are described in the review fall into the category of *recording catching point rain gauges*, with *point* referring to the fact that the gauge only measures rain at one location. Catching refers to the direct measuring of rain, as opposed to the remote sensing radar. The label *recording* is given because the measured volume is automatically recorded, instead of manually.

The 'bucket' itself is actually a small spoon-like object, which receives precipitation through a funnel, see figure 1a. Once the bucket is filled with the *tipping volume*, it tips down and drains the collected rain. While tipping, the magnet located on the other side of the bucket passes a reed switch, and the fixed rainfall amount of one tip is recorded. By counting the number of tips within an interval, the rainfall rate and accumulated rainfall are determined. Most TBRs described in the literature tip on both sides, while the Alecto WS-5500 has only one tipping bucket (figure 1b).

(a) *Schematic overview of tipping bucket mechanism (Segovia-Cardozo et al., 2021)*



(b) *Unlike other TBRs, the Alecto WS-5500 is one sided.*

Figure 1: *Tipping bucket mechanisms.*

Instrument error

It was found that total rainfall is generally underestimated by TBRs (Segovia-Cardozo et al., 2023). The main causes were given as mechanical bias and wind induced bias. The mechanical bias is introduced by incoming rain during the tipping movement, which is not recorded. With a higher rainfall intensity, and more frequent tipping, more water is lost. With an ever higher intensity, a continuous flow of water counters this underestimating effect, as the bucket tips at a high rate, and before it is completely filled, thereby reducing the tipping volume (Segovia-Cardozo et al., 2021). The magnitude of the bias thus varies with rain intensity.

Other instrument errors were reported, such as incomplete emptying of the tipping bucket, wetting losses in the funnel, losses from evaporation and splashing, as well as mechanical blockages of the funnel and logging errors (Segovia-Cardozo et al., 2023). These errors depend on the specific TBR design. Together with human errors, such as faulty set-ups or lack of maintenance, a combination of instrument errors might explain the "random" errors that were found in many of the reviewed papers. It is stated that these random errors are small in comparison to the mechanical and wind induced bias. For citizen science projects, this could however be debated, as human errors may have a significant impact on measurement results. This is especially the case when funnels are blocked, filled with debris or housing insects or birds, due to neglected maintenance duties.

To reduce instrumental errors calibration can be performed. The exact tipping volume can be determined by slowly filling the bucket until it tips. This static calibration does not account for the variable bias that is influenced by the rainfall intensity. Such a calibration curve can be determined dynamically, by applying different rainfall intensities. A precision pump and high-resolution weighing device are needed, and the testing must be carried out in a certified laboratory according to the guideline EN 17277:2020. Calibration of TBRs in the field is also possible, but this requires a portable field calibrator by instruction of the WMO. Dynamic calibration is thus not easy to perform. (WMO, 2018b)

Wind induced error

The aforementioned wind induced bias is explained by the fact that the TBR affects the surrounding wind flow, causing acceleration of wind around the funnel. This results in underestimation, as more raindrops are blown over the funnel, and are not caught by the gauge. This undercatch is found to be between 2 and 10 percent (Segovia-Cardozo et al., 2023). To avoid this effect, gauges can be placed in a pit, or hole in the ground, so that wind flow is not affected. To determine the wind induced bias experimentally, the measurements of a pit TBR and a normally placed TBR are compared.

Sampling error

If the logging frequency is higher than the time it takes to reach the tipping volume, rain at low intensity goes undetected. Subsequently, a higher rainfall rate may be recorded in a later interval, when the tipping volume is finally reached. These are *sampling errors*, as they are introduced by the sampling frequency. A solution mentioned is to use a dynamic *inter tip* time scale, where each tip is compared, rather than a set time interval. For cheap instruments as the Alecto, this would not be possible as the individual tips are not recorded.

Figure 2: *The Alecto WS-5500 (Alecto Home, 2023)*

2.1.5 Project PWS: Alecto WS-5500

WMO distinguishes three types of PWSs: compact, stand-alone instruments and all-in-one. The Alecto WS-5500 studied in this project belongs to the latter category, a single unit station that contains eight different sensors. It is a relatively cheap (commercially available for €200) instrument with the shape of a three pointed star (see figure 2). Two of the arms contain a cup anemometer and wind vane for measuring the wind speed and direction respectively. The third arm carries a tipping bucket system for measuring precipitation, and a thermometer, hygrometer and barometer shielded by a radiation screen. UV and light intensity are measured on top of the weather station. The instrument runs on solar power with backup batteries, and is wirelessly connected with a small indoor unit, on which live measurements can be monitored. The indoor unit also contains temperature and humidity sensors, as well as a pressure sensor. The data is not logged locally, but can be uploaded to aforementioned weather platforms such as WOW or Wunderground.

The Alecto's reviews by online purchasers are generally good, due to its good price-performance ratio and the easy installation of the outdoor units (KiesKeurig, 2023; Alecto Home, 2023). Some users have mentioned it needed replacement after approximately two years. The relatively quick deterioration of the plastic units by exposure to weather is one of the disadvantages of PWSs mentioned by the WMO (2018a). Due to its all-in-one and easy-to-use characteristics, separate replacement of sensors is not supported by the manufacturer, as opposed to instruments of the compact PWS category. The compact PWSs often also allow local data logging and transfer of data to a local network. With all-in-one stations like Alecto, data storage is only possible with specific software and websites, so that they are not compatible with other systems. A working internet connection is also a requirement for data collection.

2.2 Experimental site: Green Village

The Green Village is one of TU Delft's field labs, and is located on campus next to the faculty of Civil Engineering. The lab facilitates the testing of sustainable innovations for the urban environment on a real-life scale. The Green Village consists of several low rise buildings, a few streets, gardens and a square, thus representing a small part of an urban neighborhood (figure 3).

Figure 3: *The Green Village (Nationale Milieu Database, 2023)*

2.2.1 Reference PWS: Davis Vantage Pro 2

One type of reference station on the Green Village is the Davis Vantage Pro2, or Davis VP. Like the Alecto WS-5500, it is a wireless station that is easy to install and thus popular among weather amateurs. However, with prices close to 1000 euros, it is around four to five times more expensive, and thus regarded as a more advanced station. A study on the accuracy of the Davis VP versus standard meteorological equipment showed that it overestimated rainfall by approximately 1%

during the measurement period of one year (Burt, 2009). However, during single events, the over- or underestimation could be as high as 20%. The author suggests some explanations to the deviations, such as the Davis VP being sensitive to heavy dew fall. They also suspect that the Davis VP overestimates more during heavy rainfall. They conclude that the Davis' rain gauge performance is not good enough for recording daily and monthly totals, compared to manual checkgauges.

2.2.2 Ruisdael site: reference instruments

On the terrain of the Green Village, a site of the Ruisdael Observatory is located. This is an initiative to obtain better monitoring and forecasting of weather and air quality in the Netherlands (*Ruisdael Observatory*, 2023). Due to the urban environment in Delft, the Ruisdael site on the Green Village is only used to test the performance of instruments, but this nevertheless provides the Green Village with some potential reference data.

One of the instruments on the Ruisdael site is the Micro Rain Radar, produced by Metek (METEK, 2023). The Doppler radar calculates the rain rate and drop size distribution in a vertical profile above the radar. The time resolution of data is 10 seconds, and the range resolution 35 meters. The accuracy or error is not mentioned by the manufacturer.

Other meteorological instruments at the Ruisdael site on the Green Village that could be used as reference stations, are disdrometers, one Clima LPM model from Thies, and two Parsivel2 disdrometers from OTT HydroMet. With a laser beam, the disdrometer can measure the size and speed of a raindrop, as to determine the range of particle sizes, types of precipitation, radar reflectivity, visibility, and importantly, the intensity of the precipitation. Thies reports the accuracy of the precipitation measurement to be around 5%, while the error in intensity/quantity measurement is reported to be lower than 15% for rain. OTT HydroMet does not report the accuracy.

2.3 Citizen science project: Delft Measures Rain

2.3.1 Delft Measures Rain: project history

Delft Meet Regen, or Delft Measures Rain (DMR), is a citizen science project. Delft is a small city in the Netherlands, located in the densely populated and highly urbanized area between The Hague and Rotterdam. DMR started in the summer of 2020 with a precipitation measurement campaign, when nearly 100 volunteers measured daily accumulated precipitation with homemade rain gauges. The following year, the campaign was upgraded with store-bought funnels for all participants.

Data analysis of these campaigns showed spatial variability in the daily accumulated precipitation, which might be explained by the convective nature of the precipitation events. A comparison with radar data from KNMI showed that there was little deviation in the weekly average amount of precipitation measured. The manual gauge results were quite accurate, especially in case of large precipitation amounts (>4 mm/day), this might be because the effect of instrument bias and reported reading difficulties by participants was relatively smaller in these cases.

Encouraged by the positive results and enthusiastic response from participants, DMR started to focus on the use of automatic personal weather stations of the brand Alecto WS-5500, which were installed on the property of volunteers with help of the DMR staff. These PWSs measure precipitation at frequent time intervals (every 5 minutes) and other climate variables such as temperature, wind speed and humidity, provided there is a stable internet connection.

The shift to PWSs was made for several reasons. Quite some manual measurements by participants were not taken in the same agreed time interval, between 8 AM and 10 AM everyday. This led to the exclusion of a large part of data for daily accumulated precipitation analysis. Secondly, even with complete daily data, it would be challenging to find patterns or reasons for spatial variability between different areas in Delft, due to the relatively low temporal resolution. Since the PWSs collect data continuously at a high resolution, they have the potential for a more robust spatial analysis as compared to the manual observations.

2.3.2 Delft Measures Rain: project ambitions

The collection of precipitation data in the Netherlands by KNMI is covered by radars, automatic weather stations at high temporal resolution but low spatial resolution (1 per $\sim$ 1000 km^2) and manual precipitation stations (24h, 1 per $\sim$ 100 km^2). Although the data is of high quality, knowledge about spatial variability at high spatial resolution in urban areas is still limited. Official weather stations are set up outside of the city, to ensure more reliable results.

DMR was set up to gain more insight into spatial variability of precipitation in urban areas, how the urban lay-out influences the variability, how and why the results from the ground measurements differ from other precipitation data, and if there is potential to improve the quality of the radar products by using data from PWSs. Moreover, as DMR is a citizen science project, the past years have been an experiment in the social side as well: to learn more about citizen science communication, how to engage volunteers in a science project, and using their feedback and results to improve the project, for example by changing measuring instruments after the first campaign.

Before the start of the new summer campaign of 2023, the project was renamed to *Delft Meet* (DM) or Delft Measures, as more climate variables than just rain would be measured. With a rise in budget from a Climate Action research grant, more PWSs have been installed in the city of Delft. During August 2023 the number of stations have doubled from approximately 25 to 50. Some participants also installed a soil moisture sensor in their garden, which sends data to another citizen science project called *Pientere Tuinen*, or "Clever Gardens". Soil moisture sensors and additional PWSs will also be installed on the campus of TU Delft, thereby creating the opportunity for more research on the topic of the urban microclimate, such as the influence of buildings on wind and precipitation on the micro scale. All of the PWSs that have been installed during the project have also collected data on temperature and humidity, which might prove useful for future research projects.

3 Methodology

This chapter provides a detailed description of the research methods that were used in this thesis. The chapter is divided into different sections, that each describe the methodology linked to one of the research objectives. First, the experimental set up in the Green Village is presented (3.1). Second, the data set from the DMR project, and the quality control method that is applied to it is explained (3.2). Last, 3.3 contains the interaction with DMR participants and set-up of a survey.

3.1 The Green Village: experimental setup

To investigate the accuracy of precipitation measurements by the Alecto WS-5500, an experimental setup was created on the terrain of TU Delft's living lab *The Green Village*, or *TGV*. A total of eight Alecto weather stations were installed in various locations on the site, and observations were gathered in the period from June to December 2023. Additionally, several short-term experiments were conducted with the on-site weather stations. The reason to use TGV for the experimental setup was threefold. First, TGV has reference stations available on site, which allows for external validation of results. Second, TGV mimics the urban environment quite well, with low-rise buildings placed closely together and small roads and green spaces in between. Last, TGV is conveniently located on the TU Delft campus, and the managing staff allowed and encouraged the use of the site for the project of *Delft Meet*.

The objectives of the setup are explained in 3.1.1. Then, the lay-out (3.1.2) and installation process (3.1.3) are discussed in further detail. The accessing of the gathered data is explained in 3.1.4, followed by a separate section on difficulties encountered during the installation process (4.3.2). The additional experiments are discussed separately in 3.1.7.

3.1.1 The Green Village setup: objectives

There were three main objectives of the experiment at TGV. The first was to measure the accuracy of the Alecto stations, by means of comparing the measurements between the Alecto stations (see figure 4), and comparing the results to the measurements of the reference stations .
This was achieved by placing three stations close together on the Ruisdael observatory site. It was expected that the three stations receive an equal amount of precipitation since they were placed within 2 meters of one another. This way, possible variations between stations should indicate the magnitude of standard deviation of the Alecto. A consistent over- or under-estimation compared to the reference stations would indicate a relative instrument bias of the Alecto.

Figure 4: *Alecto 1, 2 and 3 aligned at the Ruisdael site*

The second objective was to investigate the effect of the micro environment on precipitation measurements. This was done by placing the stations in different locations on TGV. Possible variations between stations, larger than the standard deviation, might show the influence of the difference of environment.

The setup also created an opportunity for further testing of the stations, as they were already installed and proven functional. This allowed the tipping volume experiments to take place, which was the additional third objective.

3.1.2 The Green Village setup: lay-out

The three stations at the Ruisdael observatory were placed in a row, with approximately 85 cm in between each station. Other weather observation equipment was located at the Ruisdael site within close range (<5 meters), such as the Davis Vantage Pro2 and several disdrometers. Obstacles, including a fence, control room and trees were also close to the stations, within a range of 10 meters. The Alecto stations at the Ruisdael site were labeled 1, 2 and 3.

The other five Alectos were placed in different locations around TGV, see figure 5 for the complete overview. Station 4 was in the *Heat Square*, in one of the small green plots. The specific plot was chosen because it contained another sensor for temperature and humidity, from a different research project. By placing the Alecto in close range, it could act as a reference in future research projects.

Station 5 was placed in between two of the buildings that surround the heat square, very close (75 cm) to the wall of a two-story house. This was done to test the influence of the nearby walls.

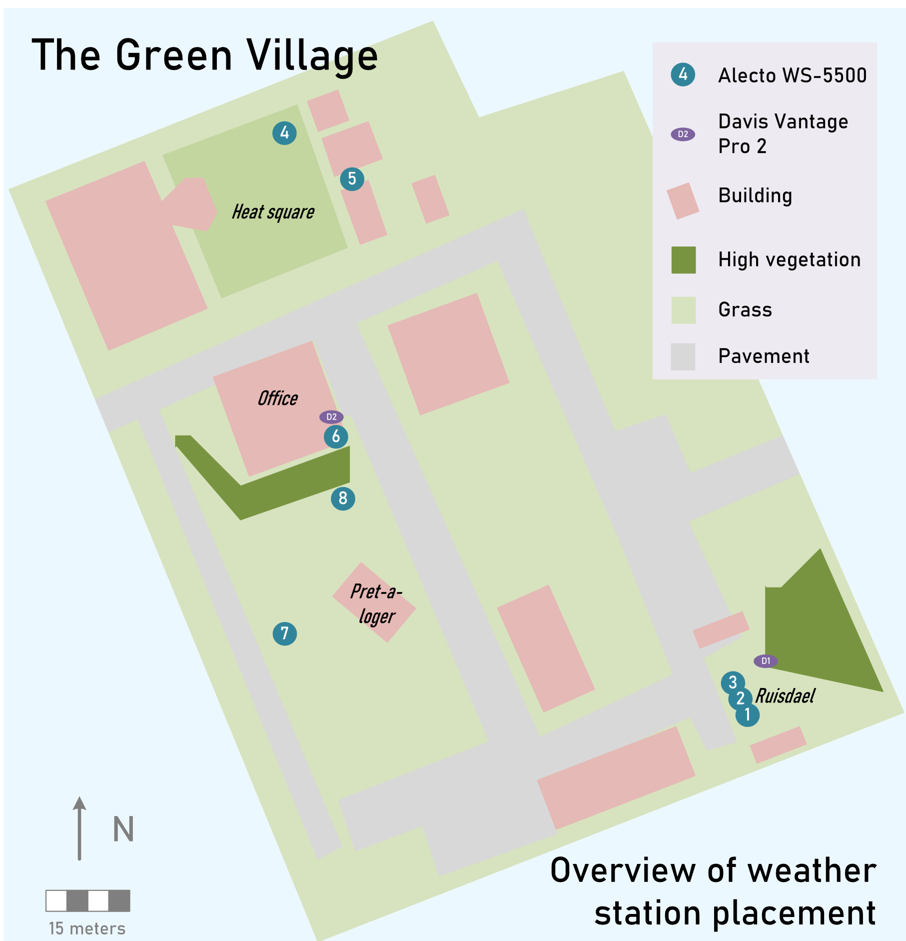


Figure 5: *Map of placement weather stations at the Green Village.*

Next, two stations were placed close to each other, around the single-story building of the office of TGV. Station 6 was placed on the roof of the office, and station 8 in the small yard next to it. The idea of this setup was to allow comparison of a roof measurement to a ground measurement, which might be influenced differently by wind. Additionally, a Davis Vantage Pro2 was already installed on the roof of the office building, providing another reference station. The yard contained some obstacles in close range (2 m), mainly trees and bushes.

Station 7 was placed on the edge of TGV, behind one of the residential buildings. This location was different than other ones, as it was less obstructed by buildings and trees, especially in the south-west direction, which is the most common wind direction in the Netherlands.

Setup changes with tilted stations

The stations' set up as described in the previous paragraphs was meant to mimic the urban garden settings of the DMR project participants. As surrounding obstacles are not the only potential source of errors, additionally, two stations were tilted to simulate a faulty set-up. Since the set-up of station 4 was comparable to that of stations 1-3 and gave similar measurements, it was decided to slightly tilt this station. The station was mounted wrongly on the pole, leaving it slightly off-level.

After a couple of months, it became clear that the tilting of station 4 did not yet result in large differences in precipitation measurements, thus a more drastic tilting was applied to station 3. During the experiment in July and August, enough data was collected to compare the three stations on the Ruisdael site. The pole of station 3 was then re-installed at an angle, leaving the station tilted parallel to the emptying direction of the tipping bucket. This was done on the 7th of September. Figures 6, 7 and 8 show stations 1, 2, and 3. While station 1 and 2 are both installed level, a small difference can be noted between 1 and 2 as well, as the bubble in the red circle is slightly off-center.

Figure 6: *Station 1*

Figure 7: *Station 2*

Figure 8: *Station 3*

3.1.3 The Green Village setup: Alecto station installation process

The Alecto stations are relatively small and light-weight at 1.09 kg (Alecto Home, 2023), with efficient packaging. The assembly of different parts, such as securely attaching the anemometer on the outside unit, is relatively simple. The inside unit should be connected to an outlet, while the outside unit is equipped with a small solar panel. Both units require batteries as back-up.

The outside unit includes a system with brackets, so that it can be easily mounted on a pole, which is not included in the package. In this set up, a PVC pole with a length of 1 meter was used. It was inserted in the ground with a depth of 30 cm, leaving the outside unit at a height of 70 cm above the ground. This length was chosen because it was the same as the other volunteers in the Delft Meet project have used in the past years. An auger was used to construct the hole in the ground. This process was more complicated, as the soil was very dry. By wetting the soil, and using a hammer, the pole was put in the ground (figure 9a).

Another more complex part of the installation process is to establish a connection between the inside and outside unit, and the inside unit to a data platform. The inside and outside unit should be able to connect with each other automatically, except when installing multiple stations at once. This issue is further discussed in section 4.3.2.

Since the inside unit does not log data locally, it needs to be connected to an online data platform. The manual provided by Alecto only briefly covers this possibility, which is why Delft Meet wrote its own manual on this process. It is possible to connect to several platforms, but because Delft Meet uses the Weather Observations Website, the stations at TGV were connected to this platform. To connect to WOW, the smartphone app WS View is needed, as well as a Met Office user account. The details of the connected weather station can be entered on the website to inform other users, such as the location, the reason for installation, photos of the station, etc. The app is necessary to connect the inside unit to the local WiFi network.

A complete list of materials can be found in the A.1.

(a) *Installation of the pole of station 3 at the Green Village.*

(b) *Testing of tipping bucket mechanism*

Figure 9: *Installation process of the Alecto WS-5500*

3.1.4 The Green Village setup: access to Alecto data

As described previously, the observations are not stored locally, but updated approximately every 5 minutes to the data platform, as long as the indoor unit is connected to the internet. The data can be viewed at `https://wow.metoffice.gov.uk`, where rain is displayed both in rainfall intensity in mm/h, and accumulated rainfall in mm, which is reset to zero at the beginning of each day (00:00 local time). The data can be exported with a limited maximum observation length of 1 month. Another possibility is to view the observations at the Dutch website, `https://wow.knmi.nl/`. There seems to be a discrepancy between the data from the two websites, as the timestamps attributed to observations are not exactly the same. Furthermore, only rainfall intensities are reported. When converting these back to accumulated rainfall, there is a high chance of calculation errors.

Another possibility to obtain data in a more efficient way, is to make use of an API. Different API calls are available on `https://mowowprod.portal.azure-api.net/`. The use of the API was not possible without signing up, and access to data seemed to be limited to a small number of stations. As the API was not used in this thesis, the topic will not be discussed in further detail.

3.1.5 The Green Village setup: installation difficulties and observed errors after installation

During the installation process, several difficulties occured, related to either material or digital malfunctioning, which could have lead to measurement errors.

