The potential of crowdsourced personal weather stations for hydrological forecasting in a Dutch lowland catchment

Romy Lammerts

July, 2021

MSc thesis

Supervisors: Claudia Brauer & Lotte de Vos
Hydrology and Quantitative Water Management Group
Wageningen University

Cover picture: Rain module of the automatic weather station by brand Netatmo. Source: http://www.atsmartliving.com/netatmo-rain-gauge-clouds-pass-but-rain-remains/. (Published: May 21, 2019).

Abstract

Making accurate predictions of rainfall extremes is challenging because rainfall fields are highly variable in space and time, which limits the quality of discharge simulations. Crowdsourced personal weather stations have a high spatio-temporal resolution but are prone to errors. The study aims to assess the accuracy of quality-controlled (QC) personal weather stations (PWS) in observing rainfall and predicting discharge in a Dutch lowland catchment, the Oude IJssel. The accuracy of rainfall observations was tested by (1) quantifying the available data before and after quality control, (2) validating individual personal weather stations, (3) validating catchment-averaged time series including a comparison with the operational weather radar (unadjusted and real-time) w.r.t. a reference radar product (gauge-adjusted and offline) and (4) investigating the effect of PWS network density. The four quantitative precipitation estimates were used as input for rainfall-runoff model WALRUS where four corresponding discharge simulations were (5) validated w.r.t. the reference input for the study period of 11 months and two precipitation events in winter and summer and (6) investigated how rainfall measurement errors propagated in the predicted discharge. The unadjusted radar systematically underestimated the reference 5 min averaged rainfall depths with a bias of -0.164 mm, while catchment-averaged rainfall depths measured by personal weather stations slightly overestimated the reference with a bias of only 0.025mm. No less bias was registered after quality control of PWS, however time series varied less and correlated better relative to the reference. Validation individual stations on the other hand yielded 10.6% bias reduction in absolute terms after quality control while 85.1% of the data remained. Discharge simulations were the best for the quality-controlled personal weather stations (NSE = 0.98, averaged over the catchments during study period; NSE = 0.91, averaged over the catchments and events), followed by the input of personal weather stations before quality control (NSE = 0.95; NSE = 0.82) and lastly the operational weather radar during both the study period and the two selected events (NSE = 0.70; NSE = 0.78) where more accurate rainfall observations resulted in more accurate discharge predictions. To conclude, quality-controlled personal weather stations better observe rainfall and predict discharge on a catchment-scale compared to the operational weather radar and thus enlarge the potential for operational hydrological applications in the Netherlands. Assessment over a full year or longer is recommended considering multiple catchments where their respective network layout is taken into account. Lastly, it is recommended to investigate the potential of quality-controlled PWSs as correction method for real-time weather radar where it could potentially observe rainfall extremes and predict local floods more accurately in the future.

Contents

1 | Introduction

Climate change is increasing the frequency and magnitude of extreme rainfall events globally (Alexander et al., 2006; Trenberth, 2011; Westra et al., 2014). and it is expected that these will increase even more in the future (Easterling et al., 2000). Lenderink and Van Meijgaard (2008) and Klein Tank et al. (2015) found that hourly precipitation extremes occur more often as a result of global warming in the Netherlands. Inevitably, short-duration extreme rainfall events could induce local floods (Madsen et al., 2014), which cause substantial damage.

Hydrometeorological forecasts with a high spatial and temporal resolution reduce damage by accurately predicting rainfall and streamflow which enable adequate warning when measures are needed. However, it is difficult to make accurate rainfall predictions as rainfall fields are highly variable in space and time (Emmanuel et al., 2012; Berenguer et al., 2005). Lobligeois et al. (2014) evaluated the conditions enhancing hydrological model performance with higher spatial resolution for a large set of catchments representing a variety of sizes and climate conditions. It was found that the impact of spatial rainfall information on discharge simulation depends on scale, catchment characteristics and event characteristics and concluded that the spatial representation of rainfall becomes more important for catchments with a high rainfall variability in the spatial domain. The temporal pattern of rainfall dominantly impacts both timing and magnitude of discharge peaks where its largest sensitivity can be found in small quickly responding catchments (e.g. Ball (1994) and Fabry et al. (1994)). For example, Berne et al. (2004) found hat a small urban catchment in the order of 10 km^2, requires rainfall sampling every 5 minutes, with a spatial resolution of 3 km indicating that the required sampling resolutions depend on catchment size relative to the spatial scale of a rainfall event. Thus, depending on the spatiotemporal requirements of a catchment, the quality of hydrometeorological forecasts is limited by the quality of their forcing input. Many hydrological studies acknowledged that inadequate streamflow simulations are caused by erroneous precipitation input (Krajewski and Smith, 2002; Borga et al., 2006; Bárdossy and Das, 2008; Moulin et al., 2009; Brauer et al., 2016).

Rain gauges and weather radar

Currently, weather radar and rain gauges are used as a common source to make quantitative precipitation estimates (QPE). Rain gauge networks provide accurate ground-based measurements, but are unable to capture the spatial variability of rainfall fields as their sampling resolutions are often coarse (Villarini et al., 2008). The national rain gauge network, employed by the Royal Netherlands Meteorological Institute (KNMI), consists of a manual gauge every 100 km^2 and an automatic rain gauge every 1000 km^2 reporting once per day and once per 10 minutes respectively. In small catchments, where it is plausible that no gauges are present at all, one is often forced to referred to the nearest available gauges outside the catchment (Brauer et al., 2016). A denser network and/or larger temporal measurement intervals would return more accurate QPEs (Villarini et al., 2008).

In contrast, weather radars are less limited by their spatial representation. In many countries, such as the Netherlands, radars retrieve rainfall information every 5 minutes on a nationwide grid with a spatial resolution of only 1 km^2. However, a major disadvantage of radar QPE is the considerable biases with respect to the true rainfall. Radars measure rainfall indirectly from reflectivity of sampled atmospheric volumes and are prone to multiple systematic and random errors (e.g. Uijlenhoet and Berne (2008), Krajewski et al. (2010), Hazenberg et al. (2013)). Usually, radar QPE underestimates real rainfall fields with a factor two (Overeem et al., 2009a,b), especially during stratiform events with low clouds in summer that are missed by the radar beam. In addition, the presence of ice crystals in the atmosphere results in systematic underestimations (e.g. Borga and Tonelli (2000)) which is often the case for convective situations with a high cloud base. These examples of biases prove that radar QPE may be less representative for ground level rainfall, while hydrologists are interested in rainfall observations at the surface.

Personal weather stations

Crowdsourcing is a non-traditional alternative which has potential to overcome issues related to spatial and temporal representativeness of rainfall observations made by traditional rainfall measurement techniques. A large and increasing amount of weather enthusiasts obtain rainfall data through non-traditional rainfall measurement sources. Crowdsourcing has already been investigated as a strategy to expand the set of traditional rainfall measurement techniques and potentially compensate for the known biases and limited availability of those techniques and equipment. An overview, state of the art and future prospective of crowdsourcing data

collection methods are presented in atmospheric sciences (Muller et al., 2015) and geophysics (Zheng et al., 2018). Amongst those are crowdsourced automatic citizen or personal weather stations (PWS) which are deployed by weather station owners and placed on their private properties. From there, stations upload real-time precipitation data every 5 minutes to online platforms such as Netatmo Weathermap, WOW-NL and Wundermap by the company Weather Underground (Muller et al., 2015). The stations measure more meteorological variables than rainfall (Meier et al., 2017), see also Section 2.4.1. Since PWSs are mainly located in densely populated regions, the highest network densities can be found in urban areas.

The accuracy of PWS had been investigated in previous studies (Jenkins, 2014; Bell et al., 2015; Meier et al., 2017; de Vos et al., 2017). Since network densities are highest in urban areas, the potential of PWS to monitor rainfall in urban areas has been explored in Amsterdam (Netherlands) (de Vos et al., 2017), the province of South Holland (Netherlands) (Golroudbary et al., 2018) and Norfolk (Virginia) (Chen et al., 2018). All studies concluded that the PWSs estimate true rainfall fields more accurately than their comparison real-time QPE.

Despite their potential for operational rainfall monitoring, the large-scale real-time application of PWS in meteorology and hydrology is constrained by limited data quality. Observation errors can be caused by instrumental errors, a compromised set-up and data processing issues (de Vos et al., 2017). For this reason, de Vos et al. (2019) developed a quality control (QC) method which is potentially applicable in real-time and excludes inaccurate measurements that are caused by typical errors of this data source with the ability to flag erroneous values and unflag once reliable data are produced again. The PWS network in Amsterdam was re-evaluated by the QC algorithm where data of one year was obtained from Netatmo Weathermap. This method improved the overall accuracy of a year of hourly rainfall depths with 11.3% bias reduction while maintaining 88% of the original dataset. In addition, nationwide application of the QC filter yielded high-resolution rainfall maps where the average density is found to be 1 station per 10 km^2 thus proving their potential to complement existing operational QPE, such as radar and traditional rain gauge networks.

Furthermore, crowdsourced measurements can be used to force hydrological models (Fletcher et al., 2013; Muller et al., 2015; Liu et al., 2016). Niemi et al. (2017) assessed the feasibility of open source rain gauges in Helsinki to force a rainfall-runoff model and Naus (2017) investigated multiple opportunistic data sources, including PWSs, to force a conceptual urban flood model in Amsterdam and Eindhoven (the Netherlands). Both studies emphasized that data quality enhancements of the crowdsourced observations are required to improve model outputs.

Although the density of PWSs is higher in urban areas and proof of its potential to be applied for urban hydrometeorological applications is given, little is known about this potential on the catchment scale, including rural regions where PWS densities are lower. van der Valk (2019) investigated the added value of calibrating QC-checked PWSs (using the method of de Vos et al. (2019)) to an operational radar product of consultancy firm Nelen & Schuurmans on national scale. He found that the with PWS calibrated product improved relative to the uncalibrated radar composite. Besides, a recent study in Germany explored the use of PWS to improve precipitation estimates and interpolation in the province of Baden-Württemberg (Bárdossy et al., 2020). This study performed an alternative QC check and concluded that filtering erroneous observations was necessary in order to make improvements relative to the German radar composite.

Recently, the Dutch water board Rijn & IJssel started a pilot study to enhance the coverage of their rain gauge network by placing extra PWSs from Netatmo in their management area (Scrumteam, 2020). With this pilot they aim to compensate for the shortcomings of the real-time available weather radar product, as explained before. Besides, the PWS are financially more attractive than traditional rain gauges. A vast network of Netatmo stations is already placed by weather enthusiasts in the region of Rijn & IJssel which are privately monitored and maintained by their owners. From the 405 stations, approximately 300 retrieve sufficient data (van den Houten, 2021). About a 100 stations will be added in 2021, leading to one station every $\sim$9 km^2.

The emergence of a QC filter for crowdsourced PWSs and growing interest of water managers to apply PWSs for operational purposes gives rise to assess the potential of quality-controlled personal weather stations (QC-PWSs) for hydrometeorological forecasting on the catchment scale. The QC filter allows for detection and filtering of erroneous data in a network of PWSs. The question remains whether data availability is sufficient after performing a QC check in a region with lower station density compared to urban areas. So far, the performance of PWSs has not been assessed within the boundaries of Dutch lowland catchments and no hydrological application studies have been performed on the catchment scale yet.

Research objectives

The objective of this study is to assess the accuracy of QC-PWSs in observing rainfall and predicting discharge in a lowland catchment. The study will focus on the Oude IJssel catchment including a sub-catchment, situated in the management area of water board Rijn & IJssel. Results contribute to making QC-PWSs operational for streamflow forecasting in the Netherlands.

Research questions

In order to reach the objective, the following supporting questions need to be answered:

Precipitation

- What is the data availability of PWS precipitation estimates in time after applying the QC filter with parameter settings suitable for the Oude IJssel catchment?
- What is the accuracy of QC-PWS precipitation estimates relative to the catchment-averaged reference radar rainfall product in the Oude IJssel catchment?
- What is the accuracy of the catchment-averaged QC-PWS precipitation estimates relative to the national operational and reference radar rainfall products in the Oude IJssel catchment and sub-catchment?
- What is the effect of network density on the data availability of PWS precipitation estimates in space and time after applying the QC filter?
- What is the effect of network density on the accuracy of QC-PWS precipitation estimates?

Discharge

- What is the accuracy of discharge predictions forced by QC-PWS precipitation estimates relative to discharge prediction forced by the raw PWS, national operational and reference radar rainfall products in the Oude IJssel catchment and sub-catchment?
- How do errors in rainfall measurements propagate in the predicted discharge during two events?

2 | Data and study area

This chapter discusses the data sources and study area consulted for this research. First, the three quantitative precipitation estimates will be described (Section 2.1) where the catchment (Section 2.2), discharge (Section 2.3) and evapotranspiration data (Section 2.4) used in this study will be presented.

2.1 Precipitation estimates

In this study, three precipitation data sources will be evaluated for the period between 01-09-2019 and 01-09-2020 within the boundaries of catchment the Oude IJssel. They originate from (1) personal weather stations (PWS), (2) real-time weather radar data and (3) weather radar data adjusted by rain gauge data (Table 2.1). The dataset length is limited by the availability of PWS data and are therefore no longer than the prescribed year that will from now on be referred as Sept. 2019 - Sept. 2020. Rainfall measured by Personal Weather Stations are subject to erroneous measurements which require elimination through quality-control filtering. Both raw and quality-controlled (QC) data from PWSs and real-time radar data were validated against the rain gauge-corrected radar and their performance was compared. Though QC-PWS data and raw PWS data were both validated, they originate from the same data source. The following sections provide a description of the three precipitation data sources.

Table 2.1: Overview the three different precipitation products, their spatiotemporal resolution and data latency.

QPE	Temporal resolution	Spatial resolution	Data latency
PWS	∼5 min	∼3 km x 3 km	Real-time
Unadjusted radar	5 min	1 km x 1 km	Real-time
Gauge -adjusted radar	5 min	1 km x 1 km	1-2 months

2.1.1 Personal weather stations

Personal weather stations are privately deployed and monitored and retrieve a time series (UTC) with corresponding rainfall estimates. The PWSs that are subject of this study are from the brand Netatmo which are equipped with sensors measuring temperature, relative humidity and barometric pressure. Additionally,

Figure 2.1: Topview of a Netatmo rain gauge module. Source: http://www.atsmartliving.com/netatmo-rain-gauge-clouds-pass-but-rain-remains/. (Published: May 21, 2019).

modules are available that measure rain and wind. According to the Netatmo product specifications the rain module is a tipping bucket with a collection funnel of 13 cm in diameter that observes in multiples of 0.101 mm, has a measurement range of 0.2-150 mm h^{-1} and a measurement accuracy of 1 mm h^{-1} according to the specifications of Netatmo.

The average density of Netatmo PWSs including a rainfall module in the Netherlands was 1 per ∼10 km ^{2} in 2018 (de Vos et al., 2019). At the time of writing, it is uncertain what the national resolution of the PWS network is for the studied year. However, the spatial resolution in 2018 is assumed to be a good approximation for its succeeding year at the start of the study period. As a growing number of PWSs are linked the Netatmo platform, it is expected that the average PWS density will increase every year. The pilot study from water board Rijn & IJssel aims to place a PWS every ∼3 by 3 km, that corresponds to 1 per ∼9 km^2, in their management area and the upstream areas in Germany.

Netatmo has its own online platform collecting and visualizing data from all operational Netatmo stations distributed globally which is called Netatmo Weathermap (https://weathermap.netatmo.com/). All Netatmo stations upload their real-time observations every

5 minutes to the Netatmo Weathermap, though exact measurement intervals vary per time step (de Vos et al., 2017, 2019).

In the Netherlands, there are two other online platforms collecting and visualizing PWS data: WOW-NL (https://wow.knmi.nl/) and Wundermap of company Weather Underground (https://www.wunderground.com/wundermap).
Weather station owners can link their device to these platforms. This results in networks of various types of PWS devices. Before 2018 there was an agreement between Netatmo and Weather Underground where all Netatmo PWS measurements were automatically linked from the Netatmo Weathermap to the Wundermap, with some processing effects like rounding and delays in the rainfall observations as presented on the Wundermap (de Vos et al., 2017). All Netatmo devices are automatically linked to the Netatmo Weathermap, and the station owners can decide to link them to other platforms manually as well. Both Netatmo Weathermap and Wundermap provide an Application Programming interface (API) to retrieve data from, which is not the case for the WOW-NL platform.

Netatmo Weathermap is favoured over the other two platforms since it provides the highest measurement and network density and the use of an API. While anyone can obtain PWS data using this API, for this study, raw PWS data is retrieved and made available by KNMI in the large quantities of the study period according to agreed terms. This holds that the dataset is not public.

2.1.2 Unadjusted radar

An unadjusted real-time radar product (R_{rad}), that is applied operationally, was consulted and used to compare with the other real-time rainfall estimate; the network of personal rain gauges. The radar composites originate from two C-band dual-polarized weather radars operated by KNMI, located in Den Helder and Herwijnen. The most detailed composite of KNMI, the NL-radar NL25, will be used for this study. It has a grid cell size of 1 km^2 and measures with 5 min intervals. The domain of this radar product extends over the Dutch national border. Besides, the water board uses this extension operationally for making hydrological forecasts for the Oude IJssel and is therefore taken in study as well. The radar product is available as open-source data that can be retrieved from https://dataplatform.knmi.nl/open-data-info/.

2.1.3 Gauge-adjusted radar

The gauge-adjusted radar product is described in Overeem et al. (2009a; 2009b) and upgraded more recently (Beekhuis & Mathijssen, 2018). This radar product is adjusted by 31 automatic rain gauges and post processed by spatial adjustments using the manual network of 325 rain gauges. These data are made available by the KNMI with a delay of 1-2 months and therefore only applicable for studies of past events. In order to validate the performance of the PWS network, this gauge-adjusted product will be used as a reference (R_{ref}). The spatiotemporal resolution is equal to the unadjusted radar product and is also available as open-source data that can be retrieved from https://dataplatform.knmi.nl/open-data-info/. For the German part of the Oude IJssel, a gauge-adjusted radar product from Germany was used for the period Sept. 2019 - Sept. 2020. This is a similar offline gauge-adjusted product, named RADKLIM-YW (YW = 5 min) which has a spatial resolution of 1 km^2. Also, the German radar composite is open source and made available by the Deutsche Wetterdienst (DWD), via their web portal (https://opendata.dwd.de/).

2.2 Oude IJssel catchment

The Oude IJssel catchment is chosen as study area. It is situated on the border between the Netherlands and Germany (Figure 2.1). Reasons for this choice are supported by the demand of Water board Rijn & IJssel to extend their rain gauge network that could compensate for the challenges they face with the operational radar products near the edge of the Dutch radar domain. Furthermore, a vast network of PWS were already placed in this catchment that has a suitable size.

The catchment is 1210 km^2 of which 363 km^2 is located in the Netherlands and 847 km^2 in Germany (en Ijssel, 2014b; Drost, 2016). The Oude IJssel springs in Germany, south of Borken. Across the border the Aa Strang and the Slinge connect to the main stream and discharge their surface water in the IJssel near Doesburg. Elevation differences are minor where the gradient gradually decreases from east to west from 70 m in Heiden and 10 m in Doesburg (AHN Viewer, 2021) where the Oude IJssel drains freely with an average stream gradient of ~ 0.8 m km $^{-1}$.

The Oude IJssel consists of 140 sub-catchments where the water board assigned 10 lumped sub-catchments for hydrological modelling. Model parameters of those lumped sub-catchments have already been estimated by the water board and are granted for this study (refer to Section 3.5). This research considered

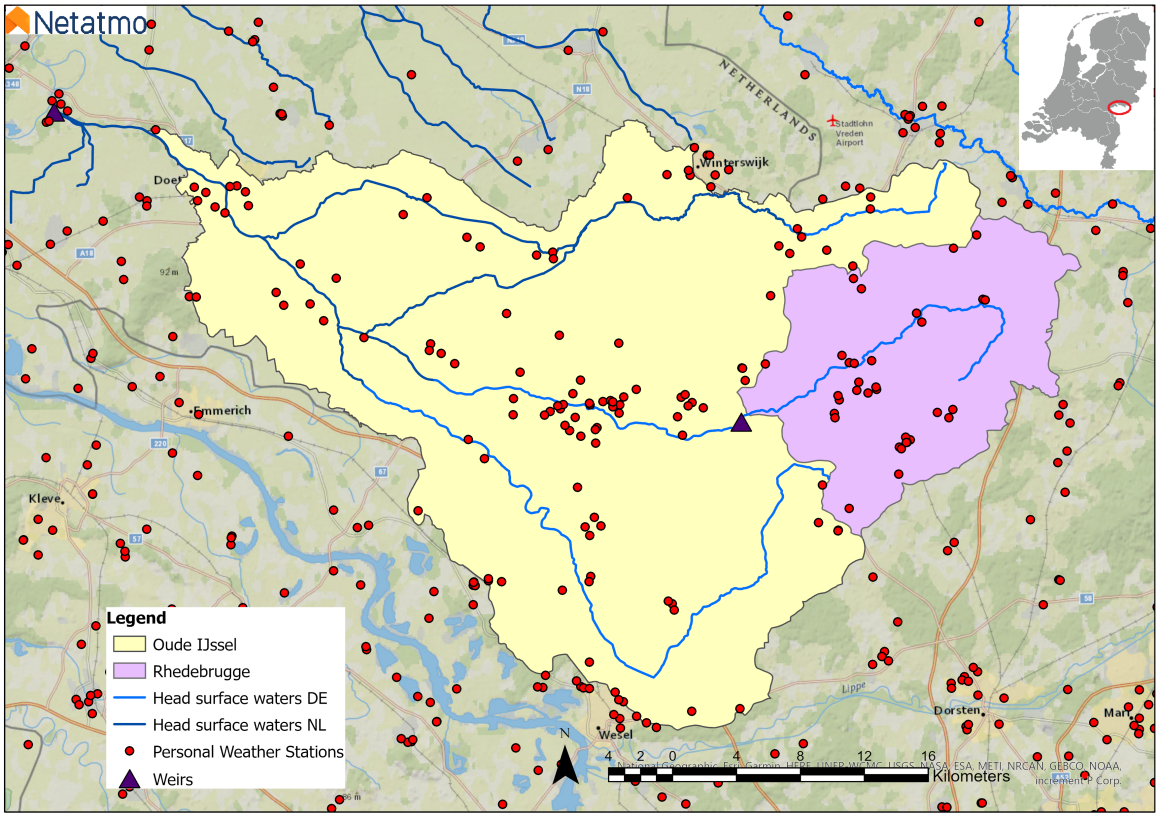


Figure 2.1: Location and spatial overview of the Oude IJssel catchment (yellow) and Rhedebrugge sub-catchment (purple), outlets (dark purple triangles and PWS locations in the region (red dots).

one sub-catchment, Rhedebrugge (Figure 2.1). Rhedebrugge is 224 km^2 and completely situated in Germany where the Aa Strang springs. This upstream part of the German stream was chosen for three reasons. Firstly, Rhedebrugge is not connected to the other head waters in the catchment. Hydrograph characteristics from this upstream region can be compared with the main catchment outlet in the downstream end. Secondly, it was the largest calibrated lumped sub-catchment available. Thirdly, most PWSs are located in Rhedebrugge sub-catchment.

