



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

Nathalie Rombeek, Markus Hrachowitz, Arjan Droste, and Remko Uijlenhoet

Department of Water Management, Delft University of Technology, Delft, the Netherlands

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

Received: 15 October 2024 – Discussion started: 8 November 2024

Revised: 29 April 2025 – Accepted: 14 June 2025 – Published: 24 September 2025

Abstract. Accurate rainfall observations with high spatial and temporal resolutions 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 has been missing so far. Here, we quantitatively compare rainfall estimates obtained from PWSs with 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 intervals (1, 3, 6, and 24 h) over a 6-year period. It was found that the average of a cluster of PWSs severely underestimates rainfall (around 36 % and 19 % for 1 h and 24 h intervals, respectively). By adjusting the data with areal reduction factors to account for the spatial variability of rainfall extremes and applying a bias correction factor of 1.22 to compensate for instrumental bias, the average relative bias decreases to -5% for 1 h intervals or almost zero for intervals of 3 h and longer. The highest correlations (0.85 and 0.86) and lowest coefficients of variation (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 a spatial resolution of the order of kilometres, such as for flood forecasting in small, fast-responding catchments. 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}) and occurred in spring and summer. However, the spatial distribution of rainfall likely played a large role in the observed differences rather than instrumental limitations. This emphasizes 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. 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 improve this further.

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 modelling (Niemczynowicz, 1999; Moulin et al., 2009; Arnaud et al., 2011; Lobligeois et al., 2014; Beven, 2016; McMillan et al., 2018). Small, fast-responding catchments especially require accurate rainfall observations with high spatial and temporal resolution for reliable predictions, such as on the order of kilometres

and minutes for 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 for providing 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 s and archived at 10 min time steps in the Netherlands) or daily, respectively. One limitation is that these measurements are strictly only representative of 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 into a measuring cylinder and read once a day, are not available in (near) real time. Weather radars, on the other hand, provide data with high spatial and temporal resolution (i.e. typically 1 km² and 5 min) that are 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, for example, the calibration of the instrument itself, signal attenuation, and 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 sensors equipped with a rain gauge has been increasing in the last decade by up to around one PWS for 9, 11, 13, and 15 km² in May 2024 in the Netherlands, Denmark, Switzerland, and Germany, respectively. Figure 1 shows several tens of thousands of PWSs with varying densities across Europe, with more than 60 % having five 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 automatic weather stations (AWSs), with even higher densities in urban areas, where AWS densities are typically low (Overeem et al., 2024b; Brousse et al., 2024). Once the PWS is connected to an online platform such as the Weather Observations Website (WOW; <https://wow.metoffice.gov.uk/>, last access: 25 August 2025), the Weather Underground website (<https://www.wunderground.com/wundermap>, last access: 25 August 2025), or Netatmo (<https://weathermap.netatmo.com/>, last access: 25 August 2025), observations are automatically uploaded to the respective platform. The PWS data are open-

access and, in the case of Netatmo, they 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 from PWSs are prone to several sources of error. These errors can be grouped into three categories. The first category consists of PWS-related errors, such as those related to inappropriate setups and a lack of maintenance of rain gauges, calibration errors, rounding due to data processing, and 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 resulting from the given volume of water that needs to be collected before the bucket tips (Habib et al., 2001; Lanza and Vuerich, 2009). A third category of errors arises due to spatial sampling uncertainties and thus differences between point rainfall estimates resulting from gauges that are not collocated (Villarini et al., 2008).

Rain gauges of PWSs typically 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 it at high intensities (Marsalek, 1981; Shedekar et al., 2009; Colli et al., 2014). In addition, these errors can be amplified if the PWSs are not installed correctly and maintained adequately.

With respect to the PWS-specific sources of error, de Vos et al. (2017) used an experimental setup to investigate part of the instrument-related 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 the Royal Netherlands Meteorological Institute (KNMI) 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. 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 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 yields an overall improvement in the quality of the PWS data (de Vos et al., 2019; Bárdossy et al., 2021;

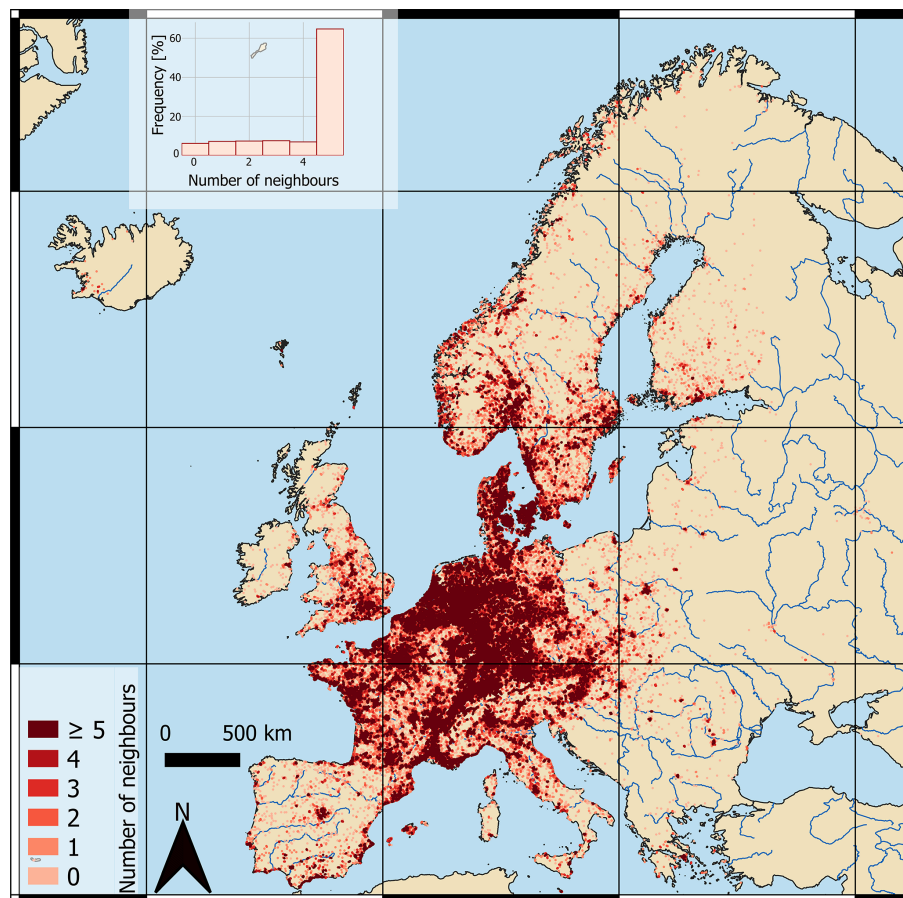


Figure 1. Indication of the density of Netatmo PWSs with rain gauges within Europe, showing the number of PWS neighbours within a radius of 10 km. Data extracted from the Netatmo API in April 2024.

