“Enhancing the usability of weather radar data for the statistical analysis of extreme precipitation events”

By Andreas Hänsler and Markus Weiler

Dear Marc Schleiss,
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

Summary:
A method to estimate spatially varying design storms with a return period of up to 100 years on the basis of short weather radar precipitation estimates and long but low spatial density gauge data is proposed. An example of application for the state of Baden Württemberg, Germany is presented. The main finding is that the spatial structure seems to be more realistic in the weather radar based product. However, despite the bias correction, the absolute magnitude of the radar-derived design storms remains too low, which should be addressed in future studies.

Evaluation:
In my opinion, the strong points of this study are:
a) Radar data are combined with gauge information to derive spatially varying design storms
b) Uncertainty is assessed through bootstrapping and ensembling.
c) The figures are nice.
d) The approach is potentially interesting.

The weaknesses are:
a) The English and the writing could be improved.
b) The paper is too long.
c) There are some parts in the methods section that need to be clarified.
d) The bias correction method used by the authors is controversial.
e) Serious doubts remain about the usefulness/reliability of the new dataset.
f) There are a lot of subjective parameters/choices (e.g., the way the radar grid cells are sampled). No proper sensitivity analysis is performed.

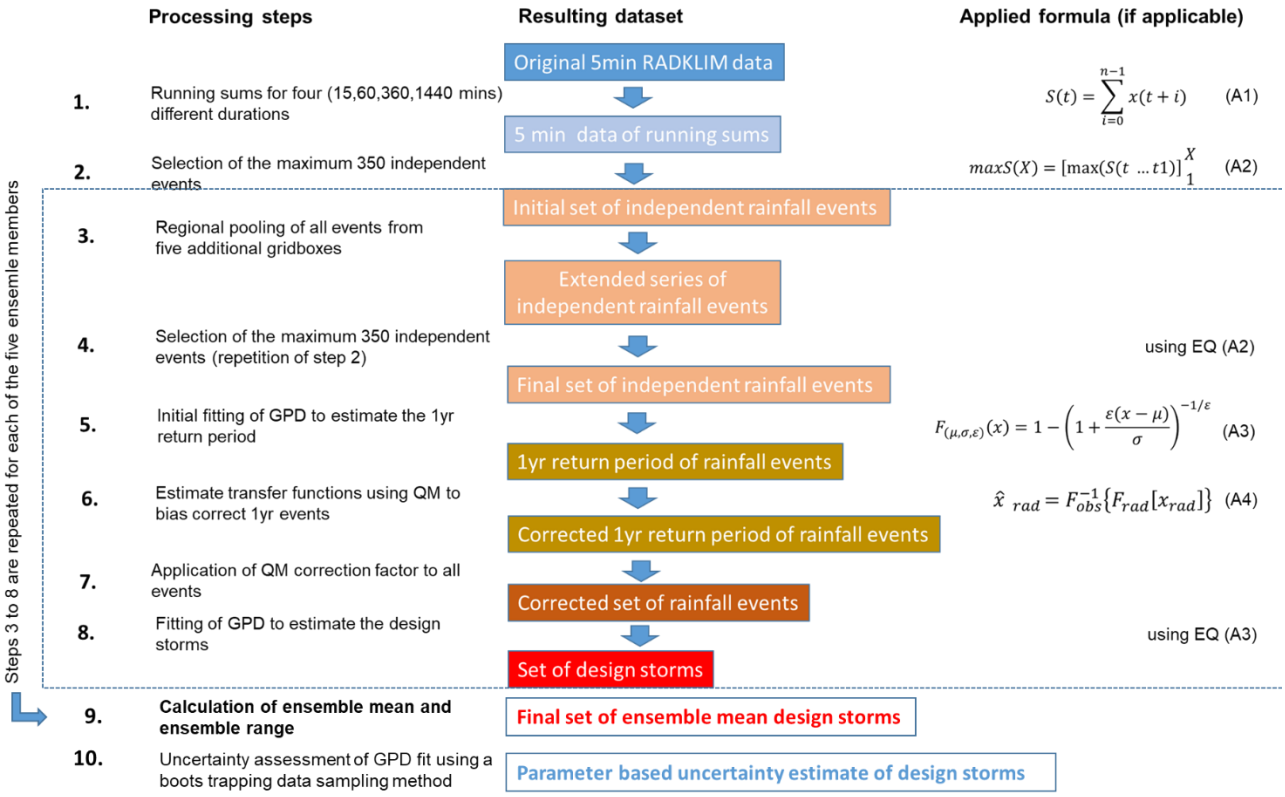
Scientific significance: Good (2)
Scientific quality: Fair (3)
Presentation quality: Good (2)
Recommendation: Reconsider after major review.