Creating a solid construction with the PVC pole was challenging, especially with dry soil. It was hard to stabilize the pole, and to make sure that the station was exactly level.

With construction activities and trees in the vicinity of the stations, the funnel can easily get blocked, and sand or other material can partially fill the tipping bucket (figures 10, 11, 12), leading to faulty measurements. During the testing of the tipping bucket measurement 9b, it was also noted that the tipping bucket did not empty completely, which may lead to overestimation.

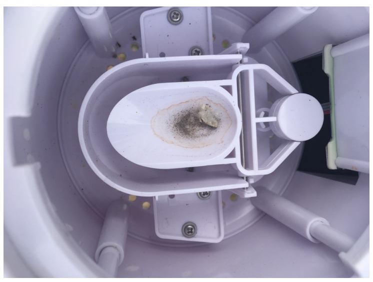


Figure 10: *Bucket with residue of polluted .*

Figure 11: *Bucket partly filled with sand and organic material.*

Figure 12: *Funnel blocked by unknown organism.*

Connecting the weather station to the internet through the mobile app was often a time consuming process, as the connection was not always established right away. The app also did not give clear indication of a successful installation, meaning that the connection has to be tested by manipulating the weather station into a measurement. For example, by giving the tipping bucket a few tips during dry weather, and checking whether the unit registered rain.

With the launch of a new app "WS View Pro", later in 2023, this problem seems to have been partly solved. The new app indicated more clearly whether the connection between station and the internet was successfully established, and was able to troubleshoot if necessary.

A larger problem presented itself when it was found that two of the inside units reported exactly the same measurements. The suspicion that both inside units were connected to the same outside unit was confirmed when tipping the bucket of only one outside station, and seeing that both units reported the same amount of rainfall. The issue was fixed by reinstalling the connection between the inside and outside units of all weather station, at a location outside the Green Village, where no other weather stations could interfere with the signal. The wind vanes were fixed with tape to a unique direction, to monitor that all inside units did indeed report the observation of the correct station. During the remainder of the experiment, the problem did not reoccur. Nevertheless, it is a serious issue and a warning for future projects. If the connection between inside and outside unit is not unique for every weather station, this leads to erroneous data.

3.1.6 The Green Village setup: processing data from WOW

The precipitation data output is given in both "Rainfall Accumulation" and "Rainfall Rate", and is given in an interval of approximately 5 minutes. The length of intervals is often 320 seconds, but varies in both directions. The reason for this variability remains unknown, but it might have to do with connectivity issues of the WiFi connection. Due to the non-constant interval length, it was decided to work with "Rainfall Accumulation" rather than "Rainfall Rate". The "Rainfall Accumulation" represents the accumulation through out the day, and resets every day at 00.00 local time, although time in the output data is given in UTC. A new data column "Rain" was added to each station to calculate the rainfall amount per interval. In this way, resampled daily rainfall sums were correct for the UTC timezone. All reference data was also supplied in UTC.

While analyzing this data, it appeared that there were also irregularities in the observed rainfall amount. For single tips, a range of numbers from 0.1778 to 0.3048 mm was reported. One would expect tipping bucket data to be a constant tipping volume multiplied by the number of tips. When observing data which was downloaded with the API instead of the user interface at WOW, it was found that the reported amounts corresponded to a range of 0.007 to 0.012 inches. This explains the high number of decimals reported in mm, but not the lack of discrete steps. The amounts of 0.2032 (0.008 inch) and 0.3048 mm (0.012 inch) were most reported. When looking at the way the Wundermap data was recorded, it seemed most logical that a tip must represent 0.01 inch (0.254 mm). To further investigate the issue, an experiment was conducted with the stations at the Green Village (3.1.7). The Alecto company was also contacted to find out how the Alecto was calibrated. At the moment of this research, the company has not been able to provide an answer.

3.1.7 The Green Village setup: tipping volume experiments

Between 4th and 15th of December 2023, two additional sets of experiments at the Green Village were conducted.

The first, "tipping volume experiment A", was designed to check the relationship between the number of tips occurring at the outdoor unit and the amount of rainfall reported by the inner unit. For each station in the experiment, a number of tips was generated manually by flipping the bucket. After a pause of more than 5 minutes, this was repeated with a different number of tips. The number of tips and time of flipping were noted, so that they could be related to the rainfall accumulation data and timestamp from the data platform. The 5 minute interval was chosen to make sure that the different accounts of flipping would not be attributed to the same timestamp in the data platform. The experiment was conducted for all stations except 6, as this is the station mounted on the roof and not easily accessible. The sequence of number of tips was slightly different for some stations and can be found in table 6. For most intervals, low number of tips were tested, as they are most common within the 5 minute data. For each station, at least one interval with a large number of tips was also tested, to check the possibility of reporting differences.

The second experiment, "tipping volume experiment B", was designed to manually determine the amount of water per tip as accurately as possible. To achieve this, tap water from a 5 mL syringe was carefully released into the tipping bucket, until the tip occurred. Due to the sensitivity of the tipping bucket mechanism, this experiment was repeated 10 times to ensure a more reliable result. As already observed during earlier maintenance rounds, the tipping bucket does not empty completely after a tip. Therefore, the experiment with the 5 mL syringe was repeated 5 more times, while completely drying the bucket with paper after each tip. This experiment was conducted for stations 1-5 at the Green Village.

3.2 Quality control of DMR 2021-2023

This section describes how the data from the DMR project is filtered by quality control (QC). The proposed QC method is initially the same as the QC method published by Vos et al. (2019), and will be further referred to as QC de Vos. However, as the DMR data set differs from the Netatmo data set, the method and its parameters are changed to investigate whether the quality control can be optimized.

3.2.1 Data availability

In QC de Vos, the PWS data set is split into a calibration data set (CAL) and a validation data set (VAL), both with a length of one year, with a 1 month "warm-up" period for the filters. The DMR project started in the fall of 2021, with less than 10 operational stations. One year later, in September 2022, almost 20 stations were available. It was therefore decided to use this second year to test the QC methods. To account for the dry October month of 2022, a warm-up period of 2 months was chosen, so that the data set runs from November 2022 until October 2023, with September and October 2022 as warm-up period. This data set will be abbreviated to 'DMR 22-23'.

It is not possible to validate the results from the QCs of the DMR 22-23 data set against a very similar data set, as there is no other set with a full year of observations for the same stations. To provide a form of comparison, QC methods are run for the period of September 2023 to November 2023. Although this is a short period, there was a lot of precipitation in these months. Additionally, approximately 20 extra stations installed in august 2023. This data set will be referred to as 'DMR 23'.

An overview of the data availability is given in figure 13.

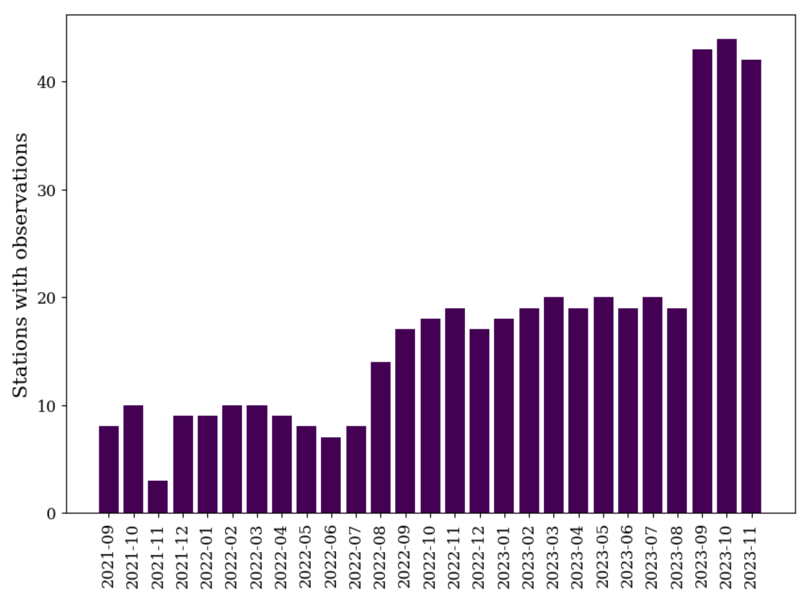


Figure 13: *Data availability from the personal weather stations in the DMR project, where stations are counted when the availability of hourly observations is above 75 percent.*

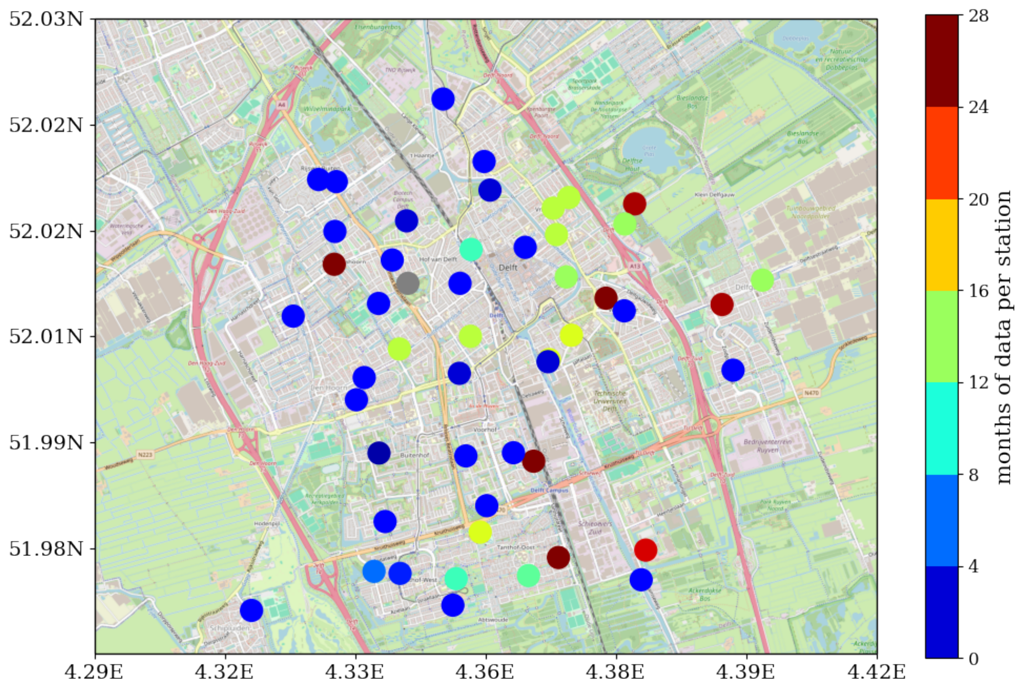


Figure 14: *Map of Alectos from the DMR project in Delft from September 2021 to November 2023, showing the data availability per station (more than 75 percent hourly observations per month). The blue stations are mostly from the last campaign, installed in 2023, but could also be stations from 2022 with little recorded data. The red stations were installed in 2021.*

3.2.2 Reference data set and validation

The reference data set is the same as used in QC de Vos, namely the climatological gauge-adjusted radar product from KNMI (KNMI, 2023b) that is adjusted with both the manual gauge network and the automatic gauge network of KNMI. This product contains the 5 minute precipitation and 60 minute accumulated depths at a 1 km grid, and is considered the best available product, even though the use of radar as reference is limited by the fact that radar and ground measurements are fundamentally different. Since February 2023, a climatological gauge-adjusted product with even better quality is available, as data from more radars were merged (Overeem & Leijnse, 2022). However, since the PWS data set spans an earlier period, this product will not be used in this analysis.

Validation in QC de Vos was done with the Pearson correlation coefficient (r), the relative bias or bias, and the coefficient of variation (CV), as described in the equations below.

Coefficient r is used to describe the strength of the linear relationship between the PWS and reference data.

$$r = \frac{\mathrm{cov}(R_{\mathrm{PWS}}, R_{\mathrm{ref}})}{\sigma(R_{\mathrm{PWS}})\,\sigma(R_{\mathrm{ref}})} \tag{1}$$

The relative bias describes the over- or underestimation of precipitation by the PWS, relative to the reference data.

$$\text{bias} = \frac{\overline{\Delta R}}{\overline{R}_{\text{ref}}} \tag{2}$$

$$\Delta R = R_{\text{PWS}} - R_{\text{ref}} \tag{3}$$

Statistic CV describes the ratio between the variance and the mean.

$$\text{CV} = \frac{\sigma(\Delta R)}{\overline{R}_{\text{ref}}} \tag{4}$$

The same descriptive statistics have been used in this thesis, to enable comparison to QC de Vos used on the Netatmo data. Other statistics reported are the the number of observations n and the percentage of data filtered when applying a QC method. A metric that was added is the root mean square error, or RMSE, which is calculated as:

$$\text{RMSE} = \sqrt{\overline{(\Delta R)^2}} \tag{5}$$

Due to the gauge-pixel discrepancy when comparing gauge data to radar data, it was noted that for accurate gauges, r does never reach the optimum value of 1, but is approximately equal to 0.75, when considering 5-minute intervals. The relative bias was found to be generally negative for Netatmo stations due to the influence of wind, as previously explained in 2.1.4. For the Netatmo data set, CV was reduced from 53.24 to 7.19, if outliers were filtered, reducing the variance drastically. A CV of 5 was considered optimally attainable, because radar data was used as validation.

3.2.3 QC de Vos: Filters and parameters

QC de Vos filters 3 types of outliers; faulty zeroes (FZ), high influxes (HI), station outliers (SO), and performs an additional bias correction (BC). The filter methods uses 11 different parameters, the concepts are explained below. Table 1 summarizes the parameters for QC de Vos. The coding flow charts of the filters can be found in A.3.

Faulty Zeroes (FZ) - d, n_{int}, n_{stat}

Faulty zeroes occur when the sensor does not measure precipitation while it should, for example when there is a blockage due to leaves, or if the station as tilted or fallen. To determine a FZ, the median of the precipitation measurements of stations within range d of the PWS is computed. FZ is true when the calculated median is larger than zero for at least n_{int} time intervals, while the station only reports zeroes. If there are fewer than n_{stat} stations within range d, FZ cannot be calculated.

High Influx (HI) - d, n_{stat}, ϕ_A, ϕ_B

High influx are the opposite of FZ, when an unrealistically large amount of rainfall is recorded, for example when the station is cleaned, tilted or water is accidentally poured into the funnel by sprinklers. When the computed median of stations within range d of a PWS is lower than a threshold value ϕ_A, the HI is true when the PWS itself reports a value above threshold ϕ_B. When surrounding stations measure higher than threshold ϕ_A, the threshold for the PWS becomes flexible, and HI is true only when the reported value is higher than the median multiplied by $\frac{\phi_B}{\phi_A}$. If there are fewer than n_{stat} stations within range d, HI cannot be calculated.

Bias Correction - DBC

A default bias correction factor (DBC) was used by Vos et al. (2019) because they found that Netatmo gauges generally underestimated precipitation, due to the fact that the PWS are not shielded from wind, and secondly because the tipping bucket volume was slightly too large for most stations. The DBC is determined during the warm-up period, in their case one month of observations, with the following equation:

$$\mathrm{DBC} = \frac{1}{1 + \mathrm{median(bias)}} \quad , \tag{6}$$

where the median of the bias is determined by comparing all PWS observations that were not flagged as FZ or HI, to the reference climatological data set. If such a warm-up period or reference data is not available, a DBC of 1 is advised.

Station Outliers (SO) - d, n_{stat}, m_{int}, m_{rain}, m_{match}, γ

Station outliers are determined by comparing a time series of m_{int} measurements, or m_{rain} nonzero measurements of that station, with n_{stat} neighboring stations within range d, with at least m_{match} simultaneous measurements. Then, r and the bias are calculated for every neighboring station. SO is true if the median of r is lower than threshold γ. If there are fewer than n_{stat} stations with m_{match} measurements within range d, SO cannot be calculated.

Bias Correction (BC) - BCF, β

The bias correction factor (BCF) is identical to the DBC at first, but changes for each PWS when threshold γ is exceeded during the SO filtering. The previously mentioned calculated median bias is used to calculate BCF_{new}. When $|\log(\mathrm{BCF}_{new} / \mathrm{BCF}_{prev}| > \log(1+\beta)$, BCF_{prev} is replaced with BCF_{new}, so each station has its own bias correction factor.

Parameter settings

The parameter values were chosen by de Vos et al. based on, first, the improvement of r, bias and CV during the validation, and, second, the amount of data filtered.

The only change in parameter value with respect to the original QC de Vos, was range parameter d. In QC de Vos, d was set to 10,000 m or 10 km. A variogram made for the 23 data set, shows that the range is not smaller than 5 km (figure 15). During convective events, d might be smaller, as also noted by van Andel (2021). The data sets used in this thesis are too short and small however to investigate the influence of such events on the range parameter. Since the largest distance between two stations in Delft is approximately 6 km, and the number of stations within 5 km range larger than $n_s tat$, d was set to 5 km.

(a) *Distances* **(b)** *Variogram*

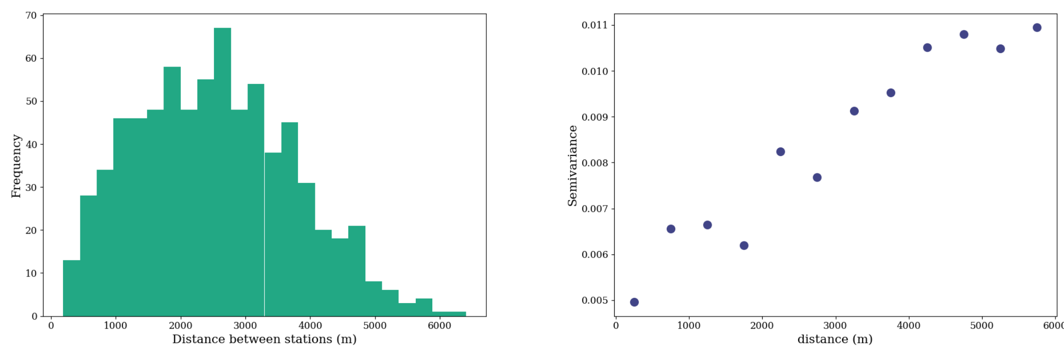


Figure 15: *Histogram of inter-station distances and semi-variogram constructed of all non-zero DMR data.*

3.2.4 QC DMR: filters and parameter updates

To investigate whether the QC method can be optimized by changing filter mechanisms or parameters, different combinations of filters and parameters were made and tested. The combining process was performed by trial-and-error on shorter and small pieces of the time series data. As QC de Vos was already proved to be functional, no major changes in the overall structure of the QC were made. The different combinations are named and explained below, and an overview of changes is given in table 1. Coding flow charts of the changed filters can be found in A.3.

QC de Vos 5 and QC de Vos 60

In the original QC de Vos, filtering only takes place on 5 minute resolution data, and to obtain hourly data, the twelve non-flagged intervals of 5-minute data were accumulated.

One of the major changes in QC in this thesis is to not only apply filters on the 5 minute data, but also on the hourly data directly. For QC de Vos, this means that for the hourly filtering, the QC's parameters are adjusted accordingly. Time dependent parameters are divided or multiplied by 12, as 12x5 minutes equals 60 minutes.

To avoid confusion on the accumulation, these QCs are names QC de Vos 5 and QC de Vos 60. When accumulating the QC de Vos 5 results to hourly resolution, the QC is named QC de Vos 5R.

FZ filter changes

In the FZ filter, an observation was flagged as a false zero, when the previous interval also recorded a false zero. In attempt to improve the FZ filter, this step was changed in DMR5-1, DMR5-3 and DMR60-1. Instead of only looking at the previous interval, the past 6 intervals were evaluated. If at least one of these flagged a false zero, the current interval would also be flagged. These 6 intervals were grouped under new parameter FZ_{int}.

For DMR5-1, the n_{int} parameter was lowered to 3. For version DMR5-2 and DMR5-3, it was lowered to 2, thus only evaluating the past 10-15 minutes instead of half an hour, when comparing observations of neighbor stations.

HI filter changes

For the HI filter, only ϕ_A and ϕ_B parameters were changed, to allow for a stricter filtering of high influxes. This was done because the initial threshold of 10 mm for ϕ_B seemed very high: 10 mm per 5 minutes equals 120 mm per hour.

SO filter changes

The original threshold for station outliers, determined by γ, was set very low at 0.15. For the 5 minute QCs, the parameter was doubled to 0.3. For the 1 hour QCs, the parameter was set to 0.5

BC filter changes

In QC de Vos, the dynamic bias update is designed to account for stations developing bias over time. The period of time that is evaluated for this bias correction, is quite short, as it is dependent on the parameters m_{int}, m_{rain} and m_{match}.

In all of the QCs DMR, the period evaluated is no longer depending on the intervals with rain, but on a constant period of 2 months, which is equal to the warm-up period. If the absolute bias is higher than 5 percent, it is updated by a 5 percent correction.

The change is made in attempt to stabilize the changes that are made to the bias correction factor, so that it is not influenced by short-term events, but only corrects for long-term bias.

3.2.5 QC DMR addition: constant mechanical bias correction

Where previously described QC variations only include changes to filters and parameters, an additional variation to the QC is made by applying a constant mechanical bias correction (CMBC).

In contrast to the dynamic bias correction of QC de Vos, the mechanical correction is a constant factor that is applied to all stations results after filtering, in order to account for the standard overestimation that was found in the tipping volume experiment B (see section 3.1.7). The data is divided by the CMBC, which is set to 1.1. This CMBC is different to the DBC parameter, as it is independent from the reference data.

