https://doi.org/10.5194/hess-2023-195
Preprint. Discussion started: 31 August 2023

Hydrology and Earth System Sciences
Open Access
Discussions

Technical note: Overview and comparison of three quality control algorithms for rainfall data from personal weather stations

Abbas El Hachem[1], Jochen Seidel[1], Tess O'Hara[3], Roberto Villalobos Herrera[4], Aart Overeem[2], Remko Uijlenhoet[5], András Bárdossy[1], and Lotte de Vos[2]

[1]Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart, D-70569 Stuttgart, Germany
[2]Royal Netherlands Meteorological Institute (KNMI), de Bilt, The Neherlands
[3]School of Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, UK
[4]School of Civil Engineering, Universidad de Costa Rica, Ciudad Universitaria Rodrigo Facio, San José, Costa Rica
[5]Department of Water Management, Delft University of Technology, The Netherlands

Correspondence: Abbas El Hachem (abbas.el-hachem@iws.uni-stuttgart.de)

Abstract. The number of rainfall observations from personal weather stations (PWSs) has increased significantly over the past years; however, there are persistent questions about data quality. In this paper, an examination and comparison of three quality control algorithms (PWSQC, PWS-pyQC, and GSDR-QC) designed for the quality control of rainfall data is presented. The focus was on a series of rainfall events occurring in the Amsterdam area between May 2017 - May 2018. Quality issues observed include faulty zeros i.e., the underreporting of rainfall, significant gaps in the dataset, and systematic bias often caused by incorrect setup and installation of the PWS. The analysis shows that all three algorithms improve PWS data quality when cross-referenced against rain radar. The considered algorithms have different strengths and weaknesses depending on PWS and official data availability, making it inadvisable to recommend one over another without carefully considering the specific setting. The need for further objective quantitative benchmarking of QC algorithms requiring freely available test datasets representing a range of environments, gauge densities, and weather patterns is highlighted.

1 Introduction

Precipitation is highly variable in space and time and thus the accurate estimation of precipitation amounts is of fundamental importance for many hydrological purposes (Estévez et al., 2011), especially on smaller scales and high temporal resolutions such as in small catchments and in the field of urban hydrology (Berne et al., 2004; Ochoa-Rodriguez et al., 2015; Cristiano et al., 2017), where typical rain gauge networks are not always capable to capture the spatio-temporal variability of precipitation. Weather radar provides rainfall estimates with good spatial coverage, but since this is an indirect measurement of atmospheric volumes, this type of data suffers from errors and uncertainties (Fabry, 2015; Rauber and Nesbitt, 2018). One approach to improve precipitation estimates is the use of personal weather stations (PWSs), which have increased in number over the past years; in many areas, they significantly outnumber rain gauges deployed by national meteorological services. Some of the most popular and widely available PWSs are simple, low-cost instruments that measure various meteorological parameters, including temperature, wind, and rainfall. Since these PWSs are installed by people who may not have access to,

https://doi.org/10.5194/hess-2023-195
Preprint. Discussion started: 31 August 2023

or knowledge of, optimal gauge placement, it is expected that many of these stations are not set up and maintained according to professional standards, therefore data from PWSs are prone to error. Overall, there is a high availability of PWS data but the expected quality of these data is fairly low. As with all weather observations, in order to make constructive use of PWS rainfall observations, the application of reliable quality control (QC) is vital. Many national meteorological services and other institutions have operational QC algorithms for their precipitation data, but these are typically not open source and are not tailored for PWS data. This can be because they assume a higher data availability and smaller bias than what is often found in PWS devices. In the past years, several QC methods for PWS rainfall data have been introduced through scientific publications, and they are typically applied to PWS datasets in different geographical areas or time periods. This lack of overlap in climate, conditions, network density, etc., can make it difficult for a reader to compare these methods.

The aim of this paper is to present and compare three different open-source QC methods designed especially for precipitation data. They can be run in the public Opportunistic Precipitation Sensing Network (OpenSense) sandbox environment (https://github.com/OpenSenseAction/OPENSENSE_sandbox) developed in the framework of the OpenSense EU COST-Action 20136 (https://opensenseaction.eu/). Two were designed to work with observations from high-density, but low-quality data. The third originated as a method to QC sub-daily rain data from various sources and can be applied to PWS data. The methods are applied to the same publicly available PWS rainfall dataset from the Amsterdam metropolitan area in the Netherlands. By showcasing the main similarities and differences in the methods this work points considerations as to the suitability of a QC method and assists interested users in selecting the most appropriate QC for their use. Lastly, by following the open data and open source concept these results are reproducible. The OpenSense COST-Action strives to promote FAIR principles in research, which are increasingly adopted and required by publishers, funding agencies, and academic institutions (Boeckhout et al., 2018).

This paper is structured as follows. Section 2 describes the study area and PWS dataset on which the three QC methods are applied. Section 3 gives a short overview of the three different algorithms. The method of this study is detailed in Section 4, followed by the Results in Section 5. This is followed by conclusions and advice on deciding which QC to use for a particular purpose in Section 6.

2 Study area and data

The study area is the Amsterdam metropolitan area in the Netherlands with a size of 575 km^2, where 134 Netatmo PWSs with a rain module are available between May 2016 and June 2018 (Fig 1). This corresponds to one PWS per 4.3 km^2. The average distance between the PWS in the study area is around 730 m. The rainfall data originate from Netatmo as time series with variable time intervals of ∼5 min. The data have been rewritten into rainfall amounts in mm within fixed 5-min time intervals by de Vos (2019). The dataset is characterized by many gaps and long periods of missing data, signified as ’NA’-values.

As a reference data set, a gauge-adjusted radar product from the Royal Netherlands Meteorological Institute (KNMI) with a 1 km spatial and 5-minute temporal resolution was used (referred to as radar reference from here on). This dataset is a radar rainfall product, corrected with validated rain gauge data from the official monitoring network, and is freely available after a delay of typically 1.5 months on the KNMI data platform (KNMI, 2023). Details on its construction can be found in Overeem et al. (2009a, b, 2011).

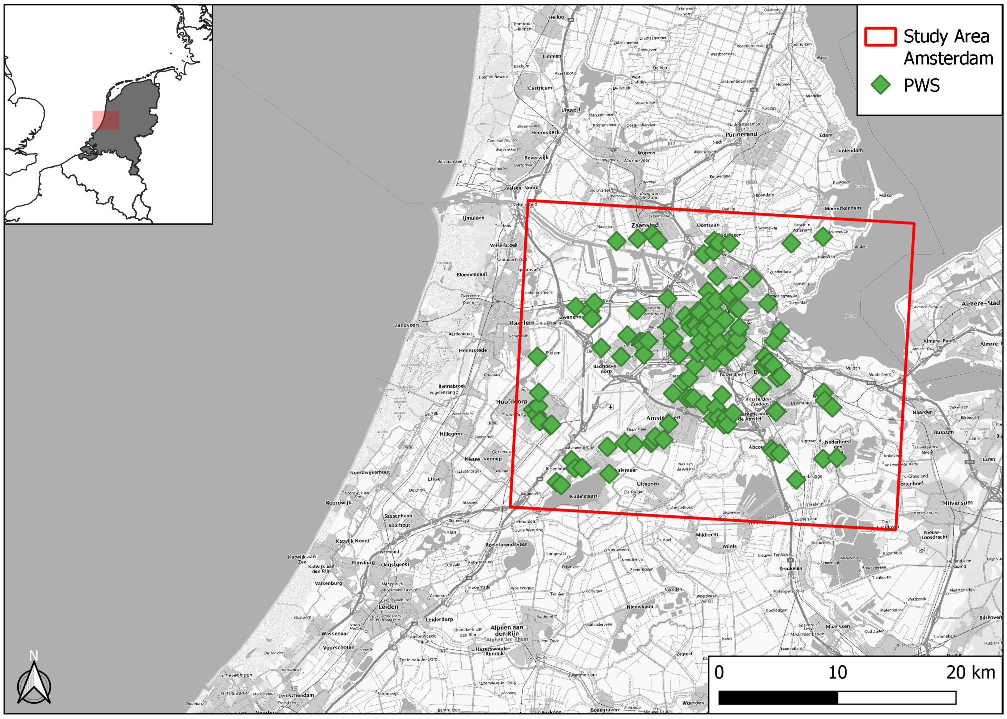


Figure 1. Overview map of the study area in the Amsterdam metropolitan area in the Netherlands. The red box depicts the domain used for the QC comparison. The PWS locations are denoted by the green dots. Background map credit to © Google Maps.

