Crowdsourced personal weather stations show great potential for operational rainfall monitoring

L.W. de Vos [1,2], H. Leijnse [2], A. Overeem [1,2], R. Uijlenhoet [2]

[1]Royal Netherlands Meteorological Institute, de Bilt, Netherlands
[2]Wageningen University & 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 nation-wide rainfall maps can be produced from the quality controlled crowdsourced personal weather stations

Corresponding author: L.W. de Vos, `lotte.devos@wur.nl`

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 prohibiting 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 0.5%, a Pearson correlation coefficient of 0.82 and a coefficient of variation of 2.87, while maintaining 88% of the original dataset. Application on a national scale (average 1 station per $\sim$10 km^2) yields high-resolution nation-wide rainfall maps, hence showing the great potential of personal weather stations for complementing existing often sparse traditional rain gauge networks.

1 Plain language summary

Rainfall measurements are needed for many applications, e.g. 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 dataset using spatial consistency checks, requiring no additional measurements, and potentially applicable in real time. The method improves the accuracy of a 1-year dataset of rainfall observations of all stations in the Amsterdam metropolitan area dramatically while removing only 12% of the raw measurements. Nation-wide 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.

2 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, Delrieu, Creutin, & Obled, 2004; Cristiano, ten Veldhuis, & van de Giesen, 2017; Emmanuel, Andrieu, Leblois, & Flahaut, 2012; Ochoa-Rodriguez et al., 2015). 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., 2016), active daily rainfall amounts reported by volunteers (Cifelli et al., 2005; Illingworth, Muller, Graves, & Chapman, 2014; Reges et al., 2016), rainfall intensities from camera images (Allamano, Croci, & Laio, 2015), rain intensity and occurrence from car sensors (Rabiei, Haberlandt, Sester, & Fitzner, 2013), derived weather information from twitter messages (Butgereit, 2014) 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 strategies to gain atmospheric data with crowdsourcing. 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 minutes. 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 ∼1000 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 (Bell, Cornford, & Bastin, 2015; De Vos, Leijnse, Overeem, & Uijlenhoet, 2017; Jenkins, 2014; Meier et al., 2015) and made use of this data source to quantify the urban heat island effect (Chapman, Bell, & Bell, 2017; Fenner, Meier, Bechtel, Otto, & Scherer, 2017; Golroudbary, Zeng, Mannaerts, & Su, 2018; Meier, Fenner, Grassmann, Otto, & Scherer, 2017; Napoly, Meier, Grassmann, & Fenner, 2018). Preliminary work on wind measurements from Netatmo and Weather Underground was performed (Droste, Heusinkveld, & Steeneveld, 2018), and 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, Behl, and Goodall (2018) (11 PWSs in Norfolk, Virginia).

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

This paper, for the first time, explores an unprecedented large dataset in terms of length (2 years), covered area and density, and shows that accurate, nation-wide 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 faulty 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 dataset show remarkable similarities with a reference radar/rain gauge dataset.

3 Method

3.1 Dataset

This study explores two extensive datasets of PWS rainfall observations obtained from the Netatmo Weathermap (`https://weathermap.netatmo.com/`). Netatmo PWSs are relatively low-cost and consist of an indoor and an outdoor module measuring temperature, relative humidity and (indoor) sound, barometric pressure and CO_2 levels, with optional additional modules for wind and rain. 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 $\sim$5 min from the moment the station becomes operational. From here they 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 weather station owners in the Netherlands calibrate their rain gauge. Netatmo rain gauges have a measurement range of 0.2 – 150 mm and an accuracy of 1 mm h^{-1} according to the manufacturer specifications. De Vos et al. (2017) show that 3 Netatmo devices in an experimental set-up with a collocated well-calibrated operational reference rain gauge measure rainfall with high accuracy when installed properly and using unrounded measurements, as in this dataset.

Two PWS datasets are analyzed in this work:

- 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 ($\sim$575 km^2) between 1 May 2016 – 1 June 2018.
- All PWSs with a rain module within the Netherlands for the month May 2018.

The first year of the urban dataset is used as calibration dataset (CAL) to design the QC algorithm. It was subsequently applied on the second year (VAL) and on the national dataset (NL) to illustrate the ability of the filters to independently identify faulty observations. For both CAL and VAL the QC starts one month before the study period of one year to allow for the warm-up period in the filters.