Table 1: *Parameters of different quality control methods*

params	Vos et al. (2019)	de Vos	DMR5-1	DMR5-2	DMR5-3	de Vos 60	DMR60-1	DM60-1
d (m)	10,000	5,000	5,000	5,000	5,000	5,000	5,000	5,000
n_{stat}	5	5	5	5	5	5	5	5
n_{int}	6	6	3	2	2	0	0	0
ϕ_A (mm)	0.4	0.4	0.508	0.508	0.508	0.4*12	2.54	0.254*6
ϕ_B (mm)	10	10	2.54	2.54	2.54	10*12	2.54*5	2* 2.54
m_{int}	4,032	4,032	4,032	4,032	4,032	4,032/12	4,032/12	4,032/12
m_{rain}	100	100	100	100	100/12	100/12	100/12	100/12
m_{match}	200	200	200	200	200/12	200/12	200/12	200/12
γ	0.15	0.15	0.3	0.3	0.3	0.3	0.5	0.5
β	0.2	0.2	-	-	-	0.2	-	-
DBC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
FZ_{int}	-	-	6	6	6		6	6
CMBC	-	-	1.1	1.1	1.1	-	1.1	1.1

3.3 DMR participants

The experiment involving the weather stations in the Green Village, as described in section 3.1, revealed that there are quite some steps and challenges to overcome when installing the stations. The main issue was the connection between the local unit and the data platform WOW. In the years 2021 and 2022, participants of the DMR project had already reported struggles with this same issue. With the expansion of the project in summer 2023, around 20 new participants received a weather station. Half of them were unable to connect their station to the internet, even with an updated manual and an additional live instruction. Additionally, some of the stations of participants from previous years had stopped sending data.

Subsequently, participants were visited at home to resolve the connectivity issues. During these visits, the reasons for the problems were identified. Moreover, participants from the 2021/2022 years were sometimes unaware that their stations had stopped sending data. A separate question section was added to the survey carried out by the DMR project management, asking participants about the installation and placement of the weather station in their garden. The section contained five closed questions on how often participants check and maintain the outside unit of their weather station, and another three closed questions about the frequency with which they check the data on the inside unit. For both the inside and outside unit, two follow-up questions were asked to discover if participants would check their weather station at particular occasions, and to give room for additional comments. The questions of the survey can be found in A.4.

The insights of visits and conversation with the participants, as well as the results of the survey, including photos of stations, will be presented in section 4.3. The insights gained through these visits can serve as recommendations for the future management of the project and future projects.

4 Results

This chapter presents the results of the research as described in the methodology. First, section 4.1 presents the results from the Green Village setup and experiments. Second, section 4.2 presents the results of the different quality control methods applied on both the 22-23 DMR data and the 23 DMR data. Third, section 4.3 presents the results of the survey and home visits to participants.

4.1 The Green Village: experimental setup

4.1.1 Tipping volume experiments

Figure 16 shows the results of tipping experiment A. The values represent the average tip volume per reported observation by each station. The boxplots summarize the reported values for the same station, except for station 6, which was excluded from the experiment.

For most stations, the median is closes to the assumed design volume of 0.254 mm, or 0.01 inch, which is shown as a horizontal line. The highest outlier values correspond to 0.3048 mm, or 0.012 inch, and the lowest to 0.2032 inch, or 0.008 inch.

From correspondence with the manufacturer of the Alecto station, it has been confirmed that the tipping bucket design volume is indeed 0.254 mm. The company also confirmed the tipping bucket works with a reed switch, meaning that one tip must represent a single value. This indicates that the variance in reported tipping volume occurs either within the software of the Alecto, or the data management of the Weather Observations Website.

A table with all results from the experiment can be found in B.1.

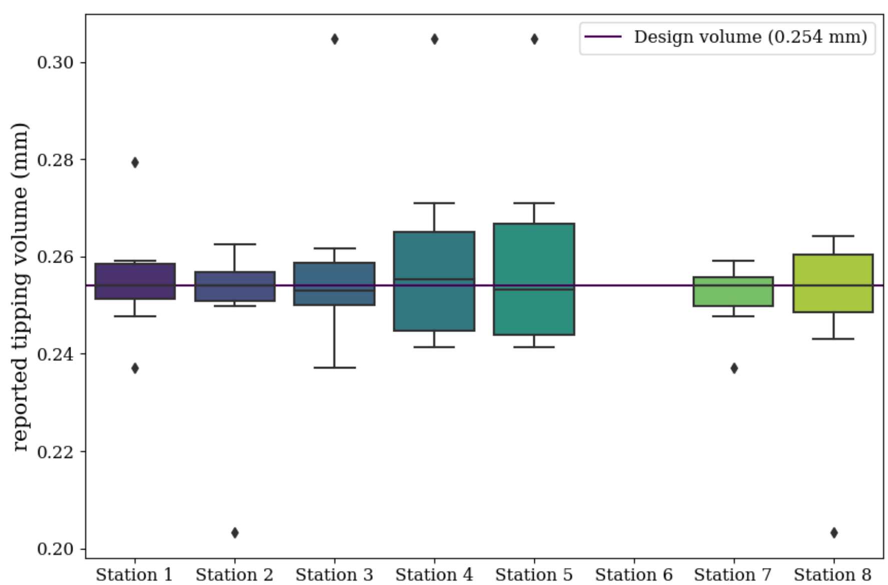


Figure 16: *Results of tipping volume experiment A. The result is shown for seven different stations, leaving station 6 untested. The box plots are based on 9-10 values per stations.The horizontal line represents the designed bucket volume of 0.254 mm, or 0.01 inch.*

Figure 17 shows the results of tipping volume experiment B. The bar graphs represent the mode of 9 or 5 values, for the categories 'unemptied bucket' and 'emptied bucket', per station. The horizontal line equals the calibrated tipping volume of 0.254 mm.

Station 1 has the highest tipping volume for the emptied bucket, followed by station 4, while station 3 has the lowest tipping volume. Station 1 also has the highest tipping volume for the unemptied bucket, followed by station 2 and 5. For all stations, the tipping volume of the unemptied bucket is higher than the tipping volume of the emptied bucket, with the exception of station 3, where both volumes are equal.

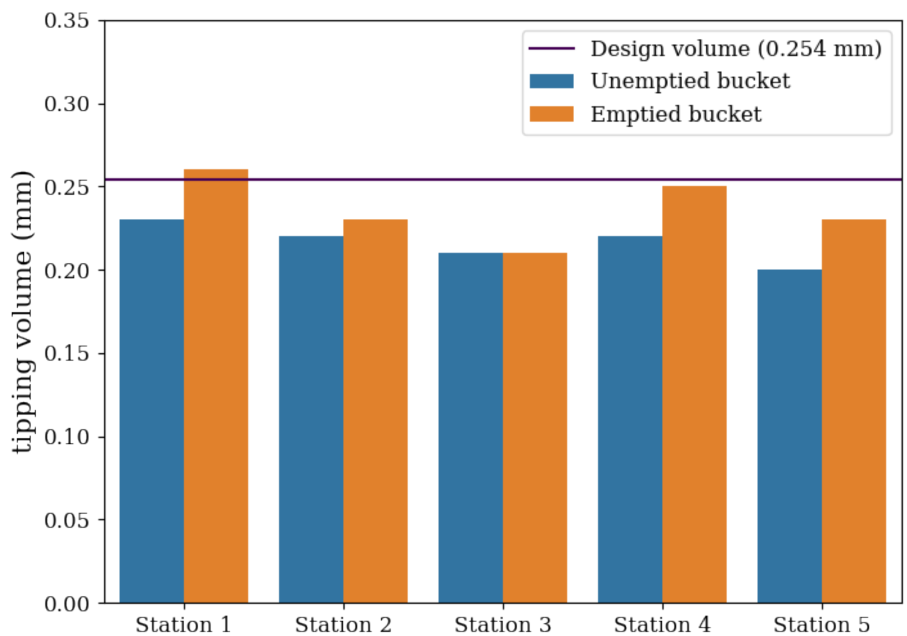


Figure 17: *Results of tipping volume experiment. The result is shown for 5 different stations, tested with unemptied and emptied buckets. The bar graphs show the mode of the experiment, that was carried out 5 times for each station.*

The measurements for the unemptied bucket and emptied bucket are different. This is because a small volume of water remains in the bucket after tipping. For the 'emptied bucket' volume, the bucket was explicitly cleaned in between the experiments. It is assumed that the 'unemptied bucket' volume is the most representative. It is unlikely that a large portion of the water remaining in the tipping bucket will evaporate, even in summer months, as it is shielded by the funnel.

When comparing stations 1, 2 and 3, the measured tipping volume of station 1 is the highest, followed by 2. Station 3 has the lowest tipping volume. This can be explained by the stations' setup. Even though station 1 and station 2 were both designed to be perfectly level, station 2 is slightly off-level, as previously seen in the methodology (figure 7). Station 3 is obviously off level, which can be observed in figure 8. Both station 2 and station 3 are tilted in the 'emptying direction' of the tipping bucket, which is why their tipping bucket holds less water before tipping.

For station 3, the tipping volume of the emptied and unemptied bucket are the same, because the station is tilted to such an extent that the bucket completely empties when tipping.

The results of station 4 are comparable to those of station 2, which makes sense as station 4 is also tilted, but in a more sideways direction. The results of station 5 are hardest to explain, as it is set up completely level. Hence, similar results to those of station 1 would be expected, but station 5 reports the lowest tipping volume. The observation that it tips at a lower volume, might be due to a manufacturing defect or deviation, or because the tipping mechanism within the outer shell of the station was damaged.

Based on the results it was decided that station 1 was the most representative for a perfect set up. From the two tested volumes, the unemptied bucket is deemed most representative of the real tipping volume. Due to imperfect emptying of the tipping bucket, the station reports 0.254 mm, while the real tipping volume is 0.23 mm. This leads to an overestimation of rainfall by approximately 10 percent. The CMBC was thus set to 1.1.

A table with all results can be found in B.1.

4.1.2 Comparison of statistics for Alecto stations

Table 2 shows the results for the comparison of station 1 to stations 2 and 3 in the period of July 1st to September 6th of 2023. In this period, the accumulated rainfall was observed to be around 270 mm. Both station 2 and station 3 report a slightly higher amount of rainfall compared to station 1, resulting in a positive bias of approximately 3 percent. The results for filtered and unfiltered data are very similar. With an r of 0.99, the correlation between the stations is high for all scenarios.

From the comparison of hourly rainfall data between stations 1, 2 and 3, in July-August, and almost perfect correlation, it can be concluded that the differences between the stations are small. The reason for the observed slightly higher recorded precipitation of stations 2 and 3 could be the earlier mentioned small differences in levelness of the stations.

Table 2: *Results of the comparison of the Alectos 2 and 3 with Alecto 1 on all statistics, with hourly unfiltered and filtered data in the period of July 1st to September 6th 2023.*

station	R (mm)	R_{ref} (mm)	ΔR (mm)	r. bias	CV	r	n	RMSE
	unfiltered, tip-corrected							
2	267	272	5	0.018	0.21	0.99	178	0.32
3	267	275	8	0.030	0.18	0.99	179	0.27
	FZ, HI, SO filtered							
2	264	269	5	0.019	0.19	0.99	162	0.32
3	266	273	7	0.027	0.17	0.99	170	0.26

Figure 18, as well as figure 19 show the difference between the comparisons of stations 1 and 2 for 5 min, 15 min and 1 h data. The correlations for the shorter time intervals are lower, especially for 5 minute data. This is probably partly caused by sampling error, as explained in 2.1.4. Another source of error for the Alecto and WOW data, is that observations for different stations are not recorded at the same time. For example, if station 1 records a full tip at 16:04:55, but station 2 records it at 16:05:12, the rainfall for station 1 and 2 are attributed to timestamps 16:05:00 and 16:10:00, respectively. This "resampling" error considerably lowers the correlation at 5 minute resolution.

As the hourly data shows the highest correlation, it was chosen to make most comparisons for the Green Village with the hourly data.

(a) *5 min, raw data* **(b)** *5 min, tip-corrected*

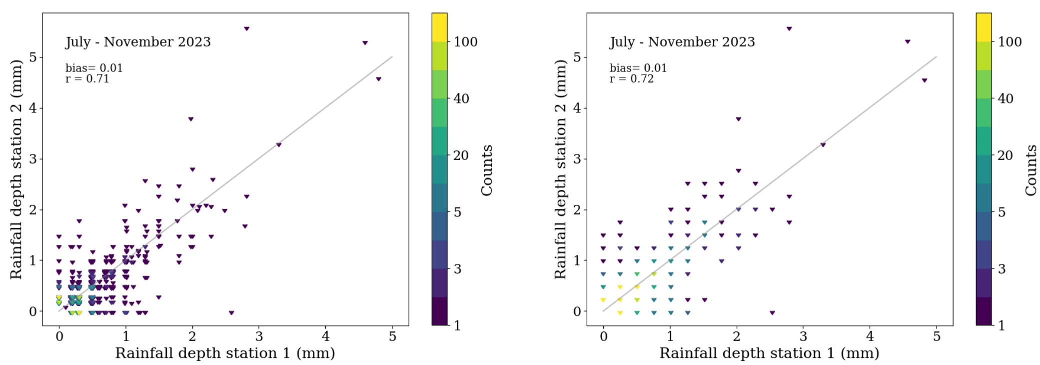


Figure 18: *Scatter density plots of unfiltered Alecto 2 rainfall measurements against the Alecto 1 rainfall measurements. The two graphs represent the 5 minute accumulation intervals, for raw and tip-corrected data respectively.*

Figure 18 shows the difference between the raw data from WOW, and the tip-corrected data. Tipping experiment A and correspondence to the Alecto company showed that the true tipping value is 0.254 mm, so tip-corrected means that all measurements are corrected to 0.254 mm or nearest multiple of 0.254 mm. The data for following analyses were also tip-corrected.

(a) *15 min, unfiltered* **(b)** *1 h, filtered*

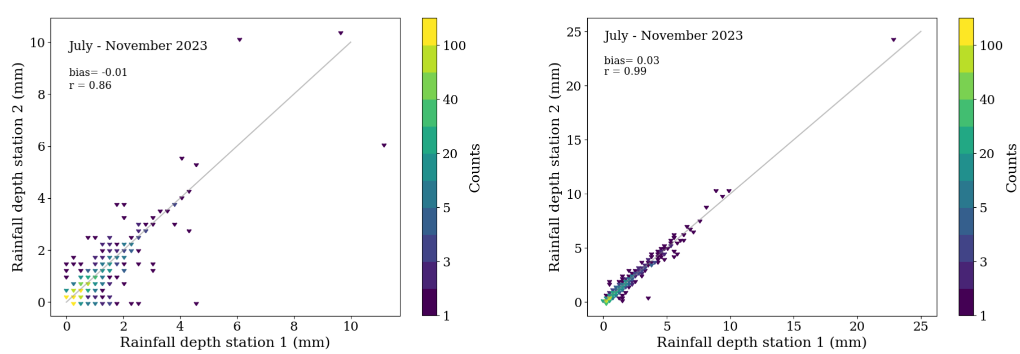


Figure 19: *Scatter density plots of Alecto 2 rainfall measurements against the Alecto 1 rainfall measurements.*

Figure 20 shows the results for the comparison of station 1 with stations 2 to 8, for the period of July 1st to November 30th of 2023. In this period, the accumulated rainfall for station 1 was observed to be around 800 mm. For station 3, as the comparison is started from September 8th due to a change in set up.

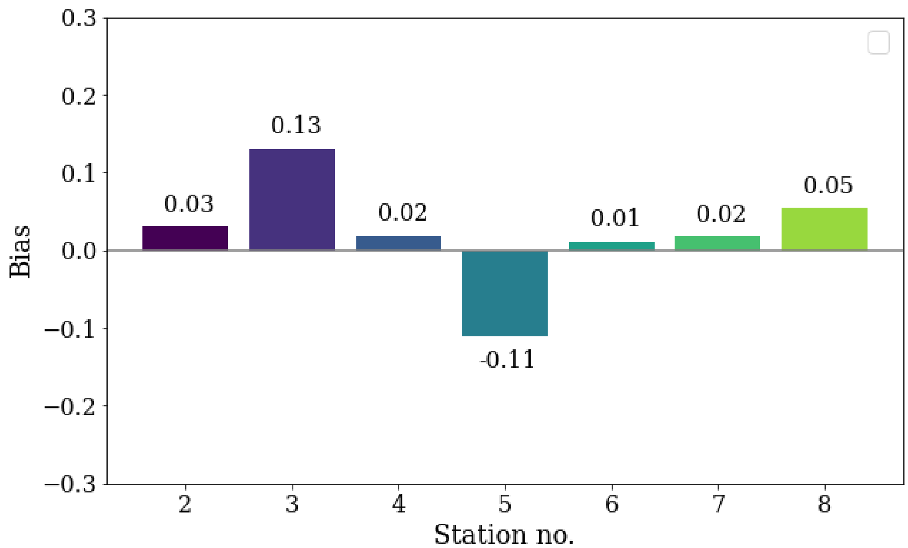


Figure 20: *Result of the comparison of the Alectos 2 to 8 with Alecto 1 as reference, with filtered hourly data in the period of July 1st to November 30th 2023. For station 3, the comparison is performed for the period of September 8th to November 30th.*

The most notable differences are found for stations 3 (tilted) and 5 (wall), with absolute biases of over ten percent and lower correlations and lower CVs as compared to the other stations. Station 3 shows a high overestimation, while station 5 exhibits a high underestimation. Station 8 reports a slightly higher amount of rainfall compared to station 1, while the rainfall sums for stations 2, 4, 6 and 7 are similar to that of station 1. For most stations, the results for filtered and unfiltered data are very similar, except for station 2, where filtering improves both r and CV, while the bias changes from negative to positive.

A table with all statistics can be found in B.1, table 8.

Figure 21 shows the double mass curves of stations 3 and 5. The result for station 3 is as expected, as the station was tilted in the direction of tipping, which leads to clear overestimation at every measurement. Although both stations have a large relative bias compared to station 1, the correlation of station 5 is much lower. The curve is less straight than that of station 3, and appears to have a gap, in an event where station 1 observed a high amount of rainfall in one hour.

The variations in the double mass plots of station 5 are more difficult to interpret. When assuming that the missing data around 400 accumulated mm can be explained by a faulty zero, it is observed that the accumulation between 100 and 500 mm is actually highly correlated to the accumulation of station 1. In the first 100 mm, there is an underestimation of rainfall, and from 500 mm, this is the case again.

The first change in underestimation might be explained by the surroundings of the station, which is visible in figure 22. The lavender plant was in full bloom in June and July, hanging over the station and possibly blocking precipitation directly. After this period, the plant shrunk and was less of an obstacle. Whether or not this is true, it is still unclear why a second change occurred at the end of the time series.

(a) *Station 3* **(b)** *Station 5*

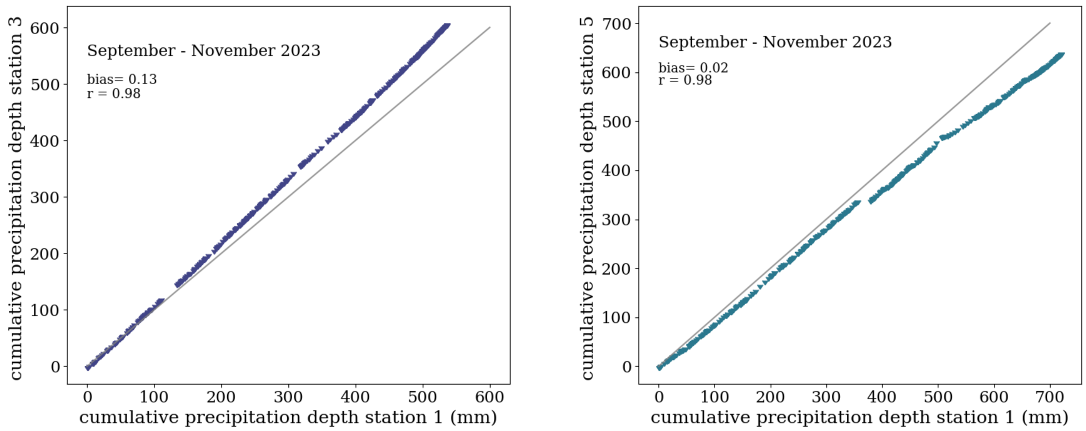


Figure 21: *Double mass plots of filtered rainfall measurements from station 3 and 5, and station 1. The 1 h accumulation data was used.*

Overall, due to the station being so close to the wall and practically squeezed in between two buildings, a higher underestimation would be expected. When reconsidering the outcomes of tipping volume experiment B, station 5 had a lower tipping volume than expected, meaning it would overestimate more than station 1 under similar circumstances. This means the expected effect of the wall might be compensated by the deviation in tipping volume caused by instrument deviation.

(a) *June* **(b)** *End of August*

Figure 22: *Station 5 and surrounding lavender plant during different stages of growth.*

4.1.3 Comparison of Alecto 1 with reference data

Figures 23 and 24 show the results for the comparison of station 1 with references: the KNMI gauge (Gauge), the KNMI radar (Radar), the local Micro Rain radar (MRR) on the Ruisdael site and the Davis Vantage Pro 2 on the office building. Comparisons are made filtered and bias corrected data, for the period of July 1st to November 30th of 2023. The comparison with both radar data sets and the Davis VP2 is done for 60 minute data. The data of station 1 was accumulated to daily data for comparison with the KNMI gauge.