3 Description of the QC algorithms

3.1 PWSQC

PWSQC was originally developed and published by de Vos et al. (2019). It consists of several QC modules, all relying on neighbour checks. A Faulty Zero filter checks for periods of 0 mm rainfall while nearby PWSs register rainfall. The High Influx filter detects unrealistic high measurements compared to its surroundings. The Station Outlier check calculates the correlation between a PWS and each neighbouring PWS and starts flagging when the correlation of most becomes too low. Finally, a dynamic bias correction factor (which differs for each PWS and can change in time) is calculated and applied to the observations. For the initial value of the bias correction factor, an auxiliary dataset can be considered to derive a proxy for the overall bias of the whole dataset. This will improve the results, but no auxiliary data is required for the application of PWSQC. The method attributes flags to individual observations that can then be filtered, it does not exclude complete time-series. PWSQC has originally been applied on the same dataset as the one used in this paper and showed promising results.

The method has been implemented in R, and is openly available (de Vos, 2021). Later, a radar version of this algorithm has been constructed in Python that makes use of unadjusted radar at the location of a PWS as input for the QC. Then neighbouring PWSs are only employed to improve the radar input data (Van Andel, 2021).

3.2 PWS-pyQC

PWS-pyQC was first introduced by Bárdossy et al. (2021). It was used in a German-wide study in Graf et al. (2021) and an event-based analysis in Bárdossy et al. (2022). The method is implemented in Python and is open-source software available in El Hachem (2022). The QC algorithm consists of three main parts. The first identifies reliable PWSs using a space-time dependence structure derived from a reference observation network (denoted as primary network). The main assumption is that the PWS values might be wrong but their order (their ranks) are correct. The second step corrects the bias in the magnitudes of the values of each PWS individually using the ranks of the PWS and the corresponding neighboring primary observations. The final step consists of an event-based filter to identify erroneous PWS observations (false zeros, false extremes) for corresponding time intervals. This event-based filter is based on a leave-one-out cross-validation approach and was further developed in El Hachem et al. (2022).

PWS-pyQC relies on a reference network (a primary network) with reliable observations to filter the PWS data. This is usually acquired from the official rain gauge network. However, in the study area, there is only one KNMI rain gauge with hourly temporal resolution available which is not sufficient for deriving a reference dependence structure. Hence, this dependence structure was derived from the radar gauge-adjusted KMNI product by taking the times series of 20 randomly chosen pixels as the primary network.

3.3 GSDR-QC

GSDR-QC is the QC algorithm developed to construct the Global SubDaily Rainfall dataset (Lewis et al., 2019), and is fully described in Lewis et al. (2021). The complete procedure relies on a two-step process, first, 25 QC checks flag suspicious data. In the original implementation, three of these checks require access to the restricted-access Global Precipitation Climatology Centre (GPCC) daily and monthly precipitation databases; however, these are not essential and are not used in this 'local' implementations of the GSDR-QC algorithm (as applied herein). Once data have been flagged, a rulebase uses 11 of the 25 QC checks to determine suspicious observations which are removed from the dataset, of these 8 checks are used in local implementations that lack access to the GPCC. The algorithm flags and removes suspicious individual observations, rather than entire gauge datasets, and does not attempt to alter (bias correct) observations, making it the most conservative of the QC methods applied.

Briefly, the eight QC checks included in the GSDR-QC rulebase used herein include checks against neighbouring gauges (×2), checks for extremely large values (×2), for long dry spells, for repeated non-zero values, and for suspect daily and monthly accumulations. Locally appropriate records of maximum daily and hourly rainfall depths are required for the extreme value checks. A key aspect of the neighbour checks is that they are applied to an aggregated daily total, to avoid any potential issue caused by the higher variability and intermittency of hourly rainfall (Lewis et al., 2021). This variability and intermittency

https://doi.org/10.5194/hess-2023-195
Preprint. Discussion started: 31 August 2023

Table 1. Comparison of the QC algorithms

Criteria	**PWSQC**	**PWS-pyQC**	**GSDR-QC Local**
QC modules	1. Faulty zeroes filter 2. High influx filter 3. Station outliers filter 4. Bias correction	1. Indicator based filter 2. Bias correction 3. Event based filter	1. Flagging of suspicious observations using defined rule base 2. Filtering of suspicious observations not meeting QC criteria
WMO QC classification types (WMO, 2021)	Consistency - Spatial, tolerance/range	consistency spatial, spike and streak	Format, completeness, consistency, tolerance/range, spike and streak
Reference data set required?	No, but optional part of initialisation of 4.	Yes, required for 1, 2 and 3	Yes, for user defined daily and hourly thresholds
Programming language	R	Python	Python
Comparisons	Neighbour PWSs comparison over various previous time intervals	PWS should fit in space-time dependence structure of reference data	Neighbouring gauges are compared to each other and optionally against a reference dataset
Level of QC-allocation	- Per individual measurement - Dynamic nature is suitable for longer time series	- Per full PWS time series - Event based	- Per individual measurement - Dynamic nature is suitable for longer time series
Output	- Original PWS dataset - 3 flag files conveying flag attribution to individual observations for all three QC - 1 file with Bias correction factors generated for each observation - Bias adjusted PWS dataset with only reliable observations	- Set of trustworthy PWS - Individual bias correction for each time series - Implausible time intervals removed for each time series	- Flag file for each gauge showing individual test results - Output file with reliable observations
Applicability notes	- Could be applied in real-time, though not in current code - Module 3 and 4 requires lead-up of 3 (rainy) weeks and are operationally slow - Can be applied on small timesteps (5min)	-Applicable firstly for static datasets - Could be applied in real-time - The data of trustworthy PWS could be further acquired and verified - Can be applied on hourly timesteps	- Intended for static datasets of up to global scale - The global GSDR-QC algorithm uses the Global Precipitation Climatology Centre (GPCC) as reference for monthly and daily values - GPCC dataset is not available for use outside of the German weather service. - Can be applied on hourly timesteps
QC methods are available in OpenSense Sandbox (https://github.com/OpenSenseAction/OPENSENSE_sandbox)			

tend to be higher in data originating from official networks (the original application of GSDR-QC) since they are much less dense than PWS networks, especially in urban areas (O'Hara et al., 2023). Up to 10 neighbouring gauges within a 50 km radius are used, originally a 3-year observation overlap is required but this has been reduced to 1 year for this implementation. Table 1 gives a summary of the three different QC methods.

4 Method

The aim of this paper is not a comprehensive validation or benchmarking of the QC methods, but a first demonstration of their applicability and performance. In order to highlight the relative performance of the three different QC methods, four 24-h rainfall events with different spatiotemporal rainfall characteristics (Table 2) were chosen to showcase the results of the

Table 2. Overview of the four rainfall events chosen for this case study

	Start	End	Event Characteristics
Event 1	12 May 2017 8:00 UTC	13 May 2017 8:00 UTC	Several showers
Event 2	27 Nov 2017 8:00 UTC	28 Nov 2017 8:00 UTC	Homogeneous rainfall, dry spell from 22:00 to 05:00
Event 3	15 Jan 2018 8:00 UTC	16 Jan 2018 8:00 UTC	Homogeneous rainfall, one very evident outlier PWS
Event 4	29 May 2028 8:00 UTC	30 May 2028 8:00 UTC	Convective rainfall from 14:00 to 22:00

different QC algorithms. These rainfall events were selected in such a way that the majority of the PWSs registered significant rainfall for a large duration of time.

PWSQC, PWS-pyQC, and GSDR-QC have all been applied to the same dataset of PWS rainfall observations. As the PWS-pyQC algorithm requires at least a 1-hour temporal resolution as data input, the 5-min PWS dataset has been aggregated to hourly values, where an hourly value can only be constructed, with the completeness condition that at least 10 out of 12 intervals were available. The GSDR-QC can be implemented on any consistent interval data, and in this instance, the same aggregated hourly dataset was used. PWSQC has been applied to that data in its raw 5-min temporal resolution. Intervals that were allocated a Faulty zero, High influx, and/or a Station outlier error were excluded. After QC, the PWS dataset was aggregated to hourly values with the same requirement on data availability.

