Comment on hess-2021-366
Anonymous Referee #2

Dear Anonymous Referee,
thank you very much for your valuable critics, remarks and suggestions in order to improve our manuscript. Please find our response to the various points you raised indicating what we adapted in the manuscript below.
Additionally, I would like to apologize for not being active in the open discussion during the review phase. The reason for this is that I was on parental leave, which actually started a bit earlier than was originally foreseen.

Best regards, also on behalf of my co-author,
Andreas Hänsler

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.4, l.121 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. 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 or the percentage of time steps with rainfall in the cell of interest but no rainfall in the other cells of the sample?

→ This 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, as the duration of an event is much shorter and hence the sample is independent of an event.

How do you make sure that the 258 events are actually taken from all 5 time series and not only taken from the 19 year time series of the COI?

→ Originally we did not investigate how often an event is pooled out of a specific cell, but we now looked at this. Indeed, we find, that each cell contributes almost the same amounts of events to the EVA (Median is 20%, 5th percentile is 13%, 95th percentile is 28%). We added this finding under section 3.1 in the manuscript.

I was also wondering if the same set of cells are used throughout the study or if the samples vary for the three durations that are considered.

→ The reason for using the same mask for all durations is that we wanted to be able to refer differences in the spatial patterns between short and long durations to the data itself. If we would have a change in the sample mask the change in spatial patterns could be mainly due to this effect.

2. The authors only consider precipitation data from April to October, because this is the main season of convective events of short durations. The statistical approach to determine designs storms

is based on a partial time series consisting of e (Euler's number) times the number of years. I was wondering, if this approach is still valid if only 7 out of 12 months of the year are considered.

→Yes, to our knowledge it is still valid. There are many studies available that use the Peak-over-Threshold (POT) approach for seasonal extreme value analysis (e.g. seasonal flood frequency analysis).

Although it is common knowledge that most of the convective storms occur during summer, some events might still be missed, especially for the design storms with 24 h durations that might also be associated with advective weather situations.

→ Yes, we agree that for the 24h case we will miss events that occurred outside the April to October season. We mainly included the 24h case since we wanted to demonstrate that there is actually a change in the spatial pattern between short duration and longer duration design storms when basing the analysis on spatially explicit data. This change in pattern is expected but not (or only to a lesser extend) visible in the interpolated station based reference datasets.

To my understanding, the reference data sets KOSTRA and BW-Stat consider all months and might not be comparable to the radar based data set. I would suggest to take all months into account or the authors should provide some kind of validation for their choice of selecting only summer months.

→ Actually the analysis design of the RADKLIM data is chosen in order to represent what was done when establishing the BW-Stat data. But this was unfortunately poorly described in the original version of the manuscript and is now changed (e.g. see Section 2.2. on dataset description but also added this information in the description of the EVA - see Section 2.4). So also in BW-Stat only April to October data is used and the method was applied for 5 minutes to 24h.

3. Section 2.2 about the reference data sets is quite short. More information about both data sets (e.g. how many stations are considered, length of the time series, interpolation methods, etc.) and on the differences in the statistical approaches to determine design storms from those data sets are 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. E.g. that a two parameter GEV distribution is used in KOSTRA, instead of GPD for BW-Stats and the radar-based data set, is only mentioned in the discussion. This is important information that should be given in the method section.

→ We completely agree with the reviewer and completely changed the section on the reference datasets.

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 (e.g. why 5 ensemble members?).

→ The second reviewer also pointed towards some shortcomings in this part of the manuscript. So we added some more information in order to clarify his remarks, and hope that this now becomes clearer.

Minor comments: -
p.3, l.75: "... which leads to a spatially..." → "...which leads to spatially..."

→ Reviewer is right – we changed it in the manuscript

p.3, l.92: To my knowledge, the KOSTRA-DWD-2010R data set has a resolution of about 8.2 x 8.2 km. Did the authors perform some kind of remapping to achieve the 5 km x 5 km resolution?

→ Reviewer is right – we changed it in the manuscript

p.5, l151-152: Almost the same sentence is repeated on p.6, l.161-162.