The reference is a climatological dataset that covers the entire land surface of the Netherlands in pixels of $\sim$1 km^2 at 5 min temporal resolution, freely accessible on `https://data.knmi.nl/datasets/rad_nl25_rac_mfbs_em_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, and Holleman (2009); Overeem, Holleman, and Buishand (2009); Overeem, Leijnse, and Uijlenhoet (2011).

3.2 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 observations by PWSs (mm) and R_{ref} are the corresponding reference rainfall observations (mm).

4 Quality control

4.1 Types of errors

Sampling and representativeness error

The crowdsourced rainfall time series have variable time intervals in which the number of tipping bucket tips since the last timestamp are reported. In addition to the intrinsic tipping bucket error where rain can be attributed to a later timestamp (Habib, Krajewski, & Kruger, 2001), additional errors result from gaps in the time series during connectivity problems. Errors in observations of tipping bucket rain gauges with tipping bucket volumes of 0.101 mm, update frequencies of ∼5 min, interpolated at a 5-min temporal resolution, have been quantified with simulated time series that yield 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, which reduces the similarity to an r of ∼0.75 and a CV of ∼5.0.

Bias

Unbiased PWS measurements rely on an unshielded set-up where all raindrops reach the collection funnel, and a level gauge so that the tipping bucket mechanism is not hindered. However, completely unshielded 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 nation-wide dataset has an estimated tipping bucket volume larger than the default (11.5% $>$ 0.101 mm and 2.0% $<$ 0.101 mm).

Faulty zeroes (FZ)

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

High influx (HI)

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

Station outlier (SO)

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.

4.2 Filter design

The CAL dataset was used to design the FZ-, HI-, SO-filter and bias correction (Section 4.1). 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 $\underline{\text{underlined}}$.

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 $\underline{n_{\text{stat}}}$ 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 $\underline{n_{\text{int}}}$ time intervals while the station itself reports zero rainfall. The FZ-flag remains 1 until the station reports nonzero rainfall.

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 ($\underline{\phi_{\text{A}}}$), the HI-flag is set to 1 for any rainfall value from the station itself above threshold $\underline{\phi_{\text{B}}}$. When the surrounding stations report moderate to heavy rainfall, the threshold becomes variable: for a median of $\underline{\phi_{\text{A}}}$ or higher, the stations' HI-flag is set to 1 when its measurements exceed median times $\underline{\phi_{\text{B}}}/\underline{\phi_{\text{A}}}$. HI-flag is set to -1 if fewer than $\underline{n_{\text{stat}}}$ neighboring stations report observations.

Bias correction & SO-filter

First, a default bias correction factor ($\underline{\text{DBC}}$) is determined to address the fact that the Netatmo rain gauges have a general tendency to underestimate rainfall. This value should compensate the average bias in the PWS network and can be determined in various ways. In the current study it was determined using auxiliary data, namely by calculating the median of all bias values of 5-min observations in the urban PWSs with respect to the radar rainfall product over the month preceding the start of the dataset. Intervals without measurements or where the FZ or HI flag are not 0 are excluded. $\underline{\text{DBC}}$

is then calculated as:

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

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 need 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 (Eq. 1) and bias (Eq. 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, a new bias correction factor is computed from the median of the bias values with the neighboring stations. If the new bias correction factor deviates more than $\underline{\beta}$ from the previous bias correction factor, this is deemed a systematic change for that station and the bias correction factor is replaced with the new value. This is hence a way to dynamically update bias correction factors for individual stations. Each individual station starts out with a bias correction factor of $\underline{\mathrm{DBC}}$.

4.2.1 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 3.2) for the CAL dataset 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 $\underline{n_{\mathrm{stat}}}$ are needed to select enough neighbor stations. The number of values used to construct the median can be limited if $\underline{n_{\mathrm{stat}}}$ is small, possibly resulting in outlier values. Higher values for $\underline{n_{\mathrm{int}}}$, $\underline{m_{\mathrm{int}}}$, $\underline{m_{\mathrm{rain}}}$ and $\underline{m_{\mathrm{match}}}$ 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 $\underline{\phi_{\mathrm{A}}}$ and $\underline{\phi_{\mathrm{B}}}$ hardly affect the results. Higher $\underline{\gamma}$

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]	1.247228
DBC [VAL]	1.176471
DBC [NL]	1.13379

Table 1. Parameter settings for QC, in detail explained in Section 4.2. The independent default bias correction (DBC) values were determined from the bias values in the urban dataset in the preceding month, i.e. May 2016 for CAL (June 2016 - June 2017), May 2017 for VAL (June 2017 - June 2018) and April 2018 for the NL dataset (May 2018).

yields more SO-flags and lower $\underline{\beta}$ results in more frequent bias correction factor adjustments (and possibly overfitting).

