Evaluation of high-intensity rainfall observations from personal weather stations in the Netherlands

Nathalie Rombeek[1], Markus Hrachowitz[1], Arjan Droste[1], and Remko Uijlenhoet[1]

[1]Department of Water Management, Delft University of Technology, the Netherlands

Correspondence: Nathalie Rombeek (n.rombeek@tudelft.nl)

Abstract. Accurate rainfall observations with a high spatial and temporal resolution are key for hydrological applications, in particular for reliable flood forecasts. However, rain gauge networks operated by regional or national environmental agencies are often sparse and weather radars tend to underestimate rainfall. As a complementary source of information, rain gauges from personal weather stations (PWSs), which have a network density 100 times higher than dedicated rain gauge networks in the Netherlands, can be used. However, PWSs are prone to additional sources of error compared to dedicated gauges, because they are generally not installed and maintained according to international guidelines. A systematic long-term analysis involving PWS rainfall observations across different seasons, accumulation intervals and rainfall intensity classes is missing so far. Here, we quantitatively compare rainfall estimates obtained from PWSs against rainfall recorded by automatic weather stations (AWSs) from the Royal Netherlands Meteorological Institute (KNMI), over the 2018-2023 period, including a sample of 1760 individual rainfall events in the Netherlands. This sample consists of the 10 highest rainfall accumulations per season and accumulation interval (1, 3, 6 and 24 h) over a 6-year period. It was found that the average of a cluster of PWSs severely underestimate rainfall (around 36% and 19% for 1 h and 24 h intervals, respectively). ~~Adjusting~~ By adjusting the data with ~~the mean field~~ areal reduction factors to account for the spatial variability of rainfall extremes and applying a bias correction factor of ~~1.24, as proposed by the PWSQC algorithm, reduces this underestimation to 21%~~ 1.22 to compensate for instrumental bias, the average relative bias reduces to -5% for 1 h intervals or almost reduces ~~it to 0~~ to zero for intervals of 3 h and longer. Largest correlation (~~0.83 and 0.83~~0.85 and 0.86) and lowest coefficient of variation (~~0.15~~ 0.14 and 0.18) were found for 24 h intervals during winter and autumn, respectively. We show that most PWSs are able to capture high rainfall intensities up to around 30 mm h^{-1}, indicating that these can be utilized for applications that require rainfall data with ~~with~~ a spatial resolution on the order of ~~kilometers~~kilometres, such as for flood forecasting in small, fast responding catchments. ~~However, PWSs severely underestimate (on average more than 50%) rainfall events~~ PWSs did not observe the most intense rainfall events which were associated with return periods exceeding 10 or 50 years (above approximately 30 mm h^{-1}) ~~, which~~ and occurred in spring and summer. ~~These underestimations are associated with large areal reduction factors, which can result in a reduction up to 17% for 1 h events with a return period of 50 years. Additionally, this~~ However, the spatial distribution of rainfall likely played a large role in the observed differences, rather than instrumental limitations. This stresses out the importance of having a dense rain gauge network. In addition, the variation in undercatch is likely partly due to the disproportional underestimation of tipping bucket rain gauges with increasing intensities. ~~We recommend additional research on dynamic calibration of the tipping volumes to further improve this.~~ Outliers during winter were likely caused by solid precipitation and can potentially be

removed using a temperature module from the PWS. We recommend additional research on dynamic calibration of the tipping volumes to further improve this.

1 Introduction

Accurate rainfall observations are essential for hydrological applications, such as flood forecasting. However, rainfall is highly variable in time and space, making it challenging to capture its dynamics accurately. Consequently, the stochastic nature of rainfall is one of the main sources of uncertainty in hydrological ~~modeling~~ modelling (Niemczynowicz, 1999; Moulin et al., 2009; Arnaud et al., 2011; Lobligeois et al., 2014; Beven, 2016; McMillan et al., 2018). Especially small, fast-responding catchments require accurate rainfall observations with ~~a~~ high spatial and temporal resolution for reliable predictions~~(e.g. ~~, such as~~ in the order of ~~kilometers~~ kilometres and minutes for ~~catchments areas of around 560 ha)~~ catchment areas of a few square kilometres (Berne et al., 2004; Ochoa-Rodriguez et al., 2015; Cristiano et al., 2017; Thorndahl et al., 2017). To reduce the uncertainty of catchment-scale rainfall estimates, accurate instruments with a high spatio-temporal resolution are required.

Rain gauges and weather radars are widely used instruments to provide rainfall information for hydrological forecasting. Each instrument has its own advantages and limitations. Rain gauges can record rainfall relatively accurately at the point-scale. These rain gauges can be automatic or manual, observing at short regular intervals (e.g. recorded every 12 sec, archived at 10 min time steps in the Netherlands) or daily, respectively. A limitation is that these measurements are strictly only representative for the orifice area of the individual recording gauge and the network density of dedicated rain gauges is not sufficient to capture small-scale rainfall dynamics (Villarini et al., 2008; Hrachowitz and Weiler, 2011; Van Leth et al., 2021). In addition, rainfall observations from manual gauges, which are emptied in a measuring cylinder and read once a day, are not available in (near) real time. Weather radars, on the other hand, provide high spatial and temporal resolution data (i.e. typically 1 km^2 and 5 min), that is available in near real-time. However, radar rainfall estimates are prone to substantial uncertainty and bias due to several sources of error. These are related to ~~both the instrument itself (e.g. calibration)~~ for example the calibration of the instrument itself, signal attenuation and to the conversion from measured reflectivities aloft into rainfall rates at the ground (Uijlenhoet and Berne, 2008; Krajewski et al., 2010; Villarini and Krajewski, 2010).

Alternatively, crowdsourced rain measurements, in the form of low-cost weather observation devices, may potentially provide accurate local rainfall observations. These devices are referred to as personal weather stations (PWSs) and can contain a rain gauge module, which records rainfall at a high temporal resolution (5 min). The popularity of these low-cost ~~rainfall sensors~~ sensors equipped with a rain gauge has been increasing in the last decade, up to around 1 PWS per 9, 11, 13 and 15 km^2 in May 2024, in the Netherlands, Denmark, Switzerland and Germany, respectively. Figure 1 shows several tens of thousands PWSs with varying densities across Europe, with more than 60% having 5 or more neighbouring stations within 10 km. In the Netherlands, these sensors currently have a network density which is about 100 times higher than that of the AWSs, with even higher densities in urban areas, where AWS densities are typically low (Overeem et al., 2024b; Brousse et al., 2024). ~~The data from these observations is automatically uploaded to different online platforms~~ Once the PWS is connected to an online platform such as the Weather Observations Website (WOW; https://wow.metoffice.gov.uk/), the Weather Underground

website (https://www.wunderground.com/wundermap) ~~and~~ or Netatmo (https://weathermap.netatmo.com/)~~.~~, observations are automatically uploaded to the respective platform. The PWS data is open access and in case of Netatmo, it can be extracted in near real-time using an application programming interface (API) every 5 min. However, since they are installed, operated and maintained largely by non-specialist citizens, the data quality of these PWSs is expected to be lower than that of professionally operated gauges of national meteorological or hydrological services.

Rain gauges ~~, both PWS and AWS,~~ from PWSs are prone to several sources of error. These errors can be grouped into three categories~~: (1) instrumental~~. The first category consists of PWS-related errors, such as ~~calibration errors or~~ those related to inappropriate setup and lack of maintenance of rain gauges, calibration errors, rounding due to data processing, as well as connectivity issues during data transfer (De Vos et al., 2017; De Vos, 2019). The second category includes general rain gauge-related errors, such as undercatch due to wind, solid precipitation or evaporation, and the intrinsic tipping bucket error, ~~meaning that a~~ resulting from the given volume of water that needs to be collected before ~~it tips (Habib et al., 2001) or; (2) data processing, such as rounding or connectivity issues during data transfer and (3) setup and maintenance. In addition, uncertainties~~ the bucket tips (Habib et al., 2001; Lanza and Vuerich, 2009). A third category of errors arise due to spatial sampling ~~errors resulting from estimating areal rainfall using point measurements (Villarini et al., 2008).~~ uncertainties resulting from gauges that are not co-located and thus differences between point rainfall estimates (Villarini et al., 2008).

Rain gauges of PWSs typically ~~uses a~~ use an unheated tipping bucket mechanism to record the rainfall volumes. The quality of rainfall intensity estimates from these mechanisms has been shown to be intensity-dependent. Tipping buckets are known to overestimate rainfall at low intensities and underestimate for high intensities (Marsalek, 1981; Shedekar et al., 2009; Colli et al., 2014)~~, which are part of the first category of errors~~. In addition, these errors can be amplified if the PWSs are not installed correctly and maintained adequately.

With respect to the ~~first two~~ PWS specific sources of error, De Vos et al. (2017) used an experimental setup to investigate part of the instrument and data processing-related errors from PWSs. They showed that, under ideal circumstances (i.e. installed and maintained according to World Meteorological Organization standards), three rain gauges, from the Netatmo brand, recorded rainfall with high accuracy. Collocating the PWSs very close to one of KNMI's automatic weather stations (AWSs), spatial sampling errors were also limited. Despite the potential of accurate rainfall measurements from PWSs, their observations are often not optimal, as the stations are not necessarily installed according to guidelines from the World Meteorological Organization. ~~To account for those, errors related to the third category are introduced.~~ For that reason, De Vos et al. (2019) developed a quality control (QC) algorithm, to filter outliers from the PWS network without using data from an official rain gauge network or weather radar. Similarly, Bárdossy et al. (2021) developed a QC algorithm, using a reference observation network to filter outliers and to correct the bias. Chen et al. (2018) assigned trust scores based on spatial consistency between stations.

While previous work has shown that implementing these QC algorithms ~~has been shown to yield~~ yields an overall improvement in the quality of the PWS data ~~(De Vos et al., 2019; Bárdossy et al., 2021; Overeem et al., 2024a; Nielsen et al., 2024; El Hachem et a~~ (De Vos et al., 2019; Bárdossy et al., 2021; Graf et al., 2021; Overeem et al., 2024a; Nielsen et al., 2024; El Hachem et al., 2024), a systematic long-term analysis of the QC algorithm of De Vos et al. (2019) for different seasons, accumulation intervals and

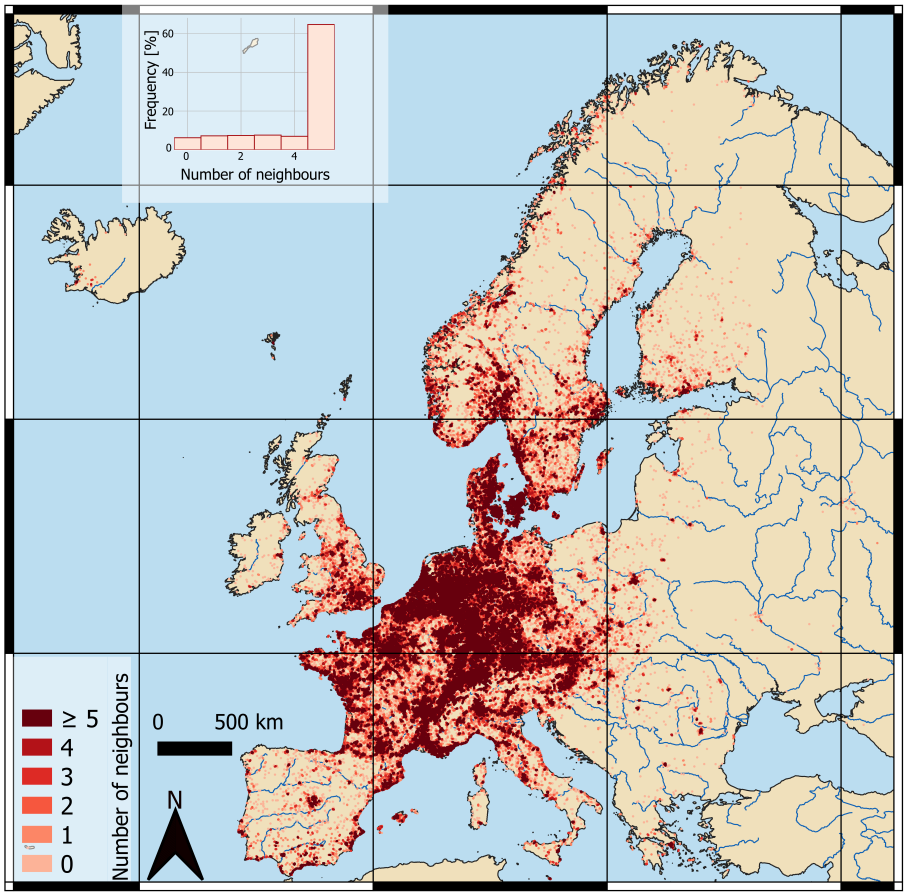

Figure 1. Indication of the Netatmo density of PWSs with rain gauges within Europe, showing number of PWS neighbours within a radius of 10 km. Data extracted from ~~netatmo~~ Netatmo API on April 2024.

intensity classes is missing so far. In particular, a focus on high rainfall intensities is important, as the undercatch of rain gauges is likely disproportional with increasing intensities. Here, we aim to quantitatively compare rain data from PWSs with AWSs, extending on the results of De Vos et al. (2017, 2019). While weather radar data has a larger spatial resolution compared to AWSs, it is not used in this research as a reference because it is prone to several sources of error and therefore significantly underestimates rainfall. Note that in previous research from Overeem et al. (2024a) and Nielsen et al. (2024) rainfall estimates from PWSs were actually used to correct a rainfall radar product.

