Royal Netherlands
Meteorological Institute
*Ministry of Infrastructure
and Water Management*

Utrecht University

Faculty of Physics

Quality control development for near real-time rain gauge networks for operational rainfall monitoring

Master Thesis

Niek van Andel

Supervisors:

Dr. A. Overeem & Dr. H. Leijnse
Royal Netherlands Meteorological Institute (KNMI)

Dr. A.J. van Delden
Institute for Marine and Atmospheric Research (IMAU)

July 2, 2021

Abstract

Water, especially fresh water, plays an unmissable role in nature, agricultural activities or as drinking water. How much precipitation actually reaches the ground is therefore important to monitor. The number of official rain gauges is however limited. If present, weather radars are able to indicate rainfall with a high spatial and temporal resolution but do not grasp actual rainfall amounts. Therefore, in recent years methods are proposed to automatically quality control (QC) measurements of third party near real-time rain gauge networks by comparison of a measurement to measurements obtained nearby. Here, a method is proposed that includes weather radar observations to quality control an unprecedented large amount of measurements of rain gauges employed by Royal Netherlands Meteorological Institute (KNMI) and third parties as water boards and Personal Weather Stations (PWSs).

Results show that the method is able to improve data quality of various real-time rain gauge networks. For PWS data of Netatmo collected in Amsterdam in 2018, the data set quality measured for 5-min rainfall accumulations is improved to a relative bias of 0.06 (-0.11 originally), a coefficient of variation of 6.8 (53.2 originally) and a Pearson correlation coefficient of 0.59 (0.07 originally) by removing 27.3% of all data. Application of the existing method however filters less data while yielding the same results. For water boards and KNMI data, this existing method cannot be applied due to the lack of nearby rain gauges. The newly proposed method however is able to improve all data sets, despite all differences in setup, maintenance and intrinsic quality. It is furthermore shown that with parameter optimization, it is able to filter 20%-point less measurements without quality loss, hence showing the wide applicability of the QC method if real-time weather radar data is present. Application of the proposed method on a national scale for 1 year of Netatmo data, shows that quality controlled third party data can be used as viable source for operational rainfall monitoring.

Layman summary (Dutch)

In dit onderzoek is gekeken hoe de kwaliteit van automatische regenmeters, of ze nu beheerd worden door het KNMI, een waterschap of een particulier, automatisch gecontroleerd kan worden. Veel processen hangen af van nauwkeurige weersinfromatie, neerslaginformatie in het bijzonder. Naast het maken van nauwkeurige weersverwachtingen op korte termijn, speelt waterbeheer in het algemeen een grote rol: steden moeten vrij blijven van overlopende straten, sloten mogen niet droog komen te staan zonder het risico over te lopen bij zware buien en grondwaterstanden behoren op peil te blijven zodat de drinkwatervoorziening niet in het gedrang komt. In praktijk wordt er echter door het KNMI op relatief weinig plaatsen automatisch neerslag gemeten, met 31 regenmeters verdeeld over heel Nederland. Dit, terwijl neerslag bij buien al op enkele honderden meters afstand significant kan verschillen en direct handelen gewenst kan zijn. Daarom wordt er ook gebruik gemaakt van neerslagradars, die elke 5 minuten voor elke kilometer neerslag kunnen detecteren. Het gecombineerde radarbeeld meet echter een signaal in de lucht en niet hoeveel neerslag er precies aan de grond terechtkomt. Verder zijn er allerlei mogelijke stoorzenders dit het signaal beïnvloeden. Daarom worden de radarbeelden onder andere aangepast op basis van de automatische officiële KNMI regenmeters, maar wordt er ook gekeken hoe andere niet officiële automatische neerslagmeters hiervoor ingezet kunnen worden. Voor het zover is, moet de data hiervan wel automatisch gecontroleerd kunnen worden.

Recent is al onderzoek verschenen waarbij de kwaliteit gecontroleerd wordt door een meting van een regenmeter te vergelijken met metingen van regenmeters uit de omgeving. Blijft een regenmeter droog terwijl anderen neerslag aangeven, dan duidt dat op een fout bijvoorbeeld. Van fabrikant Netatmo zijn er met name in dorpen en steden genoeg regenmeter 'buren' te vinden, maar op het platteland en van andere type neerslagmeters zijn die er lang niet altijd. Daarom is in dit onderzoek een methode ontwikkeld die neerslagmetingen met een direct beschikbaar radarsignaal vergelijkt. Dit radarsignaal bevat niet direct de informatie hoeveel het precies heeft geregend, maar detecteert wel of het hard, zacht of niet regent.

Resultaten laten zien dat deze methode even goede resultaten haalt als de methode die vergelijkt met de omgeving in het geval voor Netatmo regenmeters in Amsterdam. Maar, hiervoor filtert de radarmethode meer data weg dan nodig is. Voor metingen van waterschappen en het KNMI wordt ook altijd de kwaliteit van de data beter na toepassen van de filters. Zeker als de sterkte van de filters wordt aangepast op de eigenschappen van de metingen (bij het KNMI is zelfs de kleinste afwijking aanleiding om een filter toe te passen), kan er automatisch ingegrepen worden bij fouten, waardoor alle metingen die met de juiste tijd- en locatiegegevens worden aangeleverd, in de toekomst gebruikt kunnen worden om real-time in kaart te brengen hoeveel neerslag er daadwerkelijk in Nederland valt.

Contents

1 Introduction

1.1 Monitoring precipitation

Water, especially fresh water, plays an unmissable role in nature, agricultural activities or as drinking water. How much precipitation actually reaches the ground is therefore important to monitor. Irrigation, water management or urban flood forecasts in particular, depend on high resolution measurements, both in space and time [*Berne et al.*, 2004].

References to rainfall monitoring devices can be traced back to the fourth century B.C. in India, long before various types of rain gauges were invented in the 17$^{\text{th}}$ century in Europe [*Strangeways*, 2010]. Despite the urgency of measuring precipitation and the many years of rain gauge developments, approximately 100,000 'official' rain gauges contribute to the Global Precipitation Centre (GPCC) around the world, representing about 1% of the Earth's surface. Of these, only 8,000-12,000 rain gauges of the network of the World Meteorological Organisation (WMO) report in near real-time via Global Telecommunication System (GTS) [*Kidd et al.*, 2017].

Weather radars and satellites are used to obtain real-time precipitation estimates as well. Weather radars are capable of measuring rainfall intensities with high spatial and temporal resolutions [*Overeem et al.*, 2009] and are regarded as a dedicated measuring technique. In recent years, opportunistic indirect precipitation monitoring methods have been investigated, such as rainfall intensities from camera images [*Allamano et al.*, 2015; *Jiang et al.*, 2019], car sensors [*Rabiei et al.*, 2013], commercial microwave links from cellular communication networks [*Overeem et al.*, 2011] and citizen observations collected by smart phones [*Elmore et al.*, 2014; *Guo et al.*, 2019]. Quality control, combined with sophisticated algorithms are all used in order to obtain more rainfall observations. Downside is that often these data have to be adjusted by direct measurements of rain gauges. Therefore, the best estimate of precipitation amounts in space and time is often obtained by the combination of weather radar estimates and rain gauge measurements, if present as in Europe. Since official real-time rain gauge measurements are scarse, the information of third party rain gauges (partially real-time) as a potential source for extra precipitation observations is investigated in recent years [*Jenkins*, 2014; *Bell et al.*, 2015; *De Vos et al.*, 2017]. This thesis focusses on third party measurements of rain as well, in particular for the Netherlands. Since rainfall is the most important type of precipitation in the Netherlands, both terms are used interchangeably.

1.2 Real-time rain gauge networks in the Netherlands

In the Netherlands, the official network of automatic rain gauges of the Royal Netherlands Meteorological Institute (KNMI) consists of 31 rain gauge locations. This equals a density of approximately 1 rain gauge per 1,000 km^2. Gauges are placed in so-called pit gauge configurations: a 40 cm high hill with a 3 m diameter of which the collecting funnel of the gauge is placed in the open middle. An example is shown in figure 1a. An alternative method is shown in figure 1b where a wind screen is placed around the gauge. The precipitation is

collected, melted if needed and measured by a floating device connected to a potentiometer [*KNMI*, 2000]. The data is publicly available every 10 min.

(a) (b)

Figure 1: Different rain gauge setups to minimize wind influence [*Brandsma*, 2014]. (a) KNMI pit gauge setup. (b) Wind screen setup used at some KNMI stations and by several water boards.

Automatic rain gauges of water boards, agricultural collectives or personal weather stations (PWSs) are regarded as third party, nonofficial sources of real-time rain observations. Observation characteristics such as sensor type (and therefore intrinsic quality and measurement techniques), setup, availability of metadata, maintenance and measurement resolution (both in space and time) vary for every observation network source. Even within sources, characteristics are often non-uniform.

It is estimated that nearly a half of all water boards in the Netherlands employes an own network of rain gauges. These networks can consist of multiple types of rain gauges and setups. The most common mechanisms that are used (if this information is known) are floating devices and high-end tipping bucket mechanisms with a resolution between 0.01 mm and 0.2 mm. Detailed information about setup, maintenance and supervision is often scarce. However, if information is present, it is found that the use of pit gauge configurations or wind screens around the sensor is most common.

For agricultural activities, collective networks of rain gauges are on the rise. Most used are tipping bucket mechanisms, mounted on poles near agricultural fields reporting rain measurements every 10 to 30 minutes depending on the network operator. Some operators maintain and inspect gauges, but in most cases this is left to farmers. Rain gauge density estimates differ but are probably comparable to the density of the manual gauge network employed by KNMI volunteers, which is close to 1 gauge every 100 km^2.

PWSs are rain gauges installed and maintained by weather enthusiasts. Every gauge is mounted differently in a mostly urban environment. The manufacturer of the rain gauges is

variable as well, only a part of it is able to upload measurements to platforms as WOW-NL or WeatherUnderground. At 300 to 600 locations in the Netherlands, measurements are performed and uploaded to these platforms at varying time intervals. Another PWS platform is Netatmo, to which only measurements of Netatmo rain gauges can be uploaded, shown in figure 2 [*Netatmo*, 2020a]. The Netatmo rain gauge has a relatively small 13 cm diameter collection funnel with a plastic tipping bucket. It is capable of measuring rainfall of 0.1 mm for intensities between 0.2 mm/h - 150 mm/h at 1 mm/h accuracy according to the manufacturer [*De Vos et al.*, 2017]. However, information about the setup is almost non existent.

A schematic overview of different rain gauge networks with corresponding characteristics is shown in table 1.

Table 1: Overview of several automatic rain gauge networks. The official automatic KNMI rain gauge network is the best described network using high quality measurement techniques. The density of the network is relatively low compared to the third party networks. Information of rain gauges in these networks is however scarce and measurement techniques are often less reliable.

Source	Mechanism	Setup	Area/gauge (km^2)	Interval (min)
KNMI	Floating device	Pit gauge/Screen	1000	10
Water boards	Variable/unkown	Variable/ unknown	50-600	1 to 10
Agriculture	Tipping bucket	1.5 m/variable	100-200	10 to 30
PWS	Tipping bucket	Variable/unknown	75-150	Variable
Netatmo	Tipping bucket	Unknown	10	Variable

Figure 2: Netatmo rain gauge with tipping bucket mechanism, part of the popular Netatmo PWS. Setup, placement and update interval depends per rain gauge [*Netatmo*, 2020b].

1.3 Types of errors

Erroneous measurements by rain gauges could be related to characteristics shown in table 1. Overall, three types can be distinguished: instrumental errors, setup errors and data processing errors [*De Vos et al.*, 2019].

First of all, the instrument itself can be conflicted: instrumental errors can occur. Leaves can block the collecting funnel for example. Insects can cause tipping bucket mechanisms to be stuck and therefore block measurements. In general, malfunction due to dirt is a well-known effect [*Steiner et al.*, 1999]. Instruments without a heating mechanism are not able to perform measurements in snowy and frosty conditions and batteries of battery-powered gauges can run low. Small, shallow collecting funnels tend to underestimate rainfall at high rainfall intensities if rain drops are splashed by the collecting funnels [*De Vos et al.*, 2019].

Also the setup of the sensors introduces errors. Not all rain gauges of the several networks are owned by professionals with profound knowledge of optimal setup locations, leading to suboptimal measurements [*Bell et al.*, 2015]. It is known that wind speed has a significant effect on rainfall measurements. The more wind is present near the opening of the collecting funnel of the rain gauge, the less precipitation actually falls into the collecting funnel. Since the higher above the ground, the more wind is present in open field, higher placed sensors miss more precipitation than sensors that are buried in at ground level. Shadowing effects and turbulence created by nearby objects can lead to undercatch as well [*Pollock et al.*, 2018]. In urban areas setup locations matching official guidelines such as the official KNMI network are scarce and thus setup errors can be expected frequently.

Furthermore, data processing issues may lead to errors. Due to loss of internet connection or a lost connection to a base unit which is connected to the internet, precipitation events since the last measurement interval might be lost.

As a final notice: conditions leading to instrument, setup and processing errors can change in course of time. By definition due to varying weather conditions, but also for rain gauge networks where regular maintenance is not available.

1.4 Quality Control

In order to handle all types of errors, techniques have been developed in order to detect erroneous measurements: Quality Control (QC) methods. The KNMI validates all measurements on a hourly time scale afterwards primarily by hand, by comparing the measurements to radar estimates and to the manual KNMI network. Routines able to automatically detect errors are developed as well. For example by checking whether thresholds of unlikely values are violated [*Estevez et al.*, 2011], or checking for internal consistency between stations and time [*Zahumensky*, 2004; *Chen et al.*, 2018; *Bardossy et al.*, 2021]. For temperature measurements QC is performed with seven layers of quality control using metadata and comparison with nearby stations by *Napoly et al.* [2018]. The open source project TITAN [*Baserud et al.*,

2020] presents a QC method with 12 sequential checks using metadata, climatology, elevation data and a comparison with nearby stations. This algorithm is applicable to both temperature and precipitation.

Especially a real-time applicable QC method was proposed by *De Vos et al.* [2019] (from now on this method is referred to as 'QC De Vos'). Without any metadata (except location) of rain gauges part of a network, it is able to exclude inaccurate observations by comparing each observation with observations of nearby gauges. This QC method was developed using PWS data and it has been shown that the method is able to improve the overall accuracy of 2 years of Netatmo rain gauge measurements in Amsterdam. On national scale, one month of Netatmo rain gauge data was quality controlled and used to construct rainfall maps as complement to the official KNMI network measurements [*De Vos et al.*, 2019].

1.5 Objective and outline

However, the QC has not been applied to a national data set for a period longer than a month and only for the Netatmo network as source of data. Therefore, in this study, the main question to answer is: *"Is it possible to develop a quality control method based on the existing method developed by De Vos et al. [2019], that is able to detect and remove erroneous measurements for real-time rain gauge networks in the Netherlands for a period of at least one year?"* It is hypothesized that it is possible by extension of the existing method and the use of auxiliary real-time radar data, given the lower density of most rain gauge networks compared to the Netatmo rain gauge network.

In order to answer the main question, this study proposes a new QC method which uses real-time radar data, based on QC De Vos. The details of both methods are discussed in section 2. It is for the first time that a QC method is tested for various sources of rainfall measurements, by taking national third party data of Netatmo PWSs and waterboards and official measurements of KNMI into account. Therefore, the proposed QC method is tested on unprecedented large data sets, further discussed in section 2 as well. Results will be presented in section 3, followed by a discussion in section 4. Finally, in section 5, conclusions will be formulated.

2 Methodology

In order to answer the research question, it is necessary to test whether the developed QC algorithms are able to detect erroneous measurements for the entire spectrum of third party rain gauge networks as listed in table 1. First, the quality control algorithms are introduced. Next, data of as many as possible networks, both dense and sparse, is requested and processed such that the algorithms can be applied to the data.

2.1 QC De Vos algorithm

As mentioned in section 1, QC De Vos is a QC algorithm that detects and removes erroneous measurements by comparing rain gauge values to measurements of nearby rain gauges. Originally three, but here four different possible inaccuracies are detected: Faulty Zeros, High Influxes, Station Outliers and Bias Outliers. For every (constant) time interval of a given data set with measurements, each filter can flag the particular time interval for each rain gauge independently as error (code 1), no error (code 0) or not enough information (code -1). The original code used in *De Vos et al.* [2019] is publicly available via *De Vos* [2021].

2.1.1 Faulty Zero filter

A rain gauge can falsely report no rain even during rain events. This can be both due to an instrument error (obstruction) or a data processing error (lost connection to a base unit). In order to detect such a Faulty Zero (FZ), first the max_{no} nearest rain gauges to a given gauge j at time interval t are selected. Furthermore, only the rain gauges within a maximal range d to rain gauge j are taken into account. Note that the use of the limit of max_{no} nearest rain gauges is an addition to the code of the QC used in *De Vos et al.* [2019]. To determine the FZ filter, the median value of the selected neighbouring gauges is computed, but only if the number of neighbouring gauges is more than n_{stat}. If less than n_{stat} neighbouring gauges are present, the filter produces a FZ flag of -1. Otherwise, if the rain gauge j reports no rain in the last n_{int} intervals and if during those intervals, the median value is larger than zero, the measurement t of rain gauge j receives a FZ flag of 1. This, until any rain is reported again. Other situations are not flagged (FZ is set to 0).

2.1.2 High Influx filter

Unrealistically high values are regarded as High Influx (HI) measurements. If the setup location is near irrigation activities, sprinkler installations or swimming pools, there is the possibility of measuring large amounts of water unrelated to precipitation. To detect these High Influxes, first the neighbouring gauges are selected as in the FZ filter. The HI flag is set to -1 if less than n_{stat} gauges are available. If the median value of the neighbouring locations is less than ϕ_{A} and gauge j at time t reports more than ϕ_{B} rain, the HI flag equals code 1. If the median is more than ϕ_{A} and gauge j reports more than this median value times $\phi_{\text{B}}/\phi_{\text{A}}$, the HI flag is set to 1. Other situations are not flagged (HI is set to 0).

2.1.3 Station Outlier filter

Other outliers only appear over time which is the case if gauges repeatedly report deviating measurements compared to the environment. These outliers are Stations Outliers (SO). This filter is determined every time interval t for every rain gauge j by taking into account a variable period $P_{t,j}$ of at least m_{int} intervals where rain gauge j reports m_{rain} intervals of rain, FZ and HI flagged intervals excluded.

Of all neighbouring rain gauges of gauge j, it is checked whether they report rain during at least m_{match} in period $P_{t,j}$. If more than n_{stat} neighbouring gauges match this criterion, the SO flag can be calculated, otherwise the SO flag for gauge j is set to -1. The SO flag is also set to -1 if the rain gauge itself reports a non-numeric value at interval t.