5 Filter performance

When cumulative rainfall over a full year as measured by the PWS 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 dataset 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, i.e. the lines resemble the gray diagonal. The 87.2% and 88.0% of all intervals of the first and second year of the urban dataset (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 dataset with the gauge-adjusted radar yields huge improvements as compared to the raw data metrics, with a bias from -0.111 to 0.026, 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 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).

6 Discussion & 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 mea-

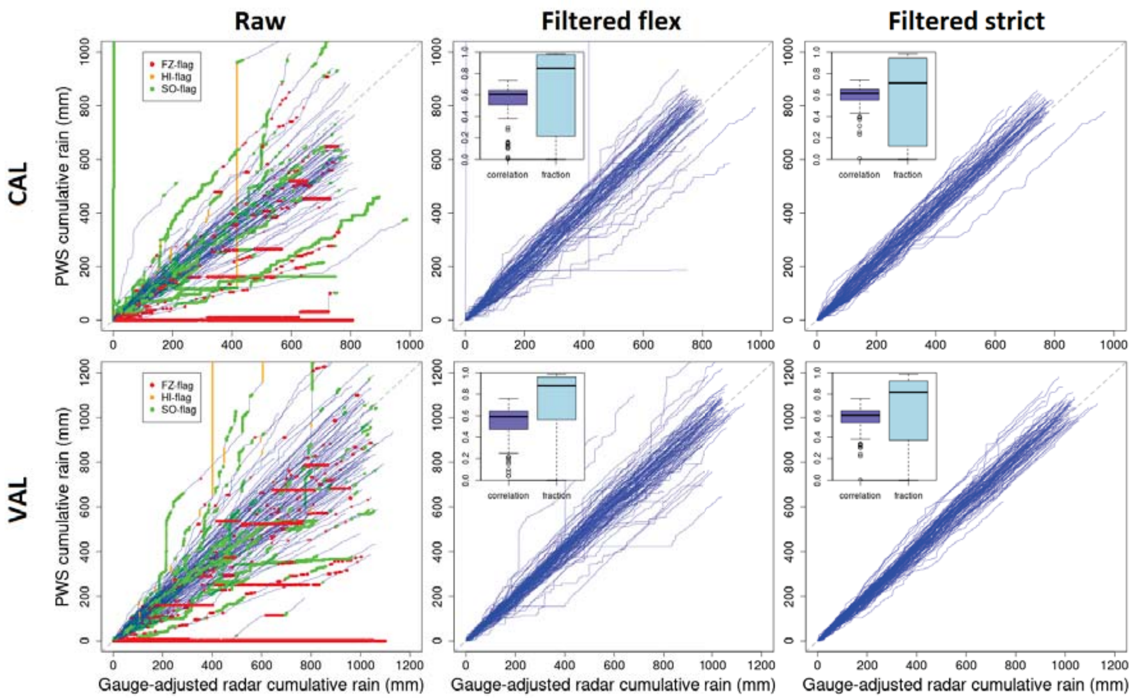


Figure 1. Double mass plots of PWS observations against their respective reference, for year 1 (CAL) and year 2 (VAL) of the Amsterdam dataset, for raw data, subsets where intervals with FZ, HI or SO flag = 1 are excluded ("Filtered Flex") and subsets where intervals with FZ, HI or SO flag not equal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

