The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam

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

[1]Hydrology and Quantitative Water Management Group, Department of Environmental Sciences, Wageningen University, 6708 PB Wageningen, the Netherlands
[2]Research and Development Observations and Data Technology, Royal Netherlands Meteorological Institute, 3732 GK De Bilt, the Netherlands

Correspondence to: lotte.devos@wur.nl

Abstract. The high density of built-up areas and resulting imperviousness of the land surface makes urban areas vulnerable to extreme rainfall, which can lead to considerable damage. In order to design and manage cities to be able to deal with the growing number of extreme rainfall events, rainfall data is required at higher temporal and spatial resolutions than those needed for rural catchments. However, the density of operational rainfall monitoring networks managed by local or national
5   authorities is typically low in urban areas. A growing number of automatic personal weather stations (PWSs) link rainfall measurements to online platforms. Here, we examine the potential of such crowdsourced datasets for obtaining the desired resolution and quality of rainfall measurements for the capital of the Netherlands. Data from 63 stations in Amsterdam (~575 km^2) that measure rainfall over at least 4 months in a 17-month period are evaluated. In addition, a detailed assessment is made of three Netatmo stations, the largest contributor to this dataset, in an experimental set-up. The sensor performance in
10 the experimental set-up and the density of the PWS-network are promising. However, features in the online platforms, like rounding and thresholds, cause changes from the original time series, resulting in considerable errors in the datasets obtained. These errors are especially large during low intensity rainfall, although they can be reduced by accumulating rainfall over longer intervals. Accumulation improves the correlation coefficient with gauge-adjusted radar data from 0.48 at 5 min intervals to 0.60 at hourly intervals. Spatial rainfall correlation functions derived from PWS data show much more small-scale variability
15 than those based on gauge-adjusted radar data and those found in similar research using dedicated rain gauge networks. This can largely be attributed to the noise in the PWS data resulting from both the measurement setup and the processes occurring in the data transfer to the online PWS-platform. A double mass comparison with gauge-adjusted radar data shows that the median of the stations resembles the rainfall reference better than the real-time (unadjusted) radar product. Averaging nearby raw PWS measurements further improves the match with gauge-adjusted radar data in that area. These results confirm that the growing
20 number of internet-connected PWSs could successfully be used for urban rainfall monitoring.

1 Introduction

Urban catchments are characterized by a high proportion of impervious surfaces, leading to a large fraction of rainfall producing direct runoff and a fast hydrological response. This makes cities especially vulnerable to flooding. The temporal and

spatial resolutions of rainfall data required for urban applications exceed those needed for rural catchments (Schilling, 1991). The rainfall information at spatial and temporal resolutions of typically 1 km by 1 km and 5 min generated by most operational weather radars is considered valuable for urban hydrological analysis and forecasting (Liguori et al., 2012). However, radar has significant limitations; rainfall is determined indirectly, over an atmospheric volume with a size depending on the distance from the radar station, which may not be representative for rainfall at ground-level (Einfalt et al., 2004; Peleg et al., 2013). Errors in rainfall estimates from radar due to sampling uncertainties can be significant. In addition, there is an optimum spatial resolution corresponding to a given temporal resolution (Fabry et al., 1994; Bell and Moore, 2000). Rain gauges, if well maintained, provide accurate ground-based measurements, although they are limited in their spatial representation; Villarini et al. (2008) showed that approximations of true spatial rainfall fields with rain gauges requires a dense network and/or large temporal measurement intervals.

Hydrological models, designed to deal with high resolution input, provide the best simulation results not just when the temporal resolution or the spatial resolution is high, but particularly when the combination thereof is optimal. The required spatio-temporal resolutions for urban applications have been studied extensively. Berne et al. (2004) determined a relation between the space-time resolution required for hydrological applications as a function of the catchment size for Mediterranean conditions. It was found that for urban catchments in the order of $10 \ \mathrm{km}^2$, rainfall data is needed at a temporal resolution of 5 min and a spatial resolution of 3 km. For urban catchments of $1 \ \mathrm{km}^2$ these resolutions were 3 min and 2 km, respectively. The space-time scales of four types of rainfall are evaluated by Emmanuel et al. (2012). With the use of variograms of 24 storm events, the spatial resolutions required to capture these types of rainfall at urban scales range from 0.8 to 3 km for instantaneous monitoring and from 2.5 to 8 km for 30 min intervals.

Gires et al. (2012) found an outflow uncertainty of up to 20% in an urban catchment of $9\mathrm{km}^2$ due to rainfall variability at scales smaller than the typical C-band radar resolution of 1 km by 1 km and 5 min. Bruni et al. (2015) addressed the loss in urban hydrodynamic model accuracy due to smoothing and smearing. Radar data of 4 storm events in 1 min temporal resolution were aggregated to various spatial and temporal resolutions (highest range resolution of 30 m) and used as precipitation input in a $3.4 \ \mathrm{km}^2$ Dutch urban catchment. Smoothing occurs when the ratio of radar resolution over catchment size becomes larger than 0.2 and storms that move near the catchment boundary are averaged partly out of the catchment. Smearing becomes significant when the ratio of the spatial resolution of radar measurements over the rainfall correlation length exceeds 0.9, leading to averaging of rainfall over the coarse spatial grid and resulting in underestimation of rainfall rates in areas within the storm cells and overestimation in the surrounding areas. Also, a runoff peak time shift of up to 6 minutes was found due to temporal aggregation (from 1 minute to 5 and 10 minutes) of rainfall input.

Ochoa-Rodriguez et al. (2015) evaluated the required spatial and temporal resolutions of rainfall in a simple spatio-temporal scaling framework. A spatial resolution of 1 km, typically found in radar, was found to give good hydrodynamic model results, although some extremes were missed. Temporal resolutions should ideally be below the 5 min intervals currently available in

most operational weather radar-products. Nevertheless, the accuracy of 5 min radar data can be improved with the use of an accumulation procedure that assumes constant velocity of the rainfall field and rainfall intensity to vary linearly in time (Fabry et al., 1994). Coarsening temporal resolution has more impact on the accuracy than coarsening spatial resolution. Initial results from an ongoing study by the authors indicate that this impact is reduced when temporal resolutions are coarsened through

5   aggregation (i.e. similar to rain gauges) instead of sampling. Lobligeois et al. (2014) evaluated the circumstances where hydrological model performance is enhanced by higher spatial resolution of rainfall. They did so by comparing lumped and semidistributed models with subcatchment sizes of 64, 16 and 4 km^2. From comparisons between the various model outputs and observations in 181 catchments in France, it was found that model accuracy improvement depends on scale, catchment and event characteristics, and that the spatial representation of rainfall can be a highly important factor in the model performance.

From these works it becomes evident that an increase of the number of measurements would yield a higher accuracy of rainfall fields and would improve hydrological applications. Adding sensors (rain gauges or others) to a network is costly, although there are alternatives. For instance, rain maps can be produced from received signal strength in cellular communication networks, as the microwave signals propagating over the link paths are attenuated by rainfall (Overeem et al., 2016). Weather

15   data can also be provided directly by crowdsourcing measurements from amateurs in various ways (Muller et al., 2015). A growing number of weather enthusiasts measure their local weather with automatic personal weather stations (PWSs). PWS accuracy on measuring temperature, relative humidity, radiation, pressure, rainfall, wind speed and direction has been evaluated for popular high-end expensive weather stations (Jenkins, 2014; Bell et al., 2015), as well as for the cheaper, user-friendly Netatmo type (temperature only) (Meier et al., 2015), which have grown rapidly in number over the past years. So far, weather

20   stations have been used to obtain air temperature data to examine the Urban Heat Island effect (Steeneveld et al., 2011; Wolters and Brandsma, 2012), although other meteorological variables, such as rainfall, are measured by some of these stations as well.