For every gauge j the correlation r over period $P_{t,j}$ with each selected neighbour is calculated. If the median value of all r-values falls short to threshold γ, the SO flag is equal 1, otherwise the SO flag is set to 0.

2.1.4 Bias correction and Bias Outlier filter

Most rain gauges tend to underestimate rainfall systematically [*Brandsma*, 2014; *Bell et al.*, 2015; *De Vos et al.*, 2017, 2019]. As discussed, surroundings, setup and wind, affect measurements. Since these change in time, the bias of rain gauges relative to the actual amount of precipitation is partially variable in time as well.

To include the systematic bias of networks of rain gauges, a default bias correction factor (DBC) is determined a priori and applied to all measurements. One can base this factor based on calibration reports or by comparing a period of measurements of other sources.

The variability of the bias of each individual rain gauge within a network, is incorporated by calculating the so-called bias correction factor (BCF) every time interval t. At $t = 0$, BCF is equal to DBC for all gauges. For other time intervals the BCF value of gauge j is based on a comparison with neighbouring stations in case the SO flag of gauge j was set to 0. For these time intervals t, the relative bias between rain gauge j and neighbouring gauges is calculated using all measurements in period $P_{t,j}$:

$$\text{bias}_{t,j,i} = \sum_{k=1}^{N_P} \frac{(R_{\text{gauge}_j} - R_{\text{gauge}_i})_k}{R_{\text{gauge}_{i,k}}}, \tag{1}$$

where $bias_{t,j,i}$ is the relative bias between gauge j and neighbouring gauge i calculated at time interval t. N_P is the number of time intervals within period $P_{t,j}$ over which the bias is calculated using time series with measurements R_{gauge}. From this, a new bias correction factor is determined:

$$BCF_{\text{new}_{t,j}} = \frac{1}{1 + \text{median}(\text{bias}_{t,j,i})}, \tag{2}$$

where i again numbers the neighbouring gauges of gauge j.

BCF_{new} replaces the bias correction factor for rain gauge j onwards from time interval t if $|\log(BCF_{\text{new}_{t,j}}/BCF_{t-1,j})| > |\log(1+\beta)|$ where β is a parameter. Then $BCF_{t,j} = BCF_{\text{new}}$, otherwise $BCF_{t,j} = BCF_{t-1,j}$.

Note that this method implies that DBC times the median of the rain gauge measurements in period P within range d, is equal to the actual rainfall amount at its location. Furthermore, it is assumed that this median value is representative for the entire area within range d at time interval t. Hence, this physically limits the choice of parameter d.

Finally, in addition to *De Vos et al.* [2019], the Bias Outlier (BO) flag is calculated. The BO filter reports 1 for measurements at time t of rain gauge j if $|\log BCF_{t,j}| > |\log 2|$. Despite these measurements have passed the FI, HI and SO filter, the measurements of rain gauge j should be at least doubled or halved to be in line with the median of neighbouring gauges and are therefore flagged.

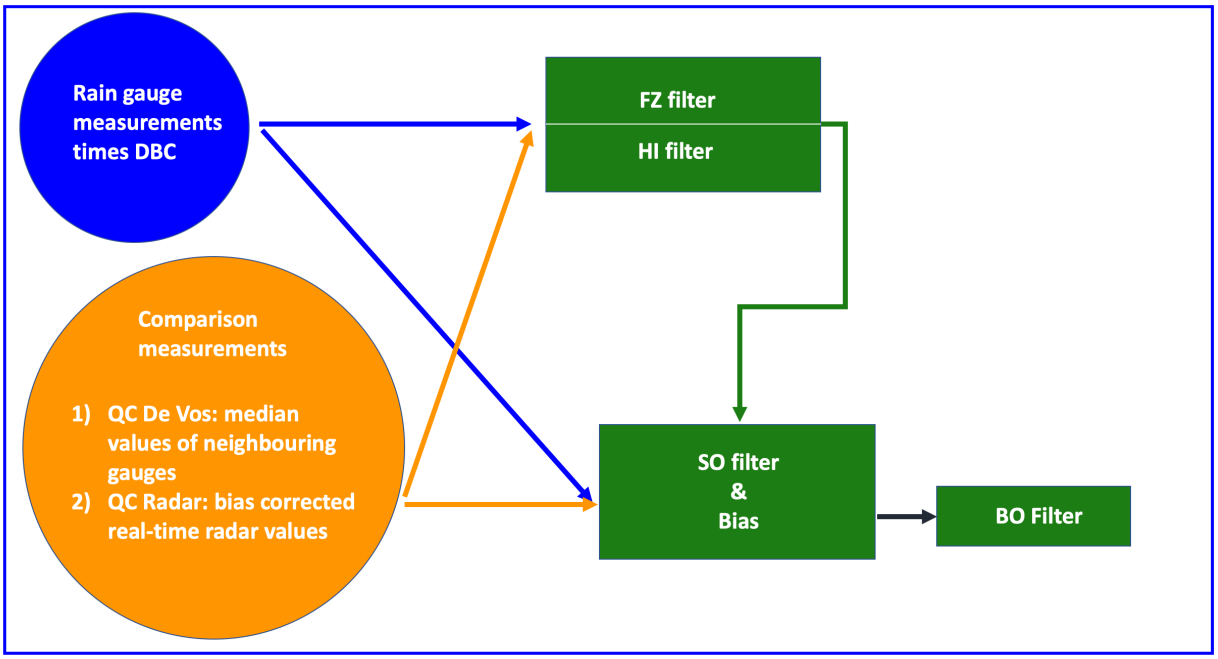


Figure 3: Schematic model overview of both QC versions. Both versions first perform the FZ and HI filter, then the SO Filter and bias correction factor, on which the BO filter is based as final step. QC De Vos uses the median value of neighbouring gauges te compare with, QC Radar uses bias corrected real-time available radar values.

2.2 QC Radar algorithm

The newly proposed QC algorithm uses real-time available weather radar values in order to quality control rain gauge data and is therefore referred to as QC Radar. The code is available via *Van Andel* [2021]. The filter design is practically identical to QC De Vos. In short, the comparison to real-time radar values at a rain gauge location replaces the comparison to the measurements of neighbouring rain gauges, after an initial adjustment of the real-time available weather radar values in QC Radar. This is needed since rainfall amounts based on weather radar estimates do not grasp actual rainfall amounts. Both algorithms can be represented as in figure 3.

2.2.1 Initial radar adjustment

First, the real-time available radar values at each gauge location are adjusted. For this, auxiliary measurements such as official rain gauges could serve. Here, the uncorrected rain gauge data (before QC) is used of which the procedure is described in this section.

In order to incorporate the systematic biases of both the uncorrected rain gauge data and the radar values, a priori calculated DBC-values are applied to the data both for the rain gauges as for the real-time radar values. Then, for each rain gauge i at time interval t, the bias between the gauge and the corresponding real-time radar value is determined:

$$\text{bias}_{t,i} = \sum_{k=1}^{N_P} \frac{(R_{\text{radar}_i} - R_{\text{gauge}_i})_k}{R_{\text{gauge}_{i,k}}}, \tag{3}$$

over a period $P_{\text{radar}_{t,j}}$. $P_{\text{radar}_{t,j}}$ is a variable period of m_{int} intervals or a longer period where the radar at the location of rain gauge j reports m_{rain} intervals. The bias correction factor for the real-time radar value at j then depends on the median of all radar-sensor biases within range d_{radar}:

$$BCF_{\text{radar}_{t,j}} = \frac{1}{1 + \text{median}(\text{bias}_{t,i})}, \tag{4}$$

where i again numbers the neighbouring gauges of gauge j.

Note that this method assumes that the median of the rain gauge measurements within range d_{radar} of rain gauge j multiplied by the DBC-value, is representative to the actual rainfall amount at its location. Then, the bias between this median value and the value measured by the real-time available radar equals the bias of the radar relative to the actual rainfall amount for that location at time t. Secondly, it is assumed that the bias of the real-time available is constant within range d_{radar} of rain gauge j. If this is true, the bias of the real-time available radar in this area is known at time interval t. Therefore, range d_{radar} has to be chosen such that the bias of the radar values to actual rainfall amounts is constant within this area. Note that this criterion is not strict, since this adjustment does not necessarily aim for obtaining

a perfectly adjusted radar product. It should assure that rain fall accumulations of the real time radar set become in line with accumulations of the median value of the rain gauge network, so that the same parameters and threshold values of QC De Vos directly can be used in QC Radar.

2.2.2 Faulty Zero filter

A rain gauge measurement at interval t of rain gauge j times DBC is flagged as 1 if the gauge reports no rain for n_{int} time intervals and the corresponding bias corrected real-time radar value does. This, until any rain is reported again. Only if the radar value is non existent, the FZ filters reports -1. All other cases are flagged as code 0.

2.2.3 High Influx filter

If the bias corrected real-time radar value at location of rain gauge j is less than ϕ_{A}, and gauge j reports more than ϕ_{B} rain, the HI flag is 1. If the radar value is more than ϕ_{A}, the HI flag is set to 1 if gauge j reports more than the radar value times $\phi_{\text{B}}/\phi_{\text{A}}$, otherwise the HI filter flags 0.

2.2.4 Station Outlier filter

The SO filter reports -1 if the rain gauge itself does not report a measurement at interval t. Furthermore, the SO flag is determined every time interval t for every rain gauge j by taking into account a period $P_{t,j}$ of at least m_{int} intervals where j reports m_{rain} rainy intervals (FZ and HI flagged intervals excluded). For the same period, the real-time radar value is checked whether during m_{match} intervals rain is reported, otherwise the SO flag becomes -1 as well. If the correlation r between the rain gauge and the radar value is less than γ, the SO flag is set to 1, otherwise the SO flag remains 0.

2.2.5 Bias correction and Bias Outlier filter

For time intervals t where the SO flag reports 0, the bias between the sensor and the radar value in period $P_{t,j}$ is calculated as well and BCF_{new} is calculated. If BCF_{new} significantly changes from a previously calculated value, BCF_{new} replaces the used bias correction factor for rain gauge j since time interval t. This criterion is expressed such, that if $|\log(BCF_{\text{new}}/BCF_{t-1,j})| > |\log(1+\beta)|$ where β is a parameter, then $BCF_{t,j} = BCF_{\text{new}}$, otherwise $BCF_{t,j} = BCF_{t-1,j}$. The BO flag then reports 1 for measurements at time t of rain gauge j if $|\log BCF_{t,j}| > |\log 2|$.

2.3 Data sets rain gauge networks

2.3.1 Netatmo Amsterdam

First of all, the urban data set described and processed by *De Vos et al.* [2019] is used available via *De Vos* [2019], the NetatmoAMS data set. It contains a time series of 5-min interval precipitation measurements of Netatmo PWS rain gauges between 1 May 2016 00:00 UTC and 1 June 2018 00:00 UTC of 134 rain gauges in the Amsterdam metropolitan area, which is defined in *De Vos et al.* [2019]. This averages a density of almost 1 per 4 km^2. On average, in 69.7% of all 5-min intervals a measurement is available.

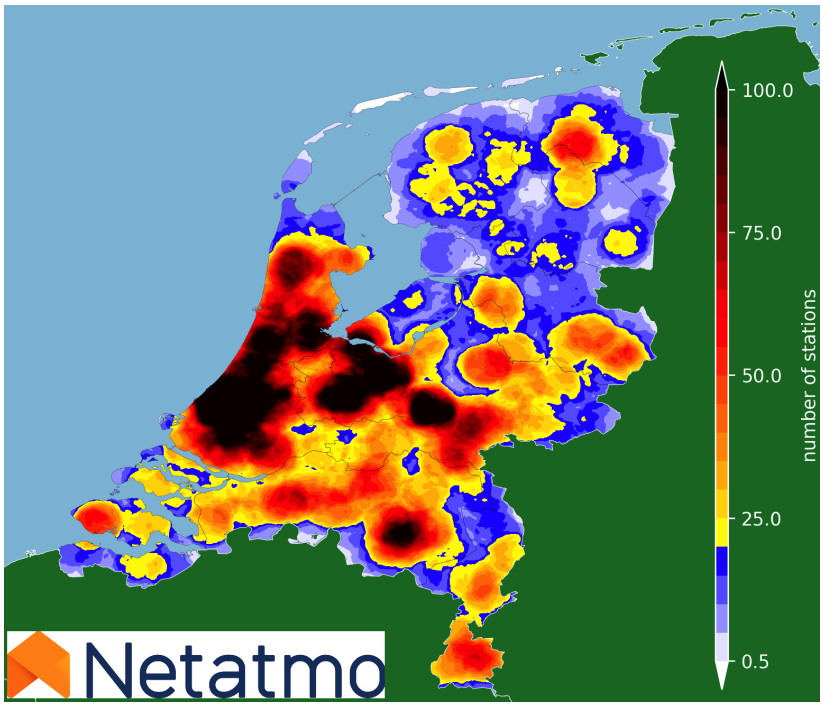


Figure 4: Netatmo rain gauge network in the Netherlands. Here, the number of rain gauges within a 10-km radius is shown for the period between 1 September 2019 and 1 September 2020. Highest density is found in urban areas, whereas in some rural areas, rain gauges are scarce [*Netatmo*, 2020b].

2.3.2 Netatmo network the Netherlands

A second data set of Netatmo observations is constructed on the national scale of the Netherlands (NetatmoNL). Between 1 September 2019 00:00 UTC and 1 September 2020 00:00 UTC, all measurements of rain gauges in the Netherlands are stored in a 5-min interval time series. On average, in 74.7% of the time a rain gauge reports measurements. This includes rain gauges that are installed within the period of the data set. The average rain gauge density is almost 1 per 10 km^2. The density however is highly variable and highest in urban regions, whereas the density in rural areas drops to 1 per 100 km^2 regionally. An overview of the variability of the Netatmo rain gauge network density is shown in figure 4.

2.3.3 Water board network

Also, data is collected from Dutch water boards. Of 6 of the 21 water boards in the Netherlands, data is available for this study: Wetterskip Fryslân (FRY), Waterschap Hunze en Aa's (HA), Hoogheemraadschap Hollands Noorderkwartier (HHNK), Waterschap Stichtse Rijnlanden (SRIJN), Waterschap Drents-Overijsselse Delta (WDOD) and Waterschap Rijn en IJssel (WRIJ). Data delivery, storage, interval and rain gauge type is not uniform for all water boards. Therefore, for each water board a data set of 5-min interval time series is constructed, for the period between 1 May 2016 00:00 UTC and 1 June 2018 00:00 UTC, following the NetatmoAMS data set. Exception are the HHNK data set which is constructed for the years 2019 and 2020 and the WRIJ data set, which is constructed at a 10-min interval due to the lack of 5-min data. A complete overview of all water board data sets is shown in table 2.

Table 2: Overview of information known about water board networks. For HHNK both floating devices as tipping buckets are present as measurement mechanism. WS stands for wind screen and PG for a pit gauge setup. 'Dens.' stands for rain gauge network density in number of gauges per area. Int.=Interval, Res.=Resolution. Avail.= Availability of data within the period of each individual data set.

Source	Mechanism	Setup	Dens. (km^{-2})	Int. (min)	Res. (mm)	Avail.
HHNK	Variable	WS/ ?	1/49	5	0.05	77.5%
WDOD	?	?	1/150	5	0.1	99.8%
FRY	Floating device	PG/WS	1/591	5	0.01	98.9%
HA	?	?	1/106	5	0.2	70.1%
SRIJN	?	?	1/200	5	0.1	45.4%
WRIJ	?	?	1/82	10	0.2	99.0%

2.3.4 KNMI network

Measurements of the 31 KNMI locations with 40 automatic rain gauges (consisting of the official operation network and rain gauges for research) are used to construct two data sets: an unvalidated 10-min time series between 1 May 2016 and 1 June 2018 (90.8% data availability) and an hourly time series within that period with validated measurements by KNMI itself. The unvalidated data set serves as input for the QC Radar calculation.

2.4 Data set real-time weather radar

In order to be able to develop and test the QC radar algorithm, a data set of radar data is chosen that represents a radar data set which would have been available in real-time.

This radar product contains rainfall estimates on a 1-km grid in the Netherlands for every 5 minutes, based on a reflectivity composite at 1,500 m height, using the Z-R relationship of $Z = 200R^{1.6}$. The reflectivity values were measured by two weather radars, before September 2016 by two C-band radars in De Bilt and in Den Helder, after January 2017 by dual polarized radars in Herwijnen and Den Helder. The reflectivity values had undergone Doppler

filtering and non-meteorological echo removal using cloud mask satellite data [*Overeem et al.*, 2011]. The data is freely accessible via [*Overeem and Imhoff*, 2020].

As input in the QC Radar algorithm, for every time interval, for every rain gauge location, the average precipitation value of the corresponding radar grid cell and the 8 directly surrounding grid cell values is stored.

2.5 Data set reference weather radar

As reference, a 1-km^2 weather radar data set covering the Netherlands every 5 minutes is used. The precipitation amounts have been adjusted using the data from the official automatic rain gauges of KNMI and from the 325 official manual rain gauges on a daily basis by KNMI. This data set is publicly available with a delay of several months [*Overeem*, 2020]. It is this data set that is used to determine DBC-values. Furthermore, the results of the quality controlled data sets are compared to this reference data set, since it is regarded as the best available estimate of rainfall amounts in both space and time. For this, for each gauge location j the precipitation amount of the overlying radar pixel is stored for each time interval of the rain gauge data set.

2.6 Algorithm parameters

The algorithm parameters have been set as in *De Vos et al.* [2019] for 5-min data sets and are all listed in table 3. For 10-min data sets some parameters have been adjusted. In contrast to *De Vos et al.* [2019], the nearest neighbours within a maximum range of $d = 10,000$ m are limited to the $max_{\text{no}} = 20$ nearest rain gauges. This, in order to reduce computing time. Extrapolated from *Van de Beek et al.* [2012], the present choice of d matches the correlation length of precipitation in convective situations. Preliminary analysis of the real-time weather radar data showed that the bias of this radar product is regionally constant. Therefore, d_{radar} is set to 50,000 m. Furthermore, DBC-values were computed for each individual data set of rain gauges.

2.7 Test cases

As basis test case, both QC algorithms are applied to the Netatmo Amsterdam (NetatmoAMS) data set described in section 2.3.1. As in *De Vos et al.* [2019], the DBC-value is set to 1.24. This was determined by a comparison of rainfall amounts of May 2016 relative to the reference gauge adjusted radar set described in section 2.5. For this, for each sensor the bias with the overlying radar pixel has been determined and hence $DBC_{\text{sensor}} = 1/(1 + \text{median(bias)})$. The DBC-value of the real-time available radar data set described in section 2.4 for the Amsterdam region was calculated using the same procedure. Since the data between 1 May 2016 and 1 June 2017 was used to design QC De Vos (calibration data set 2017), results are calculated for this period separately from the data starting at 1 June 2017 up to 1 June 2018 (validation data set 2018). Due to the addition of parameter max_{no} and the introduction of the BO filter, differences can be expected with *De Vos et al.* [2019].