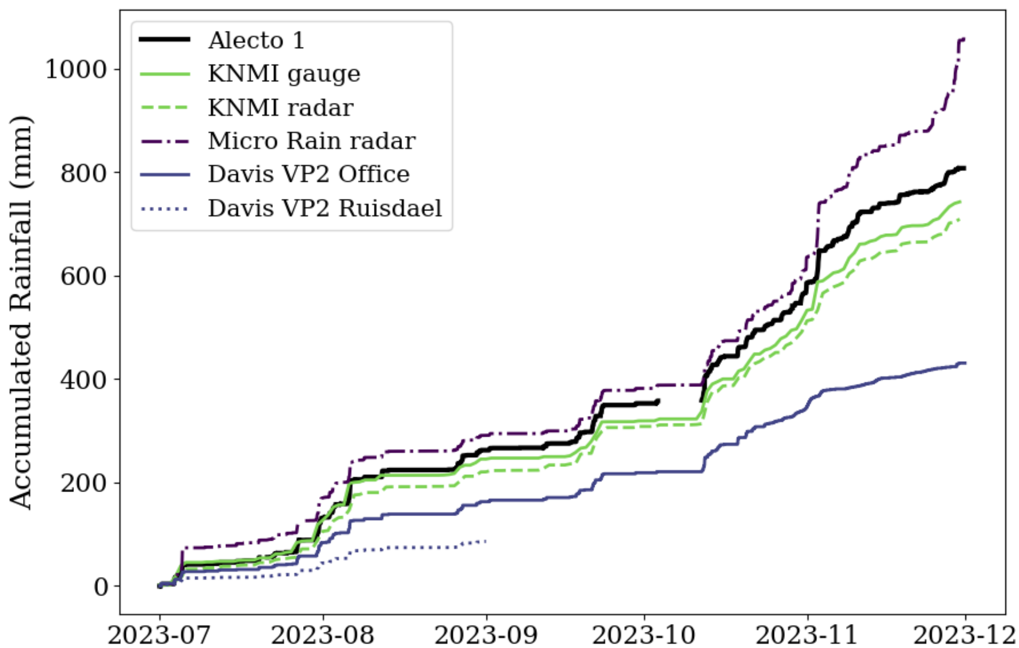


Figure 23: *Precipitation accumulation for the Alecto 1 and five references for the period of July 1st to November 30th 2023.*

Figure 24 shows that station 1's observations agree most with the gauge. Nonetheless, the overestimation of station 1 is still higher than 10 percent. Compared to the KNMI radar product, the bias is even higher. The Micro Rain Radar on the other hand, reports more rainfall than station 1, especially during large rainfall events, resulting in a negative bias of over 20 percent. The Davis VP2 reports almost half the amount of rainfall, resulting in a high bias and low correlation.

The static bias correction greatly improves the comparison for station 1 with both the KNMI gauge and the radar. Since the bias correction lowers the accumulated precipitation of station 1, the bias in comparison with MRR drops even lower.

A complete report on the statistics from the comparison to references can be found in table 24. The results between unfiltered and filtered data are not substantially different.

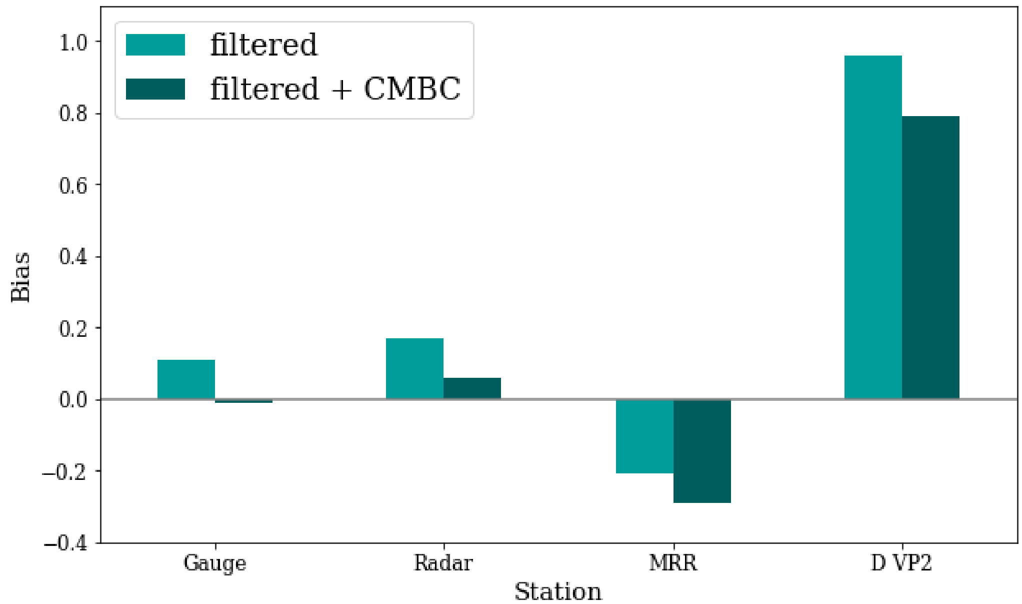


Figure 24: *Results of the comparison of the Alecto 1 to four different references, with filtered hourly data* in the period of July 1st to November 30th 2023. The four references are the KNMI gauge, the KNMI radar, the Micro Rain Radar and the Davis Vantage Pro 2. *For the KNMI gauge, daily data is used.*

Since both radar references have a fundamentally different way of measuring rain, it is not surprising that the Alecto measurements are more in line with the observations from the KNMI gauge. However, since the Davis Vantage Pro 2 is also a tipping bucket gauge, the large underestimation of the DVP2s is unexpected. The DVP2 on the Ruisdael site is partly obstructed by a tree, which might explain the large underestimation. The station was disregarded as reference due to the poor performance. The DVP2 on the office building performs better than its twin at Ruisdael, but the underestimation is still large. It was reported by TGV staff that the funnel was blocked by leaves, and was cleaned on the 29th of November. In figure 23, it is visible that the data for the month prior to this cleaning moment (end of October until end of November), the underestimation was larger than in the period of July til October.

4.2 Quality control of DMR 2021-2023

4.2.1 Validation metrics of DMR 22-23 QC methods

Figures 25, 26 and 27 demonstrate the functioning of the quality control method, with the FZ, HI and SO flags, and the BC filter. The effect of filtering the high influxes, outliers and faulty zeroes is clearly visible, although the horizontal lines in figure 26 show that filtering is not a hundred percent successful. The bias correction brings the measurements of most stations closer together, as is visible in figure 27.

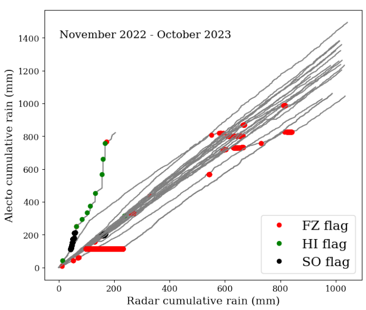


Figure 25: *Filtering of 5 min data with QC de Vos 5*

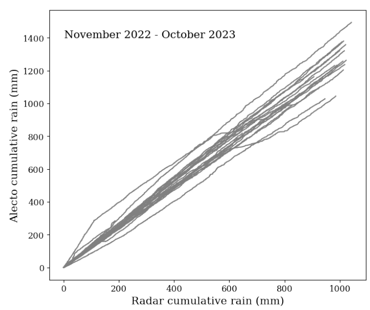


Figure 26: *Result of FZ, HI and SO filter.*

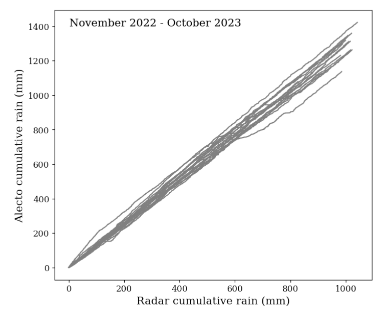


Figure 27: *Result with additional BC filter.*

Table 3 presents the results of the comparison of different quality control methods for the 2022-2023 dataset, which runs from November 1st 2022 to October 31st 2023.

When analyzing the outcomes for the 5 minute data (upper half of table 3), it is clear that for all QCs, the correlation, CV and RMSE improves compared to non-filtered data. The improvement in CV and correlation is slightly, but not substantially larger for DMR5-1 and DMR5-3. The percentage of filtered data is much higher. The type of applied filter does not lead to substantial change in the metrics.

When considering only the subset where the reference data is equal to or greater than 0.254 mm, or 1 tip (second row of table 3), a lower CV and lower correlation, and higher RMSE are obtained, because dry periods are excluded. The most striking difference is found in the bias, which has a positive value of 25 to 30 percent for the complete set, and is now found to be slightly negative with values between 5 and 10 percent.

The additional constant bias correction (CMBC) made in the DMR QCs decreases the bias with approximately 10 percent, for both the complete set and the subset. In the subset, this leads to a larger negative bias.

The large difference in biases between the entire data set and the sub can be explained by the influence of faulty zeros and high influxes. When considering the complete data set, faulty (high) influxes cause large overestimation, resulting in a high bias. When considering only the subset where the reference is equal or greater than 0.254 mm, the influence of the false influxes disappears. Instead, faulty zeros cause underestimation of rainfall, resulting in a lower bias.

The large bias shift indicates that both FZ and HI filters are not very successful. As it is hard to judge the effects of filtering by only the general validation metrics, the effects of the FZ and HI filters will be discussed separately in sections 4.2.2 and 4.2.3, respectively.

Table 3: *Validation metrics for 5 min and 1 h data running from November 1st 2022 to October 31st 2023, for different quality control methods and filter combinations. The references are the corresponding pixels of the KNMI gauge-adjusted radar. Full tables for different filter combinations are available in tables 11 and 10 in B.2.*

QC method	Applied filters	r.bias	CV	r	% filtered	RMSE	med. bias
5 min data	-	0.27	13.84	0.39	100	0.14	0.25
De Vos 5	FZ, HI, SO, BC	0.28	6.71	0.67	99.09	0.07	0.28
DMR5-1	FZ, HI, SO, BC	0.28	6.05	0.70	91.92	0.07	0.28
	all + CMBC	0.17	5.58	0.70	91.92	0.07	0.17
ref $\geq$ 0.254 mm	-	-0.09	0.91	0.57	100	0.45	-0.09
De Vos 5	FZ, HI, SO, BC	-0.06	0.95	0.57	98.36	0.47	-0.04
DMR5-1	FZ, HI, SO, BC	-0.07	0.88	0.58	95.10	0.43	-0.06
	all + CMBC	-0.15	0.81	0.58	95.10	0.43	-0.15
1h data	-	0.16	2.72	0.87	100	0.34	0.15
De Vos 60	FZ, HI, SO, BC	0.26	3.20	0.85	89.40	0.43	0.23
De Vos 5R	FZ, HI, SO, BC	0.18	2.23	0.91	99.09	0.28	0.19
DMR60-1	FZ, HI, SO, BC	0.20	1.97	0.91	76.56	0.31	0.22
	all + CMBC	0.09	1.81	0.91	76.56	0.28	0.11
DMR60-2	FZ, HI, SO, BC	0.19	2.08	0.91	90.21	0.28	0.20
	all + CMBC	0.08	1.92	0.91	90.21	0.26	0.09
ref $\geq$ 0.508	-	0.06	0.67	0.81	100	1.13	0.05
De Vos 60	FZ, HI, SO, BC	0.14	0.83	0.78	96.02	1.43	0.13
De Vos 5R	FZ, HI, SO, BC	0.09	0.58	0.86	98.46	0.98	0.08
DMR60-1	FZ, HI, SO, BC	0.08	0.57	0.85	95.49	0.97	0.10
	all + CMBC	-0.01	0.53	0.85	95.49	0.97	-0.00
DMR60-2	FZ, HI, SO, BC	0.08	0.56	0.85	96.11	0.96	0.08
	all + CMBC	-0.02	0.52	0.85	96.11	0.96	-0.02

When analyzing the outcomes for the hourly accumulated data (lower half of table 3), the improvement of validation metrics between unfiltered and filtered data is not as obvious as for the 5 minute data. For De Vos 5R, DMR60-1 and DMR60-2, the CV and RMSE lower, and correlation slightly higher. For de Vos 5R, almost no data is filtered, while 25 percent of data is filtered with DMR V1. The bias is larger when applying the BC filter, compared to the other filter combinations.

When considering only the subset where the reference data is equal to or greater than 0.508 mm, or 2 tips (bottom row of table 3), this results in a lower CV and lower correlation, and higher RMSE, similar to the 5 minute data. The bias is also lower compared to the complete subset, since the false influxes are not considered in the subset. The shift in bias is not as large as for the 5 min data, indicating a better functioning FZ filter. For all QC methods, a lower percentage data is filtered, with the largest difference for DMR60-1, where approximately 20 percent less data is filtered.

The additional correction made in the DMR QCs decreases the bias with approximately 10 percent, for both the complete set and the subset. In the subset, this leads to slightly negative biases.

4.2.2 Faulty zeros analysis

The performance of the FZ filter is individually analyzed for each QC method, by comparing the number of faulty zero measurements remaining after filtering. Faulty zero values are defined as observations where the Alectos stations did not report rainfall, but the reference did. Two subsets are added, for values where the reference was higher than 1 or 2 tips.

For the hourly data (figure 28, right), the FZ filter of DMR60-1 filters the most values across all subsets, followed by De Vos 60 and DMR60-2, which filter approximately 10 percent less. These three methods filter more than half of the faulty zeros for the two subsets. De Vos 5R filters a low number of values, and none in the last subset.

For the 5 minute data (figure 28, left), both DMR5-1 and DMR5-3 filter the most data, with 10 percent for the entire set and 25 percent for the subset. De Vos and DMR5-2 filter a low number of faulty zeros for both sets. Compared to the hourly data, the percentages of values filtered are considerably lower for all filtering methods.

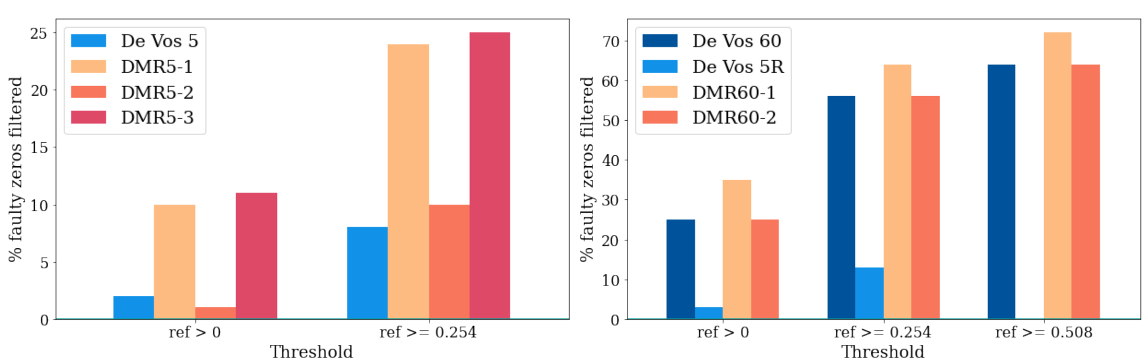


Figure 28: *Remaining faulty zero values after FZ filters from different quality control methods, for both 5 min and 1 h data of the 2022-2023 data set. Different subsets for reference threshold values are evaluated.*

Overall, DMR60-1, DMR5-1 and DMR5-3 are the best at filtering faulty zeros, at the cost of filtering out much more data in general, as seen in table 3.

That being said, none of the 5 min QCs is capable of filtering the majority of faulty zeros. This is in line with earlier speculations of poor FZ filtering in section 4.2.1. For the 5 min data, the correlation remains quite low, due to sampling and resampling errors that were also observed during the TGV experimental setup. Since the faulty zero filter relies on comparison with other stations at the same intervals, the sampling errors might be one of the causes that the filters are not working well.

For the same reason, for the hourly data, the faulty zeros for each hourly based QC method work better than the de Vos 5R, especially for the subset where the reference observations are larger.

Table 13 in B.2 shows the complete result of the FZ analysis.

4.2.3 High influx analysis

To inspect differences in high influx filtering, a visual comparison for QC de Vos 60 and QC DMR60-2 is made. Figures 29 and 30 show the scatterplots from, respectively, QC de Vos 60 filtered, and QC DMR60-2 filtered hourly data.

The high influxes are visible on the left side of the figure, where radar rainfall depth is zero or close to zero, and the station rainfall depth is high. For QC DMR60-2, these values are filtered. When zooming in (red frame in figure 29), a vertical line is still visible in the bottom left corner, where radar rainfall depth is zero. The counts for station rainfall between 1 and approximately 5 mm are between 10 and 100, indicating that a lot of false rainfall measurements remain unfiltered. Although the HI filters have improved with the changed parameters in the DMR QCs, many faulty false influxes remain unflagged.

This limited success of the HI filter was already noted in the general metrics for the full data set in section 4.2.1. Lowering the threshold parameters ϕ_A and ϕ_B could help to reduce these faulty influxes. However, lowering the parameters also increases the risk of flagging of true rainfall that goes undetected by the radar.

Figures for the 5 minute data can be found in B.2.

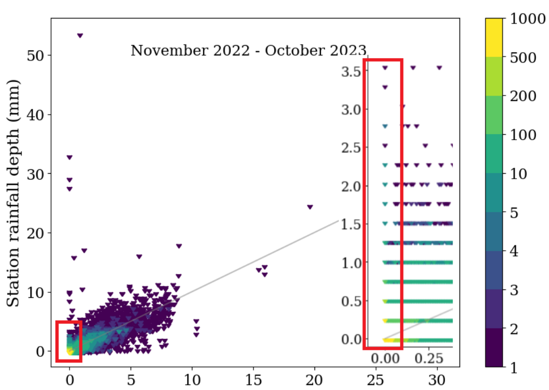


Figure 29: *HI filter QC de Vos 60.*

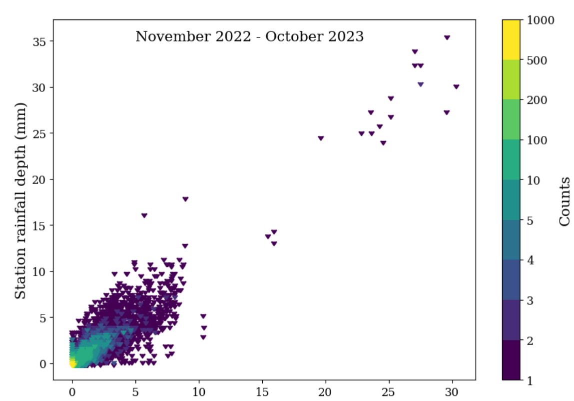


Figure 30: *HI filter QC DMR60-2.*

4.2.4 Bias correction analysis

To obtain information beyond the general validation metrics, the bias correction filter is also analyzed separately. Figure 31 provides more insight in the working of the bias filter, as the relative bias for all stations is displayed instead of only the median bias for all stations.

When comparing the two figures with box plots, it can be observed that the median of the bias increases after applying the bias correction, and this effect is stronger for QC de Vos 60. This was already visible in the table 3. However, the box plot provides extra information on the variation in bias for all stations. For both methods, the variation of in bias decreases after filtering, but this effect is stronger for QC DMR60-2.

(a) *QC de Vos60* **(b)** *QC DMR60-2*

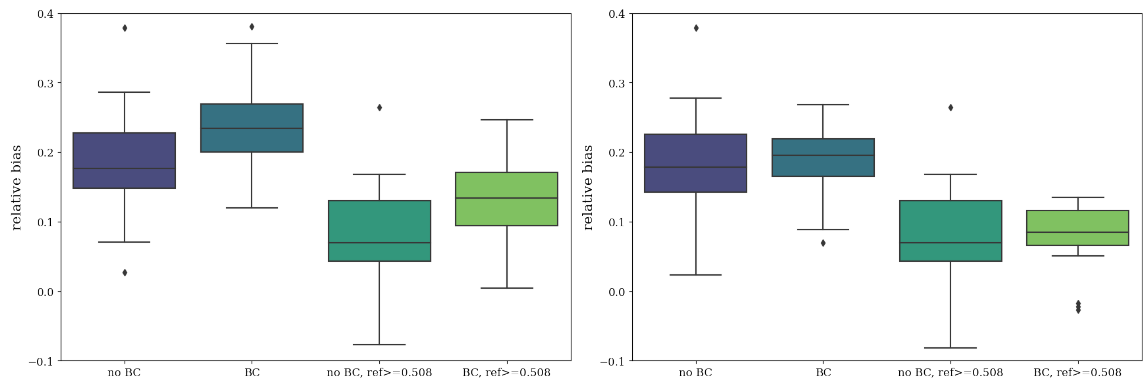


Figure 31: *Box plots of the relative bias before and after applying a bias correction filter, for QC de Vos 60 and QC DMR60-2, respectively. Subsets where ref ≥ 0.508 mm are shown for both QCs.*

The DMR60-2 is thus better at reducing large differences in bias between stations when compared to QC de Vos 60. This updated method of determining the bias correction factor (BCF) provides more stable bias correction, and thus more stable rainfall time series. An example of a time series showing the BCF difference for one station between QC de Vos 60 and DMR60-2 is given in figure 32.

(a) *QC de Vos60* **(b)** *QC DMR60-2*

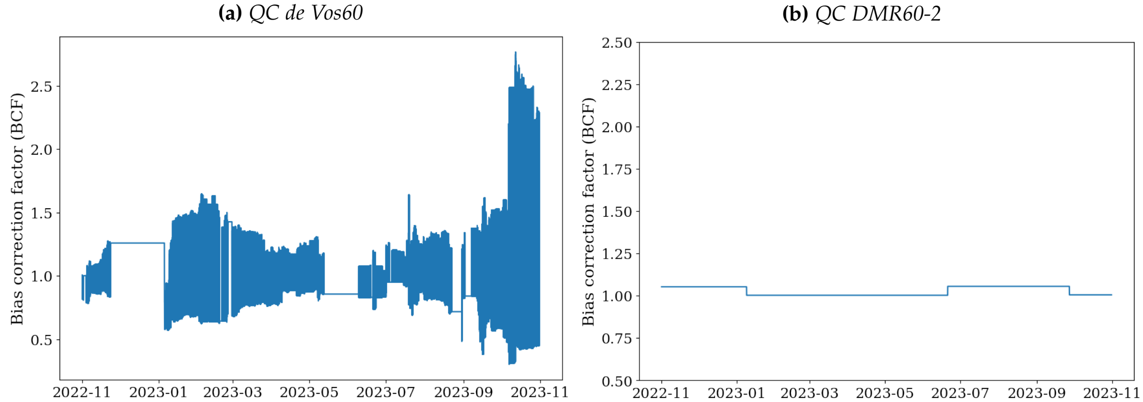


Figure 32: *Time series of the dynamic bias correction factor (BCF) for both filtering methods.*

4.2.5 DMR 23

Table 4 presents the validation metrics of two different quality control methods for the 23 datasets, which run from September 1st 2023 to November 30th 2023. The differences between the De Vos 60 and DMR60-2 QCs are comparable to those described for the 22-23 dataset. Overall, the metrics for the 23 set are somewhat better than those of the 22-23 set, considering a lower CV and higher correlation. The bias is also lower than for the 22-23 data.

