

## Review

*Title: The use of personal weather station observation for improving precipitation estimation and interpolation*

*Author(s): András Bárdossy et al.*

*MS No.: hess-2020-42*

*MS Type: Research article*

*Reviewer: Dr L.W. de Vos*

## Summary

In this paper rainfall observations from personal weather stations (PWSs) are used to describe the spatial interpolation of heavy rainfall events, while taking into consideration the typical unknown errors and biases of the observed rainfall estimations from PWSs. The authors present a novel approach to use PWS observations to better capture the variability of heavy rainfall events than a primary sensor network is able to. The method is evaluated on a substantial dataset of almost 200 intense events in Baden-Württemberg in Germany. The Netatmo PWS network can provide rainfall observations at densities that traditional operational sensor network typically lack. Making use of this data source for extreme events can therefore be highly valuable.

The paper is well-written with few grammatical errors and high-quality figures, and addresses an interesting problem (high resolution rainfall monitoring of heavy rainfall events) in a novel way. The paper would benefit from clarifying some choices and assumptions in the methodology, and describing the resulting limitations thereof. Also, some steps in the method could be presented in a more straightforward way. When these issues are addressed, the resulting paper will be highly useful and relevant.

## Major comments

P2L42-49: In the introduction, the paper of de Vos et al. (2019) is discussed, which proposes a quality control (QC) methodology for PWS rainfall observations. This QC has been applied on a Netatmo dataset in the Amsterdam metropolitan area as well as a national dataset of the Netherlands. The PWS measurements were compared with a gauge-adjusted radar product before and after the QC was applied, in order to demonstrate the improved accuracy. This validation showed that the QC was successful in flagging intervals with errors, without the need for auxiliary data, and thereby reducing the bias, increasing the Pearson correlation coefficient and reducing the coefficient of variation, by excluding the flagged intervals (12% of the original dataset of a year in the Amsterdam metropolitan area).

The gauge-adjusted radar product in that study was merely used as a ground-truth, not as a vital part in the QC. There may have been some confusion as there is one parameter in that QC methodology, the *DBC*, which is a proxy value used in the QC to compensate the overall bias in the network, that was determined in an offline exercise using the gauge-adjusted radar product. It is also clearly mentioned in the paper that this *DBC* does not need to be estimated in that manner or at all, as proven by the accuracy improvement after QC was applied when *DBC* was chosen as 1 (i.e. no proxy was made at all), see de Vos et al. (2019) Table 2.

Therefore, the statement that the QC described by de Vos et al. (2019) is "based on combined official rain gauge and radar product" is faulty, and the statements regarding the limitations and uncertainties in radar QPE in this context are irrelevant. Moreover, the statement "the study by de Vos et al. (2019) does not provide a guideline on how to use the measurements of the PWS if no radar observations are available" is wrong, and the statement "data quality issues have to be overcome" in this line of reasoning is debatable. Also in the Discussion and conclusion section it is wrongly stated "...or relies on other data

sources as reference, such as precipitation estimates from weather radars which have an appropriate spatial and temporal resolution (de Vos et al. 2019)."

The study presented in this paper does not lose its relevance by the existence of a proven QC methodology that identifies erroneous PWS rainfall observations without auxiliary information, as it has merit in its specific focus on extreme precipitation events. Nevertheless, the previous studies that are described to reflect the current state of the art in this field need to be represented accurately.

The proposed method is interesting and promising, however there are some significant limitations due to the assumptions in the filters. It can be considered contradictory that the main perceived issue with the QC in previous work (mistakenly) is its dependence on another data source, while this methodology relies on the availability of another data source itself. The PWS are used as an addition to a high quality primary rain gauge network with long observation series in the study area of interest, measuring in high temporal resolution. Such a network may not be readily available everywhere, and this should be mentioned in the discussion more broadly than it is now.

