Response to Referee #1,

Dear Theano Iliopoulou,

Thank you for taking the time to evaluate our work, and as well for your comments that will for sure improve our manuscript. Please find below our answers to the issues you have raised.

- **Comment1**: On uncertainty of DDF parameters (referred to as 'local' uncertainty)
By the investigation carried out in the first paper, the authors conclude on using a common, fixed value of the shape parameter, based on the literature for return periods up to 100 years (in particular, following Koutsoyiannis' work (2004)). Using a common value of the shape parameter is a reasonable choice given the high uncertainty involved in its estimation from single stations. However, fixing the value of the shape parameter based on the literature, instead of estimating it, is not devoid of uncertainty either. It may not be straightforward to assess the uncertainty of a fixed parameter, but on the other hand, neglecting the shape parameter from the uncertainty assessment is a limitation that I think should be discussed. Perhaps the authors could add a brief discussion of the uncertainty associated with the shape parameter, and comment on its expected impact on local uncertainty, which presumably would be exacerbated if this was also accounted for.

Response: We understand your concern regarding the fixed shape parameter to the value of 0.1. Please note that the value 0.1 was not purely theoretical (following the Koutsoyiannis work in 2004). We calculated the shape parameter for all the long stations (with long observations) and the shape parameter had a median close to 0.1. Also, Ulrich et al. 2021 computed the GEV shape parameter of annual extremes in Germany based the Koutsoyiannis mathematical framework as 0.11. Thus, the decision to fix the shape parameter at 0.1 was not trivial and not just based on the Koutsoyiannis work, but of previous analysis that we have conducted. This is something that we have tried to make clearer in the updated version of the first paper (Shehu et al. 2022). Nevertheless, thank you for pointing that out, and we will add a better clarification for this in this manuscript.

Also, we will add a description about the uncertainty of a free shape parameter (estimated independently for each station) to the local uncertainty. Please keep in mind that in the previous work of Shehu et al. 2022, we have investigated which method is more precise for local estimation of the rainfall extremes: keeping the shape parameter free or fixed. In our previous work, it was visible that the $nCI95_{width}$ was much higher for the free shape parameter than for the fixed one. Below you will see the normalized 95% Confidence Interval width ($nCI95_{width}$) obtained by local bootstrapping of annual maximum series where the shape parameter was kept constant as calculated for each station (in the Figure 1 donated as KO.FREE) and fixed to 0.1 for all stations (as implemented in this manuscript, denoted with KO.FIX). These confidence intervals widths were computed for each of the 133 long observations, for each duration and for each return period shown in the legend. It is clear that for return period higher than 20 years, the uncertainty from a free shape parameter is much higher than the uncertainty from keeping the shape parameter fixed at 0.1, which will cause the interpolation of extreme rainfall to be less certain (and with higher uncertainty ranges). Please note that the following Figure 1 will be not added at the manuscript, but serves only for this review in order to explain the difference between the two.
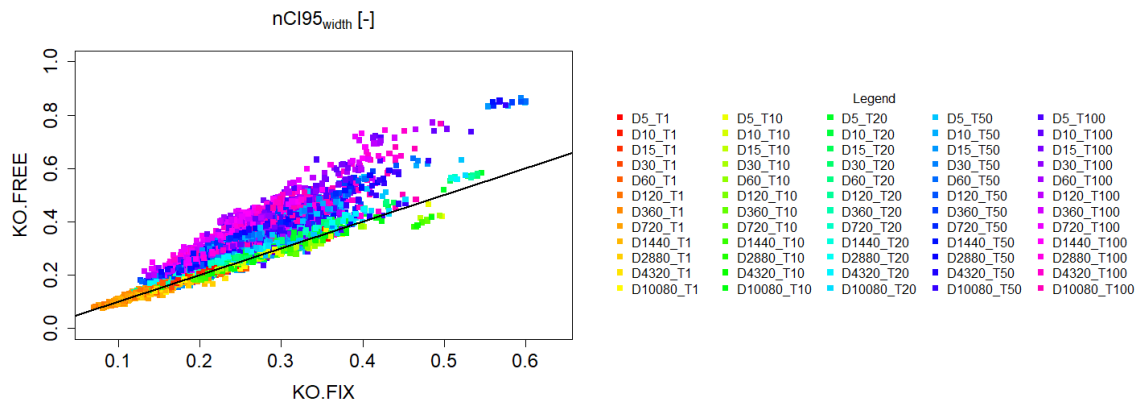
Figure 1 A direct comparison of the normalized confidence interval width for two cases of the local sample uncertainty: case 1 – KO.FIX where the GEV-shape parameter is fixed as 0.1 (x-axis) and case 2 KO-FREE where the GEV shape parameter is computed independently for each station (y-axis). The comparison is done for each station, duration and return period (shown in different colors in the legend).