4.1 Interpolation

The filtered and corrected PWS data are interpolated using Ordinary Kriging (OK) on the same grid as the reference data set. OK utilizes the spatial configuration of the points which is quantified by a fitted variogram model. The latter is derived in the rank space domain following the procedure in Lebrenz and Bárdossy (2017). The parameters were further adapted according to the aggregation interval-dependent parameters derived for the Dutch conditions in the work of Van de Beek et al. (2012). In case no suitable variogram could be derived, for example, due to the large number of zeros, an average spherical variogram was used, without a nugget value and with a sill scaled according to the data variance. For every hour of the selected daily event with positive PWS observations, the values in the domain are spatially interpolated. The number of accepted PWSs is accordingly noted. The daily map is acquired by accumulating the hourly maps.

4.2 Performance metrics

For the four events, 24h accumulations maps based on hourly interpolations were derived and compared. Subsequently, hourly areal averages over the study domain were calculated and compared. Furthermore, the number of remaining stations was calculated. A point value comparison was done by calculating the pair-wise correlation value, the bias and the coefficient of variation (CV) between the PWS and reference data for all PWS locations in the study domain.

The first evaluation metric is the Pearson correlation. It is a widely used pair-wise dependence measure to identify the presence (or absence), the strength, and the direction of a linear relationship between pairs of variables (for example, X and Y). The equation for calculating the Pearson correlation can be seen in equation 1.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{1}$$

The second metric is the relative bias defined as follows (eq. 2).

$$Bias = \frac{\overline{(x - y)}}{\bar{y}} \tag{2}$$

The third metric is the coefficient of variation (CV, eq. 3) and is used to quantify the dispersion in the data.

$$CV = \frac{\sigma(x - y)}{\bar{y}} \tag{3}$$

Where:

σ =	standard deviation
r_{xy} =	Pearson correlation coefficient
x_i =	value of x at time interval i
$\bar{x}$ =	average value of time series x
y_i =	value of y, the reference, at time interval i
$\bar{y}$ =	average value of time series y, the reference
n =	number of observations

5 Results

5.1 Spatial areal rainfall maps

Figure 2 shows the interpolated rainfall maps after the QCs have been applied to the PWS data and the radar reference for Event 4 (2018-05-29 08:00 – 2018-05-30 08:00 UTC). The highest peaks are not captured by PWSQC. The rainfall in the southwest part of the area is underestimated by all, but most severely by GSDR-QC which is least sensitive to detect faulty zeroes in the data (Fig 3). PWS-pyQC has the best metrics for this event although only 50% of the PWS are retained on average (c.f. Table 3). The corresponding maps of the other three events are shown in Appendix A. The rainfall maps after applying each of the QC algorithms show similar patterns to the radar reference.

5.2 Metrics for all four events

GSDR-QC shows the most remaining data after QC, while PWS-pyQC rejects most PWS stations on average. This is related to the faulty zero checks in the other two methods are implemented at the sub-daily timescale, whereas the GSDR-QC applies

https://doi.org/10.5194/hess-2023-195
Preprint. Discussion started: 31 August 2023

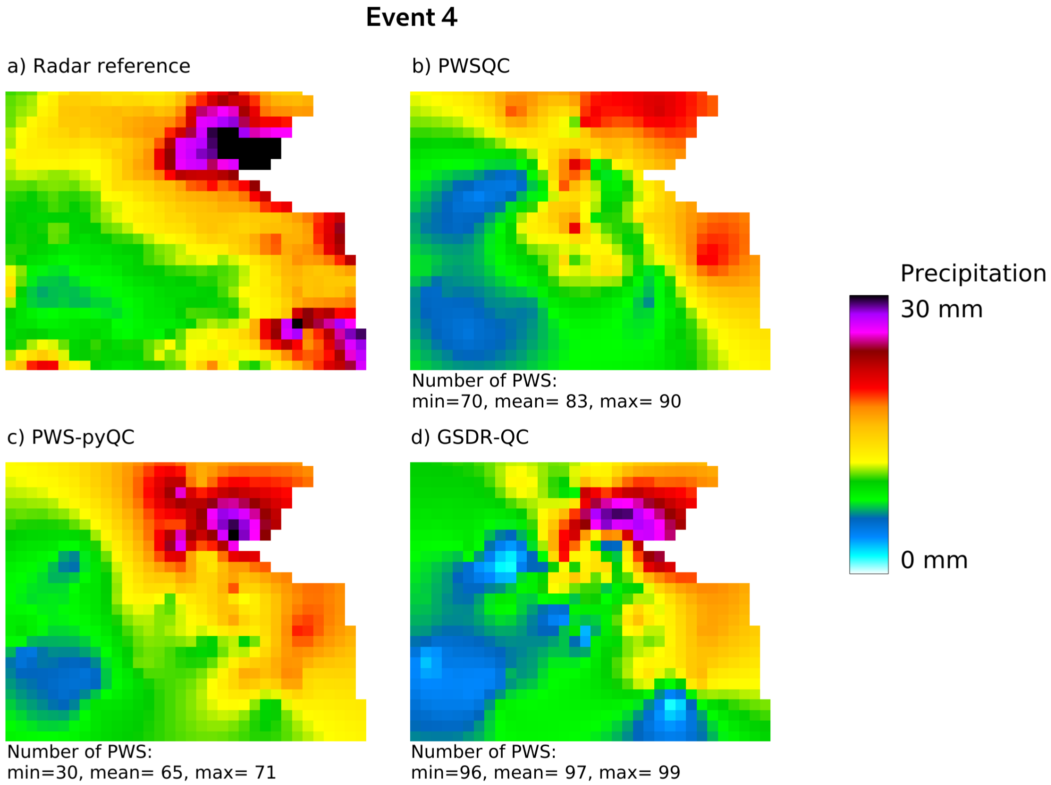


Figure 2. Rainfall maps for event 4. Panel a) shows the gauge-adjusted radar accumulation. Panels b), c), and d) show the interpolated PWS accumulations using the QC algorithms *PWSQC*, *PWS-pyQC* and *GSDR-QC*, respectively. Under each map the data availability after QC is indicated by providing the number of PWSs with hourly data, that were used to generate interpolated maps for the hour with the fewest (min) and highest (max) PWS remaining after QC, as well as the average (mean) over the 24-hour maps.

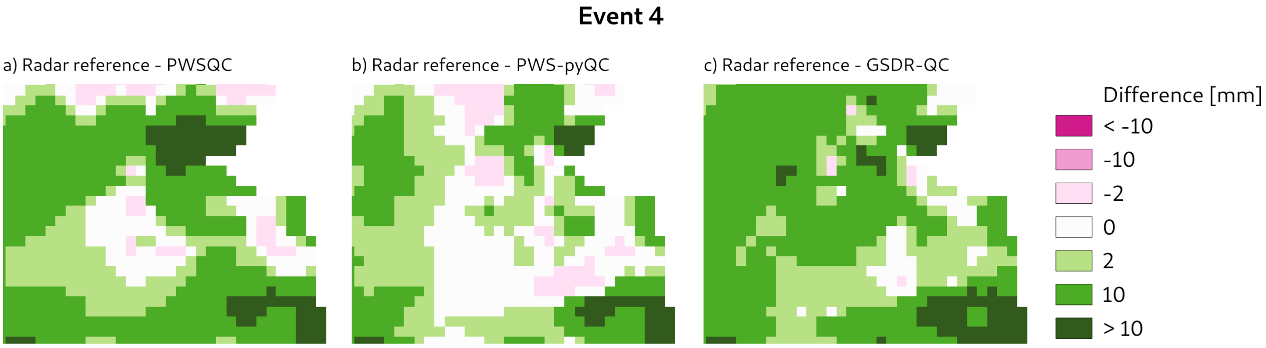

Event 4
a) Radar reference - PWSQC
b) Radar reference - PWS-pyQC
c) Radar reference - GSDR-QC
Difference [mm]
< -10
-10
-2
0
2
10
> 10

Figure 3. Differences between the radar reference and the interpolated maps from the three QC algorithms for event 4.