The objective of this study is to systematically quantify and describe the uncertainties arising from PWS rainfall estimates. By ~~analyzing~~ analysing the 10 largest rainfall accumulations, with return periods up to 50 years, during the period between 2018 and 2023, for 11 AWSs, 4 seasons and 4 time intervals, we can draw statistically meaningful conclusions on this. To the best of the authors' knowledge, such long-term study using PWS data that focuses on the most intense rainfall events, has not been performed before. Quantifying the limitations ~~the~~ of PWS rainfall observations and ~~eventually correcting it, maximizes~~

addressing them enhances the potential of PWSs for a wide range of applications, including hydrological modelling, urban hydrology and (hydrological) ~~forecasts.~~ forecasting.

2 Study area and data

This study was carried out over the period 2018-2023 in the Netherlands, which has a land surface area of approximately 35,000 km^2 (Fig. 3a). This period was chosen due to the PWS data availability, which was too low (less than 5 PWSs within 10 km distance of the AWS) before 2018 and increased over the years. The Netherlands has a maritime climate (Cfb according to the Köppen classification), where winters are mild, with an average temperature of 3.8 °C and relatively cool summers (17.2 °C) (KNMI, 2024). The average yearly rainfall between 1990-2020 is 851 mm yr^{-1} over the area (KNMI, 2024). In addition, regional variability in rainfall extremes is observed, with higher values in the western part of the country (Overeem et al., 2009a; Beersma et al., 2019). The Dutch climate has a quite uniform distribution of precipitation over the meteorological seasons, except during spring, which is the driest season and contains the driest month (i.e. April, average 41 mm) (see Fig. 2). August is on average the wettest month (average 87.4 mm) (KNMI, 2024). However, rainfall characteristics differ over the seasons. Rainfall during the summer months and beginning of autumn is characterized by shorter duration and higher precipitation intensities, as a consequence of convection during these months. In contrast, during the winter months lower intensities and more frequent and longer precipitation events occur (De Vries and Selten, 2023). These different rainfall characteristics lead to a distinct seasonal cycle in spatial rainfall correlation in the Netherlands (Van de Beek et al., 2012, Fig. 4b), with longer correlation distances for winter than summer.

2.1 Personal weather stations

For the analysis here rain gauges from the Netatmo brand of PWSs were used. These PWSs have a large coverage over the Netherlands which slightly increased since 2018 (around one PWS per 9 km^2 in 2024; Fig. 3a). This rain gauge type uses a tipping bucket mechanism with a nominal volume of 0.101 mm according to the manufacturer (Netatmo, 2024a). These gauges can also be calibrated manually by the owner by changing via software the volume per tip, resulting in deviating tipping bucket volumes (approximately 13.5% is manually calibrated according to De Vos et al. (2019)). The ~~default rain gauge processing software records the number of tips over 5 min intervals, which is communicated wirelessly to an indoor module. Next, the data is transferred, using wifi, to the Netatmo platform. The~~ diameter of the collecting funnel is 13 cm (leading to an orifice area of 133 cm^2). According to the manufacturer, the accuracy is 1 mm h^{-1} for a measurement range of 0.2 to 150 mm h^{-1} and the PWS operates best for temperatures between 0 and 50°C (Netatmo, 2024a). However, it is unclear what this accuracy exactly entails, therefore pointing out the need for this study.

The default rain gauge processing software records the number of tips over approximately 5 min intervals, which is communicated wirelessly to an indoor module. Next, the data is transmitted via wifi to the Netatmo platform. The Netatmo software resamples this to regular 5-min intervals, by assigning it to the next full five-minute interval by the Netatmo software. When within a 5-min interval no data is transferred, this time interval is not included by Netatmo (see supporting information Table A1 for

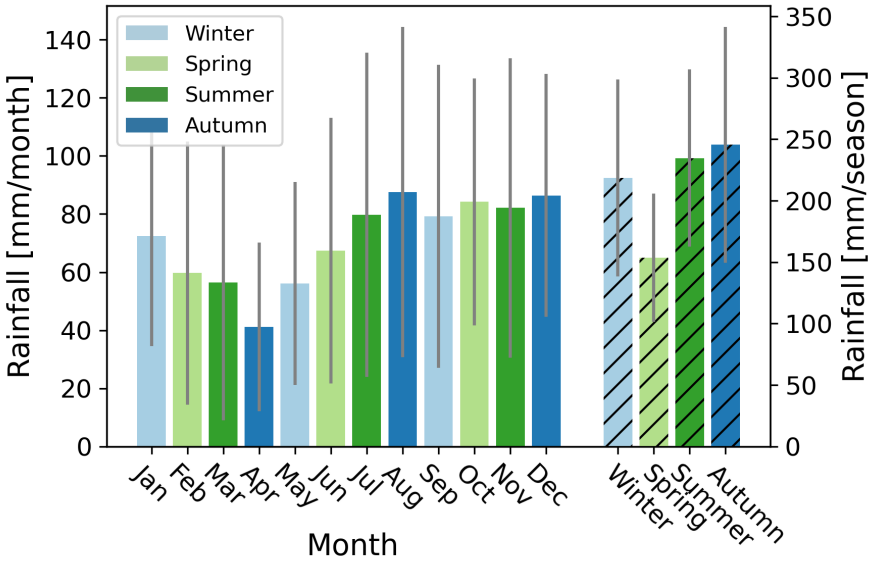

Figure 2. Average rainfall per month and season in the Netherlands over the period 1991-2020, based on data from 13 automatic weather stations spread over the country, obtained from KNMI (2024). Coloured bars indicate the average rainfall per month (~~left y-axis,~~ mm month^{-1}, left y-axis), ~~horizontal black lines~~ coloured hatched bars indicate the average rainfall ~~over~~ for each season (mm season^{-1}, right y-axis). Vertical grey lines indicate the inter quartile range.

an example). When there is a connection failure between the rain gauge module and indoor module, the rainfall will likely be attributed to a timestamp when there is a connection again, potentially aggregating it over a longer time interval than approximately 5 min (see supporting information Table A2 for an example). However, when the connection of the indoor module is also temporarily interrupted, data is lost.

2.2 Reference dataset

The PWSs were evaluated against data from automatic weather stations (AWSs) from the Royal Netherlands Meteorological Institute (KNMI). The KNMI operates a network of 33 AWSs across the Netherlands, which are relatively homogeneously distributed, with approximately one AWS per 1000 km^2 (Fig. 3a). These AWSs estimate cumulative rainfall every 12 s by measuring the displacement of a float placed in a reservoir. The data is archived at a lower resolution, i.e. every 1 min, 10 min and hourly. Here, AWS data with a resolution of 10 min and 1 h is used. The 10 min dataset contains unvalidated rainfall data, while the hourly data has been validated (Brandsma et al., 2020). The collecting funnel has a diameter of 16 cm (corresponding to an orifice area of 201 cm^2) and the device is heated for temperatures below 4°C. In addition, these stations are placed in open locations using an English setup or Ott windscreen to reduce errors from wind induced undercatch (Brandsma et al., 2020). ~~The data is archived at a lower resolution, i.e. every 1 min, 10 min and hourly. Here, AWS data with a resolution~~

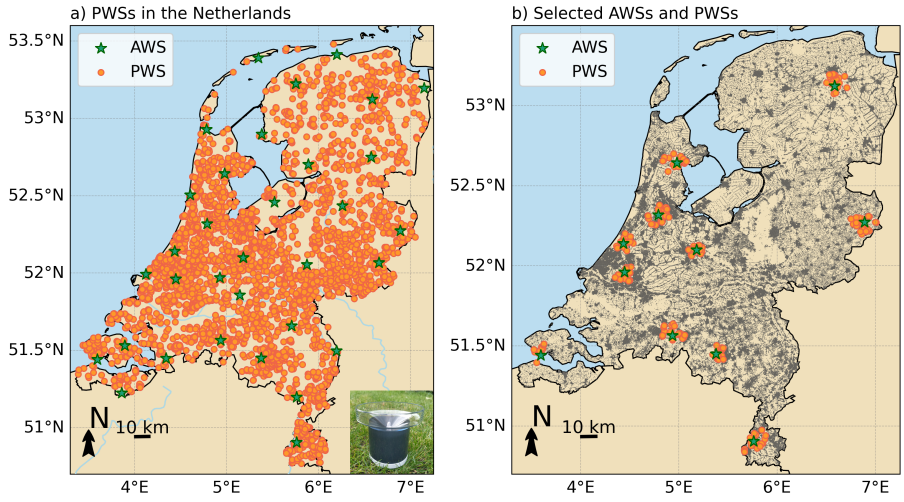

Figure 3. Map of the Netherlands showing a) the locations of the 33 AWSs employed by KNMI and the locations of available PWSs in 2024 obtained using the Netatmo API (approximately 4000 PWSs) (Netatmo (2024b) last access: 08/05/2024). The photo in the right corner shows an example of the PWS used in this research. Note that working PWSs in 2024 are not part of the dataset in this study. b) The selected AWSs and PWSs from 2018 to 2023 in this study. Built-up areas, indicated in grey, were obtained from European Environment Agency (2020).

~~of 10 min and 1 h is used. The 10 min dataset contains unvalidated rainfall data, while the hourly data has been validated (Brandsma et al., 2020).~~ Nevertheless, this data is not an absolute truth. Brandsma (2014) compared the AWS network and manual rain gauge network over the Netherlands and concluded that the AWS network underestimates rainfall with 5-8% annually, with higher underestimation in winter (7.7%) than summer (5.0%). The undercatch is nonlinear with intensity, with larger intensities resulting in less underestimation. These uncertainties are not taken into account in this research.

3 Methods

3.1 Station selection

Rainfall data at 5-min intervals from multiple PWSs were extracted using the Netatmo API (Netatmo, 2024b). ~~Netatmo limits the API requests per hour.~~ Note that the API only provides ~~PWSs that are operational at~~ access to data from PWSs that were operational at the time of access, which was in February ~~2024, meaning we~~ 2024. We do not have ~~the exact recording stations , which varied over time~~ access to data from stations that were previously in operation but are no longer online at the time of access. Two search radii were employed to find all operational PWSs within that range. One radius of 10 km around an AWS was used for quantifying the quality of PWSs and a radius of 20 km for filtering the PWS data using a quality control algorithm. Van de Beek et al. (2012) and Van Leth et al. (2021) showed that the decorrelation distance for precipitation over

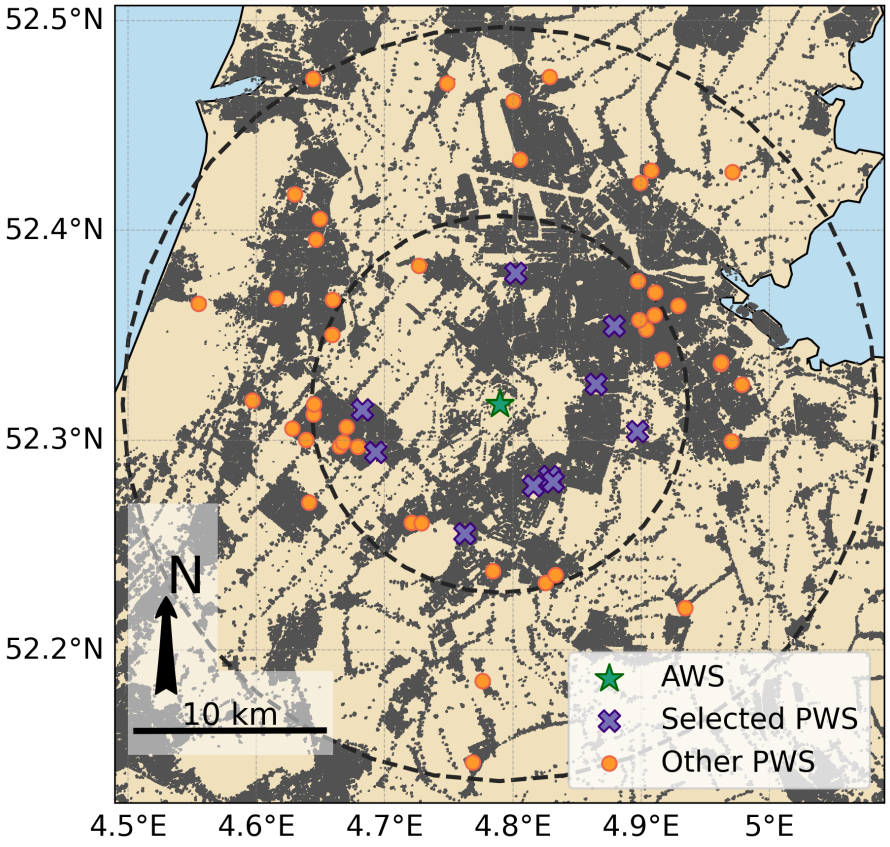

Figure 4. Example of the selection procedure for the AWS at Schiphol. The green star indicates the AWS operated by KNMI, the purple crosses indicate the 10 closest PWSs within a distance of 10 km around the AWS. The orange dots are the other PWSs within 20 km from the AWS, which are utilized for quality control. Built-up areas, indicated in grey, were obtained from European Environment Agency (2020).

the Netherlands is around 50 km for 1 h accumulation intervals. Comparing PWSs within 10 km from an AWSs can therefore be assumed to limit spatial sampling errors with respect to a larger search radius.

