AGU100 ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER
10.1029/2019GL083731

Quality Control for Crowdsourced Personal Weather Stations to Enable Operational Rainfall Monitoring

Lotte Wilhelmina de Vos[1,2], Hidde Leijnse[1], Aart Overeem[1,2], and Remko Uijlenhoet[2]

[1]Royal Netherlands Meteorological Institute, De Bilt, Netherlands, [2]Hydrology and Quantitative Water Management Group, Wageningen University and Research, Wageningen, Netherlands

Key Points:
- A real-time applicable quality control methodology for crowdsourced personal weather stations is suggested
- The quality control successfully identifies typical errors for this data source without requiring auxiliary data
- High-resolution nationwide rainfall maps can be produced from the quality-controlled crowdsourced personal weather stations

Supporting Information:
- Supporting Information S1
- Movie S1
- Movie S2

Correspondence to:
L. W. de Vos,
lotte.devos@wur.nl

Citation:
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. https://doi.org/10.1029/2019GL083731

Received 15 MAY 2019
Accepted 14 JUL 2019
Accepted article online 22 JUL 2019
Published online 5 AUG 2019
Corrected 11 SEP 2019 and 26 JUN 2021

This article was corrected on 11 SEP 2019 and 26 JUN 2021. See the end of the full text for details.

Abstract Automatic personal weather stations owned and maintained by weather enthusiasts provide spatially dense in situ measurements that are often collected and visualized in real time on online weather platforms. While the spatial and temporal resolution of this data source is high, its rainfall observations are prone to typical errors, currently preventing its large-scale, real-time application. This study proposes a quality control methodology consisting of four modules targeting these errors, applicable in real time without requiring auxiliary measurements. The quality control improves the overall accuracy of a year of hourly rainfall depths in Amsterdam to a bias of −11.3% (0.2% when a proxy for overall rainfall underestimation by personal weather stations is used), a Pearson correlation coefficient of 0.82, and a coefficient of variation of 2.70, while maintaining 88% of the original data set. Application on a national scale (average 1 station per ∼10 km^2) yields high-resolution nationwide rainfall maps, hence showing the great potential of personal weather stations for complementing existing often sparse traditional rain gauge networks.

Plain Language Summary Rainfall measurements are needed for many applications, for example, water management and weather prediction. Especially for models describing urban drainage, the resolution of rainfall data should be high and dense networks of rain gauges are often lacking. However, many citizens own personal weather stations that share weather observations in real time on online platforms. Crowdsourcing measurements from these platforms provides rainfall information at high resolutions in both space and time, although they can contain many types of errors. We propose a quality control method that detects and filters typical errors in this data set using spatial consistency checks, requiring no additional measurements, and which is potentially applicable in real time. The method improves the accuracy of a 1-year data set of rainfall observations of all stations in the Amsterdam metropolitan area dramatically while removing only 12% of the raw measurements. Nationwide quality-controlled observations are used to successfully construct rainfall maps over the Netherlands. This shows that crowdsourced personal weather stations provide a valuable source of rainfall observations.

1. Introduction

Accurate rainfall monitoring is vital in understanding hydrological and meteorological processes. Rain gauge networks provide direct point-scale measurements. However, the combined orifices of gauges routinely used to produce global precipitation products span an area smaller than a soccer field, and most of these gauge measurements consist of daily observations (Kidd et al., 2017). This is problematic as high-resolution (in space and time) flood forecasting requires high-resolution precipitation input in order to produce meaningful results, especially in urban areas (Berne et al., 2004; Emmanuel et al., 2012; Ochoa-Rodriguez et al., 2015; Cristiano et al., 2017). Crowdsourcing has been investigated as a strategy to obtain more rainfall observations, ranging from studies exploring citizen observations collected via smartphone apps (Elmore et al., 2014; Guo et al., 2019), active daily rainfall amounts reported by volunteers (Cifelli et al., 2005; Illingworth et al., 2014; Reges et al., 2016), rainfall intensities from camera images (Allamano et al., 2015; Jiang et al., 2019), rain intensity and occurrence from car sensors (Rabiei et al., 2013), derived weather information from twitter messages (De Vasconcelos et al., 2016), to simulation studies incorporating those techniques (e.g., Mazzoleni et al., 2017; Yang & Ng, 2017). Muller et al. (2015) describe some of these and other crowdsourcing strategies to gain atmospheric data. Zheng et al. (2018) present a recent overview and state of the art of crowdsourcing data collection methods in geophysics.

Recent developments enable owners of automatic weather stations to easily monitor their environment and share weather observations in real time on online platforms. Popular online platforms such as Netatmo and Weather Underground collect and visualize measurements from personal weather stations (PWSs) every ∼5 to 10 min. The average density of Netatmo PWSs measuring rainfall in the Netherlands is 1 per ∼10 km^2, while the national networks employed by the Royal Netherlands Meteorological Institute (KNMI) consist of a manual gauge every ∼100 km^2 and an automatic gauge every ∼1,000 km^2. As PWS density is correlated with population density, this provides weather observations at high temporal and spatial resolution in urban areas particularly.

Previous studies investigated the accuracy of common PWS devices (Jenkins, 2014; Bell et al., 2015; Meier et al., 2017; De Vos et al., 2017) and made use of this data source to quantify the urban heat island effect (Meier et al., 2017; Chapman et al., 2017; Fenner et al., 2017; Golroudbary et al., 2018; Napoly et al., 2018). Rain measurement from Weather Underground PWSs are explored by De Vos et al. (2017; 63 PWSs in Amsterdam), Golroudbary et al. (2018; 11 PWSs in the Netherlands), and Chen et al. (2018; 11 PWSs in Norfolk, Virginia).