Table 4: *Validation metrics for 1 h data running from September 1st 2023 to November 30th 2023, for different quality control methods and filter combinations. The references are the corresponding pixels of the KNMI gauge-adjusted radar.*

QC method	Applied filters	r.bias	CV	r	% filtered	RMSE	med. bias
	-	0.09	1.73	0.92	100	0.40	0.11
De Vos 60	FZ, HI, SO, BC	0.17	2.08	0.89	89.85	0.51	0.15
DMR60-2	FZ, HI, SO, BC	0.13	1.56	0.93	91.12	0.38	0.12
	all + CMBC	0.02	1.43	0.93	91.12	0.35	0.02
ref $\geq$ 0.508	-	0.03	0.59	0.89	100	1.10	0.04
De Vos	FZ, HI, SO, BC	0.09	0.74	0.83	96.93	1.40	0.08
DMR V2	FZ, HI, SO, BC	0.05	0.55	0.90	96.86	1.04	0.06
	all + CMBC	-0.04	0.50	0.90	96.86	0.95	-0.04

The faulty zero analysis in figure 33 shows that the percentage of faulty zeros filtered is higher than in the DMR 22-23 set, with over 80 percent of faulty zeros filtered for the subset with threshold 0.508. This might partly explain the lower bias for the 22-23 data. The results of the FZ filter of De Vos 60 and DMR60-2 are the same, which is expected as both methods are almost identical.

The better performance of the FZ filter, as compared to the 22-23 data, might be explained by the higher number of stations in the 23 set, supplying more neighbor stations for comparison and thus increasing the n_{stat} parameter, which makes the comparison to the neighbors more reliable.

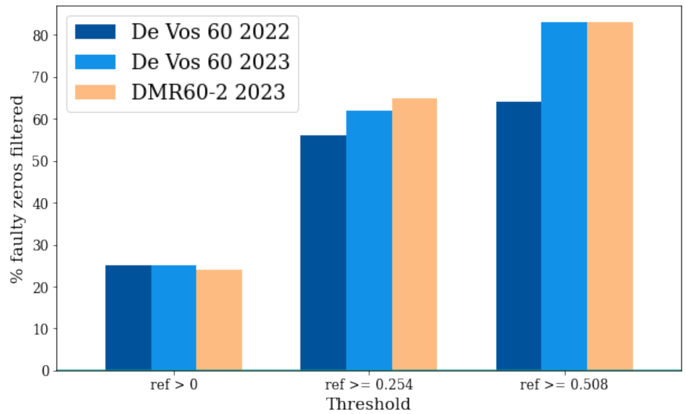


Figure 33: *Remaining faulty zero values after FZ filters from different quality control methods, for 1 h data of the 2022-2023 data set. Different subsets for reference values are evaluated.*

4.2.6 Subsets of installation years

To further analyze the lower bias that was found for the DMR 23 data validation, subsets are made, where stations are divided according to their year of installation.

For the 23 DMR data, this results in three subsets of stations installed in 2021, 2022, and 2023 respectively. In the left of figure 34, it can be observed that the newly installed stations have the lowest bias. For the same period in the year before, the analysis is done for the subsets of stations installed in 2021 and 2022. Here, the newly installed stations also had a slightly lower bias, but the biases were generally higher compared to the year after.

It is unclear why the bias for the stations installed in 2023 is lower. The boxplots with 2023 data (figure 34a) suggest some trend of sensor drift, however, when considering the subsets of stations the previous year (figure 34b), it seems that the median bias for the subsets of 2021 and 2022 was already high.

Similar to the non-bias corrected results of the 22-23 data, it is clear that the variation in bias is high for all station subsets, with a range of more than 20 percent.

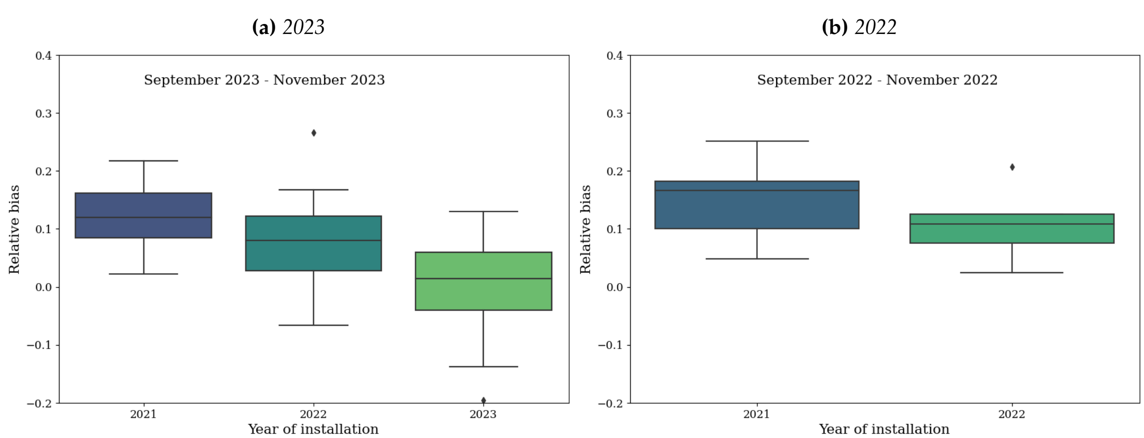


Figure 34: *Box plots of the relative bias from FZ, HI filtered data for different subsets according to the year of the installation.*

4.3 DMR participants survey and home visits

4.3.1 Participants Survey

A total of 15 participants who received a PWS in 2021 or 2022 responded to the survey. Figure 35 shows the results from the questions on the outdoor unit maintenance. It shows that with 11 out of 15 people, the majority checks whether the station is still level (and not tilted) with a frequency of at least once a month. Cleaning the entire unit was only done by the majority of participants every 6 months or less often.

In the additional questions, 8 participants reported to check the weather stations with a set frequency, while 6 participants were triggered by extreme weather conditions to check their station. Two participants said they almost never check, and three people said the reminder message from the DMR project management is a reason to check the weather station. Several participants made comments on leaves, bird and spiderwebs blocking both the funnel and inside mechanism.

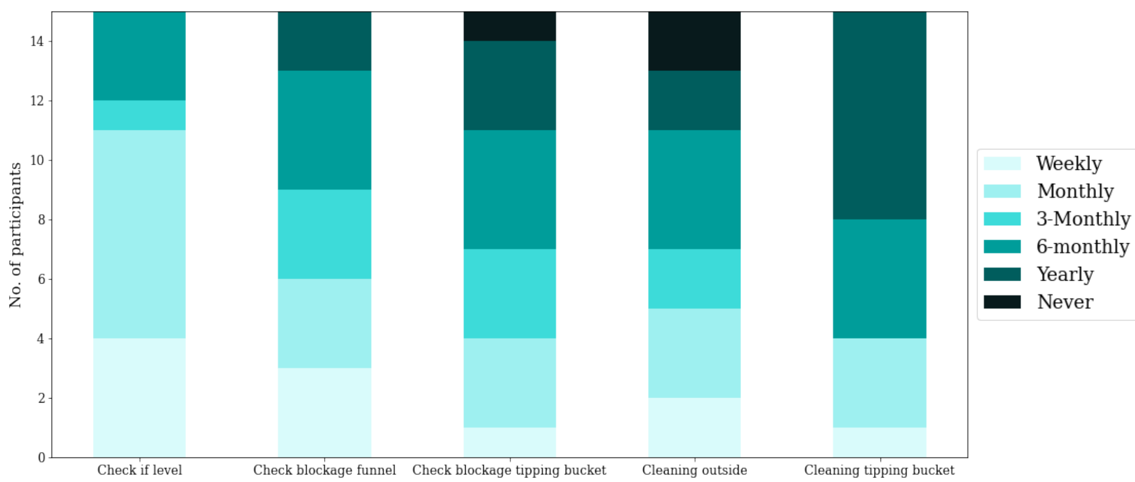


Figure 35: *Results from questions on the monitoring and maintenence of the outdoor unit of the Alecto WS-5500.*

Figure 36 shows the result from the question on the inside unit monitoring. All participants check their local inside unit at least once a month, with 80% monitoring the daily weather stats. On the other hand, most participants check the WOW website of their own unit, and of neighbor stations only once per three months or less.

In the follow-up questions, 5 respondents said they checked their connection to WOW after a reminder by DMR, and 6 said to be especially interested in the values after extreme weather. Some people emphasized their daily interest in the weather in the comments, one participant found this "fascinating". Other participants mentioned to check after re-installing the Wi-Fi router, or to check whether the back-up batteries were still working.

The answers from the questions on the outdoor maintenance of the station are not surprising, with the least complex task is being performed most often, and vice versa. It is easy to notice whether a station is still level, and most time-consuming to clean the entire unit. The other tasks mentioned in the survey lay in between these extremes in terms of task-complexity. The results show that the frequency with which these tasks were carried out, differ between participants, with the majority answering ‘monthly’ to ‘6-monthly’.

The same is true for the questions on the monitoring on the indoor unit of the Alecto. The inside unit is easy to check, and has quite a visible place in most households. The participants of

DMR are also likely to be enthusiastic about the weather. The less frequent checking of the WOW website, or neighboring stations could be explained by the higher complexity of the task, and the lower interest of participants.

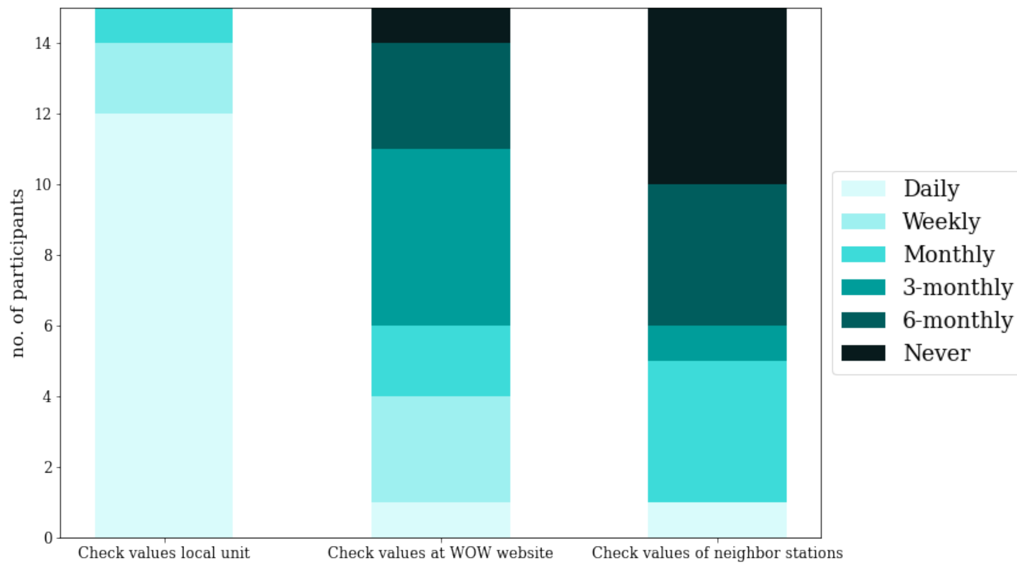


Figure 36: *Results from questions on the monitoring of the indoor unit of the Alecto WS-5500.*

4.3.2 Home visits to participants

Home visits were made to participants of the DMR project with connectivity issues. With more than 10 participants, around half of the new participants were not able to independently connect the Alecto WS5500 to the WOW data platform. Additionally, some volunteers who signed up in previous years, had lost the connection, or had never been able to complete installation.

Some participants were not aware of the lack of connection, but were reminded by the DMR project management, for example because they only checked their local unit and not the website.

With assistance of the project management, it was possible to establish the connection. The system is sensitive to human errors as there are many different steps to complete with typing of codes and passwords. This is especially difficult for participants with little knowledge of technical applications, such as computers or smartphones. Moreover, the mobile app "WS View" did not indicate which step in the process was faulty, which has been discussed in the methodology section . A new app, "WS View Plus", released in the fall of 2023, has made the process easier.

4.3.3 Participants and data quality

The results from the DMR data set show a high variability in station bias between different stations. From the station comparisons at TGV it is clear that faulty setups can lead to substantial biases. The TGV results showed that stations would normally overestimate data, while the DMR dataset contains a number of stations with negative bias.

From station photos made by participants (figures 37, 38, 39), it becomes clear that some of these negative biases can be attributed to bushes and walls that partially block the rain. However, when inspecting all the photos from the participants, it was not possible to find a clear relationship between the obstacle distance or obstacle height and the bias. From the photos it cannot be judged whether the station is perfectly level, while the TGV results suggest that tilting of the station has a large influence on the result. Furthermore, the sample size is small, as not all participants submitted photos.

Figure 37: *station 17, bias = -0.07*

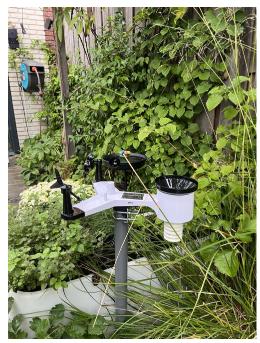


Figure 38: *station 28, bias = -0.04*

Figure 39: *station 34, bias = -0.20*

No clear relation between the maintenance and monitoring regime of the participants and the bias of the stations is not found either. For example figure 40 shows the station bias plotted against the answers for the question on how often the tipping bucket is cleaned. For the other questions, also no relation was found. As explained earlier, the sample size is very small and the impact setup differences may be too large. It could also be that the frequency of monitoring does not have an effect because the participants do not know what they should look for when checking the station.

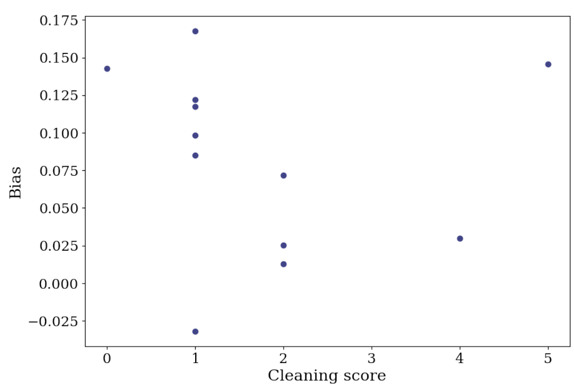


Figure 40: *Scatter plot of the station bias and tipping bucket cleaning score. The score represents the cleaning frequency from the survey answers, and ranges from 0 to 5, with 0 being 'Never' and 5 being 'Weekly'.*

5 Discussion and Outlook

This chapter discusses the results and how they agree or contrast with findings of other studies. The limitations of the methods are considered, and ideas for further research are proposed.

5.1 The Green Village: experimental setup

5.1.1 Quantifying setup differences

From the comparison of hourly rainfall data between the Alecto stations at the Green Village, it is observed that rainfall accumulations are very similar, with correlations of at least 0.98 and relative biases of below 5%. For stations 3 and 5, large biases were found, of 13% and -11%.

The overestimation for station 3, which was tilted in the direction of the emptying side of the tipping bucket, was expected and confirmed by tipping experiment B. As the correlation was still high with 0.98 for hourly data, this suggests that the bias of a tilted station is caused by a constant over- or underestimation proportional to the severity of the tilting. Subsequently, the very slight off-level installation of station 2 and 4 may also explain the slight overestimation of these stations, which was also observed in tipping experiment B. That slight tipping of the gauge leads to deviations in data, was also noted in a study on the performance DVP2, with a similar set up: "it is vital to ensure the rim is perfectly level, as this can lead to significant under- or overcatch" (Burt, 2009) Station 5 was placed in an alleyway, very close to walls. It was therefore expected that this station should underestimate, but the results could not satisfactorily be explained by the surroundings. A seasonal underestimating trend is suggested by the presence of nearby plants. The majority of the time series shows no clear underestimation, but the results of tipping experiment B suggest that the tipping bucket of station 5 has a higher overestimation tendency, even though it was perfectly level during testing. This instrumental deviation might have countered the effect of the nearby obstacles. In the last period of testing, a clear trend shift indicates that something happened that either shifted the station in its entirety, or caused a long term partial blocking to the tipping bucket. As the Green Village is a construction site, and receives many visitors, it is quite possible that someone caused a slight position shift. Changes in under- or overestimation trends were also found in studies by Bell et al. (2015) and Burt (2009).

The experiments clearly show that the exact set up and environment have a substantial impact on the measurements, but there is not enough data and experimental variation to establish a relationship between the severity of tilting and the over- or underestimation. The stations at Green Village could be used for a future study on this topic, where the exact tilting would be documented and related to the observed data. Another suggestion is to calibrate all instruments first to minimize the influence of instrument deviation,

5.1.2 Tipping volume experiments

The experiments to determine the tipping volume of stations 1 to 5 were carried out multiple times to obtain reliable results. The quality of the experiment is still limited, as the official calibration guidelines from WMO (2018b) were not followed. For example, there was for no weighing device to precisely determine the amount of water. The testing could not take place at the official testing facility of the KNMI, because the Alecto software was incompatible with the facility's software. Therefore, the simple experiments were in this case the best way to estimate the actual tipping volume.

It was decided that station 1 was the most representative station from the three stations at Ruisdael, as station 2 was slightly off level. Another option could have been to choose a median or average value from multiple stations, but as the biases of the "normal" stations were within three percent, this would not have made a large difference to the CMBC value, which was based on the overestimation measured by station 1 in tipping experiment B.

The use of the CMBC did improve results when comparing the Alecto 1 to the reference gauge and radar from the KNMI. However, apart from the non-ideal materials with which the experiment was performed, the experiment was also only performed for singular tips, representing low rainfall rates. A study by Humphrey et al. (1997) showed that TBRs tend to underestimate rainfall more with high rainfall rates. With rainfall rates up to 50 mm per hour, this may lead to an underestimation of 5 to 15 percent. When applying this knowledge to the Alecto, the CMBC might correct the data too much during convective rainfall events with high intensities.

5.1.3 Alecto vs. Netatmo

There is no previous research on the Alecto WS-5500, but the results from the setup at the Green Village can be compared to studies on other low cost PWSs. Comparison to Netatmo is relevant as their TBRs are used in other Dutch and European studies on the use of precipitation data from PWSs (Vos et al. (2019), Overeem et al. (2023), Nielsen et al. (2024)). The comparison of the results from TGV with De Vos et al. (2017) shows that both the Netatmo and Alecto stations have good correlation to nearby reference gauges and the KNMI radar product in a controlled setup, although the correlation decreases with higher resolutions. Interestingly, the Netatmo gauges were found to underestimate rainfall, while the Alectos at TGV were found to overestimate rainfall. This indicates that the Netatmo did not have issues with incomplete emptying of the tipping bucket, which may be explained by the fact that the Netatmo TBR design is two-sided, whereas the Alecto only has one bucket emptying on one side.

5.1.4 Reference stations: Davis VP2 and MRR

The Davis VP2s used in the experimental setup on the Green Village were initially intended as reference stations to the Alecto, but the results indicate that they suffer from severe underestimation, especially during large rainfall events. As mentioned earlier, this may partly be due to the reported blockage of the tipping bucket and surrounding obstacles (trees). However, both Bell et al. (2015) and Burt (2009) also reported unexplained variability of over 20% in rainfall accumulations in their reviews of the DVP2s, even though they were calibrated prior to their experiments. As the Davis is a more expensive instrument, it would be expected to have a higher quality.

The Micro Rain Radar was the other reference station on TGV. The MRR observed much more rainfall compared to the Alecto and the other references. The MRR determines drop size distributions from the Doppler spectrum, assuming zero vertical wind (Peters et al., 2002). As the vertical wind is present in reality, the calculated rain rates are biased, which is especially true for convective events with larger vertical wind (Kim & Lee, 2016). This may explain the results, where overestimation of the MRR with respect to other references was especially high during the more extreme precipitation events.

5.2 Quality control performance

For the quality control of the data, the method of Vos et al. (2019) was used and adapted. A study on comparisons of QC methods found QC De Vos "most useful where there is a dense PWS network" (El Hachem et al., 2023), which is true for the DMR data.

5.2.1 FZ filter

The structure of the new versions of the FZ, HI and SO filters was very similar to de Vos' methods. For the FZ filter, the performance of DMR60-2 and DMR5-2 were very similar to De Vos 60 and De Vos 5, respectively. The other DMR versions were better at filtering but also filtered a sizable part of data that was not faulty. The filters for hourly resolutions work much better than those at 5 minutes, due to the sampling errors and low correlation of the 5 min data. As the original method by Vos et al. (2019) only filtered on 5 min data, and accumulated results, an improvement is found by filtering faulty zeros on hourly accumulations. As the faulty zeroes mostly occur when a station is not functioning due to for example blockage, it is always for a longer period of time (e.g. hours, not minutes). Therefore, it also makes sense to filter faulty zeroes at a different time scale. This does not necessarily mean that the entire data set has to be filtered on the same time resolution. Faulty high influxes for example, are most likely to occur over the span of minutes, when people are cleaning. It would be interesting to experiment more with combining filters on different resolutions (e.g. 5, 15, 30, 60 min) in future research.