A large number of PWSs share data on online platforms, both on the owner's own initiative (Gharesifard and Wehn, 2016) or automatically as an intrinsic software feature of the product (i.e. for Netatmo). Netatmo has its own online platform collecting

25   and visualizing data from all operational Netatmo stations. The Wundermap of company Weather Underground is a similar online platform. Data from Netatmo stations are automatically linked to the Wundermap. Owners of other PWS-types can actively transmit their measurements to this platform as well. A growing number of automatic weather stations are linked to these platforms; in May 2016 there were 258 personal weather stations linked to Wundermap in the Amsterdam metropolitan area (~575 km^2) alone (239 of type Netatmo), of which 83 stations measured rainfall (64 of type Netatmo). By contrast, the

30   official national automatic weather station network in the Netherlands (~35,000 km^2) consists of 31 stations, and these are, as a rule, always located outside urban areas. Figure 1 shows the relative resolutions in the Netherlands of networks discussed in this paper. At many locations, the density of PWS-stations collecting rainfall data far exceeds that of any realistic operational network implemented by national weather services or local authorities beyond experimental campaigns. As the online platforms collecting and sharing PWS weather data are not nation-bound, global rainfall measurements have become easily

available, with especially high densities in Western-Europe, USA and Japan.

Although rainfall data availability with PWS-networks is cause for optimism for urban hydrological applications, errors are expected to be larger than those in traditional measurements. PWSs come in many types, a large fraction of which are low-cost with expected low sensor quality. In most cases there is no information available on the PWS type, the installation set-up, maintenance of the sensor, or data post processing while transferring measurements to the online platform. Bell et al. (2013) examine the potential improvement on the UK's observational network with the real-time and local weather measurements of air temperature, relative humidity and pressure collected from Wundermap. The most critical issue was found to be the estimation of data quality. Validation procedures like range tests (i.e. a check whether the measurement is within predefined extremes limits) and internal consistency tests should be applied to precipitation data from automatic weather stations (Estévez et al., 2011). The integration of crowdsourced data with variable temporal resolutions in hydrological discharge modeling by accounting for different uncertainties for data of various sources has been addressed in recent research (Mazzoleni et al., 2015).

It becomes clear that urban applications would benefit from high resolution rainfall measurements. The potential of crowdsourced PWS rainfall data for this purpose has not previously been explored. Using the existing PWS-network requires minimal financial investment, and would therefore be an economically reasonable alternative to conventional techniques to increase measurement resolutions. This study aims to determine the added value of crowdsourcing automatic weather stations for urban rainfall monitoring. For this purpose the most common PWS is tested in an experimental set-up with a high quality rain gauge reference. Additionally, a dataset of 63 crowdsourced PWS stations in Amsterdam is validated with a gridded dataset based on radar data, a manual network and a WMO certified automatic rain gauge network. These combined results provide insight in the rainfall measurement accuracy of the most commonly used PWS, as well as any issues that occur in operational crowdsourcing of PWS rain measurements. Following this Introduction is the Methods section, where Sect. 2.1 describes the data and Sect. 2.2 gives an outline to determine the achieved measurement scales and quality of PWS, respectively. The results of an experimental PWS set-up, a comparison of a larger dataset in Amsterdam with gauge-adjusted radar data, and an analysis on inter-gauge spatial correlation of this dataset are given in Sect. 3. Finally, a Discussion on the state and future role of PWS networks in (urban) hydrological applications and Conclusions are given in Sect. 4 and Sect. 5, respectively.

2 Methods

2.1 Data collection

2.1.1 Personal weather stations

From the Wundermap website, a dataset of 63 automatic weather stations located in the Amsterdam area (~575 km^2) has been retrieved. Stations were selected based on the availability of rainfall measurements, which should cover at least 4 months between December 2014 and April 2016. Of these stations, 49 are of brand Netatmo, 7 are of brand Davis, and 7 are of other

unspecified brands. No details on the devices are given. According to the product specifications provided by the manufacturer, the Netatmo rain gauges have a measurement range of 0.2-150 mm h^{-1} with an accuracy of 1 mm h^{-1}. The plastic tipping buckets have a volume of 0.1 mm and a collecting funnel with a diameter of 13 cm. The rain gauge module communicates in a wireless manner to the Netatmo indoor module over distances up to 100 m. The number of tips in the previous interval is communicated every ~5 minutes from the indoor module to the online dashboard via a WiFi-connection, where it can be monitored by the weather station owner. Simultaneously, the measurement is linked to the Netatmo Weathermap from which it is sent every ~10 minutes to the Wundermap. The Wundermap stations that contribute to the dataset are visualized in Fig. 2. The Wundermap platform collects the rainfall measurements and rewrites them into rainfall over the past hour and cumulative rainfall for that day. Daily rainfall only becomes non-zero once the 0.3 mm threshold is reached and subsequent rainfall is only reported if the rounded daily rainfall increases by at least 0.2 mm.

While Netatmo hardware can store measurements for a period of time in case of bad connectivity with the server, only real-time data is automatically transferred to the Wundermap. This causes gaps in the Wundermap datasets where there may be none in the original Netatmo data, which are only accessible to the weather station owner. Wundermap time series are characterized by (large) gaps in the dataset and irregular measurement frequencies, though often 5, 10 or 15 minutes. Also, the locations of Netatmo weather stations on the Wundermap are obtained from the settings at the Netatmo platform without notice to or confirmation from the PWS owner. Relocations of the station that are communicated to the Netatmo platform are not simultaneously adjusted on the Wundermap, potentially leading to large errors in sensor location.

We process the data obtained via Wundermap by calculating the difference in cumulative daily rainfall compared with the previous time step. Since these time steps are not fixed, this results in rainfall accumulations over time intervals of varying lengths. In order to obtain compatible time series, the rainfall is interpolated on a fixed time-line with constant steps, where constant rainfall within the original intervals is assumed. Original intervals longer than 20 min are discarded. Faulty values in precipitation data from automatic weather stations can be identified with range tests and internal consistency tests (Estévez et al., 2011). As a first quality check, values of the interpolated time series are compared with the median rainfall of all stations for each time interval. Values exceeding this median by more than 50 mm h^{-1} are excluded. Dry periods in the dataset are identified as periods of at least 24 hours where the median of all PWS measurements indicate zero rainfall. If a PWS reports continuous zero rainfall for at least 12 hours outside of this dry reference, the measurements in this dry period are considered as faulty zero rainfall measurements and are discarded. Finally, inter-gauge correlations are determined. If a low correlation (i.e. average and median < 0.21) is found between a station and all other stations the entire time series for that station is excluded. Visual comparison with corresponding radar rainfall time series showed that a filter based on these criteria was suitable in excluding obviously incorrect data from the datasets. This filter could be applied in real-time, although for operational uses beyond this dataset, adjustments will be required.

2.1.2 Radar

As rainfall reference we use gauge-adjusted radar data from a climatological rainfall dataset by the Royal Netherlands Meteorological Institute (KNMI) (Overeem et al., 2009a, b, 2011), freely available as "Radar precipitation climatology" via http://climate4impact.eu. This dataset is based on data from two C-band Doppler weather radars in De Bilt and Den Helder, has a temporal resolution of 5 min and a spatial resolution of 0.92 km^2, covering the entire land surface of the Netherlands. This radar makes volumetric scans in all directions, measuring instantaneous rainfall at a location every 5 minutes. In this product, radar composite images have been adjusted with rainfall measurements from the KNMI rain gauge networks (31 automatic and 325 manual gauges). For details on the method of adjusting, we refer to Overeem et al. (2009a, b, 2011). It should be noted that, due to their different representativeness, there can be significant differences between radar pixel areal rainfall and point rainfall (Schilling, 1991; Einfalt et al., 2004; Villarini et al., 2008; Peleg et al., 2013). Using a radar product that is adjusted with ground measurements, will likely reduce this difference.