Inaccurate rain observations can be due to (1) instrumental errors, (2) compromised setup, and (3) data processing issues. Quality control (QC) methods are designed to exclude inaccurate measurements. QC can consist of comparisons with auxiliary data (e.g., Qi et al., 2016), preset (dynamic) thresholds to exclude unlikely values (e.g., Estévez et al., 2011), or internal consistency between stations and/or in time (e.g., Zahumensky, 2004; Chen et al., 2018). PWS rainfall data are arguably highly prone to errors as the typically low-cost devices are often installed without knowledge of, or access to, optimal setup locations and are not regularly maintained. Measurement accuracy can change suddenly, for example, due to hindrance of tipping bucket mechanisms by clogging or tilted setups after windy weather, which can be resolved just as suddenly. Notwithstanding their enormous potential for operational rainfall monitoring, these sources of error currently prevent the large-scale, real-time application of PWSs in meteorology and hydrology.

This paper, for the first time, explores an unprecedented large data set in terms of length (2 years), covered area, and density and shows that accurate, nationwide rainfall maps can be constructed from crowdsourced PWS rainfall measurements. For this purpose we propose a real-time applicable QC method, consisting of a set of quality filters that excludes inaccurate observations, requiring no auxiliary data source or metadata (besides station location). We show the ability of this filter to correctly flag measurement intervals with typical errors for this data source and unflag once the PWS produces reliable values again. Nationwide rainfall maps constructed from the filtered data set show remarkable similarities with a reference radar/rain gauge data set.

2. Method

2.1. PWS Data Set

Two extensive data sets of PWS rainfall observations were obtained from the Netatmo Weathermap (https://weathermap.netatmo.com/). Netatmo PWSs consist of an indoor and an outdoor module (∼160 euro) measuring temperature, relative humidity and (indoor) sound, barometric pressure, and CO_2 levels, with optional additional modules for wind and rain (∼65 euro each). The rain module is a plastic tipping bucket with a collection funnel, 13 cm in diameter, that reports the number of tips via a wireless connection of up to 100 m to the indoor module. This indoor module broadcasts all observations to the platform every ∼5 min from the moment the station becomes operational resulting in ∼5-min time series. From here measurements can be accessed via smartphone or tablet and visualized on the online Weathermap. Netatmo gives rain observations as multiples of 0.101 mm, or, in multiples of the tipping bucket volume that is determined by the weather station owner using the calibration feature of the device. Approximately 13.5% of Dutch rain gauges are manually calibrated, as shown by deviating tipping bucket volumes. Netatmo rain gauges have a measurement range of 0.2–150 mm/hr and an accuracy of 1 mm/hr according to the manufacturer specifications. De Vos et al. (2017) show that three Netatmo devices in an experimental setup with a collocated well-calibrated operational reference rain gauge measure rainfall with high accuracy when installed properly and using unrounded measurements, as in this data set.

Two PWS data sets are analyzed in this work:

- Urban data set: All PWSs with a rain module in the Amsterdam metropolitan area, defined as the area between 4.67–5.05° longitude and 52.24–52.44° latitude (∼575 km^2) between 1 May 2016 and 1 June 2018.
- National data set: All PWSs with a rain module within the Netherlands for the month May 2018.

10.1029/2019GL083731

From PWS observations we construct time series at 5 min and at hourly intervals. The first year of the urban data set is used as calibration data set (CAL) to design the QC algorithm. The QC was subsequently applied on the second year (VAL) and on the national data set (NL) to illustrate the ability of the filters to independently identify inaccurate observations. For both CAL and VAL the QC starts 1 month before the study period of 1 year to allow for the warm-up period in the filters.

2.2. Reference Data Set

The reference is a climatological data set that covers the entire land surface of the Netherlands in pixels of ~1 km^2 at 5-min temporal resolution, freely accessible on https://data.knmi.nl/datasets/rad&urluscore;nl25&urluscore;rac&urluscore;mfbs&urluscore;em&urluscore;5min/2.0. This rainfall product is based on two C-band radars, adjusted with two rain gauge networks (31 automatic and 325 manual gauges). Detailed information on radars and processing are provided by Beekhuis and Mathijssen (2018), and on the methodology by Overeem, Buishand, et al. (2009), Overeem, Holleman, et al. (2009), and Overeem et al. (2011). This radar product, adjusted with quality-controlled high-end daily and subdaily ground observations, is considered the most accurate rainfall reference but only becomes available with a delay of 1–2 months.

2.3. Validation

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

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

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

with

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

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

where R_{PWS} are rainfall time series aggregated at 5-min or hourly temporal resolutions from PWSs (mm) and R_{ref} are the corresponding reference time series in the overlying radar pixel (mm) where the bar indicates the average of observations of all stations at all time intervals.

3. QC

3.1. Types of Errors

3.1.1. Sampling and Representativeness Error

The crowdsourced rainfall time series have variable time intervals in which the number of tipping bucket tips since the last time stamp are reported. In addition to the intrinsic tipping bucket error where rain can be attributed to a later time stamp (Habib et al., 2001), additional errors result from gaps in the time series during connectivity problems. The sampling error in rain gauges with tipping bucket volumes of 0.101 mm, measuring in ~5-min intervals, has been determined in a simulation exercise with 12-s resolution electronic rain gauge observations (Leijnse et al., 2010) as basis and ground truth. This yields an r of ~0.96 and a CV of ~2.29 (see supporting information Figure 1). Using the radar product for validation introduces an additional error due to gauge-pixel discrepancy (and sources of error in radar rainfall retrieval), which reduces the similarity to an r of ~0.75 and a CV of ~5.0.