2.3 Discharge

The catchment of the Oude IJssel has a larger buffering capacity south-east of the German border. During precipitation peak events, water is retained in the soil that causes a delay in the runoff peak. The Aa strang in Rhedebrugge on the other hand, is more canalised and therefore shows much sharper runoff peaks than the German Oude IJssel. The Slinge, located in the northern upstream part of the catchment, drains all year around. On average, discharge is 10.5 m^3 at the outlet in Doesburg and 1.6 m^3 in the Aa Strang (en Ijssel, 2014a). Hourly summations of discharge data (in mm) from the outlet and weir in Rhedebrugge are freely accessible and obtained from the open data portal of water board Rijn & IJssel (https://waterdata.wrij.nl/).

The outlet in Doesburg consists of a weir and sluice. The weir controls the water surface level upstream of the outlet meaning that the discharge is affected by this weir. Even when no rain has fallen, varying discharge peaks may occur as a result of deviating valve positions. The resulting noise in the discharge is filtered out by a moving average window of 12 hours based on the method of Drost (2016), so that visible observed discharge peaks are caused by rainfall events.

2.4 Evapotranspiration

Reference evapotranspiration (ET_{ref}) data were made available by the water board which retrieves ET data from a KNMI automatic weather station located in Hupsel. Though the location is a single ground truth measurement ~20 km away from the catchment centre, ET_{ref} in Hupsel is assumed to be the same as ET_{ref} in the Oude IJssel. Although, nearly 50 % of the land use is dominated by agriculture in the catchment (Drost, 2016), ET_{ref} was not corrected with crop factors and therefore assumed to approximate the potential evapotranspiration (ET_{pot}). Since this study used one evapotranspiration time series with four precipitation input

datasets, it was assumed that the effect of an uncorrected ET_{ref} was negligible in the validation and comparison of the four precipitation estimates.

3 | Methods

This chapter discusses the methodological steps of this study. First, the methodology of the PWS quality-control filter and how it is applied to the Oude IJssel is given in Section 3.1. Next, steps taken to derive catchment-averaged time series are given in Section 3.2. Validation methods of the rainfall products are described in Section 3.3. Then, hydrological model initiation and calibration are described in Section 3.4 followed by validation methods of discharge simulations in Section 3.5. Lastly, the effect of the PWS network density is given in Section 3.6.

3.1 Quality-controlled PWS

In this section, the methods of the QC filter for PWS are explained (3.1.1), followed by an elaboration about the filter design applied to the Oude IJssel PWS network (3.1.2).

3.1.1 Quality-control filter

For this study, the QC filter of de Vos et al. (2019) isused that is specifically designed to filter typical errors associated with crowdsourced PWS rainfall observations. The method relies on the assumption that nearby stations should measure similar rainfall dynamics, and that the network is dense enough that the majority of a cluster of stations is able to accurately capture the event. Therefore, erroneous measurements can be recognized from comparisons with nearby observations. The QC algorithm compares the observations of a given station with the median observations of its neighbouring stations and is for this reason not dependent on other data sources. The filter is designed to attribute flags to measurements where a 0 represents "no error", 1 "error" and -1 "not enough information available to determine error". The algorithm does this for four types of typical errors: faulty zeroes (FZ), high influx (HI) and station outliers (SO) and station bias (BC). The filter design consists of four corresponding modules which rely on a set of 11 self-provided parameters indicating the range within stations are considered neighbours, the minimum number of observations to determine median values of the neighbours, minimum period of comparison and other threshold values. Those four modules are:

1. Faulty zeroes (FZ) filter

   Faulty zeroes are communicated to the platform when the tipping bucket mechanism is completely obstructed due to e.g. a tilted rain gauge or physical obstructions and no tip occurs, also during rainfall. All stations within a range of d meter around a given station are selected to compute the median rainfall over the surrounding area. If fewer than n_{stat} neighbouring stations with rainfall measurements are available, the median cannot be calculated and the FZ flag is set to -1. The FZ flag is set to 1 if the median rainfall is larger than zero for at least n_{int} time intervals while the station itself reports zero rainfall. Until the stations report non-zero rainfall, the FZ flag remains 1.

2. High influx (HI) filter

   High influx measurements that are unrelated to weather, e.g. caused by the owner when liquid is poured through the rain gauge for calibration and cleaning of the device or sprinklers in the vicinity. Also, the filter for high influxes makes use of a comparison with the median rainfall from all stations within a radius of d meter around a given station. If the median amount does not exceed the threshold value ϕ_A, the HI flag is set to 1 for any rainfall value from the station itself above threshold ϕ_B. During more intense rainfall, when the median of surrounding stations report of ϕ_A or higher, the threshold becomes variable. Only if the station's measurements exceed the median times ϕ_B/ϕ_A, a 1 will be assigned to the HI flag. If fewer than n_{stat} neighbouring stations report observations, HI flag is set to -1.

3. Station outlier (SO) filter

   Station outliers measured by PWSs do not correspond with local rainfall dynamics e.g. when the reported station location is incorrect or in the rare occasion where for a period of time, rainfall is recorded in repeated daily cumulative amounts, thus resulting in far too high values. To determine whether a station yields nonsensical measurements for that location, it is compared with time series of neighbouring stations within a range d. m_{int} intervals previously to the current measurement, or any longer interval where the station reports at least m_{rain} intervals of non-zero rainfall measurements, are compared. There need to be at least n_{stat} stations with at least m_{match} intervals overlapping with the evaluated station to compute the SO flag. The Pearson correlation (r) (Equation 3.3) and bias (Equation 3.4) with all neighbouring stations are calculated. If the median of the Pearson correlation of all neighbouring stations is below threshold value γ, the SO flag is set to 1.

4. Bias Correction (BC)

Individual PWSs can systematically over- or underestimate rainfall, with a possible overall bias in the network. The filter makes use of a bias correction method to compensate for this systematic instrumental error. The filter makes use of a bias correction method to compensate for this systematic instrumental error. The initial bias correction factor (BCF) is called the default bias correction factor (DBC). The DBC is a single value proxy of the correction needed by an existing PWS network determined over a period with typical rainfall for the local climate prior to application of the QC-filter. In this study, the DBC is calculated one-off offline by the mean bias of the 5 min catchment-averaged PWS time series during the month preceding the start of study period, when compared with the catchment-averaged gauge-adjusted radar time series. Since the dataset starts at 01-09-2019, the month September was used as warm-up period for the QC filter and therefore excluded in the validation study. Prior to the derivation of the PWS catchment-averaged time series (Section 3.2.2), intervals were excluded that were flagged -1 and 1 for FZ and HI. In this way, the DBC (Table 3.1) becomes:

$$DBC = \frac{1}{1 + median(bias)} \tag{3.1}$$

where the bias is given in Equation 3.4 in Section 3.3.2. The BCF is dynamic for individual stations and in time. A new BCF of a station is calculated provided the median of the Pearson correlation from all neighbouring stations exceeds threshold γ. BCF$_{new}$ is calculated by the median bias of all neighbouring stations. If $\log \frac{BCF_{new}}{BCF_{old}} > \log(1+\beta)$, BCF$_{old}$ will be replaced by BCF$_{new}$.

The QC filter provides two options to filter data: one can decide to include all measurements unless they are flagged as erroneous by at least one module, i.e. include all intervals with flags 0 and -1 ("Filtered Flex"), or also exclude the intervals where there was too limited information to allocate a flag, i.e. include all intervals with flags 0 ("Filtered Strict"). "Filtered Flex" where all time intervals flagged with 1 and and "Filtered Strict" where all intervals flagged with both 1 and -1, are left out from the dataset and marked as 'NA'.

A detailed description of the filter, parameter definition and default settings and supporting information visualizing the iterative steps of the filter modules are provided in the documentation of de Vos et al. (2019). The code is freely accessible and can be found on https://github.com/LottedeVos/PWSQC.

Table 3.1: Parameter settings of the QC filter

Filter parameter	Value
d (m)	10,000
n_{stat}	5
n_{int}	6
ϕ_A	0.4
ϕ_B	10
m_{int}	4,032
m_{rain}	100
m_{match}	200
γ	0.15
β	0.2
DBC [September 2019]	0.92

3.1.2 Application to the Oude IJssel catchment

The QC filter is applicable to any gauge network provided that a minimal number of stations is present in the network where a group of neighbouring stations measures similar rainfall dynamics. Those conditions are defined in the filter parameter settings that should be considered carefully for each network. The default values of the filter parameter settings are based on the validation of the QC on the PWS dataset (1 May 2017 till 1 June 2018) in the Amsterdam metropolitan area and can be found in de Vos et al. (2019). This study concerns a relatively sparse network compared to the city of Amsterdam and therefore it was tested whether the default parameter settings are applicable to the network of the Oude IJssel as well. The filter requires a sufficient number of neighbours in order to attribute flags, defined by the number of observations within d distance that need to exceed at least n_{stat}.

From the PWS dataset, a subset of the national network was created covering all stations in the catchment area including the German region within the catchment boundaries. A distance of 10 km around the catchment boundaries, equal to d, was chosen as starting value of the buffer zone that created the dataset used in this study. At the time of pre-processing, only the PWS data from September until November 2019 were made available. For this reason, metadata used to determine the filter parameter settings are based on this smaller dataset and contained 300 stations.

The minimum number of stations to compare with (n_{stat}) should be sufficiently large for reliable medians that represent actual weather. The range d should not be too large as stations within this distance should represent similar rainfall patterns at these time scales. This network includes all stations within the catchment area plus a boundary equal to 10 km around the border. For

this network, the number of neighbours that each station has was calculated for a number of choices for range d. As long as a large fraction of stations has more than n_{stat} neighbours, the expectation is that for the majority of the dataset, the requirements are met for a flag to be attributed.

The default range of 10 km included a single group of stations that could not meet the minimum required condition of $n_{stat} = 5$ with only 2 neighbours present within this group (Figure 3.1). This means that if the QC filter would be applied, only 2 stations would be filtered out of the dataset regardless the quality of the observations. The majority of the stations met the requirements while still having over 20 neighbours, thus even though those stations contain intervals with no data, the requirement of 5 stations is reached. This outcome was considered sufficient and so it was concluded that all parameters, except for the DBC, were applicable for this network (Table 3.1).

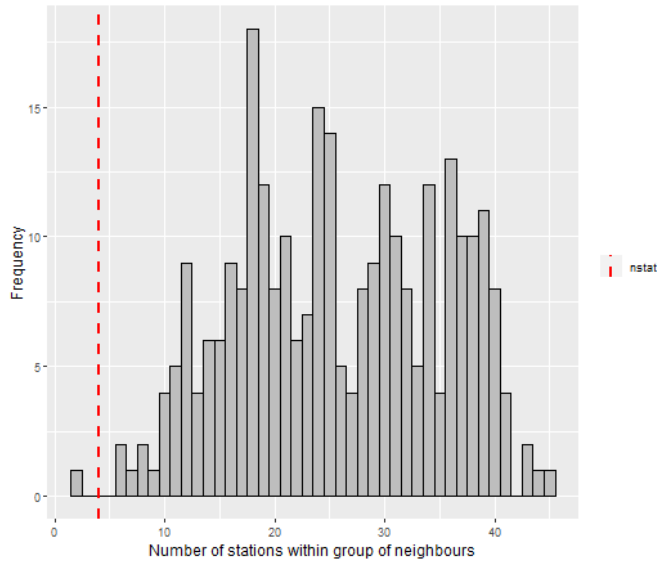


Figure 3.1: Histogram of the number of neighbours distributed per neighbour group with $d = 10$ km and $n_{stat} = 5$.

3.2 Catchment-averaged time series

For the Oude IJssel and Rhedebrugge sub-catchment a precipitation time series were made for R_{rad}, R_{ref} and R_{QC-PWS}. The two radar products, R_{rad} and R_{ref}, were clipped with the catchment boundaries of the Oude IJssel and Rhedebrugge and converted to catchment-averaged rainfall time series. Point measurements retrieved by all PWSs are interpolated in space prior to generating an averaged time series.

3.2.1 Radar products

The catchment-averaged gauge-adjusted radar combines both Dutch ($R_{ref,Dutch}$) and German reference data ($R_{ref,German}$). RADKLIM composites are projected in RADOLAN stereographic projection system (DWD, 2021). Before any clipping could have taken place, $R_{ref,German}$ first needed to be re-projected to the same coordinate reference system as $R_{ref,Dutch}$: RD new. Next, a sample of the $R_{ref,German}$ raster file around the German part of the catchment was created. Both $R_{ref,Dutch}$ and $R_{ref,German}$ could be clipped to Dutch and German part of the Oude IJssel catchment. They are respectively 388 and 914 pixels of 1 km^2 in magnitude, so that the catchment-averaged time series of R_{ref} becomes:

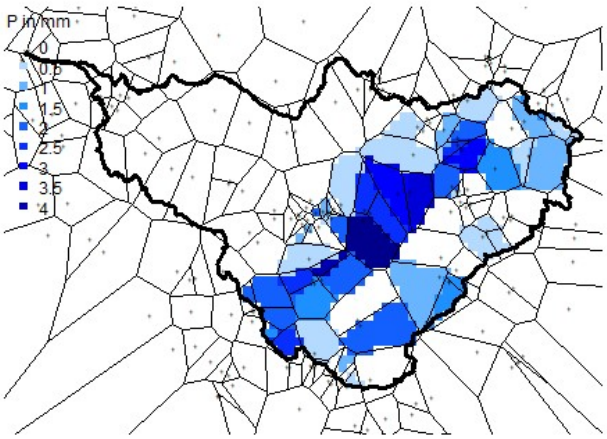
$$R_{ref} = \frac{388}{1302} R_{ref,Dutch} + \frac{914}{1302} R_{ref,German} \quad (3.2)$$

The domain of R_{rad} extends far enough over the German border to cover the entire study area. Therefore, Rraw was clipped to the complete catchment boundaries of the Oude IJssel and Rhedebrugge. Over all 1302 pixels of 1 km^2, average rainfall sums for all 5 min intervals were calculated.

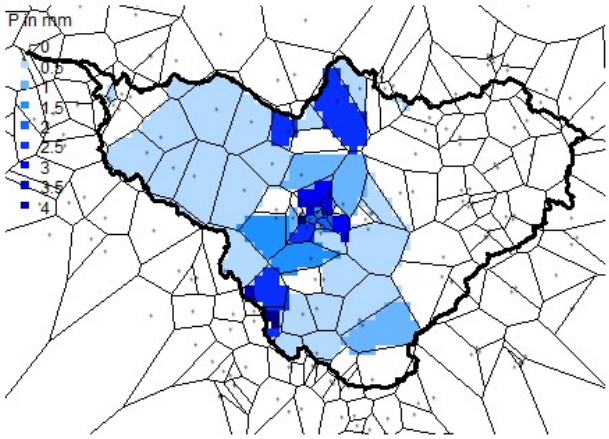
3.2.2 Interpolation of PWS

The PWS dataset over the complete study period (1 Oct. 2019 – 1 Sept. 2020), contains 5 min accumulated time series where a total of >300 PWSs are located within the area of catchment plus a boundary of distance equal to range d, with >100 PWSs within the catchment boundaries.

Rainfall measured by the extended subset were interpolated in space through the Thiessen method. Stations outside the borders can contribute to the average if their Thiessen polygon crosses the catchment boundary. Next, catchment-averaged rainfall time series were created from the area within the catchment boundaries for the raw PWS data, QC-PWS data filtered from -1 flags and QC-PWS data filtered from both -1 and 1 flags. Before average rainfall sums per interval were calculated, Thiessen polygons were drawn around the stations with remaining measurements for that time interval. Subsequently, each interpolation step is unique in time because of the unique set of the retained QC-PWS dataset where 0,-1 and 1 flags are attributed to each time interval. To illustrate, Figure 3.1 gives a visual representation of QC-PWS rainfall sums during two precipitation events for two 5 min intervals that ended at 09/02/2020 22:00 and 06/05/2020 11:00.



(a) 09/02/2020 22:00



(b) 06/05/2020 11:00

Figure 3.1: Two examples of QC-PWS 5min rainfall sums uniquely distributed in space by the Thiessen method.

3.3 Validation and comparison of rainfall products

This section first evaluates the available data before and after the quality control (Section 3.3.1). Next, validation methods that describe how well the quality-controlled PWSs measure rainfall with respect to the gauge-adjusted radar are explained (Section 3.3.2), followed by validation methods evaluating the performance of quality-controlled PWSs on a catchment scale (Section 3.3.3).

3.3.1 Data availability of quality-controlled PWS

The raw crowdsourced PWS observations are known to contain data gaps. Applying the QC filter will reduce the number of observations even more. Since catchment-averaged precipitation estimates were used to force the hydrological model (Section 3.5.1), it is important to evaluate the magnitude and variability of rainfall observations in space and time. Therefore, the data availability before and after QC were evaluated for the Oude IJssel and Rhedebrugge sub-catchment over the study period. The study period entails the length of the PWS dataset minus the warm up period which is from 1 Oct. 2019 till 1 Sept. 2020.

3.3.2 Validation of quality-controlled PWS

First, rainfall measurements made by the PWSs in the network were validated on individual basis. This holds that every 5 min accumulated rainfall measurement of each PWS (R_{PWS}) will be validated against the averaged 5 min accumulated measurement of the gauge-adjusted radar $\bar{R}_{ref}$.

In order to validate observations, the Pearson correlation (r), the relative bias (called bias from now on), and the coefficient of variation of the errors (CV) are calculated using the following equations:

$$r = \frac{cov(R_{PWS}, \bar{R}_{ref})}{sd(R_{PWS}) \cdot sd(\bar{R}_{ref})} \tag{3.3}$$

$$bias = \frac{\overline{\Delta R}}{\bar{R}_{ref}} \tag{3.4}$$

with

$$\Delta R = R_{PWS} - \bar{R}_{ref} \tag{3.5}$$

$$CV = \frac{sd(\Delta R)}{\bar{R}_{ref}} \tag{3.6}$$

where R_{PWS} represents the individual PWSs time series and $\bar{R}_{ref}$ the catchment-averaged time series of the gauge-adjusted radar. Along with these performance metrics, the percentage of available data over the complete study period was calculated before and after quality control. At last, all metrics were calculated for all individuals filters as well (Faulty zeroes, High influx, Stations outliers and Bias corrected).

3.3.3 Validation methods catchment-averaged rainfall

The catchment-averaged time series of the QC-PWS, raw PWS and unadjusted radar data are validated with respect to the gauge adjusted radar. Just as in Section 3.3.2, the r, bias and CV are calculated according to the following equations:

$$r = \frac{cov(\bar{R}_{qpe}, \bar{R}_{ref})}{sd(\bar{R}_{qpe}) \cdot sd(\bar{R}_{ref})} \tag{3.7}$$

$$bias = \frac{\overline{\Delta R}}{\bar{R}_{ref}} \tag{3.8}$$

with

$$\Delta R = \bar{R}_{qpe} - \bar{R}_{ref} \tag{3.9}$$

$$CV = \frac{sd(\Delta R)}{\bar{R}_{ref}} \tag{3.10}$$

where ($\bar{R}_{qpe}$) represent the three quantitative precipitation estimates. The discharge simulations forced by these $\bar{R}_{qpe}$ were also validated (Section 3.6).

3.4 PWS network density

To analyse the effect of network density, the spatial resolution of the existing PWS network in the study area was varied. Since the total network of existing PWSs was already employed, it was only possible to decrease the network size. A sample size of one third of the original PWS network in the study area was chosen (including the buffer zone of 10 km) through a random sample test. Multiple samples with one sample size were chosen in order to compensate for the unique network layout of the random selected samples and balance with the time needed for the datasets to be quality-controlled.

In total, 13 random samples were made and compared using histograms as was done in Section 3.1.2. The QC filter parameters of Table 3.1 were taken, so no parameter adjustments were made for the sample selection. The aim was to find samples that (1) meet the parameter requirement of n_{stat} best and that did worst in meeting the requirements of the minimum number of stations within a range of 10 km. The percentage of stations within the sample that did not meet the parameter requirement n_{stat} was calculated for all 13 samples. Samples with the best and two worst results were, from best to worst: sample 1 yielded 6.84%, sample 2 9.40% and sample 3 11.11% (Table 3.1). The histograms representing the number of stations distributed per neighbour group including maps with the spatial distribution of the three samples within the study area are presented in Figure 3.1.

