“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 second time that I review this paper. As I said, I find this topic very interesting and relevant. The large number and variety of available PWS data has created lots of new opportunities and generated a strong need for new, robust interpolation/merging methods. The authors have some good ideas for how to approach the problem. However, I don’t think that their study is ready to be published yet. The main points that need to be improved are 1) the writing, 2) the presentation of the results and 3) the description of the methods. Below, please find a list of suggestions for how to improve.

In addition, I should point out that there are 2 major comments from the previous round of review that were not fully addressed during revision. These are:
a) More details about the kriging part → The authors responded to this comment but not of all their explanations can be found in the revised paper. Please make sure that all important details are in the text so that others can reproduce what you did!
b) Comparison of kriging with simpler, faster alternatives such as inverse distance weighted interpolation or bilinear interpolation → Partially done but results are not shown and there’s only a few short sentences in the paper about this, without any numbers or critical discussion about the pros/cons.

Recommendation: Major Review

Major comments:
Note that all referenced page/line numbers refer to the revised manuscript <u>with track changes</u>.

1) Please clearly state the main main conclusion of your paper in the abstract and conclusions. Right now, this is not 100% clear. Is the conclusion that careful QC and bias-correction has to be performed before PWS precipitation data can be used? If that’s the case, then this is not really new. Other studies have already shown the same and your method is just another way to do this. So what exactly is your contribution? Please clarify!

2) Your method is rather complicated. Yet several of its components do not seem to significantly improve performance. For example, the EBF filters and the KU do not make a big difference. So why did you feel the need to include them in the methods and results? It just makes the paper longer and more complicated and forces you to introduce a lot of theory and notations for no obvious gain in performance. I suggest to shorten the paper and only keep the essential parts of the algorithm in the methods section. If you want, you can always write a short section or paragraph summarizing the results for some other options/filters that you think could be useful in other contexts.

3) The number of peer-reviewed studies about PWS and their use in hydrometeoroloy is still limited. A few of them have already been mentioned in the literature review. But overall, the introduction of the paper remains rather short. I suggest to extend this part by providing a more in-depth analysis and discussion of the state-of-the-art related to the use of citizen gauges in quantitative precipitation

estimation problems, including its challenges, similarities with other fields and open questions. For example, some parts of the Discussion (i.e., the differences/similarities with radar-gauge QPE) could be moved to the introduction. Also, I encourage the authors to explicitly state which aspect(s) of the problem their study is meant to address. What's the main contribution? Is it the method itself or is it the lessons learned and/or recommendations for a successful interpolation/merging of PWS data?

Suggested references (non-exhaustive):

- Canli, E., Loigge, B. & Glade, T. Spatially distributed rainfall information and its potential for regional landslide early warning systems. *Nat Hazards* **91,** 103-127 (2018). https://doi.org/10.1007/s11069-017-2953-9

- Reges, H. W., Doesken, N., Turner, J., Newman, N., Bergantino, A. and Schwalbe, Z. CoCoRaHS: The Evolution and Accomplishments of a Volunteer Rain Gauge Network, *Bull. Amer. Meteor. Soc.* (2016) **97** (10): 1831–1846. https://doi.org/10.1175/BAMS-D-14-00213.1

- Simpson, MJ, Hirsch, A, Grempler, K, Lupo, A. The importance of choosing precipitation datasets. *Hydrological Processes*. 31: 4600-4612 (2017). https://doi.org/10.1002/hyp.11381

- Starkey, E., Parkin, G., Birkinshaw, S., Large, A., Quinn, P. and Gibson, C. Demonstrating the value of community-based ('citizen science') observations for catchment modelling and characterisation, *Journal of Hydrology*, 548, 801-817 (2017). https://doi.org/10.1016/j.jhydrol.2017.03.019.

- Yang, P., & Ng, T. L. Fast Bayesian regression kriging method for real-time merging of radar, rain gauge, and crowdsourced rainfall data. *Water Resources Research*, 55, 3194– 3214 (2019). https://doi.org/10.1029/2018WR023857

4) The writing and structure of the Results section need to be improved. The current strategy for assessing/validating the different components of the method is not clear to me. Right now, analyses/results are presented in seemingly random order, with rather vague qualitative statements and lots of circumstantial evidence. A better, more precise, quantitative and targeted evaluation would greatly increase the quality of the paper. For example, you could consider a step-by-step, hierarchical assessment of the different components (e.g., the IBF filter, the bias correction and the interpolation/merging), with different scores and subsections for each part.