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 intensity classes has been 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 and AWSs, expanding on the results of de Vos et al. (2017, 2019). While weather radar data have a higher spatial resolution than AWSs, they are not used in this research as a reference because they are prone to several sources of error and therefore significantly underestimate rainfall. Note that in previous research by 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 analysing the 10 largest rainfall accumulations with return periods of up to 50 years during the period between 2018 and 2023, for 11 AWSs, four seasons, and four time intervals, we can draw statistically meaningful conclusions

about this. To the best of the authors' knowledge, such a long-term study using PWS data that focuses on the most intense rainfall events has not been performed before. Quantifying the limitations of PWS rainfall observations and addressing them enhances the potential of PWSs for a wide range of applications, including hydrological modelling, urban hydrology, and (hydrological) 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 km² (Fig. 3a). This period was chosen due to the PWS data availability, which was too low (less than five PWSs within a 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 summers are relatively cool (17.2 °C) (KNMI, 2024). The average yearly rainfall between 1990 and 2020 is 851 mm yr^{−1} over the area (KNMI, 2024). In addition, regional variability

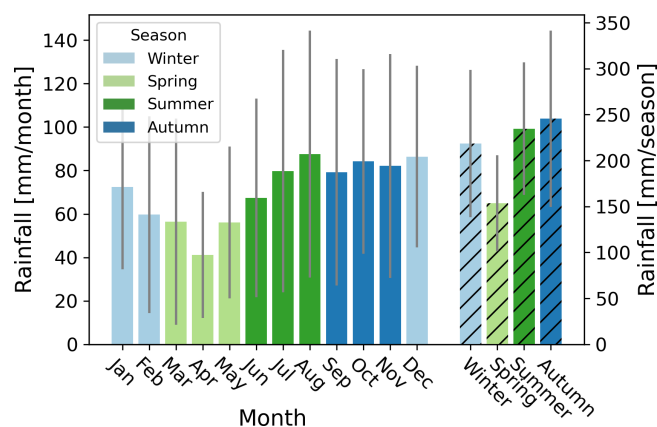


Figure 2. Average rainfall per month and season in the Netherlands over the period 1991–2020, based on data from 13 automatic weather stations spread over the country and obtained from KNMI (2024). Coloured bars indicate the average rainfall per month (millimetres per month, left y axis), and coloured hatched bars indicate the average rainfall for each season (millimetres per season, right y axis). Vertical grey lines indicate the interquartile range.

ity 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 the beginning of autumn is characterized by shorter durations 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 for 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 has slightly increased since 2018 (around one PWS per 9 km² area 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 diameter of the collecting funnel is 13 cm (leading to an orifice area of 133 cm²). According to the manufac-

turer, 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, thereby showing 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 are transmitted via Wi-Fi to the Netatmo platform. The Netatmo software resamples this to regular 5 min intervals by assigning the data to the next full 5 min interval. When within a 5 min interval no data are transferred, this time interval is not included by Netatmo (see the supporting information in Table A1 for an example). When there is a connection failure between the rain gauge module and the 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 the supporting information in Table A2 for an example). However, when the connection of the indoor module is also temporarily interrupted, data are lost.

2.2 Reference dataset

The PWSs were evaluated against data from AWSs from the KNMI. The KNMI operates a network of 33 AWSs across the Netherlands, which are relatively homogeneously distributed with approximately one AWS per 1000 km² area (Fig. 3a). These AWSs estimate cumulative rainfall every 12 s by measuring the displacement of a float placed in a reservoir. The data are archived at a lower resolution, i.e. every 1, 10 min, and hourly. Here, AWS data with resolutions of 10 min and 1 h are used. The 10 min dataset contains unvalidated rainfall data, while the hourly data have been validated (Brandsma et al., 2020). The collecting funnel has a diameter of 16 cm (corresponding to an orifice area of 201 cm²) 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). Nevertheless, these data are 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 at 5 %–8 % annually, with a higher underestimation in winter (7.7 %) than in summer (5.0 %). The undercatch is non-linear 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). Note that

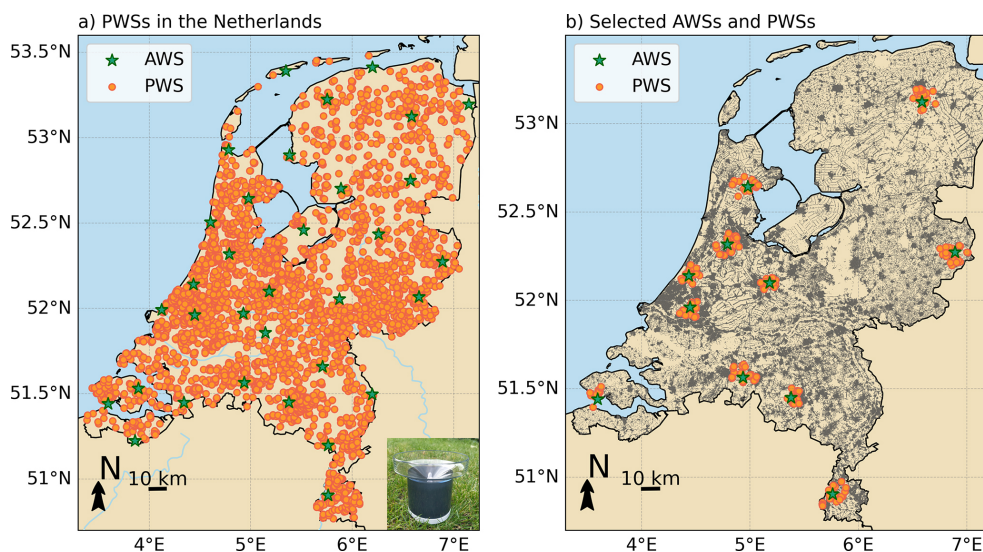


Figure 3. Map of the Netherlands showing (a) the locations of the 33 AWSs employed by KNMI and the locations of available PWSs in 2024 obtained using the Netatmo API (approximately 4000 PWSs) (Netatmo, 2024b). The photo in the right corner shows an example of the PWS used in this research. Note that 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 the European Environment Agency (2020).

the API only provides access to data from PWSs that were operational at the time of access, which was in February 2024. We do not have access to data from stations that were previously in operation but 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 to quantify the quality of the PWSs, and a radius of 20 km was used to filter 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 the Netherlands is around 50 km for 1 h accumulation intervals. Comparing PWSs within 10 km of an AWS 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. Second, AWSs that had at least five PWSs within 10 km since 1 January 2018 were kept in the dataset. Next, the 10 closest available PWSs located within 10 km of 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 (see Fig. 3b for the selected AWSs and the used PWSs). All PWSs within 20 km were used to filter 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).

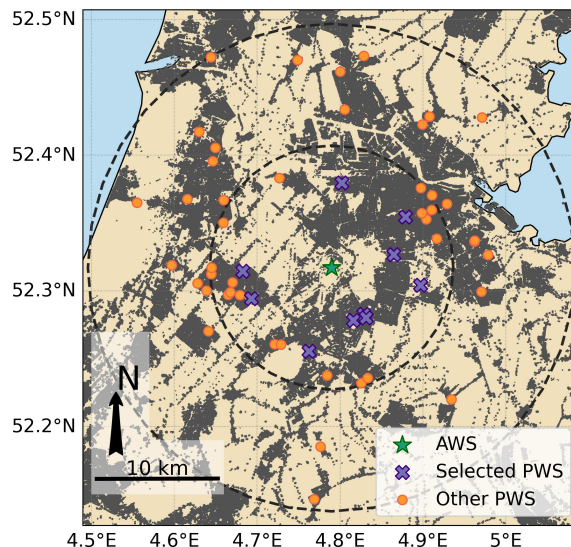


Figure 4. Example of the selection procedure for the AWS at Schiphol. The green star indicates the AWS operated by the KNMI, and the purple crosses indicate the 10 closest PWSs within a distance of 10 km around the AWS. The orange dots are the other PWSs within 20 km of the AWS, which are utilized for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

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 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 for 12.5 % of the annual rainfall. The 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 number of selected events. Note that the selected events can contain time steps without any rain observed.