Another important limitation of the applicability of this method is that the selection of stations to be included needs to be done over a considerable period for the high intensity based filtering to work. This assumes constant and steady PWS performance, and doesn't allow for changes in measurement accuracy in time (e.g. due to interference to the station or temporary blockage of the Netatmo rain gauge tipping bucket system). A station is either included or excluded in its totality, while previous research on Netatmo observations has shown that measurement accuracy can change drastically instantly for the better or worse. This should be included in the discussion as well, including the minimum required period over which the event should extend in order to reliably apply the method.

The paper is very limited in describing how the data is gathered from the Netatmo rain gauges, which measure approximately every 5 minutes. The unprocessed time series that can be collected with the Netatmo API do typically not have fixed time steps and can contain large data gaps. The paper is not clear on how these raw time series are processed into structured aggregated time series at 1, 3, 6, 12 and 24 hour time steps, but does mention in the evaluation of Netatmo data from the experimental set-up with a Pluvio sensor an error resulting from station connectivity. This error is difficult to understand without knowing the process that the authors have used.

## Minor comments

- P1L18: "in situ sensors" -> "in-situ sensors"
- P1L19: "low cost personal weather stations" -> "low-cost personal weather stations"
- P1L20: "Netamto" -> "Netatmo"
- P2L39: "...a PWS rain gauge." -> "...three PWS rain gauges."
- P2L49: "real time flood forecasting" -> "real-time flood forecasting"
- P2L50: "two fold approach" -> "two-fold approach"
- In some number values the thousands are indicated with comma, in some cases not.
- Both "Figure", "Fig" and "Fig." is used in the text.
- P4L76: "time" -> "period"
- Caption Figure 1: "Map Of" -> "Map of"
- P4L77: "one can see that many stations have less than one year of observations" -> how does that follow (from figure 2 or elsewhere), and why is the proposed methodology not able to accommodate these stations?
- Section 2 would benefit from more quantitative descriptions of the measurement uncertainty of the sensors that are mentioned, e.g. from technical documentation of these sensors from the supplier.
- P5L96: "Since is..." -> "Since it is..."

- P5L103: "Note that  $Y$  is considered to be a random field, and thus methods like Co-Kriging or Kriging with an external drift are not applicable." -> the purpose of this statement in this context is not entirely clear to me.
- P5L105: probability ( $\alpha$ ) in Eq (1) needs to be explained more fully.
- Section 3.1 describes that a secondary station is flagged as suspicious if its indicator correlations with the nearest primary network points are below the lowest indicator correlation corresponding to the primary network for the same time steps and at the same separation distance. I can imagine that not all distances between secondary station and nearest primary network points equal a separation distance between two primary network stations exactly. Is then the nearest distance used? If so, what are the largest differences between separation distances? Or is the relationship between distance and correlation ( $\rho$ ) described with a fitted relation (effectively a correlogram)? If so, what is then the meaning of "min" in Eq. (2)?
- P7L163: "... due to unforeseen events (such as battery failure or transmission errors) at certain times they may deliver individual false values." -> How is the issue of data gaps in Netatmo time series addressed? Here it seems to be referred to as "false values", however it should be evident from the Netatmo time series that an observation was lacking (due to a long duration between the timestamps of two subsequent observations). I wonder if regarding these observations as zero observations and subsequently identifying them with a simple geostatistical outlier detection method is the best approach. The author's may refer to the station in total (not a certain period in observations), which due to battery failure or transmission errors is considered to be faulty. If that is the case, which fraction of the data should be missing for a station to be considered a geostatistical outlier? A later section (P9L224-229) hints at problems due to data gaps which resulted in a large outlier, but it's not clear if these cannot be avoided by looking at the timestamps of the PWS observations. More information on how the raw irregular Netatmo PWS datasets are converted to timeseries with fixed time steps would be very helpful.
- P7L174: "...which are due to temporary loss of connection between the rain gauge module and the Netatmo base stations." There are many other possible reasons for false zeroes. The QC in de Vos et al. (2019) has dedicated an entire module to those types of errors (see the red marked observations in the left graphs in Figure 1 of that paper for an indication how often these occur). Note that those FZ-errors are not related to station outage, as those intervals are already excluded at an earlier step in the analysis.
- P8L209: "judgment" -> "judgement"
- Table 1: I would be very interested to see if all three Netatmo stations have yielded observations every  $\sim 5$  min, and if there were data gaps, how many and how long these were. I assume the stations were not calibrated and have the default tipping bucket volume of 0.101 mm. It says the statistics are based on non-0 values. Does that mean that both Pluvio and Netatmo station need to measure non-0 rainfall, or only the reference (Pluvio)? This should be specified in the footnote of the table. Here  $p_0$  likely refers to probability of precipitation, which is only mentioned later (P10L235). It should be introduced earlier in the text, including the equation to calculate it. Also consider including other metrics in the table like RMSE and correlation using the Pluvio observations as ground truth to validate the Netatmo stations.
- Consider using a different symbol than  $p_0$  for probability of precipitation (P10L235), as it is very similar to the symbol used for correlation ( $\rho$ ).
- P9L218: "on" -> "one"
- Figure 4: why are the lines of the Secondary Stations stepped and the Primary Stations not?
- P11L244: "equations (2)" -> "Eq. (2)" ?
- Caption Figure 5 -> "Xes" -> "crosses"
- P11L257: "low intensity" -> "low-intensity"
- Table 2 caption: I assume that  $p_0$  still refers to probability of precipitation. Is it then the fraction of intervals where precipitation is larger than 0.1 mm? In that case it makes more sense to change the text in the table from "<0.1 mm" to ">0.1 mm". Also, "(mean of all stations and events)" is not very clear in this context, please explain.
- P12L260: "Note the high portion of zeros" -> where can this portion be found? It doesn't seem to be provided in Table 2. Should this be portions of intervals where precipitation is <0.1 mm?