- **Comment 2:** On regionalization uncertainty

  The quantification of the spatial uncertainty presents the greatest challenge and also, the major contribution of the work. The authors could elaborate more on the reasons why they employ spatial simulations for assessing regionalization uncertainty, since this is the only component of the uncertainty analysis for which simulation is employed. For instance, Lines 87-90 could be better explained; the reasons that the authors do not employ kriging for spatial uncertainty are less clear to someone not familiar with the provided geostatistical literature. In particular, given that they consider kriging variance to be a measure of uncertainty for the unobserved locations (Line 88), why is it later claimed to be only efficient for the local (I understand this as 'at-station') uncertainty and not the spatial one? The question also arises since the Sequential Gaussian Simulation that the authors finally choose for the assessment of spatial uncertainty is based on the kriging mean and variance. Do the authors use spatial simulations mainly because they seek to produce richer spatial patterns than the ones obtained by a kriging interpolation? A brief explanation of the terms 'local' and 'spatial uncertainty' would also be helpful to the reader at this point (a definition is provided later on, in Section 3).

  Response: Following are the main reasons that justify our choice in employing simulations instead of using the kriging variance. We will update the text and make sure the reasons are clearly stated.

  1. As you have already mention, the kriging interpolation is causing smoother spatial patterns, and with the simulations we are trying to produce rich spatial patterns (Deutsch & Journel, 1998).

  2. The kriging variance depends only on the data configurations (how many stations there are and where they are in relation to each other) but they are independent of the observed values. With the simulations we try to compute uncertainty also in terms of the observed values (in addition to data configuration) (Deutsch & Journel, 1998).

  3. Spatial simulations provide a measure of uncertainty about the un- sampled values taken altogether in space rather than one by one. They can provide information regarding joint-spatial uncertainty, unlike kriging variance which provides information at one location at a time conditioned to other observations in the vicinity. Thus, under the context of the kriging systems, local refers to the uncertainty prevailing at one location, while spatial refers to the joint uncertainty prevailing at multiple locations (Deutsch & Journel, 1998). We will clarify these two terms in the lines 87-90.

4. Lastly and more importantly, to estimate confidence intervals as employed here, we should assume that the kriging prediction errors are normally distributed in order to calculate the confidence intervals from the kriging variance. But please keep in mind that we are interpolating parameters that have a specific relationship with each other. If we want to compute the 95% confidence interval for the depth-duration-frequency curves, it is not possible to combine the quantile values of each parameter, as they may not correspond to each other. That is why for this case simulation is much better than the kriging interpolation.

My understanding is that in the parts of Experiments 3-5 that involve spatial simulations, the authors produce prediction intervals, which are subtly different than the confidence intervals obtained by the bootstrapping procedure followed for the assessment of local and variogram uncertainty. Could the authors comment on this difference and clarify, if appropriate?

Response: Regarding experiments 3-5, I'm not quite sure if I understand your concern correctly. I suppose you mean why the confidence intervals obtained by the Experiments 4 and 5, have much higher values and exhibit a completely different shape across the durations, than the Experiment 1, 2 and 3 (related to Figure 10 – upper row). We have tried to explain this behavior at the lines 507-532. Additionally, please refer to our reply on the comment 3.

Line 306-307: Could the authors explain the rationale for performing the conditional simulation, i.e., maintaining in the simulation a set of observed values for the long stations, and also, explain how this set is chosen?

Response: Yes, we have performed a conditional simulation, which means that for known locations (where we have observations), we are not simulating as we have already observed those pixels. By simulated annealing the pixels are conditioned to the exact observations, while in sequential gaussian simulations, the pixels are allowed to vary from the exact observations as per the nugget coefficient of the variogram. We will add an explanation about this in the manuscript. Regarding the reason why we chosen the long station as a set in explain in the Shehu et al. 2022 manuscript, but we will add an explanation in this manuscript as well. Long stations are more reliable than short stations when estimating extreme rainfall, that is why there are recognized as the primary set of data: they are used as the ground truth (for the cross-validation), and as well the main input for the regionalization (short series are used as an external drift to the long series interpolation). In this study we have only 133 stations, because those were the stations available at the time being from German Weather Service with 1min temporal resolution for observation length more than 40 years.