For every selected AWS based on the methodology in Sect. 3.1, the 10 largest rainfall events per meteorological season (winter, spring, summer, and autumn) and accumulation interval (1, 3, 6, and 24 h) were selected to draw statistically meaningful conclusions. This results in a total of $11 \text{ stations} \times 4 \text{ seasons} \times 4 \text{ accumulation intervals} \times 10 \text{ events} = 1760$ individual events for the analysis. The events were selected in such a way that, for the same station and accumulation interval, no overlapping time series were 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 the highest rates during summer (June–July–August, JJA) and the lowest rates during winter (December–January–February, DJF) and medians of 14.65 and 6.17 mm h^{-1} , respectively. This is in line with the Dutch climate, where the 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 years, 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 (September–October–November, SON) are slightly higher than in spring (March–April–May, MAM), with medians 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 are dominated by rainfall events from March to October. Because winter has the lowest intensities, return periods then based on annual statistics are low.

3.3 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 an 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 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) of 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

As stated by El Hachem et al. (2024), the key advantage of the QC algorithm of de Vos et al. (2019) over QC algorithms such as those developed by Bárdossy et al. (2021) and Lewis et al. (2021) is that no auxiliary data are required. This makes it particularly suitable for regions lacking access to (real-time) reference data. For that reason, we decided to use the QC of de Vos et al. (2019). The high-influx (HI) and faulty zero (FZ) filters from the personal weather station quality control (PWSQC) algorithm of de Vos et al. (2019) were applied to filter the PWS dataset. HI data can be caused by sprinklers, addition of 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. Disaggregating this type of high influx using reliable rainfall time series from nearby stations could potentially solve this issue. The HI filter uses four parameters:

1. d , the maximum distance over which neighbouring PWSs are selected that likely capture similar rainfall dynamics;
2. n_{stat} , the minimum required number of neighbouring PWSs;
3. ϕ_A , a threshold value; and
4. ϕ_B , a 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

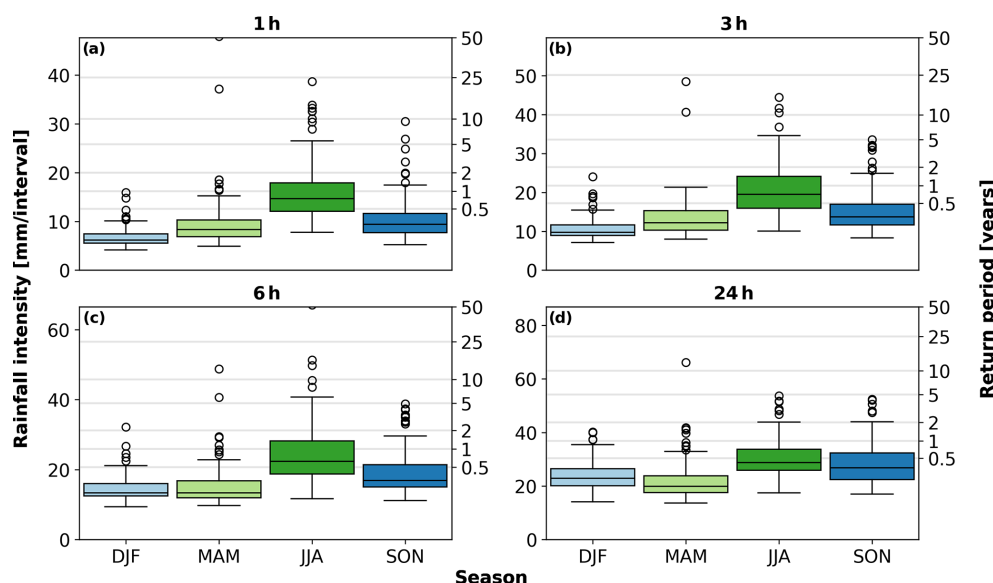


Figure 5. Mean rainfall intensity (millimetres per interval) for the different selected events per season (11 AWSs \times 10 rainfall events) for the four accumulation intervals (1, 3, 6, and 24 h). The left y axis shows the rainfall over the specific interval, and the right y axis shows the different 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 (25th–75th). During summer, outliers were present, with return periods longer than 50 years.

the median $\times (\phi_B/\phi_A)$. 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 due to, for example, a tilted rain gauge or obstructions such as leaves or insects. The FZ filter uses three parameters, i.e. the range d , n_{stat} , and n_{int} . For the n_{int} time intervals at least, the median of the neighbouring PWSs needs to be higher than zero, while the PWS itself reports zero rainfall.

The value of parameter d calibrated 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 timescale for any season in the Netherlands (Van Leth et al., 2021).

For the reasons mentioned above, the same calibrated parameters as in Table 2 of de Vos (2019) were applied. First, a list of PWS neighbours within 10 km was constructed. Second, HI and faulty zero FZ filters were computed for every time step (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 2 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 and 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 (DBC) factor was applied to the dataset. 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 analyse the availability of PWSs, a dataset comprising operational PWSs on 1 January 2018 within 10 km from an AWS was employed. This dataset contained 178 PWSs. Time series from the PWSs were extracted for each selected event according to the method described in Sect. 3.2. In total, 95.8 % of the PWSs were available for 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

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

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

irregular data transfers that were longer than 5 min. In addition, 0.7 % of the PWSs did not have five neighbours within 10 km. Applying the quality control algorithm leads to more data being discarded. For the remaining PWSs with enough neighbours, 8.9 % of the total data are either discarded (i.e. flagged as FZ or HI), have a temporal resolution beyond 5 min, or are missing.

Since we can only suspect that data are likely not missing when a 5 min time step is not included, a minimum availability criterion was set to limit a biased comparison. A minimum percentage of time steps should be valid before aggregating the data. The criterion 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 1 h) 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, the coefficient of variation (CV) of the residuals, the Pearson correlation coefficient (r), and the slope of the 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:

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

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 values below zero an underestimation of the PWS data. The CV is used to describe the dispersion of rainfall, with values closer to zero suggesting greater consistency

with the mean of the reference and higher values indicating more dispersion, defined as follows:

$$\text{CV} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (R_{\text{res},i} - \overline{R_{\text{res}}})^2}}{\overline{R_{\text{AWS}}}}, \quad (2)$$

with the overbar indicating the arithmetic mean and

$$R_{\text{res}} = R_{\text{PWS}} - R_{\text{AWS}}. \quad (3)$$

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 it was calculated between all events within a season and aggregation interval (including zeroes):

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

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

$$\overline{R_{\text{PWS}}} = a * \overline{R_{\text{AWS}}}, \quad (5)$$

with a the slope calculated over all events:

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

Values close to 1 indicate 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 time series used, at least one HI flag was attributed to a time step. In 93 % of the cases, no data were transferred for at least 15 min prior to the flagged HI time step, suggesting that these flags may result from comparing data aggregated over longer time intervals (≥ 15 min) to a 5 min time step, potentially leading to mismatches and flagging data. For 5.8 % of the time series, at least one FZ occurred. Around 15 % of the PWSs were manually calibrated, with a median tipping volume of 0.117 and 95 % of the calibrated tipping bucket volumes ranging between 0.09 and 0.203.

4.1 Spatial sampling

The PWSs in this study were selected based on the closest distance from an AWS without considering the uniformity of the distribution around the AWS. Figure 6 shows that the average cluster distance to an AWS is around 4.4 km and that

the average inter-gauge PWS distance of all pairs within a cluster is around 5 km. This indicates that, overall, the selected stations are not clustered in one location and represent a larger area. Variation in the distance to the AWSs in Fig. 6a can be explained by 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 the number of available PWSs, which increases over the years. The average number of PWSs within a cluster is 8.75.

4.2 Areal reduction effect

Figure 7 shows a substantial decline in ARFs with larger area sizes and shorter durations, with the largest reductions for short durations. The decline becomes more prominent for longer return periods. For an area of 79 km² (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 ARFs are 0.78 and 0.70 for a duration of 1 h and return periods 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).