Table 3: List of algorithm parameters.

Parameter	Algorithm	Value (5 min)	Value (10 min)
d (m)	QC De Vos	10,000	10,000
d_{radar} (m)	QC Radar	50,000	50,000
max_{no}	QC De Vos	20	20
n_{stat}	QC De Vos	5	5
n_{int}	both	6	6
ϕ_{A} (mm)	both	0.4	0.4
ϕ_{B} (mm)	both	10	10
m_{int}	both	4,032	2,016
m_{rain}	both	100	50
m_{match}	both	200	100
γ	both	0.15	0.15
β	both	0.2	0.2
DBC_{radar}	QC Radar	variable	variable
DBC_{sensor}	both	variable	variable

The results are therefore compared to the results of *De Vos et al.* [2019]. The difference between the performances of both QCs are compared as well.

Consecutively the performance of the QC methods is tested using the KNMI data set described in section 2.3.4. This data should not contain Station Outliers and should not have variable bias due to the regular maintenance and the well-described metadata of the location and setup. This experiment is performed using QC Radar only, since network densities would demand non-physically high values of range d if QC De Vos was used. For this, the 10-min parameter settings are used.

Table 4: Overview of test cases for quality control. Case 0 refers to Netatmo PWS data from Amsterdam, case 1 refers to rain gauge data from KNMI's automatic rain gauge network, case 2 to data of 6 water boards and case 3 to data of the national Netatmo PWS data.

Cases	Data set	DBC_{sensor}	DBC_{radar}
Case0	NetatmoAMS	1.24	1.42
Case1	KNMI AWS	1.11	1.63
Case2a	HHNK	1.12	2.55
Case2b	WDOD	1.22	2.02
Case2c	FRY	0.95	1.84
Case2d	HA	0.78	1.54
Case2e	SRIJN	1.13	1.68
Case2f	WRIJ	1.29	1.64
Case3	NetatmoNL	1.09	2.24

In order to analyse the performance of the QC method for sparse networks that are not necessarily well-described, QC Radar is also applied to the data sets of the water board networks described in section 2.3.3. For WSRIJ the 10-min parameter settings are used, the other data sets have been processed using the 5-min parameter settings. DBC-values are determined again based on a comparison with the reference gauge adjusted radar data set using the first month of each data set. Results are based on the other months of the data set. DBC-values per data set are listed in table 4.

Finally, QC radar is tested using the NetatmoNL data set described in section 2.3.2, in order to determine the performance for a network with varying network density in time and space. The network DBC-value is determined based on the measurements of September 2019 by comparison to the reference gauge adjusted radar set: $DBC_{\text{sensor}} = 1.09$, $DBC_{\text{radar}} = 2.24$. Results will be analysed for October 2019 to September 2020.

2.8 Validation

The QC performances are measured by comparing the data set after application of the filters to the reference weather radar data set using the following metrics: the Pearson correlation coefficient (r)

$$r = \frac{\text{cov}(R_{\text{gauge}}, R_{\text{ref}})}{\sigma(R_{\text{gauge}})\sigma(R_{\text{ref}})}, \tag{5}$$

where R_{gauge} is the rain gauge network time series and R_{ref} the constructed reference weather radar time series, cov the covariance and σ the standard deviation, the relative bias (called bias from now on)

$$\text{bias} = \sum_{j=1}^{N} \frac{(R_{\text{gauge}} - R_{\text{ref}})_j}{R_{ref,j}}, \tag{6}$$

and coefficient of variation of the errors (CV)

$$\text{CV} = \frac{\sigma(R_{\text{gauge}} - R_{\text{ref}})}{\frac{1}{N}\sum_{j=1}^{N} R_{ref,j}}, \tag{7}$$

with R_{ref} the radar data set described in section 2.5. As remark, since R_{ref} is based on radar observations a gauge-pixel discrepancy is always present. Using 12-s resolution rain gauge observations, it is shown by *De Vos et al.* [2019] that the best metrics values that can be achieved are approximately 0.75 in case of r and a CV-value of approximately 5.

Filters can be applied to the data set in several ways. 'Flex filtering' refers to application of FZ, HI and SO filter for measurements where one of these filters is set to 1 (Error). 'Strict filtering' also excludes measurements with filter flag -1 (not enough information available in order to calculate a filter). Applying all filters strict combined with the BO filter, is referred to as 'All filtering'. For Flex, Strict and All, bias correction factors are applied as well. Finally, for all filters the number of measurements flagged relative to the number of initial measurements is calculated.

3 Results

3.1 Case 0: Netatmo Amsterdam data

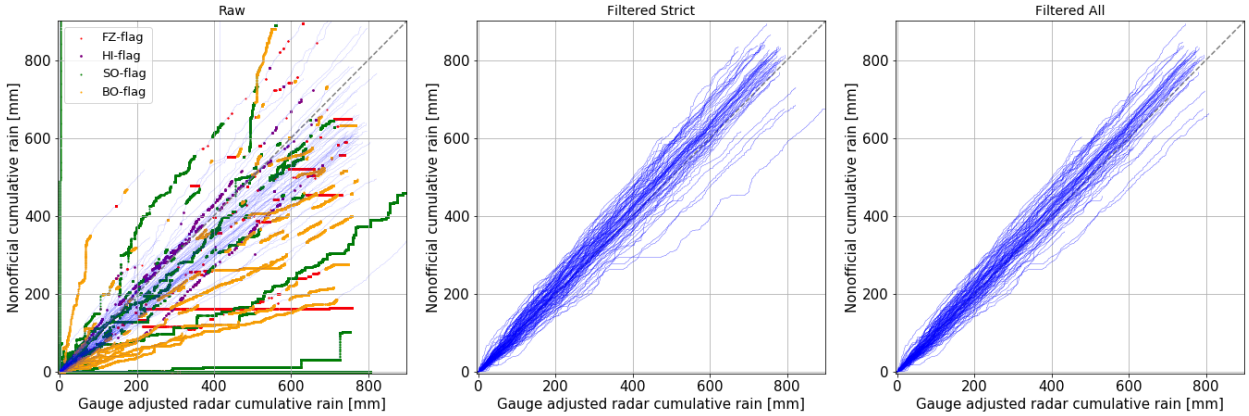
3.1.1 QC De Vos

The results of QC De Vos for the Netatmo Amsterdam data set for 2017 are shown in figure 5a. The x-axis shows the cumulative precipitation of the reference weather radar data set, the y-axis the cumulative rain gauge measurements. Every blue line on the left panel represents the cumulative raw data. The middle panel shows the rain gauge data after applying the Strict filtering method (FZ, HI, SO and bias correction), the right panel shows the data after filtering 'All'. Both are much more in line with the reference radar data. This also yields for the 'Flex' filtering method, however more outliers remain present resulting in a less robust data set based on the calculated metrics. These metrics and the metrics for other filtering methods, can be found in table 5 for calibration data set 2017 and in table 6 for validation data set 2018, both in Appendix A. All metrics are calculated based on the filtered 5-min interval measurements.

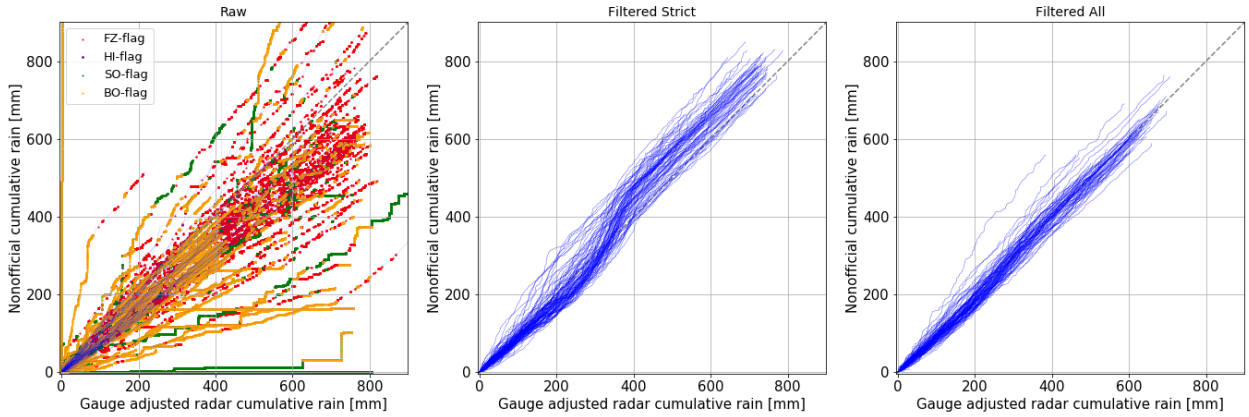
In figure 5a, measurements for which the FZ, HI, SO or BO filter reports code 1 are indicated with red, purple, green and orange dots respectively in the left panel. FZ flags are present if the cumulative reference radar value increases while the rain gauge reports no rain (in 4.7% of all measurements). Applying the FZ filter only to the raw data, the metrics do not improve significantly as is shown for r in figure 6. The HI flags are barely present (0.02%), only a few cases exceed the thresholds. However, filtering measurements for which the HI filter reports 1, results in an improvement of all metrics. The raw bias of 1.39 reduces to -0.06, CV of 147 to 12.9 and r increases from 0.04 to 0.35.

Application of the Station Outlier filter only improves the bias to 0.01, CV to 17.9 and r to 0.27, for which 9.4% is filtered. In *De Vos et al.* [2019], where max_{no} was not set, the SO filter flagged more often, in 11.7% of all measurements. This, since in the present study rain gauges are compared only to the 20 nearest neighbours instead of all rain gauges within 10 km as in *De Vos et al.* [2019]. The more nearby rain gauges are, the more the corresponding rain gauge measurements are correlated and therefore less often violate the threshold for the SO filter.

Situations in which the SO filter reports an error, can be divided into multiple categories. The first case is indicated between the points (200,400) and (400,600) in the left panel of figure 5a. Despite the accumulation of precipitation of this very gauge is equal to the reference radar pixel, the timing of the precipitation deviates from the nearest neighbours resulting in a low correlation. The SO filter is also reported in apparent FZ situations, for which the FZ filter does not flag all cases. This can be the case if a rain gauge reports rain from time to time, but only in very small amounts. The same principle holds for apparent HI situations, for example the almost vertical line at the very left of the left panel of figure 5a. Here, a gauge reports amounts of rain which repeatedly fall short of the HI threshold but therefore result in a low correlation with neighbouring gauges.



(a) QC De Vos



(b) QC Radar

Figure 5: Results of both QC methods a) QC De Vos and b) QC Radar for the Netatmo Amsterdam calibration data set 2017. The x-axis shows the cumulative precipitation of the reference weather radar data set, the y-axis the cumulative rain gauge measurements. Every blue line on the left panel represents the cumulative raw data. The middle panel shows the rain gauge data after applying the Strict filtering method (FZ, HI, SO and bias correction), the right panel shows the data after All filters, which are both much more in line with the reference radar data. QC Radar filters more measurements than QC De Vos.

Applying the bias correction factor calculated by the QC method to the residual measurements, forces precipitation amounts closer to the median of the data set. Since the (systematic) bias between this median and the reference radar is a priori corrected by the DBC-value, applying the calculated (dynamic) bias correction factor results in precipitations amounts close to the 1:1-line as shown in the middle panel of figure 5a. The data shown in this panel (Strict filtering) has 11.2% of the measurements removed for which the filters reported 1 or -1, yielding a bias of 0.05, CV of 8.9 and r of 0.59. In case of flex filtering, 2%-point less data is filtered, yielding comparable metrics. The metrics are in line with the results of *De Vos et al.* [2019].

Additionally, the BO filter can be applied as well as addition to *De Vos et al.* [2019]. Results after applying the BO filter are shown in the right panel of figure 5a. The BO filter flags 6.7% of all measurements, of which 5.8%-point is not filtered by FZ, HI or SO filter. This BO filter reduces the bias to 0.04, CV to 8.4 and increases r to 0.61 for the data set of 2017. Compared to the SO filter, the BO filter is by definition less sensitive to differences per time interval, and more sensitive to deviations in total rainfall accumulations in the period P over which both filters are calculated. To recall, using the present parameter settings this period is two weeks or longer.

Analysing the 2018 data as well leads to the same or even better metrics. Approximately 17% is filtered, leading to a data set with a bias of 0.11, CV of 7.4 and r of 0.59. Compared to the amount of flags, also in 2018 the HI filter improves the metrics the most. Visualisation of the 2018 data set results can be found in figure 18 in Appendix B.

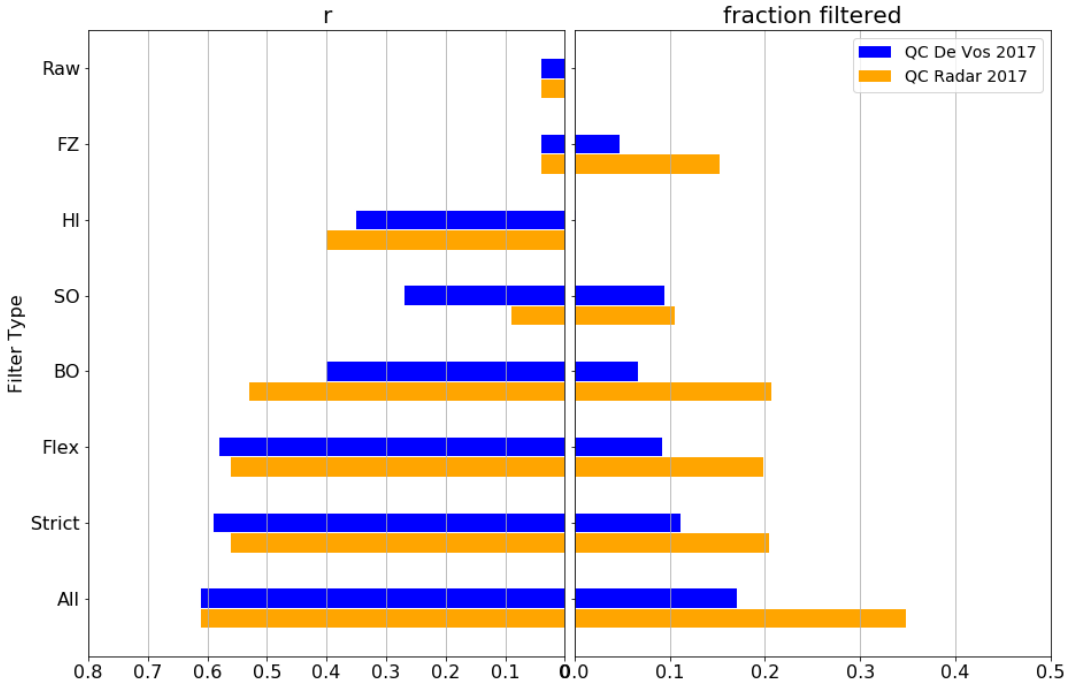


Figure 6: Correlation of the quality controlled data sets with the reference radar data set calculated at 5-min intervals (left) and fraction filtered (right) for calibration data set 2017. One can conclude directly that the HI filter is the most efficient, with almost no filtered data and still improving r significantly. The FZ filter does not improve the quality of the data a lot. QC Radar however filters 10%-point more data at the FZ filter. Residual outliers are most of the time detected by the SO or BO filter respectively, depending on the nature of the outlier and the QC method. Overall, both QC methods improve the correlation. QC Radar filters more measurements than QC De Vos, which is therefore the more efficient method to improve the data quality of these data sets. All metrics can be found in table 5 in Appendix A.

3.1.2 QC Radar

The first output of QC Radar is the adjustment of the real-time available radar data set. The result for 2017 is shown in figure 7. The x-axis shows the cumulative reference radar data described in section 2.5, the y-axis the raw real-time radar cumulation of the data set described in section 2.4. At the panel on the left, the data is shown before initial adjustment, the right hand side shows the data after adjustment. The raw data of 5-min time interval measurements underestimate rainfall accumulations by almost 50%. This underestimation is however variable in time, near $x = 100, 400, 600$ mm those variations can be observed and are linked to seasonal effects rather than the change of weather radar setup in this period. The algorithm applies the pre-determined default bias correction and furthermore is able to reduce the bias variability near $x = 100, 600$ mm by adapting the bias correction factor. Since the algorithm takes into account at least two weeks of past precipitation, a delay in the bias correction factor is present. Near 400 mm, within the period of adaptation, approximately 100 mm of precipitation fell over Amsterdam, resulting in the fact that the bias variability is still present in the bias corrected real-time available radar data.

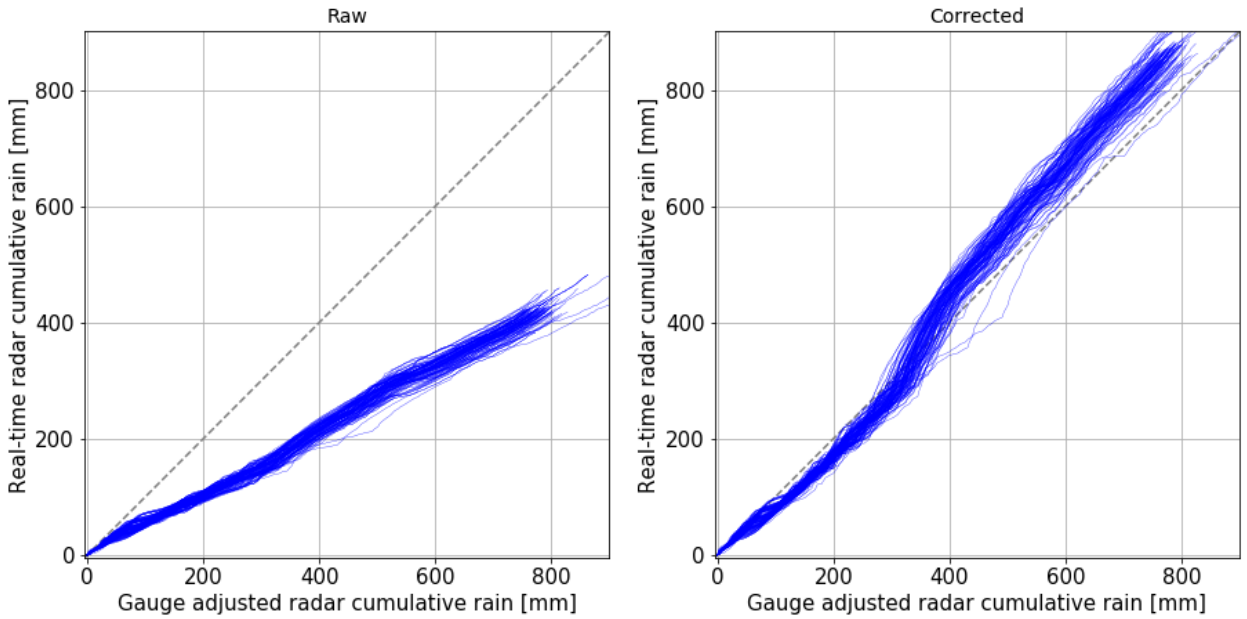


Figure 7: Raw radar adjustment by QC Radar, on the left, the raw cumulative real-time available radar data for all rain gauge locations in Amsterdam for data set 2017. At the right hand side, the bias corrected cumulative real-time radar data is shown.