- **Comment 3:** Comparison among different components of uncertainty
  The authors find that 'the spatial simulations add to the regionalization the biggest uncertainty', and based on that conclude that 'the spatial uncertainty is the main source of uncertainty when regionalizing the DDF curves'. Although this finding is reasonable and supported by the results presented, I also deem the estimation method important in the context of the inter-comparison of the different uncertainty intervals. Perhaps, this might not be the determining factor, but still could the observed difference between the spatial uncertainty and the other sources be, in part, due to the fact that (spatial) simulation produces wider intervals than bootstrapping?

Response: That is an interesting point and yes, I agree with you, the spatial simulations are in fact producing wider intervals than the local bootstrapping. This can be also seen by comparing the variance of the local bootstrapping with the variance in space of the observations. So, the parameters are varying more in space, and that is why when sampling from space (spatial simulations) the prediction intervals are higher than for the bootstrapping case. We will add a small discuss about this in the manuscript.

- **Comment 4**: Conclusion

From the large experience gained through this work, the authors could perhaps comment on whether/how the observed uncertainty could be constrained by targeted data collection. For instance, what do they think should be prioritized in terms of data collection, and could there be an added benefit from incorporating gridded precipitation/satellite products in their framework?

Response: Unfortunately, we have not yet investigated how to target the data collection in order to reduce the uncertainty. This is definitely something that we would like to try in the future. It can be that using a gridded product (like radar or satellite products) may be useful in reducing the uncertainty in the spatial simulations. Nevertheless, this might be problematic as the radar or satellite products might filter inaccurate readings when searched for extremes.

- **Comment 5**: Technical/minor comments

The phrase 'non-representativeness of point-measures' appears quite vague; I suggest adding a short explanation of what is meant by it.

Lines 58-60 has been updated to: "In this paper, the focus is on developing a framework that accounts for uncertainties due to short observation lengths and non-representativeness of point measurements for spatial dependencies of extremes."

Line 188: Typo inside the parenthesis, μ should be replaced with η.

Thank you for noticing it. The symbol was changed correctly.

Line 196: Do the authors mean a 5 km x 5 km grid here?

Yes, that is exactly what we mean. We have changed it to 5x5km.

Line 201: I suggest '...have an inadequate length for' instead of '…too little observation years for'

We accept your suggestion and we have changed it accordingly.

Line 208: Probably 'deviate' is meant instead of 'denote' here.

Yes, this is true and we have changed it accordingly.

Line 222: 'method' instead of 'moment'.

Thank you for noticing it. We have change it accordingly.

Line 297: Could the authors justify the choice of 133 stations here?

Only 133 values from all the stations were sampled here, to address the uncertainty in computing the variogram from a small dataset that corresponds with the number of the long stations that were used to compute the variogram for the KED interpolation. We will add a short explanation in the text regarding this.

Line 365: 'fixing' instead of 'fixating'

Thank you for noticing it. We have change it accordingly.

Line 508-509: Could the authors clarify what is implied here for the behavior of the Koutsoyiannis' parameters? Does the explanation provided in Lines 529-530 apply here as well?

In lines 508-509, "this behavior of the Koutsoyiannis parameters" is referring to why for experiments 1 to 3, the confidence interval width across the durations exhibits a parabolic shape. And yes, the lines 529-530 apply also here.

Figure 12: Please check subplots' numbering and consider replacing 'volume' with 'depth'.