3.1.2. Bias

Unbiased PWS measurements rely on an unshielded setup where all raindrops reach the collection funnel, and a level gauge so that the tipping bucket mechanism is not hindered. However, completely exposed rain gauges are known to suffer from wind-induced underestimation (Pollock et al., 2018). Bias can also be due to the actual tipping bucket volume of the gauge not corresponding with the reported value due to manufacturing variability or faulty calibration. Netatmo PWS owners can calibrate their gauge by pouring a known amount of water through and calculating the tipping volume from the number of tips. If water is poured too quickly, some water bypasses the tipping mechanism during each tip, resulting in overestimation of the tipping volume. The majority of calibrated Netatmo PWSs in the nationwide data set has an estimated tipping

10.1029/2019GL083731

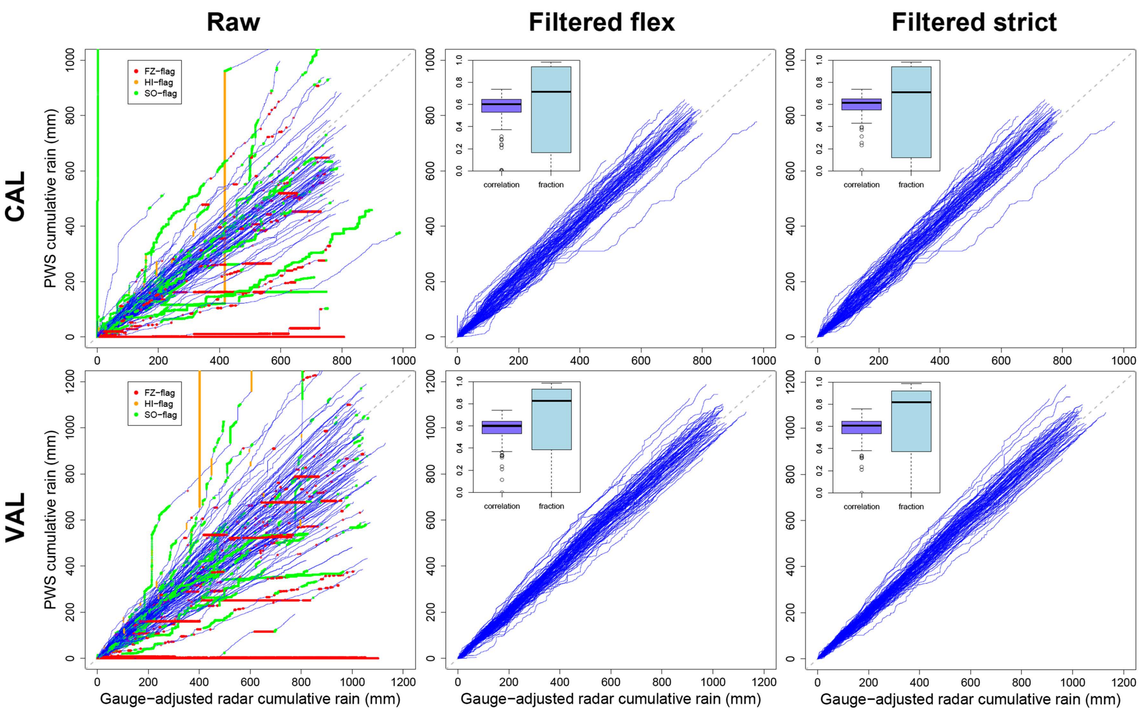

Figure 1. Double mass plots of PWS observations against their respective reference, for year 1 (CAL) and year 2 (VAL) of the Amsterdam data set, for raw data, subsets where intervals with FZ, HI, or SO flag = 1 are excluded ("Filtered Flex") and subsets where intervals with FZ, HI, or SO flag not equal to 0 are excluded ("Filtered Strict"), including box plots indicating the spread of the correlations and the fractions of the measurement intervals remaining for the stations after filtering. PWS = personal weather station; FZ = faulty zeroes ; HI = high influx; SO = station outlier.

bucket volume larger than the default (11.5% > 0.101 mm and 2.0% < 0.101 mm). While bias differs greatly between stations, the PWS network has an overall tendency to underestimate rainfall (bias of −11.1% in VAL; see Table 2).

3.1.3. Faulty Zeroes

When the tipping bucket mechanism is obstructed completely, due to tilted rain gauge or physical obstructions (leaves, insects, solid precipitation, etc.), no tip will occur. Thus, only zero amounts are communicated to the platform, even during rain events.

3.1.4. High Influx

A PWS can also report large amounts of rainfall unrelated to weather, for example, by people pouring liquids through the rain gauge for cleaning, handling of the device with tilting movements, or sprinklers in the vicinity.

3.1.5. Station Outlier

Sometimes PWS measurements do not correspond with local rainfall dynamics. This is true when the reported station location is incorrect. Also, some rare occasions have been observed where, for a period of time, rainfall is recorded in repeated daily cumulative amounts, thus resulting in far too high values.

3.2. Filter Design

The CAL data set was used to design the faulty zeroes (FZ), high influx (HI), station outlier (SO) filter, and bias correction (section 3.1). Detailed flow charts of each filter are provided in the supporting information. The general concepts are explained in this section, with the names of the 11 parameters <u>underlined</u>. Each PWS time interval receives a flag of either 0, 1, or −1 for "no error," "error," or "not enough information available to determine error," respectively.