4.3 Bias

The relative bias of the 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 decreases 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.

It is important to make a distinction between the sources of bias in order to avoid correcting non-instrumentally 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 illustrates 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 occur (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 percentage 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 DBC factor of 1.22 was applied. After application of this DBC factor, the underestimation for 1 h intervals over all seasons decreases to an average of 5 %, while for 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 DBC factor of 1.22 works effectively.

A small part of the dataset obtained from Netatmo (5.38 % of the selected events' total time steps) was not included, suggesting either that data were missing or that the system suffered from connection issues, resulting in irregular data transfer (longer than approximately 5 min). It is expected that the effect of this on the bias will be limited, as most of the data are likely not missing. Rather, the effect is caused by irregular data transfer and/or connection interruption between the rain gauge and the indoor module (see the supporting information in 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 performances of the PWSs in winter and autumn and the worst performances in summer and spring (Fig. 9). Events in winter show the lowest variability of the PWS observations compared with the AWS observations (e.g. average CV values of 0.30 and 0.21 for accumulation intervals of 1 h and 3 h, respectively). While the CV is larger for autumn (0.41 and 0.26), the correlation between the PWSs and AWSs 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 with the other seasons, showing that there is more consistency between the individual PWSs (see

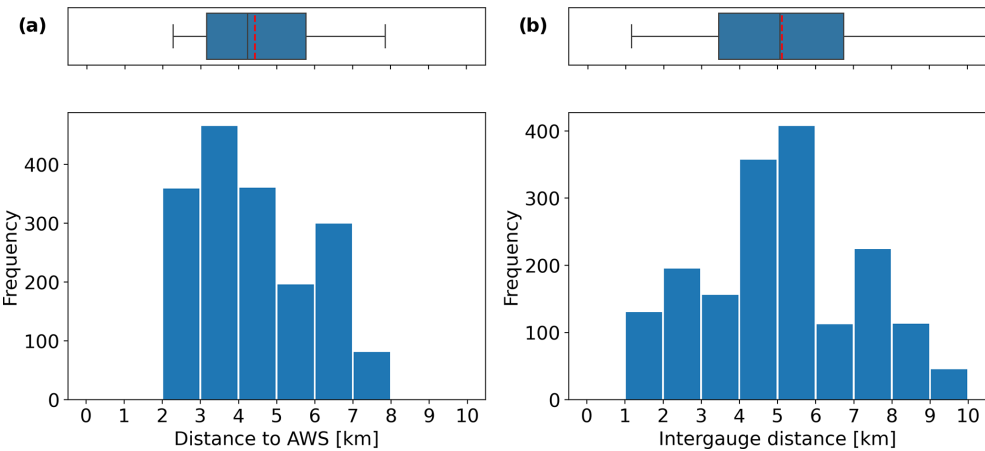


Figure 6. Histogram of (a) the distances between the PWS clusters and the AWSs per event and (b) the inter-station distances between all selected PWS pairs within a cluster per event, all based on 1760 pairs. The vertical red dashed line indicates the mean distance, the vertical black line the median,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

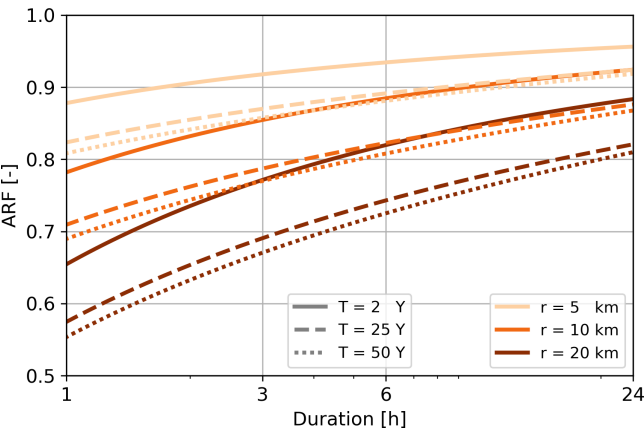


Figure 7. Areal reduction factor (left y axis) as a function of duration for area sizes with radii of 5, 10, and 20 km (converted into an area) and return periods of $T = 2$, 25, and 50 years.

Fig. 9 and the supporting information in Fig. B1 for a complete overview of all of the 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 decreases over longer accumulation intervals. For 1 h intervals the CV varies between 0.30 and 0.54 over the different seasons, while for 24 h the CV shows lower variability, with values varying between 0.14 and 0.27. Similarly, the correlation coefficient increases from values ranging from 0.43 to 0.74 for 1 h intervals to a range from 0.75 to 0.86 for 24 h intervals. This indicates that, for longer accumulation intervals, rainfall observations from PWSs exhibit less variability and more agreement with those from AWSs. Data transferring and processing errors decrease for longer accumulation intervals, as

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 \times 11 AWSs) selected rainfall events per season and interval.

Interval	Relative bias [–]			
	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

the effect of attributing rainfall to an erroneous time stamp decreases. This takes place for example when the connection between the indoor and outdoor modules is temporarily interrupted, potentially attributing rainfall to a timestamp when there is a connection again and as a consequence aggregating it over a longer time interval than approximately 5 min (see the supporting information in Table A2). Within a cluster of PWSs, variation in rainfall is observed. However, using the average rainfall of each PWS cluster shows great

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 0.97 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 good agreement with the AWSs (Fig. 9).

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 investigated the causes of some of these outliers further, which are indicated with orange circles.

It can be seen in Fig. 8c that one of the selected events showed (almost) no 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 d 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 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 time series 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 0.83 and 0.18, respectively. For 24 h this would have been $r = 0.88$ 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.

During summer and spring, the highest rainfall accumulations were observed 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 in Fig. 8b. These events likely skew the overall performance during these months. 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 AWSs are not necessarily only related to instrumental limitations but are rather due to the spatial rainfall distribution. In addition, tipping bucket mechanisms are known for having difficulties in recording higher rainfall intensities. PWSs only send rainfall data to the Netatmo platform twice every 10 min, which is considerably lower than the recording interval of AWSs (12 s). 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 whether the rain was evenly spread within these 10 min or mostly occurred during a shorter period of time and whether this measurement range of the PWSs was indeed exceeded.

5 Discussion

5.1 Bias

Adjusting the PWS dataset with ARFs and correcting for the instrumental bias using a DBC factor reduces the bias. The remaining bias can indicate either that e.g. 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 et al., 2021). The average distance of around 4.4 km from a PWS cluster to 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 will remain most pronounced for events in, for example, summer and spring 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 small-scale convective rainfall events, the distance to the closest AWS might still have been too large, resulting in the variations in the rainfall observed by the PWSs nearby. Radar images show that, for several events, the differences between AWSs and PWSs are caused by the spatial distribution of the rainfall (see the supporting information in Figs. D1, D2, and D3).

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 to 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).

