Dear reviewer,
Thank you very much for your thorough review. I have tried to reply your comments as clearly and derailed as possible. I am happy to discuss further with you any ambiguities that might still be there.
Kind regards,
Golbarg

• In the methodology, the authors fitted the IDF of Koustoyiaannis et. al. (1998) for each area separately. However, there are IDAF formulations that links all the data for different durations and different area together and fit one IDAF formulation. An advantage of this is that a constrain is already implemented in the formulation to ensure that the intensities decrease with area. Example is the IDAF formulation of De Michelle. Did the authors consider this option?

GG: This is a good point. We are aware of the analytical formulations of the IDAF/AIDF curves which include not only duration but area as well and our final goal is also to provide a formulation like this. However, the inconsistencies in IDAF/AIDF curves (crossings) are so heavy that we think it is better first to reduce the crossings based on their real cause as much as possible. For that we are looking for an optimal sampling approach. In a second step we would like to find an analytical formulation which includes the area to avoids the remaining order relation problems. This is however out of the scope of this paper.

• In the multiple location sampling, a random sampling is done. I expect that each time the sampling is repeated, a different set of locations will be selected. Would this affect the result? Have you considered using a moving window in space to capture all possibilities?

GG: Well yes, when repeating the study, the random samples of points and areas will be different. However, our samples are taken so large that it ensures all the events within the study region around each location are covered. While removing the dependencies among pooled events there are always a considerable number of events being removed. This is an indicator that the samples are covering all possibilities.

• I find it surprising that the spatial crossing does not show any pattern as a function of the topography. For instance, Melese et al. (2019) and Haruna et al. (2024) observed this behavior to depend on orography, for instance the location of the study pixel on the windward or leeward side. Could you comment on this? Would it be possible to apply the different sampling strategies to a pixel on the leeside or wind side of the mountain and to see the effect on the curves?

Melese, V., Blanchet, J., and Creutin, J.-D.: A Regional Scale–Invariant Extreme Value Model of Rainfall Intensity–Duration–Area–Frequency Relationships, Water Resources Research, 55, 5539–5558, https://doi.org/10.1029/2018WR024368, 2019.
Haruna, Abubakar, Juliette Blanchet, and Anne-Catherine Favre. "Estimation of Intensity-Duration-Area-Frequency Relationships Based on the Full Range of Non-Zero Precipitation from Radar-Reanalysis Data." *Water Resources Research* 60.2 (2024): e2023WR035902.
Rosin, T., Marra, F., and Morin, E.: Exploring patterns in precipitation intensity–duration–area–frequency relationships using weather radar data, Hydrol. Earth Syst. Sci., 28, 3549–3566, https://doi.org/10.5194/hess-28-3549-2024, 2024.

GG: Yes, this was a surprise for us as well. We are aware of the studies you mentioned and they are also cited in the manuscript. However, it is important to consider that our study area is pretty flat and there is just a small portion of it which has an overlap with the Harz mountain and due to the removal of the data at the edges of the study area to avoid including missing values and pixels

which are heavily affected by the radar errors it is not much information at hand at such locations. It is definitely interesting to investigate this further with data from other regions and climates.

• I was expecting "Summer" to have more locations with crossing, compared to winter. Since summer events are convective and tends to be isolated, while winter are stratiform and tends to cover a larger area. Could you comment? Furthermore (in Line 375), I expected frontal systems to exhibit less spatial variability compared to convective events. Since frontal storms are driven by large-scale interactions between air masses rather than localized convective processes, the intensity of precipitation and weather conditions tends to be more uniform compared to convective storms. Could you comment?

GG: The crossings appear mainly for longer durations, 4hr in winter and 12-18hr in summer. This means they are not coming from small scale thunderstorms but from events with longer durations. To produce crossing these events need to be quite heterogenous in space. So, one explanation could be that these are events where frontal systems are overlaid with convective parts. These events can occur in winter and in summer and it can be assumed that the frequency of those mixed events is increasing with global warming. There was also a misinterpretation form our side based on the cited study Kim et al 2019. We rewrote that paragraph.