3.2.1. FZ Filter

All stations within a range ($\underline{d}$) around a given station are selected to compute the median rainfall over the surrounding area. If fewer than $\underline{n_{stat}}$ neighboring 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 this median rainfall is larger than zero for at least $\underline{n_{int}}$ time intervals while the station itself reports zero rainfall. The FZ flag remains 1 until the station reports nonzero rainfall.

Table 1
Parameter Settings for Quality Control, in Detail Explained in Section 3.2

Filter parameter	Value
d (m)	10,000
n_{stat}	5
n_{int}	6
ϕ_{A} (mm)	0.4
ϕ_{B} (mm)	10
m_{int}	4,032
m_{rain}	100
m_{match}	200
γ	0.15
β	0.2
DBC [CAL]	1.24
DBC [VAL]	1.13
DBC [NL]	1.13

Note. The independent default bias correction (DBC) values were determined from the bias values in the urban data set in the preceding month, that is, May 2016 for CAL (June 2016 to June 2017), May 2017 for VAL (June 2017 to June 2018), and April 2018 for the NL data set (May 2018).

3.2.2. HI Filter

Unrealistically high rainfall amounts are determined based on a comparison with the median rainfall amount from all stations within a range ($\underline{d}$) around a given station. If the median does not exceed a threshold value ($\underline{\phi_{\mathrm{A}}}$), the HI flag is set to 1 for any rainfall value from the station itself above threshold $\underline{\phi_{\mathrm{B}}}$. When the surrounding stations report moderate to heavy rainfall, the threshold becomes variable: for a median of $\underline{\phi_{\mathrm{A}}}$ or higher, the stations' HI flag is set to 1 when its measurements exceed median times $\underline{\phi_{\mathrm{B}}}/\underline{\phi_{\mathrm{A}}}$. HI flag is set to −1 if fewer than $\underline{n_{\mathrm{stat}}}$ neighboring stations report observations.

3.2.3. Bias Correction and SO Filter

First, a default bias correction factor ($\underline{\mathrm{DBC}}$) is determined to address the fact that the Netatmo rain gauges have a general tendency to underestimate rainfall. $\underline{\mathrm{DBC}}$ is a single-value one-off proxy of the correction needed for the overall PWS network bias and can be determined a priori by comparing network measurements over a period with typical rainfall for the local climate. In this study we base it on the median of all PWS bias values of 5-min observations during the month preceding the start of the data set, when compared with the gauge-adjusted radar values, excluding intervals where FZ or HZ were not 0. $\underline{\mathrm{DBC}}$ becomes

$$\underline{\mathrm{DBC}} = \frac{1}{1 + \mathrm{median(bias)}} \tag{5}$$

These $\underline{\mathrm{DBC}}$s (Table 1) correspond with the negative bias that De Vos et al. (2017, Figure 7) find for Wunderground PWS rain data in the same area. Note that although we use another data set for this step, it occurs in an offline exercise, meaning no auxiliary data are needed for the bias correction in the operational, real-time QC. If no reference is available, $\underline{\mathrm{DBC}}$ could be set to 1 or to a value that was found elsewhere for this particular instrument in a similar climate. Initially, the bias correction factor (BCF) for each PWS is equal to $\underline{\mathrm{DBC}}$. In order to test the performance without the availability of any reference data, $\underline{\mathrm{DBC}} = 1$ is also applied.

To determine whether a station yields nonsensical measurements for that location, it is compared with time series of neighboring stations within a range ($\underline{d}$). A previous period of $\underline{m_{\mathrm{int}}}$ intervals, or any longer interval where the station has at least $\underline{m_{\mathrm{rain}}}$ intervals of nonzero rainfall measurements, is evaluated. There needs to be at least $\underline{n_{\mathrm{stat}}}$ stations with at least $\underline{m_{\mathrm{match}}}$ intervals overlapping with the evaluated station to compute the SO flag. The r (equation (1)) and bias (equation (2)) with all neighboring stations are calculated. If the median of the r values falls short of threshold $\underline{\gamma}$, the SO flag is set to 1. If this threshold is exceeded, $\mathrm{BCF_{new}}$ is computed from the median of the bias values with the neighboring stations. If $|\log(\mathrm{BCF_{new}}/\mathrm{BCF_{prev}})| > \log(1+\underline{\beta})$, this is deemed a systematic change for that station and $\mathrm{BCF_{prev}}$ is replaced with the new value. This is hence a way to dynamically update BCF for individual stations.

10.1029/2019GL083731

Table 2
Validation Metrics and Remaining Fraction of Original Observations of 5-min and Hourly Personal Weather Station 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), Also When Considering Only the Subset Where Reference Exceeds a Threshold of 0.1 and 0.5 mm for 5-min and 1-hr Values, Respectively