First, all PWSs within 20 km around each AWS were identified. Secondly, AWSs that had at least 5 PWSs within 10 km since 2018-01-01 were kept in the dataset. Next, the 10 closest available PWSs located within 10 km from the AWS were selected (Fig. 4, purple crosses) for the comparison with the AWS nearby. Note that the selection of PWSs varies per selected rainfall event, due to temporary station outages and changing data and station availability over time. This procedure resulted in 11 AWSs with a cluster of 5 to 10 PWSs around it (Fig. 3b for selected AWSs and used PWSs). All PWSs within 20 km were used for filtering the data (Fig. 4, orange dots and purple crosses) using the quality control algorithm developed by De Vos et al. (2019) and described below (Sect. 3.4).

3.2 Event selection

A similar event selection procedure was used as described by Imhoff et al. (2020), which defines an event as a certain time period, rather than by the beginning and end of rainfall. Rainfall observations from the 10-min dataset of the AWSs were employed to make a selection of events between 2018 and ~~2023. Only the~~ 2023 using a moving window approach. Only the 10 largest rainfall accumulations were selected, as these are the most important ones for pluvial flood forecasting. Analysis shows that, on average, the 10 highest 1-h rainfall accumulations for the selected AWSs account ~~or~~ for 12.5% of the annual rainfall. The ~~10-min dataset of the AWSs was used to select these events, the hourly dataset~~ hourly dataset (clock-hour) was employed to perform a consistency check on this selection. For selected events with rainfall differences of more than 10% with the validated hourly dataset, the values of the validated hourly dataset were used instead. These deviations occurred in less than 0.4% of the total selected events. Note that the selected events can contain time steps without any rain observed.

For every selected AWS based on the methodology in Section 3.1, the 10 largest rainfall events per meteorological season (winter, spring, summer and autumn) and per accumulation interval (1, 3, 6 and 24 hours) were selected to draw statistically meaningful conclusions. This results in a total of 11 stations x 4 seasons x 4 accumulation intervals x 10 events = 1760 individual events for the analysis. The events were selected in such way that for the same station and accumulation interval no overlapping time series were ~~selected~~included.

The statistics of the selected events are shown in Fig. 5. A clear seasonality is observed here, especially for 1 h intervals (Fig. 5a), with highest rates during summer (JJA) and lowest during winter (DJF), with a median of 14.65 and 6.17 mm h^{-1}, respectively. This is aligned with the Dutch climate, where highest rainfall intensities occur during summer and are typically characterized by convective rainfall (Beersma et al., 2019).

Based on the return periods provided by Beersma et al. (2019), over 75% of the selected observed rainfall accumulations in winter, spring and autumn have a return period of less than 0.5 year, while in summer this is around 25% (Fig. 5a). Most extreme events occur during the summer months, with multiple events having a return period of over 5 years, and one event exceeding a return period of 100 years. Rainfall rates during the autumn months (SON) are slightly higher compared to spring (MAM), for example with a median of 9.40 and 8.33 mm h^{-1}, respectively (Fig. 5a). However, spring appears to have two extreme outliers, with return periods that exceed 10 years. Note that these return periods are calculated based on annual statistics, which is dominated by rainfall events during March until October. Because winter has the lowest intensities, return periods in winter based on annual statistics are low.

3.3 ~~Quality control algorithm~~Areal reduction factor

The rainfall observed by a cluster of PWSs is averaged, effectively representing the rainfall over an area, while comparing it with a point measurement (AWS) which has a limited spatial footprint. With increasing domain area, the variation of areal precipitation becomes smaller than that of point precipitation. To account for the reduction in the magnitude of rainfall extremes over an area as compared to a point, areal reduction factors (ARFs) can be applied. The ARF estimates areal rainfall percentiles from point rainfall percentiles. Overeem et al. (2010) and more recently Beersma et al. (2019) parameterized the ARF based

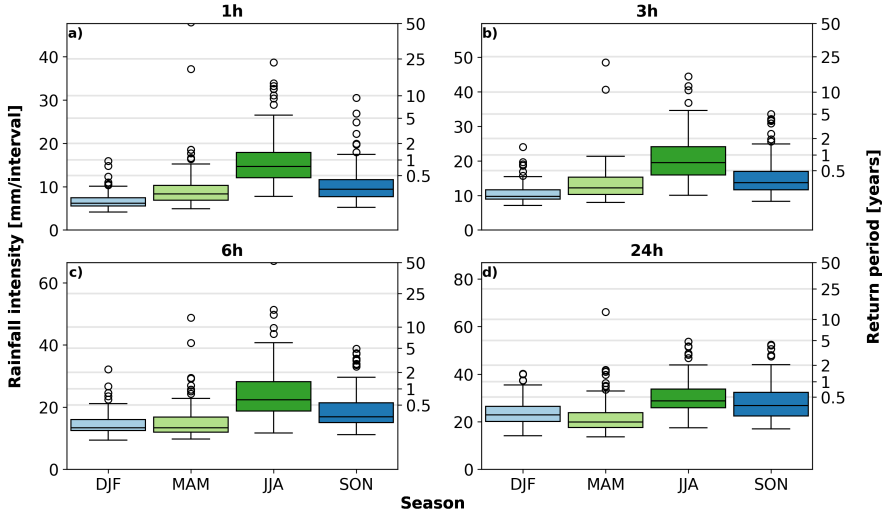

Figure 5. Mean rainfall intensity [mm/interval] for the different selected events per season (11 AWSs x 10 rainfall events) for the four accumulation intervals (1, 3, 6 and 24 h). Left y-axis shows the rainfall over the specific interval, right y-axis indicate the different corresponding return periods for the intensities reported by Beersma et al. (2019). The lower and upper whiskers indicate the minimum and maximum intensities and the boxes the inter-percentile range (25^{th}–75^{th}). During summer outliers were present, with return periods larger than 50 years.

on weather radar for the Netherlands. This reduction factor is a function of duration, area and return period, with rarer events having a stronger areal reduction. Equations 1a, 2, 3b, 4 and 5 from Beersma et al. (2019) are used to estimate the ARF. The inverse of the ARFs will be used to adjust the values of the PWSs to a point observation.

3.4 Quality control algorithm

~~Part of the quality control algorithm PWSQC from De Vos et al. (2019) was applied to filter the PWS dataset, using the same parameters. First, a list of PWS neighbours was constructed. Secondly,~~ As stated by El Hachem et al. (2024), the key advantage of the QC algorithm from De Vos et al. (2019) over QC algorithms such as developed by Bárdossy et al. (2021) and Lewis et al. (2021), is that no auxiliary data is required. This makes it particularly suitable for regions lacking access to (real-time) reference data. For that reason, we decided to use the QC from De Vos et al. (2019). The high influx (HI) and faulty zeroes (FZ) filters ~~were computed for every timestep (i.e. 5 min)~~from the quality control algorithm PWSQC from De Vos et al. (2019) were applied to filter the PWS dataset. High influx data can be caused by sprinklers, adding liquids into the gauge, or tilting of the gauge. In addition, a high influx can result from a temporary connection interruption between the rain gauge module and the indoor module, assigning the rain to the timestamp when the connection is re-established. The HI filter uses four parameters:

1. d, the maximum distance over which neighbouring PWSs are selected, which likely capture similar rainfall dynamics.
2. n_{stat}, minimum required number of neighbouring PWSs.

3. ϕ_A, threshold value.

4. ϕ_B, threshold value.

A time interval of a station is flagged as having a "high influx" if the median of the neighbouring stations does not exceed ϕ_A, while the station itself records a value above ϕ_B. When the neighbouring stations observe moderate to heavy rainfall, the station is flagged when the measurement exceeds median * (ϕ_B/ϕ_A). According to De Vos (2019) most rainfall observations that should be flagged by the HI filter, were very high. They tested different subsets of parameters and found that variations in ϕ_A and ϕ_B hardly affect the results.

Faulty zeroes can result from failure of the tipping bucket mechanism ~~. At least five neighbouring~~ due to for example a tilted rain gauge or obstructions such as leaves or insects. The FZ filter uses three parameters, the range d, n_{stat} and n_{int}. At least for n_{int} time intervals, the median of the neighbouring PWSs needs to be higher than zero, while the PWS itself reports zero rainfall.

The calibrated value of parameter d by De Vos (2019) (Table 1) is 10 km for both HI and FZ. This is the average decorrelation distance of rainfall at the 5-min time interval in the Netherlands (Van Leth et al., 2021, Fig. 4a). This same work shows that this value ranges from about 10 km in summer to about 50 km in winter. In our research, we limit the effect of spatial variability of rainfall, by selecting only the five closest neighbouring stations (this is on average a distance of 5.4 km), well within the decorrelation distance of rainfall at the 5-min time scale for any season in the Netherlands (Van Leth et al., 2021).

For the reasons mentioned above, the same calibrated parameters as in Table 2 from De Vos (2019) were applied. First, a list of PWS neighbours within 10 km was constructed. Secondly, HI and faulty zeroes FZ filters were computed for every timestep (i.e. 5 min). At least five neighbouring PWSs must be present to attribute the HI and FZ flags, otherwise the value will be eliminated from the dataset. Time steps that were flagged according to the HI or FZ flags were removed.

The station outlier (SO) filter from the PWSQC algorithm requires at least two weeks of data (or longer, if there was insufficient precipitation in this period), and is computationally expensive, which is not favourable for real-time applications. In addition, by taking the average of a cluster (minimum 5, maximum 10) of stations around an automatic weather station, the effect of individual station outliers is limited. This last step is different from the method suggested by De Vos et al. (2019). As a last step of the PWSQC algorithm, a default bias correction factor ~~of 1.24~~ (DBC) was applied to the dataset. ~~This factor is based on correcting the bias over an entire area, rather than addressing individual biases in PWSs. For this reason, we refer to it in this research as mean field bias correction (MFBC)~~De Vos et al. (2017) used an experimental setup and showed that under ideal circumstances there is on average an instrumental bias of 18% in the Netatmo PWSs, suggesting the need for a DBC factor of 1.22 to correct these instrumental biases.

3.5 Network stability

Over time the availability of PWSs changes due to factors such as connection failure. To ~~analyze~~ analyse the availability of PWSs, a dataset comprising operational PWSs at 2018-01-01 within 10 km from an AWS were employed. This dataset contained 178 PWSs. Time series from the PWSs were extracted for each selected event according to the method described

Table 1. Percentage of the PWSs available (28,480 different PWS time series) over all selected events after flagging data and setting a minimum availability criteria.

Availability criteria	Remaining PWSs
100%	40.9%
92%	79.9%
83%	83.7%
75%	85.8%

in Section 3.2. In total, 95.8% of the PWSs were available over all selected events during the 6 consecutive years. From these available PWSs, 3.5% of the total time steps did not contain any data, due to either missing data or irregular data transfers that were longer than 5-min. In addition, 0.7% of the PWSs did not have 5 neighbours within 10 km. Applying the quality control algorithm leads to discarding more data. For the remaining PWSs with enough neighbours, 8.9% of the total data is either discarded (i.e. flagged as FZ or HI) or has a temporal resolution beyond 5 min or is missing.

Since we can only suspect that data is likely not missing when a 5-min timestep is not included, a minimum availability criteria was set to limit a biased comparison. A minimum percentage of time steps should be valid before aggregating the data. The criteria set has a large impact on the availability (Table 1). By requiring a data availability of 100% before aggregation, 40.9% of the dataset is retained, while a lower required availability (92%) almost doubles (79.7%) the remaining stations of the original dataset. Lower criteria potentially result in underestimation of rainfall due to missing data. Based on these results, a data availability requirement of 83% was chosen (e.g. at least 10 out of 12 5-min intervals within one hour) to keep a large part of the original dataset (83.7%), which is also in line with Overeem et al. (2024a).