2.1.3 Netatmo experimental set-up

As the majority of the weather stations linked to the Wundermap is of type Netatmo, we examine the quality of Netatmo rain gauges in a dedicated experimental set-up, see Fig. 3, photo inset. As reference we use a high quality KNMI pit gauge at the Cabauw Experimental Site for Atmospheric Research (CESAR) (Leijnse et al., 2010), that measures cumulative rainfall in intervals of 12 seconds. This electronic rain gauge is placed in a so-called pit gauge configuration: a small hill of diameter 6.2 m with a circular pit with diameter 3 m and a depth of 40 cm in the middle. Precipitation is collected in the instrument (collecting funnel with a diameter of 16 cm, i.e. 200 cm^2) and in case of solid precipitation melted by a heating element in the funnel. The amount of liquid water is measured by the position of a floating unit connected to a potentiometer. Rainfall is measured every 12 seconds within the range of 0-0.7 mm with a resolution of 0.1 mm and an accuracy of 0.2 mm. The Netatmo sensors are placed at ~40 cm around the electronic sensor in the center of the pit in such a way that the top of each sensor is level with the rim of the pit. The period considered is from February 12th to May 25th 2016. The datasets as collected directly from the Netatmo personal account in mm rainfall per interval of typically 5 minutes, as well as via the Wundermap platform, are compared to the pit gauge reference. One of the stations was offline between April 20th and May 1st and one station could not be accessed via Wunderground.

2.2 Analysis

2.2.1 Station measurement density

As mentioned previously, the original PWS data temporal resolution from Wundermap is quite irregular. The number of stations containing rainfall measurements for time series per 5 and 10 minutes shows that the data availability is quite variable, see Fig. 4. Moreover, the fraction of the measurements over the period that is filtered out does not seem to vary significantly in time. It is not straightforward how to attribute a certain measurement resolution to a network that has highly irregular measurement

frequencies and station locations at irregular distances from one another. When the Amsterdam area is divided into grid cells, or pixels, of a certain size, the number of pixels that contain at least one measurement is an indication of the network resolution. The fraction of total pixels that contain at least one measurement has been calculated for all time steps over the entire period, for various combinations of pixel sizes and time step lengths in the scale range relevant for urban applications. It is found that for the Amsterdam dataset (before filter has been applied), the fraction of pixels containing at least one measurement is more limited by the number of stations than the measurement frequency, see Fig. 5. Only when dividing the period in time steps shorter than 10 minutes, an increase of measurement frequency will result in a higher fraction. This is unsurprising as most stations in the dataset link their measurements to the Wunderground platform approximately every 10 minutes. Adding stations will result into an increase in fraction at all time step sizes in this range. The PWS network consists of more stations than the number examined in this dataset, and continues to grow, which will have a positive effect on the PWS network measurement resolution.

2.2.2 Station measurement quality

With the Netatmo experimental set-up, the performance of this type of PWS and the consequences of transferring its data to the online platform are examined. The measurements are compared to the high resolution pit gauge as well as to the radar rainfall at the corresponding pixel. These two comparisons should give an indication of the differences due to sensor performance and those due to differences in representativeness of radar and rain gauges.

Rainfall measurements of the PWS dataset in Amsterdam are compared with the radar rainfall measurement at their corresponding radar pixels. When comparing station data with gauge-adjusted radar data, the coefficient of variation of the residuals (CV) is calculated. The standard deviation of the differences between the datasets is divided by the mean of the gauge-adjusted radar data. A low value of CV indicates a good match between the datasets. Additionally, spatial correlations between stations are estimated with the use of Pearson's product-moment correlation coefficient (r):

$$r = \frac{E[XY] - E[X]E[Y]}{\sqrt{(E[X^2] - E[X]^2) \cdot (E[Y^2] - E[Y]^2)}} \tag{1}$$

where $E[\cdot]$ is the expectation (estimated as the arithmetic mean) and (X, Y) are corresponding time series of rainfall measurements. Because of the spatial and temporal variability of rainfall, the correlation of two point-locations decreases with distance between these points. A three-parameter exponential function is suggested by Habib et al. (2001) to describe this spatial dependency relation between inter-station correlation (r) and distance (d):

$$r = r_0 \exp\left[-\left(\frac{d}{X_0}\right)^{S_0}\right] \tag{2}$$

where r_0 is the nugget parameter, X_0 is the correlation distance and S_0 is the shape factor. The nugget parameter r_0 is a measure of small scale variability and/or measurement error and is equal to 1 for perfect zero-distance correlation. Correlation

distance X_0 indicates the distance at which the rainfall decorrelates (i.e. the distance beyond which the correlation drops below e^{-1}), which should be interpreted with caution when it exceeds the investigated spatial extent.

The relationship in Eq. (2) is sensitive to rainfall extremes (Habib et al., 2001), climatic regimes (Krajewski et al., 2003) and seasonality (Van de Beek et al., 2011; Tokay and Öztürk, 2012) as well as strongly dependent on time interval (Krajewski et al., 2003; Ciach and Krajewski, 2006; Van de Beek et al., 2011; Tokay and Öztürk, 2012; Van de Beek et al., 2012; Peleg et al., 2013). For the PWS-dataset in Amsterdam, correlograms are constructed and compared with spatial dependencies found in literature. Special consideration is given to the correlations between Netatmo stations as compared to the other types of rain gauges.

3 Results

3.1 Netatmo comparison with pit gauge

The original data of three Netatmo stations (measurement frequency of ~5 min) are compared with pit gauge data (measurement frequency of 12 s) and gauge-adjusted radar data (measurement frequency of 5 min), over the period February-May 2016. Over this period, the cumulative rainfall of station 2 was lower than that of the others, see Fig. 3, left panel. This was the result of station outage. In general, the Netatmo stations measure less rainfall than the pit gauge and radar reference over this period. The scatter plots in Fig. 3, right panel, do not include the intervals where one or both of the time series contains no measurements (in case of station outage), and show a good r^2 of 0.94 between Netatmo measurements and the pit gauge reference. Even though this r^2 suggests a small measurement error in Netatmo, the comparison with radar shows significant scatter away from the perfect fit. This is inherent to comparisons between point locations and pixel averages, and the scatter plot resembles those reported in Peleg et al. (2013), though the radar value used there was an average value of 12 pixels instead of 1.

The correlation between Netatmo and the pit gauge is calculated for a multitude of accumulation intervals, see Fig. 6. This correlation reflects small-scale rainfall variability and thus is closely related to the nugget parameter in Eq. (2). As expected, an increase of correlation is found for larger accumulation intervals. However, the correlations of data from the same devices obtained via Wundermap with the same pit gauge reference show far lower values, see Fig. 6. The original Netatmo data has typical time steps of 5 min against 10 min for the Wundermap data. If this was the only difference between the time series, the correlation graphs should overlap for accumulation intervals above 10 min. As they only approach one another for hourly accumulations, it can be concluded that besides this effect, additional information is lost in the transfer of data between platforms.