Time interval	Data set	Filter type	bias	CV	*r*	Remaining
5 min	CAL	Raw	1.39	147.08	0.04	100%
		FZ-filtered	1.545	153.12	0.04	95.6%
		HI-filtered	−0.061	12.95	0.35	99.9 … %
		SO-filtered	0.003	18.1	0.27	88.3%
		bias-corrected	0.056	51.54	0.11	100%
		All filters: Flex	0.059	8.97	0.58	89.0%
		All filters: Strict	0.057	8.83	0.59	87.2%
	VAL	Raw	−0.111	53.24	0.07	100%
		FZ-filtered	−0.044	55.46	0.08	94.0%
		HI-filtered	−0.133	7.57	0.50	99.9 … %
		SO-filtered	−0.076	55.86	0.08	89.8%
		bias-corrected	−0.030	17.81	0.24	100%
		All filters: Flex	0.021	7.22	0.58	89.2%
		All filters: Strict	0.023	7.19	0.58	88.0%
		All filters: Strict; DBC = 1	−0.095	6.68	0.58	88.0%
	VAL, Ref > 0.1mm	Raw	−0.372	1.26	0.45	100%
		FZ-filtered	−0.324	1.26	0.46	92.9%
		HI-filtered	−0.373	1.25	0.45	99.9 … %
		SO-filtered	−0.329	1.24	0.47	90.4%
		bias-corrected	−0.299	1.35	0.45	100%
		All filters: Flex	−0.236	1.32	0.48	88.9%
		All filters: Strict	−0.234	1.32	0.48	88.0%
1 hr	CAL	Raw	1.302	144.37	0.03	100%
		FZ-filtered	1.475	152.03	0.03	95.6%
		HI-filtered	−0.107	9.29	0.38	99.9 … %
		SO-filtered	−0.024	15.91	0.25	88.5%
		bias-corrected	0.035	16.59	0.27	100%
		All filters: Flex	0.045	3.75	0.81	89.1%
		All filters: Strict	0.043	3.62	0.82	87.3%
	VAL	Raw	−0.167	3.74	0.68	100%
		FZ-filtered	−0.099	3.67	0.71	94.0%
		HI-filtered	−0.168	3.69	0.69	99.9 … %
		SO-filtered	−0.131	3.25	0.75	89.9%
		bias-corrected	−0.064	3.50	0.74	100%
		All filters: Flex	0.001	2.88	0.82	89.3%
		All filters: Strict	0.002	2.86	0.82	88.1%
		All filters: Strict; DBC = 1	−0.113	2.70	0.82	88.1%
	VAL, Ref > 0.5mm	Raw	−0.291	0.79	0.67	100%
		FZ-filtered	−0.231	0.74	0.71	92.4%
		HI-filtered	−0.291	0.79	0.67	99.9 … %
		SO-filtered	−0.246	0.73	0.71	90.2%
		bias-corrected	−0.199	0.83	0.68	100%
		All filters: Flex	−0.126	0.74	0.74	88.3%
		All filters: Strict	−0.125	0.73	0.74	87.5%

Note. Metrics of VAL filtered (Strict) results are included where bias correction factors are calculated with DBC = 1. DBC = default bias correction factor; FZ = faulty zeroes ; HI = high influx; SO = station outlier.

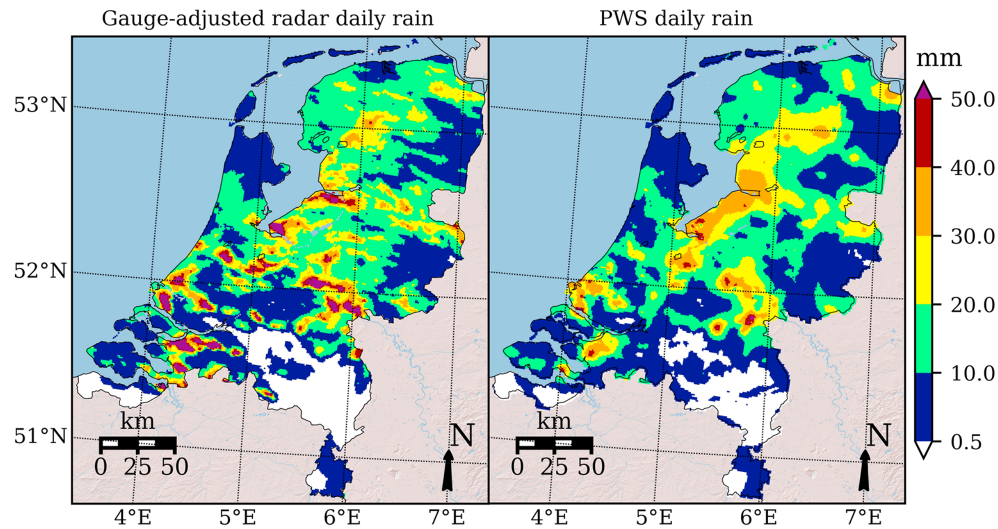

Figure 2. Daily rainfall accumulations between 29 May 2018 08:00 UTC and 30 May 2018 08:00 UTC according to the gauge-adjusted radar product (left) and Flex filtered PWSs (right), interpolated using Ordinary Kriging with fitted variograms and nugget set to zero. Only stations with at least 95% data availability after quality control during that day are included. PWS = personal weather station.

3.2.4. Parameter Choices

The chosen values for the 11 parameters are given in Table 1. Several sets of parameters were evaluated, and the best one was chosen based on the achieved improvement (see section 2.3) for the CAL data set after QC, while aiming for large applicability (i.e. fraction of flags = −1 small) and without flagging abundantly (i.e., fraction of flags = 1 small).