3.6 Validation

The data quality of the PWSs was evaluated by comparing the PWS data to those of the selected AWSs, using the relative bias, coefficient of variation (CV) of the residuals, the Pearson correlation coefficient (r), and the slope of linear regression relationship. Note that the evaluation metrics were calculated over the total rainfall within a time interval and over the average of the cluster of PWSs.

The relative bias was defined as follows:

$$\text{Bias} = \frac{\sum_{i=1}^{n} R_{\text{PWS},i}}{\sum_{i=1}^{n} R_{\text{AWS},i}} - 1, \tag{1}$$

with n the total number of events for each season and time interval, and R_{AWS} and R_{PWS} the rain recorded by the AWS and PWS respectively. Values above zero indicate an overestimation and below an underestimation of the PWS data. The CV is used to describe the dispersion of rainfall, ~~and~~ with values closer to zero suggesting greater consistency with the mean of the

reference and higher values indicating more dispersion and is defined as follows:

$$\mathrm{CV} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(R_{\mathrm{res},i} - \overline{R}_{\mathrm{res}})^2}}{\overline{R_{\mathrm{AWS}}}}, \tag{2}$$

with

$$R_{\mathrm{res}} = R_{\mathrm{PWS}} - R_{\mathrm{AWS}}. \tag{3}$$

The Pearson correlation coefficient describes the strength of the linear relation between the PWSs and the AWS with values ranging between -1 and 1 and was calculated between all events within a season and aggregation interval (including zeroes):

$$r = \frac{\mathrm{cov}(R_{\mathrm{PWS}}, R_{\mathrm{AWS}})}{\mathrm{sd}(R_{\mathrm{PWS}})\mathrm{sd}(R_{\mathrm{AWS}})}. \tag{4}$$

The linear regression line, fitted through the origin, is defined as follows:

$$\overline{R_{\mathrm{PWS}}} = a * \overline{R_{\mathrm{AWS}}}, \tag{5}$$

with a the slope calculated over all events:

$$a = \frac{\sum_{i=1}^{n} R_{\mathrm{AWS},i} R_{\mathrm{PWS},i}}{\sum_{i=1}^{n} (R_{\mathrm{AWS},i})^2}. \tag{6}$$

3.7 ~~Areal reduction factor~~

~~While the AWS represents a point measurement, which has a limited spatial footprint, it is compared to average rainfall over a domain (i.e. a cluster of PWSs around an AWS). With increasing domain area, the variation of areal precipitation becomes smaller than that of point precipitation. To account for the reduction in the magnitude of rainfall extremes over an area as compared to a point, areal reduction factors (ARF) can be applied. The ARF estimates areal rainfall percentiles from point rainfall percentiles. Overeem et al. (2010) and more recently Beersma et al. (2019) parameterized the ARF based on weather radar for the Netherlands. This reduction factor is a function of duration, area and return period, with rarer events having a stronger areal reduction. Equations 1a, 2, 3b, 4 and 5 from Beersma et al. (2019) are used to estimate the ARF~~Values close to one indicate a strong agreement with the reference dataset.

4 Results

After applying the HI and FZ filters and requiring a minimum data availability before aggregation, around 88% of the original dataset was kept. For 87 (0.5%) of the total timeseries used, at least one HI flag was attributed to a timestep. In 93% of the cases, no data was transferred for at least 15 min prior to the flagged HI timestep, suggesting that these flags may result from comparing data aggregated over longer time intervals ($\geq$ 15 min) to a 5-min timestep, potentially leading to mismatches and flagging data. For 5.8% of the timeseries, at least one FZ occurred. Around 15% of the PWSs were manually calibrated, with a median tipping volume of 0.117, with 95% of the calibrated tipping bucket volumes ranging between 0.09 and 0.203.

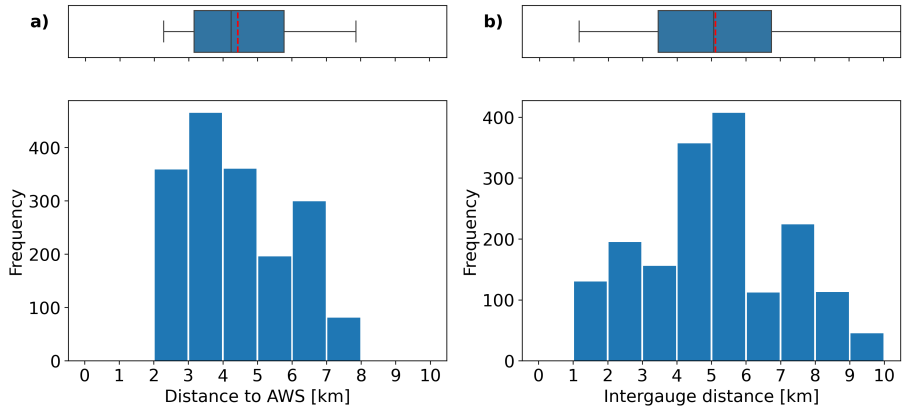

Figure 6. Histogram of a) distances between the PWS clusters and the AWSs per event and b) the inter-station distances between all selected PWS pairs within a cluster per event, both based on 1760 pairs. Vertical red dashed line indicates the mean distance, vertical black line the median, the left and right whiskers indicate the minimum and maximum distance and boxes the inter-percentile range (25^{th}–75^{th}).

4.1 Spatial sampling

The PWSs in this study were selected based on the closest distance from an AWS, without considering the uniformity of distribution around the AWS. Figure 6 shows that the average cluster distance towards an AWS is around ~~4.5~~4.4 km and that the average intergauge PWSs distance of all pairs within a cluster is around 5 km. This indicates that overall, the selected stations are not clustered at one location and represent a larger area. Variation in the distance towards the AWSs in Fig. 6a can be explained due to the location of the selected AWSs. Higher PWS network densities and associated shorter distances to the closest PWSs can be found in urban regions. However, most of the AWSs are located in rural regions. In addition, variability in Fig. 6 also partially results from fluctuating data availability and number of available PWSs, which increases over the years. The average number of PWSs within a cluster is 8.75.

4.2 ~~Bias~~Areal reduction effect

~~To highlight the severity of the bias in the PWSs, no MFBC factor was applied initially. The~~ Figure 7 shows a substantial decline in ARF with larger area sizes and shorter durations, with largest reductions for short durations. The decline becomes more prominent for longer return periods. For an area of 79 km^2 (based on a radius of 5 km towards an AWS) and a duration of 1 h, the ARF according to Beersma et al. (2019) varies between 0.88 and 0.82 for return periods of 2 and 50 years, respectively, while for 24 h the ARF varies between 0.96 and 0.92 for return periods of 2 and 50 years, respectively. The reduction becomes larger for a radius of 10 km, for example, the ARF is 0.78 and 0.70 for a duration of 1 h and a return period of 2 and 50 years, respectively. To convert the areal estimate from the PWSs into a point observation (reflecting the AWS), the PWS cluster average is adjusted using the inverse of the ARF, with the area based on a 10 km radius (the maximum distance between a PWS and an AWS used in the PWS selection procedure).

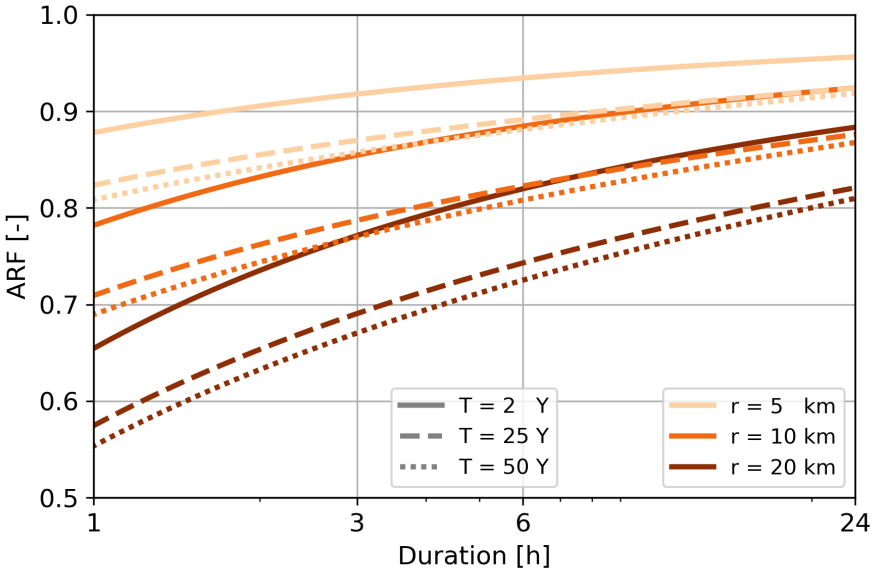

Figure 7. Areal reduction factor (left y-axis) as function of duration for area sizes with a radius of 5, 10 and 20 km (converted to an area) and return periods of T = 2, 25 and 50 years.

4.3 Bias

The relative bias of the ~~PWSs~~ non-adjusted PWS cluster average over multiple accumulation intervals was quantified by comparing it with AWSs nearby for the selected rainfall events (Table 2, each row indicating different accumulation intervals). Results indicate that without applying any ARF or DBC factor, on average, significant biases are present in rainfall observations from the PWSs. The underestimation is largest for accumulation intervals of 1 h (around 36%). The magnitude of the bias reduces over longer intervals, towards an average underestimation of around 19% for accumulation intervals of 24 h. In addition, a seasonal dependency is visible for the bias.

~~These results indicate the need for a MFBC factor, with more significant adjustment required for shorter accumulation intervals.~~ It is important to make a distinction in the sources of bias, to avoid correcting non-instrumental related errors. Due to the setup of this study, which makes use of PWSs within a maximum 10 km distance from an AWS, the bias can be divided into two categories: 1) bias resulting from the spatial variation in rainfall extremes and 2) an instrumental bias.

The ARF accounts for the spatial variability of rainfall extremes and illustrate that the bias is partially caused by category 1. The ARF was applied to compensate for the spatial variation of rainfall extremes and to fairly compare the rainfall measured by a cluster of PWSs with one AWS. The largest areal reductions are visible for 1 h accumulation (Fig. 7 and Table 2), reducing the relative bias on average over all seasons from -36% to -22%. During the winter months on average the lowest rainfall intensities and the least spatial variability of rainfall occurs (Fig.9), resulting in the smallest areal reduction. The ARFs have a

limited effect on the 24 h events, with an average reduction of 2 percent points in the bias over all seasons. The remaining bias for time intervals of 3 h over all seasons and longer is on average around 16%.

The remaining biases are part of the second category and indicate the need for a DBC factor to adjust the instrumental biases. To compensate for this instrumental bias, the ~~MFBC of 1.24, proposed by De Vos et al. (2019) as part of their quality control algorithm,~~ DBC factor of 1.22 was applied. After application of ~~the MFBC~~this DBC factor, the underestimation for 1 h intervals over all seasons reduces to an average of ~~21%~~5%, while for ~~24 h~~ 3 h and longer intervals it converges towards zero ~~.~~ or in a slight overestimation of the rain. The remaining bias is within the expected uncertainty of rainfall observations. This is supporting evidence that the ~~MFBC works effectivelyfor longer accumulation intervals, however, a significant underestimation remains for 1-h intervals, requiring a higher MFBC factor to correct for the substantial underestimation~~DBC of 1.22 works effectively.

A small part of the ~~raw~~ dataset obtained from Netatmo (5.38% of the selected events' total time steps) ~~was not available. Part of the underestimation with respect to the AWSs can be attributed to these missing time steps in the PWS dataset. Missing observations in the PWS dataset result in lower rainfall estimates compared to a complete dataset. However, the extent of the underestimation cannot be determined, as these time steps are missing~~were not included, either suggesting that data was missing or that the system suffered connection issues, resulting in irregular data transfer (longer than approximately 5 min). It is expected that the effect of this on the bias is limited, as most of the data is likely not missing, rather caused by irregular data transfer and or connection interruption between the rain gauge and indoor module (see supporting information Tables A1 and A2).

4.4 Seasonal dependence

For the selected events, a seasonal dependency is visible for the performance of the PWSs (Figs. 8 and 9). The seasonal effect is most pronounced for shorter accumulation intervals (1 h and 3 h), with the best performance of the PWSs in winter and autumn and worst in summer and spring (Fig. 9). Events in winter show the lowest variability of the PWS observations compared with the AWS observations (e.g. ~~CV of 0.27 and 0.20~~ average CV of 0.30 and 0.21 for accumulation intervals of 1 h and 3 h, respectively). While the CV is ~~slightly~~ larger for autumn (~~0.34 and 0.24~~0.41 and 0.26), the correlation between the PWSs and AWS is higher during autumn compared to winter (Fig. 9b). Winter in the Netherlands is mainly characterized by larger, more persistent rainfall systems, which have a longer decorrelation distance (80 km for 1-h aggregations) (Van de Beek et al., 2012). In addition, the error bars for winter are smaller compared to the other seasons, showing that there is more consistency between the individual PWSs (see Fig. 9 and supporting information Fig.B1 for a complete overview of all seasons and accumulation intervals). For that reason, it is expected that the spatial sampling errors were minimized during winter, indicating that other factors, such as solid rain, caused the lower correlation.