5.2.2 HI filter

For the HI filter and SO filters, not much was changed, except that parameters were somewhat stricter in the DMR versions of the filter. This filtered more high influxes. However, as noted in the results, faulty influxes are not necessarily high. Low faulty influxes are abundant, perhaps due to the cleaning and maintenance, or sprinklers, which do not cause high intensities and are subsequently not filtered. With stricter filtering this might be improved, but there would also be an increased risk of filtering true rainfall data. On the other hand, when disregarding "dry" data for rainfall analysis, by setting the reference to a minimum threshold (e.g. radar > 0.254 mm), the faulty influxes are filtered automatically.

5.2.3 Bias Correction

The bias correction from the QC de Vos was changed and expanded. From the experiments, the constant mechanical bias correction (CMBC) was introduced to reduce the overestimation of the Alectos. The original dynamic bias correction from De Vos was made less "dynamic". Earlier rain events were found to impact the bias correction a lot, making it highly variable. The updated version was only allowed to change the bias once every two months or less, depending on the rainfall data of at least two months previously. The logic behind this decision was that the other filters should be able to filter short term errors, and the bias correction is meant to correct systematic measurement deviations on the longer term. This stabilized the bias correction and also decreased the differences in bias among the different station. The chosen period of two months is of course somewhat arbitrary. A longer period provides perhaps more reliable results, with the disadvantage that a longer period of data prior to the analysis is needed.

5.2.4 QC Alecto vs QC Netatmo

When comparing the results from this study to those of Vos et al. (2019), the QC on the Alecto filters a lower amount of data than the QC on the Netatmo data, where less than 90% data remained. For the 5 min and hourly data, the correlations were higher for the Alecto filtered data, with an improvement of 0.10, and lower CVs. However, the initial quality of the Netatmo data was much worse. The unfiltered 5 min data for example had a correlation of 0.07 and a CV higher than 50. There were far more SO flags in the Netatmo QC, indicating more faulty stations. The fact that the raw data from the DMR project is much better might be explained by the fact that participants received basic instructions on station placement and maintenance.

Further research on the differences in data quality between Netatmo and Alecto, but also between DMR data and non DMR data, could help to understand better whether the interaction with participants improves data quality, or to what extent the differences are caused by the difference in the TBR's manufacturer.

5.3 PWS data quality and hydrological applications

The results from the Green Village setup are very promising when considering the price-quality ratio of the PWSs. They also indicate that the PWSs are quite sensitive to setup differences. This is confirmed when looking at the DMR data, where results from different stations vary between large positive and negative biases with regard to the reference. It is not easy to find a clear relationship between maintenance or setup variations between participants, due to the low sample size, and also because it is not possible to judge the levelness of stations from the survey questions or pictures. It seems reasonable to assume however that the high variability of the DMR data is related to the high variability of the setups. The fact that the data quality is higher than the Netatmo study from Vos et al. (2019) also suggests that engagement in citizen science project may improve results. Some recommendations to the DMR project management would be to stimulate the monitoring of the station, and advise what exactly should be checked. It would also be helpful if participants were reminded not to place their station in very close range to obstacles. Reminders could also be issued by the Alecto software. A 'cleaning mode' for temporary deactivation on the Alecto unit would be ideal to reduce faulty influxes.

Due to the variability of data between stations, it would currently not be possible to study the variation of rainfall within the urban environment from measurements of individual stations. When combining a dense set of measurements however, the data would good enough for estimates of rainfall, which was also noted by De Vos et al. (2017). This type of data could for example be used as input for hydrological models. Since gauge-adjusted radar data is not available, PWS data could provide a good alternative. However, as a high resolution is desirable, the resampling issues in the data processing are an obstacle. It would be a great improvement if the time of tipping from the stations would be transferred. Ideally, sampling intervals would be much shorter, as a study by Habib et al. (2001) showed that this improved data quality for 5 min resolution data. Alternatively it would be helpful if the time of data transfer was synchronized between stations, to reduce resampling errors.

Tipping experiment A has shown some issues with the software and data transfer with WOW which cause conversion errors in the reported rainfall correlation. Although this can be fixed by programming, it causes inconvenience to the users of the data. Similar issues with rounding and time stamp uncertainty were reported for Wundermap platform by De Vos et al. (2017). During the start of the setup on TGV it was also discovered that the data connection between station and indoor unit was not unique, which may lead to problems during experiments with multiple stations. Another issue of the Alecto is that there is no possibility of locally storing data, so that there are a lot of missing values when the internet connection is temporarily unavailable.

This thesis and previous research on PWS quality and QCs have shown that PWS data is useful in terms of quality, but that effort in future projects should also be directed towards solving the above mentioned issues with data processing and availability.

6 Conclusions

This thesis assessed the quality of precipitation measurements from the Alecto WS-5500 with three different approaches to answer three research questions. In this chapter, the answers to the research questions are provided, based on the research results, with the accompanying recommendations.

The first research question was: How accurate and reliable are the precipitation observations of the Alecto WS-5500, and is it possible to quantify setup differences?

From the experimental setup with 8 Alecto stations, it was concluded that the Alecto PWS is capable of measuring rainfall accurately. Through one of the tipping volume experiments performed, it was found that rainfall is overestimated because the actual tipping volume was lower than the reported tipping volume. A correction factor for this mechanical bias was introduced, so that the rainfall accumulations from the Alecto had low bias compared to the official KNMI gauge and KNMI gauge-adjusted radar reference.

The bias was shown to depend on the location of the station and its horizontal positioning. Substantial biases were found for the tilted station, but not for the station placed on the roof. The station close to a wall was found to underestimate substantially, although the measurements were inconsistent, possibly due to the influence of nearby vegetation and instrument bias.

For the stations with non-faulty setups, the correlation for hourly accumulations was high, but much lower for accumulations with 5 min intervals. This can be explained by sampling and resampling errors, where simultaneous rainfall is attributed to different timestamps for different stations.

The second research question was: To what extent can the quality of the DMR data set be improved by applying a QC method, and if so, how can the QC method be optimized for the Alecto WS-5500 data?

The quality control methods clearly showed that filtering improves data quality, although the success rate of filtering faulty zeros and faulty influxes was limited. QC de Vos was improved with stricter parameters for high influxes and stabilizing the dynamic bias correction process. The improvement of filtering faulty zeros was not possible without removing a substantial amount of data. The application of a constant bias correction, as found in the experiment, led to better results with a median bias close to zero.

Due to the sampling errors and lower correlation between stations at a 5 minute time scale, the filtering of the data was less successful at 5 minutes, as it relies on the comparison of observations at the same time intervals.

The third research question was: How does the monitoring and maintenance of owners of PWS influence data quality?

Although the experimental setup on the Green Village has proved the Alecto to be an accurate gauge if properly set up and maintained, and bias correction improves the data, the variation between station remained large, when comparing the biases. Inspection of station images and the survey among citizen owners of the Alecto PWS confirmed large differences in set up, surroundings, and maintenance of station, which may be the cause of this large variation in bias. A clear relationship between the station bias and setup information from DMR participants was not found, possibly due to the low sample size of the study.

The dependency on different applications complicates the installation process and leads to missing data. The dependency of data platforms and missing information on the mechanism of data processing makes it harder for researchers to access data, and to assess the source of faulty data.

The three main recommendations are:

1. A more systematic approach to the tilting experiments is recommended to be able to quantify small differences in setup. For the improvement of the experiment quality, it is recommended to explore the possibility of calibration with the Alecto stations. The non-linear relationship between rainfall intensity and bias in TBR measurements should be explored to determine whether the use of a static bias correction is justified.

2. High resolution data is the most valuable to hydrological applications, but not all filters are effective at the 5 min resolution. A suggestion for further research would be to apply filters on different resolutions and combine them, to improve the quality of the filtered data at the 5 min resolution.

3. To find out whether monitoring and maintenance regimes of PWS owners impact data quality, a higher number of stations is needed. A comparison between DMR data and other data sets from either Alecto or Netatmo stations is advised to research whether engagement with citizen science project significantly impacts data quality. The irregularities in the data processing steps that lead to lower data quality should be addressed in the development of PWSs and data platforms, to enable large scale (near) real time use of PWS data.

References

Alecto Home. (2023). *Alecto ws-5500 - professioneel 8 in 1 wifi weerstation met app en draadloze buitensensor, zwart.* Retrieved 2023-11-05, from `https://alectohome.nl/products/alecto-ws-5500-professioneel-wifi-weerstation-met-app-zwart`

Bárdossy, A., Seidel, J., & El Hachem, A. (2021). The use of personal weather station observations to improve precipitation estimation and interpolation. *Hydrology and Earth System Sciences*, *25*(2), 583–601. doi: 10.5194/hess-25-583-2021

Beele, E., Reyniers, M., Aerts, R., & Somers, B. (2022). Quality control and correction method for air temperature data from a citizen science weather station network in leuven, belgium. *Earth System Science Data*, *14*(10), 4681–4717. doi: 10.5194/essd-14-4681-2022

Bell, S., Cornford, D., & Bastin, L. (2015). How good are citizen weather stations? addressing a biased opinion. *Weather*, *70*(3), 75–84. doi: 10.1002/wea.2316

Berne, A., Delrieu, G., Creutin, J.-D., & Obled, C. (2004). Temporal and spatial resolution of rainfall measurements required for urban hydrology. *Journal of Hydrology*, *299*(3-4), 166–179. doi: 10.1016/j.jhydrol.2004.08.002

Burt, S. (2009). *The davis instruments vantage pro2 wireless aws–an independent evaluation against uk-standard meteorological instruments.*

Cristiano, E., ten Veldhuis, M.-c., Wright, D. B., Smith, J. A., & van de Giesen, N. (2019). The influence of rainfall and catchment critical scales on urban hydrological response sensitivity. *Water Resources Research*, *55*(4), 3375–3390. doi: 10.1029/2018WR024143

De Vos, H., Overeem, A., & Uijlenhoet, R. (2017). The potential of urban rainfall monitoring with crowdsourced automatic weather stations in amsterdam. *Hydrology and Earth System Sciences*, *21*(2), 765–777. doi: 10.5194/hess-21-765-2017

Droste, A. M., Heusinkveld, B. G., Fenner, D., & Steeneveld, G.-J. (2020). Assessing the potential and application of crowdsourced urban wind data. *Quarterly Journal of the Royal Meteorological Society*, *146*(731), 2671–2688. doi: 10.1002/qj.3811

El Hachem, A., Seidel, J., O'Hara, T., Villalobos Herrera, R., Overeem, A., Uijlenhoet, R., … de Vos, L. (2023). Technical note: Overview and comparison of three quality control algorithms for rainfall data from personal weather stations. *Hydrology and Earth System Sciences Discussions*, *2023*, 1–22. doi: 10.5194/hess-2023-195

Estavez, J., Gavilan, P., & Giraldez, J. (2011). Guidelines on validation procedures for meteorological data from automatic weather stations. *Journal of Hydrology*, *402*(1), 144-154. doi: 10.1016/j.jhydrol.2011.02.031

Fenner, D., Meier, F., Bechtel, B., Otto, M., & Scherer, D. (2017). Intra and inter local climate zone variability of air temperature as observed by crowdsourced citizen weather stations in berlin, germany. *Meteorologische Zeitschrift*, *26*, 525–547. doi: 10.1127/metz/2017/0861

Habib, E., Krajewski, W. F., & Kruger, A. (2001). Sampling errors of tipping-bucket rain gauge measurements. *Journal of Hydrologic Engineering*, *6*(2), 159-166. doi: 10.1061/(ASCE)1084-0699(2001)6:2(159)

Humphrey, M., Istok, J., Lee, J., Hevesi, J., & Flint, A. (1997). A new method for automated dynamic calibration of tipping-bucket rain gauges. *Journal of Atmospheric and Oceanic Technology*, *14*(6), 1513–1519. doi: 10.1175/1520-0426(1997)014<1513:ANMFAD>2.0.CO;2

KiesKeurig. (2023). *Alecto ws-5500 - overzicht.* Retrieved 2023-11-05, from `https://www.kieskeurig.nl/weerstation/product/3588501alecto-ws-5500`

Kim, D.-K., & Lee, D.-I. (2016). Raindrop size distribution properties associated with vertical air motion in the stratiform region of a springtime rain event from 1290 mhz wind profiler, micro rain radar and parsivel disdrometer measurements. *Meteorological Applications*, *23*(1), 40–49. doi: 10.1002/met.1518

KNMI. (2023a). *Over wow-nl.* Retrieved 2023-05-02, from `https://wow.knmi.nl/over-wow-nl`

KNMI. (2023b). *Precipitation - 5 minute precipitation accumulations from climatological gauge-adjusted radar dataset for the netherlands (1 km) in knmi hdf5 format.* Retrieved 2023-10-23, from `https://dataplatform.knmi.nl/dataset/rad-nl25-rac-mfbs-5min-2-0`

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., ... Sherstyukov, B. (2014). Flood risk and climate change: global and regional perspectives. *Hydrological Sciences Journal*, *59*(1), 1-28. doi: 10.1080/02626667.2013.857411

Meier, F., Fenner, D., Grassmann, T., Otto, M., & Scherer, D. (2017). Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Climate*, *19*, 170-191. doi: 10.1016/j.uclim.2017.01.006

METEK. (2023). *Micro rain radar mrr.* METEK Meteorologische Messtechnik GmbH. Retrieved 2024-02-13, from `https://metek.de/product-group/micro-rain-radar-24-ghz/`

Nationale Milieu Database. (2023). *The green village, proeftuin voor duurzame innovaties.* Retrieved 2024-02-20, from `https://milieudatabase.nl/nl/actueel/nieuws/the-green-village-proeftuin-voor-duurzame-innovaties/`

Netatmo. (2023). *Netatmo weathermap.* Retrieved 2023-05-02, from `https://weathermap.netatmo.com/`

Nielsen, J., van de Beek, C., Thorndahl, S., Olsson, J., Andersen, C., Andersson, J., ... Nielsen, J. (2024). Merging weather radar data and opportunistic rainfall sensor data to enhance rainfall estimates. *Atmospheric Research*, *300*, 107228. doi: 10.1016/j.atmosres.2024.107228

Ochoa-Rodriguez, S., Wang, L.-P., Gires, A., Pina, R. D., Reinoso-Rondinel, R., Bruni, G., ... others (2015). Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation. *Journal of Hydrology*, *531*, 389–407. doi: 10.1016/j.jhydrol.2015.05.035

Office, M. (2024). *Weather observations website.* Retrieved 2023-05-10, from `https://www.wunderground.com/about/our-company`

Overeem, A., & Leijnse, H. (2022). *Rapportage wiwb onderzoek deelaspecten 5 & 7:kwaliteitsinformatie gebruiken voor het mengen van verschillende radars en voor het mengen van radars en regenmeters* (Tech. Rep.). KNMI.

Overeem, A., Leijnse, H., van der Schrier, G., van den Besselaar, E., Garcia-Marti, I., & de Vos, L. W. (2023). Merging with crowdsourced rain gauge data improves pan-european radar precipitation estimates. *Hydrology and Earth System Sciences Discussions*, *2023*, 1–29. doi: 10.5194/hess-2023-122

Peters, G., Fischer, B., & Andersson, T. (2002). Rain observations with a vertically looking micro rain radar (mrr). *Boreal environment research*, *7*(4), 353.

Pollock, M. D., O'donnell, G., Quinn, P., Dutton, M., Black, A., Wilkinson, M., ... others (2018). Quantifying and mitigating wind-induced undercatch in rainfall measurements. *Water Resources Research*, *54*(6), 3863–3875. doi: 10.1029/2017WR022421

Praskievicz, S., & Chang, H. (2009). A review of hydrological modelling of basin-scale climate change and urban development impacts. *Progress in Physical Geography*, *33*(5), 650–671. doi: 10.1177/0309133309348098

Rosenzweig, B. R., McPhillips, L., Chang, H., Cheng, C., Welty, C., Matsler, M., ... Davidson, C. I. (2018). Pluvial flood risk and opportunities for resilience. *Wiley Interdisciplinary Reviews: Water*, *5*(6), e1302. doi: 10.1002/wat2.1302

Ruisdael observatory. (2023). Ruisdael Observatory. Retrieved 2024-02-13, from `https://ruisdael-observatory.nl/`

Rözer, V., Müller, M., Bubeck, P., Kienzler, S., Thieken, A., Pech, I., ... Kreibich, H. (2016). Coping with pluvial floods by private households. *Water*, *8*(7). doi: 10.3390/w8070304

Schilling, W. (1991). Rainfall data for urban hydrology: what do we need? *Atmospheric Research*, *27*(1-3), 5–21. doi: 10.1016/0169-8095(91)90003-F

Segovia-Cardozo, D. A., Bernal-Basurco, C., & Rodríguez-Sinobas, L. (2023, Jun). Tipping bucket rain gauges in hydrological research: Summary on measurement uncertainties, calibration, and error reduction strategies. *Sensors*, *23*(12), 5385. doi: 10.3390/s23125385

Segovia-Cardozo, D. A., Rodríguez-Sinobas, L., Díez-Herrero, A., Zubelzu, S., & Canales-Ide, F. (2021). Understanding the mechanical biases of tipping-bucket rain gauges: a semi-analytical calibration approach. *Water*, *13*(16), 2285. doi: 10.3390/w13162285

van Andel, N. (2021). *Quality control development for near real-time rain gauge networks for operational rainfall monitoring* (Unpublished master's thesis). Utrecht University.

Vos, D., 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

Weather Underground. (2023). *Wundermap.* Retrieved 2023-05-10, from `https://www.wunderground.com/about/our-company`

WMO. (2018a). Guide to instruments and methods of observation. volume iii: Observing systems. In (chap. ANNEX. AUTOMATIC WEATHER STATIONS – LOW COST). World Meteorological Organization. Retrieved from `https://library.wmo.int/index.php`

WMO. (2018b). Guide to instruments and methods of observation, volume i: Measurement of meteorological variables. In (chap. ANNEX 6.C. STANDARDIZED PROCEDURE FOR LABORATORY CALIBRATION OF CATCHMENT TYPE RAINFALL INTENSITY GAUGES). World Meteorological Organization. Retrieved from `https://library.wmo.int/index.php`

A Methodology appendix

A.1 The Green Village set up: materials

- Alecto WS-5500 box contents: main frame outdoor unit, anemometer, wind vane, tipping funnel, screws, mounting brackets, indoor unit+ power cable, installation manual
- 3 AAA batteries (outdoor unit)
- 2 AA batteries (indoor unit)
- Small phillips screwdriver
- PVC/steel pole, 50 mm diameter, 100 cm length.
- Smartphone or tablet with WS View or WS View Plus app installed
- Auger
- Hammer
- Measuring tape
- Notebook
- Compass (smartphone)
- Labelprinter
- WiFi secured with WPA2-PSK (AES), 2.4GHz connection

Additional Experiments:

- Stopwatch and clock (smartphone)
- Measuring tape
- 5 mL syringe (or alternative water measuring device)
- Notebook
- Paper towels

A.2 The Green Village set up: Tipping experiment A

round no.	Station no. 1	2	3	4	5	7	8
1	1	1	1	1	1	-	1
2	3	3	3	2	2	3	2
3	4	4	4	3	3	4	3
4	5	5	5	4	4	10	4
5	6	6	6	5	5	11	5
6	7	7	7	6	6	12	7
7	8	8	8	7	7	13	7
8	9	9	9	8	8	14	8
9	10	10	10	9	9	15	10
10	30	22	23	24	25	37	38