The QC principle applies to gauge networks in general, although the parameter values should be considered carefully for each network separately. For a sparser network, a larger $\underline{d}$ parameter and/or lower $\underline{n_{stat}}$ are needed to select enough neighbor stations. The number of values used to construct the median can be limited if $\underline{n_{stat}}$ is small, possibly resulting in outlier values. Higher values for $\underline{n_{int}}$, $\underline{m_{int}}$, $\underline{m_{rain}}$, and $\underline{m_{match}}$ result in more robust subsets of data on which flags and bias corrections are determined, at the cost of a longer unflagged warm-up period and more cases in which flags cannot be attributed. Most rainfall observations that should be targeted by the HI filter were found to be very high; thus, small variations in $\underline{\phi_A}$ and $\underline{\phi_B}$ hardly affect the results. Higher $\underline{\gamma}$ yields more SO flags and lower $\underline{\beta}$ results in more frequent BCF adjustments (and possibly overfitting).

4. Filter Performance

When cumulative rainfall over a full year as measured by the PWSs is plotted against the cumulative amount according to the reference, FZ, HI, and SO errors are shown as horizontal line segments, vertical line segments, and fluctuating lines deviating from the diagonal, respectively. Figure 1 shows that the QC attributes flags to the time intervals causing these horizontal, vertical and fluctuating line segments. Stations can be susceptible to bias, seen in Figure 1 as lines with slopes differing from the gray diagonal line.

The data set can be filtered in two ways: retaining only intervals where no flag is 1 (“Flex”), or, retaining only intervals where all flags are 0 (“Strict”). After QC is applied, which includes station-specific bias correction, the remaining measurements correspond far better with the reference, that is, the lines resemble the gray diagonal, especially for the Strict version. The difference between the Strict and Flex versions of the filter performance is very small for the Amsterdam metropolitan area, where the network is dense. In sparser networks it is likely that the filter criteria are met less often, resulting in a larger difference between Strict and Flex filter results. The 87.2% and 88.0% of all intervals of the first and second years of the urban data set (CAL and VAL, respectively) without any error flag show a dramatic improvement in accuracy regarding bias, CV, and r (Table 2). Each filter yields accuracy improvement, with the largest effect in the HI filter given the small number of flagged intervals (Table 2). The comparison of the 5-min Strict-filtered VAL data

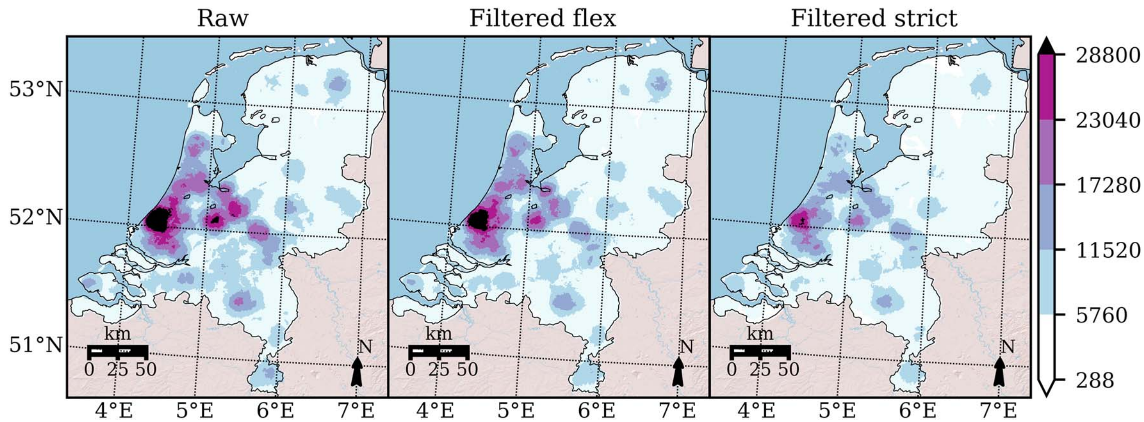

Figure 3. Density maps of the number of 5-min station measurements within 10 km/day, calculated between 29 May 2018 08:00 UTC and 30 May 2018 08:00 UTC, for all measurements ("Raw"), excluding intervals with FZ, HI, or SO flag = 1 ("Filtered Flex") and excluding intervals where FZ, HI, and SO flags are not equal to 0 ("Filtered Strict"). FZ = faulty zeroes ; HI = high influx; SO = station outlier.

set with the gauge-adjusted radar yields huge improvements as compared to the raw data metrics, with a bias from −0.111 to 0.023, a CV from 53.24 to 7.19, and an r from 0.07 to 0.58, thus more closely resembling the upper limit of accuracy of rainfall data sampled in this manner of CV of 5.0 and r of 0.75 (see supporting information).

The filters can be applied on a national scale. Rainfall patterns found by the PWSs correspond well with those from gauge-adjusted radar (Figure 2, and movies in the supporting information). As the filters rely on neighbor checks, the QC is best applicable in the urban areas in the west where the PWS network is densest (Figure 3).

5. Discussion and Conclusions

This study proposes a real-time applicable QC algorithm that does not require auxiliary or metadata. Rather than stations, time intervals are identified and flagged for errors related to faulty zero observations, high influxes, and station outliers. Additionally, dynamic bias correction is performed. The QC was designed on a full year of PWS measurements in the Amsterdam metropolitan area, and applied on measurements in the same study area during the subsequent year, as well as on a month of all PWS measurements covering the Netherlands. Results show large improvements of filtered data over raw measurements in the calibration, and even more in the validation data set (bias of 0.023, CV 7.19 and r of 0.58, while retaining 88% of the original data set at 5-min resolution), likely due to the higher data availability in the second year.