Results of applying QC Radar to the Netatmo Amsterdam data set, are visualised in figure 5b. Calculation of the metrics are listed in table 5 for calibration data set 2017 and in table 6 for validation data set 2018, both in Appendix A. Results for r in combination with the amount of filtering, are explicitly shown in figure 6. This, since if a higher correlation is realised, often values for bias and CV improve as well.

In figure 5b, the FZ filter is directly remarked and flags 15.2% of all measurements in 2017 (16.2% in 2018) which is significantly more than QC De Vos (4.7% in 2017, 6% in 2018). This, since the real-time radar data contains 5-min precipitation accumulations less than 0.1 mm, i.e., the resolution of the Netatmo rain gauge. This understandably does not effect the calculation of the HI filter, which is again only applied on a few measurements, increasing the data set quality from $r = 0.04$ to $r = 0.40$, CV from 147 to 11.4 and bias from 1.39 to -0.09 for 2017. The SO filter flags 10.6% of all measurements which results in a data set with a bias of 0.73, CV 57.1 and r of 0.09.

The BO filter however filters 20.7% of the measurements, yielding a bias of -0.16, CV of 8.4 and r of 0.53, yet filtering cases not detected by the SO filter in QC Radar. To recall, the BO filter filters measurements for which the bias between the gauge measurements and the adjusted real-time available radar data exceeds the threshold. Therefore measurements can be filtered as well if the adjusted real-time available radar data itself is still biased rather than the measurements. For example in the 2017 data shown in the left panel of figure 5b, the data between $x = 300$ and $x = 400$ is flagged by the BO filter more often since this coincides with the period in which the radar data is shown to be less optimal than other periods. In 2018, the variability in the adjusted real-time radar data is less and the BO filter filters only 13.3% of the measurements.

Combining some or all filters improve the quality of the data set as can be seen in figure 6. Due to the almost continuous availability of the raw real-time available radar data, the difference between filtering 'Flex' or 'Strict' is almost vanishing. The strict application of the FZ, HI, SO filter combined with the bias correction factor results in an improvement of the bias from 1.39 to 0.09, CV of 147 to 9.5 and r from 0.04 to 0.56 by filtering 20.4% of the measurements for 2017. Adding the BO filter (All-filtering), a bias of 0.01, CV of 7.9 and r of 0.61 is achieved, filtering 34.8%. The middle panel in figure 5b clearly shows that the gauge data is bias corrected using the real-time available radar data showing the same bias variability compared to the reference radar data set (figure 5b). Applying the BO filter removes this effect. For 2018, applying the BO filter only filters an extra 7% of the data.

Overall, both QC methods are able to improve the Netatmo Amsterdam data set to bias, CV and r values that are almost the same. QC Radar detects more FZs which is the main reason more data is filtered in QC Radar. Even more data can be unnecessarily filtered by QC Radar if the BO filter is applied due to the bias of the radar as well. The better the real-time available radar data set, the more these results would resemble QC De Vos. These QC De Vos results are furthermore comparable to *De Vos et al.* [2019]. However, here QC De Vos filters less data and achieves the same metrics. This, since only the 20 nearest neighbours are taken into account which was done to limit computation time. It appears however that this also results in less SOs increasing the algorithm efficiency.

3.2 Case 1: KNMI AWS

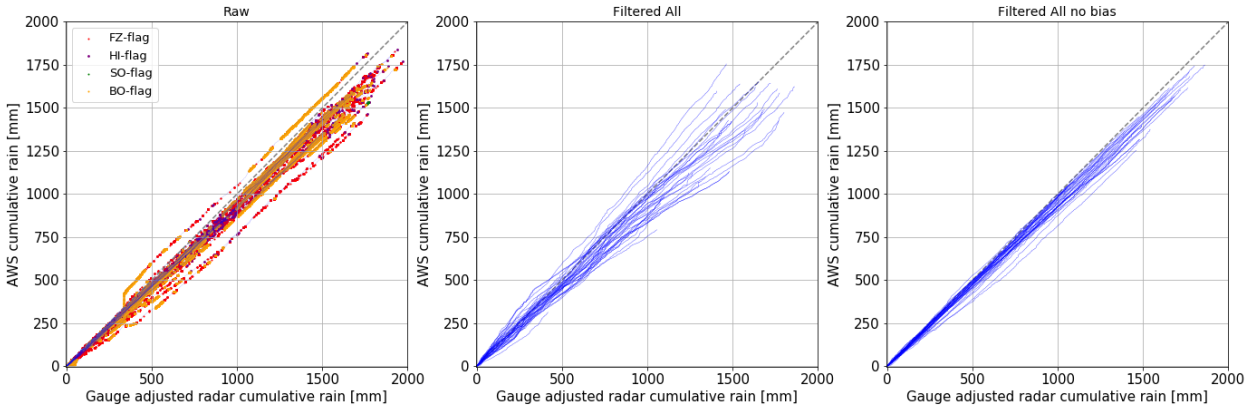
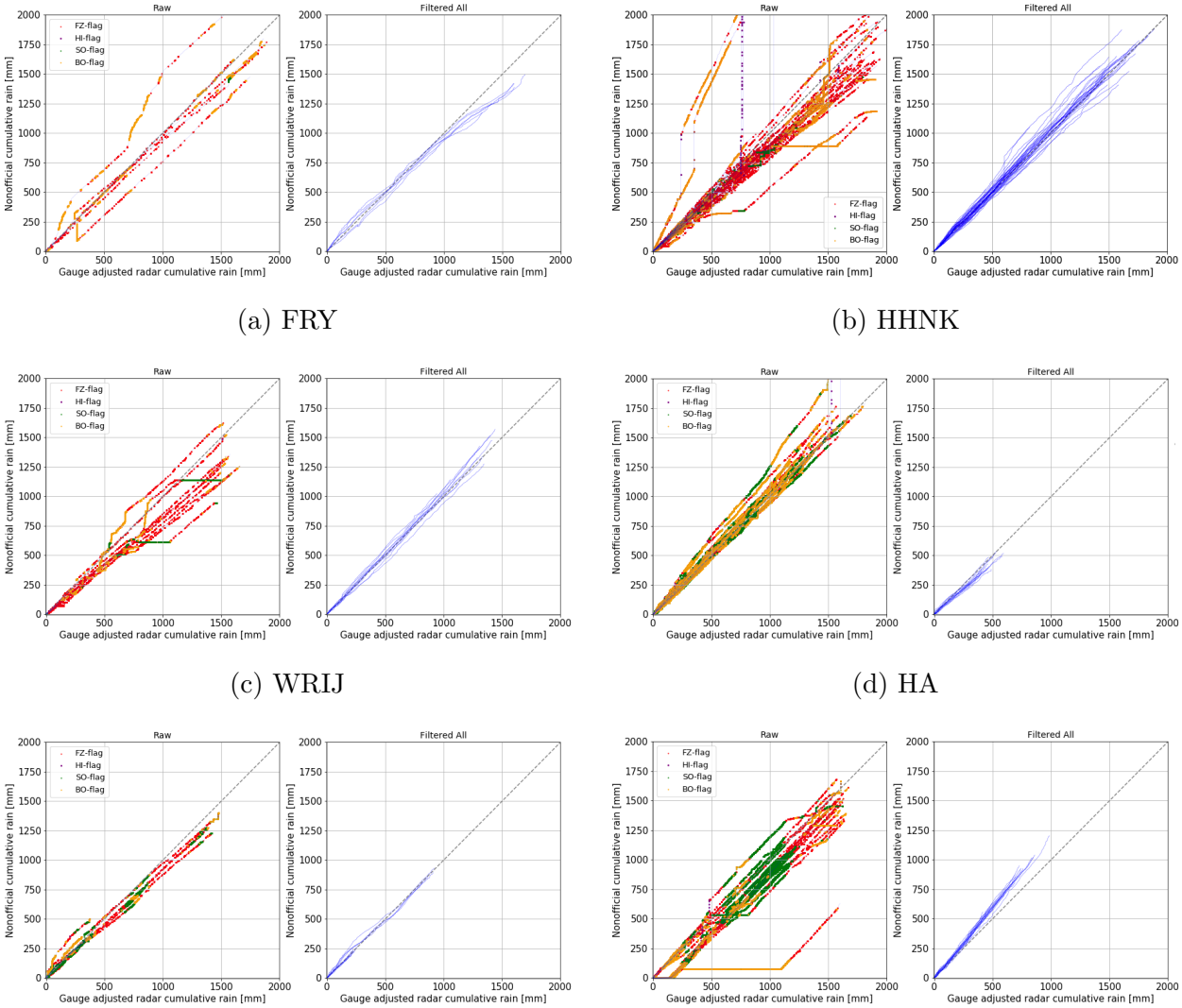


Figure 8: KNMI results before filtering (left), after All filtering (middle) and after applying All filters without applying the bias correction factor to the data for the period between 1 June 2016 and 1 June 2018. Without the bias correction factor applied, the quality of the KNMI is slightly enhanced.

Applying QC Radar to the KNMI data set, results in figure 8 and metrics that can be found in table 7 in Appendix A. The unvalidated data itself is of high quality. The networks underestimates precipitation by 7% (bias -0.07) compared to the reference radar set. This, since the radar set is adjusted not only by the present KNMI data set, but also by the network of manual rain gauges. Relative to these manual gauges, the official automatic rain gauges underestimate daily rainfall amounts [*Brandsma*, 2014]. Since the manual gauges are dominant in the adjustment procedure of the reference radar data set, the same underestimation of the official automatic rain gauges to the reference radar set is present. The CV of the raw data is 5.8 and with $r = 0.71$ it already approaches the optimal metric values. Based on figure 8, the data contains only one outlier overestimating precipitation and two rain gauges underestimating the data. The QC however also detects in almost 10% of the cases a FZ, due to the discrepancy in measurement resolution (radar also detects rainfall with an intensity less than 0.1 mm at 5-min intervals). The HI filter is not applied. The overestimating outlier contains several measurements of 3 mm to 5 mm, which is below the HI threshold of 10 mm. Since these measurements are only present at one day, the period is too short to be detected by the SO filter. Overall, the SO filter is applied in 0.5% of all cases, which is in line with what can be expected. The outlier is however filtered by the BO filter, together with 16.4% of the measurements. The vast majority is probably caused by the bias of the real-time available radar set rather than the bias of the rain gauges. After all, including the bias correction factor deteriorates the quality of the network as is shown in the middle panel of figure 8. Using QC Radar, the data overall can be improved to a bias of -0.06, CV of 5.6 and r of 0.72, for which as many as 24.6% of all data is excluded by FZ, HI, SO and BO filters, if the determined bias correction factor is not applied.



(a) FRY

(b) HHNK

(c) WRIJ

(d) HA

(e) SRIJN

(f) WDOD

Figure 9: Water board data filtering results using QC Radar. The results are highly variable depending on the properties of each individual data set. Results are shown for the period between 1 June 2016 and 1 June 2018, except for HHNK with a period between 1 February 2019 and 1 January 2021.

3.3 Case 2: Water board data

Results of the QC Radar applied to water board data are shown in figure 9. All calculations of the metrics can be found in table 7, results for the correlation r and the fraction of filtered measurements is shown in figure 10.

The FZ filter flags between 7% and 22% of the data sets. The least flags are found for data set FRY, which reports rainfall with a resolution of 0.01 mm. For WRIJ and WDOD the most data is removed by the FZ filter, improving the correlation and other metrics only slightly. For WDOD, a delay between measurement value and time stamp is suspected causing the majority of FZ flags. A deeper discussion can be found in section 4.2.

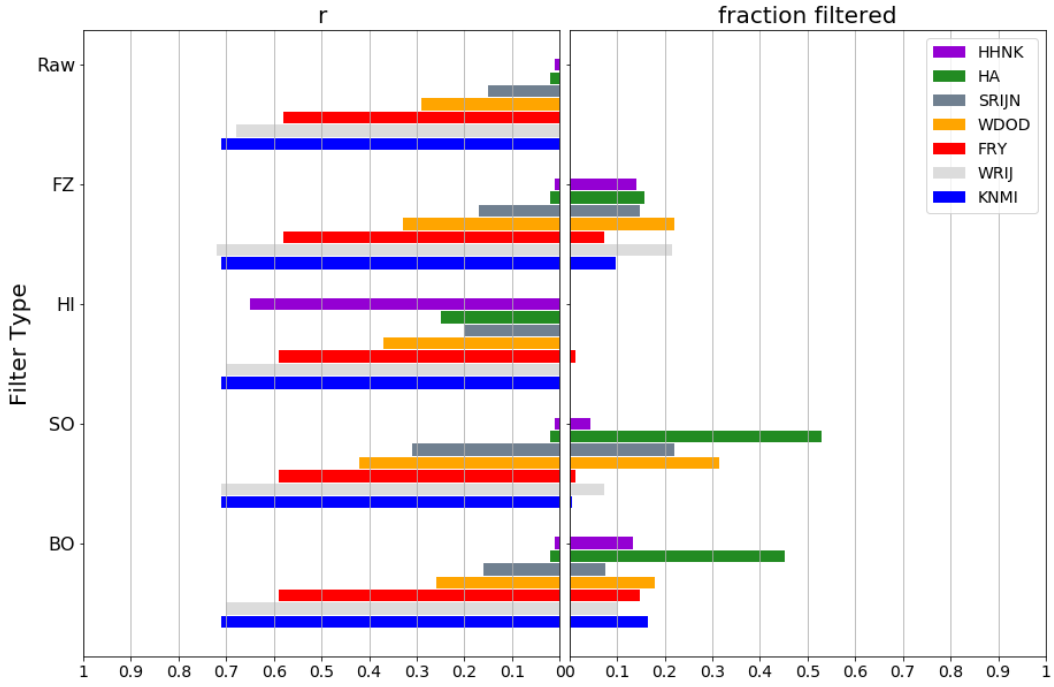


Figure 10: Correlation of the quality controlled data sets with the reference radar data set calculated at 5-min (10-min for WRIJ and KNMI) intervals (left) and fraction filtered (right) for filters FZ, HI, SO and BO for all water board data sets and the KNMI data set. Of all filters, the HI filter is applied the least, however causing the most improvement of all filters. Results are shown for the period between 1 June 2016 and 1 June 2018, except for HHNK with a period between 1 February 2019 and 1 January 2021. All metrics can be found in table 7 in Appendix A.

For all data sets, only several HI flags are present but cause a significant improvement of the metrics, the most for data set HHNK. In figure 9b several vertical lines before QC stand out. Filtering those, improves the bias from 12.97 to -0.05, CV from 1791.5 to 6.9 and r to 0.65

whereas the uncorrected data barely correlates with the reference radar data with $r = 0.01$.

The impact of the SO and BO filters are different depending on each water board. For HHNK, FRY and WRIJ, the percentage SO filters is relatively low. However, the metrics barely improve after applying the SO filter for these water boards: where r does not change for HHNK (0.01 to 0.01), it changes from 0.58 to 0.59 for FRY and 0.68 to 0.71 for WRIJ. Outliers, such as the upper line in figure 9a for FRY are detected by the BO filter rather than the SO filter. A calibration report employed by the Frysian water board revealed that this very rain gauge is the only rain gauge of the network of which the setup does not match the WMO-criteria.

For the data of SRIJN and WDOD, the SO filter is explicitly present in 22% and 31.4% of all measurements respectively. Moreover, for these data sets the metrics improve after applying the SO filter. For example WDOD reports a bias of 0.00 instead of -0.11, CV of 9.8 instead of 12.0 and r of 0.42 instead of 0.29. For these water boards, the BO filter is subsequently used in less situations. At data set HA however, both SO and BO filter are applied to 53% and 45.1% (partially overlapping) of the measurements respectively, both even without improving r. For all those three data sets, it is suspected that there is a discrepancy between the time stamp of the measurement and the actual time of the measurement. A deeper discussion can be found in section 4.2. Despite the fact that the data sets do not deviate significantly visually in figure 9 compared to other data sets, the QC method detects a discrepancy and applies the SO filter, as is shown in figure 10.

Results in correlation of applying all filters are shown in figure 11. The improvement found for HHNK is similar to earlier found results for the Netatmo Amsterdam data set. The unvalidated data with a bias, CV and r of 12.97, 1791.5 and 0.01, are improved to 0.01, 6.1 and 0.67 respectively by filtering 25.1% of the data. Metrics of the unvalidated data for FRY and most of all WRIJ were comparable to the metrics of KNMI unvalidated data on the other hand. By filtering, again the quality is further improved to, in case of WRIJ, a bias of 0.04, CV of 5.4 and r of 0.74 which practically equals the optimum $r = 0.75$ and CV= 5 by comparing rain gauge measurements to the reference radar data set at 5-min time intervals. For this improvement, 29.6% of the data is filtered. For WDOD, HA and SRIJN however, even after filtering, correlation values do not exceed $r = 0.5$. Besides remarkable is that only for the SRIJN data set, applying the bias correction factor further improves the quality. Especially for FRY, with bias correction, the data in combination with all filters is improved to a bias of -0.12, CV of 6.7 and r of 0.58. In contrast to without bias correction: values of bias, CV and r become -0.01, 6.4 and 0.60 respectively.

Overall, results variate strongly between the water board data sets. As the KNMI data set, applying a bias correction factor is not necessarily needed to improve the data. For data sets with comparable raw values of bias, CV and r as for the discussed Netatmo data sets, the same improvement can be reached using QC Radar (provided that the time stamps of the measurements match the timing of the measurements). For data sets of which the metrics match the quality of the KNMI data set, improvements can be made as well however again filtering more than 20% of the data using the current parameter settings.

