Comment on hess-2021-640

Anonymous Referee #3

Referee comment on "Enhancing the usability of weather radar data for the statistical analysis of extreme precipitation events" by Andreas Hänsler and Markus Weiler, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-640-RC3, 2022

General comments:

The authors provide a quite novel approach to estimate design rainfall from weather radar in Baden-Württemberg (BW). The main idea is the pooling of data from radar pixels in the proximity of the target cell to increase the sample size beyond the 19 years available record length. The radar data are bias corrected and compared against two station-based data sets, the German design storm standard KOSTRA and a regional data set from BW.

Although the approach is quite heuristic with several arbitrary assumptions and decisions (e.g., search radius, local estimation method, interpolation of parameters over durations, independence assumptions, etc.) it is practically pragmatic and statistically satisfactory. There are some major issues which need further attention and discussion. The first is the selection of events within a small search radius with the assumption of spatial independence. Second, the many arbitrary assumptions need to be better justified. Third, there is a need for better and more formal description of the methods.

Altogether the idea is good, the results are interesting and plausible. The text is significant and reads quite well although the English could be improved. I would recommend publication after major revision.

Specific comments:

- Title: The title suggests a postprocessing of radar data for a later statistical analysis. However, you have done already the analyses. I would recommend to adjust the title e.g. something like "A pooling approach for design storm estimation using weather radar data – a case study for BW"
- Lines 31: The non-stationarity is not considered in the approach. Of course, with only 19 years of observations this is hardly feasible. However, at least a brief discussion or an outlook should be included.
- Line 89: "Reassembling via running sums" becomes not clear. Usually, the highest temporal resolution of 5 min is used to build the extreme values series for all durations by calculating sums over a moving window with width equal to the duration and moving step of 5 min? Is that the procedure used here?
- Lines 146 ff: I think I could finally figure out how the sampling locations are selected but the description is weak. Please, reformulate and explain better. There are several arbitrary assumptions: why have you selected the normal distribution, how did you define its parameters, how did you select the 0.8 threshold, etc.? These need to be justified and discussed.
- Line 157: Include equation for normalisation.
- Line 168: The "sub-sampling is not adapted for different event durations"; I guess you mean by that, that the same locations have been sampled independent of durations?
- Line 189: What about spatial independence? Is this minimum separation time of 48 hours between events applied on the whole compiled data set from all 5 locations together? Only that way a spatial independence can be assumed. On the other hand, in that case considering the small search radius I would assume, that the sample from the five locations is not really comparable with a real 100-year sample; it probably will contain less extreme events and finally lead to an underestimation, which partly may explain the results.
- Line 197 ff: In the independent fitting of distributions for different durations order relations problems may occur. This is accounted for in DWA (2012) by smoothing the parameters over the durations, which is a bit "old-fashion". Please, explain more in detail which method has been applied here and discuss also alternatives.
- Line 217 ff: Please provide equations for the quantile mapping approach.
- Fig. 3: Are the probability distributions compiled from all stations/locations together? If yes how many stations are included?
- Line 290: What is meant by spatial correlation analyses? Do you refer to correlations between rainfall and elevation? If yes, this is a cross-correlation but not a spatial correlation, which is usually used to quantify spatial persistence by correlation-distance relationships like the variogram, which by the why could have been employed for a more objective selection of the neighbourhood for sampling.
- Fig. 5: Same question as for Fig. 3.
- Line 315: Why are you using an 80% confidence interval here; usually a 90% interval between 5% and 95% quantiles is used?
- Discussion/ conclusion: The new product has been compared against 2 reference data sets, but no strict validation has been carried out as usually desired. This is of course difficult since the truth is not known. However, often the long-term observations (>30 years) are applied as truth in a cross-validation mode. The application of this is also difficult here since the RADKLIM data set itself is a merged product involving these stations which makes this infeasible. At least a discussion of this problematic is required and optimal would be some suggestions for further research.