The QC is successful in flagging observations that are inaccurate, although as it relies on neighbor comparison it is better applicable on the urban data set in Amsterdam than for other areas in the Netherlands with fewer PWSs, where the difference between filtered Flex and filtered Strict will be larger than indicated in Figure 1 due to more frequent allocations of −1 flags. Also, it inherently assumes that nearby PWSs measure the same rainfall dynamics, which may be true in a relatively flat country like the Netherlands, but is far less likely in areas with systematic spatial rainfall gradients, e.g. due to mountains. This could be addressed by introducing elevation dependent neighbor selection, which only selects neighbor PWSs at similar elevations or orientations with respect to mountain ranges. This would require additional meta-data.

Depending on the requirements of the resulting rainfall data set, one may choose to make the QC more selective (at the expense of observation density) by decreasing $\underline{d}$, increasing $\underline{n_{stat}}$, $\underline{n_{int}}$, $\underline{m_{int}}$, $\underline{m_{rain}}$, and $\underline{m_{match}}$, or more inclusive the other way around (at the expense of accuracy). These parameters are related to the spatial and temporal scales of rainfall events, and should therefore correspond to typical rainfall variability in the local climate.

Although the QC does not need auxiliary data operationally, the $\underline{\mathrm{DBC}}$ parameter was determined offline, based on bias between PWSs and gauge-adjusted radar reference. Ideally, $\underline{\mathrm{DBC}}$ should be redetermined with reliable local rainfall information approximately annually to address changes in overall network bias (due to loss and additions of PWSs, and accuracy changes over lifetime). If the bias correction module is implemented without this proxy ($\underline{DBC} = 1$) the QC results in rainfall estimates that are as good as with a calibrated value of $\underline{DBC}$, except, of course, for the overall bias (see Table 2).

In this study the filter was applied on crowdsourced PWS rainfall observations every 5 min, although the QC (with adjusted settings) will also be applicable at other time scales, resulting in time series with that time interval. The QC targets the errors that are typical for crowdsourced PWS networks measuring at variable time intervals. However, it can be applied successfully on any gauge networks with active periodic measurements. Sparser networks may be processed employing a longer comparison range, thus at longer time intervals. The next phase for this work is to make raw data, collected on platforms maintained by commercial organizations, accessible in real time for PWS networks in order for them to become a viable data source for rainfall monitoring. If this issue is addressed, a huge number of in situ rainfall observations available in real time (mostly in developed regions of the world) can be used for various (operational) purposes, for example, for PWS-adjusted radar products that, contrary to our validation data set, can be available in (near) real time.

Acknowledgments
This work was financially supported by KNMI MSO Project 2013.09, cofunded by the Amsterdam Institute for Advanced Metropolitan Solutions (AMS; Project VIR17006) and Het Waterschaphuis (project Onderzoek Neerslagmetingen). The authors thank all PWS owners who contribute to the Netatmo Weathermap platform, as well as the Netatmo company for making the data set available to us. The PWS data sets are freely available via 4TU: doi.org/10.4121/uuid:6e6a9788-49fc-4635-a43d-a2fa164d37ec. See also https://dev.netatmo.com/resources/technical/reference/weatherapi/getpublicdata for information on how to obtain Netatmo data. The gauge-adjusted radar rainfall data set is freely available in netCDF4 format via http://climate4impact.eu as "Radar precipitation climatology," or in HDF5 format via the KNMI Data Centre (https://data.knmi.nl/datasets/rad&urluscore;nl25&urluscore;rac&urluscore;mfbs&urluscore;5min/2.0?q=radar). The authors acknowledge the fruitful discussions with Rinske Hutten, Marie-claire ten Veldhuis, and Marc Schleiss of TU Delft while collaborating on a project focusing on faulty zero detection in PWS data.

References

Allamano, P., Croci, A., & Laio, F. (2015). Toward the camera rain gauge. *Water Resources Research*, *51*, 1744–1757. https://doi.org/10.1002/2014WR016298

Beekhuis, H., & Mathijssen, T. (2018). From pulse to product, highlights of the upgrade project of the Dutch national weather radar network, Extended abstract for ERAD2018 conference (pp. 960–965).

Bell, S., Cornford, D., & Bastin, L. (2015). How good are citizen weather stations? Addressing a biased opinion. *Weather*, *70*(3), 75–84.

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

Chapman, L., Bell, C., & Bell, S. (2017). Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations. *International Journal of Climatology*, *37*(9), 3597–3605.

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*, ACM (pp. 25–28).

Cifelli, R., Doesken, N., Kennedy, P., Carey, L. D., Rutledge, S. A., Gimmestad, C., & Depue, T. (2005). The community collaborative rain, hail, and snow network: Informal education for scientists and citizens. *Bulletin of the American Meteorological Society*, *86*(8), 1069–1078.

Cristiano, E., ten Veldhuis, M.-C., & van de Giesen, N. (2017). Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas-A review. *Hydrology and Earth System Sciences*, *21*(7), 3859–3878.

De Vasconcelos, L. E. G., dos Santos, E. C. M., Neto, M. L. F., Ferreira, N. J., & de Vasconcelos, L. G. (2016). Using tweets for rainfall monitoring, *Information technology: New generations* (Vol. 448, pp. 1157–1167). Cham, Switzerland: Springer International Publishing.

De Vos, L. W., 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*(2), 765–777.

Elmore, K. L., Flamig, Z., Lakshmanan, V., Kaney, B., Farmer, V., Reeves, H. D., & Rothfusz, L. P. (2014). MPING: Crowd-sourcing weather reports for research. *Bulletin of the American Meteorological Society*, *95*(9), 1335–1342.