Sample 1, 2 and 3 were chosen for QC filter application. First, datasets of selected samples were created that were one third in size of the total PWS dataset. Then, all steps to filter the raw PWS data were repeated according to Section 3.1. As explained in Section 3.1.1, the Default Bias Correction (DBC) is a unique correction factor for each network layout. Therefore, it was calculated for samples 1, 2 and 3 (Table 3.1) according to Equation 3.1. This meant that the interpolated averaged time series of the sampled PWS data was excluding -1 and 1 flagged intervals for FZ and HI was used to calculate the DBC. After full application of the QC filter, a selection of the best and worst sample was made after assessment of available data (Section 3.1.1.) and validation of quality-controlled PWS (Section 3.3.2). Rainfall measurements of the two remaining PWS samples are interpolated in space (Section 3.3.2) after which the raw ($\bar{R}_{PWS,SMP}$) and filtered ($\bar{R}_{QC-PWS,SMP}$) catchment-averaged time series are validated with respect to the gauge-adjusted radar ($\bar{R}_{ref}$) (Section 3.3.2) and compared with the original PWS network ($\bar{R}_{QC-PWS}$)and $\bar{R}_{rad}$.

Table 3.1: The percentage of stations within the sample that did not meet parameter requirement n_{stat}; DBC values of the sample datasets calculated over the $\bar{R}_{PWS}$ time series in September 2019 exluding intervals that were flagged -1 and 1 for FZ and HI,.

Sample number [-]	Percentage [%] $< n_{stat}$	DBC [-]
Sample 1	6.8	0.88
Sample 2	9.4	0.96
Sample 3	11.1	0.95

3.5 Hydrological application

3.5.1 WALRUS

To make discharge predictions from the QC-PWS and radar precipitation estimates, the Wageningen Lowland Runoff Simulator (WALRUS) was used as hydrological model. There are multiple reasons for the choice for this model. First, WALRUS is a lumped rainfall-runoff model especially designed for lowland catchments such as the Oude IJssel. Important couplings and processes for lowland catchments that are included in WALRUS are the groundwater-unsaturated zone coupling, wetness-dependent flow routes, groundwater-surface water feedbacks, seepage and surface water supply (Figure 3.1) (Brauer et al., 2014). Secondly, water board Rijn & IJssel uses WALRUS when making hydrological forecasts and already estimated the model parameters for Rhedebrugge sub-catchment. WALRUS requires at least rainfall (P) and potential evapotranspiration (ET_{pot}) as input and simulates groundwater depth (dG), actual evapotranspiration (ET_{act}) and discharge (Q). The latter output variable, discharge, is the only variable of interest during this study.

3.5.2 Calibration

Model parameters for the Oude IJssel and Rhedebrugge sub-catchment had already been estimated (Drost, 2016; WRIJ, 2021). Four parameters and one initial condition required automatic recalibration. This was done for

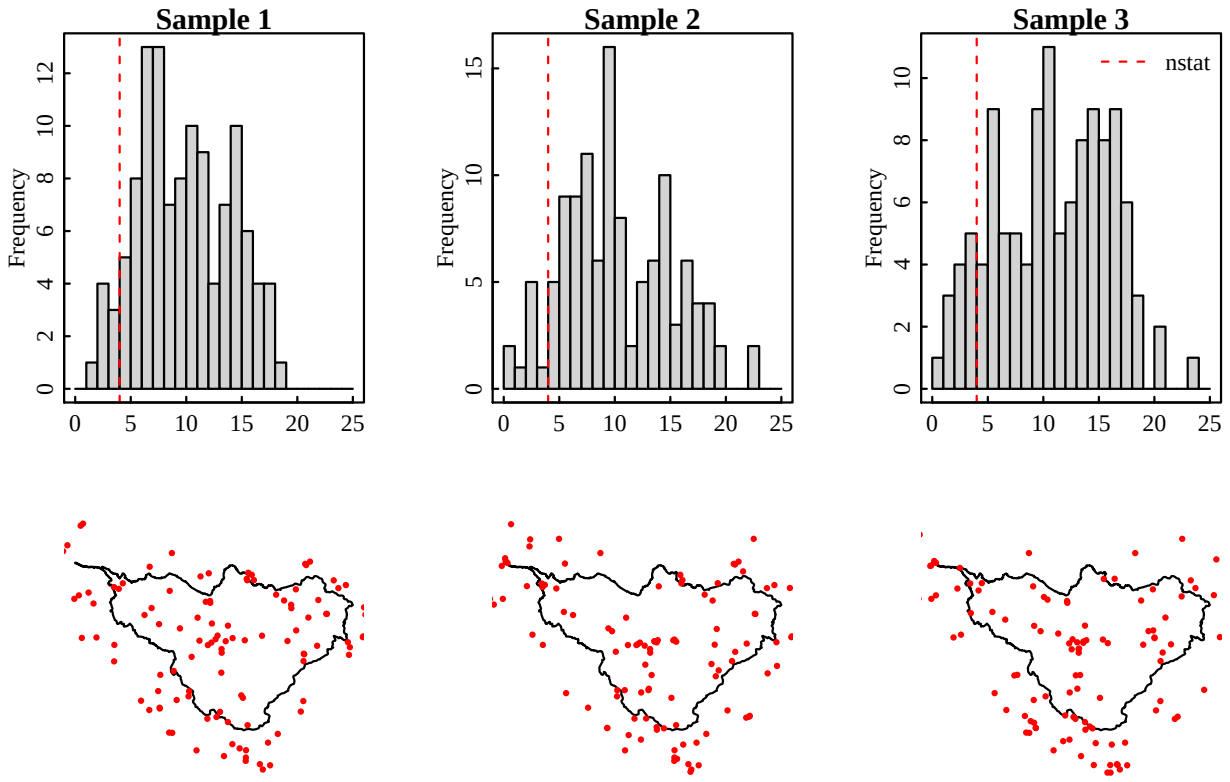


Figure 3.1: Histograms of the number of neighbours distributed per neighbour group with $n_{stat} = 5$ and the spatial distribution of the PWS locations of samples 1, 2 and 3 .

both catchments for the wetness index paramater cW, the groundwater reservoir constant cG, the quickflow reservoir constant cQ and the surface water parameter cS. Since no groundwater data were available, an indirect value of the initial groundwater table was supplied and included in the calibration which is the fraction of initial discharge originating from drainage G_{frac} (Table 3.1). Since the model is relatively unsensitive to the cV parameter, it was decided to not automatically recalibrate cV. The precalibrated values from literature were used as starting values for the automatic recalibration. For the calibration and validation of WALRUS, $\bar{R}_{ref}$ and ET_{pot} were used as input data and Q as observed target. The calibration started at 01-01-2019 and lasted till 01-04-2020 and the validation period was between 01-09-2019 and 01-09-2020. Despite the overlap with the validation period, it was decided to include the autumn season of 2019 and summer season of 2020. A calibration period of one year could give well calibrated parameter values (Brauer et al., 2014b). Besides, calibration results improved significantly when two summer seasons were included in the calibration period. Automatic calibration was executed applying the Levenberg-Marquardt optimisation algorithm on a temporal resolution of 1 hour. The four parameters were slightly adjusted during 6 iteration steps retrieving the least deviation between observed and modelled discharge. The Nash-Sutcliffe efficiency (NSE) was used as measure of goodness of fit of the WALRUS model (Brauer, 2017). A NSE value of 0.93 and 0.86 were retrieved for the main and sub-catchment respectively. Graphical output including performance metrics of the calibration can be found in Appendix A.1.

Table 3.1: Calibrated WALRUS parameters for the Oude IJssel and Rhedebrugge sub-catchment. * = *values retrieved by automatic calibration.*

Parameter	Oude IJssel (Drost, 2016)	Rhedebrugge (WRIJ, 2021)
cW [mm]	472.27*	449.30*
cV [h]	2	51
cG [mm h]	1.00*10^6*	17*10^6*
cQ [h]	33.57*	3.64*
cS [mm h^{-1}]	0.495*	0.63*
cD [mm]	1600	2300
G_{frac} [-]	0.1*	1*
a_s [-]	0.01	0.01
s_t	Sand	Loamy sand

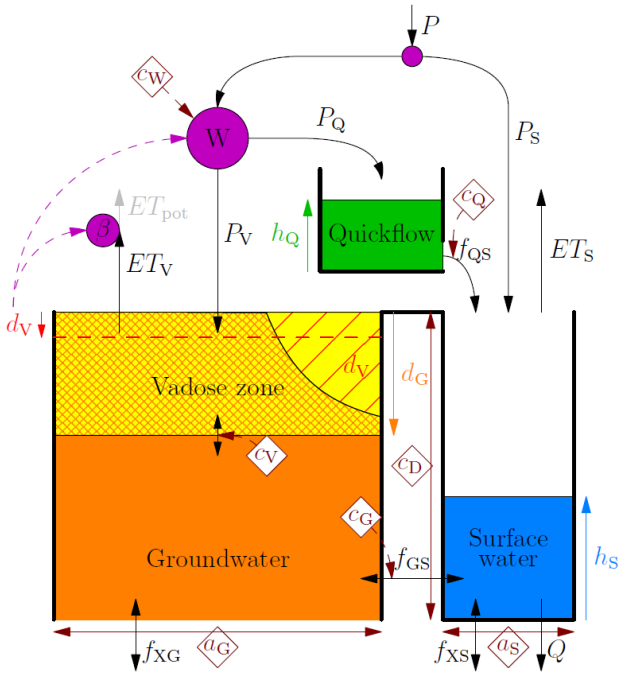


Figure 3.1: Overview of the model structure with the five compartments: land surface (purple), vadose zone within the soil reservoir (yellow/red hatched), groundwater zone within the soil reservoir (orange), quick flow reservoir (green) and surface water reservoir (blue). Fluxes are black arrows, model parameters brown diamonds and states in the colour of the reservoir they belong to (Brauer et al., 2014a)

.

3.5.3 Event Selection

Two precipitation events were selected during the study period, one during summer and one during summer season. Visual assessment of gauge-adjusted radar data resulted in the selection of two precipitation peaks: on February 23rd and on June 5th. Though June 5th does not fall in summer yet, the simulated groundwater table was low at the start of the event with 2.15 m in depth. On February 23rd the simulated groundwater level was just 1.16 m below surface. The dates, duration and precipitation sums of the selected event are visualized in Table 3.2.

The lag time was computed in two ways. For event 1, it was the time difference between the centre of mass of the rainfall event ($TP_{c,m}$) and the moment of maximum discharge (TQ_{max}). For event 2 it was the time difference between the centre of mass of the rainfall event and the centre of mass of the discharge peak $TQ_{c,m}$. The centre of mass is defined as half the precipitation or discharge sum during the event of target. Different methods were chosen because the maximum discharge peak of the second event already had reached after one hour and would have been too short for analysis. The events' precipitation sums were calculated over the start of the event till $TQ_{c,m}$ or TQ_{max}. The lag time, timing of the discharge peak and sum during the runoff period are visualized in Table 3.3.

Table 3.2: Start and end dates, duration and precipitation sums (P_{sum}) of the two selected precipitation events in winter (event 1) and summer (event 2).

	Dates	Duration [h]	Oude IJssel P_{sum} [mm]	Rhedebrugge P_{sum} [mm]
Event 1	23-02-2020 00:00 – 16:00	16	28.1	27.9
Event 2	05-06-2020 08:00 – 20:00	12	15.4	16.5

Table 3.3: Dates at which the maximum discharge ($TQ_{max/cm}$) occurred and discharge sums for the Oude IJssel and Rhedebrugge during events 1 and 2.

Event 1	Lag time [h]	TQ_{max} [date]	Q_{sum}
Oude IJssel	14	23-02-2020 22:00	2.35
Rhedebrugge	11	23-02-2020 19:00	4.24
Event 2	Lag time [h]	TQ_{cm} [date]	Q_{sum}
Oude IJssel	13	06-06-2020 00:00	0.175
Rhedebrugge	11	05-06-2020 22:00	0.501

3.6 Validation methods discharge

The hydrological validation of quality-controlled PWS consisted of investigating its performance as rainfall input data in WALRUS hydrological model relative to that of the gauge-adjusted radar. Just as for the validation of rainfall data, the catchment-averaged gauge-adjusted radar precipitation time series ($\bar{R}_{ref}$) served as reference input data for hydrological modelling. Along with the catchment-averaged QC-PWS precipitation time series ($\bar{R}_{QC-PWS}$), the catchment-averaged raw PWS ($\bar{R}_{PWS}$) and the unadjusted radar ($\bar{R}_{rad}$) precipitation time series were validated against $\bar{R}_{ref}$ as well. WALRUS ran with hourly rainfall accumulations aggregated from the catchment-averaged 5 min time series of $\bar{R}_{ref}$, $\bar{R}_{QC-PWS}$, $\bar{R}_{PWS}$ and $\bar{R}_{rad}$ using the same potential evapotranspiration as forcing and identical initial conditions. The difference between using hourly instead of a 5 min resolution is negligible given the response time of the catchment (Table 4.1). Besides, none of the hydrological processes in the model depend on rainfall intensity directly (Brauer et al., 2016).

3.6.1 Study period

The study period for discharge validation is between 01-10-2019 and 09-09-2020. The simulated groundwater depth at the start of the validation period, retrieved through calibration, was taken as initial condition for d_G. The start of the validation period is at the end of the summer where groundwater level dropped below 2.35 m according to the calibrated output. A known model limitation of WALRUS is that its reservoirs could drop too far in summer when little rainfall input is given and too little reduction of evapotranspiration is taken into account (Lubben, 2020). Therefore, a groundwater level 20 cm closer to land surface, 2.15 m, was chosen as initial value of d_G in the study period forecast of the Oude IJssel catchment. For Rhedebrugge a 20 cm higher groundwater level was set at 2.20 m.

Four simulations were made over the study period with the four rainfall time series so that four hydrographs were obtained. To validate these discharge simulations, the Nash-Sutcliffe Efficiency (NSE) was taken as measure of goodness of fit. The NSE measures the ability to predict variables different from the mean observation, and gives the proportion of the initial variance accounted for by the model (Nash and Sutcliffe, 1970). During calibration, the performance of the reference run was validated against the observed discharge. This study aims to quantify the accuracy of discharge prediction forced by $\bar{R}_{PWS}$ relative to $\bar{R}_{ref}$. Where the Nash-Sutcliffe Efficiency is originally defined with the difference between observed and simulated data by Nash and Sutcliffe (1970), it is now defined with respect to the simulated reference run. Thus, NSE becomes:

$$NSE = 1 - \frac{\sum_{t=1}^{n}(Q_{ref}(t) - Q_{qpe}(t))^2}{\sum_{t=1}^{n}(Q_{ref}(t) - \bar{Q}_{ref}(t))^2} \qquad (3.11)$$

where t is the time [T] in hours, n the total number of time steps [-], Q_{ref} the simulated reference discharge and Q_{ref} the simulated discharge forced by the other three quantitative precipitation estimates [mm].

3.6.2 Events

The two selected events were simulated with WALRUS. A larger forecast period than the rainfall-runoff event was chosen to allow the model to warm up (~ 1 day for both events) and a few days after the right time boundary of peak discharge (TQ_{max}; event 1 and $TQ_{c,m}$; event 2) to allow visualization of the recession period. Two forecasts were made between 22-02-2020 till 29-02-2020 (event 1) and 04-06-2020 till 08-06-2020 (event 2). For these runs, the calibrated parameters from section 3.3.2 were used as well. Initial groundwater depths at the event start dates were taken for each QPE that were retrieved through the study period validation runs. The performance of simulations were also evaluated by calculating the NSE over the forecast period as is described in Equation 3.11.

3.7 Error propagation rainfall and discharge

We are also interested in how well the simulated discharge peak approximates the simulated discharge peak that was forced by the reference rainfall product, which is assumed to describe true rainfall fields the best. This is done by evaluating how errors in rainfall measurements propagated through the hydrological system. The Relative Rainfall volume Error (RRE) and Relative Discharge volume Error (RDE) were chosen as error metric. The RRE was calculated between the start of the precipitation event and the time of the defined discharge peak and defines the percentage difference in hourly accumulated rainfall sums between R_{ref} and the other QPE:

$$RRE = 100 \cdot \frac{\sum_{t=1}^{n}(\bar{R}_{qpe}(t) - \bar{R}_{ref}(t))}{\sum_{t=1}^{n}\bar{R}_{ref}(t)} \qquad (3.12)$$

The RDE was calculated over the defined lag time period and defines the percentage difference in the discharged volume of water between Q_{ref} and the discharge simulations of the other QPE:

$$RDE = 100 \cdot \frac{\sum_{t=1}^{n}(Q_{qpe}(t) - Q_{ref}(t))}{\sum_{t=1}^{n}Q_{ref}(t)} \qquad (3.13)$$

Note that the time frame over which RRE and RDE were calculated is not unique for a single event due to the recognition of lag time between a precipitation event that is always prior to the discharge event. For this reason, the method that evaluates rainfall error propagation in the simulated discharge is only applied to the selected events.

4 | Results

This chapter reviews the results of the retained data available after quality control (Section 4.1.1), validation and comparison of rainfall products (Section 4.1.2 and 4.1.3), the effect of PWS network density on the retained quality-controlled PWS dataset (Section 4.2), the validation and comparison of discharge simulations during the study period and selected events (Section 4.3) and analysis of how errors in rainfall measurements propagated in the predicted discharge (Section 4.4).

4.1 Validation and comparison of rainfall products

4.1.1 Data availability of quality-controlled PWS

A 24-hour moving average of the number of available measurements over the study period is calculated for the raw PWS and QC-PWS data that were placed in the study area (Figure 4.1). In other words, the data availability refers to how much data were not "Not a Value ('NA')" in the PWS dataset. After application of the QC filter, the number of observations is reduced by filtering measurements that were flagged with 1 (QC-PWS flex) or flagged with 1 and -1 (QC-PWS strict). The attribution of flags was done for each time interval per station meaning that the available data per station is distributed over each time interval.

The number of available measurements is logically the highest for the raw PWS data, followed by the QC-PWS flex and QC-PWS strict. Important note is that the number of PWSs in the study area increased by ∼20% from December 2019 till September 2020. That includes the stations placed in the boundary of 10 km around the Oude IJssel catchment (Section 3.1.2). From the start in October 2019 till about March 2020, the raw PWS data obtained ∼ 250 measurements per time step. The flex filtered observations move in a range of ∼220-240 stations were ∼0-10 observations less per time step were recorded by the strict filter with many location deviations per time step. Note that Figure 4.1 is a daily moving average of the 5 min time intervals that showed much more noise in the available data. This proves that each time step is uniquely evaluated in space on the presence of $nonsensical$ data. From March till the end of the study period in September, the number of raw observations gradually increased to ∼300 measurements over time. The flex and strict QC-PWS start to observe less from the spring season with a minimum around the start of June where a deviation of ∼80 stations between the raw PWS and QC-PWS strict can be observed. After a strong inclination of both QC-PWS flex and strict growing towards a deviation of ∼10 observations again with raw PWS, a steep decline started around 08-07-2020. The decay lasted till 01-09-2020 where only half of the observations, ∼150 measurements, were left in QC-PWS left and strict.

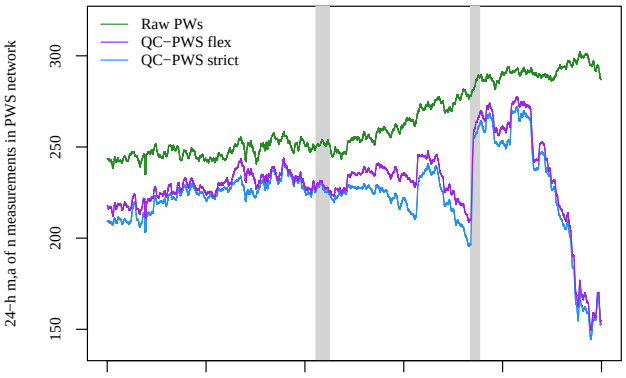


Figure 4.1: PWS data availibility over time for the raw (green) and strict quality-controlled (blue) PWSs in the Oude IJssel catchment including the time frame of two selected precipitation events (grey), Section 3.6.2

4.1.2 Validation quality-controlled PWS

The quality control filter was applied to the PWS dataset over the complete study period. Cumulative sums of the individual PWS time series R_{PWS} are plotted against the catchment-averaged gauge-adjusted radar $\bar{R}_{ref}$ (Figure 4.2). In the raw double mass curve, faulty zeroes (FZ) measurements are visualized as horizontal line segments (red) and are flagged successfully by the FZ filter. The vertical line segments (orange) indicate measured rainfall by a personal weather station during the time intervals averaged reference product did not register rainfall which were successfully flagged by the high influx filter (HI). At last, the station outlier (SO) filter successfully flagged the rainfall measurements that deviated from the one-to-one line with the reference and are visualized as fluctuating lines (green). Stations with a bias deviate from the grey one-to-one line with the reference data. Though not all horizontal, vertical and fluctuating line segments are flagged by the QC filter, less bias is visible after the application of the QC filter in the right panel of Figure 4.2. This indicates that the deviating segments were corrected by the dynamical bias correction (BC) in the quality control filter.