Table 3. Comparison metrics calculated for the four events

		Event 1	Event 2	Event 3	Event 4
Remaining PWS [%]	PWSQC	61	67	63	62
	PWS-pyQC	44	43	42	50
	GSDR-QC	74	71	70	73
Pearson Correlation	PWSQC	0.71	0.90	0.83	0.84
	PWS-pyQC	0.73	0.94	0.87	0.92
	GSDR-QC	0.64	0.83	0.48	0.86
Bias	PWSQC	-0.01	0.02	-0.13	-0.20
	PWS-pyQC	0.12	0.08	-0.03	-0.07
	GSDR-QC	-0.21	-0.09	-0.17	-0.25
CV	PWSQC	1.67	1.01	0.75	1.58
	PWS-pyQC	1.53	0.82	0.62	1.11
	GSDR-QC	1.64	1.29	1.70	1.44

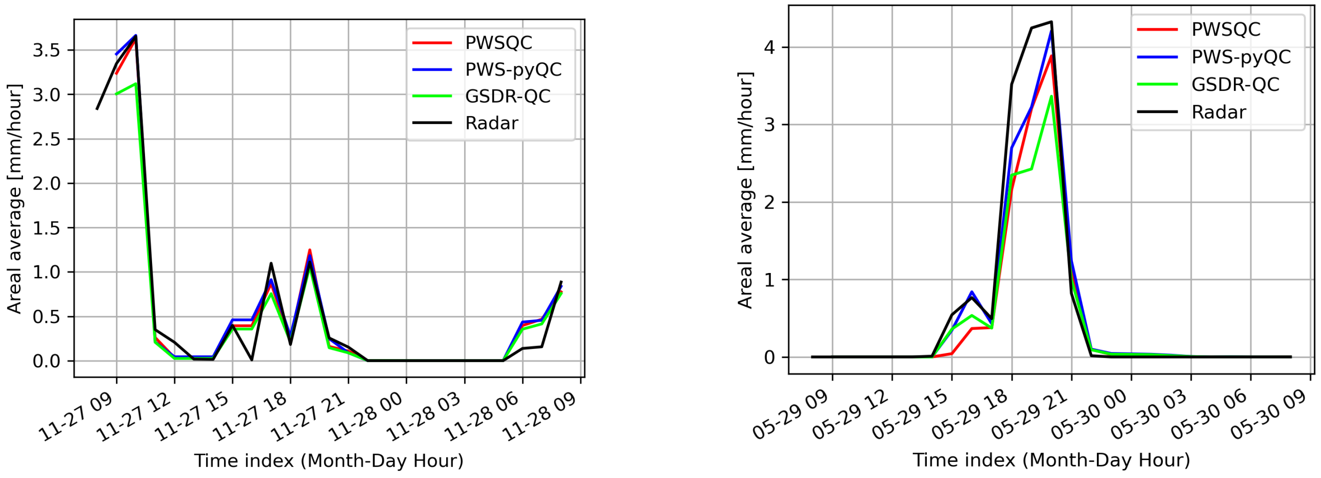
the check to daily aggregated data, resulting in reduced sensitivity to missing observations (see also scatterplots in Appendix B). The tendency of a higher negative bias in GSDR-QC was to be expected as no bias correction is implemented.
Overall, the metrics do not show a clear picture as they differ from event to event. Results after PWS-pyQC yield similar values for bias and Pearson correlation as PWSQC, and values for the coefficient of variation smaller than the other two QC methods.

5.3 Average areal rainfall values

For every hour in the selected events, the domain (see Figure 1) rainfall average value is calculated from each data set individually. Figure C1 shows an example of two daily events. For both cases, the hourly average rainfall values for all datasets show high similarities, where the peak for the first event is underestimated, especially for the GSDR-QC. Even though the spatial rainfall maps present large differences as seen in Figure 2, this is largely averaged out over the domain. Overall, the areal average timelines of all QC results match radar reference well although the individual rainfall patterns from the interpolated maps are different for some of the events.

5.4 Validation over longer periods

To investigate the difference between the performance of the three QC algorithms a cross-validation of an independent observation data set was carried out. From the radar reference product, 15 locations within the study area were randomly selected and used as validation locations for the interpolated rainfall maps of each QC method. The output of each QC method for hourly and daily time intervals was interpolated and compared to the radar reference at the 15 validation locations using the Pearson correlation as the target metric. Figure 5 shows slightly better results for daily data than for hourly data. Overall, the

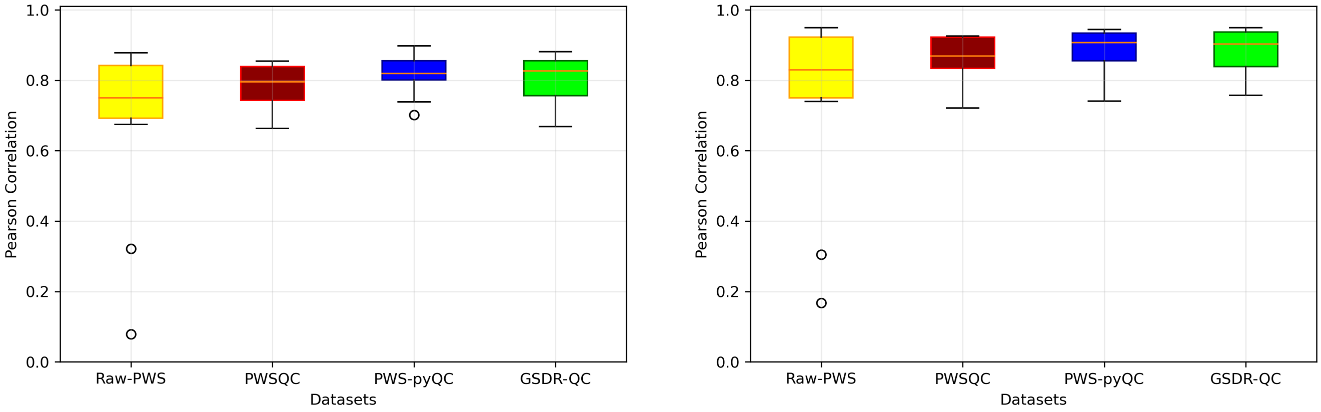


(a) Areal average for event 2

(b) Areal average for event 4

Figure 4. Panel (a) shows an example of areal rainfall over the Amsterdam metropolitan area for event 2, panel (b) for event 4.

variance after applying the QC methods is smaller than the raw PWS data but there is no clear difference between the individual methods. However, the correlations are higher after QC of the PWS data.



(a) Pearson correlation - Hourly data

(b) Pearson correlation - Daily data

Figure 5. Panels (a) and (b) show box plots of the Pearson correlation values for the 15 pixels between the interpolated and observed time series for hourly and daily data, respectively.

6 Discussion

As shown in previous studies, PWS may provide reliable rainfall data if these stations are set up and maintained correctly (Bárdossy et al., 2021; de Vos et al., 2017); however, as this is often not the case, thorough QC is required before PWS data can be used for hydrological applications. This is also confirmed by the results of the validation over longer periods as shown in section 5.4.

The three QC methods show that faulty zeroes are problematic to detect. Furthermore, there is a tendency that local rainfall peaks as shown in Figure 2 to be missed or not captured in their spatial extent. This can either be due to the fact that the density of PWS is too low and it also depends on whether such local rainfall maxima are captured by one or more PWSs and not discarded as high influx or outlier.

The GSDR-QC method was not developed with a specific focus on observations from Netatmo PWSs. This is most evident from a lower sensitivity to faulty zero observations and the absence of automated correction of bias in GSDR-QC, corresponding with the after-QC results found in the interpolated maps with "dry" spots and overall higher bias. This is a limitation of the GSDR-QC that makes it less well suited to the QC of PWS observations where faulty zeros are a common feature of the datasets. PWS-QC and PWS-pyQC were designed to work with data from (Netamo) PWS. Specific error types like faulty zeroes or high influxes that are often found in this type of PWS data are accounted for by these two algorithms.

PWS-QC does not require any additional information from more reliable observations and can thus be used in areas without reference data and only PWS data is available. PWS-pyQC requires a reference data set (primary stations) set to derive information about the spatial pattern of indicator correlations. Such a data set can either be a dense rain gauge network or as shown in this study, a gauge-adjusted radar product. In the absence of such a data set, PWS-pyQC cannot be used. GSDR-QC requires a reference data set as well, however, as this is used for hourly and daily thresholds gridded precipitation data sets can also be used. PWS-pyQC typically retains the smallest number of stations compared to PWSQC and GSDR-QC. PWSQC has been applied conservatively where if not enough data was available to determine a flag, the data is not excluded. Given that PWSQC is applicable to 5 min time series and PWS-pyQC and GSDR-QC to hourly, the observations remaining calculation is slightly different. Also, the indicator correlation filter of PWS-pyQC rejects the complete PWS series whereas the other QC methods flag and/or remove suspicious individual observations.