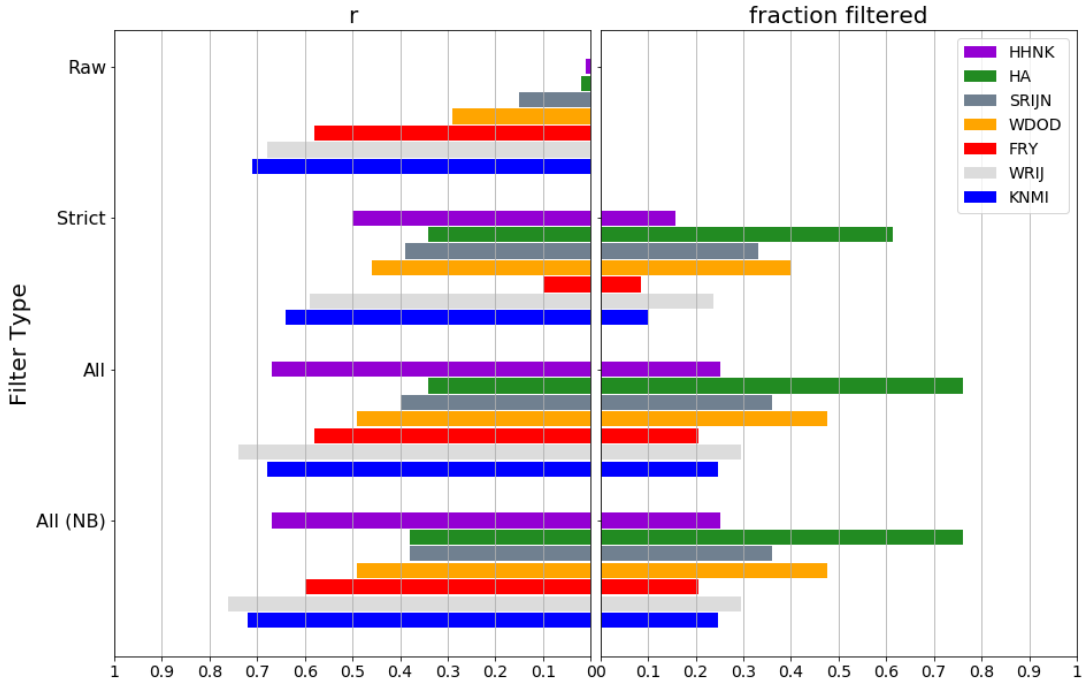


Figure 11: Correlation of the quality controlled data sets with the reference gauge adjusted radar data set calculated at 5-min (10-min for WRIJ and KNMI) intervals (left) and fraction filtered (right) applying the filters Strict, All and All (without applying bias correction factor). Overall, filtering All without applying the bias correction factor results in the highest correlation. The data quality for HHNK is improved the most. Results are shown for the period between 1 June 2016 and 1 June 2018, except for HHNK with a period between 1 February 2019 and 1 January 2021. All metrics can be found in table 7 in Appendix A.

3.4 Case 3: NetatmoNL data

Finally, the results of testing QC Radar using the Netatmo data set of 1 year starting in September 2019 for the entire Netherlands, are shown in figure 12. Metrics are shown in table 8 in Appendix A. The algorithm succeeds in improving the data quality drastically. The value for CV drops from 66.6 before filtering to 7.0 after filtering, with an increasing correlation from 0.07 to 0.58. For this, 31% of all measurements is filtered. Overall, the underestimation remains present with a bias of -0.04 before filtering and -0.08 after filtering. This, since more overestimations are filtered. Again, the HI filter is only applied in a few cases, however improving the data set to a CV of 9.4 and correlation of 0.47. The FZ filter flags 15.5% of all measurements, the SO filter 11.1% and the BO filter 16.2%. Hence, QC Radar performs equally if applied to national measurements (with varying spatial measurement density and varying quality of the real-time available radar data) as in the NetatmoAMS data set case. In figure 13, the summed filtered measurements for November 2019 are used to obtain a national overview of rainfall accumulations, by taking into account all rain gauges with more than 85% data availability, which is valid for 45% of all rain gauges of the raw

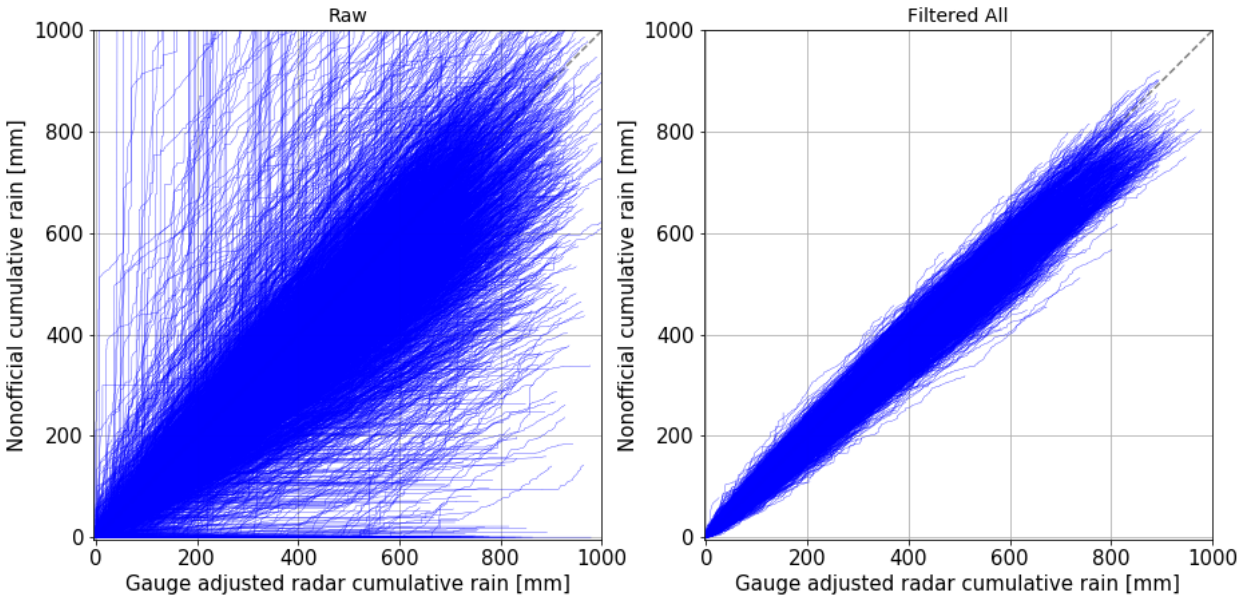
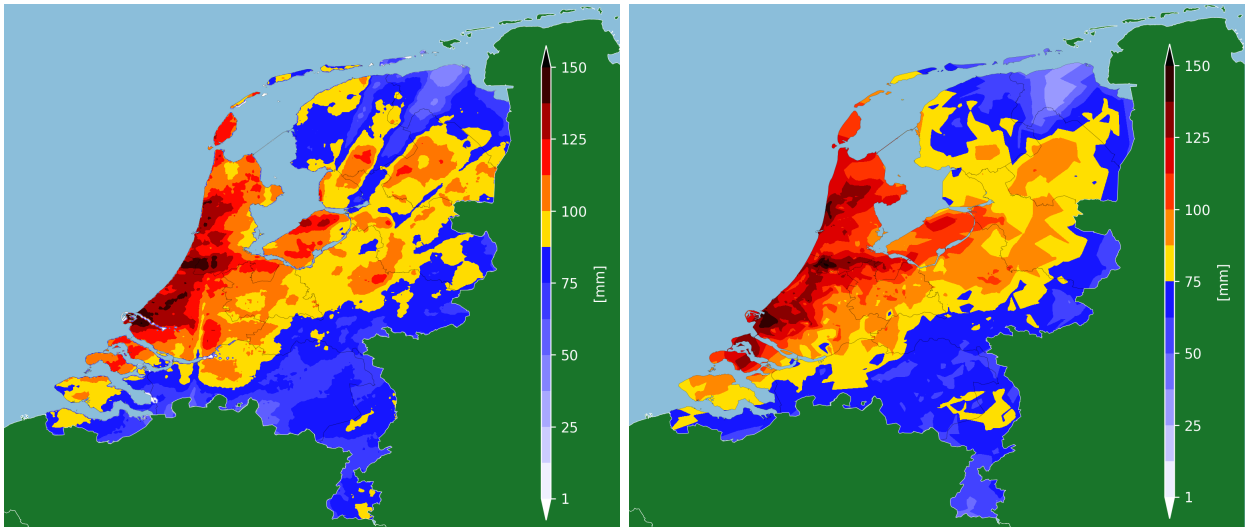


Figure 12: Results of QC Radar using Netatmo data. The QC method filters outliers, resulting in rain gauge measurements that are in line with the reference radar data set.

data set. Compared to the reference gauge adjusted radar image, both rainfall patterns as accumulations are produced by the filtered measurements.



(a) Reference radar accumulations (b) All-filtered rain gauge accumulations

Figure 13: Rainfall accumulations for November 2019 according to the reference radar product (a) and All filtered Netatmo rain gauges (b), linearly interpolated [*Scipy*, 2021]. Only rain gauges with at least 85% data availability after filtering are used. After filtering, still most rain gauges with enough measurements available are found in urban areas as is shown in figure 23 in Appendix B.

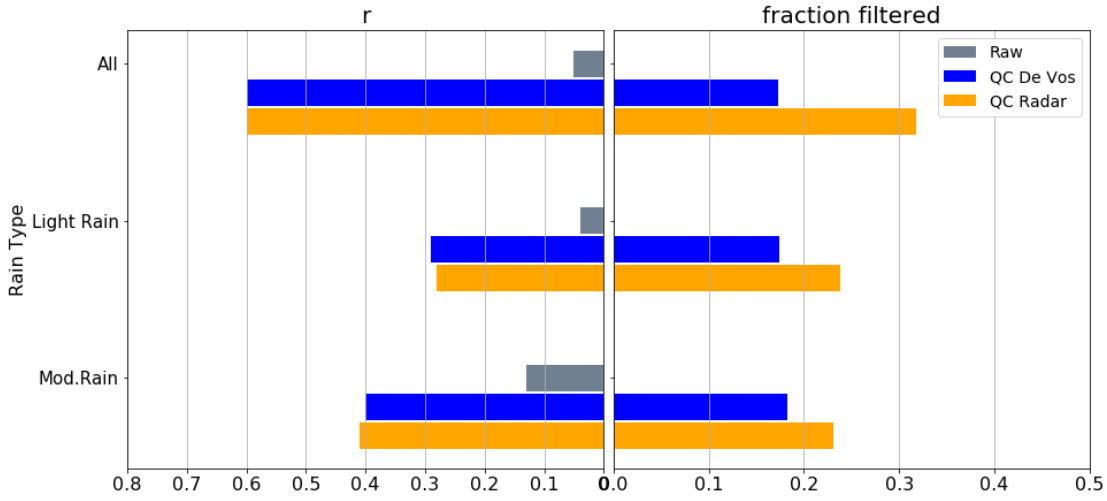
4 Discussion and outlook

4.1 Algorithm limitations

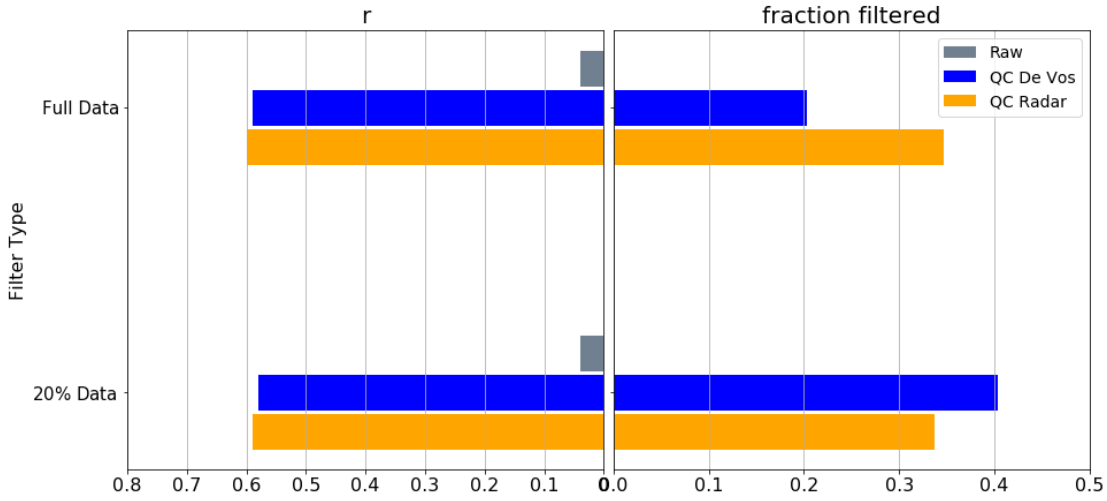
As described in section 2.1, QC De Vos inherently assumes that the median of the selected neighbouring rain gauges, represents the rainfall dynamics for the entire area within range d of a rain gauge. Here $d = 10,000$ m is chosen to be constant based on extrapolation of the results of *Van de Beek et al.* [2012]. However, in convective rainfall situations, this assumption can be violated since in these cases, rainfall deviates strongly locally. In contrast, in QC Radar a measurement is compared to a local radar pixel value, so that one can expect that QC De Vos is more sensitive to this criterion than QC Radar.

To demonstrate, results for Netatmo Amsterdam for 2017 and 2018 were filtered based on the rainfall dynamics, i.e. in case of moderate to heavy rain (here defined as time intervals in which the reference radar reports more than 0.5 mm per 5-min interval, 50,485 intervals in total summed for all rain gauges) and light rain events for which 0.1 mm $<$ measurement $<$ 0.5 mm is valid (438,109 intervals). Raw data shows that rain gauges increasingly underestimate precipitation at higher rainfall intensities as expected, as can bee seen in table 9 in Appendix A. Including all intervals, the median bias is positive for the years 2017 and 2018 combined and decreases to -0.26 and -0.40 for light and moderate to heavy rainfall respectively. Applying both QC algorithms to all data for 2017 and 2018 combined, it is again shown that the metrics provided by both QCs are in agreement, see figure 14a. However, QC Radar filters more data due to the more sensitive FZ filter. The same conclusion is valid for light rain, despite the fact that the difference in amount of filtering is reduced to a difference of 16.4%-point instead of almost 15%-point. For moderate to heavy rain it appears that QC Radar filters in less cases than light rain, whereas QC De Vos filters in more cases, due to an increase of SO flags. Hence, during moderate to heavy rain which is assumed to accommodate with more local rainfall dynamics, a real-time available radar pixel becomes relatively more representative of these rainfall dynamics than neighbouring rain gauges which can be on distance d of a rain gauge. In order to improve QC De Vos, one can incorporate a dynamic choice of d based on climatology in case the number of neighbouring gauges is sufficient.

Furthermore, if all neighbouring rain gauge measurements deviate from the actual rainfall dynamics (for instance if all neighbouring rain gauges are stuck or placed incorrectly), applying QC De Vos is less favoured over QC Radar as well. Already QC De Vos could not be applied to water board data since neighbouring stations were not present. In order to demonstrate a comparison with QC Radar, again a selection of the NetatmoAMS data set is made. Now, only data of 20% of all rain gauges (randomly selected) is taken into account. This resembles a rain gauge density of approximately 1/20 km^2 which is twice as dense as the densest rain gauge network of the water boards. Metrics after filtering for these selected stations are hardly affected at QC Radar by excluding the majority of the neighbouring gauges. In both cases, approximately 35% of the measurements of these rain gauges is filtered and a correlation of 0.59 instead of 0.6 is obtained as can be seen in figure 14b. For QC De Vos however, 40.4% of all data was filtered in the sparser data set compared to 20.3% for the selected gauges using all neighbours. Of the 40.4% filtered, 24.6%-point is explained by the



(a) Rainfall dynamics sensitivity



(b) Rain gauge density sensitivity

Figure 14: Correlation of the quality controlled data sets with the reference radar data set calculated at 5-min intervals (left) and fraction filtered (right) a) for several rainfall dynamics in terms of intensity. Light rain is defined for rainfall events where the reference radar set measures 0.1 mm up to and including 0.5 mm during a 5-min interval. More than 0.5 mm is regarded as moderate to heavy rainfall (Mod.). QC Radar filters less measurements for moderate to heavy rainfall than in case of light rainfall, or in case all measurement are taken into account. All metrics can be found in table 9 in Appendix A. b) For a reduced data set with less rain gauges, results shown for the rain gauges based on the computation with the full data set and the reduced data set where 20% of all rain gauges were selected before QC. All metrics can be found in table 10 in Appendix A. Data is shown for the period between 1 June 2016 and 1 June 2018.

absence of neighbouring measurements. Despite the fact that the metrics of the resulting data set are not affected, the result of filtering more measurements can conflict with possible applications, for example constructing rainfall maps which then become less detailed. Comparing the yielded metrics between QC Radar and QC De Vos for the sparse data set, QC Radar filters less measurements and metric results are equal or better.

At which exact rain gauge density QC Radar is preferred over QC De Vos depends on the precise selection of rain gauges. As shown, rainfall dynamics play a role as well since QC Radar becomes more efficient at higher rainfall intensities. This is explicitly shown by calculating the amount of filtered measurements for the various rainfall intensities if only 20% of all rain gauges is used. Results can be found in table 10 in Appendix A. Evaluations using several other rain gauge network density selections indicate that for those selections, the rain gauge density had to decrease by 50% to 80% in order for QC Radar to become more favourable than QC De Vos in terms of data removal and the same performance of the metrics. Extra comparison results could be obtained if QC De Vos is tested using the national Netatmo data set. Within this data set, (much) stronger variations in rainfall dynamics as in local rain gauge density are present. By comparing results of both methods, it could be analysed which method performs better as function of the local rain gauge network density, which might lead to combining both methods. Due to the relatively long computation time of QC De Vos for large data sets as NetatmoNL, this analysis is left out of the present study but is highly recommended for future research.

As a final remark, note that the above discussion is most relevant for data sets with short time intervals, such as data sets with a 5 to 10-min interval as used in the present study. For example in case of regional hydrological applications, a 1-hour interval would be sufficient. For climatological applications, daily accumulations would also be valuable. Using those longer time intervals, the choice of d is less strict reducing the sensitivity of QC De Vos to rainfall dynamics or a limited number of neighbouring rain gauges.

4.2 Water board rain gauge data issues

As mentioned in section 3, for the water board data sets SRIJN, HA and WDOD much more data is filtered compared to HHNK, FRY and WRIJ, due to the FZ and SO filter. Manual inspection of the measurement data sets of the water boards and the corresponding radar values of the reference radar data set, reveals a discrepancy between the timing of precipitation for SRIJN, HA and WDOD. In practice, the time stamp of a measured amount of precipitation is not congruent with the time stamp found in the reference radar data set. This is an example of a data processing error and can be due to an error of the gauge, the data transfer from water board to KNMI or due to the processing in order to use the data for quality control.

This last processing step is carried out as follows. Incoming measurements, accompanied with a certain time stamp, are saved or transformed into a regular UTC time series of 5 or 10 minutes. For example, data with a regular 1-min frequency is summed every 5 minutes over the past 5 measurements. If for a certain gauge in the past 5 minutes no measurements

are present, a NaN-value is filled in instead. A priori known error codes are transformed into NaN as well. However, if this information is not delivered, these error codes are included into the time series file, which presumably explains the negative values in the FRY time series.