Besides the Netatmo dashboard (available to the station owner) and Wundermap, Netatmo data is also accessible from the Netatmo Weathermap platform. In this research, real-time measurements from the three stations in the experimental set-up were obtained with from this platform with an API. It was found that rainfall measurements from this dataset were attributed with a time stamp of the moment the data was collected, instead of the time stamp of the measurement itself. In case of sensor

outage, the last available measurement was collected repeatedly. These artifacts resulted in faulty interval attribution of rainfall and negatively affected the correlations with the original dataset as well as with gauge-adjusted radar data. An API containing such processing errors will result in datasets that contain considerable errors, though these errors are easily overlooked without the original data. Fortunately, the original data can also be obtained from the Netatmo platform. These time series are identical to the data from the Netatmo dashboard (N1, N2 and N3 from the experimental set-up) and can be obtained in real time.

3.2 Amsterdam weather station comparison with radar

Figure 7 shows the double mass plot of the filtered PWS dataset in Amsterdam, as well as the unadjusted (real-time) radar with the gauge-adjusted radar reference at the same locations. The only intervals considered are those where both time series contain measurements. Even though individual stations often do not follow the diagonal line representing a perfect match, the median of all available stations only shows a slight underestimation as compared to the gauge-adjusted radar rainfall data. This underestimation is far greater in the real-time radar product. Though large deviations occur, the median of the stations resembles the reference quite well.

When comparing station rainfall against corresponding gauge-adjusted radar rainfall data over the entire period with the condition that the radar measures non-zero rainfall, a better correspondence is found for longer time steps, see Fig. 8. A similar scatter as in Fig. 8 is found by Peleg et al. (2013). At longer accumulation intervals, the averages resemble each other more, the CV decreases and the r increases, indicating a better resemblance between gauge adjusted radar and station datasets.

3.3 Amsterdam center average comparison

In order to investigate whether the generally poor quality of individual PWS measurements can (partly) be compensated by the generally high quantity of measurements, averages of unfiltered PWS measurements are compared with radar pixel averages over a small area in Amsterdam. The selected area is the region with highest parking rates; the densely populated and touristic area of the city center and Museum square, as floods in this area will heavily impact residents, businesses and tourism alike. This region of ~20 km^2 is shown in Fig. 9, where the cumulative rainfall of each station relative to the mean of the 12 stations is shown. From Fig. 9 the variation between station measurements becomes evident. Some stations measure highly unlikely values considering the measurements of their nearby stations, such as station 3, 9 and 12.

The means of all possible subsets of the 12 PWSs are compared with the average of the 20 radar pixels over the selected Amsterdam center region. For each subset, the correlation, standard deviation and coefficient of variation of the residuals (CV) of rainfall intensity is calculated over all intervals where each station contains measurements. The resulting outcomes of each subset are represented with boxplots in Fig. 10 per number of stations contributing to the PWS-mean. The correlation increases and the standard deviation and CV decrease when averaging multiple stations, even when some of the station time series consist of obviously faulty measurements, see Fig. 10. By averaging the unfiltered measurements of a dozen stations,

crowdsourced measurements seem to be able to describe rainfall in the city center. As expected, the values based on 60 minute rainfall intensities show a better correspondence with gauge-adjusted radar data than 5 minute rainfall intensities.

3.4 Amsterdam weather station spatial correlations

Rainfall variability is often described with correlograms, see Sect. 2.2.2, describing Pearson's product-moment correlation between station pairs as a function of distance. Correlograms of PWS-data at longer accumulation intervals show higher inter-station correlations and the decrease with distance is not as steep, see Fig. 11. This is similar to the results reported by Villarini et al. (2008), Peleg et al. (2013) and Tokay and Öztürk (2012). Especially in winter, see upper panels of Fig. 11, and for short accumulation intervals, the non-Netatmo pairs show higher correlation with one another. However, the goodness-of-fit of the correlograms differs significantly than those found by Villarini et al. (2008), Peleg et al. (2013) and Tokay and Öztürk (2012).

The correlations of all station pairs in the dataset are fitted with the relation in Eq. (2). Fitting was done by determining the nonlinear (weighted) least-squares estimates of the parameters with the Gauss-Newton algorithm. The resulting parameters for the total dataset, as well as winter and summer individually, are given in Fig. 12. The graphs for winter show the most deviating response, suggesting irregularities in this subset in particular. The nugget parameter r_0 of the total dataset varies between 0.50 and 0.67 for this accumulation interval range. Villarini et al. (2008) found a similar nugget parameter of 0.51 for 1 minute accumulations, though far larger values at higher accumulation intervals. The nuggets found by Krajewski et al. (2003): 0.95-0.97 for 15 minutes and longer, Ciach and Krajewski (2006): 0.995 and higher for 1 minute and longer, Tokay and Öztürk (2012): 0.97 and higher for 5 minutes and longer and Peleg et al. (2013): 0.92 and higher for 1 minute and longer, are all considerately higher than the nugget parameters found here. This is not surprising as the gauges in the networks evaluated in those papers are carefully controlled and of higher sensor quality than typical PWSs.

The correlation distance of the total PWS dataset increases with interval size in a similar manner as in previous research, see Fig. 12. The erratic response of the winter graphs suggests a poor fit resulting from other factors than rainfall variability. Likely the correlation distance of stratiform winter rainfall is larger than the spatial scale examined here. The shape parameters do not seem to follow an obvious dependence, similar to Peleg et al. (2013), though other research found this parameter to increase with interval size (Krajewski et al., 2003; Ciach and Krajewski, 2006; Villarini et al., 2008; Tokay and Öztürk, 2012).

4 Discussion

In the experimental set-up in Cabauw, the immediate overlying radar pixel that was first considered as reference turned out to show a significant bias as compared to gauge-adjusted radar rainfall data in all neighboring pixels. The next nearest pixel to the set-up was then used as reference instead. The distance between radar pixel center and experimental set-up thereby increased slightly from 428.9 m to 473.5 m. Faulty measurements can occur in the gauge-adjusted radar dataset, which should be kept in mind when it is used as a reference. When comparing the Amsterdam area radar pixels used in this research to their combined

mean value over the 17-month period, individual time series showed up to 10% consistent higher or lower values. Biases in gauge-adjusted radar could result in a larger spread in Fig. 8, although they have a far smaller influence on the results found in Fig. 10 as in that case the values are averaged.

5 Each aspect of this research, i.e. the Netatmo experimental set-up, the analysis of the station data obtained with the Netatmo API and the Amsterdam PWS dataset from Wundermap, concerned time series over a different, though partly overlapping, time period. As the shorter time series were examined with the purpose of identifying artifacts in the data, those conclusions can be carried over to the longer more robust analyses. The results on PWS data availability, see Fig. 1 and Fig. 5, do not take measurement quality into account. Because of the faulty attribution of rainfall to measurement intervals due to rounding in the

10 data transfer, the measurements in the current form should be accumulated to larger intervals to reduce errors, although this reduces the temporal resolution appreciably. More desirable would be to address the collection method of the PWS data in the platforms in order to maintain the quality of the original PWS rainfall measurements before data transfer.

The filter applied on the PWS dataset in this paper was based on all stations in the dataset. For operational purposes, the

15 median value that is used as a selection criterion should be based on nearby stations only. Large rainfall values were excluded based on a limit on maximum rainfall of 50 mm h^{-1} above the median rainfall at all PWSs at that interval. This potentially excludes rainfall with plausible return times: we take the example of a 10 minute interval during which the median rain intensity of the stations is 4 mm h^{-1}. A measurement of 54 mm h^{-1} and higher would then be excluded, though this corresponds to an event that would occur statistically every 1.5 years (Buishand and Wijngaard, 2007). Because of the small spatial scales and

20 the lack of extremely heavy precipitation in this dataset, the current filter was applicable, as confirmed by visual comparison with gauge-adjusted radar data.

Although a large fraction of the PWS-networks consists of Netatmo stations, this does not imply similar performance of these datasets. Factors like placement and maintenance are unknown and not necessarily equally interfering with the measurements.