7 Conclusions and outlook

In this work, we presented an inter-comparison study of open-source QC algorithms on PWS rainfall data. The aim was to introduce different concepts on how QC for PWS stations can be used and to give an overview of additional input data and strengths and limitations of the individual QC methods. Overall, all presented QC methods can improve the quality of PWS rainfall data, but the underestimation of small-scale rainfall peaks which typically occur during convective events, and the correct attribution of faulty zeroes are two aspects that require further improvement of PWS QC algorithms. Nevertheless, studies like the ones from Bárdossy et al. (2022) and Overeem et al. (2023) have shown the added value of PWS for improving rainfall estimates for extreme events and quantitative precipitation estimation on a European scale, respectively.

In conclusion, a QC algorithm has to be selected based on the available data, and whilst the subsequent dataset might not be perfect, there is an improvement from the raw data. Based on these example events and previous work, the following suggestions emerge: the PWS-pyQC algorithm is best suited to an area with a widely spaced but comprehensive official monitoring network, the PWSQC algorithm is most useful where there is a dense PWS network, and the GSDR-QC is most appropriate in locations where the PWS network is sparse and comprises rain gauges from a range of manufacturers (resulting in a range of potential errors). Further work is required for comprehensive sensitivity testing across a range of environments, monitoring networks, and weather patterns to provide more quantitative guidance on the most appropriate QC method.

We plea for making open opportunistic sensing data on a European or even global level (or restricted access for research purposes), which would foster the development and improvement of QC and rainfall retrieval algorithms. Eventually, this will lead to improved precipitation products and applications such as validation of weather/climate models, hydrological modelling, nowcasting, etc. Furthermore, there is a need for large benchmark radar and rain gauge datasets from different regions and climates. Such benchmark datasets would facilitate a fair intercomparison of QC algorithms and even different opportunistic sensor rainfall estimates from commercial microwave links (CMLs) and satellite microwave links (SMLs). Intercomparison studies also require appropriate metrics and the aforementioned datasets. A discussion on standardized benchmark metrics to be used for intercomparison studies is needed. Benchmarking and intercomparison of algorithms for opportunistic sensor data, merging of opportunistic sensor data and traditional data from rain gauges and radars, and the integration of these data into standard observation systems are objectives that are currently being addressed in the OpenSense COST Action (https://opensenseaction.eu/).

Data availability. The gauge-adjusted radar product from the Royal Netherlands Meteorological Institute (KNMI) is freely available on the KNMI data platform (KNMI, 2023). The employed PWS dataset is publicly available (de Vos, 2019)

Code and data availability. The corresponding code for the analysis is available upon request from the contact author. All QC software is available as open source and can be accessed as follow: **PWSQC** is available as R code under https://github.com/LottedeVos/PWSQC. **PWS-pyQC** is available as Python code under https://github.com/AbbasElHachem/pws-pyqc. **GSDR-QC** is available as Python code under https://github.com/nclwater/intense-qc and https://pypi.org/project/intense/.

Author contributions. AEH implemented PWS-pyQC to the raw PWS data and did the analysis following the outputs of the 3 QC methods. JS assisted in analysing and displaying the results. TOH and RHV implemented GDDR-QC to the raw PWS data. AO, RU, and AB aided in reviewing and structuring the manuscript. LdV organized the workflow and implemented PWSQC to the raw PWS data. All authors contributed to writing the manuscript.

Competing interests. At least one of the (co-)authors is a member of the editorial board Hydrology and Earth System Sciences.

Acknowledgements. This work was supported by the COST Action Opportunistic Precipitation Sensing Network (OpenSense; CA20136; https://opensenseaction.eu/). We thank the private company Netatmo for providing the PWS dataset and all citizen scientists for buying, installing, and maintaining these stations.

References

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

Bárdossy, A., Seidel, J., Eisele, M., Hachem, A. E., Kunstmann, H., Chwala, C., Graf, M., Demuth, N., and Gerlach, N.: Verbesserung der Abschätzung von Gebietsniederschlägen mittels opportunistischer Niederschlagsmessungen am Beispiel des Ahr-Hochwassers im Juli 2021, Hydrologie und Wasserbewirtschaftung, 66, 208 – 214, 2022.

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

Boeckhout, M., Zielhuis, G. A., and Bredenoord, A. L.: The FAIR guiding principles for data stewardship: fair enough?, European journal of human genetics, 26, 931–936, 2018.

Cristiano, E., ten Veldhuis, M.-C., and van de Giesen, N.: Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas – a review, Hydrology and Earth System Sciences, 21, 3859–3878, https://doi.org/10.5194/hess-21-3859-2017, 2017.

de Vos, L. W.: Rainfall observations datasets from Personal Weather Stations, https://doi.org/10.4121/uuid:6e6a9788-49fc-4635-a43d-a2fa164d37ec, 4TU.ResearchData. Dataset., 2019.

de Vos, L. W.: PWSQC Code, https://github.com/LottedeVos/PWSQC, 2021.

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

de Vos, L. W., Leijnse, H., Overeem, A., and Uijlenhoet, R.: Quality Control for Crowdsourced Personal Weather Stations to Enable Operational Rainfall Monitoring, Geophysical Research Letters, 46, https://doi.org/10.1029/2019GL083731, 2019.

El Hachem, A.: AbbasElHachem/pws-pyqc: OpenSense Integration, https://doi.org/10.5281/zenodo.7310212, 2022.

El Hachem, A., Seidel, J., Imbery, F., Junghänel, T., and Bárdossy, A.: Technical Note: Space–time statistical quality control of extreme precipitation observations, Hydrology and Earth System Sciences, 26, 6137–6146, https://doi.org/10.5194/hess-26-6137-2022, 2022.

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

Fabry, F.: Radar Meteorology: Principles and Practice, Cambridge University Press, Cambridge, U.K., https://doi.org/10.1017/CBO9781107707405, 2015.

Graf, M., El Hachem, A., Eisele, M., Seidel, J., Chwala, C., Kunstmann, H., and Bárdossy, A.: Rainfall estimates from opportunistic sensors in Germany across spatio-temporal scales, Journal of Hydrology: Regional Studies, 37, 100 883, https://doi.org/https://doi.org/10.1016/j.ejrh.2021.100883, 2021.

KNMI: Precipitation - 5 minute precipitation accumulations from climatological gauge-adjusted radar dataset for The Netherlands (1 km) in NetCDF4 format, retrieved July 2023, 2023.

Lebrenz, H. and Bárdossy, A.: Estimation of the Variogram Using Kendall's Tau for a Robust Geostatistical Interpolation, Journal of Hydrologic Engineering, 22, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001568, 2017.

Lewis, E., Fowler, H. J., Alexander, L., Dunn, R., Mcclean, F., Barbero, R., Guerreiro, S., Li, X. F., and Blenkinsop, S.: GSDR: A global sub-daily rainfall dataset, Journal of Climate, 32, 4715–4729, https://doi.org/10.1175/JCLI-D-18-0143.1, 2019.

Lewis, E., Pritchard, D., Villalobos-Herrera, R., Blenkinsop, S., McClean, F., Guerreiro, S., Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rustemeier, E., and Fowler, H. J.: Quality control of a global hourly rainfall dataset, Environmental Modelling & Software, 144, 105 169, https://doi.org/10.1016/j.envsoft.2021.105169, 2021.

Ochoa-Rodriguez, S., Wang, L.-P., Gires, A., Pina, R. D., Reinoso-Rondinel, R., Bruni, G., Ichiba, A., Gaitan, S., Cristiano, E., van Assel, J., et al.: Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation, Journal of Hydrology, 531, 389–407, 2015.

O'Hara, T., McClean, F., Villalobos Herrera, R., Lewis, E., and Fowler, H. J.: Filling observational gaps with crowdsourced citizen science rainfall data from the Met Office Weather Observation Website, Hydrology Research, 54, 547–556, https://doi.org/10.2166/nh.2023.136, 2023.