Emmanuel, I., Andrieu, H., Leblois, E., & Flahaut, B. (2012). Temporal and spatial variability of rainfall at the urban hydrological scale. *Journal of Hydrology*, *430-431*, 162–172.

Estévez, J., Gavilán, P., & Giráldez, J. V. (2011). Guidelines on validation procedures for meteorological data from automatic weather stations. *Journal of Hydrology*, *402*(1-2), 144–154.

Fenner, D., Meier, F., Bechtel, B., Otto, M., & Scherer, D. (2017). Intra and inter local climate zone variability of air temperature as observed by crowdsourced citizen weather stations in Berlin, Germany. *Meteorologische Zeitschrift*, *26*(5), 525–547.

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*(2), 95–109.

Guo, H., Huang, H., Sun, Y.-E., Zhang, Y., Chen, S., & Huang, L. (2019). Chaac: Real-time and fine-grained rain detection and measurement using smartphones. *IEEE Internet of Things Journal*, *6*(1), 997–1009.

Habib, E., Krajewski, W. F., & Kruger, A. (2001). Sampling errors of tipping-bucket rain gauge measurements. *Journal of Hydrologic Engineering*, *6*(2), 159–166.

Illingworth, S. M., Muller, C. L., Graves, R., & Chapman, L. (2014). UK citizen rainfall network: A pilot study. *Weather*, *69*(8), 203–207.

Jenkins, G. (2014). A comparison between two types of widely used weather stations. *Weather*, *69*(4), 105–110.

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

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, how much of the Earth's surface is covered by rain gauges? *Bulletin of the American Meteorological Society*, *98*(1), 69–78.

Leijnse, H., Uijlenhoet, R., Van De Beek, C. Z., Overeem, A., Otto, T., Unal, C. M. H., et al. (2010). Precipitation measurement at CESAR, the Netherlands. *Journal of Hydrometeorology*, *11*(6), 1322–1329.

Mazzoleni, M., Verlaan, M., Alfonso, L., Monego, M., Norbiato, D., Ferri, M., & Solomatine, D. P. (2017). Can assimilation of crowdsourced data in hydrological modelling improve flood prediction? *Hydrology and Earth System Sciences*, *21*(2), 839–861.

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.

Muller, C. L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S. M., Foody, G., et al. (2015). Crowdsourcing for climate and atmospheric sciences: Current status and future potential. *International Journal of Climatology*, *35*(11), 3185–3203.

Napoly, A., Meier, F., Grassmann, T., & Fenner, D. (2018). Development and application of a statistically-based quality control for crowdsourced air temperature data. *Frontiers in Earth Science*, *6*, 118.

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

Overeem, A., Buishand, T., & Holleman, I. (2009). Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. *Water Resources Research*, *45*, W10424. https://doi.org/10.1029/2009WR007869

Overeem, A., Holleman, I., & Buishand, A. (2009). Derivation of a 10-year radar-based climatology of rainfall. *Journal of Applied Meteorology and Climatology*, *48*(7), 1448–1463.

Overeem, A., Leijnse, H., & Uijlenhoet, R. (2011). Measuring urban rainfall using microwave links from commercial cellular communication networks. *Water Resources Research*, *47*, W12505. https://doi.org/10.1029/2010WR010350

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

Qi, Y., Martinaitis, S., Zhang, J., & Cocks, S. (2016). A real-time automated quality control of hourly rain gauge data based on multiple sensors in MRMS system. *Journal of Hydrometeorology*, *17*(6), 1675–1691.

Rabiei, E., Haberlandt, U., Sester, M., & Fitzner, D. (2013). Rainfall estimation using moving cars as rain gauges - laboratory experiments. *Hydrology and Earth System Sciences*, *17*(11), 4701–4712.

Reges, H. W., Doesken, N., Turner, J., Newman, N., Bergantino, A., & Schwalbe, Z. (2016). COCORAHS: The evolution and accomplishments of a volunteer rain gauge network. *Bulletin of the American Meteorological Society*, *97*(10), 1831–1846.

Yang, P., & Ng, T. L. (2017). Gauging through the crowd: A crowd-sourcing approach to urban rainfall measurement and storm water modeling implications. *Water Resources Research*, *53*, 9462–9478. https://doi.org/10.1002/2017WR020682

Zahumensky, I. (2004). Guidelines on quality control procedures for data from automatic weather stations. *World Meteorological Organization, Switzerland*, *WMO-No 955*.

Zheng, F., Tao, R., Maier, H. R., See, L., Savic, D., Zhang, T., et al. (2018). Crowdsourcing methods for data collection in geophysics: State of the art, issues, and future directions. *Reviews of Geophysics*, *56*, 698–740. https://doi.org/10.1029/2018RG000616

Erratum

In the originally published version of this article, two supporting information files were omitted. There was also an error in the link to a data set associated with the paper. In addition, the affiliation listed for coauthor Hidde Leijnse was incorrect. Hidde Leijnse's correct affiliation is Royal Netherlands Meteorological Institute, De Bilt, Netherlands. Subsequently, a bug was found in the openly available code (https://github.com/LottedeVos/PWSQC) that was used to visualize the results. The "Filtered flex" double mass plots for the Amsterdam metropolitan area are actually more similar to the "Filtered strict" plots in Figure 1. The open code, Figure 1, and some values in Table 2 have been corrected, and some nuances in the text discussing the differences between the settings "Flex" and "Strict" have been included. The underlying method to filter the rainfall observations is unaffected, as are the conclusions drawn in the original publication, and the present version may be considered the authoritative version of record.