Figure 9a shows that for all seasons, the CV reduces over longer accumulation intervals. For 1-h intervals the CV varies over the different seasons between ~~0.27 and 0.55~~0.30 and 0.54, while for 24 h the CV shows lower variability, with values varying between ~~0.15~~ 0.14 and 0.27. Similarly, the correlation coefficient increases from values ranging from ~~0.36 to 0.71~~ 0.43 to 0.74 for 1-h intervals to a range from ~~0.70 to 0.83~~ 0.75 to 0.86 for 24-h intervals. This indicates that, for longer

Table 2. Relative bias calculated before and after applying an areal reduction factor based on Beersma et al. (2019) and correcting for the instrumental bias over the 110 (i.e. 10 rainfall events x 11 AWSs) selected rainfall events per season and interval.

	Relative bias [-]			
Interval	**DJF**	**MAM**	**JJA**	**SON**
No ARFs or DBC applied				
1 h	−0.33	−0.39	−0.38	−0.34
3 h	−0.22	−0.26	−0.27	−0.21
6 h	−0.20	−0.23	−0.21	−0.20
24 h	−0.18	−0.21	−0.19	−0.18
ARFs applied				
1 h	−0.20	−0.26	−0.22	−0.19
3 h	−0.15	−0.19	−0.16	−0.13
6 h	−0.16	−0.19	−0.12	−0.14
24 h	−0.17	−0.20	−0.16	−0.16
ARFs and DBC applied				
1 h	−0.02	−0.10	−0.05	−0.01
3 h	0.04	−0.01	0.03	0.07
6 h	0.03	−0.01	0.07	0.05
24 h	0.01	−0.02	0.03	0.03

accumulation intervals, rainfall observations from PWSs exhibit less variability and more agreement with those from AWSs. Data transferring and processing errors ~~reduces~~ reduce for longer accumulation intervals, as the effect of ~~incorrectly~~ attributing rainfall to ~~a certain~~ an erroneous time stamp decreases. This takes place for example when the connection between the indoor and outdoor module is temporarily interrupted, potentially attributing rainfall to a timestamp when there is a connection again, as a consequence aggregating it over a longer time interval than approximately 5 min (see supporting information Table A2). Within a cluster of PWSs, variation in rainfall is observed, however, using the average rainfall of each PWS cluster shows a large agreement with the AWS. Overall, the average of the cluster of PWSs largely follows the 1:1 line, with slopes of the fitted lines indicating slight underestimation or overestimation by varying between ~~0.95 and 1.02~~ 0.97 and 1.03 for 24 h. A seasonal effect is limited on the slope for durations of 3 h and longer. Furthermore, high correlations, low CV-values and low biases are found for both winter and autumn, indicating that there is a good agreement with the AWSs (Fig. 9).

~~Relative bias calculated before and after applying the MFBC of 1.24 from De Vos et al. (2019) over the 110 (i.e. 10 rainfall events x 11 AWSs) selected rainfall events per season and interval. **Interval** 1 h −0.33 −0.39 −0.38 −0.34 3 h −0.22 −0.26 −0.27 −0.21 6 h −0.20 −0.23 −0.21 −0.20 24 h −0.18 −0.21 −0.19 −0.18 1 h −0.17 −0.24 −0.23 −0.18 3 h −0.03 −0.08 −0.09 −0.03 6 h −0.01 −0.05 −0.02 −0.01 24 h −0.02~~

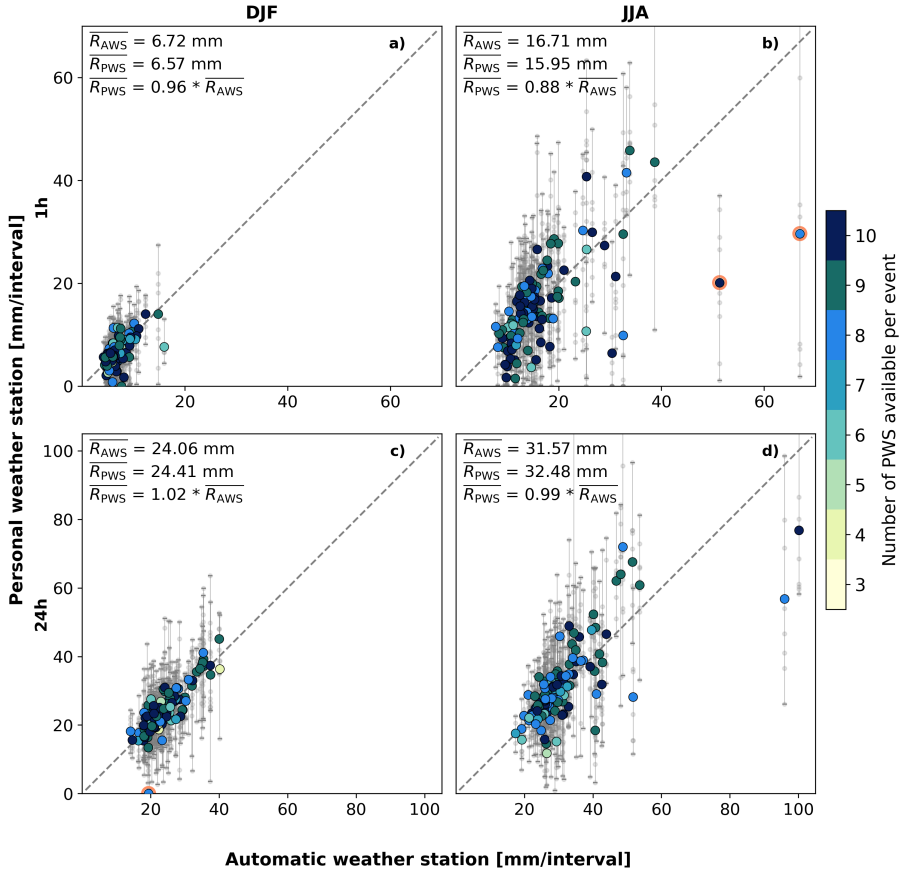

Figure 8. Scatter plots of filtered PWS rainfall accumulations against AWS records for the winter (a, c) and summer (b, d) seasons and accumulation intervals of 1 h (a, b) and 24 h (c, d). The large ~~blue~~ coloured dots indicate the average of a cluster of PWS against one AWS, the vertical bars indicate the minimum and maximum of that cluster of PWSs. The ~~blue gradient indicates~~ colours indicate the number of PWSs used to calculate the mean, minimum and maximum rainfall. The small grey dots indicate one individual PWS against an AWS. Orange circles indicate examples of outliers. The $\overline{R_{\mathrm{AWS}}}$ and $\overline{R_{\mathrm{PWS}}}$ represent the average rainfall over the selected events recorded by the AWS and PWS, respectively. $\overline{R_{\mathrm{PWS}}} = a * \overline{R_{\mathrm{AWS}}}$ represents the linear regression line, fitted through zero, with a indicating the slope.

4.5 Outlier identification

Figure 8 shows that, for certain individual events, the PWSs report rainfall amounts which deviate strongly from those observed by the AWSs. We further investigated the cause of some of these outliers, which are indicated with orange circles.

It can be seen in Fig. 8c that one of the selected events showed (almost) zero precipitation measurements according to the cluster of PWSs, while the AWS nearby recorded precipitation. These outliers occurred during winter and were also observed for the 6 h accumulation interval. In the Netherlands, winter has on average around 34 days with a minimum temperature below 0°C (KNMI, 2024), which is outside the optimal temperature range of the PWSs. For the event with the outliers in

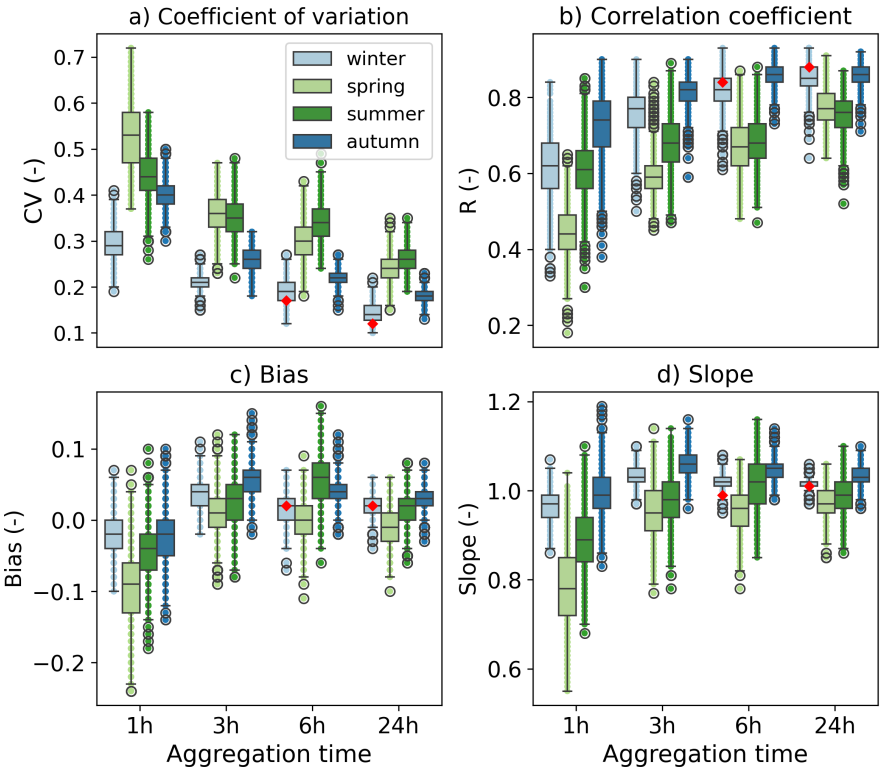

Figure 9. Coefficient of variation (a), correlation coefficient (b), bias (c) and slope (d) of filtered PWS rainfall accumulations against AWS for different seasons and accumulation intervals. ~~Hollow circles~~ The lower and upper whiskers indicate the minimum and maximum of each metrics and the boxes the inter-percentile range (25^{th}-75^{th}). The red diamonds indicate the values in winter after ~~applying a~~ discarding events based on attributed temperature ~~filter~~flags. The estimated uncertainty is obtained using bootstrapping (1000 iterations, with replacement).

Fig. 8c, the maximum temperature was below -1.4 °C. It is not possible to unambiguously determine whether precipitation is solid based solely on temperature. However, a temperature-based flag can provide end users with an indication that the rainfall observations may be subject to uncertainty. Flags were assigned to timeseries where the corresponding AWS recorded temperatures below freezing. If these events were discarded by a temperature filter that filters stations when temperatures are below a certain threshold for a certain duration, the values of r and CV for 6 h for winter would have improved to ~~0.81 and 0.17~~0.83 and 0.18, respectively. For 24 h this would have been r = ~~0.86~~ 0.88 and CV = 0.12. However, as these are only two points and the two intervals have some overlap in time at the same location, no statistically valid conclusion can be drawn from this.

~~The~~ During summer and spring the highest rainfall accumulations were observed ~~during summer and spring, with an average intensity of more than~~ by the AWSs, with intensities exceeding 35 mm h^{-1}. An example of the low performance of PWSs for two high rainfall events in summer can be seen from Fig. 8b. These events ~~influence~~ likely skew the overall performance during

these months. ~~This suggests PWSs have difficulties recording higher rainfall intensities, which is known for~~ A large spread is observed within the cluster, especially for the highest event in summer in Fig. 8b, indicating a large variation in the spatial rainfall distribution. This suggests that the differences between the cluster of PWSs and AWS are not necessarily only related to instrumental limitations, rather due to the spatial rainfall distribution. In addition, tipping bucket mechanisms ~~. In addition, are known for having difficulties in recording higher rainfall intensities.~~ PWSs only sends rainfall data to the ~~netatmo~~ Netatmo platform twice per 10 min, which is considerable lower than the recording interval of AWSs (12 sec). More than one third of the rainfall during these events fell within 10 min, exceeding intensities of 75 mm h^{-1} during that interval. However, it is unknown if the rain was evenly spread within these 10 min, or mostly occurred during a shorter period of time and if this measurement range of the PWSs was indeed exceeded.

5 Discussion

5.1 ~~Areal reduction effect~~Bias

~~Figure 7 shows a substantial decline in ARF with larger area sizes and shorter durations, with largest reductions for short durations. The decline becomes more prominent for longer return periods. For an area of 64 km^2 (based on the~~

Adjusting the PWS dataset with ARFs and correcting for the instrumental bias using a DBC factor, reduces the bias. The remaining bias can either indicate that for example a higher DBC factor is required to correct for the substantial underestimation or that the ARFs are not able to fully account for the spatial variability of the rainfall.