Overeem, A., Buishand, T. A., and Holleman, I.: Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar, Water Resour. Res., 45, n/a–n/a, https://doi.org/10.1029/2009WR007869, W10424, 2009a.

Overeem, A., Holleman, I., and Buishand, A.: Derivation of a 10-year radar-based climatology of rainfall, J. Appl. Meteor. Climatol., 48, 1448–1463, https://doi.org/10.1175/2009JAMC1954.1, 2009b.

Overeem, A., Leijnse, H., and Uijlenhoet, R.: Measuring urban rainfall using microwave links from commercial cellular communication networks, Water. Resour. Res., 47, n/a–n/a, https://doi.org/10.1029/2010WR010350, w12505, 2011.

Overeem, A., Leijnse, H., van der Schrier, G., van den Besselaar, E., Garcia-Marti, I., and de Vos, L.: Merging with crowdsourced rain gauge data improves pan-European radar precipitation estimates, Hydrol. Earth. Syst. Sc. Discuss. [preprint], https://doi.org/10.5194/hess-2023-122, submitted, 2023.

Rauber, R. M. and Nesbitt, S. L.: Radar Meteorology: A First Course, John Wiley & Sons, Hoboken, NJ, U.S.A., https://doi.org/10.1002/9781118432662, 2018.

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

Van de Beek, C., Leijnse, H., Torfs, P., and Uijlenhoet, R.: Seasonal semi-variance of Dutch rainfall at hourly to daily scales, Advances in Water Resources, 45, 76–85, 2012.

WMO: Guidelines on Surface Station Data Quality Control and Quality Assurance for Climate Applications, Tech. rep., MWO Geneva, Switserland, 2021.

https://doi.org/10.5194/hess-2023-195
Preprint. Discussion started: 31 August 2023

295 Appendix A: Additional Rainfall Maps

Figures A1-A3 show the interpolated 24h rainfall map after the corresponding QC algorithms have been applied. For event 2 (Fig. A2), all QC methods show more spatial variability than the radar reference, which is caused by some faulty zeros which are not detected by PWSQC and GSDR. Also, some higher values appear which were not obviously not identified as outliers by the QC methods. For event 3 (Fig. A2), there's a very evident outlier PWS with high rainfall amounts over 30mm. This 300 outlier was not detected by PWSQC and GSDR.

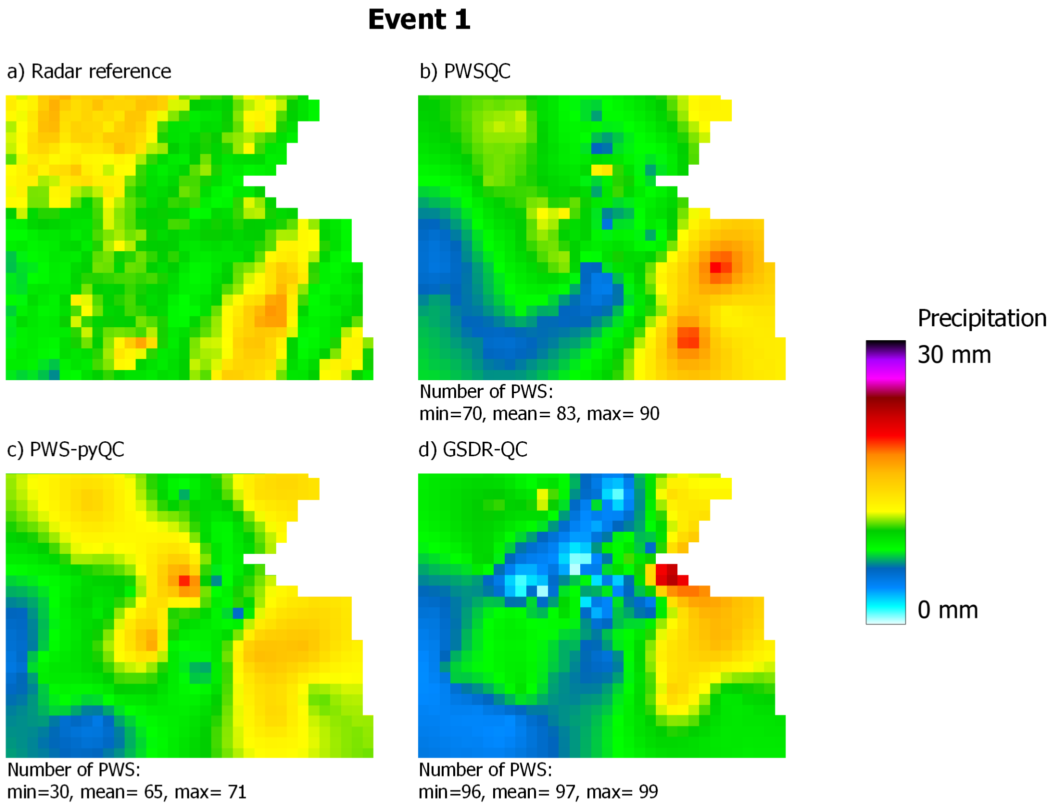


Figure A1. Rainfall maps for event 1. Panel a) shows the gauge-adjusted radar accumulation. Panels b), c), and d) show the interpolated PWS accumulations using the QC algorithms *PWSQC*, *PWS-pyQC* and *GSDR-QC*, respectively. Under each map the data availability after QC is indicated by providing the number of PWSs with hourly data, that were used to generate interpolated maps for the hour with the fewest (min) and highest (max) PWS remaining after QC, as well as the average (mean) over the 24-hour maps.

https://doi.org/10.5194/hess-2023-195
Preprint. Discussion started: 31 August 2023

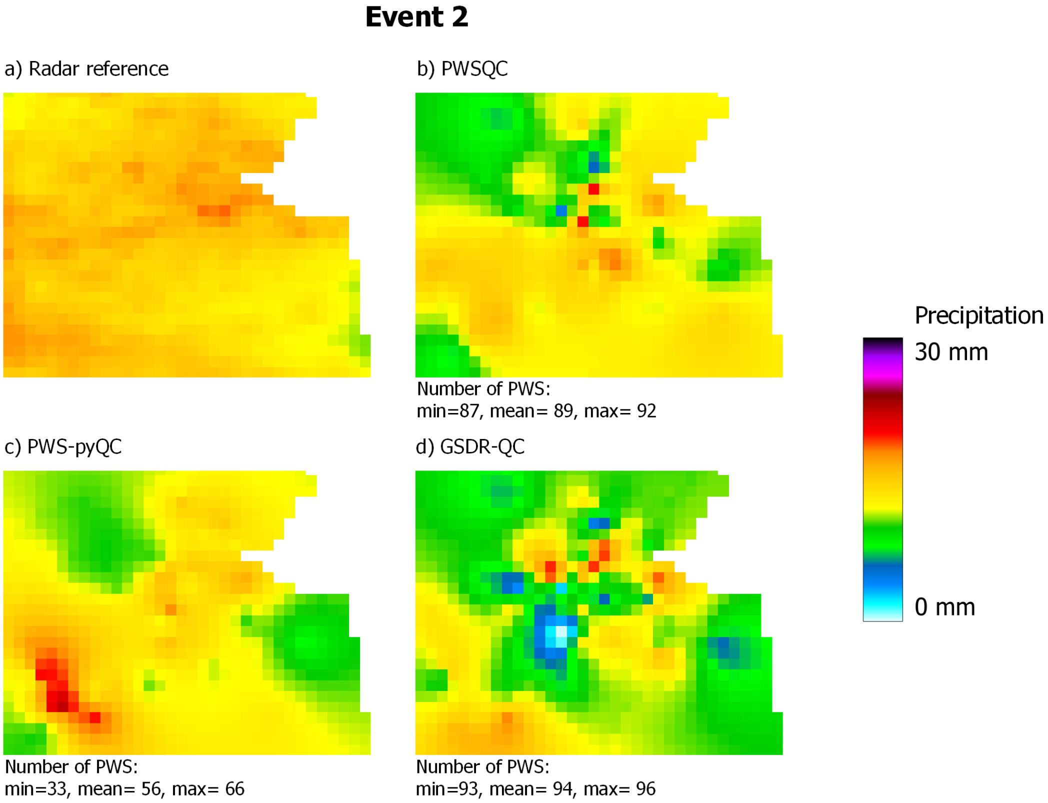


Figure A2. Rainfall maps for event 2. Panel a) shows the gauge-adjusted radar accumulation. Panels b), c), and d) show the interpolated PWS accumulations using the QC algorithms PWSQC, PWS-pyQC and GSDR-QC, respectively. Under each map the data availability after QC is indicated by providing the number of PWSs with hourly data, that were used to generate interpolated maps for the hour with the fewest (min) and highest (max) PWS remaining after QC, as well as the average (mean) over the 24-hour maps.