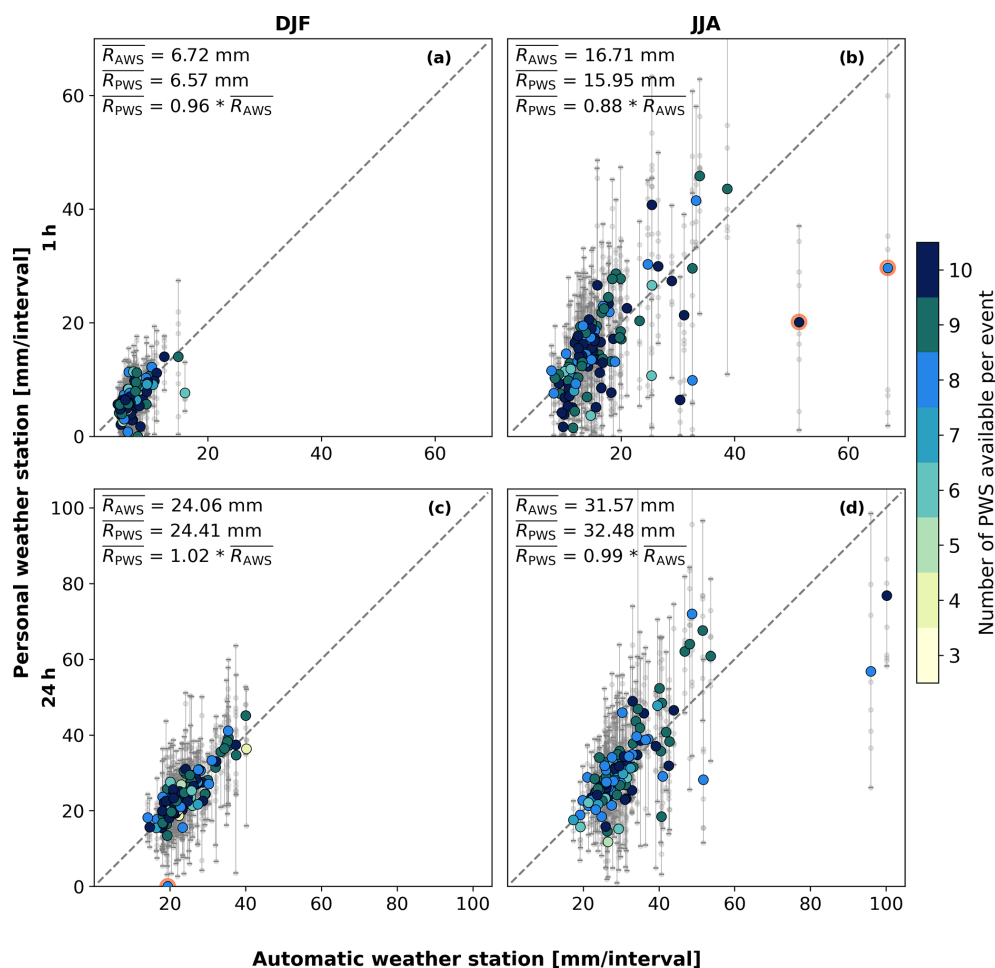


Figure 8. Scatterplots of filtered PWS rainfall accumulations against AWS records for the winter (a, c) and summer (b, d) seasons and accumulation intervals of 1 h (a, b) and 24 h (c, d). The large coloured dots indicate the average of a cluster of PWSs against one AWS, and the vertical bars indicate the minimum and maximum of that cluster of PWSs. The 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 $\overline{R_{AWS}}$ and $\overline{R_{PWS}}$ represent the average rainfall over the selected events recorded by the AWS and PWS, respectively. $R_{PWS} = a * R_{AWS}$ represents the linear regression line fitted through zero, with a indicating the slope.

5.1.2 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 for 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 (DBC factor of 1.24). Alternatively, Overeem et al. (2024a) used a DBC factor of 1.063 for a pan-European dataset. Neither of the 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 DBC factors were bulk cor-

rection factors tuned on different reference datasets. de Vos et al. (2019) utilized gauge-adjusted radar values. Radars indirectly measure rainfall, which might not be representative of 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 neither de Vos et al. (2019) nor Overeem et al. (2024a) distinguished between certain types of rain events, looking rather 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.

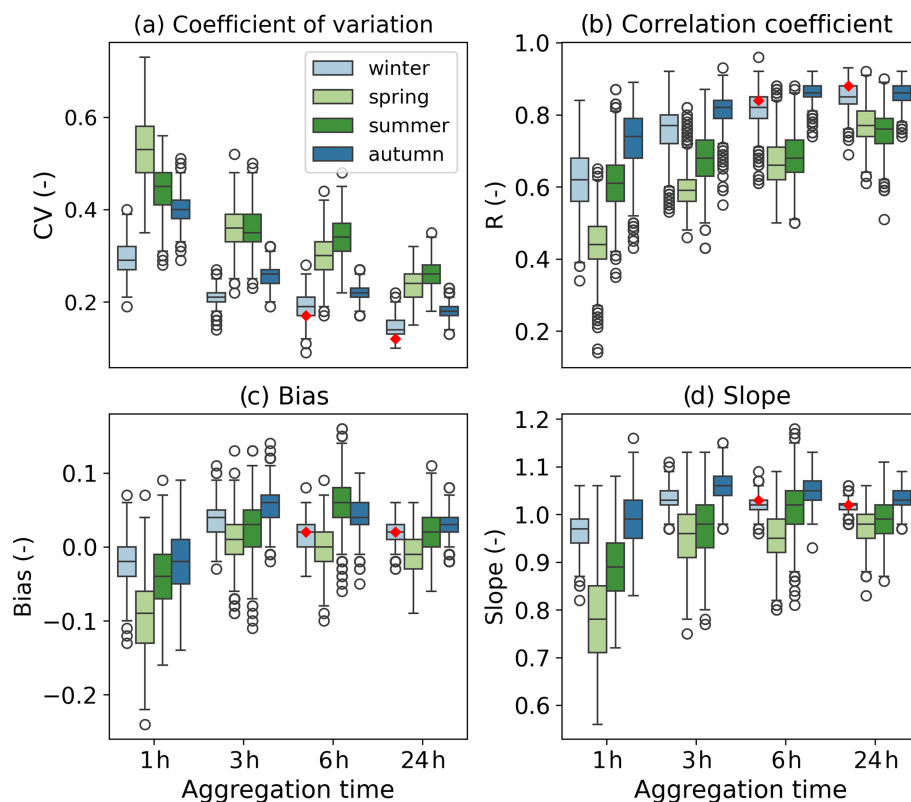


Figure 9. Coefficient of variation (a), correlation coefficient (b), bias (c), and slope (d) of filtered PWS rainfall accumulations against AWS for different seasons and accumulation intervals. The lower and upper whiskers indicate the minimum and maximum of each metric and the boxes the inter-percentile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

Therefore, the type of dataset and the included rainfall events play a role in the performance. Applying a DBC factor of 1.22 over the dataset almost eliminates the bias. 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.

5.1.3 Manual calibration

Around one-seventh of the PWSs used in this study (15 %) were manually calibrated by their owners. 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 into the original default value of 0.101 mm. On average, there is a 4 % decrease in the rainfall observed 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 the supporting information in Table C1).

5.2 Quality control

The quality control algorithm of de Vos (2019) utilized in this research improves the overall performance of the PWSs (see the supporting information in Fig. E1). However, there are some limitations to this algorithm. The algorithm works only if there are enough neighbouring stations within 10 km, limiting its 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 those 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 be-

come 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 the 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 the 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.

6 Conclusions

This study provides insights into the systematic errors across the personal weather station (PWS) network during high-intensity rainfall events by performing a comprehensive analysis over 6 years. The analysis focuses on the most intense rainfall observations for a large number of events (1760) over 6 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, and autumn), durations (1, 3, 6, and 24 h), and AWSs (11 locations spread across the country). PWS data were filtered with a quality control (QC) algorithm utilizing neighbouring PWSs. After QC, around 88 % of the original dataset was kept. To reduce uncertainty from single stations, metrics were calculated over a cluster of PWSs rather than individual stations.