25 Even less meta-data is available on the other PWS-types in the dataset, since information on data transfer and the sensors used are not provided for those PWSs. It is expected that there is a positive correlation between the purchase costs of the PWS and the importance of maintenance and high quality measurements to its owner, although this assumption could not be examined with our dataset. Furthermore, the location of the station is based on the setting provided by the PWS-owner, although these may be faulty due to inaccurate localization, rounding of the longitude and latitude or relocation of the station at a later time.

30 Even when relocations of PWSs are accurately provided to the Netatmo platform, this is not automatically communicated to Wundermap, resulting in inaccurate time series for that location. This issue is found to arise in the PWS-dataset, though the filter criterion regarding minimum correlation with the other stations excludes time series of those stations entirely.

Different spatial correlation parameters between studies are to be expected due to different climates, rainfall types, gauge

35 network density and quality. However, the nugget-parameter r_0 (1 for perfect correlation between time series) found here is

significantly lower than in other studies. Additionally, the nugget values of the Amsterdam dataset are significantly lower than the correlation found between the Netatmo datasets with the electronic rain gauge reference in the experimental set-up when the data was obtained via the Wundermap platform, see also Fig. 6. This suggests the interference of additional factors besides sensor measurement errors and data transfer rounding when rainfall measurements are gathered in a less controlled manner. Such factors could be measurement errors due to station placement and poor maintenance.

It is important to note that, even though gauge-adjusted radar rainfall is used as a rainfall reference, differences with point measurements are to be expected because of representativeness errors. Ideally, a high-density gauge network could be used to improve this rainfall product in the future. A non-identical match should therefore not directly be interpreted as negative. However, as the nugget parameter from the station analyses was considerably lower than could be explained by rainfall variability alone, differences with gauge-adjusted radar data here are likely mainly caused by errors in the PWS-dataset. Besides data transfer errors that heavily influence the nugget parameter, the installation errors (e.g. due to shielding), that are minimized in the experimental set-up, further decrease the nugget in the Amsterdam dataset. When comparing nuggets from the experimental set-up and the Amsterdam dataset in the left panel in Fig. 12, the correlations found in the former do indeed reach higher values than those influenced by installation errors in the latter.

5   Conclusions

The resolution and quality of crowdsourced PWS rainfall measurements from the platform with the most dense PWS-network were analyzed to establish whether this data source allows urban hydrological applications. Although the required resolutions (as described by Schilling (1991), Berne et al. (2004), Emmanuel et al. (2012) and Ochoa-Rodriguez et al. (2015)) are not yet achieved by the current PWS-networks, the density of these networks is expected to increase. As the resolution of the current network in Amsterdam is more limited in the spatial resolution than the temporal resolution, the expected continued growth of PWSs that share rainfall measurements via online platforms will yield a network approaching the desired resolutions. This offers a vast contrast compared to KNMI's automatic rain gauge network which, in the Amsterdam metropolitan area, only measures rainfall at one location outside of the city (at Schiphol airport).

From comparisons between Netatmo rainfall time series in an experimental set-up that reduces the errors due to faulty installation to a minimum, the measurements closely resemble those from the high resolution electronic rain gauge. Larger differences are found with radar rainfall, likely due to differences in representativeness between pixels and point measurements. Although the sensor performance of this largest contributor of data in the PWS-network considered in this research looks promising, there is a significant loss in accuracy due to transfer of data to the online platform Wundermap. In this study, the daily cumulative rainfall values as obtained from Wundermap are rewritten as the difference in rainfall as compared to the previous time step. Wundermap cumulative daily rainfall can only become non-zero when at least 0.3 mm rainfall has been collected, and later increases are only registered if they amount to at least 0.2 mm. Especially in case of light rain, rainfall could

occur for a longer period than the interval length in which the daily cumulative rainfall increases. The rainfall is then attributed to a single interval instead of all previous intervals in which it may have been raining as well. This causes significant errors at small timescales. These errors result in inter-station correlations that were considerably poorer than those found in literature, especially in winter and at short accumulation intervals.

The median rainfall of the Amsterdam PWS-dataset shows less systematic bias than the real-time available radar product. Averaging PWS time series further improves correlation, standard deviation and coefficient of variation with the averaged gauge-adjusted radar rainfall in a certain region (~20 km^2). Provided that the degree and likelihood of overestimation of rainfall by PWSs is similar to the degree and likelihood of rainfall underestimation, as was the case in our Amsterdam city center dataset, a dense subset of PWSs can provide good rainfall estimation over a small area, even for intervals of 5 min and without applying a quality filter.

The largest obstacles for the use of crowdsourced PWS datasets are the errors resulting from data transfer, errors due to poor maintenance and faulty installations (i.e. at shielded locations). The rounding of cumulative daily rainfall measurements occurring in the Wundermap platform and the time stamp uncertainty of measurements obtained from platforms with faulty API's lead to considerable errors in the time series, which are only reduced at large accumulation intervals. For the purpose of a high-quality rainfall measurement network with PWS-data, these issues need to be addressed first. Processing errors can be avoided by obtaining raw data from the Netatmo Weathermap platform, though the station density is slightly lower than that of the network linked to the Wundermap. When the processing of data is no longer interfering with the quality of the datasets, the potential of PWS-platforms becomes significant. It provides rainfall measurements from all over the world that are easy to collect, located in rural areas as well as in cities, with station densities and coverage exceeding those from national weather services, and growing towards a level matching the reported resolutions that are required for urban hydrological applications.

Acknowledgements. This research was performed as part of the RainSense project, funded by the Amsterdam Institute for Advanced Metropolitan Solutions (AMS) and the SMART city project (project number 13760) funded by Netherlands Technology Foundation (STW). The data was made available by Weather Underground, and subsequently the weather enthusiasts sharing their weather data with the online community at the online platform Wundermap. The authors would like to thank Marcel Brinkenberg of KNMI for his assistance with the experimental set-up of the weather stations at the Cabauw Experimental Site for Atmospheric Research (CESAR). Thanks is also due to Tom de Ruijter from MeteoGroup for providing data and insight in weather measurements obtained with the Netatmo API.

5 References

Bell, S., Cornford, D., and Bastin, L.: The state of automated amateur weather observations, Weather, 68, 36–41, 2013.

Bell, S., Cornford, D., and Bastin, L.: How good are citizen weather stations? Addressing a biased opinion, Weather, 70, 75, 2015.

Bell, V. A. and Moore, R. J.: The sensitivity of catchment runoff models to rainfall data at different spatial scales, Hydrology and Earth System Sciences, 4, 653–667, 2000.

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.

Bruni, G., Reinoso, R., van de Giesen, N. C., Clemens, F. H. L. R., and ten Veldhuis, J. A. E.: On the sensitivity of urban hydrodynamic modelling to rainfall spatial and temporal resolution, Hydrology and Earth System Sciences, 19, 691–709, 2015.

Buishand, T. A. and Wijngaard, J.: Statistiek van extreme neerslag voor korte neerslagduren [Statistics of extreme rainfall for short durations], Royal Netherlands Meteorologic Institute, 2007.

Ciach, G. J. and Krajewski, W. F.: Analysis and modeling of spatial correlation structure in small-scale rainfall in Central Oklahoma, Advances in Water Resources, 29, 1450–1463, 2006.

Einfalt, T., Arnbjerg-Nielsen, K., Golz, C., Jensen, N.-E., Quirmbach, M., Vaes, G., and Vieux, B.: Towards a roadmap for use of radar rainfall data in urban drainage, Journal of Hydrology, 299, 186–202, 2004.

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

Estévez, J., Gavilán, P., and Giráldez, J. V.: Guidelines on validation procedures for meteorological data from automatic weather stations, Journal of Hydrology, 402, 144–154, 2011.

