“The use of personal weather station observation for improving precipitation estimation and interpolation”

by András Bárdossy, Jochen Seidel and Abbas El Hachem

Review by Dr. Marc Schleiss, Dept. of Geoscience and Remote Sensing, Delft University of Technology, the Netherlands

General assessment:
This is the third time that I review this paper. The paper has improved. However, there are still a lot of typos and unclear sentences and the writing could be improved further. Most of the major issues I raised during the previous rounds were (partially) addressed. The only major points of criticism that I have left are:
- the justification of the assumptions in Section 4.1, which could be more quantitative and exhaustive.
- the conclusion section, which is too short and does not include all major findings.
- the structure of the paper. In particular, the event selection procedure and cross-validation strategies which should not be in the results part but introduced earlier in the text, in the methods section.

Minor comments and typos:

- Conclusions: Your conclusion section is really short and does not do justice to all the work that you have done. I suggest to extend it. For example, you do not mention two crucial points which are: a) the gain in performance when using KU compared with OK and b) the fact that important assumptions where made during the development of the filters, such as stationarity, preservation of the ordering of the data and neglecting of other sources of errors such as local wind effects.

- “*Although this is only one example with a relatively short time period it does support our assumption that the quantiles between primary and secondary stations are similar for higher precipitation intensities. However, one secondary device (N10) delivered data which deviates substantially from the other measurements. This was caused by an interrupted connection between the rain sensor and the base station. In this case, the total sum of precipitation over a longer time period was transferred at once (i.e. in one single measurement interval) when the connection was established again. This leads to an extreme outlier which falsifies the results.*”

> Not sure to understand your argument here. According to your assumption, the points in Figure 4b should align with each other (though not necessarily along y=x). Still, there seems to be substantial residual scatter and uncertainty due to quantization effects (especially for Netatmo). Please provide some quantiative metrics to judge the degree of linear relationship and highlight which data point in 4b corresponds to the “extreme outlier”. In addition, it would be worth commenting on the discretization effects you see in the Netatmo stations.

- ll.73-74: “*The number of secondary stations is higher in densely populated areas ~~are~~ such as in the Stuttgart metropolitan area and the Rhine-Neckar Metropolitan Region between Karlsruhe and Mannheim.*”

- ll.103-104: “*It is assumed that this precipitation is measured by <u>the</u> primary network [..]*”

- ll.105-107: *"The basic assumption for the suggested quality control and bias correction method is that the measured precipitation data from the secondary network may be biased in their values* <u>*but correct in terms of their order*</u> *~~they are good in their order~~"*

- l.119: <u>"As a first step in quality control, all PWS ~~locations~~ with ~~notoriously contradicting~~ notoriously inconsistent rainfall values are removed."</u>

- ll.149-150: *"In order to have a sufficient sample size and to have robust results, high α values and low temporal aggregations Δt are preferred."*

Can you be more specific? What are sufficiently large values for alpha and delta t?

- ll.163-164: *"Another possibility is to interpolate the quantiles corresponding to selected non percentiles or interpolating percentiles for selected precipitation amounts."*

Not clear. Please reformulate.

- ll.272-273: *"Furthermore, one can observe that the* <u>*differences*</u> *~~deviation~~ between the reference and the Netatmo gauge are not linear, ..."*

- l.276: *"Figure 4 shows that for high percentiles their occurrence is the same for the primary and the secondary devices."*

This sentence is not clear. Please reformulate.

- ll.289-291 *"While the distributions differ, the probability of no precipitation p0 (defined as precipitation < 0.1 mm) ranges from 0.90 to 0.91 and is thus very similar for both types of stations indicating that the occurrence of precipitation can be well detected by the secondary network."*

Actually, in Table 1, the percentages p0 (at 1h resolution) are 0.84 for N07 and N10, which is 7% lower than for N11 (0.91) and 8% lower than for the Pluvio (0.92). Please explain!

- l.302 *"In our case, 862 secondary stations remained after the application of the IBF."*

In addition to the number, please specify the percentage of stations that were removed.

- ll.316-317 *"The secondary station in the centre recorded 1.7 mm* <u>*of*</u> *rainfall"*

- ll.332-334 *"The cross validation was carried out for a set of different temporal aggregations Δt and a set of selected events. Only times with intense precipitation were selected, as for low-intensity cases the interpolation based on the primary network is sufficiently accurate"*

Actually, you did not show any evidence that the interpolation for lower intensities is accurate. Please provide some numbers or reformulate this sentence.

