Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-42-AC4, 2020

Hydrology and Earth System Sciences Discussions
Open Access

Interactive comment on "The use of personal weather station observation for improving precipitation estimation and interpolation" *by* András Bárdossy et al.

András Bárdossy et al.

jochen.seidel@iws.uni-stuttgart.de

Received and published: 27 March 2020

We thank Lotte de Vos for taking the time to review our manuscript thoroughly. Regarding the summary, we' like to clarify that we investigated 955 individual events (about 200 for each duration), not only 200.

Our response to the major comments:

P2L42-49 ff.

We apologize for the misinterpretation of the paper of de Vos et al. (2019). After careful rereading we recognized that our interpretation was wrong, and we'll correct

Printer-friendly version

Discussion paper

the corresponding paragraphs in the revised version of the paper. The filtering is in fact not requiring the actual radar product. On the other hand the bias correction filter SO requires the radar product for the previous time period. This is itself is subject of errors. Further please note that the validation of the precipitation amounts is done on the basis of the radar product, for which the uncertainty and inaccuracy plays an important role. The SO filter provides a kind of regional bias correction, our transformation is correcting each station individually as we have observed that even within a small region significant positive and negative biases may occur. The filters FZ and HI are very similar to our second event based filters. The first filter requires at least a few months of observations - this is a disadvantage, but on the other hand it provides an overall judgement of the individual PWS. As the second filter is applied for each event to all stations which passed the first filter. Thus there is little risk that occasionally bad measurement are not rejected. Our filter is in fact rather strict (conservative) as we remove many stations. We need further work to find the best selection of useful PWS and for the bias correction.

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.

It is true that high quality primary measurements might not be available everywhere. We are testing the methodology on smaller primary datasets to quantify the usefulness of the PWS network.

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.

We will describe the data used and the processing more clearly in the revised manuscript. The data we downloaded using the Netatmo API did have regular 5-min timesteps, however these we're not always continuous. Such gaps in the data were filled with NaNs. These data were then aggregated to 1h sums and by keeping the NaNs, i.e. any 1h-aggregation with NaNs in-between was considered as NaN. We compared the frequencies of the zero observations of the primary and secondary network and did not find significant differences. This means that the problem of providing 0-s for nan-s was negligible in our case (but we did find occasional occurrences of false zeroes when comparing the 3 Netatmos with the reference at our weather station). Moreover, since each PWS station was verified individually, the missing data were always taken into account and the corresponding data from the primary network were considered. All analyses in the study are based on hourly precipitation sums, and all other aggregations were based upon these.

Minor comments:

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?

We will clarify this in the revisions by adding a figure showing a histogram of the time lengths of the PWS stations. Furthermore, a certain time length (2 months excluding the winter months) is required for the filters to work.

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.

We will address this aspect in the revisions.

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.

Correctly: Y is not a stationary random field as the measurement bias and uncertainty can differ from one station to the other. Meanwhile we found a way to use Co-Kriging - after using a transformation. The reference for non-collocated Kriging is in our response to Reviewer #3. The manuscript will be modified accordingly.

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 (ρ) described with a fitted relation (effectively a correlogram)? If so, what is then the meaning of "min" in Eq. (2)?

Each secondary station has a single closest primary station. The indicator correlations are calculated based on the whole time series (after removal of the NaNs) of these pairs . The indicator correlations using all pairs of primary stations are also calculated using exactly the same timesteps. We assume that the indicator correlations of the primary stations represent the *true* spatial variability of precipitation. Thus we compare these clouds and reject all secondary stations where the correlations are below those primary pairs within a distance window with a tolerance. The tolerance is needed for close pairs of primary and secondary stations. We do not calculate indicator correlations for pairs of secondary stations.

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 timesteps would be very helpful.

As mentioned above, all missing time stamps in the downloaded data were flagged as NaN (not 0). The timesteps from the data we downloaded are in regular 5-min intervals. We will describe our data processing more clearly in the revised manuscript. In table 1, only 1h timesteps where all devices (i.e. the three Netatmo and the Pluvio reference) have valid data were considered.

Figure 4: why are the lines of the Secondary Stations stepped and the Primary Stations not?

Because of the different resolution of the rain gauges, i.e. Netamo 0.1mm and Pluvio 0.01mm

Table 2 caption: I assume that p0 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?

We will clarify this in the revision.

Table 2: what was the procedure to select these events?

The intense rainfall events were selected from the observation of the primary network. For each temporal aggregation, we investigated the highest 200 intense events. These were selected regardless of the observed location or time. For the cross validation procedure, only events without nugget variograms were chosen, this is why for each temporal resolution the final number of events was slightly less than 200.

P12L274: "Pearson (r) and Spearman (ρ) 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 ρ 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?

As the distribution of precipitation amounts is skewed the Pearson correlation may be strongly influenced by a few high values. The Spearman correlation is independent of the distribution and shows whether the ranks of the observations were correctly reproduced. As our method is strongly based on rank based assumptions it is reasonable to consider it. The text will be revised to recognize which correlation was actually used.

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?

This seems to be a misunderstanding. These sets are not references these are the interpolations - we used a cross validation approach and both interpolations are compared on the observed primary dataset (every time for the stations not considered).

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.

Our wording is in fact confusing - we meant with the increase of aggregation (the number of 5 min data considered) the fields become smoother. We'll correct this in the manuscript.

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?

The quantiles are derived from the distributions based on the time series of the primary stations. For this example we assumed that these are the corresponding values.

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.

Thank you for pointing out these references, we will consider them in the Introduction.

The other minor remarks will be considered while preparing the revised manuscript.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-42, 2020.