→ Not exactly sure which sentence is repeated. On page 5 we talk about estimating the threshold value for the partial series, on page 6 we talk about parameter fitting for the GPD.

p.6, l.174-176: The radar data are adjusted to the 1 year design storms of the station-based BW-Stat data set. In the results section both data sets are also compared to design storms with 100 year return period derived from KOSTRA. For a better assessment of the differences between 100 year design storms from KOSTRA and the other two data sets it would be beneficial to also compare the 1 year design storms. Do they show the same features in the spatial pattern? How large are the differences?
→ We added a figure showing the spatial pattern of a 1yr event in the appendix and briefly discussed it in the text. Actually, the spatial patterns in the 1yr design storm of BW-Stat are similar to the ones of RAD-BC.

p7., l.205: "... located more the centre..." → "...located more to the centre..."
→ Reviewer is right – we changed it in the manuscript

p.8, l.222: "...time steps can attributed..." → "...time steps can be attributed..."
→ Reviewer is right – we changed it in the manuscript
What is meant by "lower spatial distribution"? Lower spatial resolution?
→ Yes, we wanted to talk about the spatial resolution. We changed it in the manuscript

p.8, l.230: "...as well as the for the bias-corrected..." → "...as well as for the bias-corrected..."
→ Reviewer is right – we changed it in the manuscript

p.8, l.238: Isn't the 24 h design storm from KOSTRA shown in Fig.4? Or is that something else the authors refer to here?
→ Yes, it is shown. But the 360 minute storm is actually interpolated between 1 and 12h and the latter is not shown. We changed it accordingly and actually also pointed to the fact, that in the latest version also the 24h design storm is interpolated (between 12h and 72h).

p.8, l.249: REGNIE is first mentioned here and should be explained.
→ REGNIE is now introduced earlier in section 2.2

p.9, l.257: Which figure do the author refer to here regarding the 20 year design storms?
→ Figure 5, like the rest of the paragraph

p.9, l.269: 10 th → 10th
→ Reviewer is right – we changed it in the manuscript

p.9, l.276: "... in in the case..." → "...in the case..."
→ Reviewer is right – we changed it in the manuscript

p.9, l.283: "...relatively larger uncertainty..." → "...relatively large uncertainty..." or "...larger uncertainty..."
→ Reviewer is right – we changed it in the manuscript to „...relatively large uncertainty..."

p.10, l.299-303: This might fit better in the result section.
→ We actually would like to leave it in the discussion section since it fits to the overall discussion of the remaining biases

p.10, l.310: "...can be seen as a rather robust..." → "...can be seen as rather robust..."

→ Reviewer is right – we changed it in the manuscript

p.11, l.318: "...difference...become..." → "...difference...becomes..."
→ we wanted to talk about differences – not a difference – so we changed it to „differences … become ...“

p.11, l.319: "...between in rainfall estimates..." → "...between rainfall estimates..
→ Reviewer is right – we changed it in the manuscript

p.11, l.331: "...in future" → "...in the future"
→ Reviewer is right – we changed it in the manuscript

p.14, l.417-418: "...based on distance to cell of interest..." → "...based on the distance to the cell of interest..."
→ Reviewer is right – we changed it in the manuscript

p.14., l.420: "...is marked with in red..." → "...is marked in red..."
→ Reviewer is right – we changed it in the manuscript
April to October of which years?
→ The most recent 30 year period from 1991 to 2020 – we added this information to the figure caption

p.15, l.426: It should either be "for a single member" or "for single members"
→ Reviewer is right – we changed it in the manuscript to for a single member.

p.15, l.430: Remove one of the brackets.
→ Done

p.16, Figure 4: It would also be interesting to see the differences between KOSTRA and RAD-BC.
→ we decided not to plot the 1 to 1 differences as the methodology and the spatial resolution is very different to BW-Stat and RAD-BC. We included it rather as an independent reference. We hope that this becomes clearer now, since we substantially changed the description of the datasets.

p.17, l.440: What is meant by "four different event"? I assume it is supposed to be "for different event durations"?
→ Reviewer is right – we changed it in the manuscript

p.17, l.443: Why is the 96th percentile chosen here instead of the 95th percentile?
→ it is the 95th percentile - we changed it.

P18, l.450: There is no comma needed after "regions".
→ Reviewer is right – we changed it in the manuscript