PWS data can be filtered in flex and strict manner

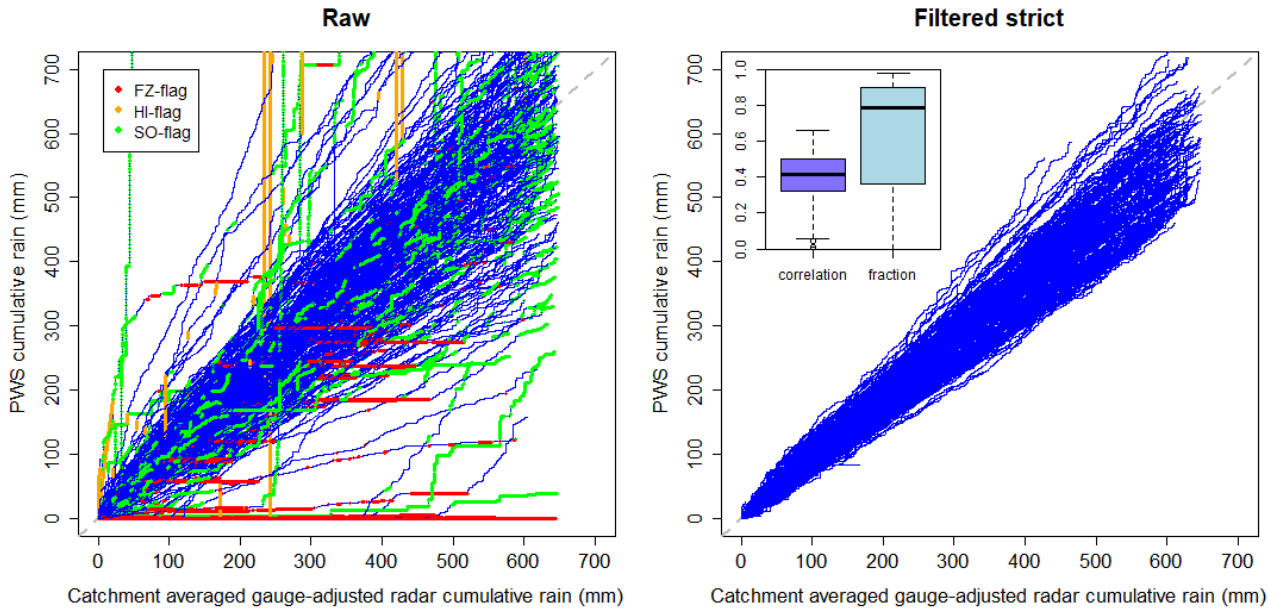


Figure 4.2: Double mass curves of $>$ 300 personal weather stations situated in the study area of the Oude IJssel catchment for the raw dataset flagged by FZ, HI, and SO filters and are shown as horizontal line segments, vertical line segments, and fluctuating lines deviating from the diagonal, respectively. After removal of the flagged time interval, the remaining segments are visualized under Filtered Strict in the right panel.

where less data is retained by applying the strict filter, because only measurements attributed with 0 remained in the strict filtered dataset (Section 4.1.1). The validation metrics of and remaining fraction of the original observations of the 5 min PWS time series, before and after quality control were calculated of the individual filters (in strict manner), and all combined filters applied (in both flex and strict manner) for the PWS network in the Oude IJssel catchment including a 10 km boundary. (Table 4.1). The flex filtered dataset retained 87.49% and the strict filtered dataset 85.11% of all intervals during the study period (Table 4.1). The small percentage difference indicates that only 2.38% of all available rainfall measurements, i.e. raw data that were not 'NA', got attributed with a flag expressing that the interval did not have sufficient information available to determine an error. The 85.11% of strict-filtered and 87.49% of flex-filtered intervals without any error flag reveal great improvement in bias, coefficient of variation (CV) and Pearson correlation (r) as compared to the metrics of the raw 5 min PWS data. The bias improved from 0.104 to -0.092 (flex) and -0.093 (strict), the CV from 130.7 to 7.095 (flex) and 7.003 (strict) and r from 0.025 to 0.411 (flex) and 0.415 (strict). Only the bias improved slightly more for the flex filtered dataset than for the strict filtered dataset. The individual filters, that were applied in strict manner, yielded an accuracy improvement too except for the faulty zeroes (FZ) filter. The largest impact was made by filtering high influx (HI) intervals given the small fraction of excluded HI intervals (Table 4.1).

Given the low percentage difference between flex and strict filter options, and to guarantee that no erroneous data is remained in the quality-controlled data, it was decided to only include the strict filtered PWS dataset in validating the catchment-averaged rainfall and discharge simulations. From now, strict quality-controlled PWS rainfall time series will be referred to as quality-controlled PWS rainfall time series (R_{QC-PWS}).

4.1.3 Validation and comparison of catchment-averaged rainfall

The 5 min PWS time series before (Raw PWS) and after quality-control (QC-PWS) have been averaged in space through the Thiessen method and yielded catchment-averaged time series $\bar{R}_{PWS}$ and $\bar{R}_{QC-PWS}$. An average rainfall sum of the pixels of the unadjusted ($\bar{R}_{rad}$) and gauge-adjusted radar 5 min time series ($\bar{R}_{ref}$) were calculated for both the main catchment the Oude IJssel and sub-catchment Rhedebrugge. Subsequently, the validation metrics of $\bar{R}_{rad}$, $\bar{R}_{PWS}$ and $\bar{R}_{QC-PWS}$ were calculated over the study period Oct. 2019 – Sept 2020 over the intervals where both PWSs and reference radar and unadjusted radar and reference radar contained measurements (Table 4.2). When reviewing both the 5 min

Table 4.1: Validation metrics and remaining fraction of original observations of the 5 min PWS time series, before (Raw) and after quality control of the individual filters (in strict manner), and all combined filters applied (in both flex and strict manner) for the PWS network in the Oude IJssel catchment including a 10 km boundary.

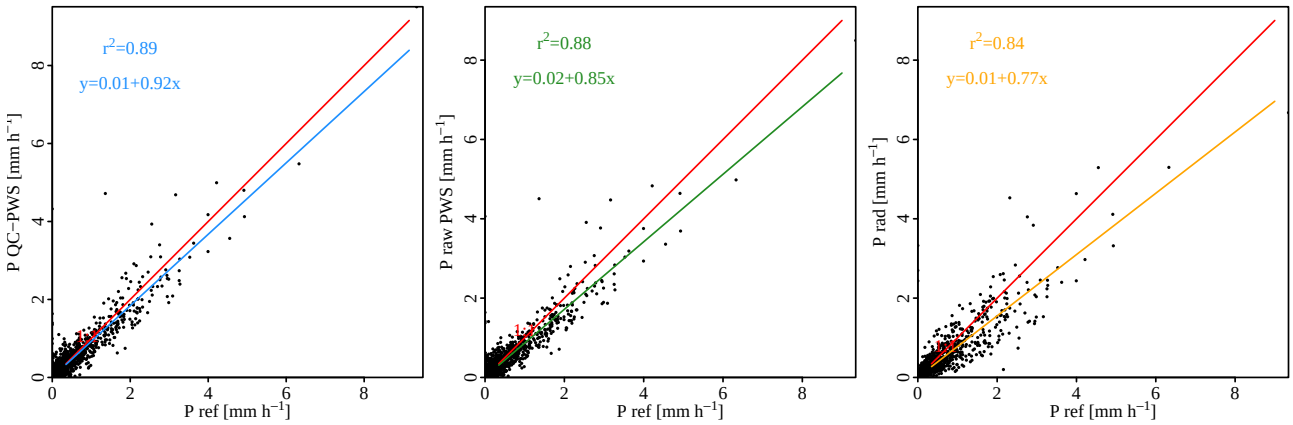
Time Interval	Dataset	Filter type	Bias	CV	r	Remaining (%)
5 min	Oct 2019 - Sept 2020	Raw	0.104	130.7	0.025	100.0
		FZ-filtered	0.170	136.1	0.026	95.74
		HI-filtered	0.033	10.17	0.307	99.57
		SO-filtered	0.031	113.6	0.030	86.46
		Bias-corrected	-0.064	83.84	0.035	100.0
		Flex	-0.092	7.095	0.411	87.49
		Strict	-0.093	7.003	0.415	85.11

Table 4.2: Validation metrics of the catchment-averaged 5 min and hourly PWS time series, before (Raw PWS) and after quality control (QC-PWS) and the unadjusted radar in the Oude IJssel catchment and Rhedebrugge sub-catchment. Only the intervals where 1) both PWSs and the reference and 2) unadjusted radar and the reference contain measurements.

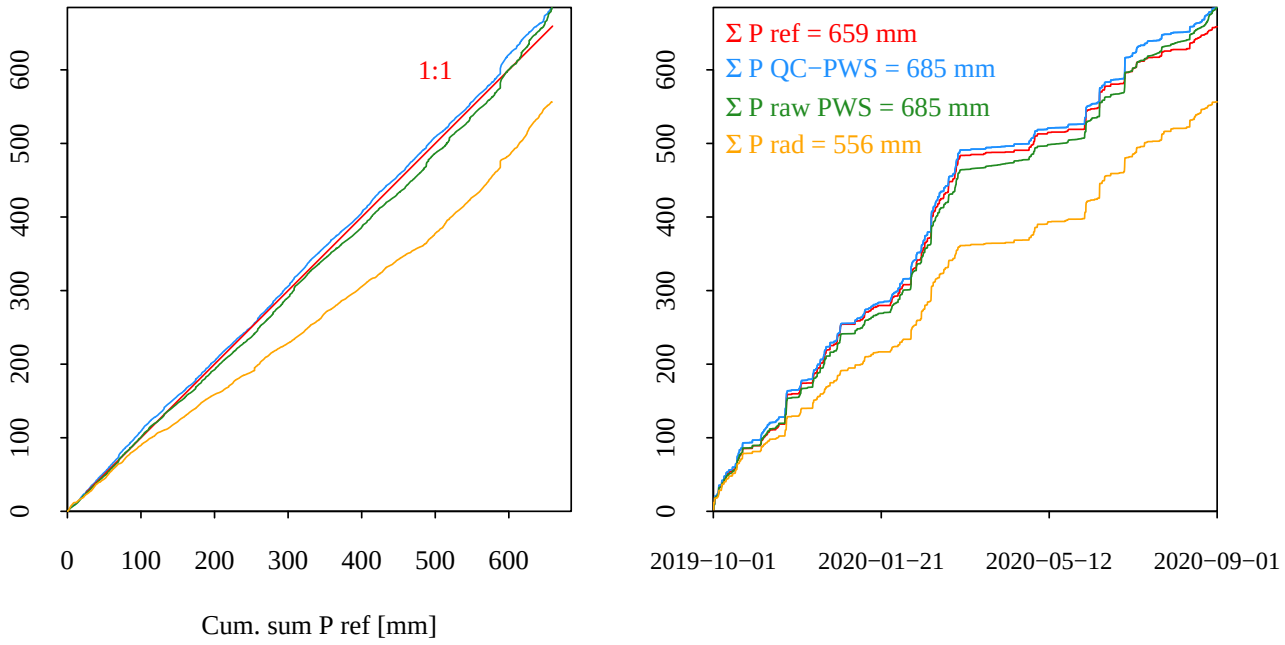
			Oude IJssel			Rhedebrugge		
Time Interval	Dataset	QPE	Bias	CV	r	Bias	CV	r
5 min	Oct 2019 - Sept 2020	Unadjusted radar	-0.164	1.875	0.914	-0.167	2.962	0.812
		Raw PWS	0.025	2.048	0.893	0.097	4.100	0.629
		QC-PWS	0.025	1.857	0.914	-0.036	3.091	0.790
1 hour	Oct 2019 - Sept 2020	Unadjusted radar	-0.155	1.717	0.916	-0.156	2.270	0.863
		Raw PWS	0.040	1.474	0.937	0.112	2.448	0.844
		QC-PWS	0.040	1.421	0.942	-0.019	1.952	0.905

and hourly accumulated time series, one can see there is no difference in bias between the catchment-averaged raw and quality-controlled PWS where they both slightly overestimated the gauge-adjusted radar. Though, the quality control improved the CV and r of the PWS 5 min (from 2.048 to 1.857) and hourly (from 1.474 to 1.421) accumulations on a catchment-scale. The unadjusted radar correlated well with the gauge-adjusted radar too on both temporal resolutions (r of 0.914 and 0.916), but underestimated the reference with greater magnitude on both a 5 min and hourly aggregations with a bias of -0.164 and -0.155. Averaged rainfall time series of the sub-catchment show similar results, only the differences between $\bar{R}_{QC-PWS}$ and $\bar{R}_{PWS}$ were far greater where $\bar{R}_{PWS}$ varied twice as much and correlated 26% less than in the Oude IJssel.

The hourly catchment-averaged time series and double mass curves of $\bar{R}_{QC-PWS}$, $\bar{R}_{PWS}$ and $\bar{R}_{rad}$ are plotted against $\bar{R}_{ref}$ (Figure 4.3a,b) along with cumulative time series of all four quantitative precipitations estimates (Figure 4.3b). Just as the Pearson correlation, the coefficient of determination (R^2) (Figure 4.3a) is highest for $\bar{R}_{QC-PWS}$, followed by $\bar{R}_{PWS}$ and $\bar{R}_{rad}$ where all catchment-averaged rainfall products underestimate the reference product. Though, the cumulative sums of both $\bar{R}_{QC-PWS}$ and $\bar{R}_{PWS}$ over the study period are equal and 26 mm more than of $\bar{R}_{ref}$. Those sums coincide with the equal biases that were calculated for the quality-controlled and raw PWS data. The cumulative time series of $\bar{R}_{PWS}$ and especially $\bar{R}_{QC-PWS}$ caught up on the $\bar{R}_{ref}$ from $\sim$ start of July 2020 which is also visible in the double mass curve on the left (Figure 4.3b). Likewise, more deviation from the one-to-one line can be seen for the raw PWS data which yielded a higher coefficient of variation. On contrary to the overestimation of catchment-averaged PWSs, the $\bar{R}_{rad}$ registered 103 mm less rainfall than $\bar{R}_{ref}$ which is visible in both the double mass curve and the cumulative time series (Figure 4.3b). The unadjusted radar systematically reports less rainfall over the complete study period that is in correspondence with the found bias of -0.155 for hourly time series (Table 4.2).



(a) The hourly accumulated time series plotted against $\bar{R}_{ref}$.



(b) Double mass curve of hourly cumulative sums (left) plotted against $\bar{R}_{ref}$ and hourly cumulative time series plotted along with $\bar{R}_{ref}$ inluding total sums over the study period of 11 months.

Figure 4.3: Time series comparison between hourly rainfall accumulations of rainfall registered by catchment-averaged QC-PWS (blue), raw PWS(green) and the unadjusted radar (orange) w.r.t. the gauge-adjusted radar (red) over the study period.

4.2 Effect of PWS network density

4.2.1 Quality-controlled PWS

The effect of PWS network density was analysed by decreasing the size of the PWS dataset to one third of the stations that already were employed in the Oude IJssel catchment including the boundary of 10 km. The three selected samples were quality-controlled by the filter of de Vos et al. (2019) over the study period. Cumulative sums of the individual 5 min PWS time series R_{PWS} are plotted against the catchment-averaged gauge-adjusted radar $\bar{R}_{ref}$ (Figure 4.2).

In the raw double mass curve, FZ, HI, and SO errors are shown as horizontal line segments (red), vertical line segments (orange), and fluctuating lines deviating from the diagonal (green), respectively. Different stations were selected in sample 1 (Figure 4.2a) and sample 2 (Figure 4.2b) given the unique cumulative patterns of the raw time series. Both samples show significantly less deviating from the one-to-one line after application of the quality control, in strict manner.

The strict filtered dataset of sample 1 retained 72.01% while only 65.60% of the raw data in sample 2 were remained after quality control over the study period (Table 4.3). In contrast, 90.19% and 87.59 % remained in the flex filtered datasets of samples 1 and 2 respectively. In Appendix A.3, one can see that more cumulative time series of the stations significantly deviated from the perfect fit where in general, PWSs in sample 1 underestimated and PWSs in sample 2 overestimated the gauge-adjusted radar. Those lines represent the intervals that did not contain sufficient information to determine error. More stations in sample 2 overestimated the reference than sample 1 underestimated the reference. However, reducing the dataset of sample 2 to strict filtered data still yielded a bias of only -0.068 over 65.60% of the raw data relative to sample 1 that retrieved a bias of -0.136 over 72.01% of the raw data which is exactly twice as negative. Most SO flagged intervals were excluded and the least for HI flagged intervals. Excluding High influx errors made the largest impact given the smallest fraction of excluded intervals and most improvement in coefficient of variation (CV) and the Pearson correlation r relative to the FZ and SO flagged intervals.

4.2.2 Data availability in space and time

As mentioned before, the difference in flex and strict filtered data is attributed to the number of observations measured by PWSs in the network that did not contain sufficient information to determine an error or not. Given that two samples were randomly selected in space out of all stations within the study area, differences between samples 1 and 2 were found in meeting the parameter requirement of n_{stat} in Section 3.4. 6.8% and 9.4% of the PWSs did not have at least 5 neighbouring stations within a range of 10 km around the station. The percentage of available data measured by the thinned PWS networks over the study period is given per station for sample 1 (Figure 4.3a) and 2 (Figure 4.3b. Each station yielded a percentage of available data for the raw PWS measurements (green), flex filtered (purple) and strict filtered (blue) sample data. The number of neighbouring stations within range d (10 km) were found in Section 3.4 for samples 1 and 2 (Figure 3.1). The percentages of available data per station are plotted against the neighbour groups which are sorted on size from smallest to largest in Figure 3.1. From these bar plots it can be seen that more stations contained low percentages of strict filtered data ($<$ 60 %) from 1 till 7 neighbouring stations in sample 2 compared to sample 1. This larger reduction of measurement intervals means that more of those intervals contained data which did not have sufficient information to determine error.

It should be noted that differences in data availability in space between samples after quality control strongly depend on the data available of the raw PWS measurements that were selected during sampling. Therfore, it was investigated how the share in available data before and after quality control differed for the two samples. This is done in strict manner where the difference is expressed as the relative error between the percentages of available data before and after quality control over time. A comparison of the relative error found for the total PWS network is calculated accordingly and presented in Figure 4.1. The time series show that the two samples systematically retrieved a larger error, i.e. filter more data over time compared to the total PWS dataset. The sudden decline in available data from July 2020 was also observed in the two samples. Most of the time, sample 2 filtered more data per time step than sample 1 did which corresponds to the observation that sample 2 filtered more data in spatial context; less measurements remained in the strict filtered dataset of stations that contained less than 7 neighbours over a radial distance of range d.

Table 4.3: Validation Metrics and remaining fraction of original observations of 5 min PWS time series, before (Raw) and after quality control of the individual filters (in strict manner), and all combined filters applied (in both flex and strict manner) for the PWS network in the Oude IJssel catchment including a 10 km boundary.