However, for the SRIJN data set, the data is accompanied with UTC time stamps according to the metadata and thus the data is treated as such. In reality, this is only valid for measurements during the winter, since after switching to daylight saving time, precipitation has a 1-hour delay compared to the radar data set. Since this conclusion was only found after manual analysis and could not be made based on the a priori information, this error was left into the data set. For WDOD, the data is delivered with local time stamps, according to the metadata. However, during the analysis it appeared that time stamps are continuous during the transition to daylight saving time (instead of a 1-hour gap), leading again to a delay during the summer months. For both data sets, in figure 9 this daylight saving time delay in 2017 is clearly visible, since in this period of time the most SO flags are present.

For data set HA this is not the case. Again, the data is delivered with local time stamps according to the metadata. However, for data set HA, during the transition to daylight saving time most rain gauges pass two hours instead of one hour, leading to filtering of almost all data of these rain gauges.

No information is available whether these discrepancies are due to a sampling error of the rain gauge itself or due to the data processing and transfer carried out by the water boards. For WDOD, no information was present at the water boards, for HA and SRIJN no answers were retrieved. However, in either case, the manually found discrepancies are detected by the QC, most of all by the SO filter.

4.3 Parameter optimization

Algorithm parameters used in order to obtain all results, shown in table 3, were based on the parameters proposed by *De Vos et al.* [2019]. These parameters were found to achieve the best performance for data set 2017 of the Netatmo Amsterdam data set, alongside with large applicability and little data loss. Results can therefore be improved for each individual data set when parameters are optimized.

In general, parameters depend on local climatology and on rain gauge network characteristics, such as rain gauge type or on the purpose of the quality control (filtering all suspect measurements or only obvious errors). Especially n_{int}, m_{int}, m_{rain} and m_{match} can be set to higher values in order to retrieve more robust neighbour reference data in case of QC De Vos. If climatology describes local convective precipitation, d has to be chosen smaller in order to match the criterion that all rain gauge within range d measure similar rainfall dynamics. It is also possible to increase n_{stat} on the other hand in QC De Vos, so that the reference data is robust enough to perform the filters. This however comes with the cost of more -1 flags. Parameter changes are furthermore required for regions with elevation differences leading to large gradients in rainfall dynamics. Also if the rain gauge quality is known to be poor, a higher value of n_{stat} can be selected in order to obtain enough robust data. Regarding the HI

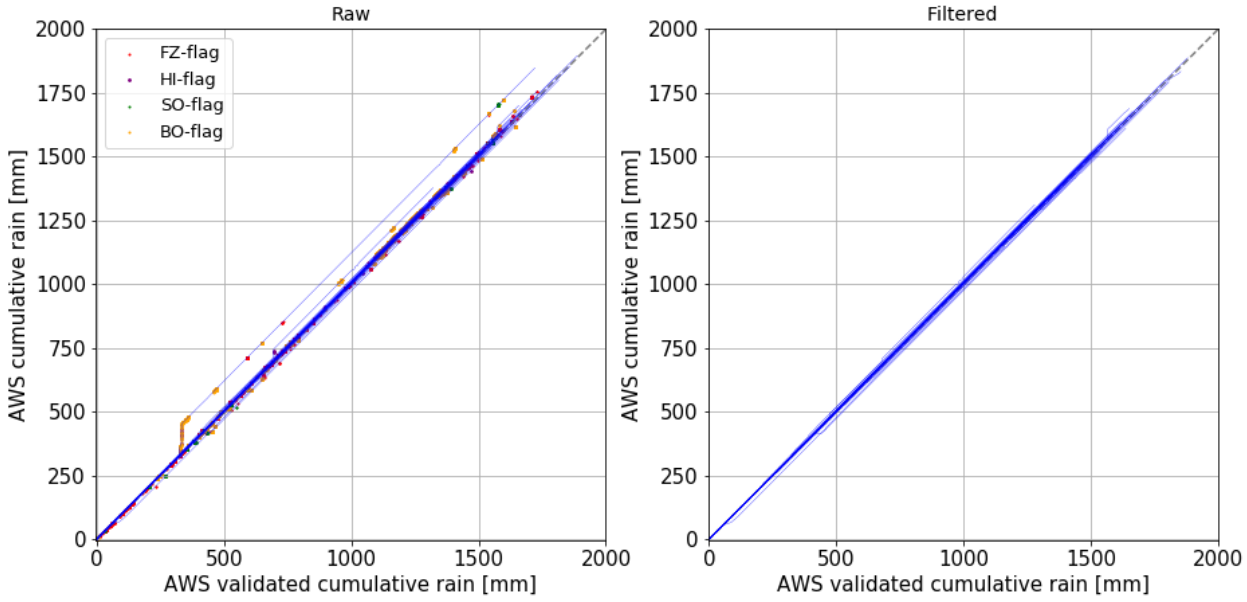
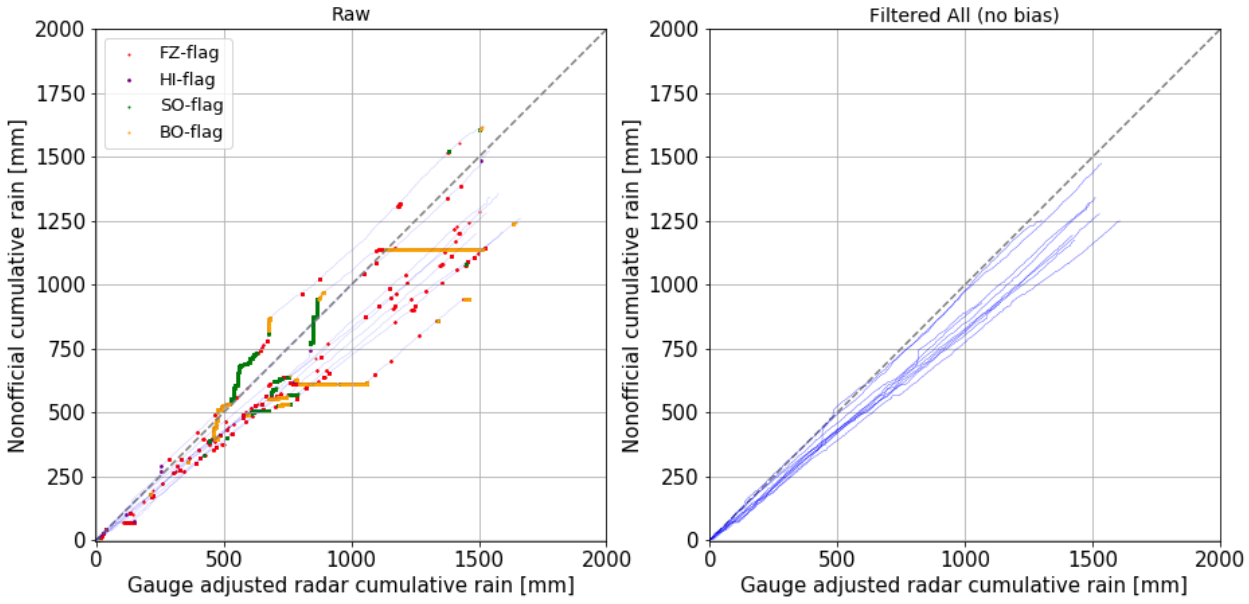


Figure 15: KNMI data set compared to the manually validated KNMI measurements as result of the calibration of parameters. After calibration, the result of the QC Radar is almost identical to the manually validated KNMI measurements in the shown period between 1 June 2016 and 1 June 2018.

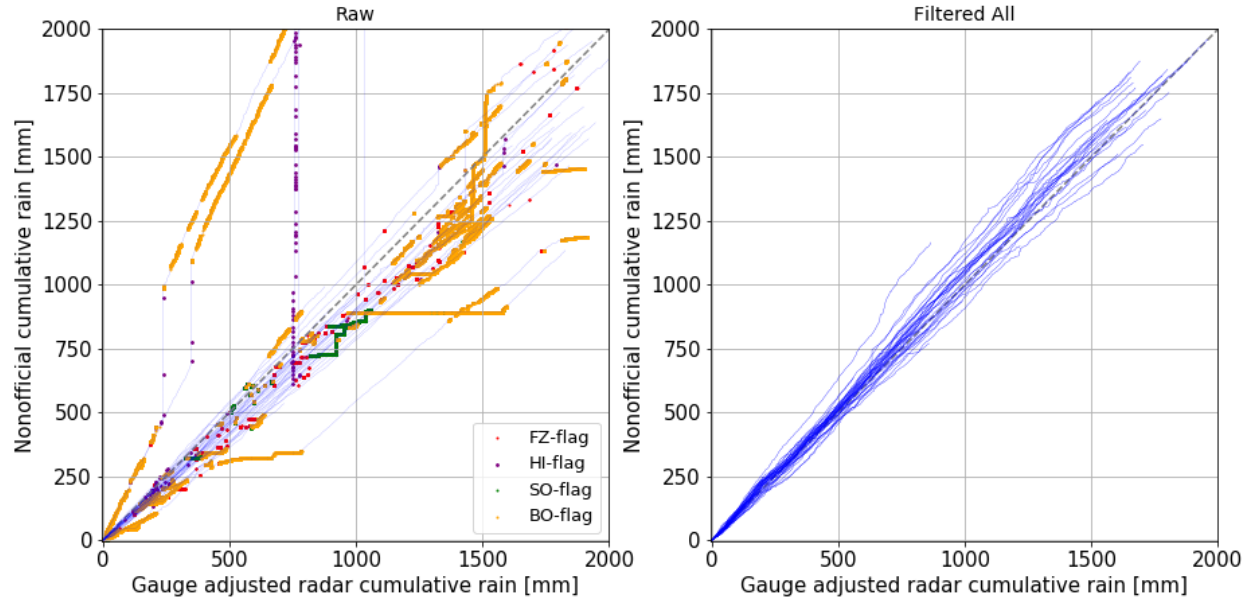
filter, the thresholds ϕ_A en ϕ_B are set relatively high based on the Netatmo data set, since HI outliers were found to be relatively high. For water board and KNMI data, some data sets contained measurements that can be regarded as HI, however were not selected by the QC Radar since the thresholds were set relatively high. It furthermore yields for QC Radar that n_{int} can be increased in order to be less sensitive for FZ filters in case of the Netherlands. For other less sensitive weather radars this might not be required. Weather radar characteristics are necessary to include by determining d_{radar}, which can be smaller if weather radar ranges are limited due to elevation or lower wave length of the radar signal.

As an example of parameter optimization, the parameters of QC Radar were optimized using the entire KNMI data set by comparing the output to the official manually validated hourly data set of KNMI for the same period. So, for the entire period between 1 June 2016 and 1 June 2018 with the period between 1 May 2016 and 1 June 2016 as warm-up period, results were compared to the validated hourly data in order to calibrate parameters. This leads to a decrease of ϕ_B to 3 mm, while preserving the fraction ϕ_B/ϕ_A by adjusting ϕ_A to 0.12 mm, for a more sensitive HI filter. FZ filter sensitivity is decreased by increasing n_{int} to 2 hours, that is to 12 intervals for a 10-min data set. Also, the BO filter is only applied to measurements with a bias correction factor less than 0.25 or more than 4. The bias correction factor itself, however, is not applied to the measurements.

The result of the calibration is shown in figure 15. Application of these parameters leads to a filtered data set almost identical to the manually validated hourly data set. Compared to the



(a) WRIJ



(b) HHNK

Figure 16: Results of quality control using optimized parameters of the KNMI data set for the WRIJ and HHNK data. By opimization, less measurements are filtered, whereas quality of the filtered data still improves. For WRIJ, the data is shown for the period between 1 June 2016 and 1 June 2018, for HHNK the period between 1 February 2019 and 1 January 2021.

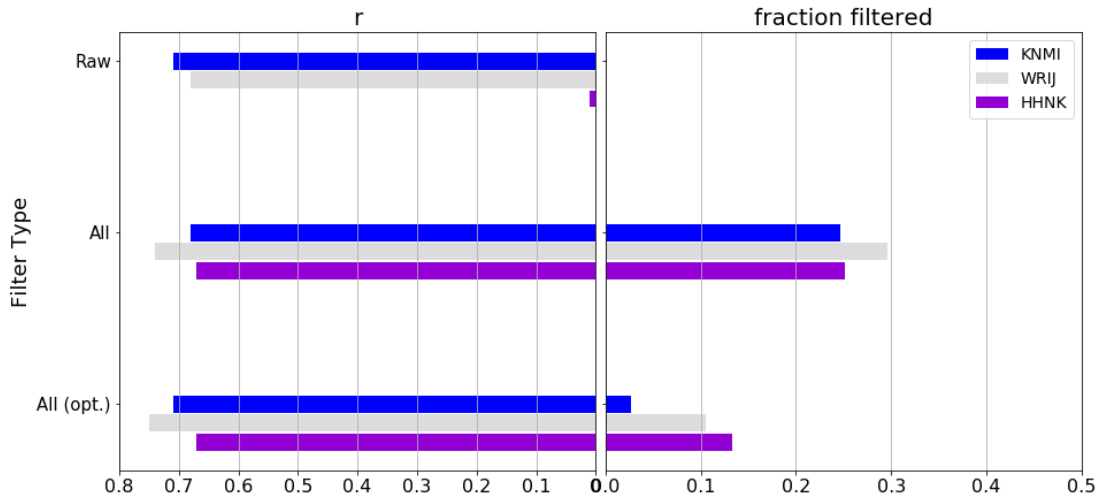


Figure 17: Correlation of the quality controlled data sets with the reference gauge adjusted radar data set calculated at 5-min intervals for HHNK and 10-min intervals for WRIJ and KNMI (left) and fraction filtered (right) applying All filters as in section 3 and in the optimized cases for KNMI, WRIJ and HHNK. Here, the KNMI data set can be regarded as calibration data set, since parameters are optimized using the KNMI data set with the official manually validated data set as reference. Results are shown for the period between 1 June 2016 and 1 June 2018, except for HHNK with a period between 1 February 2019 and 1 January 2021. All metrics can be found in table 11 in Appendix A.

reference radar data for 5-min time intervals, the parameter settings yield a bias of -0.07, a CV of 5.7 and r of 0.71. These results are highly comparable to the results described in section 3 (bias of -0.06, CV of 5.6, r of 0.72). However, with the optimized parameter settings only 2.7% of the data is filtered, whereas the first parameter settings filtered 24.6% of the data. In particular the FZ filter only filters 1% of the data compared to almost 10% earlier, and the BO filter 1.5% instead of 16.4%. As remark, note that these results are all obtained within the period that served as calibration period, further research has to test these parameters to other years in order to formally validate the parameters for the KNMI rain gauge network.

However, one can already assume that the found optimized parameters are applicable to other data sets. Since it was concluded that for water board data set WRIJ especially the raw data yields metrics comparable to the metrics of the raw KNMI data, the found KNMI parameters are consecutively applied to the WRIJ data set as validation of the parameters. Here all filters lead to a bias of -0.13, CV of 5.1 and r of 0.75 by filtering 10.5% of all data, instead of -0.09, 4.7 and 0.76 respectively by filtering 29.6%. The resulting data set is shown in figure 16a. It can be concluded that the KNMI parameter settings are able to produce comparable results with less abundant flagging also for this particular water board data set. For data set HHNK however, better results in terms of metrics were yielded by still applying the bias correction factor and applying the BO filter for measurements with a bias correction factor less than 0.5 or more than 2 as for the Netatmo data set. Then, a bias of -0.05, CV of 6.1, r of 0.68 by filtering 13.3% of the data instead of 25.1% (see figure 17) due to the less

strict FZ filter. The resulting data set is shown in figure 16b, an overview of all metrics is shown in table 11 in Appendix A. A detailed parameter optimization routine for every single data set can be the next step in improving the QC method. Optimized quality controlled measurements from the entire spectrum of third party rainfall data are namely qualified to be used in further research of improving real-time available radar products.

4.4 Validation remarks

As briefly discussed in section 2.8, the use of the gauge adjusted radar data set as reference data set is not optimal. This leads to the fact that the best metrics values that can be achieved are approximately 0.75 in case of r and a CV-value of approximately 5 according to *De Vos et al.* [2019]. On one side, the radar data set is considered as the most accurate rainfall reference, as the radar data is gauge adjusted using the 31 automatic KNMI rain gauges and most of all the 325 manual gauges. Hereby, this radar data set is particularly adjusted in order to match with daily rain accumulations rather than 5-min rain accumulations. For example, extreme 5-min rain accumulations could be underestimated or even missed by the radar if radar beams are blocked due to high precipitation intensities shielding off the radar location. Furthermore, within the presented validation method, the filtered rain gauge measurements are compared to radar pixels representing an area of approximately 1 km^2, whereas rain gauges provide point measurements leading to underestimation of extremes. One also has to consider that the precipitation reflectivity is measured above the ground leading to discrepancies as well. All these representativity differences are the largest for shorter time intervals. Therefore, the choice of validating results at this 5-min resolution pushes the validation to the limit. Evaluation of the metrics on a 1-hour time interval after quality control at 5-min interval results in metrics with a higher correlation and a lower value of CV as was shown by *De Vos et al.* [2019]. Further research is recommended to include calculations of the metrics at a 1-hour interval.

5 Conclusions

The main objective was to apply Quality Control to various sources of rainfall measurements with diverse characteristics, by taking national third party data of Netatmo PWSs and water boards and official KNMI measurements into account. For this, a new QC method based on QC De Vos which uses real-time weather radar data was proposed, as QC De Vos depends on nearby rain gauges which can be nonexistent for several rain gauge networks. Both methods are tested using data sets of at least one year.

It is shown that both QC methods are able to improve the Netatmo Amsterdam data set to almost identical bias, CV and r values for 5-min time intervals. QC Radar however filters more measurements so that QC De Vos is the preferred method in this case. QC Radar is relatively less sensitive to increasing rainfall intensities. Comparing the yielded metrics between QC Radar and QC De Vos for a sparse data set, QC Radar filters less measurements and metric results are equal or better. Further research is however recommended to include calculations of the metrics at a 1-hour interval reducing the effect of rainfall representativity discrepancies.

Secondly, results show that QC Radar is applicable to rain gauge data of water boards as well. For data sets with a high initial quality, the QC filters abundantly if parameter settings based on Netatmo data sets are used. Based on an official KNMI validation data set, it is shown that parameter settings can be chosen such that the loss of measurements is reduced without any costs in terms of the quality of the filtered data set. Further study for parameter optimization can utilize this potential improvement.

Finally, QC Radar is able to improve a Netatmo data set with varying rain gauge density on a national scale, hence showing that quality controlled third party data can be used as a viable source for rainfall monitoring. For further research, evaluation of the national Netatmo data set using the QC De Vos is highly recommended in order to correlate the performance of both QC methods with respect to local rain gauge network densities, as is exploring the potential of improving real-time radar products using quality controlled data sets.