• The authors compare the quantiles from the various sampling strategies to those from KOSTRA. Due to the inherent differences between the two, I don't understand how KOSTRA values could serve as benchmark for preferring one method over the other. Should the best method agree with KOSTRA values? Why?

GG: Of course, only the point-related IDF curves (a = 1 km^2) can be compared between KOSTRA and the radar product. This is what is done here. The radar product is a merged data set considering station data as truth and so is supposed to resample the station statistic at a point quite well (see Fig. 3). For both products KOSTRA and the merged radar data the extreme value analyses is done using the same approach. In KOSTRA an additional regionalization step is included which makes things more uncertain than for a direct point estimation. On the other hand, KOSTRA is based on 60 years data and the radar analyses only on 20 years, which is probably the largest difference between the two. Altogether we think the point radar IDF curves should approach the KOSTRA IDF curves from below due to the difference in observation length.

• Is there any motivation for the choice of the different areas and durations, and more precisely the upper bounds?

GG: nothing specific. R = 18 km gives an area of roughly 1000 km^2 and area sizes ranging up to 1000 km^2 are representative of the catchments in the studies area. We have to assume stationarity within the area, which is as less guaranteed as larger the area becomes.

• Do you observe the same "crossing" based on simple exploratory analysis of the annual maxima series (without fitting GEV).

GG: yes, we do, even more complex crossings.

• Line 93: I don't understand the sentence "The region has been observed for the time period of 2000 to 2019…" Could you rephrase.

GG: done

• Line 101. Do you mean "final merged radar data"?

GG: yes, corrected

• Line 155: Eq 7 instead of 6

GG: actually, it refers to both of them, corrected

• Line 194: "…actual intensities". Do you mean "intensities with durations d"?

GG: I don't get your question exactly but I tried to make the sentence a bit clearer.

• Line 207. The largest area (R=36 km), any justification for this choice? Would it affect the result?

GG: the largest radius we worked with in the base method SLS was 18 km because it gives an area of roughly 1000 km^2. And doubling that to 36 km was to ensure we can take enough samples from the surrounding area. No specific reason for these numbers in first place. The analysis was done for whole Germany with the station dataset used in section 4.1.2. and we saw the same pattern for an area as big as Germany and station data.

• Line 415: "...smaller values than KOSTRA all durations." - > "... for all durations"

GG: corrected

• Figures

- In Figure 1, it is difficult to contextualize the location of the study area with respect to the map of Germany, would it be possible to add a locator map?
- What are the grey colored points that are randomly located in Figure 4b. They seem to be independent of the circles.
- Figure 4: Add that the circles are colored according to the radius(area)
- Caption of Figure 4, the phrase "In both schemes the outer most circles are to the center of the study location." Seems not complete.

GG: Fig 4 adapted

- Figure 5 and 9. The choice of the color palette (seems to be discrete/qualitative) makes it difficult to track the changes of the quantiles as functions of the area. Since the area are increasing, a "sequential" or "diverging" palette would be better (Eg. Melese et al. (2019)).

GG: your concern is valid. The initial color palette was rainbow which showed the areal dimension pretty well, but that palette is not colorblind friendly and that is one of the requirements of this journal. We tried so many different palettes, sequential etc. this was at the end the best compromise.

- Figure 6 B is mentioned earlier (Line 284) than 6 A (Line 300). Consider switching the two?? (Moving 6 B to the right)

GG: corrected

• Equations

- Check equation 11 carefully, a power is missing

GG: corrected

- the transition from Eq 8 to 9 is not very clear. For example, the index *a* is suddenly dropped. I suggest rewriting it for clarity.

GG: corrected. Not the equation, but how the text leads up to it. Hopefully it is easier to understand now?

- In Eq 14, since dependent events are removed, then $n_{amMLS} \leq n_t \times n_{sp}$. Am I correct?

GG: correct, thanks