The y-axis labels for the Figure 12 have changed from Volume to Depth. Please see below the updated Figure 12.
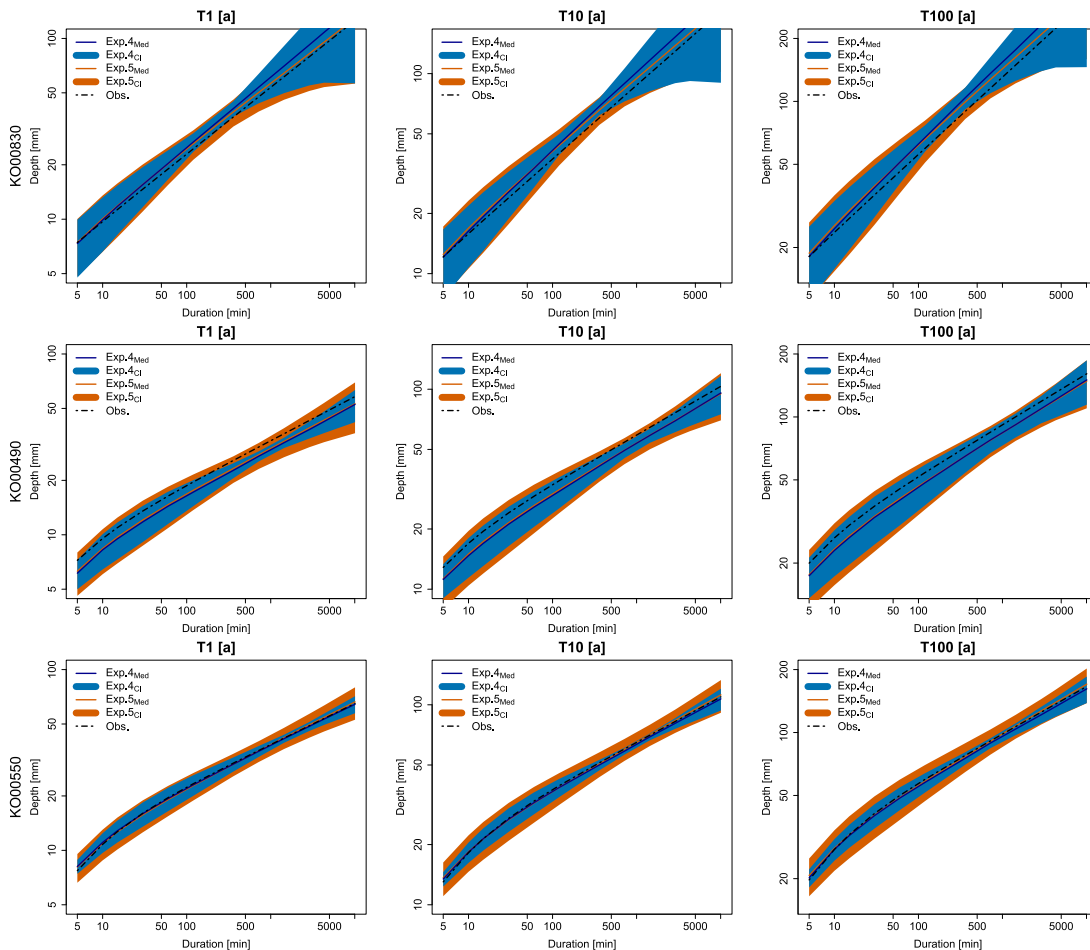

Figure 12 Examples of DDF estimates from observed data and predicted by simulations of Exp. 4 and 5 in a cross-validation mode: as median over all simulations and as 95% tolerance ranges from all simulations: upper row for return period T=1years, middle row for T=10years and lower row for return period T=100years. Three stations are shown here: K000830 located in the German-Alps, KO00490 location in Lower Saxony, and KO00550 located in Black Forest.

Figure 13: 'with' instead of 'will'

Thank you for noticing it. We have change it accordingly.

Please check numbering of the sections; there are two sections '2.1'.

Thank you for noticing it. We have change it accordingly and checked the rest of the numbering.

With kind regards,

Bora Shehu

References:

Deutsch, C. V. and Journel, A. G.: GSLIB: geostatistical software library and user's guide. Second edition., 1998.

Koutsoyiannis, D.: Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation, Hydrol. Sci. J., 49(4), 575–590, doi:10.1623/hysj.49.4.575.54430, 2004a.

Koutsoyiannis, D.: Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records, Hydrol. Sci. J., 49(4), 591–610, doi:10.1623/hysj.49.4.591.54424, 2004b.

Shehu, B., Willems, W., Stockel, H., Thiele, L.-B. and Haberlandt, U.: Regionalisation of Rainfall Depth-Duration-Frequency curves in Germany, Hydrol. Earth Syst. Sci., [preprint], 2022

Ulrich, J., Fauer, F. S. and Rust, H. W.: Modeling seasonal variations of extreme rainfall on different timescales in 774 Germany, Hydrol. Earth Syst. Sci., 25(12), doi:10.5194/hess-25-6133-2021, 2021.