References

Allamano, P., A. Croci, and F. Laio, Toward the camera rain gauge, *Water Resources Research*, *51*, 1744–1757, https://doi.org/10.1002/, 2015.

Bardossy, A., J. Seidel, and A. E. Hachem, The use of personal weather station observations to improve precipitation estimation and interpolation, *Hydrol. Earth Syst. Sci.*, *25*, 105–110, https://doi.org/10.5194/hess-25-583-2021, 2021.

Baserud, L., C. Lussana, T. Nipen, I. Seierstad, L. Oram, and T. Aspelien, TITAN automatic spatial quality control of meteorological in-situ observations, *Adv. Sci. Res.*, *17*, 153–163, https://doi.org/10.5194/asr-17-153-2020, 2020.

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

Berne, A., G. Delrieu, J.-D. Creutin, and C. Obled, Temporal and spatial resolution of rainfall measurements required for urban hydrology, *Journal of Hydrology*, *299*, 166–179, https://doi.org/10.1016/s0022-1694(04)00363-4, 2004.

Brandsma, T., Comparison of automatic and manual precipitation networks in the netherlands, *Tech. rep.*, KNMI, 2014.

Chen, A., M. Behl, and J. Goodall, Trust me, my neighbors say it's raining outside: Ensuring data trustworthiness for crowdsourced weather stations, *In Proceedings of the 5th Conference on Systems for Built Environments, ACM*, pp. 25–28, 2018.

De Vos, L., Rainfall observations datasets from personal weather stations, `https://data.4tu.nl/articles/dataset/Rainfall_observations_datasets_from_Personal_Weather_Stations/12703250`, doi:10.4121/uuid:6e6a9788-49fc-4635-a43d-a2fa164d37ec, [4TU.ResearchData. Dataset.], 2019.

De Vos, L., PWSQC Code, `https://github.com/LottedeVos/PWSQC`, 2021.

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

De Vos, L., H. Leijnse, A. Overeem, and R. Uijlenhoet, Quality control for crowdsourced personal weather stations to enable operational rainfall monitoring, *Geophysical Research Letters*, *46*, 8820–8829, https://doi.org/10.1029/2019GL083731, 2019.

Elmore, K., Z. Flamig, V. Lakshmanan, B. Kaney, V. Farmer, H. Reeves, and L. Rothfusz, MPING: Crowd-sourcing weather reports for research, *Bulletin of the American Meteorological Society*, *95*, 1335–1342, https://doi.org/10.1175/BAMS-D-13-00014.1, 2014.

Estevez, J., P.Gavilan, and J. Giraldez, Guidelines on validation procedures for meteorological data from automatic weather stations, *Journal of Hydrology*, *402*, 144–154, https://doi.org/10.1016/j.jhydrol.2011.02.031, 2011.

Guo, H., H. Huang, Y.-E. Sun, Y. Zhang, S. Chen, and L. Huang, Chaac: Real-time and fine-grained rain detection and measurement using smartphones, *IEEE Internet of Things Journal*, *6*, 997–1009, doi:10.1109/JIOT.2018.2866690, 2019.

Jenkins, G., A comparison between two types of widely used weather stations, *Weather*, *69*, 105–110, https://doi.org/10.1002/wea.2158, 2014.

Jiang, S., V. Babovic, Y. Zheng, and J. Xiong, Advancing opportunistic sensing in hydrology: A novel approach to measuring rainfall with ordinary surveillance cameras, *Water Resources Research*, *55*, 3004–3027, https://doi.org/10.1029/2018WR024480, 2019.

Kidd, C., A. Becker, G. Huffman, C. Muller, P. Joe, G. Skofronick-Jackson, and D. Kirschbauw, So, how much of the earths surface is covered by rain gauges?, *Bulletin of the American Meteorological Society*, *98*, 69–78, https://doi.org/10.1175/BAMS-D-14-00283.1, 2017.

KNMI, *Handbook for the Meteorological Observation*, 91-110 pp., De Bilt, 2000.

Napoly, A., T. Grassmann, F. Meier, and D. Fenner, Development and application of a statistically-based quality control for crowdsourced air temperature data, *Front. Earth Sci.*, *6*, 118, https://doi.org/10.3389/feart.2018.00118, 2018.

Netatmo, Netatmo weather, `https://netatmo.com/weather`, 2020a.

Netatmo, Smart rain gauge, `https://shop.netatmo.com/en-us/weather/accessories/rain-gauge`, 2020b.

Overeem, A., Precipitation - 5 minute precipitation accumulations from climatological gauge-adjusted radar dataset for the Netherlands (1 km) in NetCDF4 format, `https://dataplatform.knmi.nl/dataset/rad-nl25-rac-mfbs-5min-netcdf4-2-0`, 2020.

Overeem, A., and R. Imhoff, Archived 5-min rainfall accumulations from a radar dataset for the Netherlands., https://doi.org/10.4121/uuid:05a7abc4-8f74-43f4-b8b1-7ed7f5629a01 [4TU.ResearchData. Dataset.], 2020.

Overeem, A., I. Holleman, and A. Buishand, Derivation of a 10-year radar-based climatology of rainfall, *Journal of applied meteorology and climatology*, *48*, 1448–1463, doi:10.1002/wea.548, 2009.

Overeem, A., H.Leijnse, and R. Uijlenhoet, Measuring urban rainfall using microwave links from commercial cellular communication networks, *Water Resources Research*, *47*, https://doi.org/10.1029/2010WR010350, 2011.

Pollock, M. D., et al., Quantifying and mitigating wind-induced undercatch in rainfall measurements, *Water Resources Research*, *54*(6), 3863–3875, https://doi.org/10.1029/2017WR022421, 2018.

Rabiei, E., U. Haberlandt, M. Sester, and D.Fitzner, Rainfall estimation using moving cars as rain gauges - laboratory experiments, *Hydrology and Earth System Sciences*, *17*, 4701–4712, doi:10.5194/hess-17-4701-2013, 2013.

Scipy, Scipy interpolate interp2d, `https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.interp2d.html#scipy.interpolate.interp2d`, 2021.

Steiner, M., J. Smith, S. Burges, C. Alonso, and R. Darden, Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation, *Water Resour. Res.*, *35*, 2487–2503, doi:10.1029/1999WR900142, 1999.

Strangeways, I., A history of rain gauges, *Weather*, *65*, 133–138, doi:10.1002/wea.548, 2010.

Van Andel, J., QC radar, `https://github.com/NiekvanAndel/QC_radar`, 2021.

Van de Beek, C., H. Leijnse, P. Torfs, and R. Uijlenhoet, Seasonal semi-variance of Dutch rainfall at hourly to daily scales, *Advances in Water Resources*, *45*, 76–85, doi:https://doi.org/10.1016/j.advwatres.2012.03.023, 2012.

Zahumensky, I., Guidelines on quality control procedures for data from automatic weather station, WMO-No 955., 2004.

A Tables

Table 5: List of NetatmoAMS data set results for 2017 (calibration set). Metrics evaluated on a 5-min interval using the pixel value of the gauge adjusted radar data set as reference at each rain gauge location.

Data set	QC	Filter type	bias	CV	r	loss
NetatmoAMS 2017	De Vos	Raw	1.39	147.0	0.04	0.0%
		Bias	0.06	91.4	0.06	0.0%
		FZ (strict)	1.55	153.2	0.04	4.7%
		HI (strict)	-0.06	12.9	0.35	0.2%
		SO (strict)	0.01	17.9	0.27	9.4%
		BO (strict)	-0.09	11.7	0.40	6.7%
		Flex	0.05	9.0	0.58	9.2%
		Strict	0.05	8.9	0.59	11.2%
		All	0.04	8.4	0.61	17.0%
		All (no bias)	-0.13	7.5	0.61	17.0%
NetatmoAMS 2017	Radar	Raw	1.39	147.0	0.04	0.0%
		Bias	0.10	383.9	0.01	0.0%
		FZ (strict)	1.75	155.7	0.04	15.2%
		HI (strict)	-0.09	11.4	0.40	0.0%
		SO (strict)	0.73	57.1	0.09	10.6%
		BO (strict)	-0.16	8.4	0.53	20.7%
		Flex	0.09	9.5	0.56	19.8%
		Strict	0.09	9.5	0.56	20.4%
		All	0.01	7.9	0.61	34.8%
		All (no bias)	-0.07	7.3	0.62	34.8%

Table 6: List of NetatmoAMS data set results for 2018 (validation set). Metrics evaluated on a 5-min interval using the pixel value of the gauge adjusted radar data set as reference at each rain gauge location.

Data set	QC	Filter type	bias	CV	r	loss
NetatmoAMS 2018	De Vos	Raw	-0.11	53.2	0.07	0.0%
		Bias	0.07	21.4	0.22	0.0%
		FZ (strict)	-0.05	55.4	0.08	6.0%
		HI (strict)	-0.13	7.6	0.50	0.0%
		SO (strict)	-0.08	55.3	0.08	8.4%
		BO (strict)	-0.10	8.1	0.47	10.3%
		Flex	0.12	7.8	0.57	9.7%
		Strict	0.13	7.8	0.57	10.3%
		All	0.11	7.4	0.59	17.6%
		All (no bias)	-0.06	6.8	0.58	17.6%
NetatmoAMS 2018	Radar	Raw	-0.11	53.2	0.07	0.0%
		Bias	0.00	37.3	0.12	0.0%
		FZ (strict)	0.01	55.6	0.08	16.2%
		HI (strict)	-0.14	7.4	0.51	0.0%
		SO (strict)	-0.07	55.0	0.08	10.3%
		BO (strict)	-0.08	55.8	0.07	13.3%
		Flex	0.12	12.4	0.39	19.5%
		Strict	0.09	7.6	0.56	20.1%
		All	0.06	6.8	0.59	27.3%
		All (no bias)	-0.03	6.5	0.59	27.3%

Table 7: List of KNMI and water board data set results. Metrics evaluated on a 5-min interval (10-min interval for WRIJ and KNMI) using the pixel value of the gauge adjusted radar data set as reference at each rain gauge location. Data is shown for the period between 1 June 2016 and 1 June 2018, except for HHNK with a period between 1 February 2019 and 1 January 2021.

Experiment	Data set	Filter type	bias	CV	r	loss
Case 1	KNMI	Raw	-0.07	5.8	0.71	0.0%
		Bias	-0.14	7.2	0.64	0.0%
		FZ (strict)	-0.05	5.6	0.71	9.8%
		HI (strict)	-0.07	5.8	0.71	0.0%
		SO (strict)	-0.07	5.8	0.71	0.5%
		BO	-0.08	5.8	0.71	16.4%
		Strict	-0.12	6.9	0.64	10.0%
		All	-0.09	6.3	0.68	24.6%
		All (no bias)	-0.06	5.6	0.72	24.6%
Case 2	WDOD	Raw	-0.11	12.0	0.29	0.0%
		Bias	0.05	48.2	0.08	0.0%
		FZ (strict)	0.17	13.2	0.33	22.1%
		HI (strict)	-0.13	9.8	0.37	0.0%
		SO (strict)	0.00	10.9	0.42	31.4%
		BO	-0.06	11.6	0.26	17.9%
		Strict	0.17	10.9	0.46	40.0%
		All	0.21	9.7	0.49	47.6%
		All (no bias)	0.13	8.5	0.49	47.6%
	FRY	Raw	-0.01	6.8	0.58	0.0%
		Bias	-10.79	834.9	0.01	0.0%
		FZ (strict)	0.01	6.6	0.58	7.3%
		HI (strict)	-0.01	6.8	0.59	1.3%
		SO (strict)	0.00	6.7	0.59	1.3%
		BO	-0.02	6.6	0.59	14.7%
		Strict	-0.17	45.4	0.10	8.5%
		All	-0.12	6.7	0.58	20.6%
		All (no bias)	-0.01	6.4	0.60	20.6%
	HA	Raw	0.54	188.9	0.02	0.0%
		Bias	-0.11	126.7	0.02	0.0%
		FZ (strict)	1.05	231.1	0.02	15.8%
		HI (strict)	0.06	12.9	0.25	0.0%
		SO (strict)	0.32	154.3	0.02	53.0%
		BO	0.24	136.4	0.02	45.1%
		Strict	-0.18	10.0	0.34	61.3%
		All	-0.09	9.5	0.34	76.1%
		All (no bias)	0.20	10.2	0.38	76.1%

(Table 7 continues at the following page)

Experiment	Data set	Filter type	bias	CV	r	loss
	SRIJN	Raw	-0.05	13.1	0.15	0.0%
		Bias	-0.15	10.8	0.19	0.0%
		FZ (strict)	0.14	13.7	0.17	14.7%
		HI (strict)	-0.07	10.7	0.20	0.0%
		SO (strict)	-0.08	8.7	0.31	22.0%
		BO	-0.07	12.6	0.16	7.5%
		Strict	0.01	7.7	0.39	33.1%
		All	0.00	7.6	0.40	36.1%
		All (no bias)	0.04	7.9	0.38	36.1%
	WRIJ	Raw	-0.17	5.8	0.68	0.0%
		Bias	-0.04	9.7	0.55	0.0%
		FZ (strict)	-0.04	5.4	0.72	21.5%
		HI (strict)	-0.17	5.5	0.70	0.0%
		SO (strict)	-0.14	5.5	0.71	7.3%
		BO	-0.19	5.4	0.70	10.0%
		Strict	0.09	9.3	0.59	23.7%
		All	0.04	5.4	0.74	29.6%
		All (no bias)	-0.09	4.7	0.76	29.6%
	HHNK	Raw	12.97	1791.5	0.01	0.0%
		Bias	12.21	1671.5	0.01	0.0%
		FZ (strict)	13.92	1773.0	0.01	14.1%
		HI (strict)	-0.05	6.9	0.65	0.0%
		SO (strict)	2.77	895.9	0.01	4.5%
		BO	14.88	1914.8	0.01	13.4%
		Strict	0.13	13.3	0.50	15.7%
		All	0.01	6.1	0.67	25.1%
		All (no bias)	-0.01	6.1	0.67	25.1%

Table 8: List of NetatmoNL data set results. Metrics evaluated on a 5-min interval using the pixel value of the gauge adjusted radar data set as reference at each rain gauge location for the period between 1 October 2019 and 1 September 2020.

Data set	QC	Filter type	bias	CV	r	loss
NetatmoNL	Radar	Raw	-0.04	66.5	0.07	0.0%
		FZ (strict)	0.07	67.9	0.07	15.5%
		HI (strict)	-0.07	9.4	0.47	0.0%
		SO (strict)	-0.05	66.1	0.07	11.1%
		BO (strict)	-0.10	51.8	0.09	16.2%
		All	-0.08	7.0	0.58	31.0%

Table 9: List of results for several rainfall dynamics. Metrics evaluated on a 5-min interval using the pixel value of the gauge adjusted radar data set as reference at each rain gauge location. Light rain is defined for rainfall events where the reference radar set measures more than 0.1 mm and less than or equal to 0.5 mm during a 5-min interval. More than 0.5 mm is regarded as moderate to heavy rainfall. Results are shown for the period between 1 June 2016 and 1 June 2018.

Data set	QC	Filter type	bias	CV	r	loss
NetatmoAMS All data		Raw	0.48	94.9	0.05	
NetatmoAMS light rain		Raw	-0.26	6.92	0.04	
NetatmoAMS moderate to heavy rain		Raw	-0.40	2.3	0.13	
NetatmoAMS All data	De Vos	All	0.09	7.8	0.60	17.3%
NetatmoAMS light rain	De Vos	All	-0.14	1.2	0.29	17.4%
NetatmoAMS moderate to heavy rain	De Vos	All	-0.24	1.0	0.40	18.3%
NetatmoAMS All data	Radar	All	0.04	7.3	0.60	31.8%
NetatmoAMS light rain	Radar	All	-0.18	1.1	0.28	23.8%
NetatmoAMS moderate to heavy rain	Radar	All	-0.29	0.9	0.41	23.1%

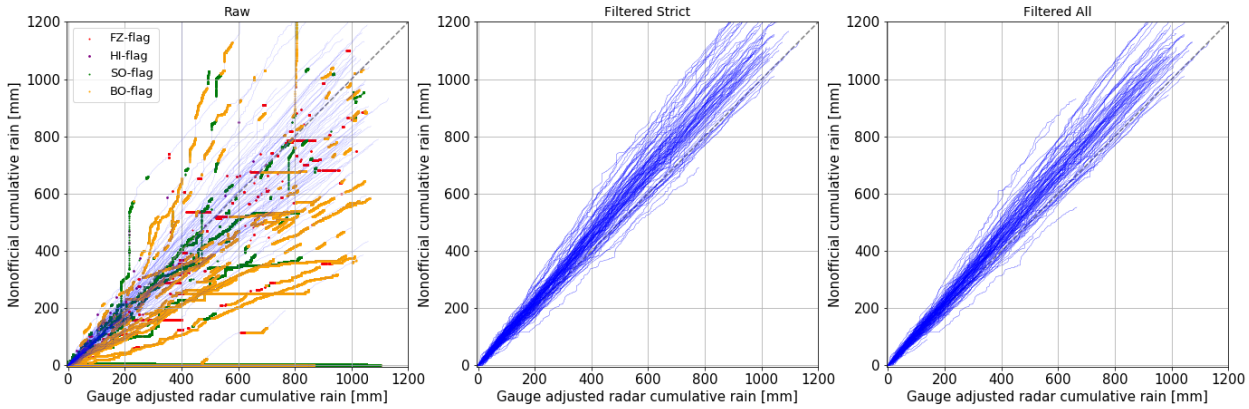
Table 10: List of results for several rainfall dynamics for a selection of rain gauges. Metrics evaluated on a 5-min interval using the pixel value of the gauge adjusted radar data set as reference at each rain gauge location. During calculation of all metrics, only a selection of 20% of the original NetatmoAMS data set is used, hence 'NetatmoAMS 20% as description of the data sets. In case of 'QC Radar 20%' and 'QC De Vos 20%', the filters were furthermore calculated based on only 20% of the NetatmoAMS data set. This in contrast to 'QC Radar' and 'QC De Vos', where filter calculation was performed including the entire NetatmoAMS data set. Results are also filtered for several types of rainfall dynamics. Light rain is defined for rainfall events where the reference radar set measures more than 0.1 mm and less than or equal to 0.5 mm during a 5-min interval. More than 0.5 mm is regarded as moderate to heavy rainfall. Results are furthermore shown for the period between 1 June 2016 and 1 June 2018.