- ll.338-348: the detailed description of the CV method and different configurations and metrics used during evaluation could be moved to the methodology section.

- ll.359-360: *"The measured and interpolated results were also compared for each event in space and (r) and (rS) and the observed the interpolated spatial patterns were calculated as well"*

This sentence makes no sense. Please reformulate!

- l.376 *"The use of KU for interpolation resulted only <u>in</u> a minor improvement"*

- l.379: *"In this case, OK with secondary data did not lead <u>to</u> an improvement"*

- ll.380-381: *"Stations located very close to each other can cause instabilities in the solution of the Kriging equations leading to high positive and negative weights"*

Are you referring to the screening effect? Please clarify and provide a reference to a textbook to clarify what you mean by "stabilizes the solution" on l.382. KU. Would adding a nugget effect in the variograms help model the small-scale differences you see between PWS data? Please discuss!

ll.387-388: *"The poor <u>performance</u> ~~peformance~~ of Co-Kriging is surprising, but an appropriate selection of the co variable (for example transformed rank) may improve the results."*

Too speculative. Please provide more details or reformulate this sentence. One explanation could be that co-kriging makes rather strong modeling assumptions (stationarity of both primary and secondary variable). It also requires the estimation and fitting of 3 (cross-)variograms, which increases uncertainty (especially in small samples). You make some other interesting comments about an extension of co-kriging toward the end of the paper. Perhaps you could include these here as well.

- ll.410-411: *"This is a typical case where all methods yield unbiased <u>results</u> ~~resluts~~"*

- ll.446-447: *"However, the results from this study <u>as</u> well as the ones from de Vos et al. (2019)"*

- ll.463-465: *"A detailed cross-validation of different filter combinations and temporal aggregations shows that the IBF is the most important step <u>and</u> ~~as~~ yields the highest improvement in interpolation quality"*

- ll.465-466: *"Furthermore, the performance of the presented method is better ~~a~~ <u>at</u> smaller temporal aggregations"*

- ll.484-485: *"Problems occur if the order of the observations is influenced by wind effects, but due to the highly skewed distribution of the precipitation amounts the problem mainly occurs for small precipitation amounts."*

I don't understand your last argument. Please explain! The way I see it, the wind-induced bias mostly affects high rainfall intensities. Also, its effect will become more visible when quantities are aggregated over time. Wind-induced biases can represent 20-30% and are the main source of uncertainty in in-situ rainfall measurements. PWS tend to be installed in weird places and are particularly prone to this type of errors/biases.

- l.488 *“Furthermore, the near real-time availability of the data of secondary networks may help to improve the quality of flood forecasts.”*

- *l.490 “on the ~~in~~ contrary it often increases uncertainty”*

- l.502 *“In this study, The number of primary stations ~~in this~~ was sufficient to improve the interpolation quality”*

- ll.506-508 *“By applying a rather strict threshold of 5 C average daily temperature, many rainfall events ~~are~~ were rejected. It would be conceivable to include the hourly temperature data from PWS in order to estimate whether a given precipitation event corresponds to ~~of~~ rain or snow ~~at a specific location.~~”*

- Figure 2: Change axis labels. For 2a, put “years” on the x-axis and “number of stations” on the y-axis.

- Figure 4: The axis labels for 4b should be “Quantile Pluvio [-]” and “Quantile Netatmo [-]”

- Equation 1: You need to specify in the equation that this only applies for Y above a certain threshold.

- Table 2: this table shows some basic statistics of the selected events and could be moved to the methods section, together with the text explaining how events were selected and how cross-validation was performed. I don’t think that putting it in the results section is a good choice.

- Table 7 is interesting. But the discussion going with it is very short. You could expand this part and provide more discussion about the pros/cons of your approach compared with other faster, simpler and deterministic alternatives. I’m relieved to see that KU performs better than IDW and NN. But it’s a close call and the lower performances of NN and IDW are mostly due to their higher biases compared with KU. If you would compare the methods on a fair basis, for a similar level of bias, would you still see significant differences in RMSE? Indeed, the bias in IDW can easily be reduced by performing hyperparameter optimization of the distance decay parameter or choosing a different distance metric. So there’s definitively room for improvement. On the other hand, there also seems to be some room left for further optimization of the KU technique. For example, you could optimize the uncertainty parameter linked to the PWS data. In the paper, you arbitrarily use 10% but this could be tuned to the dataset as well (using LOOCV).