Review of „Enhancing the usability of weather radar data for the statistical analysis of extreme precipitation events"

Dear Anonymous Reviewer,
thank you very much for your very detailed review of our manuscript and your very valuable suggestions to improve it.
Please find our response to the various points you raised below (in red).
Best regards,
Andreas Hänsler and Markus Weiler

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:

1. 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"

   Thank you very much for this suggestion. However, we would rather like to keep the title since we want to focus rather on the methodological aspects that are needed to post-process rainfall radar to be used for statistical extreme value analysis than on the results of the EVA itself. However, we probably have to make this more clear in our manuscript.

2. 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.

   Thank you very much for pointing this out – we will include this in a revised version of the manuscript.

3. 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?

   Yes, exactly that is what has been done. We will add some flow charts (suggestion of Reviewer 1) to improve the description of the sampling and data processing.

4. 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.

   Yes, we agree that this should be improved. We will also add a flow chart on this to improve the description of the sampling and data processing. The choice of parameters defining the boundaries is indeed somehow subjective. They were basically chosen in a way that we give the sampling process a high degree of freedom adapted to the specific local conditions (so no fixed sampling area like circles or other structures have been prescribed). On the other hand, the parameter choice takes care of the fact that only cells are sampled that are somehow regionally representative for the COI. The resulting combination of probability distribution and threshold gave the best results (see Fig 2), but others were tested.

5. Line 157: Include equation for normalisation.

   We will add this to the flow chart

6. 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?

   Correct, that is what we meant. We will adapt the sentence.

7. 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.

   Yes, for the final dataset of resampled events we applied the 48h criteria.

   The search radius is a compromise between the spatial representativeness of the COI and the inclusion of additional extreme events. We find that through the sampling process a significant increase in the rainfall amount of the top events could be reached. But we also believe that the general bias of the RAD-BC events would still remain, even we would enlarge the larger sampling radius. This is mainly due to the known underestimation of high intensity rainfall events in the radar.

8. 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.

So far, we have not applied any smoothing across the different event durations. This was done to be in close agreement with the method used to generate the BW-Stat dataset. Also, the focus of the paper is more on the combined pooling process and bias-correction as well as the resulting changes in spatial patterns. But of course we could briefly discuss the cross-duration relation (at least briefly in the discussion section) although a parameter smoothing might impact the positive effects the approach has on the varying spatial patterns of design rainfall events for different durations.

9. Line 217 ff: Please provide equations for the quantile mapping approach.
   We will include the equation in the respective flow chart

10. Fig. 3: Are the probability distributions compiled from all stations/locations together? If yes, how many stations are included?
   They are calculated for all grid boxes. BW-Stat and RAD-BC have the same number of grid boxes since they are on the same grid. KOSTRA has a substantially lower number of grid boxes (hence the pdf is less smooth). We will include this information in the figure caption.

11. 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.
   It is actually the cross correlation between the spatial patterns of REGNIE precipitation and the RAD-BC data. We will change this in the manuscript.
   Regarding the variogram, we agree that this could also have been a valuable approach to identify a potential sampling area. However, also with this approach some subjective assumptions have to be made, since the correlation will be highest closest to the COI. But we want to sample in certain boundaries (linked to the size of convective cells). Based on the suggestions of Reviewer 1 we will move the figure describing the sampling statistics into the attachment. But we can add some comparable analysis in the respective figure on how the sampling statistics would change, if a variogram approach would be used.

12. Fig. 5: Same question as for Fig. 3.
   see reply above

13. Line 315: Why are you using an 80% confidence interval here; usually a 90% interval between 5% and 95% quantiles is used?
   Thank you very much for pointing this out. We will change it to 5/95%.

14. 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.

Thanks for pointing this out. Since the methodological differences and the much lower spatial resolution of KOSTRA a one-to-one validation with KOSTRA is not possible. And the BW-Stat dataset is actually a pooled dataset itself.
Furthermore, we already know that both station based datasets have deficits in their spatial patterns caused by (i.) the limited number of stations included and (ii) the explicit consideration of the topography in the interpolation, which is (at least for short duration events) somehow questionable. We mention this problem of not having a clear validation dataset already in the manuscript, but we will make sure that it is better reflected in the discussion of the results.