- Table 2: what was the procedure to select these events?
- P12L265: "the temporal filter" and "the event based spatial filter" probably refer to 3.1 and 3.3 respectively. It would be helpful to name those two filters explicitly in the method section and uses those names throughout .
- P12L274: "Pearson ( $r$ ) and Spearman ( $\rho$ ) correlation" -> up until now I would have assumed the correlation that was introduced in section 3.1 to be the Pearson correlation. However, as the symbol  $\rho$  was used in that section, that was likely actually Spearman. Either way, it should be specified in section 3.1. Also, what is the motivation to evaluate two types of correlation?
- Section 4.2: It is explained that two references are constructed using cross validation. Reference 1 is constructed by interpolating the subsets with only primary network stations, and Reference 2 is constructed by interpolating the subsets with primary and secondary network stations. What is the reason for constructing two references? From their captions it seems that Table 3 and 4 are based on comparisons with Reference 1. Is Reference 2 used somewhere else?
- P14L315: "(Fig. 6 e))" -> avoid double brackets.
- P15L321: "on" -> "in"
- P17L333: "(panel b))" -> avoid double brackets.
- P17L334: "This is caused by the reduction of the variability with increasing number of observations" -> Is that true? Why would the variability of a rainfall event be dictated by the number of observations in space? It seems to refer to the more smooth rainfall patterns found at daily scales compared to hourly scales, but this phrasing is confusing.
- P19P370: "in higher" -> "at higher"
- P19P370: "point precipitation" -> "precipitation as a point value"
- P20L381: "real time availability" -> "real-time availability"
- P20L399: "This corresponds to the 0.99 non-exceedence probability of precipitation for the specific secondary station." -> how does this follow from the information that is provided? Or is this provided information?
- P20L400: "The precipitation quantiles at the primary stations corresponding to the 0.99 probability are 3.2, 3.5, 3.1 and 3.0 mm." -> how does this follow from the information that is provided? Or is this provided information?
- Some interesting additional literature to refer to could be: <https://www.nat-hazards-earth-syst-sci.net/20/299/2020/nhess-20-299-2020.pdf> on the use of Netatmo data for describing deep convection features. Also, the QC method <https://github.com/metno/TITAN> could be mentioned in addition to the QC method of Napoly et al. in the introduction. Finally, Chen et al. (2018) "Trust me, my neighbors say it's raining outside: Ensuring data trustworthiness for crowdsourced weather stations." is an example for quality estimation of PWS rainfall data from the Wundermap platform.