https://doi.org/10.5194/hess-2023-195
Preprint. Discussion started: 31 August 2023

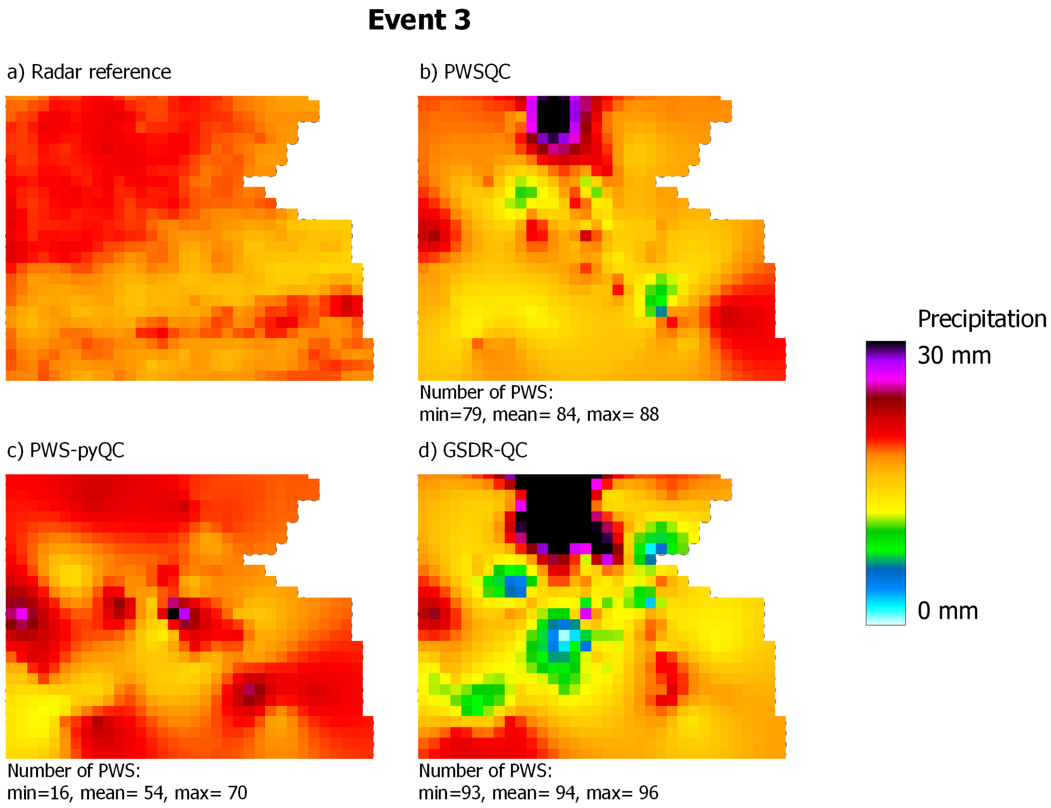


Figure A3. Rainfall maps for event 3. Panel a) shows the gauge-adjusted radar accumulation. Panels b), c), and d) show the interpolated PWS accumulations using the QC algorithms *PWSQC*, *PWS-pyQC* and *GSDR-QC*, respectively. Under each map the data availability after QC is indicated by providing the number of PWSs with hourly data, that were used to generate interpolated maps for the hour with the fewest (min) and highest (max) PWS remaining after QC, as well as the average (mean) over the 24-hour maps.

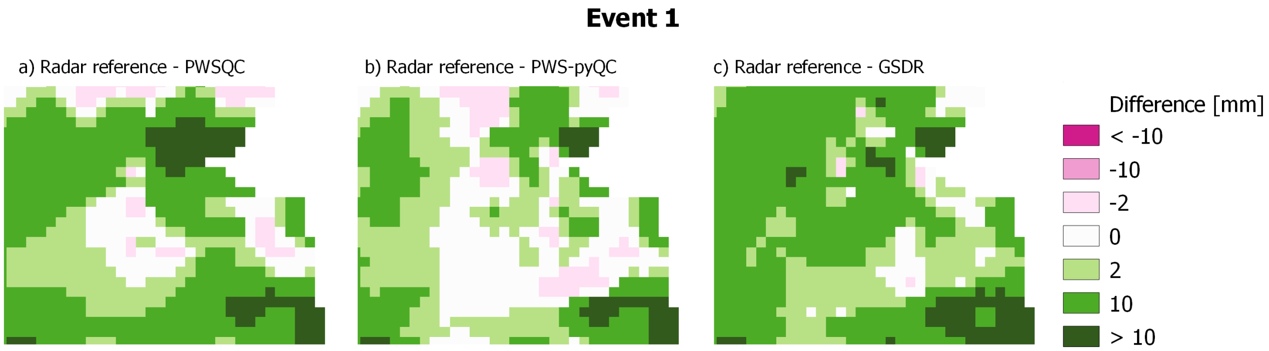

Event 1
a) Radar reference - PWSQC
b) Radar reference - PWS-pyQC
c) Radar reference - GSDR
Difference [mm]
< -10
-10
-2
0
2
10
> 10

Figure A4. Differences between the radar reference and the interpolated maps from the three QC algorithms for event 4.

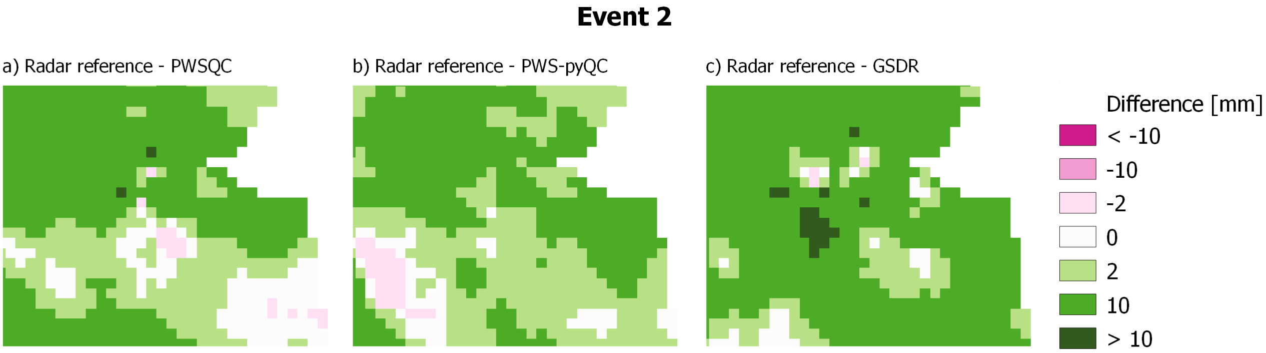

Event 2
a) Radar reference - PWSQC
b) Radar reference - PWS-pyQC
c) Radar reference - GSDR
Difference [mm]
< -10
-10
-2
0
2
10
> 10

Figure A5. Differences between the radar reference and the interpolated maps from the three QC algorithms for event 4.