Fabry, F., Bellon, A., Duncan, M. R., and Austin, G. L.: High resolution rainfall measurements by radar for very small basins: the sampling problem reexamined, Journal of Hydrology, 161, 415–428, 1994.

Gharesifard, M. and Wehn, U.: To share or not to share: Drivers and barriers for sharing data via online amateur weather networks, Journal of Hydrology, 535, 181–190, 2016.

Gires, A., Onof, C., Maksimovic, C., Schertzer, D., Tchiguirinskaia, I., and Simoes, N.: Quantifying the impact of small scale unmeasured rainfall variability on urban runoff through multifractal downscaling: A case study, Journal of Hydrology, 442, 117–128, 2012.

Habib, E., Krajewski, W. F., and Ciach, G. J.: Estimation of rainfall interstation correlation, Journal of Hydrometeorology, 2, 621–629, 2001.

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

Krajewski, W. F., Ciach, G. J., and Habib, E.: An analysis of small-scale rainfall variability in different climatic regimes, Hydrological Sciences Journal, 48, 151–162, 2003.

Leijnse, H., Uijlenhoet, R., Van de Beek, C. Z., Overeem, A., Otto, T., Unal, C. M. H., Dufournet, Y., Russchenberg, H. W. J., Figueras i Ventura, J., Klein Baltink, H., et al.: Precipitation measurement at CESAR, the Netherlands, Journal of Hydrometeorology, 11, 1322–1329, 2010.

Liguori, S., Rico-Ramirez, M. A., Schellart, A. N. A., and Saul, A. J.: Using probabilistic radar rainfall nowcasts and NWP forecasts for flow prediction in urban catchments, Atmospheric Research, 103, 80–95, 2012.

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, p–575, 2014.

5 Mazzoleni, M., Verlaan, M., Alfonso, L., Monego, M., Norbiato, D., Ferri, M., and Solomatine, D. P.: Can assimilation of crowdsourced streamflow observations in hydrological modelling improve flood prediction?, Hydrology and Earth System Sciences Discussions, 12, 11 371–11 419, 2015.

Meier, F., Fenner, D., Grassmann, T., Jänicke, B., Otto, M., and Scherer, D.: Challenges and benefits from crowd-sourced atmospheric data for urban climate research using Berlin, Germany, as testbed, in: ICUC9 - 9th International Conference on Urban Climate jointly with

10 12th Symposium on the Urban Environment, 2015.

Muller, C. L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A., and Leigh, R. R.: Crowdsourcing for climate and atmospheric sciences: current status and future potential, International Journal of Climatology, 35, 3185–3203, 2015.

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

15 investigation, Journal of Hydrology, 531, 389–407, 2015.

Overeem, A., Buishand, T. A., 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, T. A.: Derivation of a 10-year radar-based climatology of rainfall, Journal of Applied Meteorology and Climatology, 48, 1448–1463, 2009b.

20 Overeem, A., Leijnse, H., and Uijlenhoet, R.: Measuring urban rainfall using microwave links from commercial cellular communication networks, Water Resources Research, 47, 2011.

Overeem, A., Leijnse, H., and Uijlenhoet, R.: Two and a half years of country-wide rainfall maps using radio links from commercial cellular telecommunication networks, Water Resources Research, 52, 8039–8065, doi:10.1002/2016WR019412, 2016.

Peleg, N., Ben-Asher, M., and Morin, E.: Radar subpixel-scale rainfall variability and uncertainty: lessons learned from observations of a

25 dense rain-gauge network, Hydrology and Earth System Sciences, 17, 2195–2208, 2013.

Schilling, W.: Rainfall data for urban hydrology: what do we need?, Atmospheric Research, 27, 5–21, 1991.

Steeneveld, G. J., Koopmans, S., Heusinkveld, B. G., Van Hove, L. W. A., and Holtslag, A. A. M.: Quantifying urban heat island effects and human comfort for cities of variable size and urban morphology in the Netherlands, Journal of Geophysical Research: Atmospheres, 116, 2011.

30 Tokay, A. and Öztürk, K.: An experimental study of the small-scale variability of rainfall, Journal of Hydrometeorology, 13, 351–365, 2012.

Van de Beek, C. Z., Leijnse, H., Torfs, P. J. J. F., and Uijlenhoet, R.: Climatology of daily rainfall semi-variance in The Netherlands, Hydrology and Earth System Sciences, 15, 171–183, 2011.

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

480 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.