Major Comments: (MC)

MC1: A lot of the methods are described in plain text, without the use of any equations or formal mathematical notations. As a result, it is hard to understand what exactly has been done to the data. For example: ll.157-158 ``In a final step, both spatial sampling probabilities are normalized with the respective maximum probability, added together and again normalized by the maximum to generate the final spatial probability distribution for the sampling (Fig. 1, panel bIII).''

Suggestion: Add some flowcharts or equations to complement the text and clarify the different steps, transformations or models used, especially in the methods section. Every step needs to be clearly described and reproducible.

Thank you very much for this very valuable suggestion. Also the other reviewers pointed out that the method section still needs some improvement. So we will add a (i) flow chart to summarize the main data processing (including relevant equations) as well as (ii) a flow chart to detail the sampling a bit more. A quick draft for (i) (just for illustration, still needs some work) is added below. This will certainly also help to shorten the text and clarify the sampling process as well as the data analysis.



MC2: The English is poor and the writing is overly complicated. The paper is too long for its content. Suggestion: substantially shorten the text by removing unnecessary words, expressions or sentences. Get rid of fillers such as “certainly”, “basically”, “already”. Write shorter sentences and go to the point. Refrain from adding too much information in parentheses. Before including parentheses, check if they are essential.

Also the other reviewers pointed out that the language should be improved. We will carefully recheck what we have written and shorten some of the text passages. Since we will include some flow charts we definitely can shorten some of the text passages substantially.

MC3: It is still unclear whether RAD-BC can be used to build reliable regional design storms. Yes, the radar data are spatially more explicit, which removes the need for interpolation. But radar data have two huge disadvantages: 1) they are much shorter and 2) they are biased and more uncertain than gauge data. Given these drawbacks, I would argue that it’s probably safer to build design storms based on a long rain gauge time series, even if the gauge is several hundreds of kilometers away and not very representative of the region of interest. But I might be wrong and the challenge is to find a good balance between having enough data for extreme value analysis and being spatially representative. So which approach is better, and under what circumstances? You could do more to answer this important question! For example, spatial variations in small-scale extremes (15 min) are likely to be small, due to the random nature of convection. It’s only at the daily aggregation time scales and beyond that the

spatial structure really starts to matter (due to orographic enhancement). Does this mean that RAD-BC only has value for longer duration extremes?

This is actually an example where our dataset delivers an added value, but we obviously have to make it more clear in the manuscript. Both station based products that are included for comparison, show some spatial patterns linked to topography already for the short event durations. This is not surprising when considering the fact that the topography was explicitly included in the spatial interpolation of the data (and in the case of the BW-Stat dataset, additionally in the selection of stations to extend the data series). So we believe that especially for shorter durations, our dataset provides an added value (at least in the spatial patterns).

> Suggestion(s): Improve the validation part and clearly explain why/where you think your new RAD-BC dataset is valuable. Look for new ways to assess the effect of interpolation and bias correction. Look at differences in return values as a function of elevation and/or distance to gauges. Consider adding a case-study where you focus on specific locations/areas for which you know (from literature) that current methods provide bad estimates and compare them to what you get using RAD-BC. You need to convince the reader that RAD-BC offers real tangible advantages compared to simply interpolating design storms from gauges.

Thank you very much for the suggestions. Given what we wrote above, we clearly think that there is an added value in the spatial representation of design storms when using radar data. An additional information we can provide, is to compare our remaining bias with the interpolation error for the BW-Stat data, since we have the possibility to redo the spatial interpolation of the station based product and therefore can compare the potential interpolation error at each station location.

MC4: the method used to identify the peaks over threshold for fitting the GPD using partial series on ll.180-189 is not explained very well.

> Suggestion: Clarify Equation 1 by mentioning how the data from the 5 time series corresponding to the 5 radar cells were combined to estimate the GPD. Give a step-by-step description of the data extraction procedure, using equations and formal mathematical notations.

Thank you very much for the valuable suggestion. We will include this information in the respective flow chart (see response to first comment).