5.1.1 Spatial sampling errors

Rainfall exhibits spatial variability, which is related to the temporal scale and rainfall intensity, with shorter temporal resolutions and higher intensities often associated with greater heterogeneity. The decorrelation distance of rainfall is typically much higher in winter compared to summer in the Netherlands. Specifically, for shorter aggregation times this holds (Van de Beek et al., 2012; Van Leth . The average distance of around 4.4 km from a PWS cluster ~~to the corresponding AWS) and a duration of~~ towards the closest AWS was below the decorrelation distance corresponding to a 1-h aggregation interval found by Van Leth et al. (2021) and Van de Beek et al. (2012). While this limits the errors related to spatial sampling, it is expected that this effect remains most pronounced for events in for example summer and spring and at shorter aggregation intervals (i.e. 1 h~~, the ARF according to Beersma et al. (2019) varies between 0.89 and 0.83 for return periods of 2 and 50 years, respectively, while for 24 h the ARF varies between 0.96 and 0.93 for return periods of 2 and 50 years, respectively. The larger areal reductions for shorter intervals are also visible in the average bias in Fig. 7. ~~). Lowering the radius reduces the number of available PWSs and consequently increases the uncertainty. From the error bars in Fig. 8 variation within a cluster of PWSs was observed, with the largest variation found in spring and summer. For small-scale convective rainfall events the distance towards the closest AWS might still have been too large, resulting in variations of the observed rainfall by the PWSs nearby. Radar images support

that for several events the difference between AWS and PWSs are caused by spatial distribution of the rainfall (see supporting information Figs.D1, D2 and D3).

~~Areal reduction factor (left y-axis) as function of duration for area sizes with a radius of 5, 10 and 20 km (converted to an area) and return periods of T = 2, 25 and 50 years. The right y-axis shows the averaged relative bias per duration.~~ To account for spatial variability, an ARF was estimated for each event. However, such factors are based on areal rainfall climatology, representing an average behaviour and not tied to one specific event. The radius of 10 km, which was based on the maximum possible distance of a PWS towards an AWS, was used for the area to estimate the ARFs. For each event, a different PWS cluster was used, potentially representing a larger or smaller area and therefore requiring a different corresponding adjustment of the ARF. A smaller or larger radius has a large effect on the ARF (Fig.7).

6 ~~Discussion~~

5.1 ~~Bias correction factor~~

~~A significant bias is present in the PWS dataset.~~

5.0.1 Instrumental bias

An instrumental bias of -18% was identified by De Vos et al. (2017) using an experimental setup that minimized the spatial sampling errors. This suggests a DBC factor of 1.22 to compensate the instrumental bias. However, other studies came up with different DBC factors. From De Vos (2019) a bias correction factor of 1.13 came out for a 1-month dataset covering the Netherlands, which is different from the 1-year calibration set from the same study for Amsterdam only (~~MFBC~~ DBC of 1.24)~~,~~ ~~which is part of the PWSQC algorithm and was used in this study~~. Alternatively, Overeem et al. (2024a) used a ~~MFBC~~ DBC of 1.063 for a pan-European dataset. Neither of the ~~MFBC~~ DBC factors (1.063 or 1.13) would have been able to fully compensate for the bias present in the dataset used in this research. This difference might be caused by two main reasons. First, both ~~MFBC factors were based~~ DBC factors were a bulk correction factor tuned on different reference datasets. De Vos et al. (2019) utilized gauge-adjusted radar values. Radars indirectly measure rainfall, which might not be representative for rainfall at the ground. In addition, radars do not measure on a grid, rather the values are interpolated, adding extra uncertainty. Spatially adjusting radars with rain gauges improves the overall quality, however, substantial errors remain. Secondly, this research focused on the highest rainfall events over a longer period of time, while both De Vos et al. (2019) and Overeem et al. (2024a) did not distinguish between certain types of rain events, rather looking at a full month or year of rainfall. The performance of tipping buckets is a non-linear function of rainfall intensity (Niemczynowicz, 1986; Humphrey et al., 1997). It requires time for the tipping bucket mechanism to reposition itself after a tip. Higher intensities enhance this problem, resulting in an increased undercatch. Therefore, the type of dataset and the included rainfall events play a role in the performance. Applying a ~~MFBC factor of 1.24~~ DBC factor of 1.22 over the dataset almost eliminates the bias~~for accumulation intervals of 24 h. The remaining bias in 1 h intervals may be attributed to predominating high-intensities~~. The study of De Vos et al. (2017) considered only

a few months around the spring period, specifically from 12 February to 25 May 2016, which may have influenced the bias reported in that study.

5.0.2 Manual calibration

Around 1/7 of the PWSs used in this study (15%) were manually calibrated by their owner. However, it is unknown what the accuracy of such a manual calibration is. The number of tips was determined for each manually calibrated PWS and converted to the original default value of 0.101 mm. On average, there is a 4% decrease in the observed rainfall by the PWS cluster, resulting in a slightly increased underestimation or slightly decreased overestimation by the PWSs. The CV values slightly improve with an average of 0.01, while the change in r is negligible, see supporting information, Table C1.

5.1 Quality control

The quality control algorithm from De Vos (2019) utilized in this research improves the overall performance of the PWSs (see supporting information, Fig.E1). However, there are some limitations to this algorithm. The algorithm works only if there are enough neighbouring stations within 10 km, limiting the usefulness for less densely populated regions. That said, for this study only a small percentage (0.66%) was discarded from this dataset due to an insufficient number of neighbouring PWSs. While the number of currently active PWSs is quite high in Europe, they are not evenly distributed (see Fig.1 for the network density of PWSs across Europe). For that reason, regions with a less dense network of PWSs are expected to have a higher percentage of discarded stations due to insufficient neighbours (around 35% within Europe). Alternatively, other data sources (such as gauges or weather radars operated by national meteorological or hydrological services) can be employed for quality control, such as employed in the QC algorithm by Bárdossy et al. (2021). In addition, insufficiently calibrated PWSs which record higher rainfall at each time stamp, are not discarded by the HI filter if a certain threshold is not reached. These differences become more apparent when accumulated over longer periods. With a dynamic bias correction factor this could have been adjusted.

Another limitation of a quality control algorithm that does not use auxiliary data is that if all PWSs provide faulty observations, these timestamps are not flagged, consequently giving a wrong signal. This was observed for two events during winter for 6 and 24 h accumulation intervals, when none of the stations recorded any precipitation during an event, while the AWSs did observe precipitation (e.g. Fig. 8c). During winter, solid precipitation can occur, which can result in an undercatch by the PWSs, as these are not heated and consequently work best for temperatures above freezing point. Results from Overeem et al. (2024a) also suggested that PWSs are not able to capture solid precipitation. Quality control algorithms based on a reference dataset, such as from Bárdossy et al. (2021), would have filtered these PWSs. Alternatively, a temperature filter could be developed, without using auxiliary weather stations. The temperature module present at the PWS can be utilized for this.

5.2 ~~Spatial sampling errors~~

~~The decorrelation distance of rainfall is typically much higher in winter compared to summer in the Netherlands. Specifically, for shorter aggregation times this holds (Van de Beek et al., 2012; Van Leth et al., 2021). The average distance of around 4.5 km from a PWS cluster towards the closest AWS was below the decorrelation distance corresponding to a 1-h aggregation interval found by Van Leth et al. (2021) and Van de Beek et al. (2012). While this limits the errors related to spatial sampling, it is expected that this effect remains most pronounced for events in summer and at shorter aggregation intervals (i.e. 1 h). Lowering the radius reduces the number of available PWSs and consequently increases the uncertainty. From the error bars in Fig. 8 variation within a cluster of PWSs was observed, with the largest variation found in spring and summer. For very small-scale convective rainfall events the distance towards the closest AWS might still have been too large, resulting in variations of the observed rainfall by the PWSs nearby. This is also visible from the ARF in Fig. 7, which declines more for rarer events.~~

5.2 ~~Uncertainty of AWSs~~

~~The 10-min AWS dataset is used as a reference, however, it is not an absolute truth. Brandsma (2014) compared the AWS network and manual rain gauge network over the Netherlands and concluded that the AWS network underestimates rainfall with 5-8% annually, with higher underestimation in winter (7.7%) than summer (5.0%). The undercatch is nonlinear with intensity, with larger intensities resulting in less underestimation. These uncertainties are not taken into account in this research.~~

6 Conclusions

This study provides insight into the ~~uncertainty arising from personal weather stations (PWSs) rainfall estimates~~ systematic errors across the personal weather station (PWS) network during high-intensity rainfall events ~~,~~ by performing a ~~systematic long-term analysis~~ comprehensive analysis over six years. The analysis focuses on the most intense rainfall observations for a large number of events (1760) over six years (2018-2023) in the Netherlands. PWS data were evaluated against rainfall measurements from automatic weather stations (AWSs). These events were selected over different meteorological seasons (winter, spring, summer, autumn), durations (1, 3, 6 and 24 h) and AWSs (11 locations spread over the country). ~~While PWSs are subject to connection failure which reduces data availability, around 96% of the stations were available over all events.~~ PWS data were filtered with a quality control (QC) algorithm, utilizing neighbouring PWSs. After QC, around 88% of the original dataset was kept. ~~Metrics~~ To reduce uncertainty from single stations, metrics were calculated over a cluster of PWSs, rather than individual stations~~, reducing the impact of uncertainties arising from single stations~~.

Results showed that PWSs severely underestimate rainfall. ~~However, applying a mean field~~ A seasonal effect is visible in the bias, specifically for shorter accumulation intervals, with largest biases for summer and spring. A part of this bias can attributed to the spatial distribution of rainfall. To account for this, areal reduction factors (ARFs) were applied. This seasonal dependence minimizes after applying ARFs. In addition, a default bias correction (~~MFBC~~DBC) factor of ~~1.24 as part of the PWSQC algorithm~~ 1.22 was used to compensate for the instrumental bias. The DBC factor substantially reduces this bias for intervals of 3 h and longer, indicating that the ~~MFBC~~ DBC of 1.063 by Overeem et al. (2024a) (European dataset) and 1.13 proposed by De Vos et al. (2019) (Netherlands dataset) are not able to account for high-intensity rainfall events. A seasonal ~~effect is~~

visible in the bias, specifically for shorter accumulation intervals, with largest biases for summer and spring. This seasonal and and temporal dependence is also seen from the correlation coefficient (r) , and coefficient of variation (CV)and the slope of the fitted linear regression line. Outliers in winter seemed to have been caused by freezing temperatures (solid precipitation). For that reason, it is recommended to further analyse the impact of temperature or solid precipitation on the performance of the PWS. The temperature module available in PWSs can be used for this. In addition, PWSs were not able to capture did not observe the most intense rainfall events, with high intensities over a relative short amount of time (e.g. > 75 mm h^{-1} within 10 min). These highest intensities occurred during summer and spring, with events that typically occur once in 10 years or even longer return periods. Rain gauges from PWSs used in this research utilize a tipping bucket mechanism, which are known to suffer from non-linear underestimation errors with increasing intensities, potentially causing these outliers. In addition, the The spatial footprint of these high-intensity rainfall events is unknown, potentially often small, influencing errors related to spatial sampling due to the average distance of 4.54.4 km towards the nearest AWS. The areal reduction factor shows a strong decline for longer return periods, which partially explains the larger bias for rarer events, such as events with a return period of 50 years, for which the reduction can be up to 17% for 1 h durations. This reduction decreases to 7.0% for a duration of 24 hThis suggests the need for a high-density observation network to reliably capture localized rainfall extremes. Rain gauges from PWSs used in this research utilize a tipping bucket mechanism, which are known to suffer from non-linear underestimation errors with increasing intensities, potentially contributing in these outliers. To quantify this, experimental studies are necessary to limit other sources of errors. Performance of the PWSs improved over longer accumulation intervals, resulting in a r of 0.83 and 0.83 0.85 and 0.86 and a CV of 0.15 0.14 and 0.18 for 24 h for winter and autumn, respectively.

With the high density of PWSs in the Netherlands (around 1 PWS per 9 km^2), there is a clear potential of using PWSs. This will also be the case for other regions in Europe that have a relative high coverage of PWSs. The accuracy however depends on the desired temporal resolution, season and intensity. Although applying a MFBC DBC factor reduces or even completely compensates for the underestimation, we recommend to further investigate the dynamic response of these stations at different intensities to enable dynamic calibration and consequently minimize non-linear errors related to this.

Code and data availability. The automatic weather stations from the Royal Netherlands Meteorological Institute (KNMI) is freely available on the KNMI data platform for 10 min intervals: https://dataplatform.knmi.nl/dataset/neerslaggegevens-1-0. The hourly validated dataset is available at: https://www.daggegevens.knmi.nl/klimatologie/uurgegevens. Part of the quality control of the Netatmo gauge data, i.e. the faulty zeroes and high-influx filters, are based on the PWSQC algorithm written in R language and can be found at: https://github.com/LottedeVos/PWSQC. Dataset with Netatmo PWS rainfall data can be found at: https://doi.org/10.4121/caa0a93a-effd-4574-95ec-cd874ca97c05.v1 (Rombeek et al., 2024).