5) Figure A1 is crucial for understanding how the bias adjustment method works. I suggest to move this from the Appendix to the main text, together with the corresponding explanations. Actually, I don't think you need an appendix at all!

6) Table 3 does not show correlations (which should be between -1 and 1). Please correct.

7) The step-by-step description of the algorithm is a good idea. But it's really hard to follow, even for somebody familiar with the geostatistical jargon. More work is needed to streamline this and make it clear. A flowchart of the whole method would help, with different symbols for filters, adjustments and interpolations! Also, you could shorten the text by grouping some of the smaller steps together into larger modules or tasks. The details of each task can be given in the different subsections of the methodology.

8) The crucial assumption behind your method is that for high precipitation intensities, the ranks of the secondary stations are correct. Some superficial analyses in Section 4.1 suggest that this assumption is probably not too bad. But since this is such a critical hypothesis, it should be assessed in much more detail. Please extend Section 4.1 and perform more tests designed to assess how good this ordering assumption really is. For example, your could compute rank correlation coefficients for different thresholds, stations and lengths of time series. Or you could look at fluctuations over time or as a function of distance. To better understand the limitations of your method, it could also be good to show a few cases for which the assumption does not hold.

9) I have some issues with the terminology chosen by the authors, especially regarding the EBF (Event-based filter). I think this is a poor choice of words. In reality, the EBF filter is a spatial filter for one particular aggregation time period (and not an event). More generally, I don't think that it is a good idea to use the word "event" to refer to a particular aggregation time periods. This is not standard practice and might be confusing to many readers. Please modify accordingly.

10) Regarding the bias correction scheme: If I understood the approach correctly, the idea is to use the percentile of the PWS observations (secondary network) to estimate the equivalent precipitation estimates of the professional gauges (primary network) and then spatially interpolate this value to the location of the PWS using kriging. On top of the large uncertainty that comes with estimating a percentile from a short PWS series, one problem with this approach is that it uses the ordering assumption multiple times (i.e., once for each pair or PWS and professional gauge). This greatly increases the chances of errors during bias correction due to imperfect modeling assumptions. Also, the final spatial interpolation may re-introduce bias due to smoothing and/or modeling choices. So my question is: why don't you just pool the professional rain gauge data together into a single distribution and directly adjust the PWS observations using quantile-quantile mapping on the pooled data? In this way, you would use the ordering assumption only once and you would not have to interpolate at all, which is likely to be faster and more robust. By the way, you can pool data even if the time series of the professional gauges have different lengths. Please explain why you think the current approach is better!

11) A substantial part of Section 5 (Discussion) from lines 434-455 is not a discussion but just a summary of the method and therefore should be moved to the conclusions. The last part of the discussion (ll.467-475) about the similarities/differences of PWS with radar measurements. This is out of scope here because not part of the analyses. I suggest to shorten this and/or move it to the introduction. Please use the discussion section to analyze pros/cons, mention alternatives or new ideas for follow-up studies.

12) Conclusions, ll.501-503: Wind has a major effect on precipitation measurements, leading to a systematic undercatch. This may influence the order of data, but the effect is the same for the primary and secondary network."

> I do not agree with this statement. Literature shows that wind effects tend to be very local. Sometimes, both gauges will be affected by the same bias. But often, it's likely that the PWS and professional gauges will have different biases. More importantly, wind-induced biases will fluctuate over time and space, which affects the rank statistics and the performance of the IBF and bias correction schemes. There's not much that you can do about this. But at least, you should properly acknowledge the problem and discuss its possible consequences in the text. I suggest to do this in Section 5 (Discussion) rather than the conclusions.

13) On a personal note: PWS stations tend to cluster in/around urban areas. Spatial interpolation methods such as kriging do not always perform optimally on highly clustered data. For example, it is well known that clustering can lead to screening effects and highly negative kriging weights. This does not necessarily lead to wrong estimates but decreases robustness and accuracy. I am aware that this goes beyond the scope of this study. Still, I invite the authors to briefly mention this issue in the Discussion section and to point to possible ways to overcome it in future work. This is particularly relevant for small-scale estimates of heavy precipitation.

Minor comments and typos:

- In the abstract, please specify what you mean by secondary observations. I assume it's the observations from the PWSs.