Data set	QC	Filter	bias	CV	r	loss
NetatmoAMS 20% All data		Raw	-0.11	102.9	0.04	
NetatmoAMS 20% light rain		Raw	-0.36	1.15	0.23	
NetatmoAMS 20% mod. to heavy rain		Raw	-0.45	0.9	0.35	
NetatmoAMS 20% All data	De Vos	All	0.08	8.0	0.59	20.3%
NetatmoAMS 20% light rain	De Vos	All	-0.14	1.2	0.28	20.6%
NetatmoAMS 20% mod. to heavy rain	De Vos	All	-0.25	1.0	0.39	20.4%
NetatmoAMS 20% All data	De Vos 20%	All	0.08	8.0	0.58	40.4%
NetatmoAMS 20% light rain	De Vos 20%	All	-0.15	1.2	0.28	41.4%
NetatmoAMS 20% mod. to heavy rain	De Vos 20%	All	-0.27	1.0	0.38	41.3%
NetatmoAMS 20% All data	Radar	All	0.05	7.4	0.60	34.7%
NetatmoAMS 20% light rain	Radar	All	-0.17	1.2	0.28	27.7%
NetatmoAMS 20% mod. to heavy rain	Radar	All	-0.27	1.0	0.41	25.8%
NetatmoAMS 20% All data	Radar 20%	All	0.05	7.5	0.59	33.7%
NetatmoAMS 20% light rain	Radar 20%	All	-0.17	1.2	0.27	27.5%
NetatmoAMS 20% mod. to heavy rain	Radar 20%	All	-0.28	1.0	0.42	25.0%

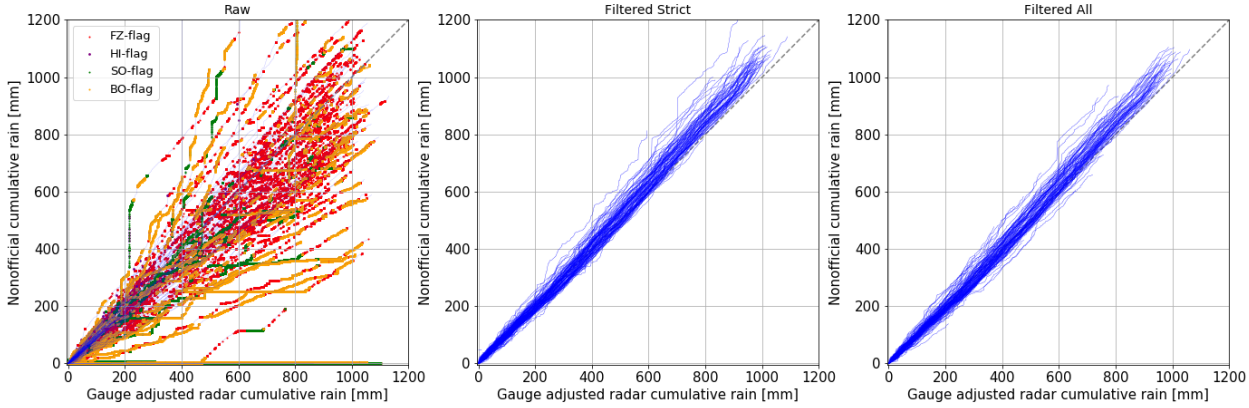
Table 11: List of data set results using optimized parameters. Here, the KNMI data set can be regarded as calibration data set, since parameters are optimized using the KNMI data set with the official manually validated data set as reference. Metrics were evaluated on a 5-min interval for HHNK and on a 10-interval for WRIJ and KNMI using the pixel value of the gauge adjusted radar data set as reference at each rain gauge location. Results are shown for the period between 1 June 2016 and 1 June 2018, except for HHNK with a period between 1 February 2019 and 1 January 2021.

Data set	QC	Filter type	bias	CV	r	loss
KNMI	Radar	Raw	-0.07	5.8	0.71	0.0%
		FZ (strict)	-0.07	5.8	0.71	1.0%
		HI (strict)	-0.08	5.8	0.71	0.0%
		SO (strict)	0.07	5.8	0.71	0.4%
		BO (strict)	-0.08	5.8	0.71	1.5%
		All (no bias)	-0.07	5.7	0.71	2.7%
WRIJ	Radar	Raw	-0.17	5.8	0.68	0.0%
		FZ (strict)	-0.11	5.7	0.71	5.9%
		HI (strict)	-0.17	5.5	0.71	0.0%
		SO (strict)	-0.15	5.5	0.72	7.7%
		BO (strict)	-0.13	5.6	0.71	6.2%
		All (no bias)	-0.13	5.1	0.75	10.5%
HHNK	Radar	Raw	12.97	1791.5	0.01	0.0%
		FZ (strict)	13.29	1811.6	0.01	2.3%
		HI (strict)	-0.06	6.6	0.67	0.0%
		SO (strict)	10.13	1661.9	0.01	2.7%
		BO (strict)	11.22	1753.5	0.01	12.1%
		All	0.02	6.5	0.67	13.3%

B Figures



(a) QC De Vos



(b) QC Radar

Figure 18: Results of both QC methods a) QC De Vos and b) QC Radar for the Netatmo Amsterdam data set 2018. The x-axis shows the cumulative precipitation of the reference weather radar data set, the y-axis the cumulative rain gauge measurements. Every blue line on the left panel represents the cumulative raw data. The middle panel shows the rain gauge data after applying the Strict filtering method (FZ, HI, SO and bias correction), the right panel shows the data after All filters, which are both much more in line with the reference radar data.

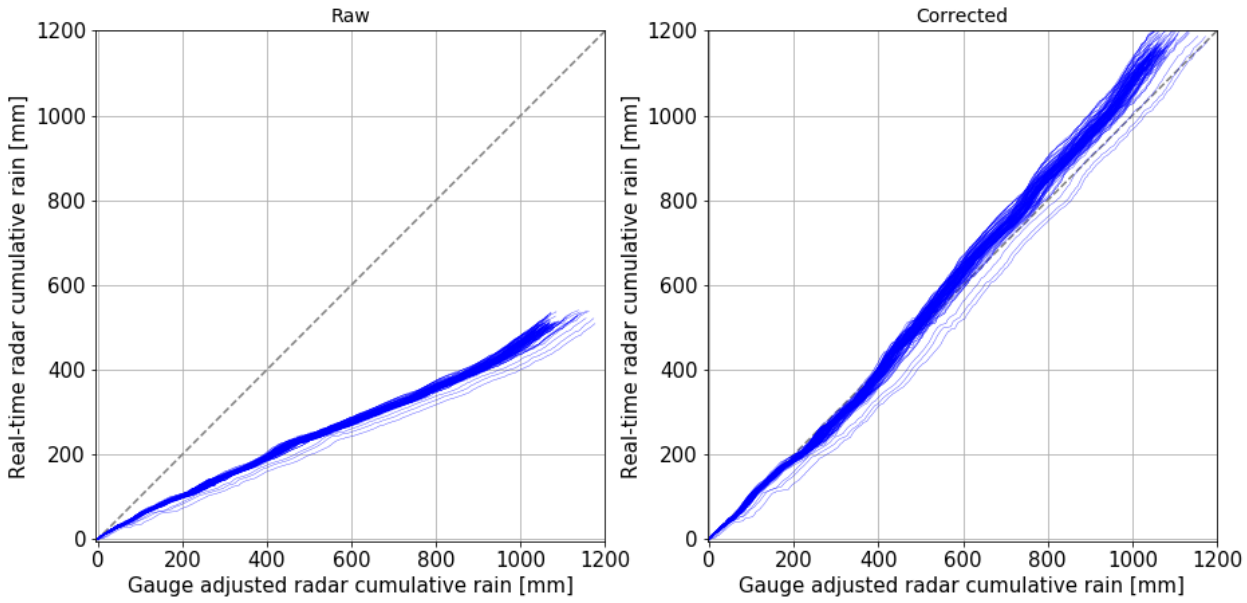


Figure 19: Raw radar adjustment by QC Radar, on the left, the raw cumulative real-time available radar data for all rain gauge locations in Amsterdam for data set 2018. At the right hand side, the adjusted cumulative real-time radar data is shown.

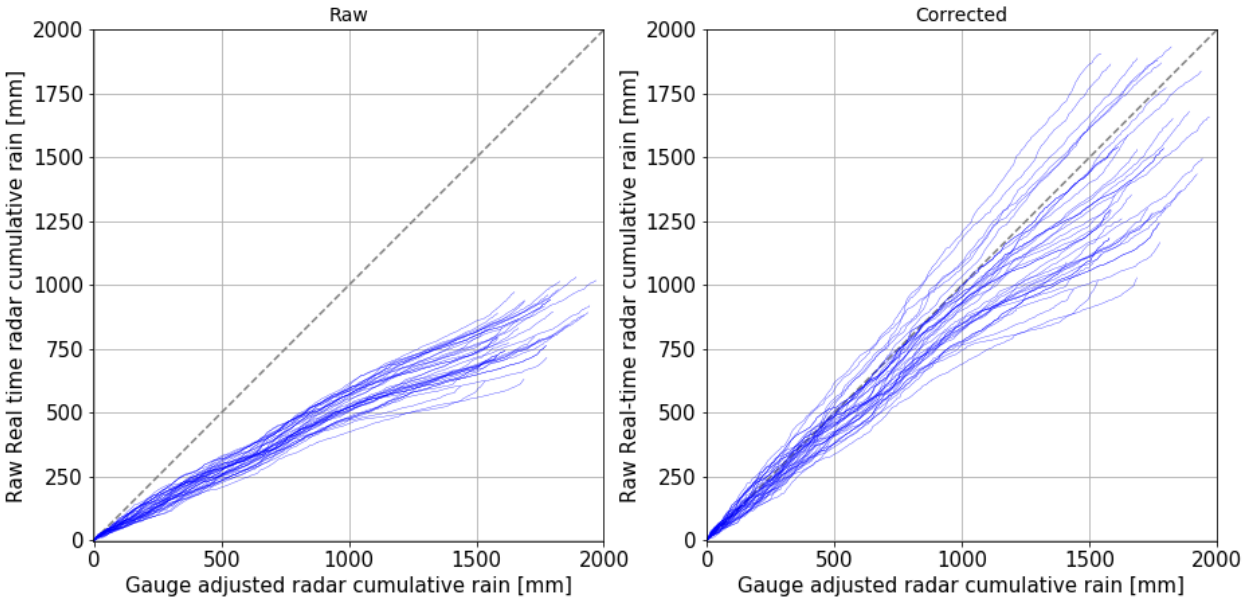
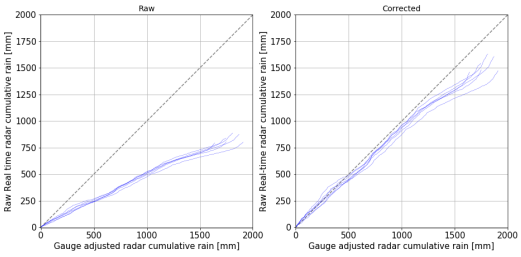
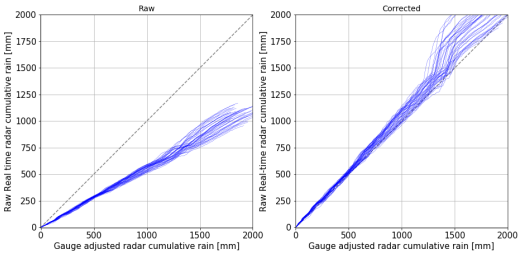


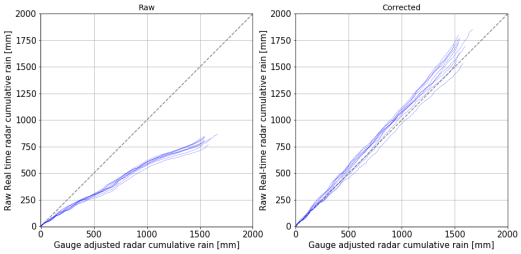
Figure 20: Raw radar adjustment by QC Radar, on the left, the raw cumulative real-time available radar data for all rain gauge locations of the KNMI data set between 1 June 2016 and 1 June 2018. At the right hand side, the adjusted cumulative real-time radar data is shown.



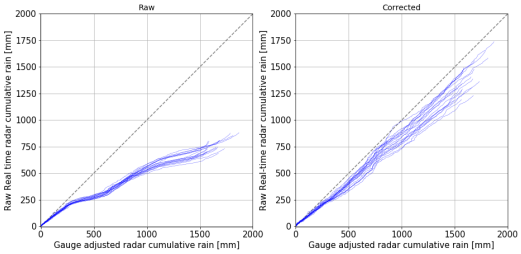
(a) FRY



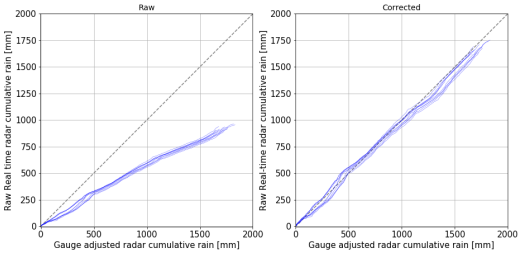
(b) HHNK



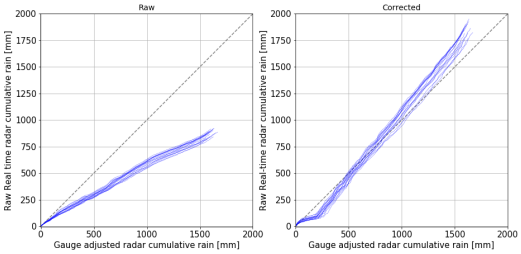
(c) WRIJ



(d) HA



(e) SRIJN



(f) WDOD

Figure 21: Raw radar adjustment by QC Radar, on the left, the raw cumulative real-time available radar data for all rain gauge locations for 6 water boards. At the right hand side of each panel, the adjusted cumulative real-time radar data is shown. Results are shown for the period between 1 June 2016 and 1 June 2018, except for HHNK with a period between 1 February 2019 and 1 January 2021.

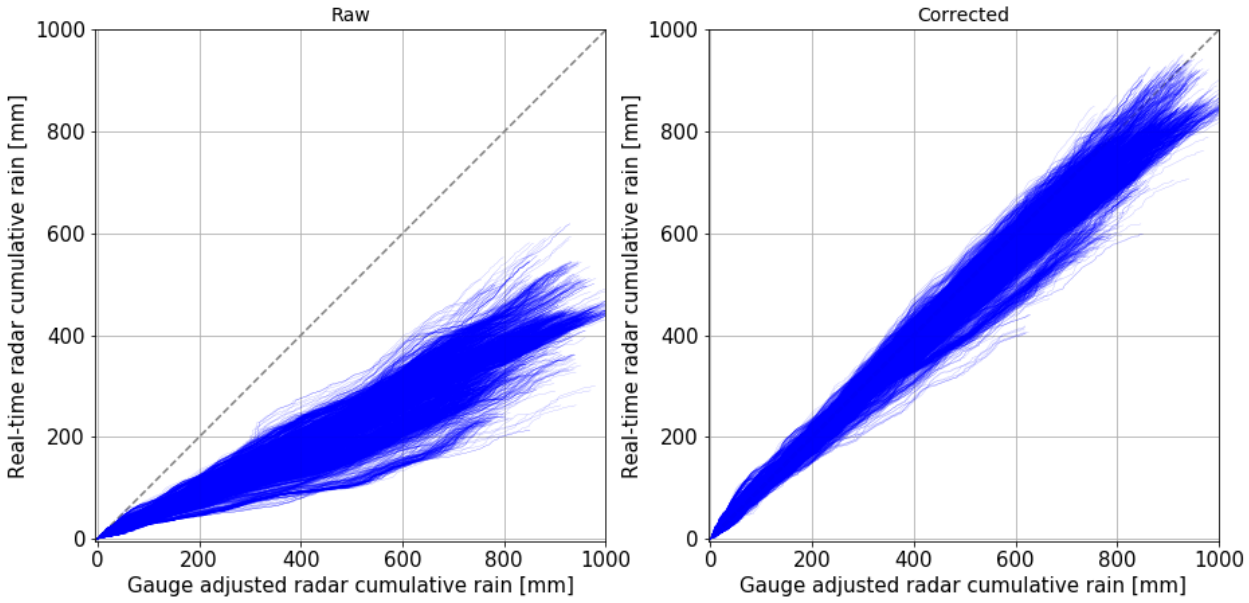
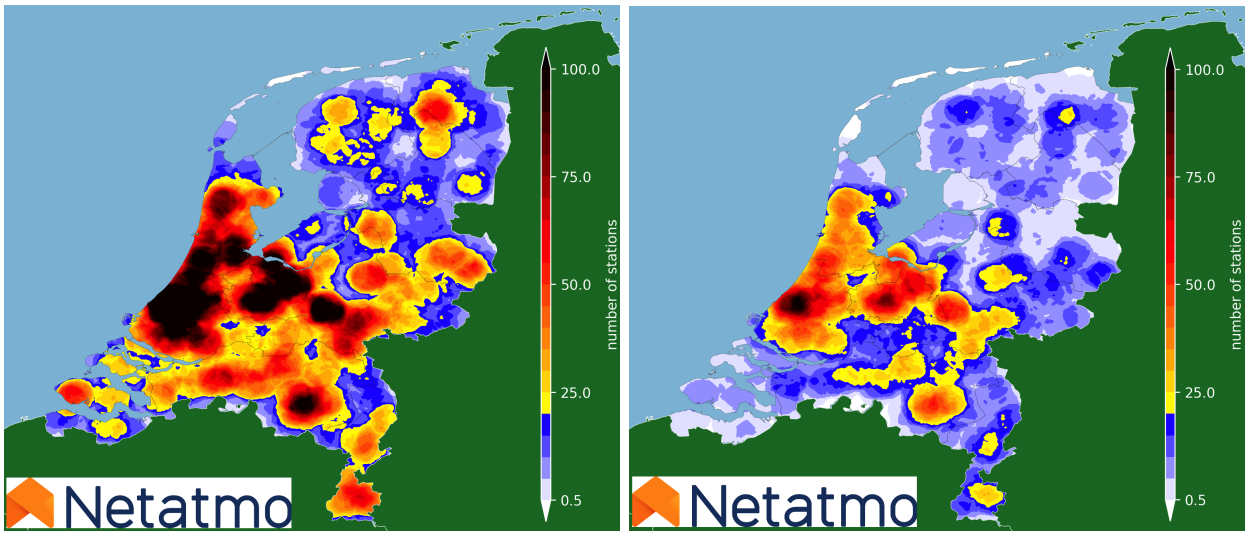


Figure 22: Raw radar adjustment by QC Radar, on the left, the raw cumulative real-time available radar data for all rain gauge locations of the NetatmoNL data set. In contrast to regional rain gauge networks, the bias of the real-time available radar accumulation is not constant over the entire Netherlands. At the right hand side, the adjusted cumulative real-time radar data is shown for the period between 1 October 2019 and 1 September 2020.



(a) Rain gauge density (b) Rain gauge density used for map

Figure 23: Rain gauge density before filtering (a) and after filtering (b) in November 2019 using rain gauges for which in 85% of the time, measurements are available [*Netatmo*, 2020a].