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

The manuscript deals with the important and timely topic of determining design storms with return periods of up to 100 years from rather short time series of precipitation data from radar observations. The authors present a method to statistically extend time series of weather radar rainfall estimates by combining regional frequency analyses with subsequent bias correction. The results show improvement over the sampling approach by Goudenhoofdt et al. (2017) that is used as basis for their method, but uncertainties, e.g. a bias in the radar data for design storms with large return periods, still remain.

The study fits in the scope of HESS and is of interest to the research community. I have already reviewed an earlier version of the manuscript and the authors have taken some of my suggestions into account. However, some major concerns remain and new questions came up, that I listed below. I recommend major revisions of the manuscript before publishing it. I'd be happy to discuss my suggestions with the authors in the open discussion and clear up possible misunderstandings.

Major comments:

1. A major concern is the minimum distance of the radar cells that are considered to statistically extend the time series of the cell of interest. As far as I understand the cells have to be at least 4 km apart. The authors mention that the typical size of a convective cell in Germany is 40 km for hourly events according to Lengfeld et al. 2019 (p.5, l.150 in this manuscript). Therefore, the minimum distance of 4 km seems a bit too small to me, especially when considering also daily events that have a much larger typical spatial extent. The authors mention that more or less the same amount of events have been taken from all pixels, but did the authors perform any kind of independence check for the time series from the cells that are combined to a long time series, e.g. the correlation of the time series?

   It is true that we sample rather close to the COI, in order to mainly sample cells that have similar rainfall characteristics. As shown in Figure 2b, the majority of sampled cells are in a distance range between 8 and 12 km. The events of the sampled cells will definitely have a certain amount of correlation (actually it is intended that they have) to the events in the COI - especially for the longer durations. However, since we have as prerequisite that single events (independent of the cell they are sampled from) have to be at least two days apart, we assume that we can ignore the autocorrelation effect in the EVA, since the duration of an event is much shorter.

2. The aim of the study is to determine design storms with a return period of 100 years. Therefore, a method to extend rather short time series (19 years) from radar data by using additional data

from similar regions is presented and compared to a station-based interpolate product. It makes sense to have a time series of more or less the length of the return period for the radar data. Therefore, a length of 95 years has been chosen which equals a combination of 5 pixels. But there is no information on the length of the station-based products that are used as references here. To my knowledge KOSTRA is contains 60 years of data. How reliable are the estimations of design storms with a return period of 100 years from KOSTRA? How many years of data are included in BW-stat? I was wondering how fair the comparison is when using data sets with different lengths. It would be beneficial to the manuscript if the authors add some information about this.

Most of the stations included in BW-Stat have actually a similar length than the radar data series and only a few are more than 40 yrs. Due to the short data series available, a spatial pooling approach was applied to construct the BW-Stat dataset, as well. We will include the information on the length of the data series in the manuscript.

3. Although the authors extended Section 2.2 about the reference data sets some information are still missing (e.g. how many stations are considered, length of the time series, interpolation methods, etc.) A more detailed description of methods used in BW-stats and KOSTRA as well as the differences in the statistical approaches to determine design storms from those data sets is also desirable. The method for BW-Stat is briefly described in section 2.4. Maybe it would be better to have a general section about the methods first and then describe the data sets and their differences. Important information about the methods are missing that are crucial to understand the results and differences between the datasets.

   We will add more information on the method behind the two station based datasets. However, we also have been asked by Reviewer 1 to substantially shorten the manuscript, so we will keep it short and refer to the respective reports. Considering the restructuring of data and method section we believe that we need to first describe the radar dataset before we can discuss the methods. Hence, we would prefer to leave the order of data and method description in its current state.

4. A more detailed description of the sampling process, the generation of the ensemble members, the bootstrapping method and the bias correction is needed to allow for better understanding of the results and of the choices made by the authors.

   Reviewer 1 suggested to include some flow charts to make the sampling processes more clear. This suggestion we will follow and we will also include the respective equations in the charts.

5. Some findings are mentioned in the result section, but not sufficiently discussed in the discussion section. E.g. Why the spatial pattern in BW-stat is following the behaviour of RAD-BC for a return period of 1 year (p.10, l.295). Extending the discussion section and analysing the results in more depth is necessary to enhance the quality of the manuscript.

Thanks a lot for pointing this out. We will work more on the discussion section and also add some more analyses regarding the uncertainty of our radar-based data set when compared to the interpolation error of the BW-Stat dataset.

Minor comments:

p.1, l.6-8: This sentence sounds odd to me. Please rephrase.

We will try to make it more clear.

p.1, l.17: A bracket seems to be missing here.

The reviewer is right. We will change it accordingly.

p.2, l.49: times series --> time series

The reviewer is right. We will change it accordingly.

p.2, l.62: …might not sufficient… --> …might not be sufficient…

The reviewer is right. We will change it accordingly.

p.4, l.107: What exactly are the methodological differences the authors mention here?

They are basically (i) the use of a different extreme value distribution (2 parameter GEV vs 3 parameter GPD) as well as (ii) that for some durations, the design storms in KOSTRA are interpolated from design storms of neighboring durations, while we calculate them explicitly for each of the durations. We will make this more clear

p.4, l.126: A word seems to be missing in this sentence.

The reviewer is right. The word ‘station’ is missing here

p.5, l.127: Is the limitation to values between the 5^{th} and 95^{th} percentile really necessary? How large the outliers? Please justify this decision.

This decision was not made by us, but the developers of the BW-Stat dataset. We will make this more clear.

p.12, l.371: …the also the… --> ….also the…

The reviewer is right. We will change it accordingly.

p.12, l.372: approached --> approaches

The reviewer is right. We will change it accordingly.