Wolters, D. and Brandsma, T.: Estimating the Urban Heat Island in residential areas in the Netherlands using observations by weather amateurs, Journal of Applied Meteorology and Climatology, 51, 711–721, 2012.
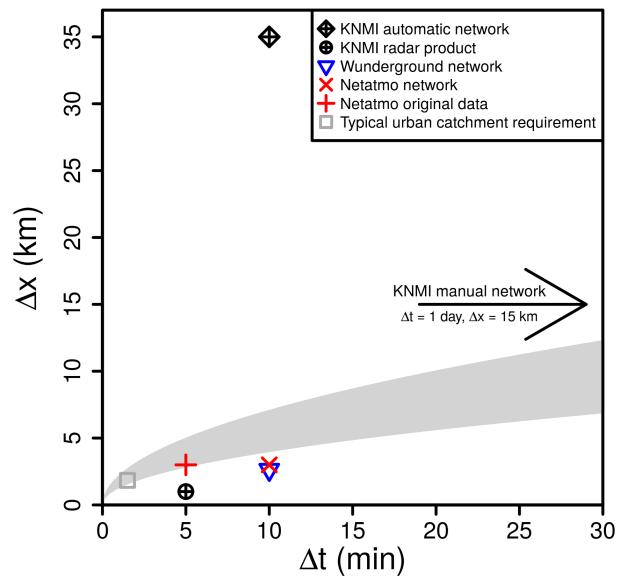
Figure 1. Temporal and spatial resolution of unfiltered rainfall measurements in the Netherlands with PWS network obtained via Netatmo API, Wundermap API and the potential availability of Netatmo measurements, as well as the resolution of KNMI's automatic and manual rainfall measurement network and radar product. The curve represents a relation between the temporal and the spatial resolution of rainfall measurement required for urban hydrology as determined by Berne et al. (2004) for Mediterranean climate, where the square represents the value for an urban catchment with surface area of $0.1 \ \mathrm{km}^2$.
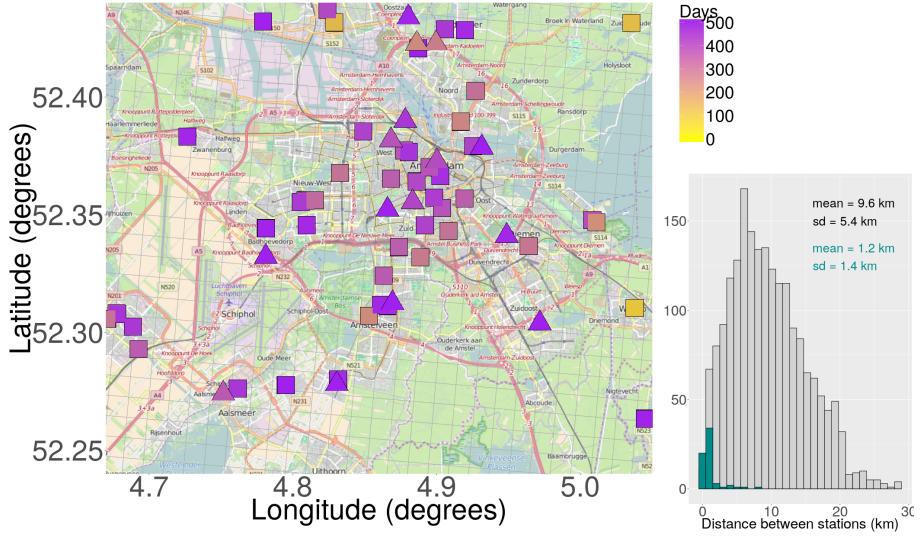
Figure 2. Locations and operational days (i.e. days with measurements) of Netatmo (squares) and other types (triangles) of PWSs, with the radar pixel grid in the Amsterdam metropolitan area. Inter-station distances are represented in the histogram, colored green for nearest neighbor distances.
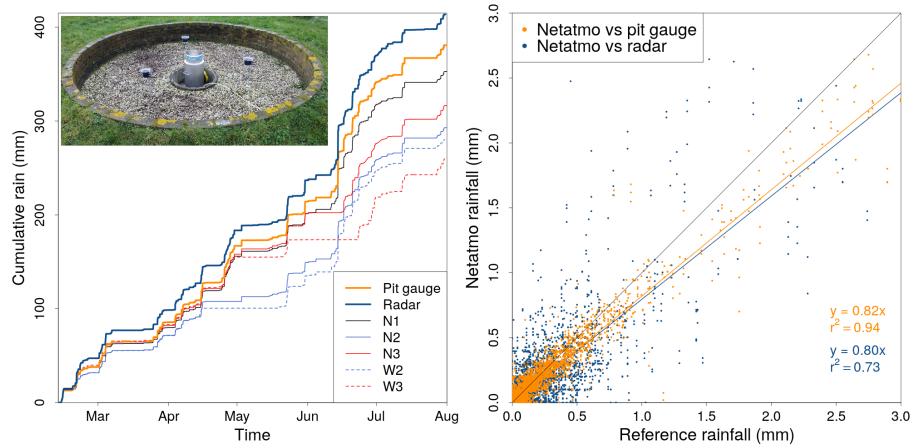
Figure 3. Left panel: Cumulative rainfall according to reference pit gauge, gauge-adjusted radar, Netatmo stations (N1, N2 and N3) and Netatmo stations obtained via Wundermap (W2 and W3). N2 (and as a consequence W2) was offline between April 20th and May 1th. Photo shows the experimental set-up of the rain gauges in the pit gauge configuration. Right panel: scatter plots of 10 minute rainfall and linear fits of rainfall according to N1, N2 and N3 as compared to reference pit gauge (orange) and gauge-adjusted radar (blue).
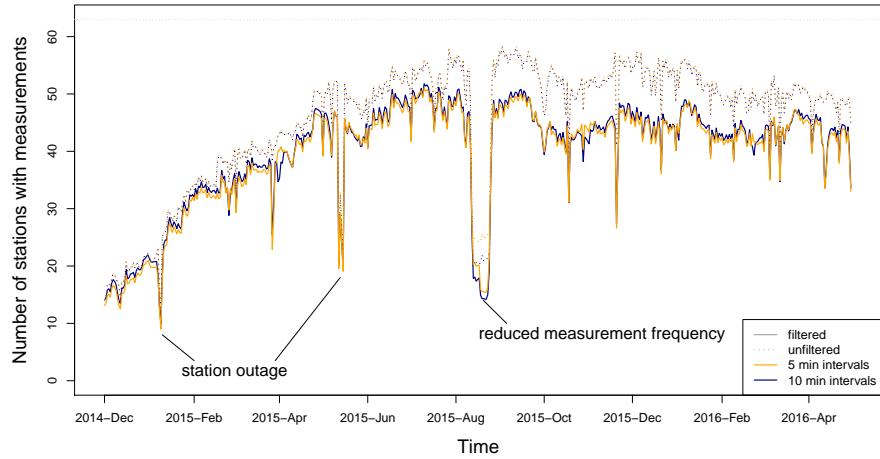
Figure 4. Number of stations with rainfall data from the PWS dataset, before and after applying filter, for every 5 and 10 minute interval over the entire period, smoothed per day. The two indicated dips correspond to complete outage of stations, the third with a longer period of fewer measurements in all stations.
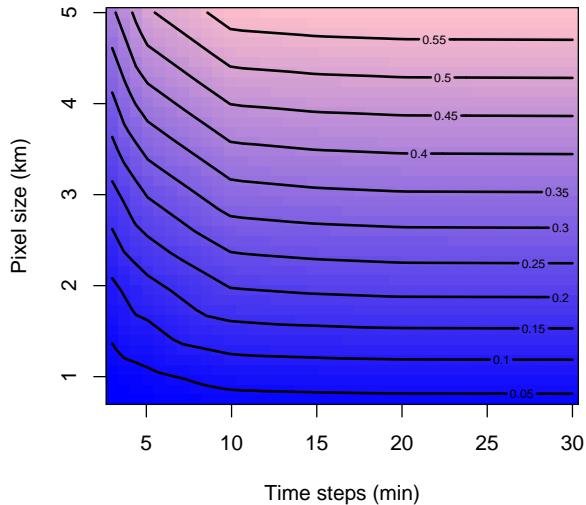
Figure 5. Indicated with curves as well as colors are the fractions of pixels containing at least one measurement of the unfiltered dataset for combinations of time step length and pixel grid size over the Amsterdam area between December 2014 and April 2016.
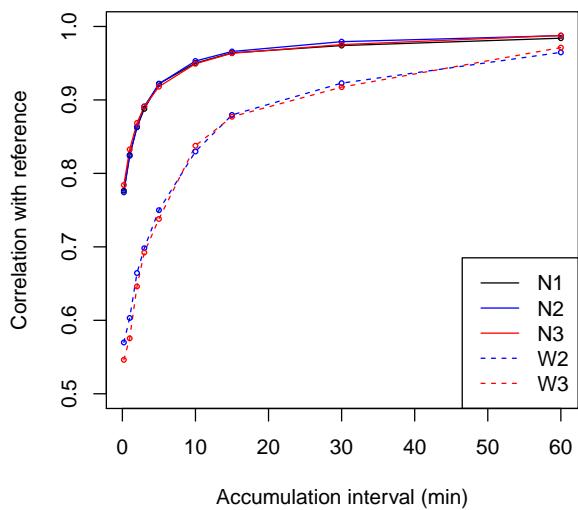
Figure 6. Correlation between rainfall measurements by Netatmo stations as obtained via personal dashboard (N1, N2 and N3), also obtained via Wundermap (W2 and W3), and the pit gauge reference for various accumulation time steps.
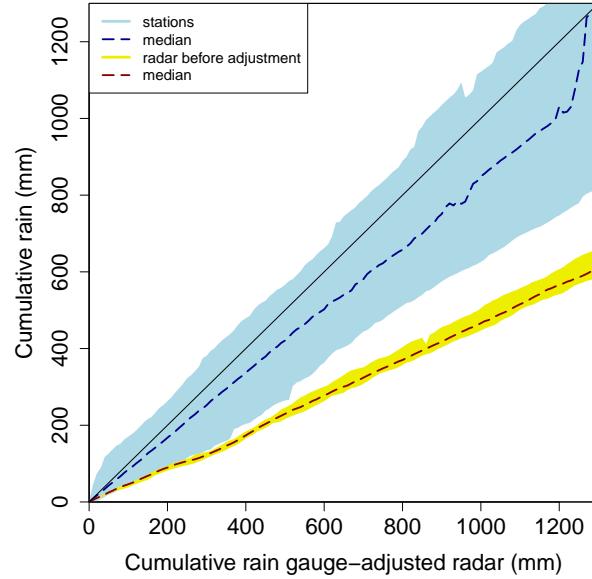