Time Interval	Dataset	Filter type	Sample 1 Bias	CV	R	Sample 2 Bias	CV	R	Sample 1 Remaining [%]	Sample 2 Remaining [%]
5 min	Oct 2019 - Sept 2020	Raw	0.113	178.569	0.018	0.206	180.588	0.017	100	100
		FZ-filtered	0.203	203.664	0.018	0.308	188.189	0.018	80.869	76.909
		HI-filtered	0.039	10.786	0.303	0.109	12.769	0.242	83.736	80.167
		SO-filtered	0.088	199.753	0.017	0.053	183.585	0.018	75.675	68.728
		Bias-corrected	-0.076	84.72	0.033	0.016	101.483	0.029	100	100
		Flex	-0.142	6.662	0.419	-0.031	86.192	0.036	90.185	87.592
		Strict	-0.136	6.618	0.435	-0.068	7.249	0.416	72.005	65.596

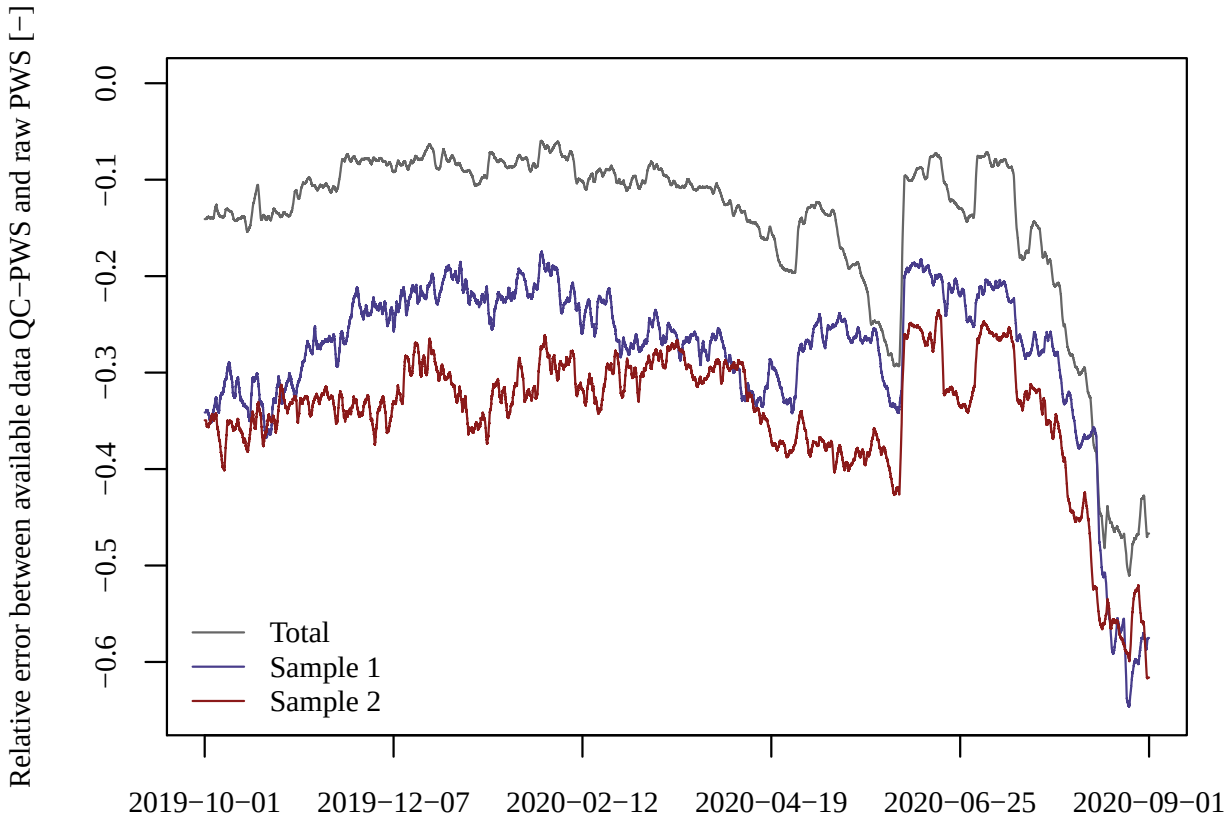
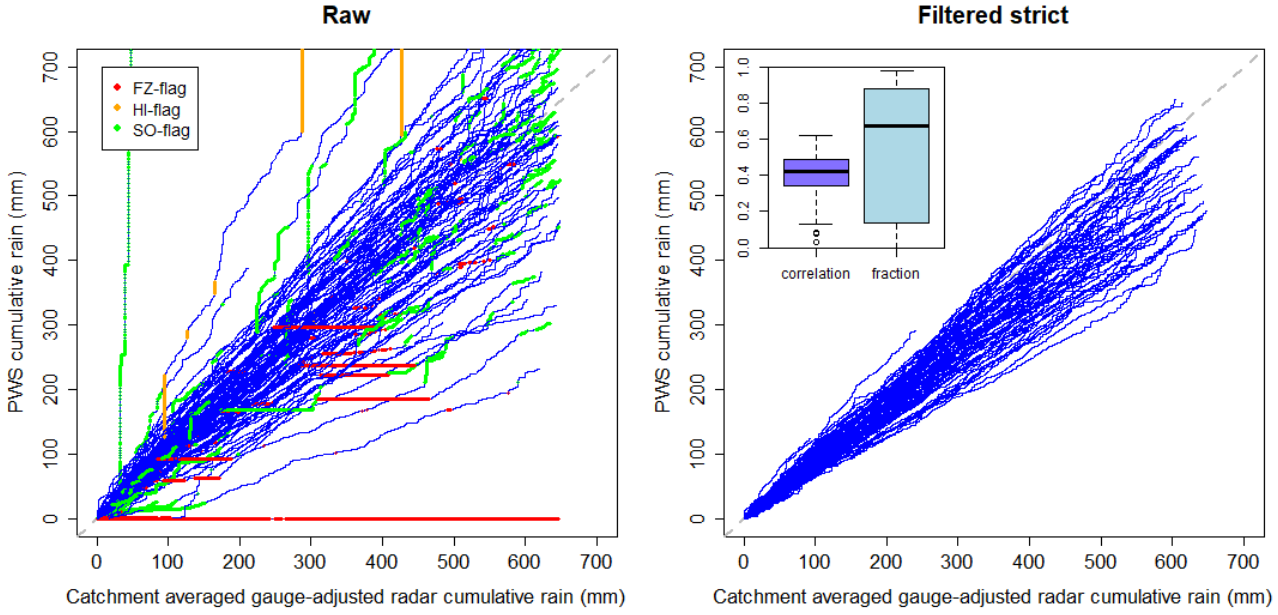
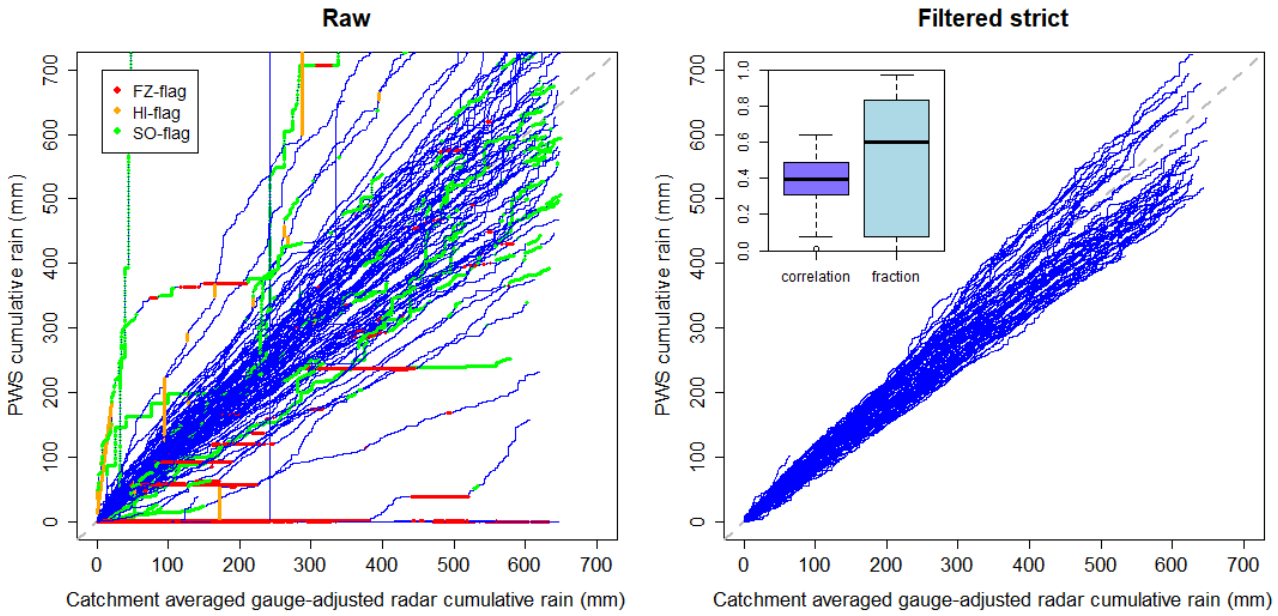


Figure 4.1: Relative error between Strict filtered and raw PWS data availability for the total PWS dataset and samples 1 and 2 over the study period.

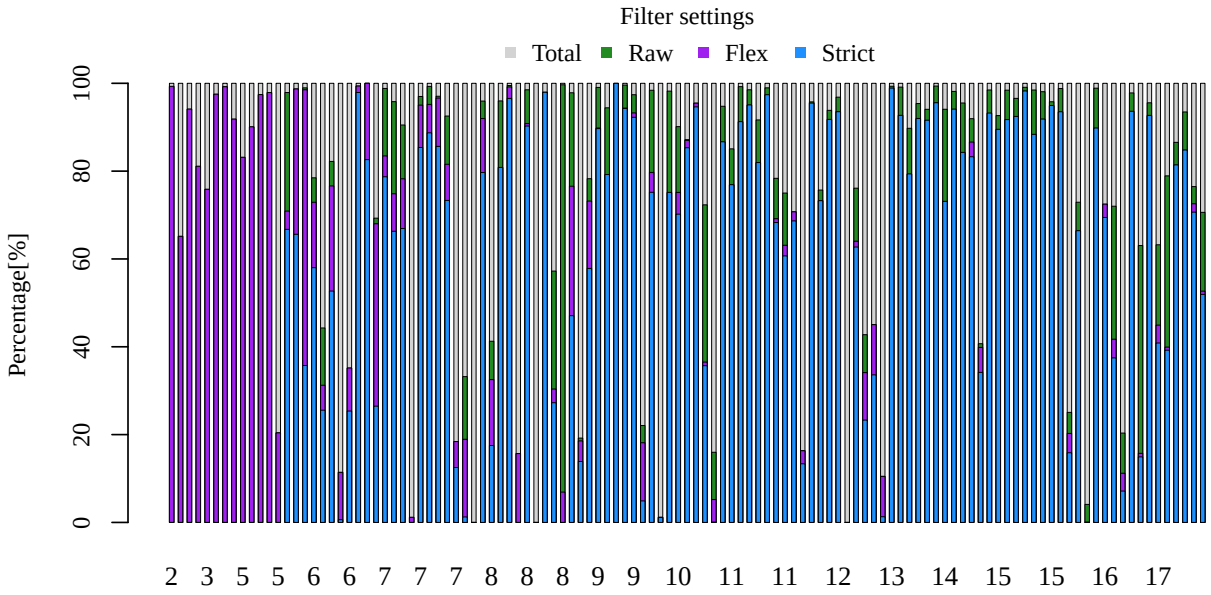


(a) Sample 1

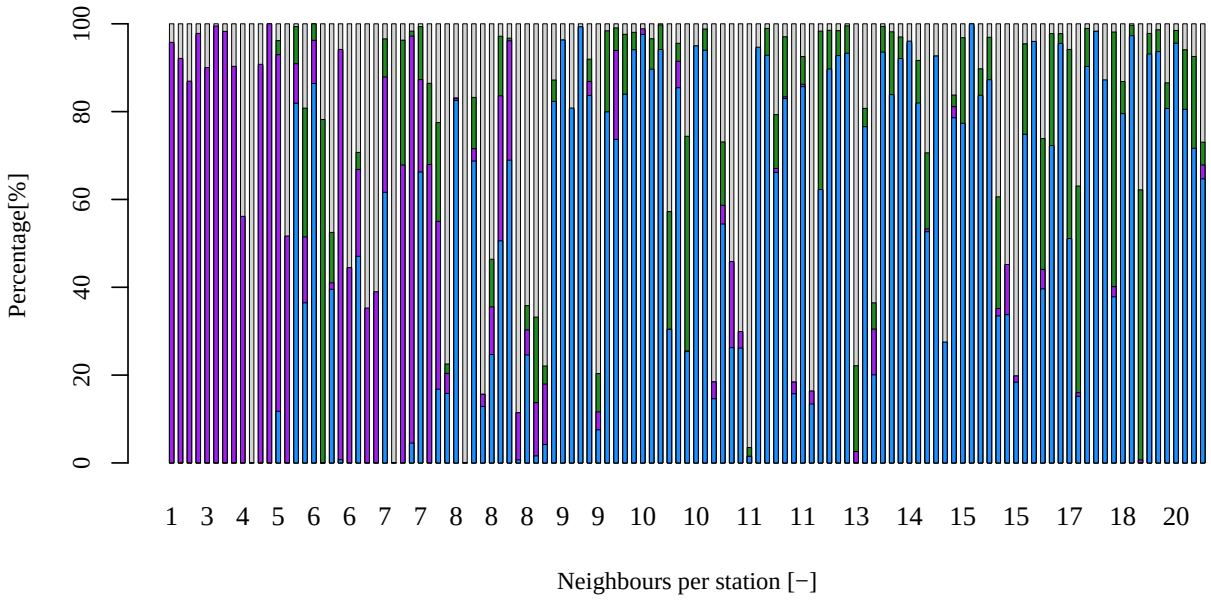


(b) Sample 2

Figure 4.2: Double mass curves of $>$ 100 personal weather stations situated in the study area of the Oude IJssel catchment for the raw dataset flagged by FZ, HI, and SO filters and are shown as horizontal line segments, vertical line segments, and fluctuating lines deviating from the diagonal, respectively. After removal of the flagged time interval, the remaining segments are visualized under filtered strict in the right panel.



(a) Sample 1



(b) Sample 2

Figure 4.3: Bar plots of percentage available data measured by the thinned PWS networks (samples 1 and 2) over the study period calculated per station with the total number of time intervals (grey), the raw available data (green), the flex filtered (purple) and strict filtered (blue) available data. Percentage bars overlap with the total dataset in the back and the strict filtered dataset in the front. These percentages are plotted over the number of neighbouring stations within range d sorted from smallest to largest neighbour group.

4.3 Validation and comparison of discharge simulations

The accuracy of discharge predictions forced by catchment-averaged QC-PWS is determined by calculating the Nash-Sutcliffe efficiency relative to the simulation with catchment-averaged gauge-adjusted radar used as reference forcing data. Accordingly, a comparison with the raw PWS and unadjusted radar as input was made. The validation and comparison were executed over the study period of 11 months (Section 4.3.1) and the two selected events during summer and winter (Section 4.3.2). Furthermore, the simulated events were examined on whether errors in rainfall measurements propagated through the hydrological system (Section 4.3.3).

4.3.1 Study period

The catchment-averaged hourly precipitation sums of the gauge-adjusted radar ($\bar{R}_{ref}$), strict QC-PWS ($\bar{R}_{QC-PWS}$), raw PWS ($\bar{R}_{PWS}$) and the unadjusted radar ($\bar{R}_{rad}$) were compared for the Oude IJssel and Rhedebrugge. In Section 4.1.3 it was found that the hourly accumulations of both $\bar{R}_{PWS}$ and $\bar{R}_{QC-PWS}$ slightly overestimated $\bar{R}_{ref}$ (mean bias = 0.04). $\bar{R}_{rad}$ underestimated the reference product with a larger magnitude (mean bias = -0.155) on the catchment-scale of the Oude IJssel. In time, this mean bias was reflected in the hourly accumulations of the three quantitative precipitation estimates where the highest rainfall peaks were mostly dominated by the gauge-adjusted radar and QC-PWS (Figure A.1). The mean bias of the average hourly accumulations deviated slightly for Rhedebrugge (Table 4.2) which can also be seen in the average hourly accumulated rainfall values over time (Figure A.1). During higher rainfall peak events, $\bar{R}_{ref}$ dominated even more in the sub catchment. With the bare eye, hardly any differences can be observed between the residuals of $\bar{R}_{QC-PWS}$, $\bar{R}_{PWS}$ and $\bar{R}_{rad}$. For further information on validation of the rainfall estimates one is referred to Sections 4.1 and 4.2.

The observed discharge at the outlet in Doesburg is plotted in time along with the base run Q_{ref}, Q_{QC-PWS}, Q_{PWS} and Q_{rad} for the Oude IJssel (Figure 4.1a) and Rhedebrugge (Figure 4.1b). The large discharge peaks of the sub-catchment ($>$ 0.10 mm/h) are much sharper and approximately twice as high as of the main catchment. When looking at the modelled discharge, the simulation of the three QPEs show large deviations in their residuals w.r.t. the base run (Figure 4.1 panels 2-4). In agreement with the bias of the catchment-averaged QPEs, the discharge simulation forced by the unadjusted radar product underestimated the base run the most. This is especially true for the autumn, summer and early spring months, followed by Q_{PWS} and Q_{QC-PWS} at last. For Rhedebrugge, this observation also holds, only anomalies were even sharper and more negative during the large and sharp discharge peaks, as was mentioned before. However, all simulations did perform well as they retrieved a NSE of 0.70 or higher taken over the complete study period where Q_{QC-PWS} is most similar to the base run with an NSE value of nearly 1 on the main catchment scale.

4.3.2 Events

Discharge simulations of Q_{ref}, Q_{QC-PWS}, Q_{PWS} and Q_{rad} for the Oude IJssel and Rhedebrugge catchments were made for the winter event (Figure 4.2a) and the summer event (Figure 4.2b). The precipitation and discharge sums during events 1 and 2 have been calculated for both catchments (Table 4.1). During the winter event, the main catchment outlet registered a discharge sum of 2.21 mm and the weir in Rhedebrugge a sum of 3.94 mm. The peak registered at the outlet is topped off at a maximum of 0.21 mm h^{-1} (Figure 4.2b, bottom left panel) while the peak in the sub-catchment is sharper with a maximum of 0.43 mm h^{-1}. During the summer event, a discharge sum of 0.182 and 0.482 were measured by the Oude IJssel and Rhedebrugge respectively. While multiple peaks were registered by the outlet with a maximum of 0.023 mm h^{-1}, a single peak with a maximum of 0.052 mm h^{-1} was seen in Rhedebrugge.

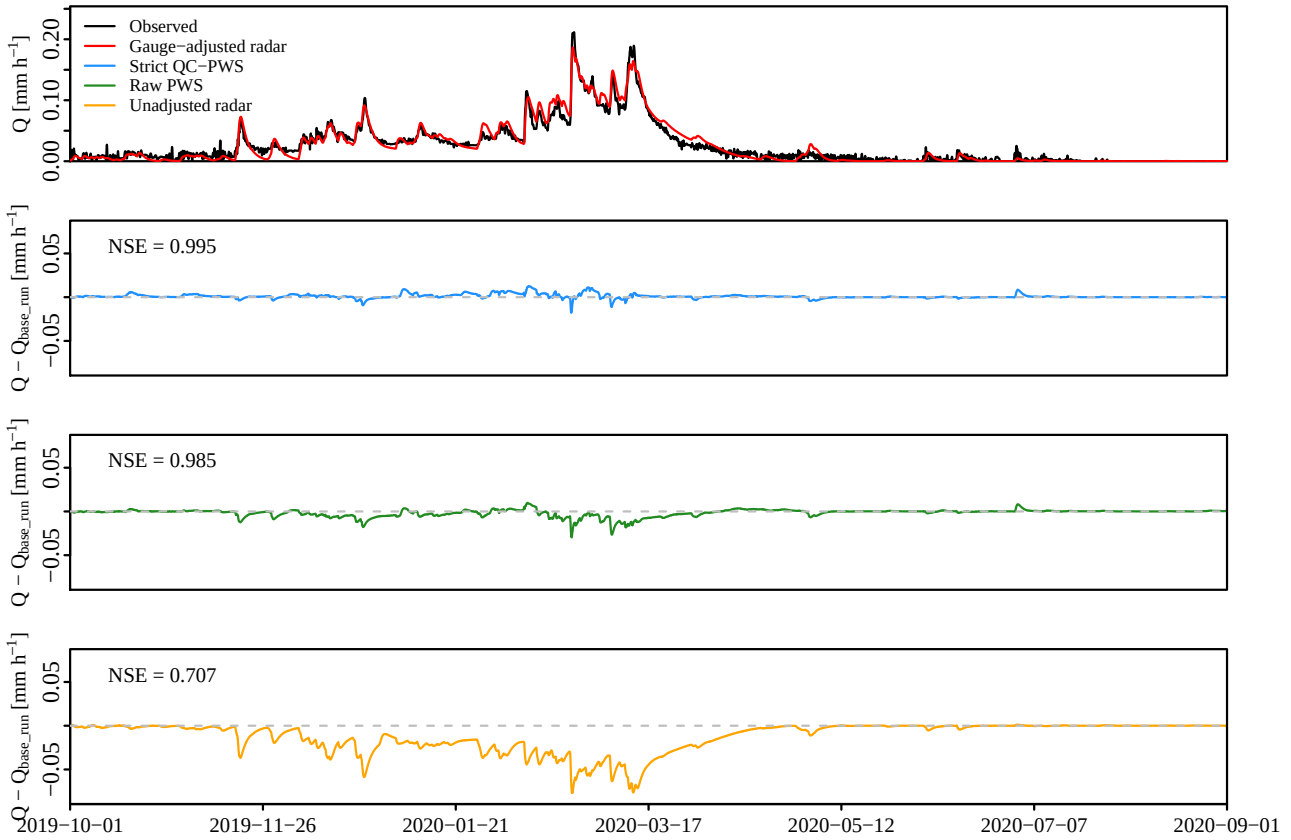
In Section 4.1.3, it was found that cumulative summations of $\bar{R}_{QC-PWS}$ were slightly higher than $\bar{R}_{ref}$ during the major part of the study period where it was followed by $\bar{R}_{PWS}$ and $\bar{R}_{rad}$. During both events and in both catchments the averaged rainfall sums of the gauge-adjusted radar dominates followed by the QC-PWS, raw PWS and unadjusted radar product. The same order can be noticed for the simulated maximum discharge peak (Figure 4.2) and sums (Table 4.1. Though, two exceptions can be noted. During the last day of the winter event in the Oude IJsel, QC-PWS

caught up with the gauge-adjusted radar which is also reflected in the recession period of the discharge event where higher values were simulated for Q_{QC-PWS} than for Q_{ref}. The other remarkable observation is that the $\bar{R}_{rad}$ underestimated $\bar{R}_{ref}$ less than $\bar{R}_{PWS}$ in the Oude IJssel during the summer event and also $\bar{R}_{QC-PWS}$ in the sub-catchment. $\bar{R}_{rad}$ of Rhedebrugge measured about 22.5 mm of rainfall between the start of the simulation and the moment a maximum peak discharge of 0.41 mm h^{-1} got reached relative to $\bar{R}_{QC-PWS}$ that measured 21 mm until a peak of 0.039 mm h^{-1} was predicted.