Time interval	Dataset	Filter type	bias	CV	r	Remaining
		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... %
	CAL	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 %
		Raw	-0.111	53.24	0.07	100 %
		FZ-filtered	-0.044	55.46	0.08	94.0 %
5 min		HI-filtered	-0.133	7.57	0.5	99.9... %
	VAL	SO-filtered	-0.076	55.86	0.08	89.8 %
		bias-corrected	-0.028	13.4	0.32	100 %
		All filters - Flex	0.025	7.23	0.58	89.2 %
		All filters - Strict	0.026	7.19	0.58	88.0 %
		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... %
	VAL, Ref > 0.1mm	SO-filtered	-0.329	1.24	0.47	90.4 %
		bias-corrected	-0.297	1.35	0.45	100 %
		All filters - Flex	-0.234	1.32	0.48	88.9 %
		All filters - Strict	-0.233	1.32	0.49	88.0 %
		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... %
	CAL	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 %
		Raw	-0.167	3.74	0.68	100 %
		FZ-filtered	-0.099	3.67	0.71	94.0 %
1 hour		HI-filtered	-0.168	3.69	0.69	99.9... %
	VAL	SO-filtered	-0.131	3.25	0.75	89.9 %
		bias-corrected	-0.06	3.55	0.73	100 %
		All filters - Flex	0.003	2.89	0.82	89.3 %
		All filters - Strict	0.005	2.87	0.82	88.1 %
		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... %
	VAL, Ref > 0.5mm	SO-filtered	-0.246	0.73	0.71	90.2 %
		bias-corrected	-0.197	0.83	0.68	100 %
		All filters - Flex	-0.124	0.74	0.74	88.3 %
		All filters - Strict	-0.123	0.73	0.75	87.5 %

Table 2. Validation metrics and remaining fraction of original observations of 5 min and hourly PWS time series, before (Raw) and after QC of the individual filters (in Strict manner) as well as all combined filters applied (in both Flex & Strict manner), also when considering only the subset where reference exceeds a threshold of 0.1 mm and 0.5 mm for 5-min and 1-hour values respectively.

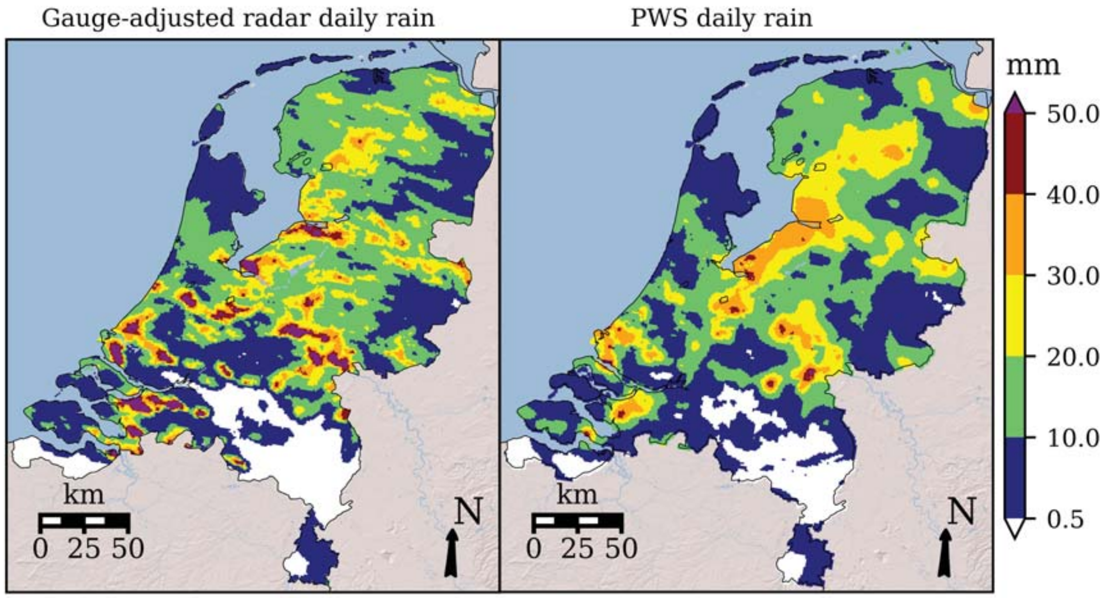


Figure 2. Daily rainfall accumulations between 2018-05-29 08:00 UTC and 2018-05-30 08:00 UTC according to the gauge-adjusted radar product (left) and Flex filtered PWSs (right), interpolated using Ordinary Kriging with fitted variograms and nugget set to zero. Only stations with at least 95% data availability after QC during that day are included.