Figure 7. Double mass plots of station filtered rainfall measurements and real-time radar data with gauge-adjusted radar rainfall at the corresponding location in the period between December 2014 and March 2016. Only intervals where both radar and station contain measurements are taken into account. Colored regions indicate the range between double mass plot of stations with minimum and maximum steepness and dashed lines represent the median of the combined datasets.
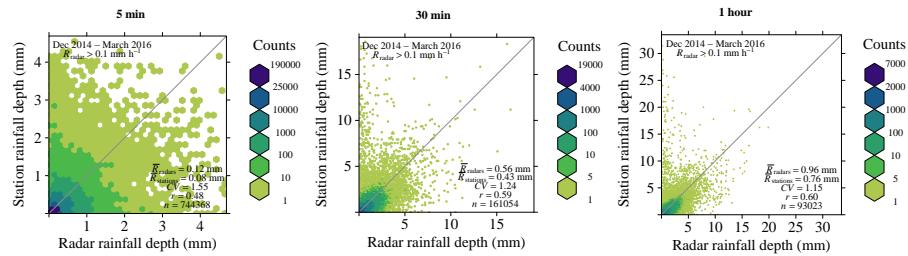
Figure 8. Scatter density plots of all station rainfall measurements against the gauge-adjusted radar rainfall data in the corresponding radar pixel when radar reported non-zero rainfall (> 0.1 mm). The $\bar{R}_{\text{radars}}$, $\bar{R}_{\text{stations}}$, CV, r and n values in the panels represent the average rainfall according to the gauge-adjusted radar data, the average rainfall according to the stations, the coefficient of variation of the residuals, the correlation and the number of intervals, respectively. Graphs are made for 5 min, 30 min and hourly accumulation intervals.
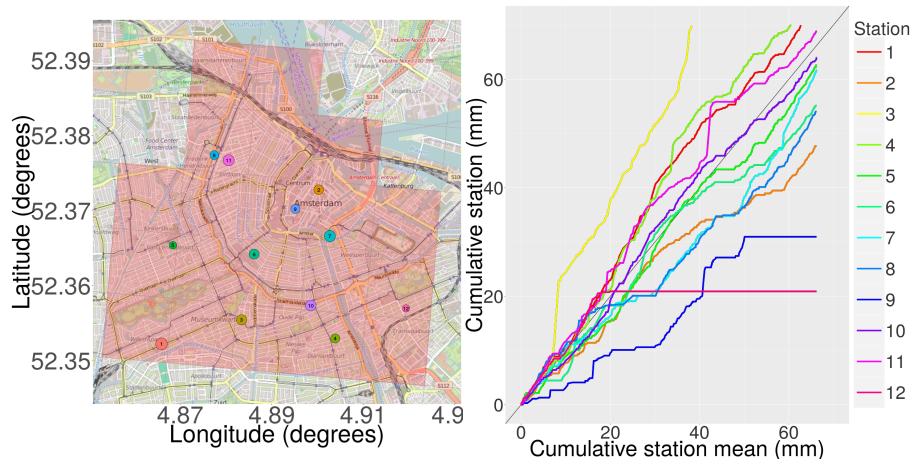
Figure 9. Left panel: Locations of 12 stations and 20 radar pixels in the city center of Amsterdam, where symbol size represents the number of unique days with measurements by the station (range 371-514 days). Right panel: Double mass plots of the station measurements as compared to the mean of all 12 stations over the intervals where all stations contain measurements.
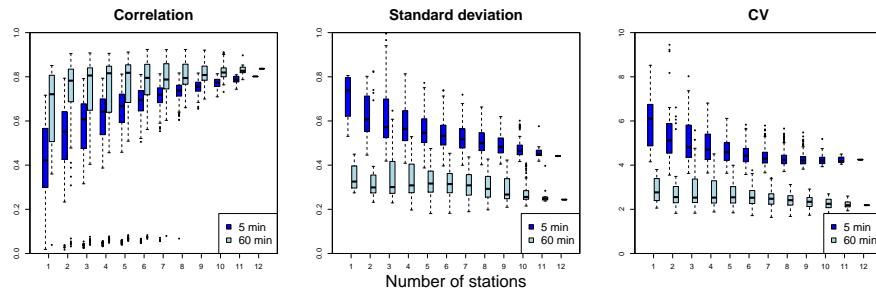
Figure 10. Boxplots of correlation, standard deviation and coefficient of variation of residuals (CV) of averaged rainfall intensity time series. The boxplots contain the outcomes for all possible subsets within the 12 stations in Amsterdam city center, as compared to the gauge adjusted radar rainfall intensity 20 pixel mean for interpolated 5 minute and hourly time series.
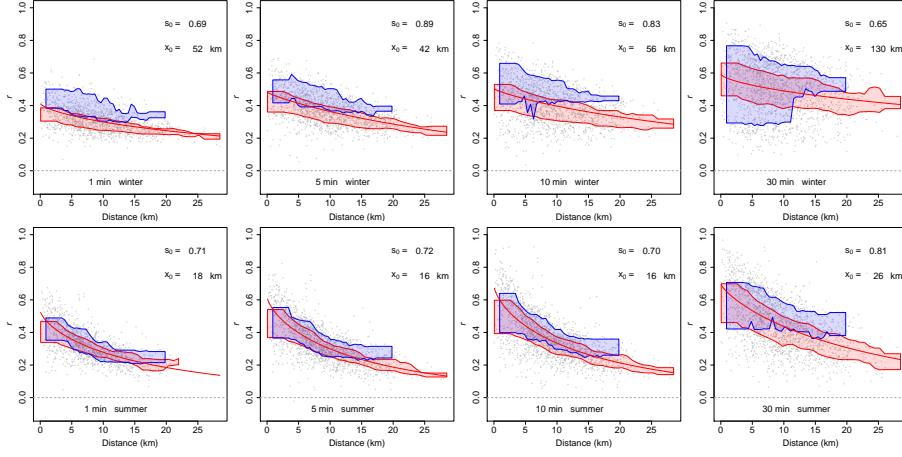
Figure 11. Correlograms of all stations after filtering at various accumulation intervals for winter (top panels) and summer (bottom panels). The red and blue areas represent the interquartile range of the Netatmo stations and non-Netatmo stations, respectively. The areas are constructed with a moving window of width 5 km. The scatter plots are fitted with the exponential relation of Eq. (2), parameters of which are given in the panels.
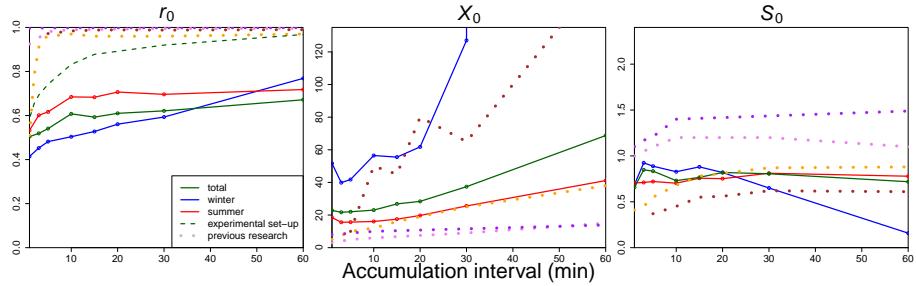
Figure 12. Time scale dependency of nugget (left panel), correlation (middle panel) and shape (right panel) parameters from fit described in Eq. (2), for total PWS dataset, as well as winter and summer only. Dotted lines represent values found in previous research by Peleg et al. (2013) (violet), Villarini et al. (2008) (orange), Tokay and Öztürk (2012) (brown) and Ciach and Krajewski (2006) (purple), where the dashed line in the first panel shows the timescale dependency of the Netatmo station nugget found in the experimental set-up as previously shown in Fig. 6.