MC5: the quantile matching method you use for the location parameter in section 2.5 is questionable: In particular, I wonder why you decided to only bias-correct the location parameter but not the scale parameter as well, since the two are usually correlated. By only changing the location parameter, you are at risk of under/overestimating the spread of the GPD. An underestimated scale parameter might explain why RAD-BC has lower return values compared to BW-Stat, even after bias correction.

> Suggestion(s): Explain why you think it is justified to only adjust the location parameter during the bias correction. Explore what would happen if you adjust both the location and scale parameters simultaneously. If possible, explore alternatives for how to bias correct the radar-derived GPD parameters without depending too much on the 1-year return period.

Thank you very much for the valuable suggestions. In the manuscript we describe two BC options, both based on the 1-year return period. By using a multiplicative approach, we indirectly adapted also the scale and shape parameter, since we multiply all events with a certain factor leading to a larger absolute precipitation amount added to the largest events.

We believe that an explicit correction of the shape parameter is of course doable, however, given the nature of the station data (most of them are of rather short duration of ~20yrs and have been resampled itself), we think that an explicit correction of the scale factor is not justified.
We are meanwhile working on a more complex BC-correction method based on the individual events, which looks promising but still has some major drawbacks and needs some more testing – and this is probably a study for itself.

MC6: The structure of the paper could be improved. For example, one could argue that section 3.1, which presents the sampling statistics, is not really a result (in the scientific sense) but merely a continuation of section 2.4.

Suggestion: consider moving 3.1 to the methods section and merging it with 2.4 while shortening the text. There's nothing really surprising or valuable in this section that the reader can learn and all the statistics you present highly depend on the way you select the cells in the first place.

Thank you very much for this suggestion that we will follow. Actually section 3.1. was intended to proof that the spatial sampling is doing what it was expected to do, even we leave some degree of freedom. Applying an ensemble sampling approach could theoretically lead to a repeated sampling of the same cells/events, which would bias the result. Nevertheless, we agree that the respective figure is not a key element of the manuscript. Therefore, we will move the respective figure to the attachment section, which will also shorten the manuscript.

Minor comments :

ll.127-128 : ``It further has to be mentioned that in the final product all design rainfall values below/above the 5th/95th percentile have been set constant (to the 5th/95th percentile) in order to prevent for extremely low/high outliers.'' Not clear. Please reformulate. What quantiles are you referring to? What do you mean by extremely low/high outliers?

This sentence actually refers to the BW-Stat dataset, were all values below/above the 5th/95th percentile (spatially) were capped. We can reformulate this, but it was not applied to our radar-based dataset and it was not done by us.

ll. 149-151 : ``The maximum sampling radius was set to 25 km (cells). These numbers are chosen in order to reflect the typical size of a convective cell in Germany (~25 to 40 km for hourly events in the summer season in BaWu, Lengfeld et al., 2019).'' The actual size of a convective cell is much smaller. But because of the hourly aggregation time scale and cell motion, the impacted area is in the order of 25-40 km. Please reformulate the sentence to convey the right meaning.

We will reformulate this.

ll.160-163 : ``In order to prevent that neighbouring cells are sampled (which actually would limit the number of additional rainfall events), the sampling probabilities of the cells in a radius of 4 km (cells) of the cell are reduced below the threshold value after each sample is drawn (Fig. 1, pnel bIV).''