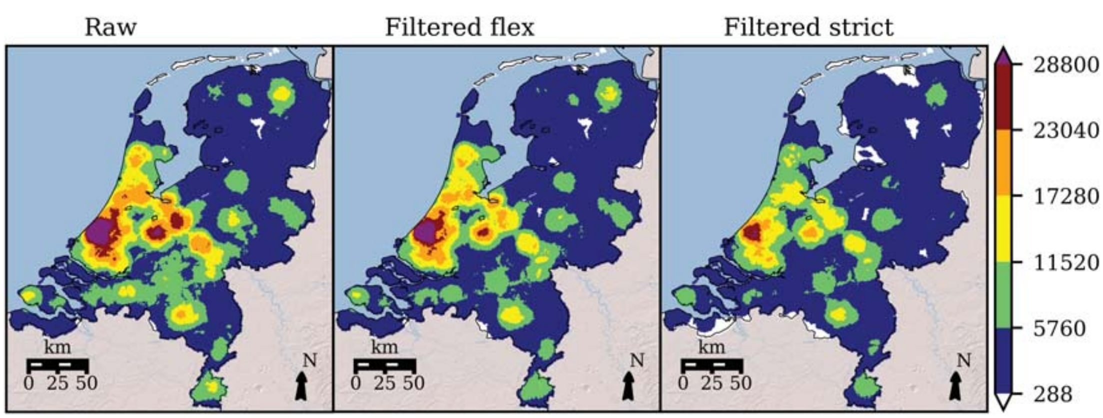


Figure 3. Density maps of the number of 5 min station measurements within 10 km per day, calculated between 2018-05-29 08:00 UTC and 2018-05-30 08:00 UTC, for all measurements ("Raw"), excluding intervals with FZ, HI or SO flag = 1 ("Filtered Flex") and excluding intervals where FZ, HI and SO flags are not equal to 0 ("Filtered Strict").

surements 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 dataset (bias of 0.026, CV 7.19 and r of 0.58, while retaining 88% of the original dataset at 5-min resolution), likely due to the higher data availability in the second year.

The QC is successful in flagging observations that are faulty, although as it relies on neighbor comparison it is better applicable on the urban dataset in Amsterdam than for other areas in the Netherlands with fewer PWSs. Depending on the requirements of the resulting rainfall dataset, one may choose to make the QC more selective (at the expense of observation density) by decreasing d, increasing n_{stat}, n_{int}, m_{int}, m_{rain} and 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, we did use a gauge-adjusted radar product to determine the DBC parameter, based on the overall bias in the complete network. Bias changes for individual stations are adjusted with the dynamic bias correction, but DBC needs to be redetermined with an independent reference periodically to address changes in overall network bias (due to loss, additions and accuracy changes over lifetime of PWSs).

In this study the filter was applied on crowdsourced PWS rainfall observations every 5 minutes, 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. The next phase for this research is to develop tools that can employ the QC algorithm fast enough to apply on PWS observations in real-time and at larger spatial scales. Moreover, raw data is collected on platforms that are maintained by commercial organizations. They need to be made accessible in real-time for PWS networks to become a viable data source for rainfall monitoring, mostly in developed regions of the world. If these issues are addressed, a huge number of in-situ rainfall observations available in real time can be used for various (operational) purposes, especially for improving radar rainfall products.

Acknowledgments

This work was financially supported by KNMI MSO project 2013.09, co-funded 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 dataset available to us. See `https://dev.netatmo.com/resources/technical/reference/weatherapi/getpublicdata` for information on how to obtain Netatmo data. The gauge-adjusted radar rainfall dataset 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_nl25_rac_mfbs_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*(3), 1744–1757.

Beekhuis, H., & Mathijssen, T. (2018). From pulse to product, highlights of the upgrade project of the Dutch national weather radar network. In (pp. 960–965). (Extended abstract for ERAD2018 conference) doi: 10.18174/454537

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

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.

Butgereit, L. (2014). Crowdsourced weather reports: An implementation of the μ model for spotting weather information in twitter. In *IST-Africa conference proceedings, 2014* (pp. 1–9).

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* (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 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.

Droste, A., Heusinkveld, B., & Steeneveld, G.-J. (2018). Crowdsourcing the urban wind climate using private weather stations. In *EGU General Assembly Conference Abstracts* (Vol. 20, p. 8833).

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*, 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*, 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., Wang, J., Tang, S., Zhao, Z., Sun, Z., … Liu, H. (2016).

Tefnut: An accurate smartphone based rain detection system in vehicles. In *International Conference on Wireless Algorithms, Systems, and Applications* (pp. 13–23).

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.

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.

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).

Meier, F., Fenner, D., Grassmann, T., Jänicke, B., Otto, M., & Scherer, D. (2015). Challenges and benefits from crowd-sourced atmospheric data for urban climate research using Berlin, Germany, as testbed. In *ICUC9 - 9th International Conference on Urban Climate jointly with 12th Symposium on the Urban Environment.*

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., ... Leigh, R. R. (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., ... others (2015). Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation. *Journal of Hydrology*, *531*, 389–407.