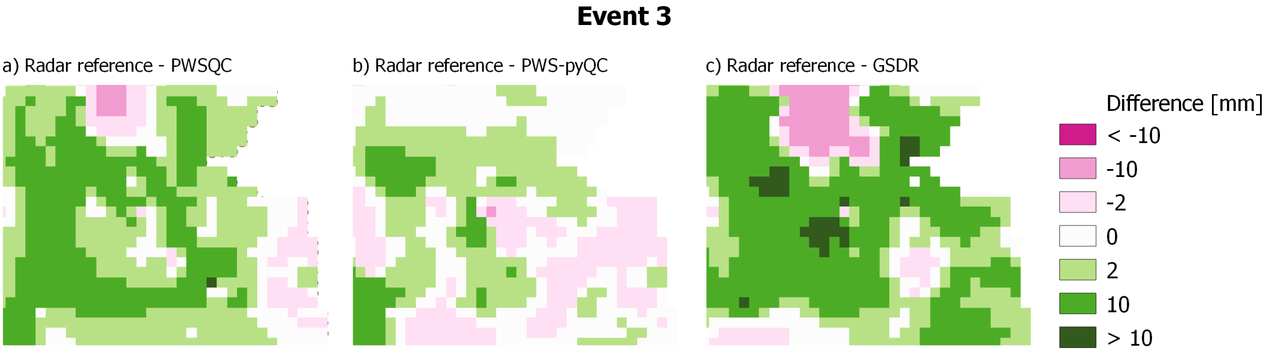
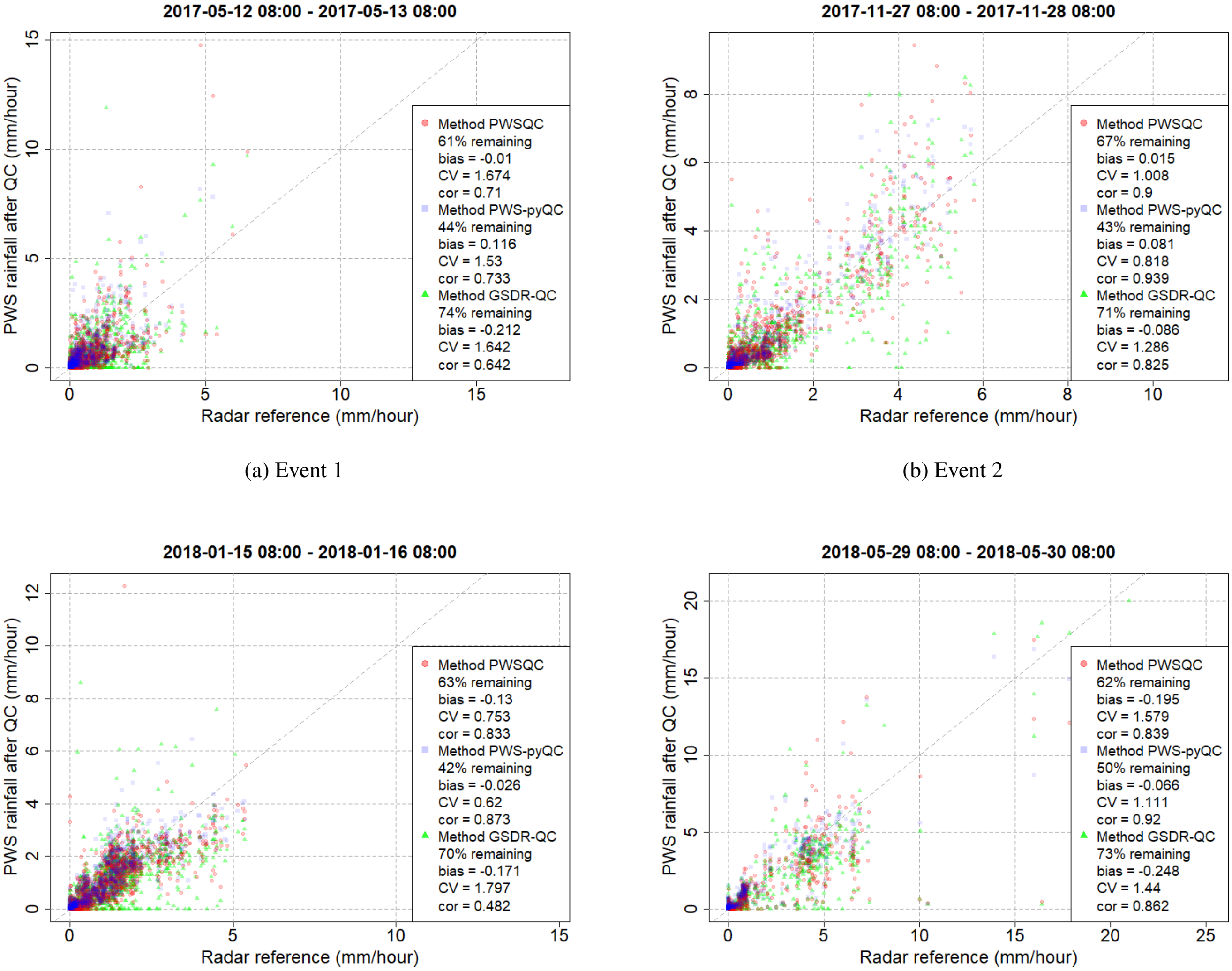

Event 3
a) Radar reference - PWSQC
b) Radar reference - PWS-pyQC
c) Radar reference - GSDR
Difference [mm]
< -10
-10
-2
0
2
10
> 10

Figure A6. Differences between the radar reference and the interpolated maps from the three QC algorithms for event 4.

Appendix B: Scatter plots for the four selected events

Figure B1 shows four scatter plots for the chosen events. The scatter plots are derived by comparing the hourly PWS data after QC has been applied, with a gauge-adjusted radar product, more specifically the overlying pixel of these PWS locations. Only the remaining hourly intervals for every QC method were considered. The data of PWSQC are displayed by the red dots, those of PWS-pyQC by the blue squares, and the GSDR-QC results by the green triangles. For every event, several metrics are calculated and showcased within each plot. For each QC method, the number of total data points in the event (134 PWSs * 24 hours) that is covered after filtering is provided as a percentage. Given that we did not start off with 100% data availability in the original PWS dataset, this should only be interpreted relative to the other QC method outcomes. This shows that after PWS-pyQC, most data is rejected.

GSDR-QC shows more remaining data after QC, evident are the 0 mm precipitation records in PWS data, while the radar reference records rainfall (the dots spread out horizontally on the x-axis). This is due to faulty zero checks in the other two methods being implemented at the sub-daily timescale, whereas the GSDR-QC applies the check to daily aggregated data, resulting in reduced sensitivity to missing observations.



(a) Event 1

(b) Event 2

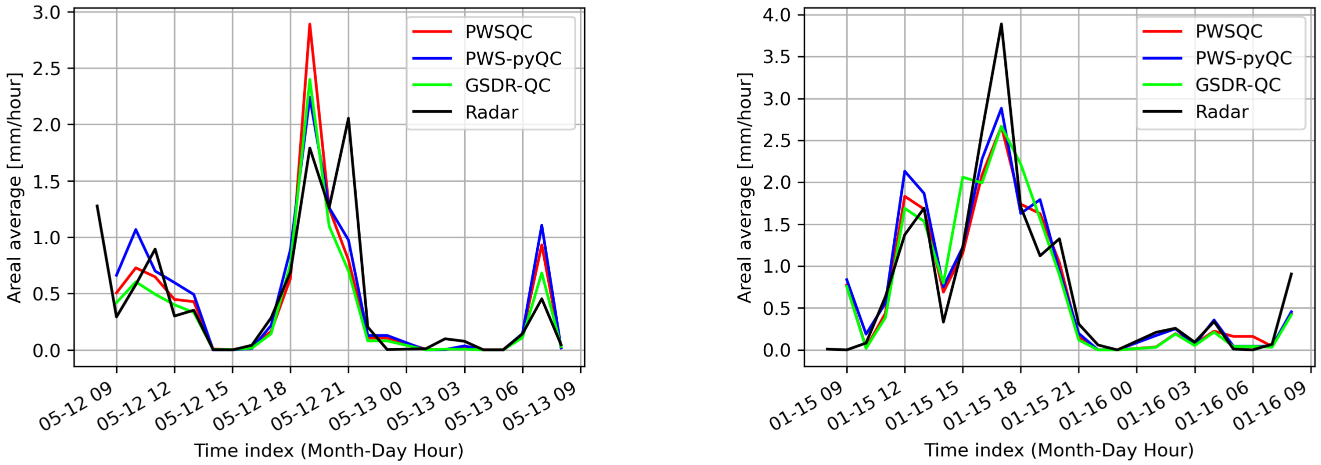
(c) Event 3

(d) Event 4

Figure B1. The scatterplots of hourly rainfall amounts of PWS after QC is applied against gauge-adjusted radar reference at the PWS location, including metrics for each of the three QC algorithms.

https://doi.org/10.5194/hess-2023-195
Preprint. Discussion started: 31 August 2023

Appendix C: Areal rainfall



(a) Areal average for event 1

(b) Areal average for event 3

Figure C1. Panel (a) shows an example of areal rainfall over the Amsterdam metropolitan area for event 1, panel (b) for event 3.