Results showed that PWSs severely underestimate rainfall. A seasonal effect is visible in the bias, specifically for shorter accumulation intervals, with the largest biases for summer and spring. A part of this bias can be 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 (DBC) factor of 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 DBC factors of 1.063 proposed 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 and temporal dependence is seen from the correlation coefficient (r) and coefficient of variation (CV). Outliers in winter seemed to

have been caused by freezing temperatures (solid precipitation). For that reason, we recommended further analysing 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 did not observe the most intense rainfall events, with high intensities over a relatively short amount of time (e.g. $> 75 \text{ mm h}^{-1}$ 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. The spatial footprint of these high-intensity rainfall events is often small, influencing errors related to spatial sampling due to the average distance of 4.4 km to the nearest AWS. This suggests the need for a high-density observation network to reliably capture localized rainfall extremes. Rain gauges from the PWSs used in this research utilize a tipping bucket mechanism, which is known to suffer from non-linear underestimation errors with increasing intensities, potentially contributing to these outliers. To quantify this, experimental studies are necessary to limit other sources of errors. The performance of the PWSs improved over longer accumulation intervals, resulting in r values of 0.85 and 0.86 and CV values of 0.14 and 0.18 for 24 h in winter and autumn, respectively.

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

Appendix A: Netatmo data processing

Table A1. Example of the Netatmo software that resamples data to regular 5 min intervals by assigning them to the next full 5 min interval.

Raw		Aggregated	
Time (dd-mm-yyyy hh:mm:ss)	Rain (mm)	Time (dd-mm-yyyy hh:mm:ss)	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
		29-10-2024 06:35:00	1.616

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

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

Appendix B: Overview of all seasons and accumulation intervals

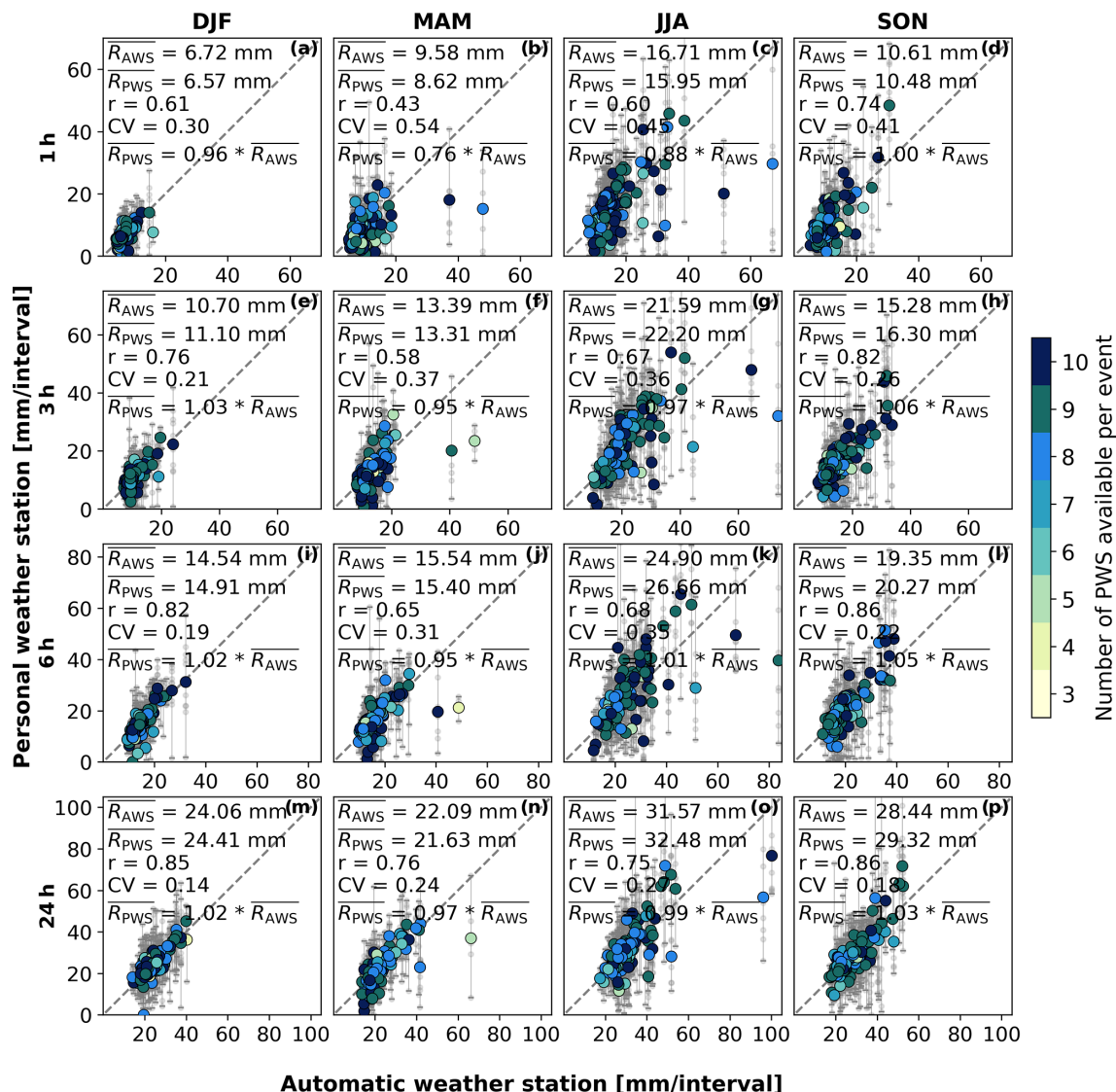


Figure B1. Scatterplots of filtered PWS rainfall accumulations against AWSs for the winter (a, e, i, m), spring (b, f, j, n), summer (c, g, k, o), and autumn (d, h, l, p) seasons and accumulation intervals of 1 h (a–d), 3 h (e–h), 6 h (i–l), and 24 h (m–p). The large coloured dots indicate the average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

Appendix C: Calibration effect

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

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

Appendix D: Spatial distribution of rainfall

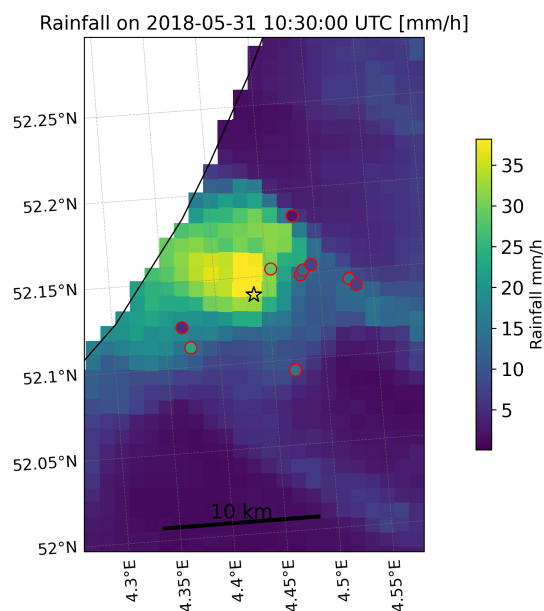


Figure D1. Rainfall distribution on 31 May 2018 at 10:30 UTC, based on 1 h of accumulated rainfall from the gauge-adjusted radar product (Overeem et al., 2009b). The asterisk indicates the location of the AWS which measured 39 mm in 1 h, whereas the circles with red borders represent the locations of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