Overeem, A., Buishand, T., & Holleman, I. (2009). Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. *Water Resources Research*, *45*(10).

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*(12).

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

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.

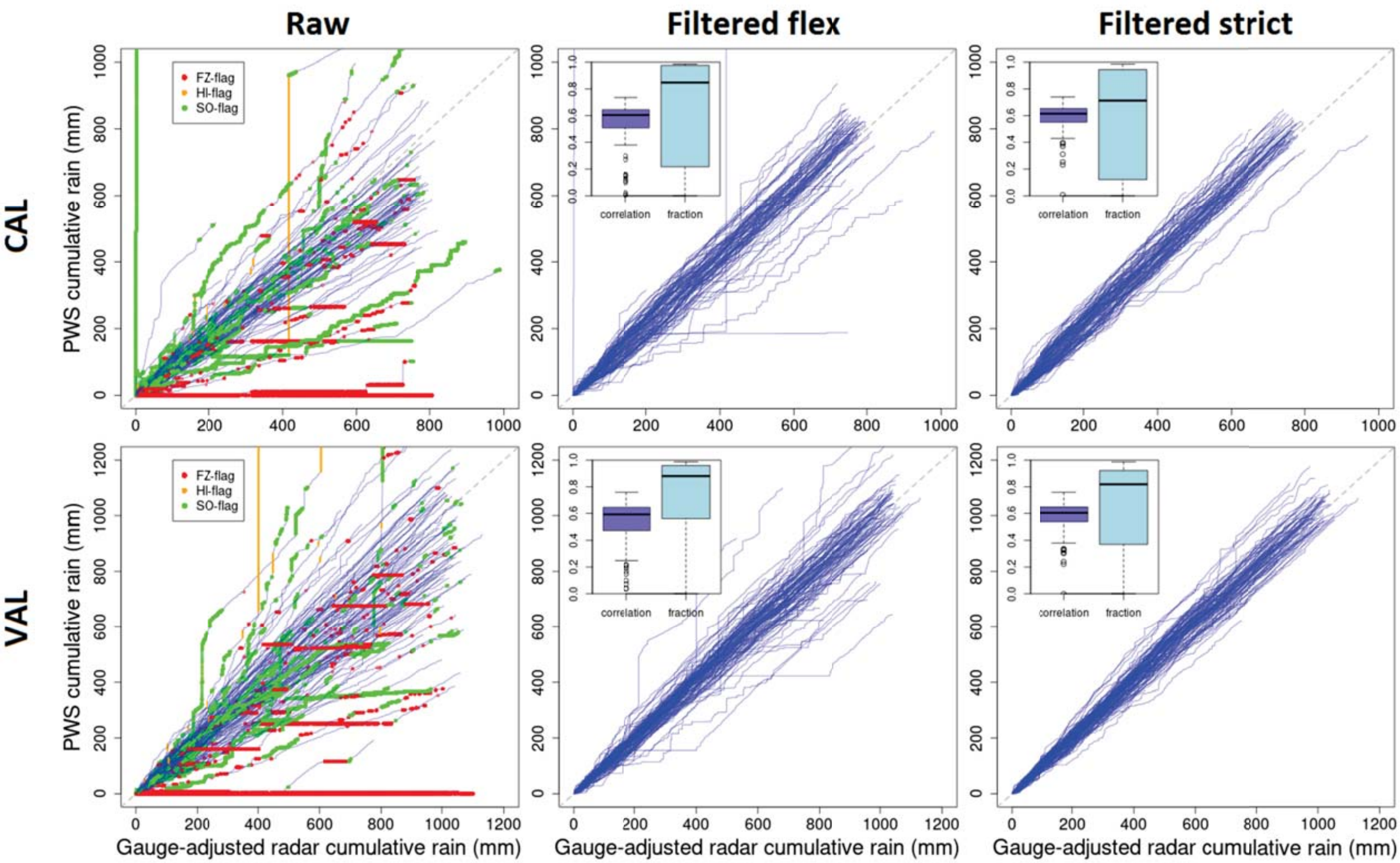
Reges, H. W., Doesken, N., Turner, J., Newman, N., Bergantino, A., & Schwalbe, Z. (2016). COCORAHS: The evolution and accomplishments of a volunteer rain gauge network. *Bulletin of the American Meteorological Society*, *97*(10), 1831–1846.