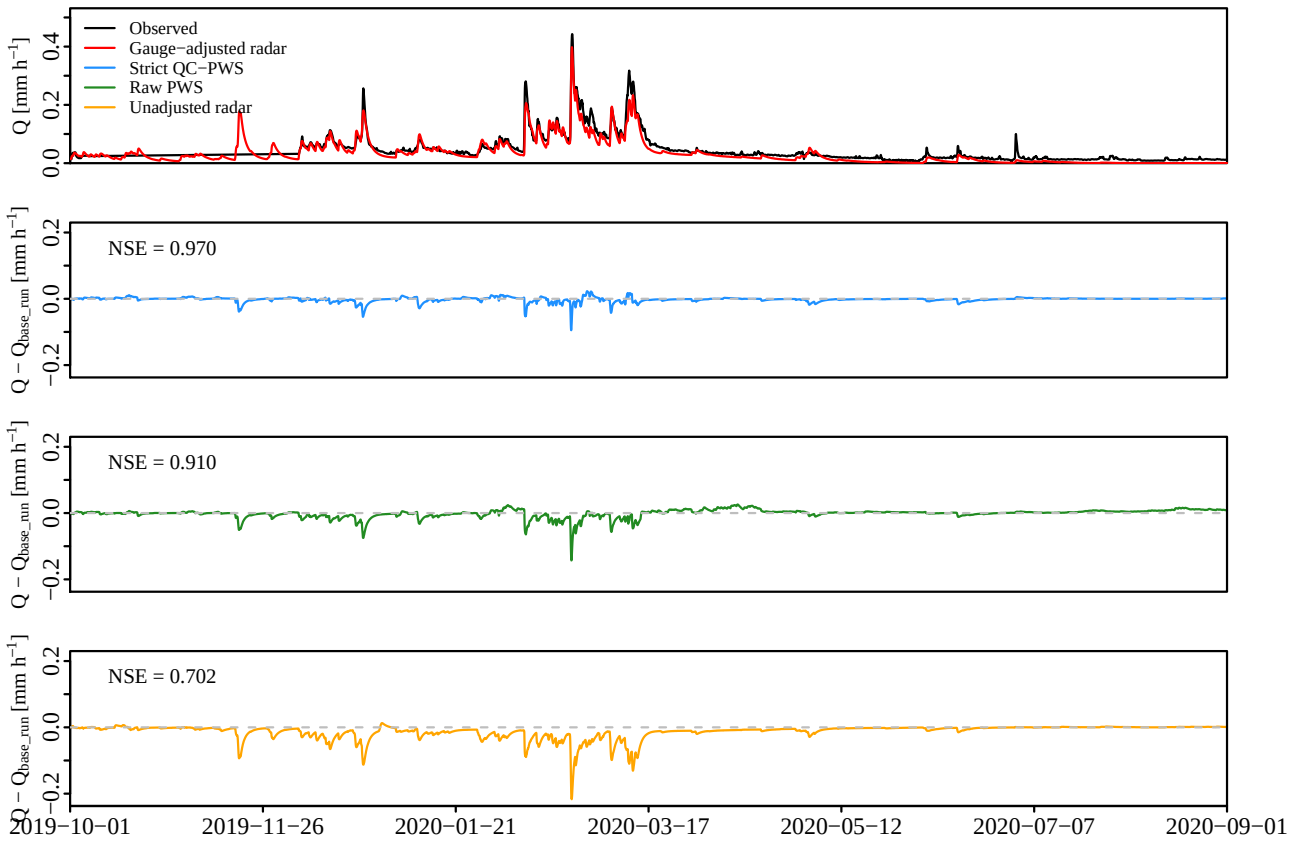
This indicates that the magnitude of the cumulative precipitation sums directly influences the magnitude of the simulated discharge forced by the four quantitative precipitation estimates. Similar to the discharge volume results, best performance was retrieved by Q_{QC-PWS} with a NSE value of 0.97 and 0.93 for the Oude IJssel and Rhedebrugge during event 1 and a NSE of 0.97 during event 2 in the main catchment while Q_{rad} did best with a NSE of 0.88 in the sub-catchment.

Table 4.1: Sums of precipitation and discharge during events 1 and 2 in the Oude IJssel and Rhedebrugge (refer to Section 3.5.3 for definitions of event duration).

Event 1	Oude IJssel		Rhedebrugge	
	P_{sum} [mm]	Q_{sum} [mm]	P_{sum} [mm]	Q_{sum} [mm]
Observed discharge	-	2.21	-	3.94
Gauge-adjusted radar	28.2	2.10	28.1	3.57
Strict QC-PWS	27.2	1.87	26.4	2.82
Raw PWS	24.7	1.80	22.8	2.46
Unadjusted radar	19.9	1.65	17.7	2.05
Event 2				
Observed discharge	-	0.182	-	0.482
Gauge-adjusted radar	15.4	0.116	16.5	0.445
Strict QC-PWS	13.9	0.098	13.5	0.340
Raw PWS	12.8	0.089	11.3	0.313
Unadjusted radar	13.2	0.094	13.6	0.370

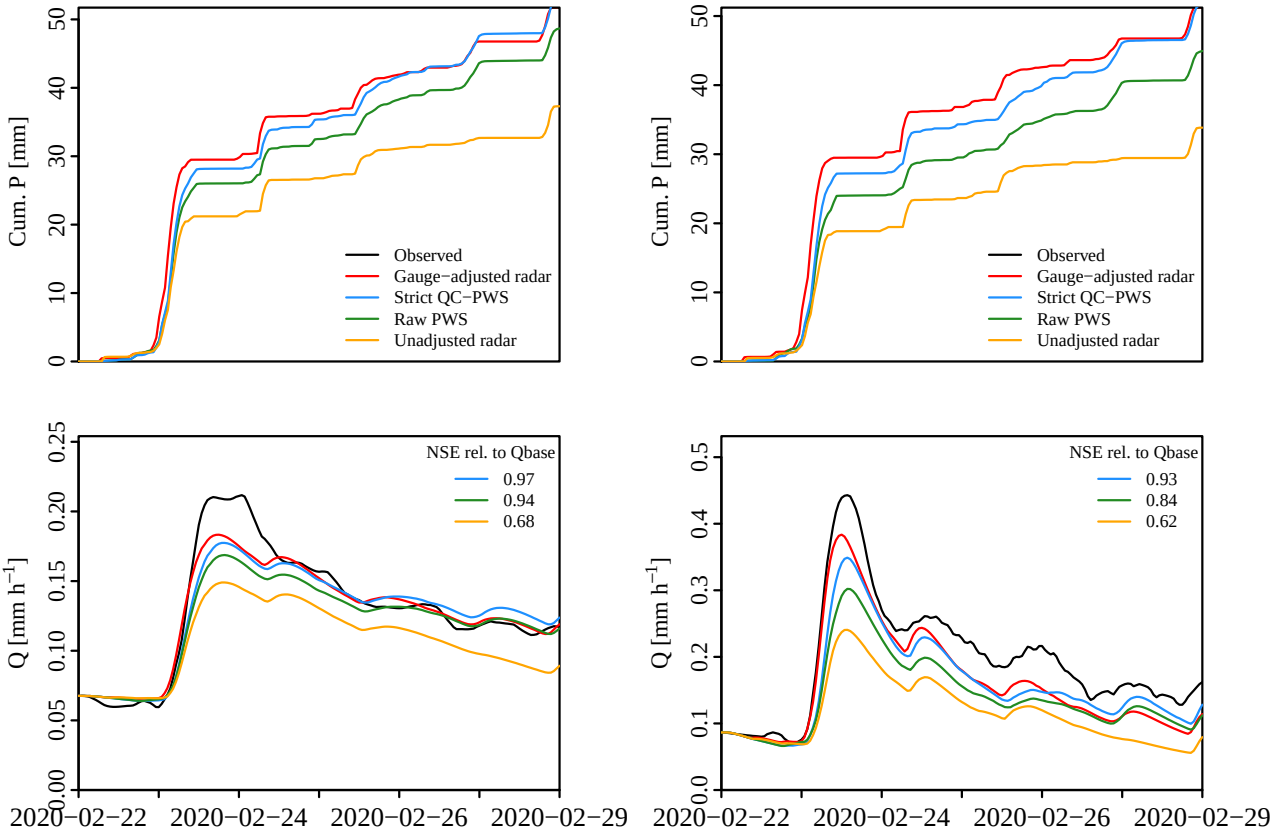


(a) the Oude IJssel

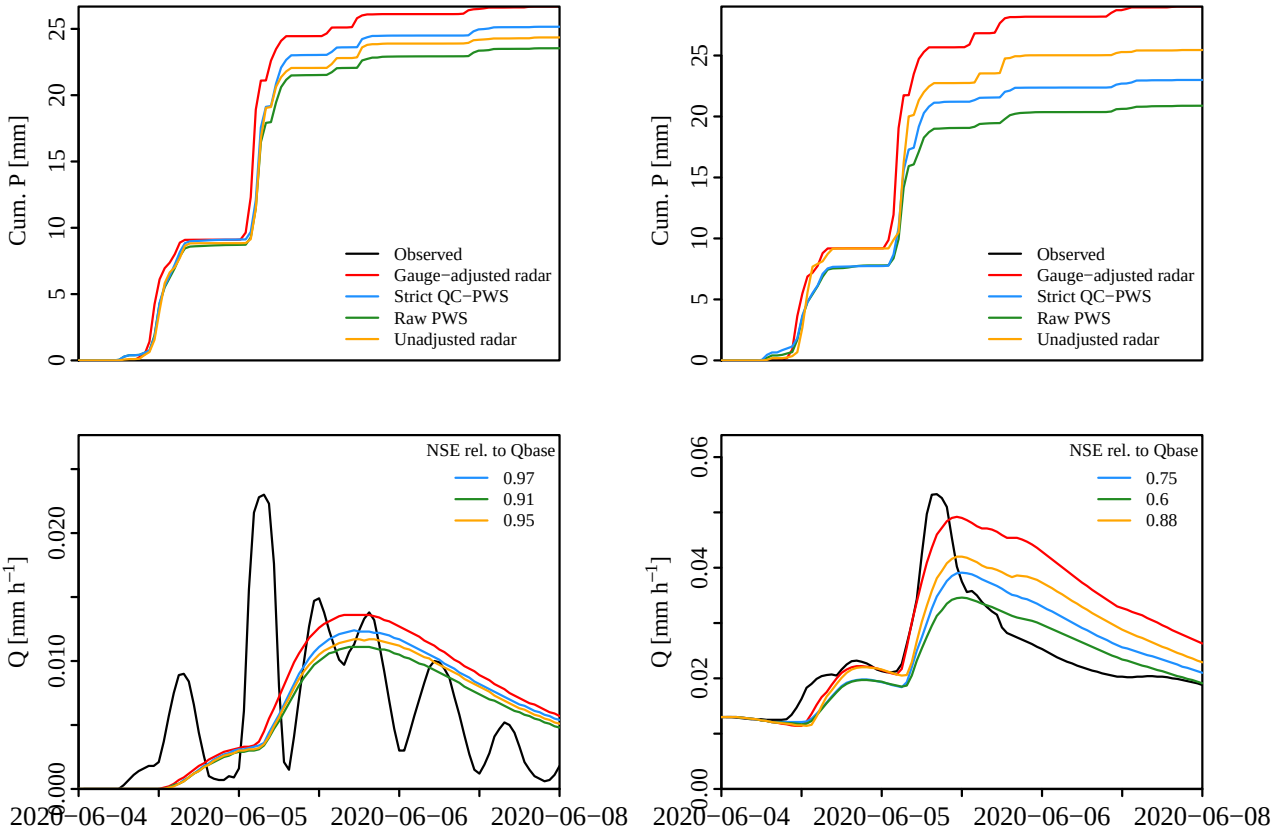


(b) Rhedebrugge

Figure 4.1: Comparing the simulated discharge of rainfall inputs QC-PWS strict, raw PWS and the unadjusted radar to the modelled discharge of the base run (with the gauge-adjusted radar data). In each of the three sub figures, the base situation is plotted on top and the deviation from the base run (i.e. the residuals) for each of the three QPE below it.



(a) Event 1: winter



(b) Event 2: summer

Figure 4.2: Hourly catchment-averaged precipitation cumulative sums (top) and simulated discharge (bottom) of the Oude IJssel (left) and Rhedebrugge (right).

4.4 Error propagation rainfall and discharge

The relative rainfall volume error and the relative peak volume error percentages were calculated by the sums of $\bar{R}_{QC-PWS}$, $\bar{R}_{PWS}$ and $\bar{R}_{rad}$ and Q_{QC-PWS}, Q_{PWS} and Q_{rad} (from Table 4.1) for the Oude IJssel and Rhedebrugge catchments during the winter and summer events visualized in Figure 4.1. The symbols indicate the name of the quantitative precipitation estimates and the colours the events and catchment names.

All catchment-averaged rainfall sums underestimated $\bar{R}_{ref}$ and all simulated discharge sums underestimated Q_{ref} during event 1 and 2. There is a positive correlation between the relative rainfall volume error (RRE) and relative peak volume error (RDE) in Figure 4.1, where Q_{rad} (triangles) showed the largest errors followed by Q_{PWS} and Q_{QC-PWS}. The winter event in the Oude IJssel showed the lowest relative errors where QC-PWS underestimated the least with a RRE value of 3.6% and a RDE of 11.1%. On the other hand, Rhedebrugge gave the largest negative percentage values with a RRE value of 37% and a RDE of 42.4% calculated for the unadjusted radar during the winter event. This simulation also gave the largest differences between the three QPE. The relative error percentages of the Oude IJssel and Rhedebrugge deviated less during the summer event with a RRE value of 9.29% and a RDE value of 15.2% for Q_{QC-PWS} while the Q_{PWS} gave the largest errors with a RRE of 31.6% and a RDE of 29.5%.

A one-to-one line is plotted in Figure 4.1. Though each calculated error percentage is represented by different events, catchments and model forcing data, it can be noted that the relative peak volume error percentage has a tendency to have a larger negative magnitude than the relative rainfall volume error percentage since 9 out of the 12 RRE percentage values are lower than the 9 corresponding RDE percentages.

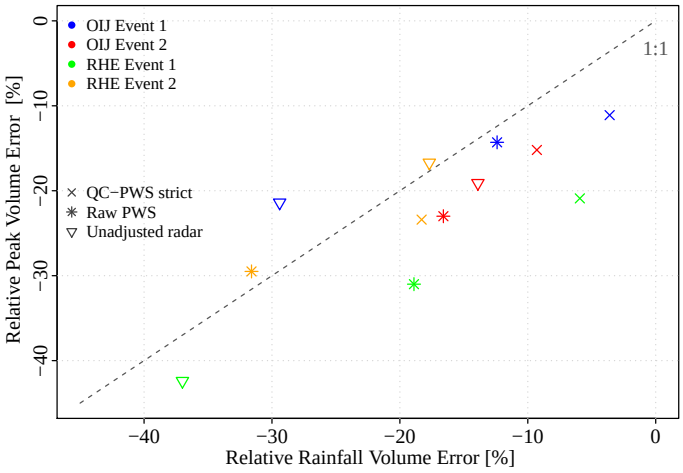


Figure 4.1: The relative rainfall volume error against the relative peak volume error during the summer event(1) and summer event (2) for the Oude IJssel (winter: blue, summer: red) and Rhedebrugge (winter: green, summer: orange). QC-PWS strict, raw PWS and the unadjusted radar are distinguished by a cross, star and triangle respectively.

5 | Discussion

5.1 Comparison and validation of rainfall products

The comparison and validation of rainfall products were done for individual 5 min PWS time series and for the catchment-averaged 5 min and hourly rainfall sums. A difference of only 2.38% in the retained available data between filtered strict and flex was found indicating that there were only few measurements that did not contain sufficient information to determine the error (Section 4.1.2). The difference between the two filter options completely rely on parameters n_{stat} and range d in the quality control filter (de Vos et al., 2019). The value of n_{stat} should be evaluated carefully for a sparse PWS network before quality control. Validation of the national PWS dataset by de Vos et al. (2019) found that fewer observations (before and after quality control) were made by PWS in regions with a lower population density. The Oude IJssel catchment is a rural region in East Netherlands with only 9 % of the land covered with urban area (Drost, 2016) which is low relative to the West Netherlands and thus less Personal Weather Stations are placed there by citizens. Yet, an average network density of $\sim$9 km^2 in the Oude IJssel catchment proved to remain sufficient data distributed in space given the small percentage difference of 2.38% between the two filter options that is comparable to the percentage difference of 1.20% found for PWS network in Amsterdam (de Vos et al., 2019). Furthermore, a bias reduction of 10.6% was yielded in absolute terms after quality control of individual PWS time series compared to 11.3% improvement for the city of Amsterdam. Therefore, the application of the PWS quality control filter to rural lowland catchment the Oude IJSsel did not underperform the application of the filter to a city with high PWS network density.

The quality control filter did not make any difference in the bias of catchment-averaged time series w.r.t. the catchment-averaged gauge-adjusted radar (Section 4.1.3). More measurement intervals were taken into account for generating the raw catchment-averaged PWS timeseries which could have had a smoothening effect on systematic deviation of raw PWS w.r.t. the reference product. Both the filtered and unfiltered data even yielded a slight positive bias w.r.t. the gauge-adjusted radar where earlier research by de Vos et al. (2017) and de Vos et al. (2019) found that PWS have a general tendency to underestimate rainfall. No evident explanation can be given but a notifiable observation in the PWS data availability over time (Section 4.1.1) could explain why the filtered catchment-averaged PWS time series measured almost 4% more rainfall during the study period. After quality control, the amount of measurements by PWSs over time suddenly decreased massively from July 2020 till September 2020 after quality control. The cumulative precipitation sums of catchment-averaged quality-controlled PWS catch up on the gauge-adjusted the moment available data declined significantly for the filtered dataset.

Compared to the unadjusted radar, which is used in real-time, significantly less bias was registered by the raw PWSs per time step and also over the study period which is reflected in the cumulative rainfall sums of the catchment-averaged time series. This suggests that unfiltered PWS data, also available in real-time, could be chosen over the unadjusted radar product for rainfall monitoring. The application of the quality control filter requires prerequisite knowledge about the PWS network layout and a reference dataset to determine the default bias correction factor (DBC). Although, no auxiliary data are needed and the DBC can be determined offline, no DBC is taken into account if no reference is available (DBC is set to 1) (de Vos et al., 2019). The quality-control filter can also be applied in real-time and is preferred over the raw PWSs given their better performance over all other metrics for both individual and catchment-averaged time series, but under two conditions: a reference dataset and sufficient data after quality control are available.

5.2 Effect of PWS network density

The quality-controlled PWS are expected to be influenced by the PWS network density in space, since the filter inhibits spatial parameters which determine whether a rainfall observation by PWS is excluded or remained in the dataset. The effect of PWS network density is analysed by randomly taking three samples of the existing PWS network which were one third of the original network size. Two samples were validated with the methodological steps of Sections 3.3.1 and 3.3.2 were more data were filtered and larger differences were found between the flex and strict filtered data compared to the existing network of PWSs. Since the difference between flex and strict filtered data only depends on the filter parameters having a spatial character, n_{stat} and range d, this difference can completely be attributed to the network layout of of samples 1 and 2. Thus, the degree of equal distribution of PWSs in space might cause the share of measurement intervals flagged with -1. Figure 3.1 shows

that some stations are less equally distributed in space for the second sample compared to the first sample, especially in the centre of the catchment.

Even though, more data were filtered in the second sample where stations were distributed less equally in space, the validation metrics proofed that sample 2 yielded similar results with a bias that was even twice as small compared to sample 1. Accordingly to de Vos et al. (2018), the CV and bias would grow and the r decline when less stations are present in the PWS network. Results did not provide the same relation between network density and performance metrics. Though it is important to note that this study only compared the original network to two thinned networks with the same sample size, while de Vos et al. (2018) selected 50 random samples of n PWSs present in the total network in Amsterdam.

5.3 Methods rainfall

Multiple methodological assumptions were made including limitations for execution during this study w.r.t. the quality control, preprocessing of catchment-averaged time series, validation of rainfall observations and the analysis of PWS network density. This research interplayed between individual, - and catchment averaged time series during validation of Personal Weather Stations in the PWs network of the Oude IJssel. Since the aim of this study was to validate and compare rainfall time series and relate it to hydrological application, it was decided to validate individual PWS 5 min time series against the catchment-averaged gauge-adjusted. These averaged reference data were assumed to be a proxy of the observed rainfall registered by radar pixels within the catchment. Furthermore, catchment-averaged PWS time series know the advantage that they do not contain 'NA' data unless there were intervals were not any station reported measurements. validating catchment-averaged PWS against catchment-averaged reference radar resulted in a larger validation dataset relative to the validation dataset of individual PWSs.

Secondly, the run time of the quality control took over 3 days to completely filter the PWS time series over the study period of 11 months. The long run time is especially caused by flagging the station outlier (SO) intervals and dynamical bias correction which were evaluated for every time step per station in the network. The long run time of the quality control filter made me decide to not filter the stations within Rhedebrugge sub-catchment. Instead, individual time series located in Rhedebrugge including the 10 km boundary were extracted from the quality-controlled PWS time series in the total network of the Oude IJssel. Therefore, results for the observed rainfall by PWS in Rhedebrugge were only validated on a catchment-scale.

The analysis of the network density only included thinning of the existing network to multiple samples with one sample size. Since each sample did have to be quality-controlled with the filter, it was decided to reduce the uncertainty in network layout by analysing the extent of three samples before quality control. This method was chosen rather than sampling over n stations present in the existing network to save much time in the preprocessing of quality-controlled PWS data.

5.4 Precipitation data

This research was limited in the availability and source of quantitative precipitation estimates data.

Firstly, only one year of PWS data was available that limited the temporal extend of this study. No seasonality analysis could have taken place while this is important for hydrological forecasting like was investigated in Brauer et al. (2016). Besides, the number of PWS in the network had grown during the study period, thus no independent judgement could be made of the PWS data availability over time and validation of quality-controlled PWS time series.