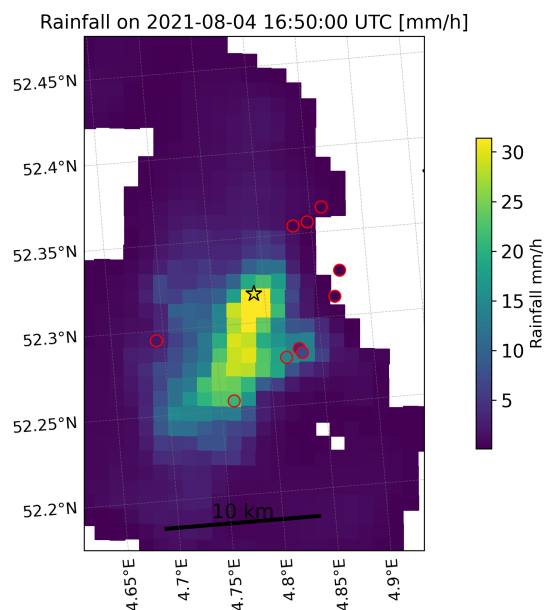


Figure D2. Rainfall distribution on 4 August 2021 at 16:50 UTC, based on 1 h of accumulated rainfall from the gauge-adjusted radar product (Overeem et al., 2009b). The asterisk indicates the location of the AWS which measured 30 mm in 1 h, whereas the circles with red borders represent the locations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

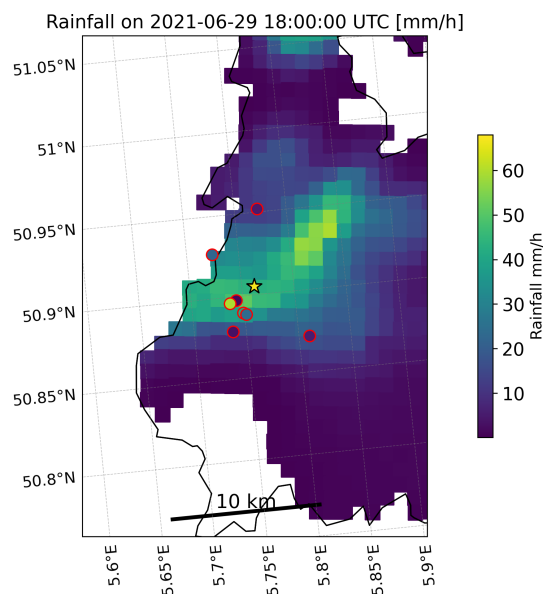


Figure D3. Rainfall distribution on 29 June 2021 at 18:00 UTC, based on 1 h of accumulated rainfall from the gauge-adjusted radar product (Overeem et al., 2009b). The asterisk indicates the location of the AWS which measured 67 mm in 1 h, whereas the circles with red borders represent the locations of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

Appendix E: Raw PWS data

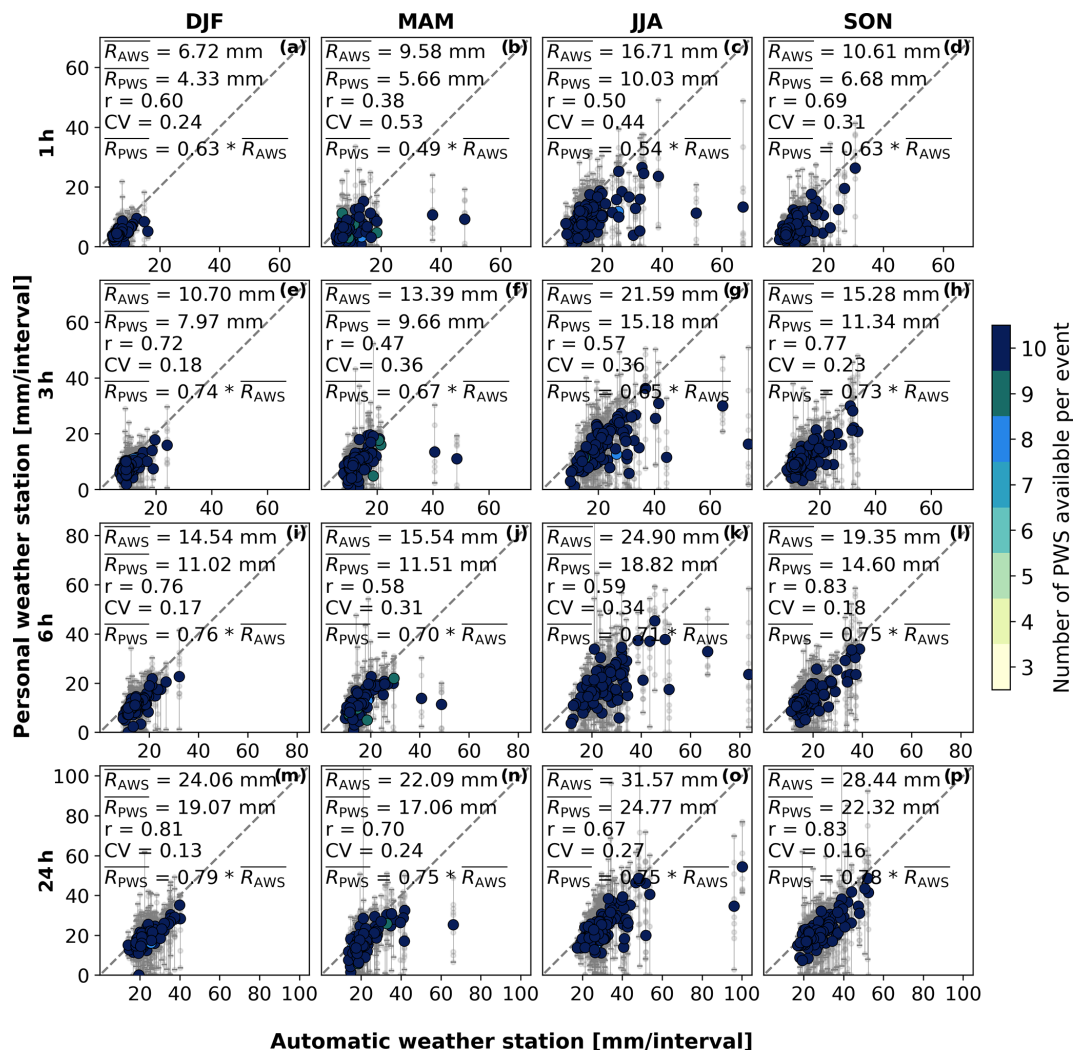


Figure E1. Scatterplots of raw PWS rainfall accumulation data against AWSs for the winter (a, e, i, m), spring (b, f, j, n), summer (c, g, k, o), and autumn (d, h, l, p) seasons and accumulation intervals of 1 h (a–d), 3 h (e–h), 6 h (i–l), and 24 h (m–p). The large coloured dots indicate the average of a cluster of PWSs against one AWS, and the vertical bars indicate the minimum and maximum of that cluster of PWSs. The colours indicate the number of PWSs used to calculate the mean, minimum, and maximum rainfall. The small grey dots represent one individual PWS against an AWS. $\overline{R_{AWS}}$ and $\overline{R_{PWS}}$ 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. $\overline{R_{PWS}} = a * \overline{R_{AWS}}$ represents the linear regression line fitted through zero, with a indicating the slope.

Code and data availability. The data from automatic weather stations from the KNMI are freely available on the KNMI data platform for 10 min intervals: <https://dataplatforn.knmi.nl/dataset/neerslaggegevens-1-0> (KNMI, 2025a). The hourly validated dataset is available at <https://www.daggegevens.knmi.nl/klimatologie/uurgegevens> (KNMI, 2025b). Part of the quality control of the Netatmo gauge data, i.e. the faulty zeroes and high-influx filters, is based on the PWSQC algorithm written in the R language and can be found at <https://doi.org/10.5281/zenodo.10629488> (de Vos, 2024). The dataset with the Netatmo PWS rainfall data can be found at <https://doi.org/10.4121/caa0a93a-effd-4574-95ecd874ca97c05.v1> (Rombeek et al., 2024).