Table 5: *Number of tips in tipping volume experiment A.*

A.3 Quality control of DMR 2021-2023 : coding flow charts

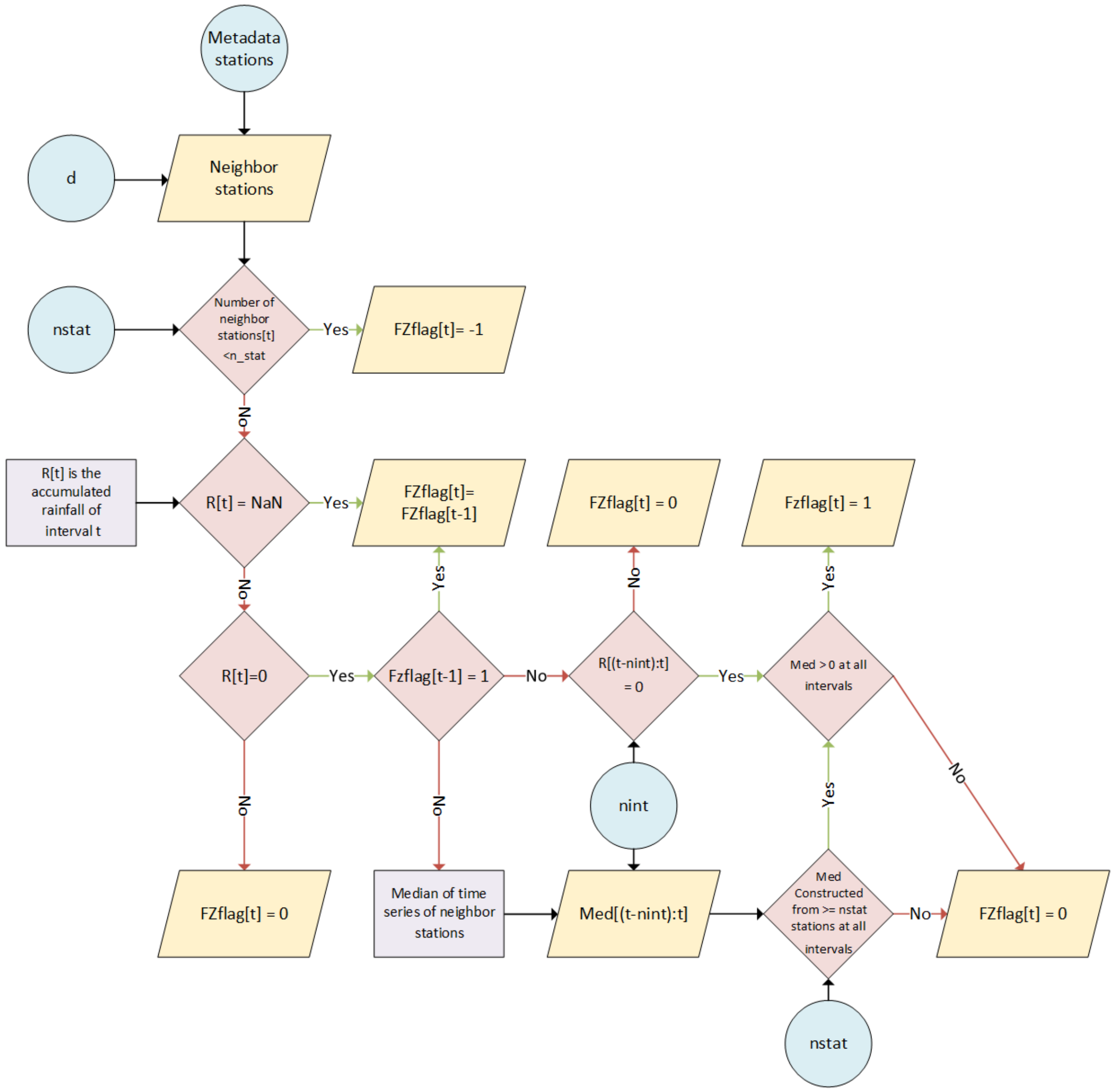


Figure 41: *FZ filter QC de Vos*

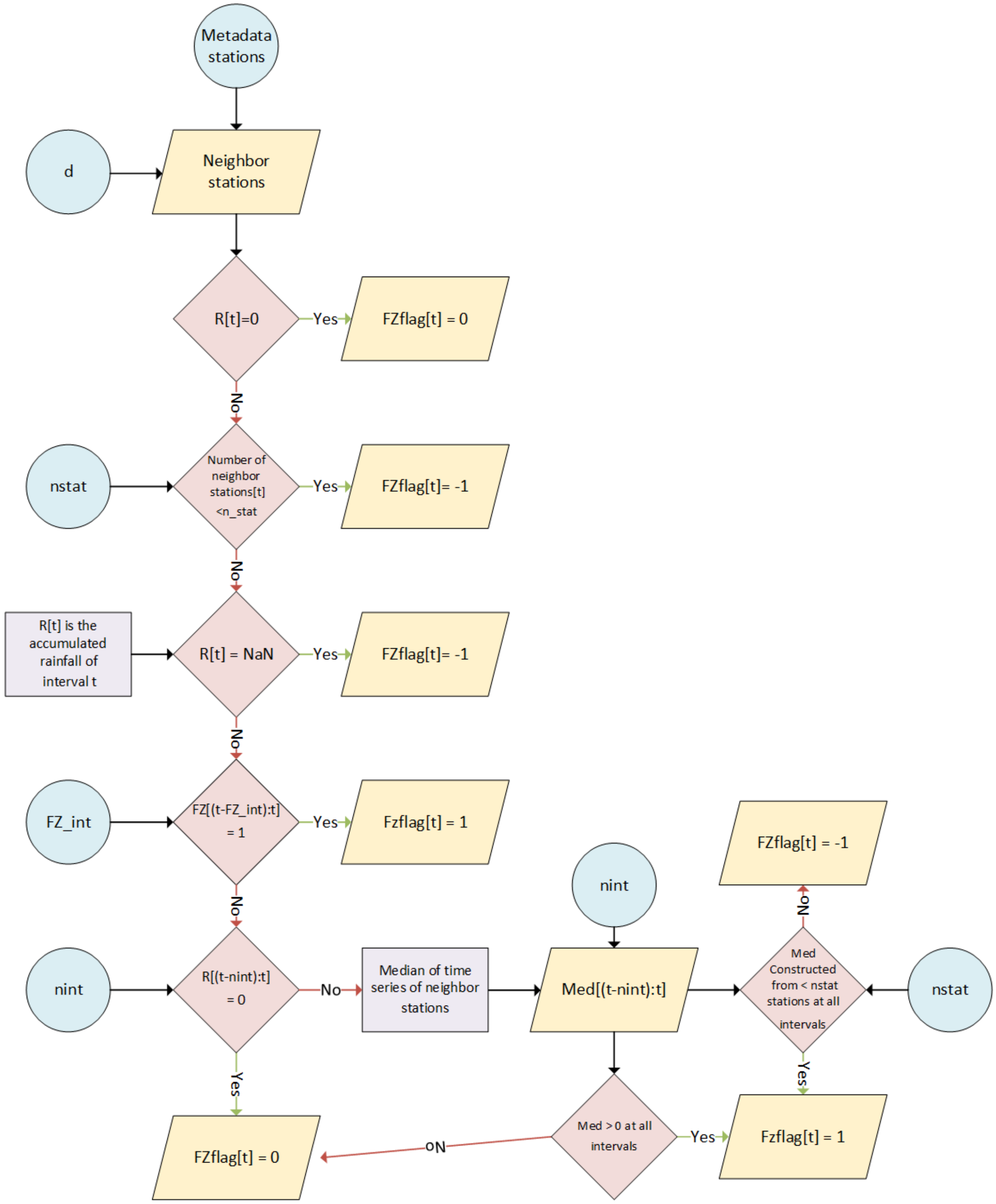


Figure 42: *FZ filter QC DMR*

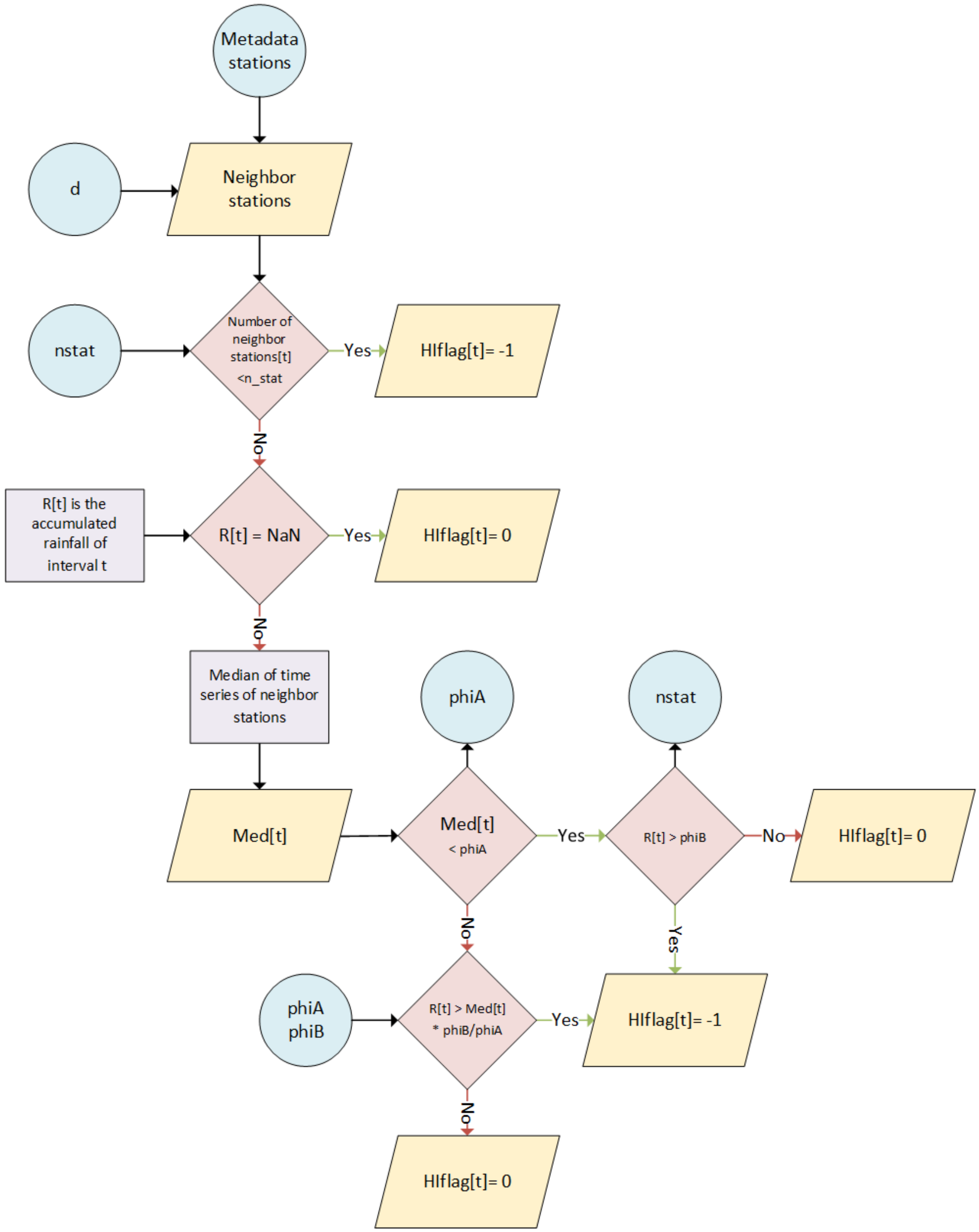


Figure 43: *HI filter QC de Vos*

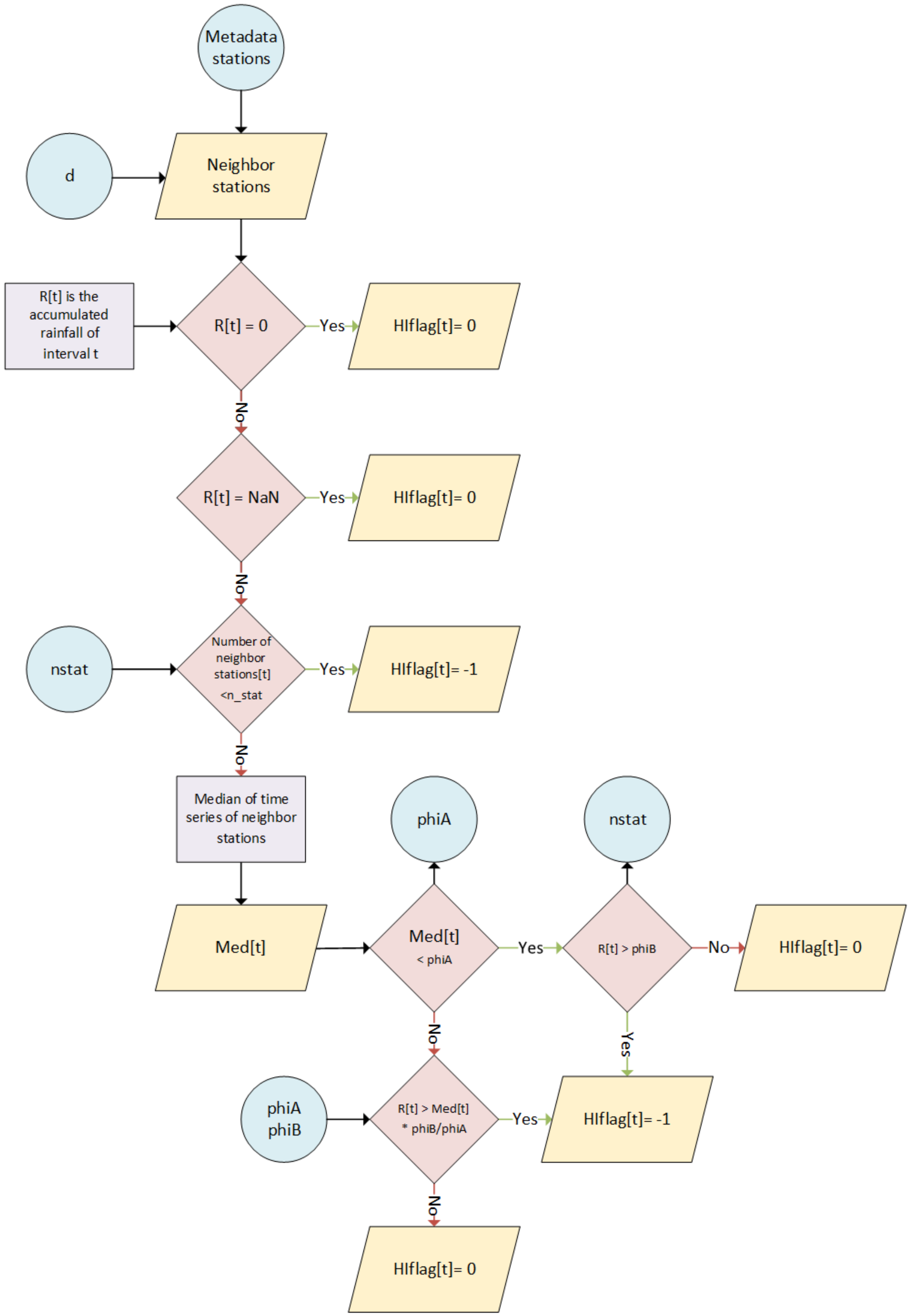


Figure 44: *HI filter QC DMR*

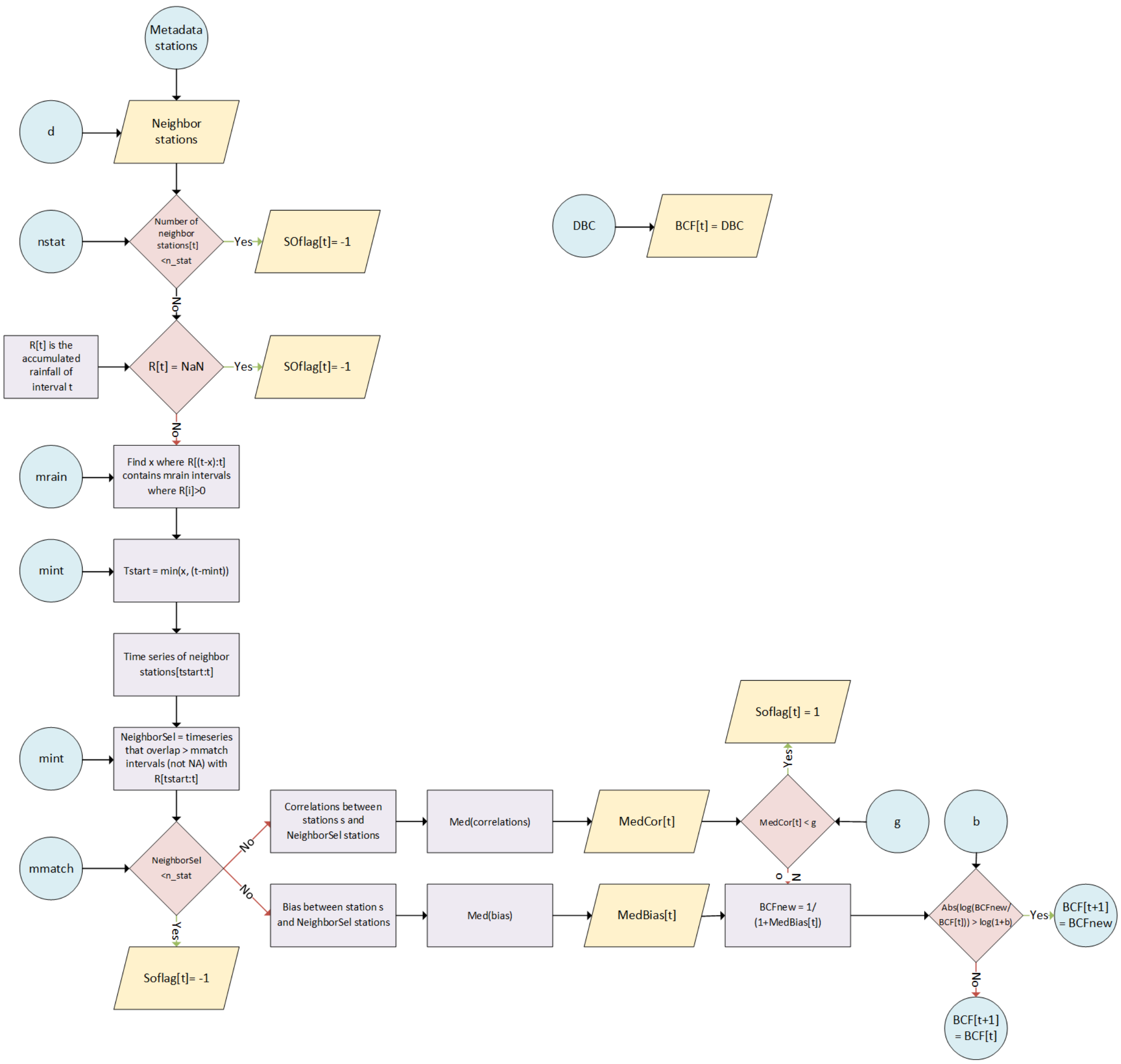


Figure 45: *SO filter and BC QC de Vos*

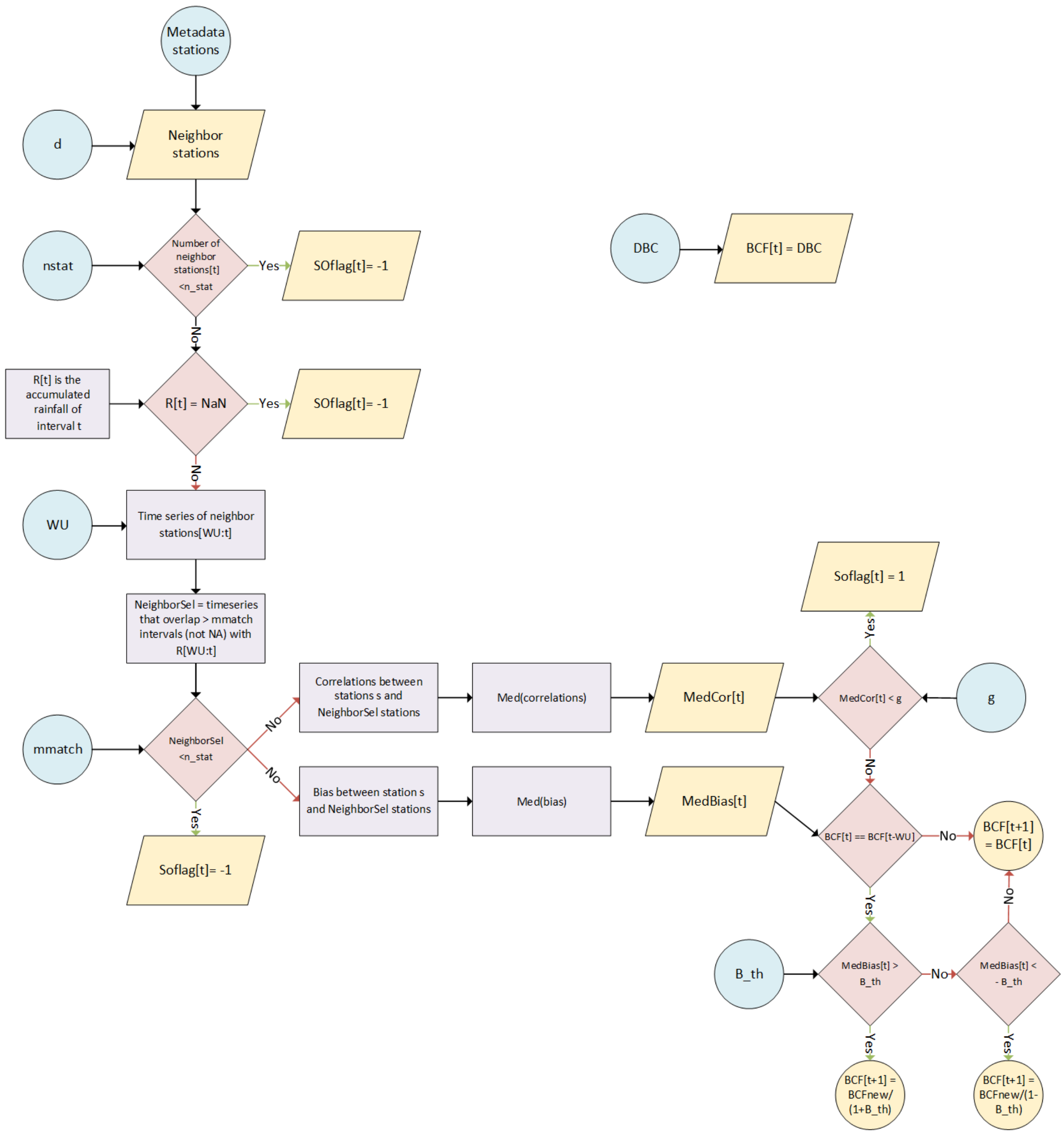


Figure 46: *SO filter and BC QC DMR*

A.4 Participants' survey

Question 1: This question is about the monitoring and maintenance of the outdoor unit of your weather station. How often did you...

1. ... check whether the station was still level?
2. ... check for a blockage of the funnel of the rainfall sensor?
3. ... check for a blockage or dirt in the tipping bucket by screwing off the funnel from the rainfall sensor?
4. ... clean the outside of the weather station, without screwing off the funnel of the rainfall sensor?
5. ... clean the rainfall sensor, including the tipping bucket beneath the funnel?