- Introduction: “In recent years, the amount of low-cost personal weather stations (PWS) has increased with an incredible speed”.

    Incredible is not a good choice of words here. Please reformulate.

- Introduction, ll.24-26, “This is potentially very useful to complement systematic weather observations of national weather services, especially with respect to precipitation, which is highly variable in space and time”.

    Please add a few references at the end of this sentence to support your statement.

- Introduction, ll.28-29, “In consequence, the number of interpolated precipitation products with sub-daily resolution is low, but such data ~~would be~~ are required for many hydrological applications (Lewis et al., 2018)”

- Introduction, ll.29-31,“Additional information such as radar measurements can improve interpolation (Haberlandt, 2007), however, radar rainfall estimates are ~~is~~ still highly prone to different kinds of errors (Villarini and Krajewski, 2010) [...]”

- Introduction, ll.33-34, “However, one of the major drawbacks from PWS precipitation data is their trustworthiness”

    Please add a few references to support this statement.

- Introduction, ll.36-37, “The measured data itself may have unknown errors which can be biased and contain independent measurement errors, too.”

    This sentence is not clear. Please reformulate and be more specific.

- Introduction, ll.45-47, In a more recent study, de Vos et al. (2019) developed a QC methodology of PWS precipitation measurements based on filters which detect faulty zeroes, high influxes and stations outliers based ~~ona~~ on a comparison between neighbouring stations.

- Section 2, l.69, “The gauges used in this network are typically weighing gauges”.

Do you mean predominantly? In addition, please specify the type of weighing gauges (e.g., the model, brand or serial number).

- On l.119, you mention that the random variable Y is not stationary. Yet, on ll.144-145 and Equation 2, you refer to its cumulative distribution function F, without any dependence on time. Please clarify this apparent contradiction.

- Equation 4, what's your definition of "nearly" at the same separation? Please specify!

- l.168, "Under the assumption that the temporal order of precipitation at secondary locations is correct"

- ll.168-172, "Under the assumption that the temporal order of precipitation at secondary is correct (eq.1), one could have used rank correlations instead of the indicator correlations. The indicator approach is preferred however, as the sensitivity of the devices of the primary and secondary networks is different and this would influence the order of the small values strongly. Furthermore, random measurement errors would also influence the order of low values. In order to have a sufficient sample size and to have robust results, high α values and low temporal aggregations Δt are preferred."

Or you could just say that the ordering between the primary and secondary networks needs to be the same for values above a certain threshold.

- Section 3.3, ll.215-216 "Instead one can use rank based methods for this purpose as suggested in Lebrenz and Bárdossy (2017) and rescale the rank based variogram ~~variogramm~~"

- Section 3.5, ll.279, "Interpolate precipitation for target grid using all remaining values using OK or KU."

Bad English, please reformulate

- Section 4.2, ll.375-376 "Decreasing spatial variability and increasing regularity with increasing time aggregation is the reason for these differences."

I am not sure to understand what you mean by regularity. Please reformulate to make this clear.

- Section 4.3, ll.395-397 "Note that for this data the cross validation based on the primary observations showed an improvement of r from 0.36 to 0.77, of r S from 0.55 to 0.76 and a reduction of the RMSE from 12.5 to 8.2."

Units for the RMSE values are missing. Same for line 409.

- Section 4.3, ll.398-399 "Figure 7 shows the distributions of the cross validation errors for the different interpolations for this event. This is a typical case where all methods yield unbiased results ~~resluts~~"

- Section 5, ll.440-442 "This approach uses a comparison of the data with those of the nearby stations to remove unreasonable values, a separate procedure to identify and remove false zeros and another ~~one~~ filter to find unreasonably high values."

- Section 5, ll.452-455 “The use of secondary stations after filtering and data transformation improves the results of interpolation for other possible interpolation methods, such as nearest neighbour or inverse distance weighting. However, in this study these methods yield worse results than OK (results not shown here).”

> Not clear. Please provide more details. For example, you could give the average reduction in terms of RMSE or increase in correlation for each interpolation method.

- Figure 1: Please add a scale! Same comment for figures 6, 8, 9,10

- Figure 3: Please use different symbols for N07, N10 and N11 to better distinguish the points.

- Figure 4: Please specify the 3 primary and 4 secondary stations in the caption and how far away they are from each other.