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

Competing interests. At least one of the (co-)authors is a member of the editorial board of *Hydrology and Earth System Sciences*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. The authors would like to thank Davide Wüthrich for the discussions and for proofreading the manuscript as well as Claudia Brauer for providing insightful feedback on the results, which helped to improve this paper.

Financial support. This work is part of the Perspectief research programme Future Flood Risk Management Technologies for rivers and coasts with project no. P21-23. This programme is financed by Domain Applied and Engineering Sciences of the Dutch Research Council (NWO).

Review statement. This paper was edited by Efrat Morin and reviewed by two anonymous referees.

References

- Arnaud, P., Lavabre, J., Fouchier, C., Diss, S., and Javelle, P.: Sensitivity of hydrological models to uncertainty in rainfall input, *Hydrol. Sci. J.*, 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, *Hydrol. Earth Syst. Sci.*, 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, Amersfoort, ISBN 9789057738609, 2019.
- Berne, A., Delrieu, G., Creutin, J.-D., and Obled, C.: Temporal and spatial resolution of rainfall measurements required for urban hydrology, *J. Hydrol.*, 299, 166–179, 2004.
- Beven, K.: Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication, *Hydrol. Sci. J.*, 61, 1652–1665, 2016.
- Brandsma, T.: Comparison of automatic and manual precipitation networks in the Netherlands, Technical report TR-347, KNMI, De Bilt, <https://cdn.knmi.nl/knmi/pdf/bibliotheek/knmipubTR/TR347.pdf> (last access: 22 September 2025), 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, <https://cdn.knmi.nl/knmi/pdf/bibliotheek/knmipubTR/TR387.pdf> (last access: 22 September 2025), 2020.
- Brousse, O., Simpson, C. H., Poorthuis, A., and Heaviside, C.: Unequal distributions of crowdsourced weather data in England and Wales, *Nat. Commun.*, 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, 25–28, <https://dl.acm.org/doi/pdf/10.1145/3276774.3276792> (last access: 25 August 2025), 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, *Atmos. Res.*, 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, *Hydrol. Earth Syst. Sci.*, 21, 3859–3878, <https://doi.org/10.5194/hess-21-3859-2017>, 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.: LottedeVos/PWSQC: PWSQC (v.1.0). Zenodo [code], <https://doi.org/10.5281/zenodo.10629488>, 2024.
- de Vos, L., Leijnse, H., Overeem, A., and Uijlenhoet, R.: The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam, *Hydrol. Earth Syst. Sci.*, 21, 765–777, <https://doi.org/10.5194/hess-21-765-2017>, 2017.
- de Vos, L. W., Leijnse, H., Overeem, A., and Uijlenhoet, R.: Quality control for crowdsourced personal weather stations to enable operational rainfall monitoring, *Geophys. Res. Lett.*, 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> (last access: 25 August 2025), 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, *Hydrol. Earth Syst. Sci.*, 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, EEA [data set], <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, *J. Hydrol.*, 37, 100883, <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, *J. Hydrol. Eng.*, 6, 159–166, 2001.
- Hrachowitz, M. and Weiler, M.: Uncertainty of precipitation estimates caused by sparse gauging networks in a small, mountainous watershed, *J. Hydrol. Eng.*, 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, *J. Atmos. Ocean. Technol.*, 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 Resour. Res.*, 56, e2019WR026723, <https://doi.org/10.1029/2019WR026723>, 2020.
- KNMI: Climate Viewer, KNMI [data set], https://www.knmi.nl/klimaat-viewer/grafieken-tabellen/meteorologische-stations/stations-maand/stations-maand_1991-2020 (last access: 3 July 2024), 2024.
- KNMI: Precipitation - duration, amount and intensity at a 10 minute interval, KNMI [data set], <https://dataplatform.knmi.nl/dataset/neerslaggegevens-1-0>, last access: 25 August 2025a.
- KNMI: Weerstations - Uurwaarnemingen, KNMI [data set], <https://www.daggegevens.knmi.nl/klimatologie/uurgegevens>, last access: 25 August 2025b.
- Krajewski, W. F., Villarini, G., and Smith, J. A.: Radar-rainfall uncertainties: Where are we after thirty years of effort?, *Bull. Am. Meteorol. Soc.* 91, 87–94, 2010.
- Lanza, L. G. and Vuerich, E.: The WMO field intercomparison of rain intensity gauges, *Atmos. Res.*, 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, *Environ. Model. Softw.*, 144, 105169, <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, *Hydrol. Earth Syst. Sci.*, 18, 575–594, <https://doi.org/10.5194/hess-18-575-2014>, 2014.
- Marsalek, J.: Calibration of the tipping-bucket raingage, *J. Hydrol.*, 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, <https://doi.org/10.1002/wat2.1319>, 2018.
- Moulin, L., Gaume, E., and Obled, C.: Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations, *Hydrol. Earth Syst. Sci.*, 13, 99–114, <https://doi.org/10.5194/hess-13-99-2009>, 2009.
- Netatmo: Smart Rain Gauge, <https://www.netatmo.com/en-eu/smart-rain-gauge> (last access: June 2024), 2024a.
- Netatmo: Welcome aboard, <https://dev.netatmo.com/apidocumentation> (last access: 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, *Atmos. Res.*, 300, 107228, <https://doi.org/10.1016/j.atmosres.2024.107228>, 2024.
- Niemczynowicz, J.: The dynamic calibration of tipping-bucket rain-gauges, *Hydrol. Res.*, 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., Kroll, S., Murlà-Tuyls, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P., and Ten Veldhuis, M. C.: Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation, *J. Hydrol.*, 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 Resour. Res.*, 45, W10424, <https://doi.org/10.1029/2009WR007869>, 2009a.
- Overeem, A., Holleman, I., and Buishand, A.: Derivation of a 10-year radar-based climatology of rainfall, *J. Appl. Meteorol. Climatol.*, 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 Resour. Res.*, 46, 09S14, <https://doi.org/10.1029/2009WR008517>, 2010.
- Overeem, A., Leijnse, H., van der Schrier, G., van den Besseelaar, E., Garcia-Marti, I., and De Vos, L. W.: Merging with crowdsourced rain gauge data improves pan-European radar precipitation estimates, *Hydrol. Earth Syst. Sci.*, 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*, Vol. 3, edited by: Bringi, V., Mishra, K., and Thurai, M., Vol. 3, 27–68, Institution of Engineering and Technology, ISBN 9781839536267, 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 [data set], <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, 21 June–24 June 2009, p. 1, American Society of Agricultural and Biological Engineers, American Society of Agricultural and Biological Engineers (ASABE), <https://doi.org/10.13031/2013.27308>, 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, *Hydrol. Earth Syst. Sci.*, 21, 1359–1380, <https://doi.org/10.5194/hess-21-1359-2017>, 2017.
- Uijlenhoet, R. and Berne, A.: Stochastic simulation experiment to assess radar rainfall retrieval uncertainties associated with attenuation and its correction, *Hydrol. Earth Syst. Sci.*, 12, 587–601, <https://doi.org/10.5194/hess-12-587-2008>, 2008.
- Van de Beek, C., Leijnse, H., Torfs, P., and Uijlenhoet, R.: Seasonal semi-variance of Dutch rainfall at hourly to daily scales, *Adv. Water Res.*, 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, *J. Hydrometeorol.*, 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, *Surv. Geophys.*, 31, 107–129, 2010.
- Villarini, G., Mandapaka, P. V., Krajewski, W. F., and Moore, R. J.: Rainfall and sampling uncertainties: A rain gauge perspective, *J. Geophys. Res.-Atmos.*, 113, D11102, <https://doi.org/10.1029/2007JD009214>, 2008.