Table A1. Example of the Netatmo software that resamples data to regular 5-min intervals, by assigning it to the next full five-minute interval.

Raw		Aggregated	
Time	Rain (mm)	Time	Rain (mm)
29-10-2024 05:54:48	0.0	29-10-2024 05:55:00	0.0
29-10-2024 05:59:56	0.303	29-10-2024 06:00:00	0.303
29-10-2024 06:05:03	1.313	29-10-2024 06:05:00	Not included by Netatmo
29-10-2024 06:09:58	2.626	29-10-2024 06:10:00	3.393
29-10-2024 06:15:07	2.626	29-10-2024 06:15:00	Not included by Netatmo
29-10-2024 06:20:01	1.111	29-10-2024 06:20:00	2.626
29-10-2024 06:25:09	0.505	29-10-2024 06:25:00	1.111
29-10-2024 06:30:17	1.616	29-10-2024 06:30:00	0.505
		2024-19-29 06:35:00	1.616

Appendix A: Netatmo data processing

Table A2. Example of when there was likely a temporary connection interruption between the indoor module and rain gauge. Rainfall will likely be attributed to a timestamp when there is a connection again, resulting in a longer temporal resolution, e.g. at 20:26:21 the rainfall is likely aggregated over 22 min.

Raw		**Aggregated**	
Time	Rain (mm)	Time	Rain (mm)
2024-29-10 19:50:59	0.303	2024-29-10 19:55:00	0.303
2024-29-10 19:56:07	0.101	2024-29-10 20:00:00	0.101
2024-29-10 20:01:02	0.404	2024-29-10 20:05:00	1.111
2024-29-10 20:04:08	0.707	2024-29-10 20:10:00	Not included by Netatmo
2024-29-10 20:26:21	7.777	2024-29-10 20:15:00	Not included by Netatmo
		2024-29-10 20:20:00	Not included by Netatmo
		2024-29-10 20:25:00	Not included by Netatmo
		2024-29-10 20:30:00	7.777

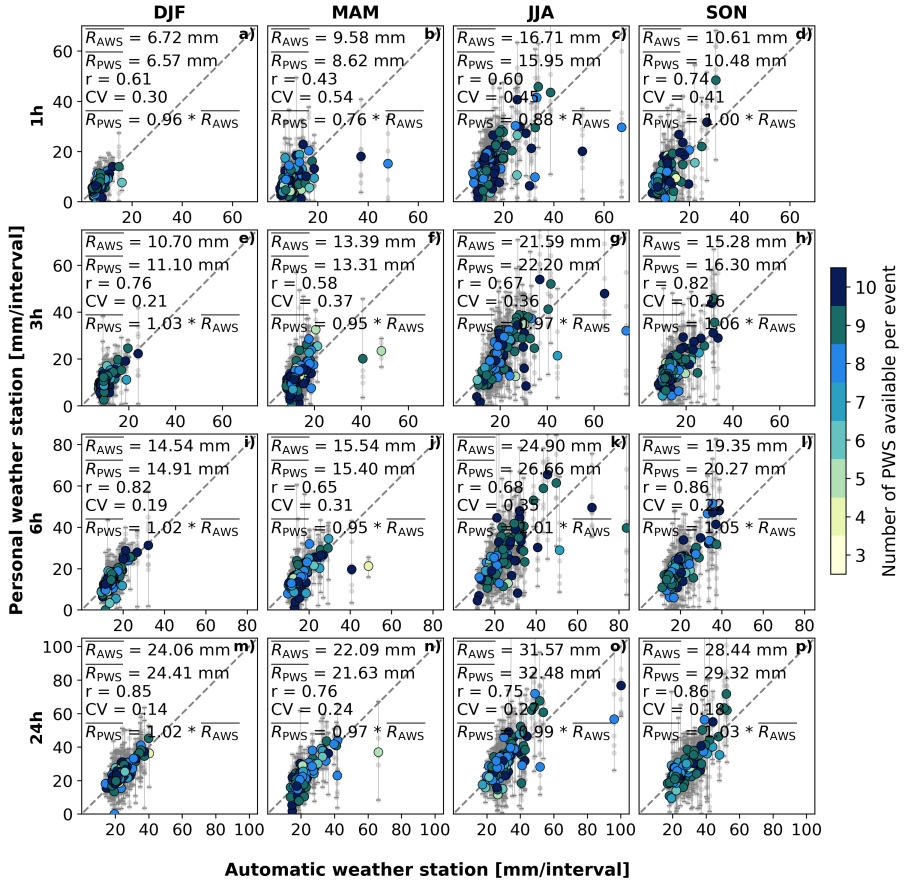

Figure B1. Scatter plots of filtered PWS rainfall accumulations against AWS for the winter (a, e, i and m), spring (b, f, j and n), summer (c, g, k and o) and autumn (d, h, l and p) seasons and accumulation intervals of 1 h (a-d), 3 h (e-h), 6 h (i-l) and 24 h (m-p). The large ~~blue~~ coloured dots indicate the average of a cluster of PWS against one AWS, the vertical bars indicate the minimum and maximum of that cluster of PWSs. The ~~blue gradient indicates~~ colours indicate the number of PWSs used to calculate the mean, minimum and maximum rainfall. The small grey dots resemble one individual PWS against an AWS. The $\overline{R_{\mathrm{AWS}}}$ and $\overline{R_{\mathrm{PWS}}}$ represent the average rainfall over the selected events recorded by the AWS and PWS, respectively. $\overline{R_{\mathrm{PWS}}} = a * \overline{R_{\mathrm{AWS}}}$ represents the linear regression line, fitted through zero, with a indicating the slope.

Appendix B: Overview of all seasons and accumulation intervals

Table C1. Relative bias, coefficient of variation (CV) and correlation (r) of filtered PWS rainfall accumulations against AWS for different seasons and accumulation intervals. PWSs with manual calibrated tipping volumes were converted to the original default value of 0.101 mm.

Interval	DJF	MAM	JJA	SON
		Bias		
1h	−0.06	−0.13	−0.08	-0.05
3h	0.0	−0.04	0.0	0.03
6h	−0.01	−0.05	0.04	0.01
24h	−0.02	−0.05	−0.01	0.0
		CV		
1h	0.29	0.54	0.44	0.39
3h	0.20	0.36	0.35	0.26
6h	0.18	0.30	0.34	0.22
24h	0.14	0.24	0.26	0.18
		r		
1h	0.62	0.43	0.61	0.75
3h	0.77	0.57	0.67	0.81
6h	0.82	0.65	0.67	0.86
24h	0.86	0.76	0.75	0.86

Appendix C: Calibration effect

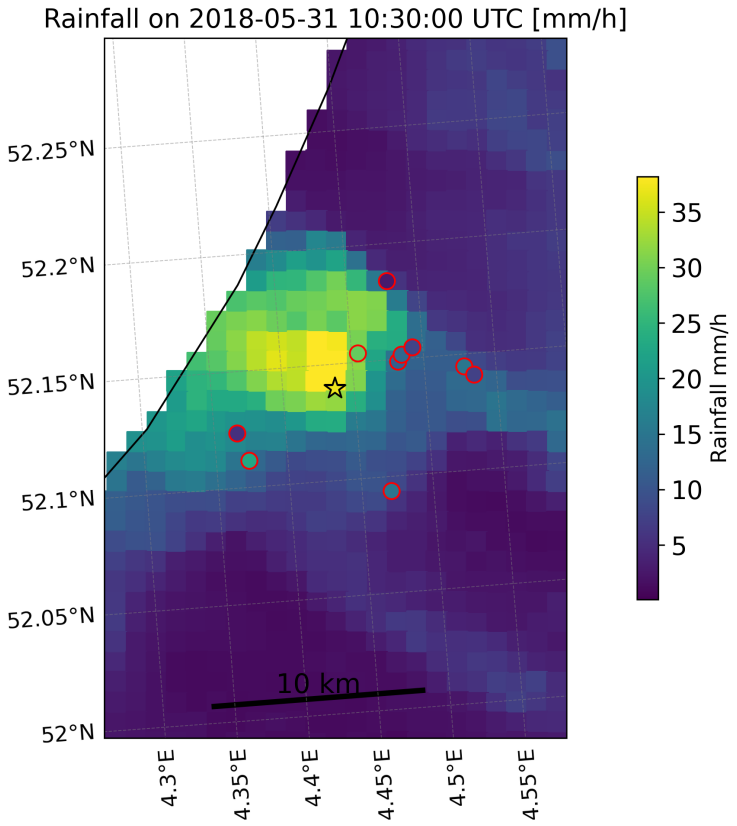

Figure D1. Rainfall distribution on May 31st, 2018 at 10:30 UTC, based on 1-h accumulated rainfall from the gauge-adjusted radar product (Overeem et al., 2009b). The asterisk indicates the location of the AWS which measured 39 mm in one hour, whereas circles with red borders represent the locations of the PWSs. The fill colour of both the asterisk and circles represents the recorded rainfall at the specific rain gauge.

Appendix D: Spatial distribution rainfall

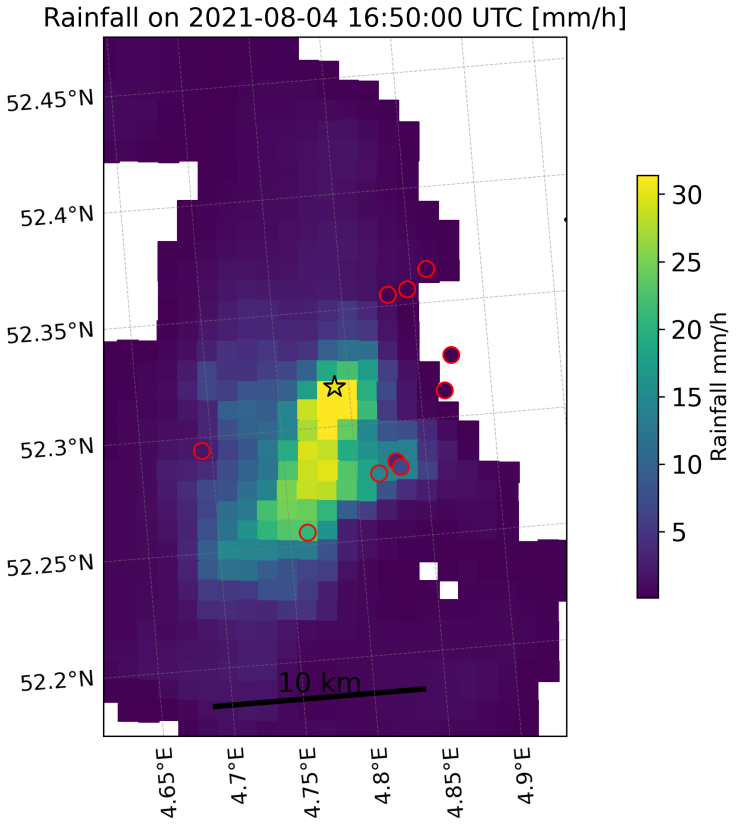

Figure D2. Rainfall distribution on August 4th, 2021 at 16:50 UTC, based on 1-h accumulated rainfall from the gauge-adjusted radar product (Overeem et al., 2009b). The asterisk indicates the location of the AWS which measured 30 mm in one hour, whereas circles with red borders represent the locations of the PWSs. The fill colour of both the asterisk and circles represents the recorded rainfall at the specific rain gauge.

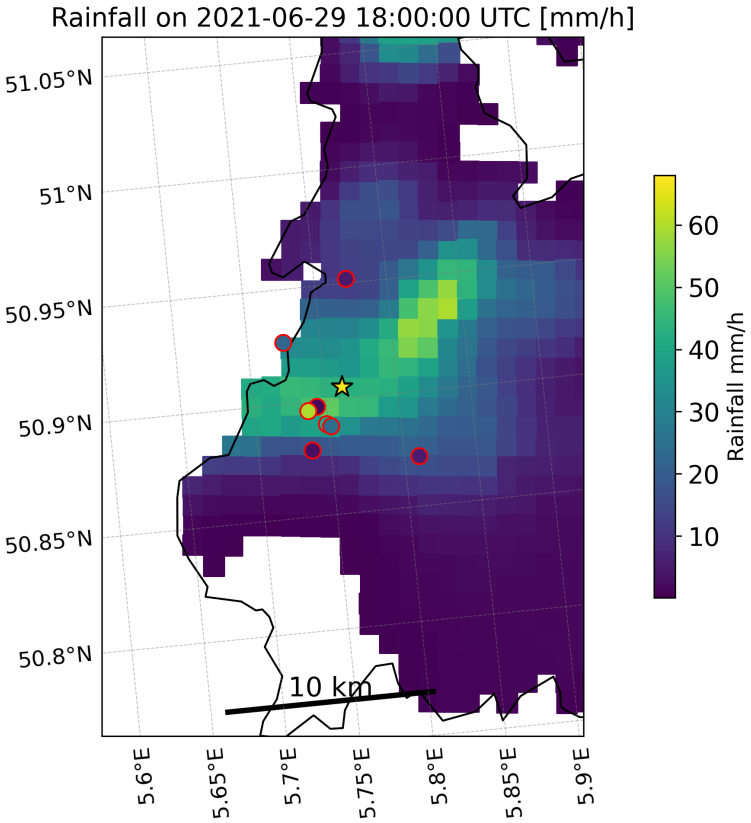

Figure D3. Rainfall distribution on June 29th, 2021 at 18:00 UTC, based on 1-h accumulated rainfall from the gauge-adjusted radar product (Overeem et al., 2009b). The asterisk indicates the location of the AWS which measured 67 mm in one hour, whereas circles with red borders represent the locations of the PWSs. The fill colour of both the asterisk and circles represents the recorded rainfall at the specific rain gauge.