Secondly, the gauge-adjusted radar data were used as the reference rainfall product during this study. Using these data as reference data source knows two disadvantages. First disadvantage is that this study assumed that the gauge-adjusted radar product currently is the most accurate method available to describe rainfall fields in space and time (van Beekhuis et al., 2018; Overeem et al., 2009a, 2009b), while the method proposed in this study has potential to outperform the reference product. Only 7 rain gauges are placed within the Oude IJsel catchment with the nearest automatic rain gauge placed 20 km away from the catchment's centre (https://www.knmi.nl/kennis-en-datacentrum/uitleg/vrijwillige-neerslagmeters). One does not know how the network of PWSs ability to measure rainfall relates to that of the rain gauge network deployed and used by the KNMI for weather radar correction. Certain is that Personal Weather Stations are placed with a far greater spatial resolution than the traditional rain gauges are. Therefore, additional research would be needed to investigate how well rain gauges used for correction measure rainfall relative to the network of PWSs where ultimately investigation to the potential of Personal Weather Stations as correction for weather radar images is required to answer this question. Secondly and a more explicit disadvantage of the reference data used is that they were not independent. Data from both the Dutch Royal meteorological Institute

(KNMI) in the Netherlands and the Deutsche Wetterdienst (DWD) in Germany were consulted which are two different data sources retrieved by different radars and methods. Furthermore, it took some extra preprocessing steps to convert the German reference data to data with the same properties to that of the KNMI.

At last, the unadjusted radar was retrieved from the Dutch institute completely which was assumed to observe similar rainfall as the German radar east of the German borders in the catchment.

5.5 Comparison and validation of discharge

Discharge simulations were made forced by catchment-averaged time series of Personal Weather Stations before and after quality control and for the unadjusted real-time radar over the complete study period and the two selected events in the Oude IJssel and sub-catchment Rhedebrugge. The residuals of the discharge simulations w.r.t. the reference run forced by the catchment-averaged time series were the smallest for the filtered PWS followed by the input of PWSs before quality control and lastly the unadjusted weather radar. This result is what you would expect to find given that the operational real-time observes rainfall the least accurate. Brauer et al. (2016) found that the unadjusted radar also systematically underestimates the reference run, especially during peak discharges. In this study, residuals became larger during peak discharge for all input data for main and sub-catchment.

During the winter and summer event it becomes clear why one should compare different rainfall input data w.r.t. a reference run instead of the observed discharge. The second maximum peak during the winter event at the outlet in Doesburg is not what you would expect given the cumulative rainfall time series during the event. It is unclear why the hydrological model did not simulate this second peak, though it is unlikely that the underestimation is caused by systematic errors in the rainfall input data because the unadjusted radar and Personal Weather Stations are independent rainfall data sources. Modifications of the weir at the outlet or influence further upstream could cause this horizontal peak where more water was retained in the system preventing the evolution of a higher discharge peak than was observed. During the summer event though, it is evident that the observed discharge is influenced by modifications of the weir valve in Doesburg. The simulated discharge forced by personal weather stations underestimates the simulated discharge forced by the unadjusted radar during the summer event. No evident cause can be attributed to this unexpected output, however the PWs data availability before and after quality control might play a role since the peaks of both inputs understimated the real-time radar in Rhedebrugge.

Errors in rainfall measurements propagated in the predicted discharge over the two events, the two catchments and all input data. Likewise to Brauer et al. (2016), the errors in predicted discharge are systematically lower than the reference run. Furthermore, the relative predicted discharge peak volume errors showed a tendency to take a larger negative magnitude than the relative rainfall volume errors as was found in the simulations of Brauer et al. (2016). An average of -14.2% in the relative rainfall volume error percentage resulted in an average discharge peak volume error percentage of -17.3% in the Oude IJssel catchment and an average percentage of -21.6% observed rainfall error resulted in -27.3% predicted discharge error in Rhedebrugge sub catchment during the two selected events. Hence, the lower reliability of catchment-averaged rainfall time series generated for the sub catchment propagated in the hydrological system. Though, it should be noted that this research included an independent investigation of the error propagation of rainfall by introducing imposed errors in the rainfall forcing data.

5.6 Hydrological model and input data

The Wageningen Lowland Simulator (WALRUS) is a lumped rainfall-runoff model. Despite, this study focussed on precipitation data with high spatial resoluions, a lumped model was preferred over a spatially distributed model. This choice was mainly based on the significantly shorter run and calibration times for a lumped model. The run time of the PWS quality control filter was a constraining factor in the preprocessing steps of this study, thus reducing the run time of the hydrological model was highly preferred. Supporting the choice for WALRUS were that parameters were already automatically calibrated for the weir in Rhedebrugge by water board Rijn & IJssel and by Drost (2016) for the outlet in Doesburg both retrieving good results.

Furthermore, assumptions were made for the evaporation and discharge data that functioned as input for the hydrological model. Reference evaporation data from a single automatic weather station (monitored by the KNMI) was assumed to approximate the evaporation in the Oude IJssel catchment. Besides, these ET_{ref} data were not corrected by crop factors to retrieve the potential evaporation and were assumed to be a proxy of ET_{pot} in the study area. Since no seasonality analysis was inclueded in this study, the uncorrected evaporation data were assumed to approximate the potential evaporation sufficiently.

Thirdly, discharge data were measured at the outlet in Doesburg that was heavily influenced by valve modifications done by local water managers. Although a moving average of 12 hours was calculated over the hourly discharge sums, the influence of of weir effects was not integrated in the hydrological model.

5.7 Methods discharge

Methodological assumptions in simulating the discharge time series were made.

Firstly, the short study period of PWS data made me decide to overlap the calibration period in WALRUS with the validation period. Furthermore, a calibration period of one year could give well calibrated parameter values (Brauer et al., 2014b). Besides, calibration results improved significantly when two summer seasons were included in the calibration period and the calibration did not end during a peruiod with little water avaiable in virtual reservoirs of WALRUS.

Four state variables ($dG0$, $dV0$, $hQ0$ and $hS0$) could have been attributed as initial conditions in WALRUS. The warm-up period of the validation runs were short at the end of the summer (study period). Water levels can drop too far in summer when little rainfall input is given and too little reduction of evapotranspiration is taken into account. Since, the groundwater level is most sensitive to the length of the warm-up period, it was decided to only set the $dG0$ as initial condition before running WALRUS.

6 | Conclusion and recommendations

In this study, the accuracy of quality-controlled (QC) Personal Weather Stations (PWS) in observing rainfall and predicting discharge were assessed in a Dutch lowland catchment. The QC filter developed by de Vos et al. (2019) was applied to the network of PWS located in a defined boundary around the the Oude IJssel catchment. All stations were evaluated in time where the QC filter identified and filtered intervals containing erroneous measurements.

Rainfall measurements by PWSs were evaluated in space and time before (raw PWS) and after quality-control (QC-PWS). Validation studies of both individual and catchment-averaged time series of raw and QC-PWS were performed including a comparison with catchment-averaged time series of the operational weather radar (unadjusted and real-time) w.r.t. a reference radar product (gauge-adjusted and offline). Accordingly, the effect of a thinned PWS network density on individual and catchment-averaged rainfall measured by PWSs was investigated.

The Wageningen Lowland Runoff Simulator (WALRUS) was used as rainfall-runoff model to make discharge simulations using catchment-averaged time series of raw PWS, QC-PWS, the unadjusted and gauge-adjusted radar as rainfall forcing where simulations were validated w.r.t. the reference input for the main catchment and a sub catchment situated in the upstream region.

Quality control yielded a bias reduction of 10.6% in the 5 min rainfall sums measured by Personal Weather Stations while 85.1% of the original data remained which are similar results that were found in the PWS network in the capital city of the Netherlands that had a far greater network density.

The spatial density of PWS in the existing network was sufficient for application of a QC check, since only 2.38% of the original data in the network was filtered because spatial requirements of the quality control filter were not met. 20.0% of the data did not meet the spatial requirements on average when 1/3rd of the original network size was taken. Yet, a bias reduction of 20.6 % was yielded on average. Though the thinned PWS networks are not neutrally chosen and contain much sample uncertainty, metrics have shown that no significant difference in QC filter performance was found when network density reduced considerably.

The unadjusted radar systematically underestimated the reference 5 min averaged rainfall depths with a bias of -0.164 mm, while catchment-averaged rainfall depths measured by personal weather stations slightly overestimated the reference with a bias of only 0.025mm. No less bias was registered after quality control of PWS, however time series varied less and correlated better per time step and over the study period relative to the reference.

Discharge simulations were made over the study period of 11 months and two precipitation events in winter and summer and best simulations were made forcedby the quality-controlled personal weather stations (NSE = 0.98, averaged over the catchments during study period; NSE = 0.91, averaged over the catchments and events), followed by the input of personal weather stations before quality control (NSE = 0.95; NSE = 0.82) and lastly the operational weather radar during both the study period and the two selected events (NSE = 0.70; NSE = 0.78) where more accurate rainfall observations resulted in more accurate discharge predictions. Residuals became larger during peak discharge events for all input data for main and sub-catchment that resulted in lower NSE values. Only the unadjusted radar did better describe the peak discharge during the summer event resulting in a higher average NSE of the events. Errors in rainfall measurements propagated in the predicted discharge over the two events, the main and sub catchment and all input data with a relative higher error found in the peak discharge volumes. Hence, a lower reliability of the catchment-averaged rainfall time series for the sub catchment resulted in higher relative peak discharge volume error percentages.

In a broader perspective, it is concluded that quality-controlled personal weather stations observe rainfall and predict discharge far more accurate on the catchment-scale compared to the operational weather radar and thus enlarge the potential for operational hydrological applications in the Netherlands. Even without quality control, PWSs outperformed the operational weather radar on the catchment scale. Proof was found that quality of discharge simulations is strongly influenced by the quality of their forcing input as was acknowledged by multiple hydrological studies in the past.

Two considerations should be made for future research where firstly explicit methodological recommendations are made and secondly an implicit research extension is suggested. First, assessment over a full year or longer is recommended considering multiple catchments where their respective network layout is taken into account. Enabling seasonality analysis and the inclusion of multiple events where error propagation analysis by introducing imposed errors in the rainfall forcing data are recommended. On the other hand, multiple catchments

and a longer study period could compensate for the measurement uncertainty in observed discharge and can be insightful in the ability of PWS in accurately predicting the observed discharge.

Secondly, the gauge-adjusted radar was used as a reference precipitation data source, while the method proposed in this study has potential to outperform the reference product. Personal weathers stations, which have proof to provide high-quality rainfall measurements in real-time when undergone a quality control check, have a similar temporal and far greater spatial resolution than traditional rain gauges have in the Neterlands. Furthermore, the worldwide network of deployed stations is growing till the present day that raises ability and potential of this research to extend over Dutch national borders. Hence, investigating the potential of PWS as correction method for real-time weather radar is strongly recommended. Eventually a PWS-adjusted radar product may be developed that is, in contrast to the gauge-adjusted radar dataset used for validation, applicable in near real-time that could potentially observe rainfall extremes and predict local floods more accurately in the future.

Acknowledgements

I would like to thank multiple people and parties who helped me throughout my Master's thesis. First and most of all, my supervisors Claudia Brauer and Lotte de Vos who supported me whenever they could during these times of working in our home offices and who provided me with constructive feedback during our meetings. Secondly, I would like to thank Aart Overeem (KNMI) who granted me the Personal Weather Station data from brand Netatmo and the KNMI for freely distributing the unadjusted and gauge-adjusted radar data and evapotranspiration data. Thirdly, I would like to thank water board Rijn & IJssel for our fruitful discussion at the start of my thesis and in special Eoin Burke who provided me with geospatial vector data of the Oude IJssel and its tributaries and the calibrated parameter values of the weir Rhedebrugge. Next, I would like to thank my fellow student Bram Wijnants who put major effort in supporting and providing me with the generation of clipped and re-projected grids of German RADOLAN data. At last, I would like to thank my other fellow students who took part in the weekly thesis rings and helped me improving the quality of my work by giving critical advise and from whom I learned by giving constructive feedback to themselves.

References

Alexander, L.V., Zhang, X., Peterson, T.C., Caesar, J., Gleason, B., Klein Tank, A., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., et al., 2006. Global observed changes in daily climate extremes of temperature and precipitation. Journal of Geophysical Research: Atmospheres 111.

Ball, J.E., 1994. The influence of storm temporal patterns on catchment response. Journal of Hydrology 158, 285–303.

Bárdossy, A., Das, T., 2008. Influence of rainfall observation network on model calibration and application. Hydrology and earth system sciences 12, 77–89.

Bárdossy, A., Seidel, J., El Hachem, A., 2020. The use of personal weather station observation for improving precipitation estimation and interpolation. Hydrology and Earth System Sciences Discussions 2020, 1–23.

Bell, S., Cornford, D., Bastin, L., 2015. How good are citizen weather stations? addressing a biased opinion. Weather 70, 75–84.

Berenguer, M., Corral, C., Sánchez-Diezma, R., Sempere-Torres, D., 2005. Hydrological validation of a radar-based nowcasting technique. Journal of Hydrometeorology 6, 532–549.

Berne, A., Delrieu, G., Creutin, J.D., Obled, C., 2004. Temporal and spatial resolution of rainfall measurements required for urban hydrology. Journal of Hydrology 299, 166–179.

Borga, M., Degli Esposti, S., Norbiato, D., 2006. Influence of errors in radar rainfall estimates on hydrological modeling prediction uncertainty. Water resources research 42.

Borga, M., Tonelli, F., 2000. Adjustment of range-dependent bias in radar rainfall estimates. Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere 25, 909–914.

Brauer, C., Teuling, A., Torfs, P., Uijlenhoet, R., 2014a. The wageningen lowland runoff simulator (walrus): a lumped rainfall–runoff model for catchments with shallow groundwater. Geoscientific model development 7, 2313–2332.

Brauer, C., Torfs, P., Teuling, A., Uijlenhoet, R., 2014b. The wageningen lowland runoff simulator (walrus): application to the hupsel brook catchment and the cabauw polder. Hydrology and Earth System Sciences 18, 4007–4028.

Brauer, C.C., Overeem, A., Leijnse, H., Uijlenhoet, R., 2016. The effect of differences between rainfall measurement techniques on groundwater and discharge simulations in a lowland catchment. Hydrological Processes 30, 3885–3900.

Brauer, C., T.P.T.A.U.R.., 2017. The Wageningen Lowland Runoff Simulator WALRUS. Wageningen University Research.

Chen, A.B., Behl, M., Goodall, J.L., 2018. 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, pp. 25–28.

Drost, J., 2016. Modellering van het hydrologische effect van klimaatverandering in de oude ijssel met walrus. Master's Thesis, Wageningen University. .

Easterling, D.R., Meehl, G.A., Parmesan, C., Changnon, S.A., Karl, T.R., Mearns, L.O., 2000. Climate extremes: observations, modeling, and impacts. science 289, 2068–2074.

Emmanuel, I., Andrieu, H., Leblois, E., Flahaut, B., 2012. Temporal and spatial variability of rainfall at the urban hydrological scale. Journal of hydrology 430, 162–172.

Fabry, F., Bellon, A., Duncan, M.R., Austin, G.L., 1994. High resolution rainfall measurements by radar for very small basins: the sampling problem reexamined. Journal of Hydrology 161, 415–428.

Fletcher, T.D., Andrieu, H., Hamel, P., 2013. Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art. Advances in water resources 51, 261–279.

Golroudbary, V.R., Zeng, Y., Mannaerts, C.M., Su, Z.B., 2018. Urban impacts on air temperature and precipitation over the netherlands. Climate Research 75, 95–109.

Hazenberg, P., Torfs, P., Leijnse, H., Delrieu, G., Uijlenhoet, R., 2013. Identification and uncertainty estimation of vertical reflectivity profiles using a lagrangian approach to support quantitative precipitation measurements by weather radar. Journal of Geophysical Research: Atmospheres 118, 10–243.

Jenkins, G., 2014. A comparison between two types of widely used weather stations. Weather 69, 105–110.

Klein Tank, A., Beersma, J., Bessembinder, J., Van den Hurk, B., Lenderink, G., 2015. Knmi'14 climate scenarios for the netherlands: A guide for professionals in climate adaptation.

Krajewski, W., Smith, J.A., 2002. Radar hydrology: rainfall estimation. Advances in water resources 25, 1387–1394.

Krajewski, W.F., Villarini, G., Smith, J.A., 2010. Radar-rainfall uncertainties: Where are we after thirty years of effort? Bulletin of the American Meteorological Society 91, 87–94.

Lenderink, G., Van Meijgaard, E., 2008. Increase in hourly precipitation extremes beyond expectations from temperature changes. Nature Geoscience 1, 511–514.

Liu, J., Li, J., Li, W., Wu, J., 2016. Rethinking big data: A review on the data quality and usage issues. ISPRS journal of photogrammetry and remote sensing 115, 134–142.

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., Loumagne, C., 2014. When does higher spatial resolution rainfall information improve streamflow simulation? an evaluation using 3620 flood events. Hydrology and Earth System Sciences 18, 575–594.

Lubben, R., 2020. Improving flow forecast skill by assimilating groundwater observations in walrus. Master's Thesis, Wageningen University. .

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., Kjeldsen, T., 2014. Review of trend analysis and climate change projections of extreme precipitation and floods in europe. Journal of Hydrology 519, 3634–3650.

Meier, F., Fenner, D., Grassmann, T., Otto, M., Scherer, D., 2017. Crowdsourcing air temperature from citizen weather stations for urban climate research. Urban Climate 19, 170–191.

Moulin, L., Gaume, E., Obled, C., 2009. Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations. Hydrology and Earth System Sciences 13, 99–114.

Muller, C., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A., Leigh, R., 2015. Crowdsourcing for climate and atmospheric sciences: Current status and future potential. International Journal of Climatology 35, 3185–3203.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part i—a discussion of principles. Journal of hydrology 10, 282–290.

Naus, L., 2017. Improving urban flood modelling using opportunistic data. Master's Thesis, Wageningen University. .

Niemi, T.J., Warsta, L., Taka, M., Hickman, B., Pulkkinen, S., Krebs, G., Moisseev, D.N., Koivusalo, H., Kokkonen, T., 2017. Applicability of open rainfall data to event-scale urban rainfall-runoff modelling. Journal of hydrology 547, 143–155.

Overeem, A., Buishand, T., Holleman, I., 2009a. Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. Water resources research 45.

Overeem, A., Holleman, I., Buishand, A., 2009b. Derivation of a 10-year radar-based climatology of rainfall. Journal of Applied Meteorology and Climatology 48, 1448–1463.

Scrumteam, 2020. Pilot netatmo-neerslagstations. URL: https://storymaps.arcgis.com/stories/d7687c547e1446ae

Trenberth, K., 2011. Changes in precipitation with climate change. Climate Research 47, 123–138.

Uijlenhoet, R., Berne, A., 2008. Stochastic simulation experiment to assess radar rainfall retrieval uncertainties associated with attenuation and its correction. Hydrology and Earth System Sciences 12, 587–601.

van der Valk, J., 2019. Het kalibreren van de nationale regenradar met persoonlijke weerstations. Eindrapport stage Nelen Schuurmans, Wageningen University. .

Villarini, G., Mandapaka, P.V., Krajewski, W.F., Moore, R.J., 2008. Rainfall and sampling uncertainties: A rain gauge perspective. Journal of Geophysical Research: Atmospheres 113.

de Vos, L., Leijnse, H., Overeem, A., Uijlenhoet, R., 2017. The potential of urban rainfall monitoring with crowdsourced automatic weather stations in amsterdam. Hydrology and Earth System Sciences 21, 765–777.

de Vos, L., Raupach, T., Leijnse, H., Overeem, A., Berne, A., Uijlenhoet, R., 2018. High-resolution simulation study exploring the potential of radars, crowdsourced personal weather stations, and commercial microwave links to monitor small-scale urban rainfall. Water Resources Research 54, 10–293.

de Vos, L.W., Leijnse, H., Overeem, A., Uijlenhoet, R., 2019. Quality control for crowdsourced personal weather stations to enable operational rainfall monitoring. Geophysical Research Letters 46, 8820–8829.

Westra, S., Fowler, H.J., Evans, J.P., Alexander, L.V., Berg, P., Johnson, F., Kendon, E.J., Lenderink, G., Roberts, N.M., 2014. Future changes to the intensity and frequency of short-duration extreme rainfall.

WRIJ, 2014a. Afvoerkarakteristieken. URL: https://www.wrij.nl/statisch/oude-ijssel/kopie-2/artikel/.

WRIJ, 2014b. Gebiedsbeschrijving en indeling. URL: https://www.wrij.nl/statisch/oude-ijssel/kopie-algemene-0/gebiedsbegrenzing/.

Zheng, F., Tao, R., Maier, H.R., See, L., Savic, D., Zhang, T., Chen, Q., Assumpção, T.H., Yang, P., Heidari, B., et al., 2018. Crowdsourcing methods for data collection in geophysics: State of the art, issues, and future directions. Reviews of Geophysics 56, 698–740.