Possible answers (multiple choice):

- Once a week
- Once a month
- Once every 3 months
- Once every half year
- Once a year
- Never

Question 2: Did the monitoring or maintenance of the outside take place because of a specific event?
Possible answers (multiple choice, multiple answers possible):

- Yes, after an update or webinar from the Delft Measures project.
- Yes, during or after a period with high precipitation.
- Yes, after storm.
- Yes, after a period with snow and/or subzero temperatures.
- Yes, during or after a period with high temperatures.
- No, I checked my weather station periodically.
- No, I (almost) never check my weather station.

Question 3: This question is about the monitoring of the indoor unit of your weather station. How often did you...

1. ... use the display to read observations (e.g. temperature, rainfall) from your weather station?
2. ... use the websites wow.knmi.nl of wow.metoffice.gov.uk to read observations (e.g. temperature, rainfall) from your weather station?

3. ... use the websites wow.knmi.nl of wow.metoffice.gov.uk to read observations (e.g. temperature, rainfall) from other weather station?

Possible answers (multiple choice):

- Daily
- Once a week
- Once a month
- Once every 3 months
- Once every half year or less
- Never

Question 4: Did the monitoring of your indoor unit happen because of a specific event?
Possible answers (multiple choice, multiple answers possible):

- Yes, after an update or webinar from the Delft Measures project.
- Yes, during or after a period with high precipitation.
- Yes, during or after storm.
- Yes, after a period with snow and/or subzero temperatures.
- Yes, during or after a period with high temperatures.
- No, I checked my weather station periodically.
- No, I (almost) never check my weather station.

B Results appendix

B.1 The Green Village set up: tipping volume experiments

Table 6: *Number of tips in tipping volume experiment A.*

round no.	Station no. 1	2	3	4	5	7	8
1	1	1	1	1	1	-	1
2	3	3	3	2	2	3	2
3	4	4	4	3	3	4	3
4	5	5	5	4	4	10	4
5	6	6	6	5	5	11	5
6	7	7	7	6	6	12	7
7	8	8	8	7	7	13	7
8	9	9	9	8	8	14	8
9	10	10	10	9	9	15	10
10	30	22	23	24	25	37	38

Table 7: *Number of tips in tipping volume experiment A.*

Round	time interval (UTC)	station no. 1	2	3	4	5	7	8
		no.of tips						
1	15.56-16.01	1	1	1	1	1	-	1
2	16.02-16.07	3	3	3	2	2	3	2
3	16.08-16.13	4	4	3	3	4	3	3
4	16.14-16.19	5	5	5	4	4	10	4
5	16.21-16.26	6	6	6	5	5	11	5
6	16.27-16.32	7	7	7	6	6	12	6
7	16.33-16.38	8	8	8	7	7	13	7
8	16.43-16.47	9	9	9	8	8	14	8
9	16.48-16.53	10	10	10	9	9	15	9
10	16.56-17.01	30	22	23	24	34	37	48

Table 8: *tipping volume experiment A results from WOW*

Station 1				Station 2			
t_{obs}	p_t	x	p_x	t_{obs}	p_t	x	p_x
16:02:07	0.2794	1	0.2794	16:04:31	0.2032	1	0.2032
16:07:27	0.7112	3	0.2371	16:09:52	0.7874	3	0.26247
16:12:47	0.9906	4	0.24765	-	-	-	-
16:23:27	1.2954	5	0.2591	16:21:04	2.3114	9	0.25682
16:28:47	1.5240	6	0.254	16:31:44	1.4986	6	0.249767
16:34:07	1.7780	7	0.254	16:37:04	1.8034	7	0.2576
16:39:27	2.0066	8	0.2508	16:42:24	2.0066	8	0.2576
16:50:07	2.3114	9	0.2568	16:47:44	2.2860	9	0.254
16:55:27	2.5908	10	0.2591	16:58:24	2.5146	10	0.25146
17:06:07	7.5946	30	0.25315	17:03:44	5.5880	21	0.254

Station 3				Station 4			
t_{obs}	p_t	x	p_x	t_{obs}	p_t	x	p_x
16:01:24	0.3048	1	0.3048	15:58:44	0.3048	1	0.3048
16:06:44	0.7112	3	0.237067	16:04:04	0.4826	2	0.2413
16:12:04	0.9906	4	0.24765	16:09:24	0.8128	3	0.2709
16:22:44	1.2954	5	0.25908	16:14:44	0.9906	4	0.24765
16:28:04	1.4986	6	0.249767	16:25:24	1.2192	5	0.24384
16:33:24	1.8034	7	0.25762	16:30:44	1.6002	6	0.26667
16:38:44	2.0066	8	0.2508	16:41:24	1.7018	7	0.24311
16:49:24	2.2860	9	0.2540	16:46:44	2.0828	8	0.26035
16:54:44	2.6162	10	0.2616	16:52:04	2.3114	9	0.26582
17:05:24	5.7912	23	0.25179	17:02:44	6.0960	24	0.254

Station 5				Station 7			
t_{obs}	p_t	x	p_x	t_{obs}	p_t	x	p_x
16:01:51	0.3048	1	0.3048	-	-	-	-
16:07:11	0.4826	2	0.2413	16:07:03	0.7112	3	0.23706
16:12:31	0.8128	3	0.27093	16:17:59	0.9906	4	0.24764
16:17:51	0.9906	4	0.24765	16:23:19	2.5908	10	0.25908
16:28:31	1.2192	5	0.24384	16:28:31	2.8194	11	0.2563
16:33:51	1.6002	6	0.26667	16:39:03	2.9972	12	0.249766
16:39:11	1.7018	7	0.24311	16:44:23	3.3020	13	0.254
16:44:31	2.0828	8	0.26035	16:49:43	3.5814	14	0.25581
-	-	-	-	17:00:23	2.5146	15	0.16764
17:00:47	8.6106	34	0.25325	17:05:43	9.3980	37	0.254

Station 7			
t_{obs}	p_t	x	p_x
16:03:31	0.2032	1	0.2032
16:08:54	0.5080	2	0.254
16:14:14	0.7874	3	0.2624666
16:24:54	0.9906	4	0.24765
16:30:14	1.3208	5	0.26416
' 16:35:31	1.7018	6	0.2431
16:46:14	1.7780	7	0.253999
16:51:31	2.1082	8	0.263525
16:56:54	2.5146	9	0.25146
17:07:31	12.1920	48	0.254

Table 6: *Tipping volume experiment B: emptied bucket results.*

Round	station no. 1	2	3	4	5
	tipping volume (mL)				
1	2.8	2.4	2.2	2.6	2.4
2	2.6	2.5	2.3	2.6	2.4
3	2.7	2.5	2.1	2.6	2.4
4	2.7	2.4	2.2	2.7	2.3
5	2.7	2.3	2.2	2.4	2.4

Table 7: *Tipping volume experiment B: unemptied bucket results.*

Round	station no. 1	2	3	4	5
	tipping volume (mL)				
1	-	2.2	2.1	1.7	1.8
2	2.5	2.2	2.1	2.1	2.2
3	2.6	2.3	2.2	2.3	2.2
4	2.4	2.3	2.0	2.3	2.1
5	2.4	2.3	2.0	2.2	2.2
6	2.4	2.1	2.2	2.3	2.1
7	2.5	2.2	2.0	2.1	2.1
8	2.3	2.3	2.2	2.2	2.1
9	2.5	2.3	2.2	2.3	2.0
10	2.4	2.2	2.3	2.2	2.2

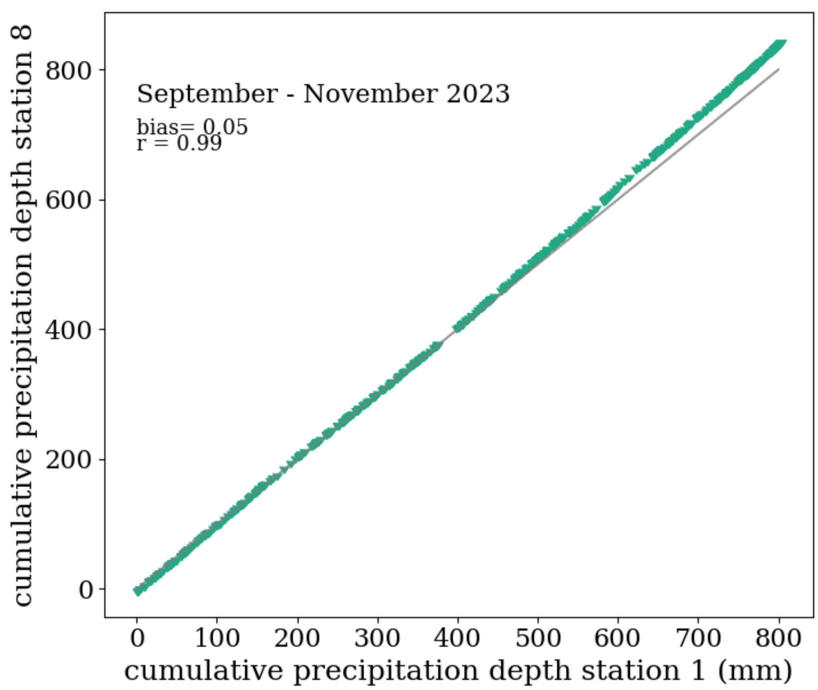


Figure 47: *Double mass plots of filtered rainfall measurements from station 8 and station 1. The 1 h accumulation data was used.*

Table 8: *Result of the comparison of the Alectos 2 to 8 with Alecto 1 as reference, with hourly data in the period of July 1st to November 30th 2023*. Comparison is done for unfiltered and filtered data. *For station 3, the comparison is performed for the period of September 8th to November 30th.*

station	R (mm)	R_{ref} (mm)	ΔR (mm)	r. bias	CV	r	n	RMSE
	Tip-adjusted, unfiltered							
2	796	808	-11	-0.014	0.47	0.94	638	0.59
3	613	540	73	0.14	0.34	0.98	463	0.43
4	812	796	15	0.019	0.26	0.98	641	0.33
5	684	788	-105	-0.133	0.80	0.82	623	1.03
6	819	808	11	0.014	0.24	0.99	645	0.30
7	821	808	13	0.017	0.27	0.98	657	0.33
8	856	808	49	0.060	0.26	0.98	642	0.34
	FZ, HI, SO filtered							
2	788	761	26	0.035	0.23	0.99	561	0.32
3	607	536	71	0.13	0.33	0.98	429	0.44
4	803	788	14	0.018	0.24	0.98	584	0.33
5	640	720	-80	-0.111	0.76	0.81	514	1.08
6	811	802	9	0.011	0.23	0.99	595	0.31
7	812	799	13	0.017	0.25	0.98	598	0.33
8	846	803	43	0.054	0.24	0.99	589	0.34

Table 9: *Results of the comparison of the Alecto 1 to four different references, with hourly data* in the period of July 1st to November 30th 2023. Comparison is done for unfiltered, filtered data and bias corrected data. *For the KNMI gauge, daily data is used. The four references are the KNMI gauge, the KNMI radar, the Micro Rain Radar and the Davis Vantage Pro 2.*

Ref.	R (mm)	R_{ref} (mm)	ΔR (mm)	r. bias	CV	r	n	RMSE
				unfiltered, tip-adjusted				
Gauge	819	740	79	0.11	0.31	0.98	107	2.26
Radar	821	709	112	0.16	0.80	0.92	890	0.65
MRR	820	1055	-235	-0.22	2.69	0.76	3454	0.82
D VP2	819	431	386	0.90	2.14	0.65	677	1.48
				FZ, HI, SO filtered				
Gauge	805	727	78	0.11	0.31	0.98	103	2.29
Radar	821	699	122	0.17	0.78	0.92	855	0.65
MRR	820	1045	-223	-0.21	2.67	0.76	3331	0.84
D VP2	817	416	401	0.96	2.16	0.66	662	1.49
				FZ, HI, SO filtered + CMBC				
Gauge	724	727	-3	-0.00	0.27	0.98	103	1.90
Radar	739	699	40	0.06	0.70	0.92	855	0.58
MRR	738	1044	-305	-0.29	2.7	0.76	3331	0.85
D VP2	743	416	327	0.79	1.95	0.66	662	1.32

B.2 Quality control of DMR

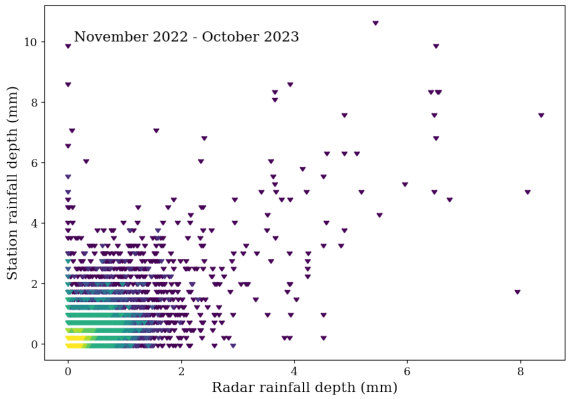


Figure 48: *HI filter QC De Vos 5.*

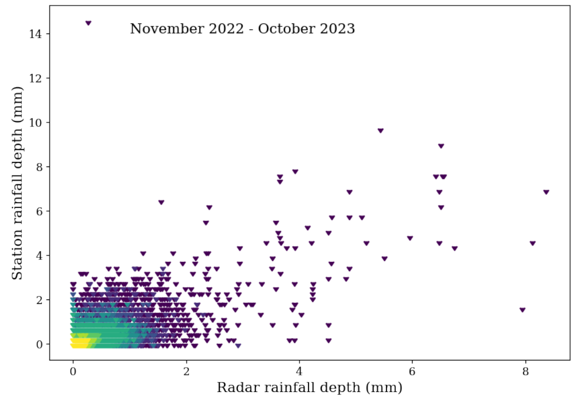


Figure 49: *HI filter QC DMR5-2.*

Table 10: *Validation metrics for 1 h data running from November 1st 2022 to October 31st 2023, for different quality control methods and filter combinations. The references are the corresponding pixels of the KNMI gauge-adjusted radar.*

QC method	Applied filters	r.bias	CV	r	% filtered	RMSE	med. bias
	-	0.16	2.72	0.87	100	0.34	0.15
De Vos 60	FZ, HI	0.20	2.61	0.87	89.57	0.35	0.18
	FZ, HI, SO	0.20	2.51	0.88	89.40	0.34	0.18
	FZ, HI, SO, BC	0.26	3.20	0.85	89.40	0.43	0.23
De Vos 5R	FZ, HI	0.16	2.21	0.91	99.18	0.28	0.15
	FZ, HI, SO	0.16	2.21	0.91	99.09	0.28	0.15
	FZ, HI, SO, BC	0.18	2.23	0.91	99.09	0.28	0.19
DMR60-1	FZ, HI	0.20	1.99	0.91	76.61	0.31	0.18
	FZ, HI, SO	0.20	1.97	0.91	76.56	0.31	0.18
	FZ, HI, SO, BC	0.20	1.97	0.91	76.56	0.31	0.22
	all + CMBC	0.09	1.81	0.91	76.56	0.28	0.11
DMR60-2	FZ, HI	0.19	2.10	0.91	90.23	0.28	0.18
	FZ, HI, SO	0.19	2.10	0.91	90.21	0.28	0.18
	FZ, HI, SO, BC	0.19	2.08	0.91	90.21	0.28	0.20
	all + CMBC	0.08	1.92	0.91	90.21	0.26	0.09
ref $\geq$ 0.508	-	0.06	0.67	0.81	100	1.13	0.05
De Vos 60	FZ, HI	0.09	0.65	0.82	96.24	1.12	0.07
	FZ, HI, SO	0.09	0.65	0.82	96.02	1.11	0.07
	FZ, HI, SO, BC	0.14	0.83	0.78	96.02	1.43	0.13
De Vos 5R	FZ, HI	0.07	0.58	0.85	98.64	0.98	0.05
	FZ, HI, SO	0.07	0.58	0.85	98.46	0.98	0.05
	FZ, HI, SO, BC	0.09	0.58	0.86	98.46	0.98	0.08
DMR60-1	FZ, HI	0.09	0.57	0.85	95.80	0.98	0.07
	FZ, HI, SO	0.09	0.57	0.85	95.49	0.98	0.06
	FZ, HI, SO, BC	0.08	0.57	0.85	95.49	0.97	0.10
	all + CMBC	-0.01	0.53	0.85	95.49	0.97	-0.00
DMR60-2	FZ, HI	0.08	0.57	0.86	96.20	0.97	0.06
	FZ, HI, SO	0.08	0.57	0.86	96.11	0.97	0.06
	FZ, HI, SO, BC	0.08	0.56	0.85	96.11	0.96	0.08
	all + CMBC	-0.02	0.52	0.85	96.11	0.96	-0.02

Table 11: *Validation metrics for 1 h data running from November 1st 2022 to October 31st 2023, for different quality control methods and filter combinations. The references are the corresponding pixels of the KNMI gauge-adjusted radar.*

QC method	Applied filters	r.bias	CV	r	% filtered	RMSE	med. bias
	Raw	0.27	13.84	0.39	100	0.14	0.25
De Vos 5	FZ, HI	0.26	6.38	0.68	99.18	0.07	0.25
	FZ, HI, SO	0.26	6.38	0.68	99.09	0.07	0.25
	FZ, HI, SO, BC	0.28	6.71	0.67	99.09	0.07	0.28
DMR5-1	FZ, HI	0.30	6.10	0.70	91.94	0.07	0.31
	FZ, HI, SO	0.29	6.11	0.70	91.92	0.07	0.30
	FZ, HI, SO, BC	0.28	6.05	0.70	91.92	0.07	0.28
	all + CMBC	0.17	5.58	0.70	91.92	0.07	0.17
DMR5-2	FZ, HI	0.25	6.20	0.69	99.89	0.06	0.25
DMR5-3	FZ, HI	0.30	6.08	0.70	90.94	0.07	0.31
ref $\geq$ 0.254 mm	-	-0.09	0.91	0.57	100	0.45	-0.09
De Vos 5	FZ, HI	-0.08	0.88	0.59	98.55	0.43	-0.09
	FZ, HI, SO	-0.08	0.88	0.59	98.36	0.43	-0.09
	FZ, HI, SO, BC	-0.06	0.95	0.57	98.36	0.47	-0.04
DMR5-1	FZ, HI	-0.06	0.89	0.58	95.39	0.44	-0.06
	FZ, HI, SO	-0.06	0.89	0.58	95.10	0.44	-0.06
	FZ, HI, SO, BC	-0.07	0.88	0.58	95.10	0.43	-0.06
	all + CMBC	-0.15	0.81	0.58	95.10	0.43	-0.15
DMR5-2	FZ, HI	-0.08	0.90	0.58	97.98	0.44	-0.09
DMR5-3	FZ, HI	-0.05	0.89	0.59	95.13	0.44	-0.05

Table 12: *Validation metrics for 1 h data running from September 1st 2023 to November 30th 2023, for different quality control methods and filter combinations. The references are the corresponding pixels of the KNMI gauge-adjusted radar.*

QC method	Applied filters	r.bias	CV	r	% filtered	RMSE	med. bias
	Raw	0.09	1.73	0.92	100	0.40	0.11
De Vos 60	FZ, HI	0.12	1.64	0.92	90.07	0.40	0.13
	FZ, HI, SO	0.12	1.64	0.92	89.85	0.40	0.13
	FZ, HI, SO, BC	0.17	2.08	0.89	89.85	0.51	0.15
DMR60-2	FZ, HI	0.12	1.55	0.93	91.13	0.38	0.13
	FZ, HI, SO	0.12	1.55	0.93	91.12	0.38	0.13
	FZ, HI, SO, BC	0.13	1.56	0.93	91.12	0.38	0.12
	all + CMBC	0.02	1.43	0.93	91.12	0.35	0.02
ref $\geq$ 0.508	-	0.03	0.59	0.89	100	1.10	0.04
De Vos	FZ, HI	0.05	0.57	0.89	96.97	1.08	0.06
	FZ, HI, SO	0.05	0.57	0.89	96.93	1.08	0.06
	FZ, HI, SO, BC	0.09	0.74	0.83	96.93	1.40	0.08
DMR V2	FZ, HI	0.05	0.55	0.90	96.91	1.03	0.04
	FZ, HI, SO	0.05	0.55	0.90	96.86	1.03	0.04
	FZ, HI, SO, BC	0.05	0.55	0.90	96.86	1.04	0.06
	all + CMBC	-0.04	0.50	0.90	96.86	0.95	-0.04

Table 13: *Remaining faulty zero values after FZ filters from different quality control methods, for both 5 min and 1 h data of the 2022-2023 data set. Different subsets for reference values are evaluated.*

QC method	n	% filtered	n	% filtered	n	% filtered
	1h, ref > 0		1h, ref $\geq$ 0.254		1h, ref $\geq$ 0.508	
-	10663	0	1538	0	641	0
De Vos 60	8005	25	684	56	228	64
De Vos 5R.	10345	3	1339	13	641	0
DMR60-1	6984	35	552	64	181	72
DMR60-2	8048	25	689	56	230	64
	5 min, ref > 0		5 min, ref $\geq$ 0.254			
Raw	92897	0	3643	0	-	-
De Vos 5	91079	2	3364	8	-	-
DMR5-1	83860	10	2773	24	-	-
DMR5-2	91602	1	3272	10	-	-
DMR5-3	82744	11	2723	25	-	-

Table 14: *Remaining faulty zero values after FZ filters from different quality control methods, for 1 h data of the 2022-2023 data set. Different subsets for reference values are evaluated.*

Filter	n	% filtered	n	% filtered	n	% filtered
	1h, ref > 0		1h, ref $\geq$ 0.254		1h, ref $\geq$ 0.508	
-	7783	0	1053	0	372	0
De Vos	5815	25	398	62	62	83
DMR V2	5872	24	401	62	63	83