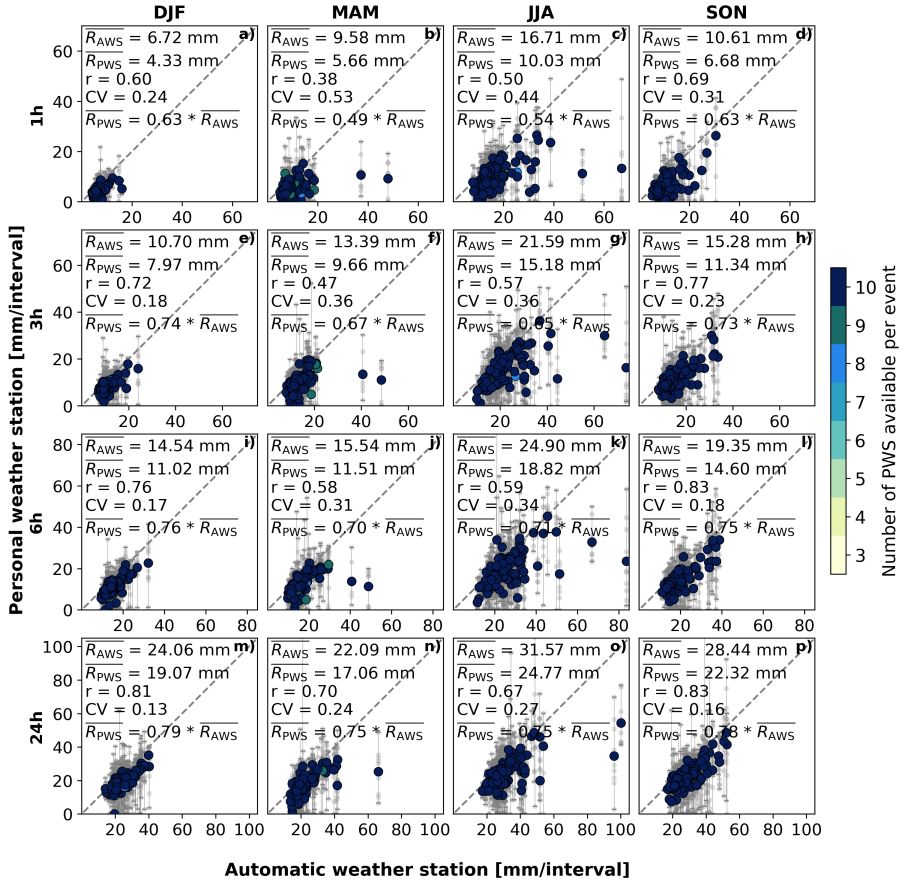

Figure E1. Scatter plots of raw PWS rainfall accumulation data against AWSs for the winter (a, e, i and m), spring (b, f, j and n), summer (c, g, k and o) and autumn (d, h, l and p) seasons and accumulation intervals of 1 h (a-d), 3 h (e-h), 6 h (i-l) and 24 h (m-p). The large ~~blue~~ coloured dots indicate the average of a cluster of PWS against one AWS, the vertical bars indicate the minimum and maximum of that cluster of PWSs. The ~~blue gradient indicates~~ colours indicate the number of PWSs used to calculate the mean, minimum and maximum rainfall. The small grey dots resemble one individual PWS against an AWS. The $\overline{R_{\mathrm{AWS}}}$ and $\overline{R_{\mathrm{PWS}}}$ represent the average rainfall over the selected events recorded by the AWS and PWS, respectively. *r* and CV indicate the correlation coefficient and the coefficient of variation. $\overline{R_{\mathrm{PWS}}} = a * \overline{R_{\mathrm{AWS}}}$ represents the linear regression line, fitted through zero, with ~~a~~ *a* indicating the slope.

Appendix E: Raw PWS data

565 *Author contributions.* NR: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualisation, writing. MH: conceptualization, funding acquisition, methodology, writing (review & editing). AD: software, writing (review & editing). RU: conceptualization, funding acquisition, methodology, writing (review & editing)

Competing interests. One of the authors (MH) is member of the editorial board of HESS.

Acknowledgements. The authors would like to thank Davide Wüthrich for discussions and for proofreading the manuscript and Claudia Brauer for providing insightful feedback on the results, which helped to improve this paper. This work is part of the Perspectief research programme Future flood risk management technologies for rivers and coasts with project number P21-23. This programme is financed by Domain Applied and Engineering Sciences of the Dutch Research Council (NWO).

References

Arnaud, P., Lavabre, J., Fouchier, C., Diss, S., and Javelle, P.: Sensitivity of hydrological models to uncertainty in rainfall input, Hydrological Sciences Journal–Journal des Sciences Hydrologiques, 56, 397–410, 2011.

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.

Beersma, J., Hakvoort, H., Jilderda, R., Overeem, A., and Versteeg, R.: Neerslagstatistiek en-reeksen voor het waterbeheer 2019, vol. 19, STOWA, 2019.

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, 2004.

Beven, K.: Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication, Hydrological Sciences Journal, 61, 1652–1665, 2016.

Brandsma, T.: Comparison of automatic and manual precipitation networks in the Netherlands, KNMI, 2014.

Brandsma, T., Beersma, J., van den Brink, J., Buishand, T., Jilderda, R., and Overeem, A.: Correction of rainfall series in the Netherlands resulting from leaky rain gauges, Tech. rep., Technical Report TR-387. De Bilt: KNMI, 2020.

Brousse, O., Simpson, C. H., Poorthuis, A., and Heaviside, C.: Unequal distributions of crowdsourced weather data in England and Wales, Nature Communications, 15, 1–11, 2024.

Chen, A. B., Behl, M., and Goodall, J. L.: 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, https://dl.acm.org/doi/pdf/10.1145/3276774.3276792, 2018.

Colli, M., Lanza, L., La Barbera, P., and Chan, P.: Measurement accuracy of weighing and tipping-bucket rainfall intensity gauges under dynamic laboratory testing, Atmospheric research, 144, 186–194, 2014.

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, 2017.

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

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, 8820–8829, 2019.

De Vries, H. and Selten, F.: Zomer Bijna Net Zo Nat Als Winter, https://www.knmi.nl/over-het-knmi/nieuws/zomer-bijna-net-zo-nat-als-winter, published: 25 May 2023, 2023.

El Hachem, A., Seidel, J., O'Hara, T., Villalobos Herrera, R., Overeem, A., Uijlenhoet, R., Bárdossy, A., and de Vos, L.: Technical note: A guide to using three open-source quality control algorithms for rainfall data from personal weather stations, Hydrology and Earth System Sciences, 28, 4715–4731, https://doi.org/10.5194/hess-28-4715-2024, 2024.

European Environment Agency: Impervious Built-up 2018 (raster 10 m), Europe, 3-yearly, Aug. 2020, https://doi.org/10.2909/3e412def-a4e6-4413-98bb-42b571afd15e, 2020.

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.

Habib, E., Krajewski, W. F., and Kruger, A.: Sampling errors of tipping-bucket rain gauge measurements, Journal of Hydrologic Engineering, 6, 159–166, 2001.

Hrachowitz, M. and Weiler, M.: Uncertainty of precipitation estimates caused by sparse gauging networks in a small, mountainous watershed, Journal of Hydrologic Engineering, 16, 460–471, 2011.

Humphrey, M., Istok, J., Lee, J., Hevesi, J., and Flint, A.: A new method for automated dynamic calibration of tipping-bucket rain gauges, Journal of Atmospheric and Oceanic Technology, 14, 1513–1519, 1997.

Imhoff, R., Brauer, C., Overeem, A., Weerts, A., and Uijlenhoet, R.: Spatial and temporal evaluation of radar rainfall nowcasting techniques on 1,533 events, Water Resources Research, 56, e2019WR026 723, 2020.

KNMI: Climate Viewer, https://www.knmi.nl/klimaat-viewer/grafieken-tabellen/meteorologische-stations/stations-maand/stations-maand_1991-2020, last accessed: 3 July 2024, 2024.

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

Lanza, L. G. and Vuerich, E.: The WMO field intercomparison of rain intensity gauges, Atmospheric Research, 94, 534–543, https://doi.org/10.1016/j.atmosres.2009.06.012, 2009.

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/https://doi.org/10.1016/j.envsoft.2021.105169, 2021.

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., and Loumagne, C.: When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events, Hydrology and Earth System Sciences, 18, 575–594, https://doi.org/10.5194/hess-18-575-2014, 2014.

Marsalek, J.: Calibration of the tipping-bucket raingage, Journal of Hydrology, 53, 343–354, 1981.

McMillan, H. K., Westerberg, I. K., and Krueger, T.: Hydrological data uncertainty and its implications, Wiley Interdisciplinary Reviews: Water, 5, e1319, 2018.

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

Netatmo: Smart Rain Gauge, https://www.netatmo.com/en-eu/smart-rain-gauge, last accessed: June 2024, 2024a.

Netatmo: Welcome aboard, https://dev.netatmo.com/apidocumentation, last accessed: June 2024, 2024b.

Nielsen, J., van de Beek, C., Thorndahl, S., Olsson, J., Andersen, C., Andersson, J., Rasmussen, M., and Nielsen, J.: Merging weather radar data and opportunistic rainfall sensor data to enhance rainfall estimates, Atmospheric Research, 300, 107 228, 2024.

Niemczynowicz, J.: The dynamic calibration of tipping-bucket raingauges, Hydrology Research, 17, 203–214, 1986.

Niemczynowicz, J.: Urban hydrology and water management–present and future challenges, Urban water, 1, 1–14, 1999.

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.

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

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

Overeem, A., Buishand, T., Holleman, I., and Uijlenhoet, R.: Extreme value modeling of areal rainfall from weather radar, Water Resources Research, 46, 2010.

Overeem, A., Leijnse, H., van der Schrier, G., van den Besselaar, E., Garcia-Marti, I., and De Vos, L. W.: Merging with crowdsourced rain gauge data improves pan-European radar precipitation estimates, Hydrology and Earth System Sciences, 28, 649–668, https://doi.org/10.5194/hess-28-649-2024, 2024a.

Overeem, A., Uijlenhoet, R., and Leijnse, H.: Quantitative precipitation estimation from weather radars, personal weather stations and commercial microwave links, in: Advances in Weather Radar. Volume 3, edited by Bringi, V., Mishra, K., and Thurai, M., vol. 3, pp. 27–68, Institution of Engineering and Technology, https://doi.org/10.1049/SBRA557H_ch2, 2024b.

Rombeek, N., Hrachowitz, M., Droste, A. M., and Uijlenhoet, R.: Rainfall observations dataset from personal weather stations around automatic weather stations in the Netherlands, https://doi.org/10.4121/caa0a93a-effd-4574-95ec-cd874ca97c05.v1, 2024.

Shedekar, V. S., Brown, L. C., Heckel, M., King, K. W., Fausey, N. R., and Harmel, R. D.: Measurement errors in tipping bucket rain gauges under different rainfall intensities and their implication to hydrologic models, in: 2009 Reno, Nevada, June 21-June 24, 2009, p. 1, American Society of Agricultural and Biological Engineers, 2009.

Thorndahl, S., Einfalt, T., Willems, P., Nielsen, J. E., Ten Veldhuis, M.-C., Arnbjerg-Nielsen, K., Rasmussen, M. R., and Molnar, P.: Weather radar rainfall data in urban hydrology, Hydrology and Earth System Sciences, 21, 1359–1380, 2017.

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

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.

Van Leth, T. C., Leijnse, H., Overeem, A., and Uijlenhoet, R.: Rainfall Spatiotemporal Correlation and Intermittency Structure from Micro-γ to Meso-β Scale in the Netherlands, Journal of Hydrometeorology, 22, 2227–2240, 2021.

Villarini, G. and Krajewski, W. F.: Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall, Surveys in Geophysics, 31, 107–129, 2010.

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