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*(11), 9462–9478.

Zahumenskỳ, 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., ... others (2018). Crowdsourcing methods for data collection in geophysics: State of the art, issues, and future directions. *Reviews of Geophysics*, *56*. doi: 10.1029/2018RG000616

Figure1.



Raw
Filtered flex
Filtered strict
CAL
VAL
FZ-flag
HI-flag
SO-flag
PWS cumulative rain (mm)
Gauge-adjusted radar cumulative rain (mm)
correlation
fraction

Figure2.

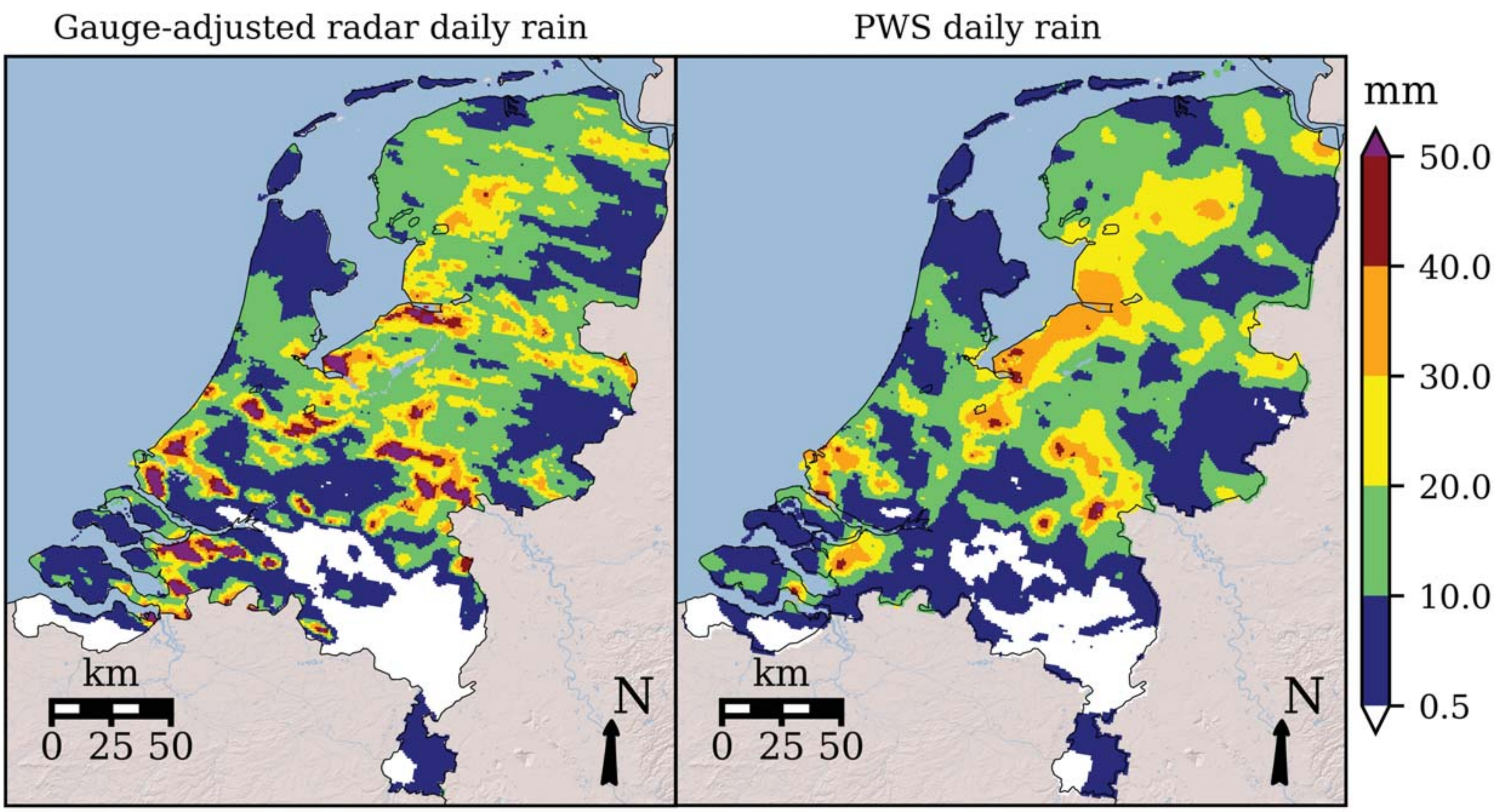
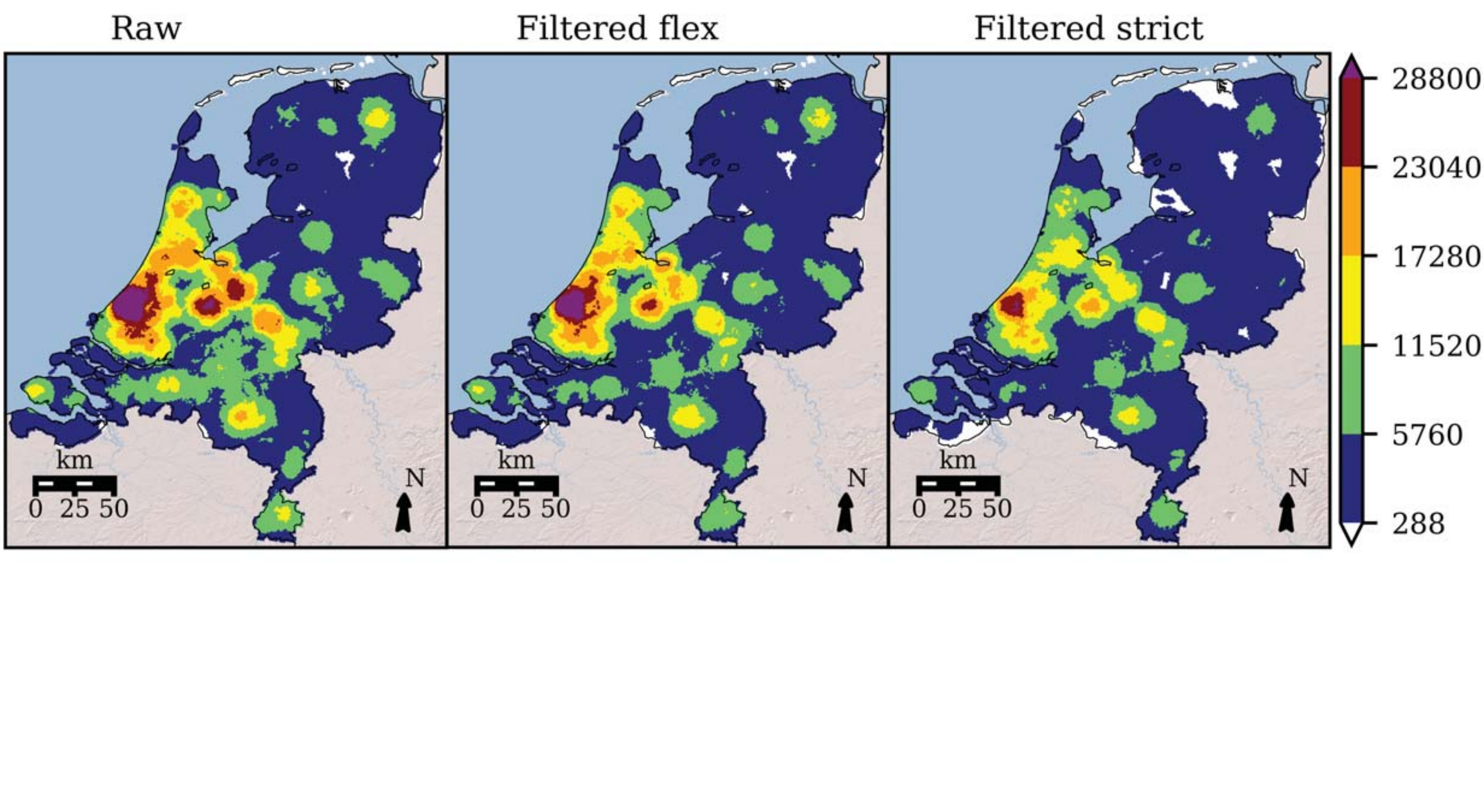

Gauge-adjusted radar daily rain
PWS daily rain
mm
50.0
40.0
30.0
20.0
10.0
0.5
km
0 25 50
N
km
0 25 50
N

Figure3.


Raw
Filtered flex
Filtered strict
km
0 25 50
N
28800
23040
17280
11520
5760
288