A | Appendices

A.1 Calibration in WALRUS

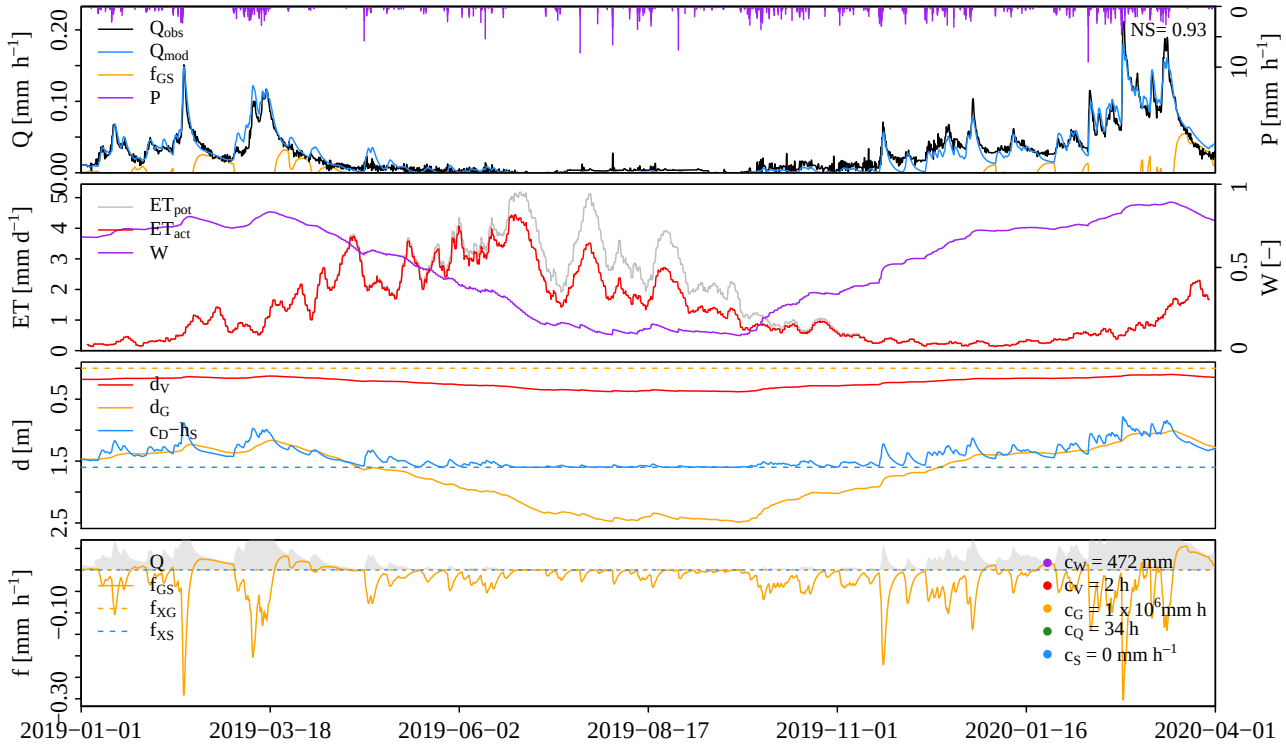


Figure A.1: WALRUS output for the automatic calibration run for the Oude IJssel catchment.

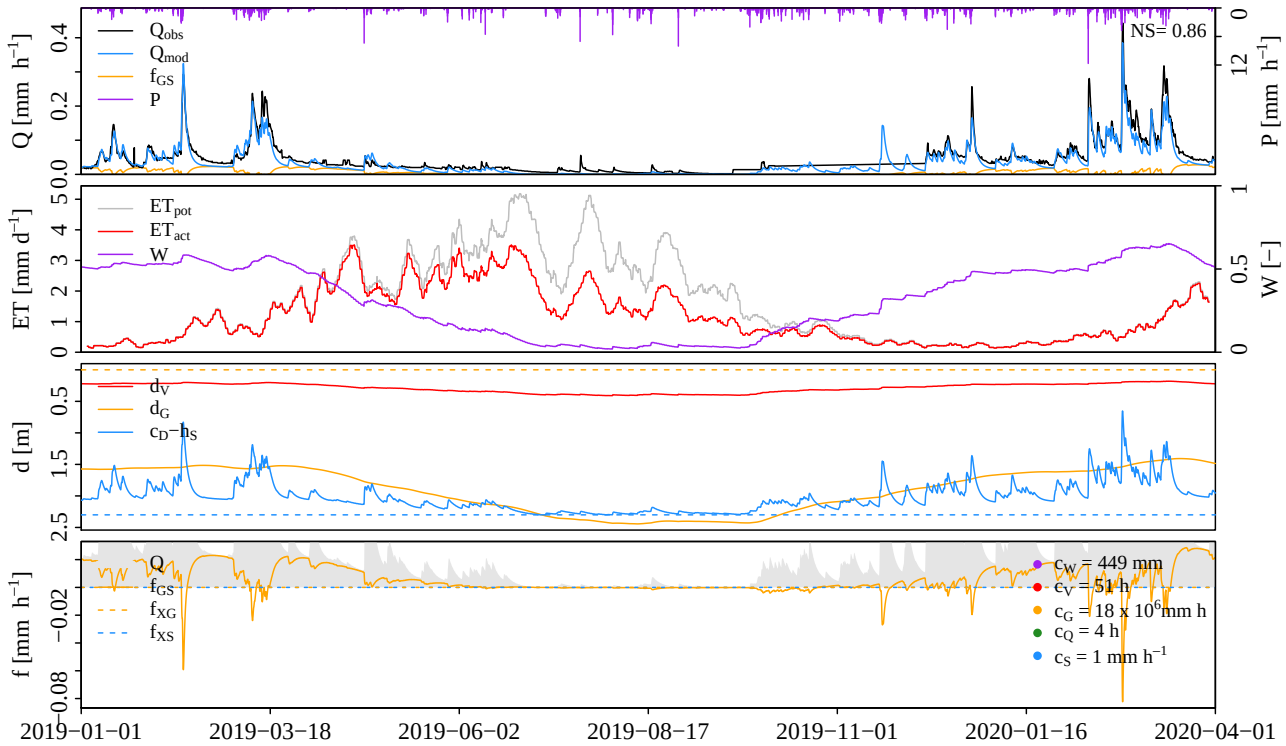
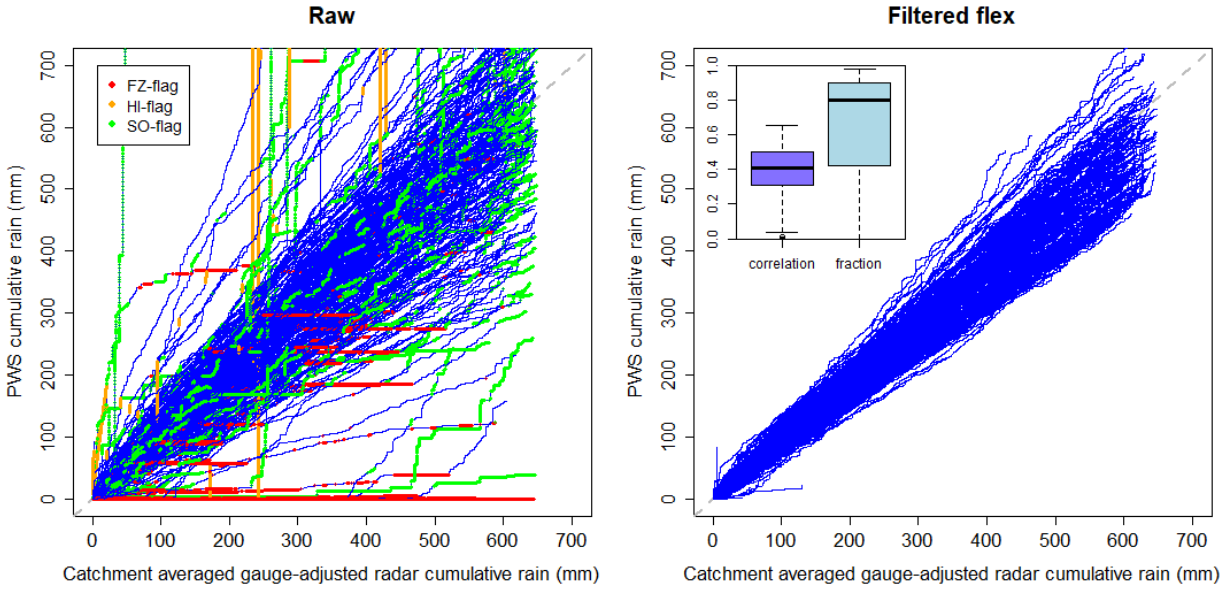
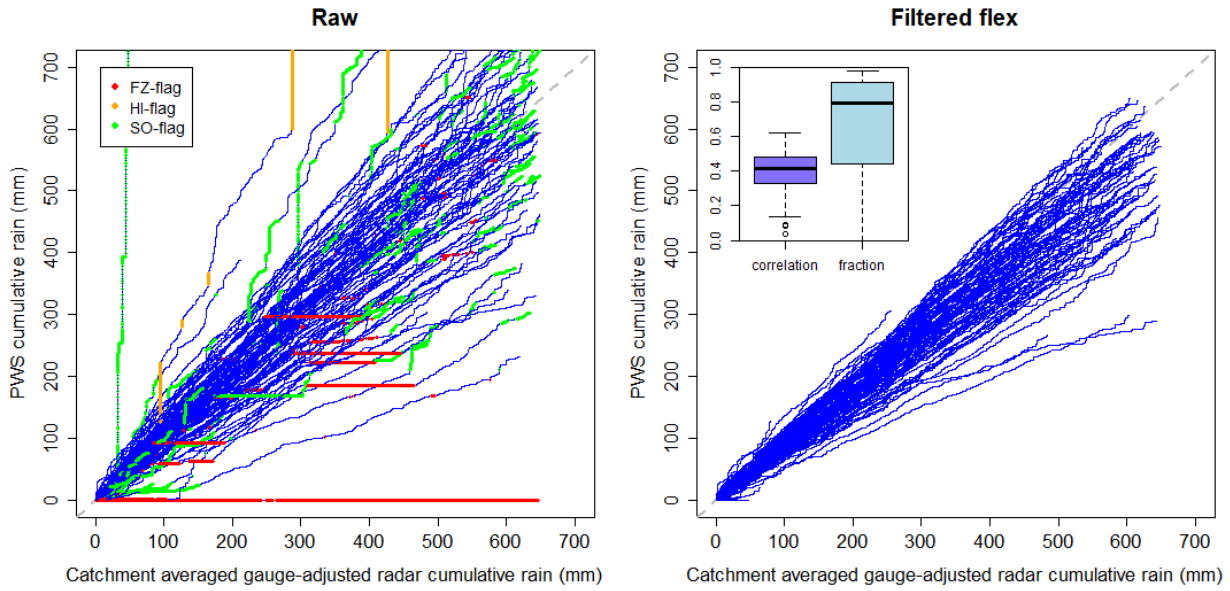


Figure A.2: WALRUS output for the automatic calibration run for Rhedebrugge sub-catchment.

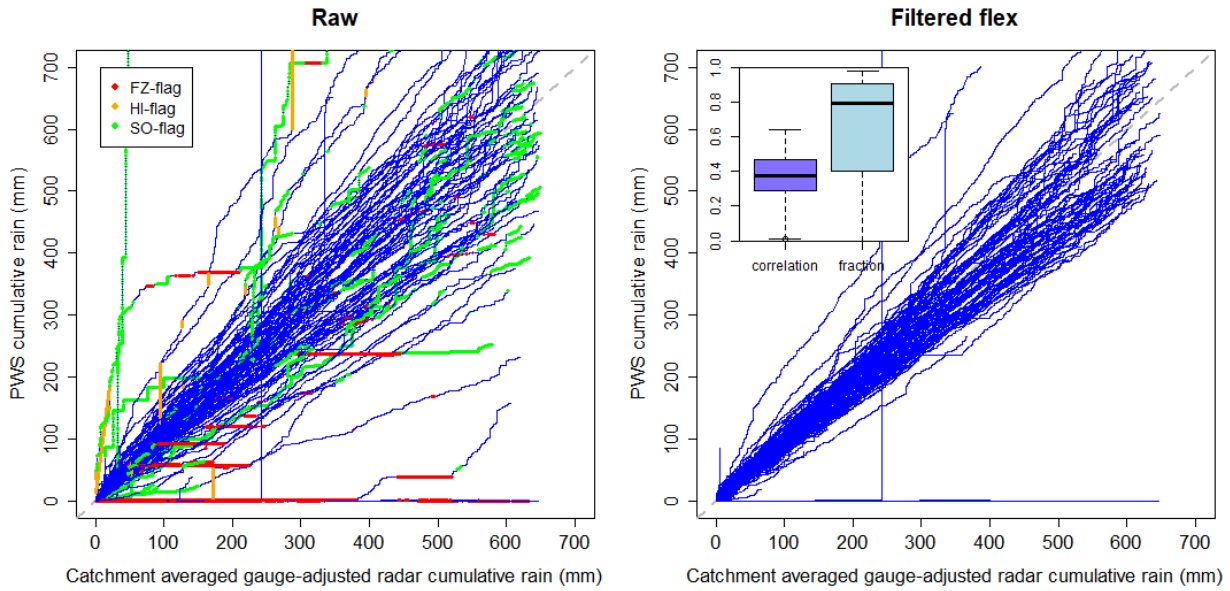
A.2 Double mass curves of total PWS dataset: flex filtered



A.3 Double mass curves of samples 1 and 2: flex filtered

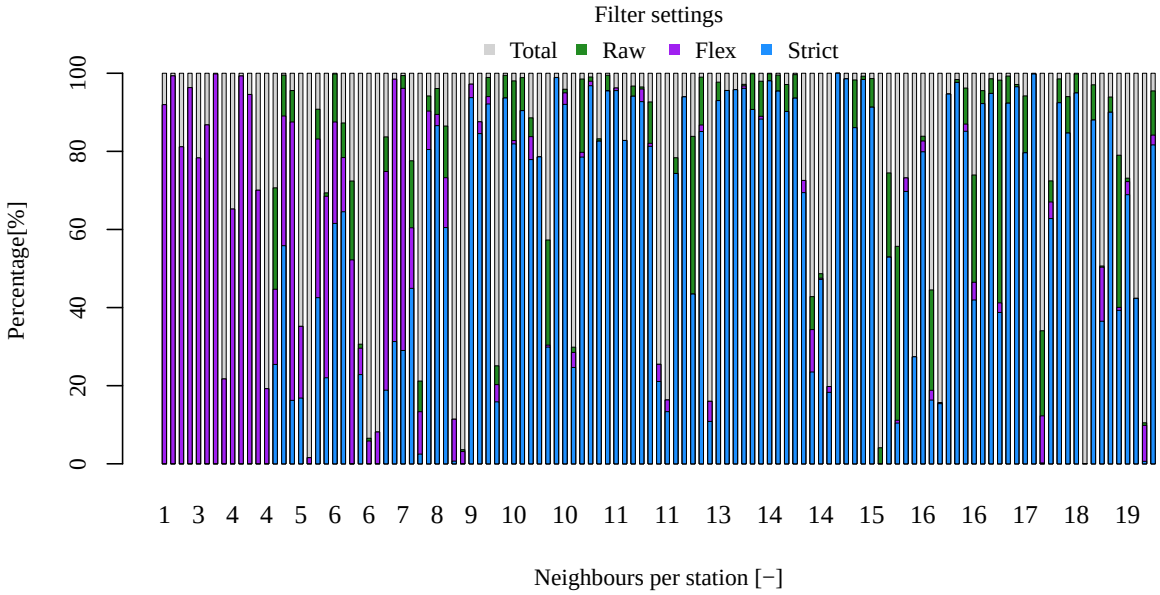


(a) Sample 1

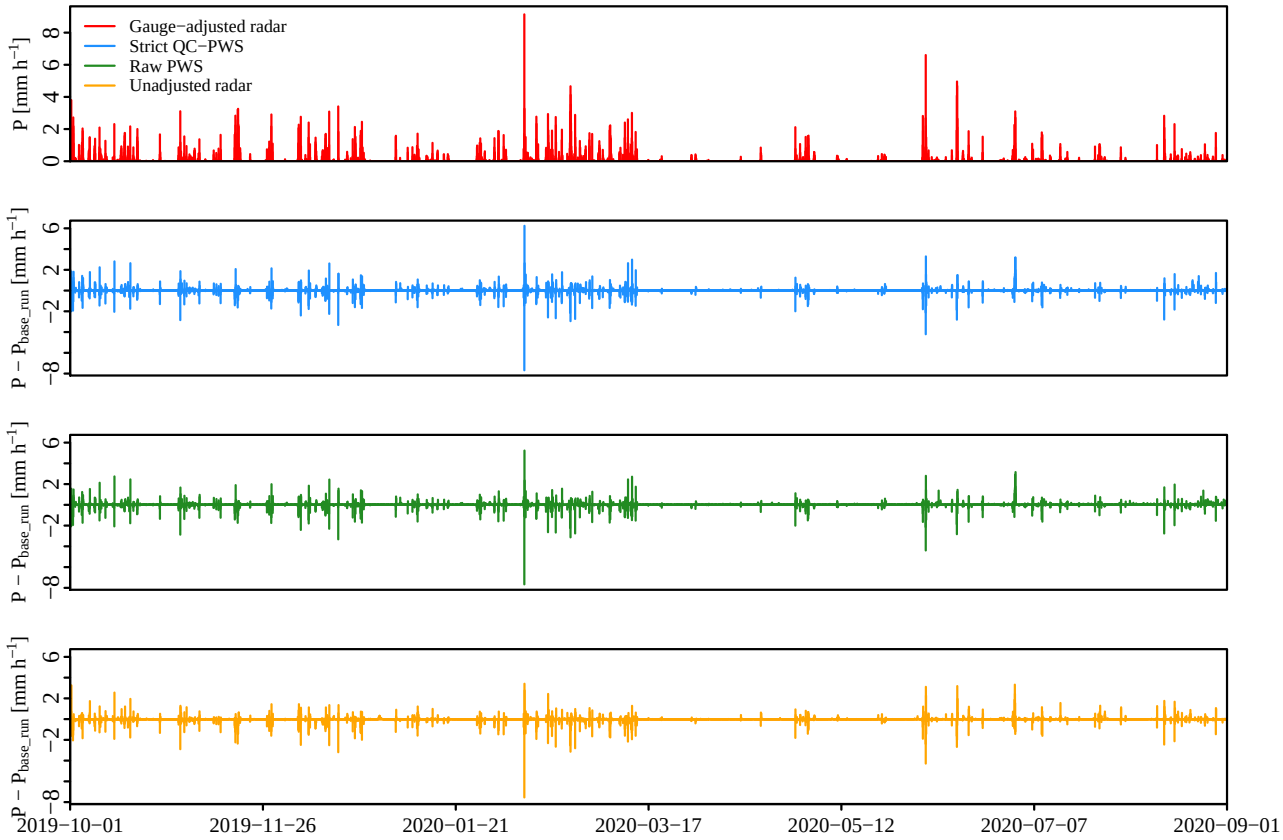


(b) Sample 2

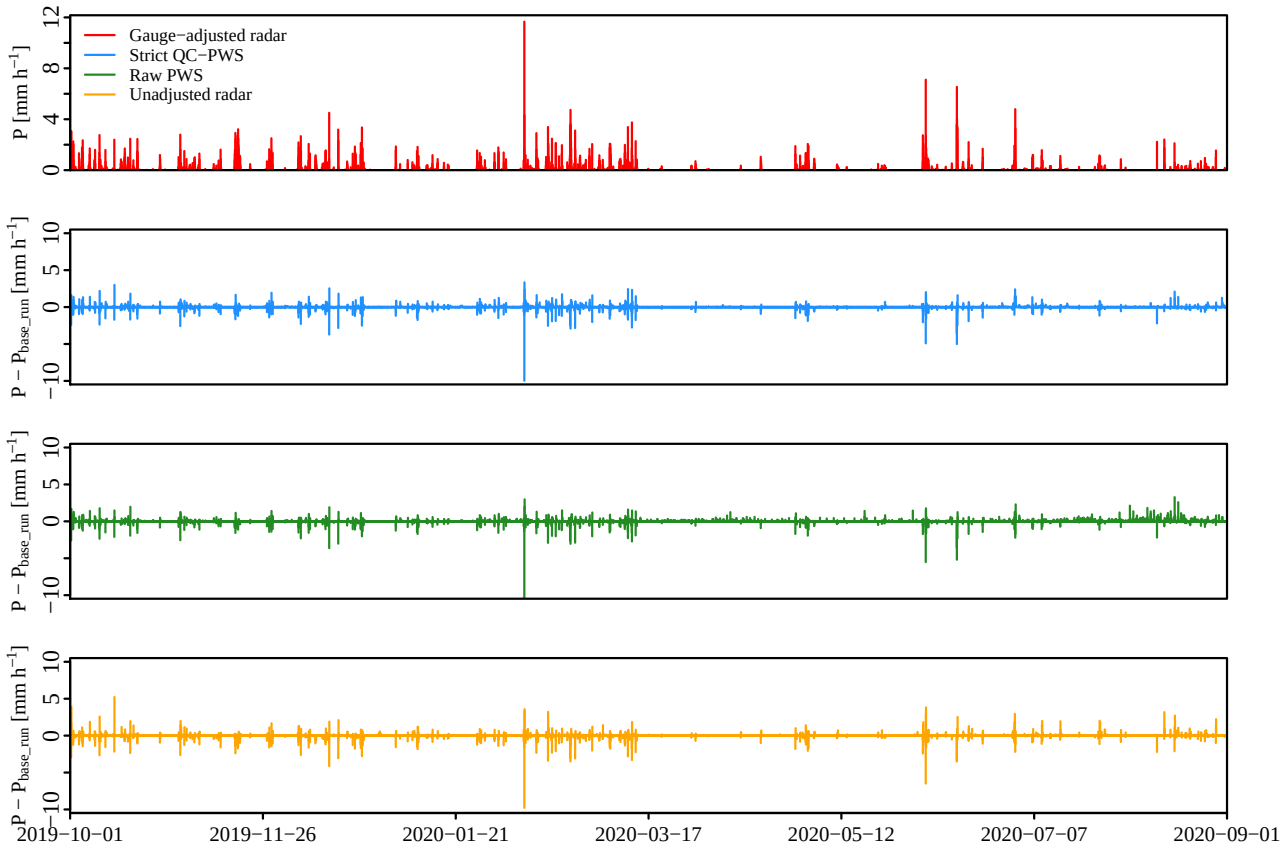
A.4 Effect of PWS network density: sample 3



A.5 Precipitation time series



(a) the Oude IJssel



(b) Rhedebrugge

Figure A.1: Comparing the residuals of QC-PWS strict, raw PWS and the unadjusted radar precipitation time series in the three sub figures relative to the gauge-adjusted radar in the top panel.