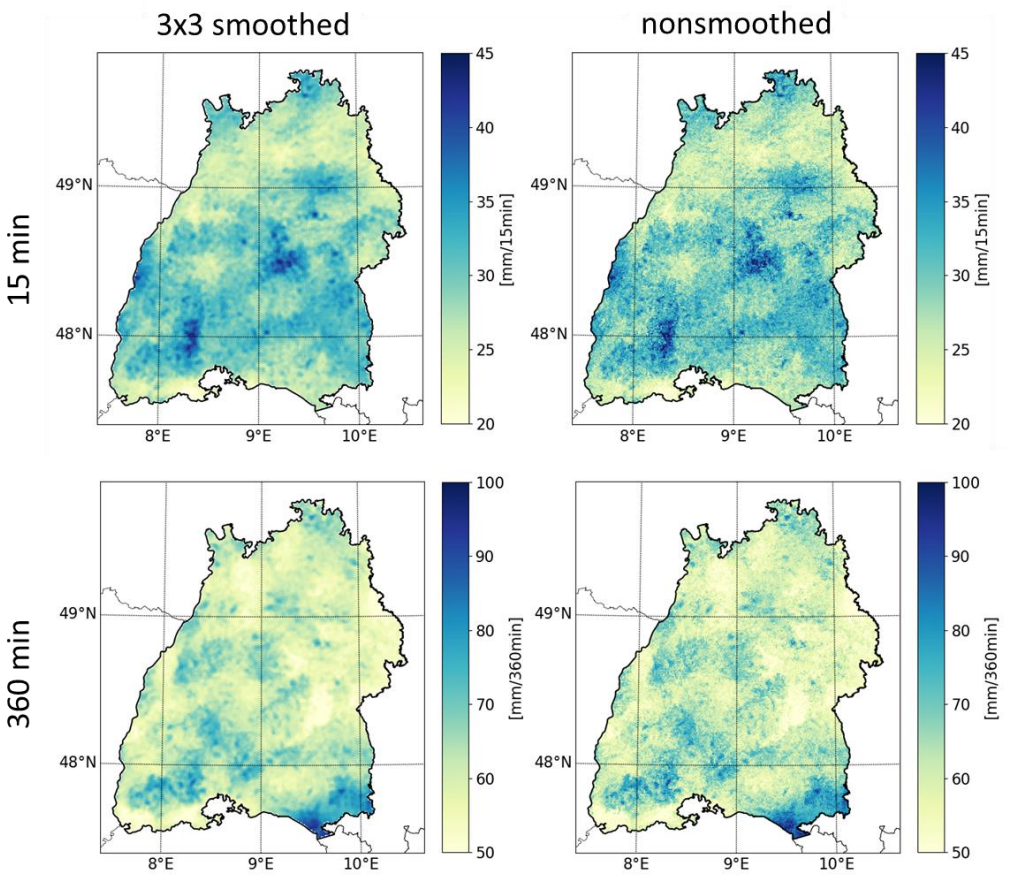
Not clear. Please reformulate the sentence to make sure the reader can understand and reproduce what you have done. Suggestion : first explain what you have done, then mention the rational behind it in a separate sentence/paragraph.

We will reformulate this.

ll.272-273 : ``Note that the RAD-BC dataset represents the ensemble mean of the five individual sample products and that the data is spatially smoothed with a 3 by 3 cell filter to avoid single outliers.''

What happens if you do not smooth? What are these «single outliers» exactly and how large are they?

They are basically rather small (see figure below). We just smoothed it since the KOSTRA and BW-Stat datasets are highly smoothed due to the interpolation.



ll.325-330: "In order to reveal the uncertainty contribution resulting from the ensemble sampling we highlighted regions with a large (> 75% of the range) contribution of the sampling uncertainty. Generally, the contribution of the sampling uncertainty is larger in regions with a lower overall uncertainty range. However, there are various spots with relatively larger uncertainty that are dominated by the sub-sampling uncertainty. The previously mentioned enhanced uncertainty in the northern Black Forest case seems to be substantially influenced by sampling uncertainty in its eastern parts, [...]"

This is a good example of a passage that is difficult to read/understand and for which the writing could be improved.

We will reformulate this.

l. 348-349 : ``The lower values for the scale/shape parameters of RAD-BC can partly be attributed to the fact, that for high rainfall intensities radar data is known to underestimate rainfall amounts due to the ~~reflexivity~~ reflectivity bounds (e.g. Schleiss et al., 2020).''

No idea what you are referring to here. In the paper you cite, I show that radar rainfall

estimates during heavy rain exhibit a strong (increasing) conditional bias with intensity. This bias persists despite the radar operators' best effort to combine/merge radar data and perform frequent bias correction (e.g., hourly mean field bias adjustments). It's likely that the German radar network suffers from the same problems. The fact that the bias increases with intensity means that a simple quantile mapping based on the 1-year return period probably won't be enough to reliably estimate longer return periods using radar. Note that the conditional bias is probably due to multiple factors, including signal attenuation, range effects, and natural variations in the raindrop size distributions with intensity, which affects the Z-R relationship (locally).

We actually wanted to point out the use of a fixed Z-R relationship in the radar data. We will reformulate this.

- ll.372-373 : ``A detailed comparison of the different bias correction approached and their implications on the derived design storms is currently in progress.'' This is a good example of a sentence that could be deleted to shorten the text.

We will remove this sentence

Typos :

Sorry for the typos and thank you very much for pointing them out. We, of course, will correct them