Response to Reviewer #2

Thank you for your comments that has helped us to improve the manuscript. We hope that the following changes that we have made, make the manuscript now easier to read and show the novelty we are bringing more clearly. The responses to the comments that you have listed are given below.

Major comments:

1. Eq. 3 vs Eq. 5: in Eq.3 the IPF of the events is obtained by dividing the MDF by the 'linear model' while the MDF statistics is multiplied by the linear model in Eq.5. is it a typo (in line 103 the authors say the two models are analogous) or there is a reason behind this difference in the structure of the two corrections? Why do the authors use this correction type? Is the linear regression an appropriate model?

   This is actually a typo, both in Eq. 3 and 5 we are dividing the MDF statistics with the linear model.

   Regarding the other question, if the linear model is appropriate or not, then we will discuss the following. Typically, also in the literature, the ratio between the IPF and MDF statistics (also called the peak ratio) is modelled by a linear model based on different catchment characteristics. This has proven to be successfully for many applications. On the other hand, we have applied these models, because as in in Figure 3-4 there appears to be a high correlation between the error IPF-MDF and the logarithm of the catchment area (or even gauge elevation).

2. Terminology: throughout the manuscript the authors refer to the proposed correction method as "linear models" or "linear regression models". This is somehow confusing. I suggest finding another name for the correction method.

   Thank you for your remark. We have decided to use the term "p/V approach /method". We have made the necessary changes in the updated version of the manuscript.

Specific comments:

- Line 145: I disagree that with the bootstrapping/resampling we measure the 'uncertainty due to distribution fitting'. In my opinion it is the sampling uncertainty / parameter uncertainty. Same in section 4.5.

  Thank you for your comment. We have updated it as "sample and parameter uncertainty" in the new version of the manuscript.

- Lines 233-236: the authors use both '(a)synchronous occurrence' and 'temporal overlap'. Do they refer to the same thing or not? If yes, please use consistent terminology. How is the temporal overlap measured/identified?

  Yes, they refer to the same thing. Temporal overlap is measured in days, if the maxima of the two series are on the same day, or on a difference of some days. We have clarified this in the new version of the manuscript.

- Lines 237-238: "…may belong to significantly different events and thus to different populations". This is not clear in this context and clarification.

  As they are caused by different processes, their extreme events come from different samples and hence they can be described by different probability distributions. This is described for instance in Fischer et al. (2016). We have added a small explanation in the updated version of the manuscript.

- Line 256: I suggest adding "percentage" in front of "change".

  We will change it as you proposed in the updated version of the manuscript.

- In the results section there are often reported considerations that would better fit into the discussion section (e.g. lines 274-280).

  Yes, it is true, nevertheless the same topic has already been discussed in the "ranges of application and limitation" section. Therefore, to void redundancy we would like to keep it here as it currently is.

- It is not always clear which tables / figures refer to the calibration or validation set of gauges. II suggest clarifying this in the figure captions and in the relative text.

  Yes, we have clarified this in the updated version of the manuscript. Additionally, to the Figure caption and throughout the text.

- Figure 8: it would be interesting to see a similar figure for IPF vs the corrected MDF

  The corrected parameter distribution values are shown in boxplots in Figure 9 for annual and seasonal series and for all the methods. Since there are 4 correcting methods, 3 parameters and 3 maximum series, we thought that the best way to compare them is through boxplots. Also, it is easier to draw conclusions. However, we are happy to provide here the equivalent of Figure 8 for two other correction methods: slope-event as the reference method- Figure 1 below and LM-Lmoms method as the best method – Figure 2 below. As it is seen from Figure 2, the LM-Lmoms estimates successfully the location parameter and there is no overall bias in underestimation (as in the case of the MDF or slope-event method shown in Figure 1). The scale parameter is also better estimated on average by the LM-Lmoms method, however there is a slight underestimation for few sites in the scale parameter of the summer series. Overall, the results agree with Figure 9, where the location and scale parameters are systematically underestimated by the slope-event method, while the location from LM-Lmoms is unbiased.

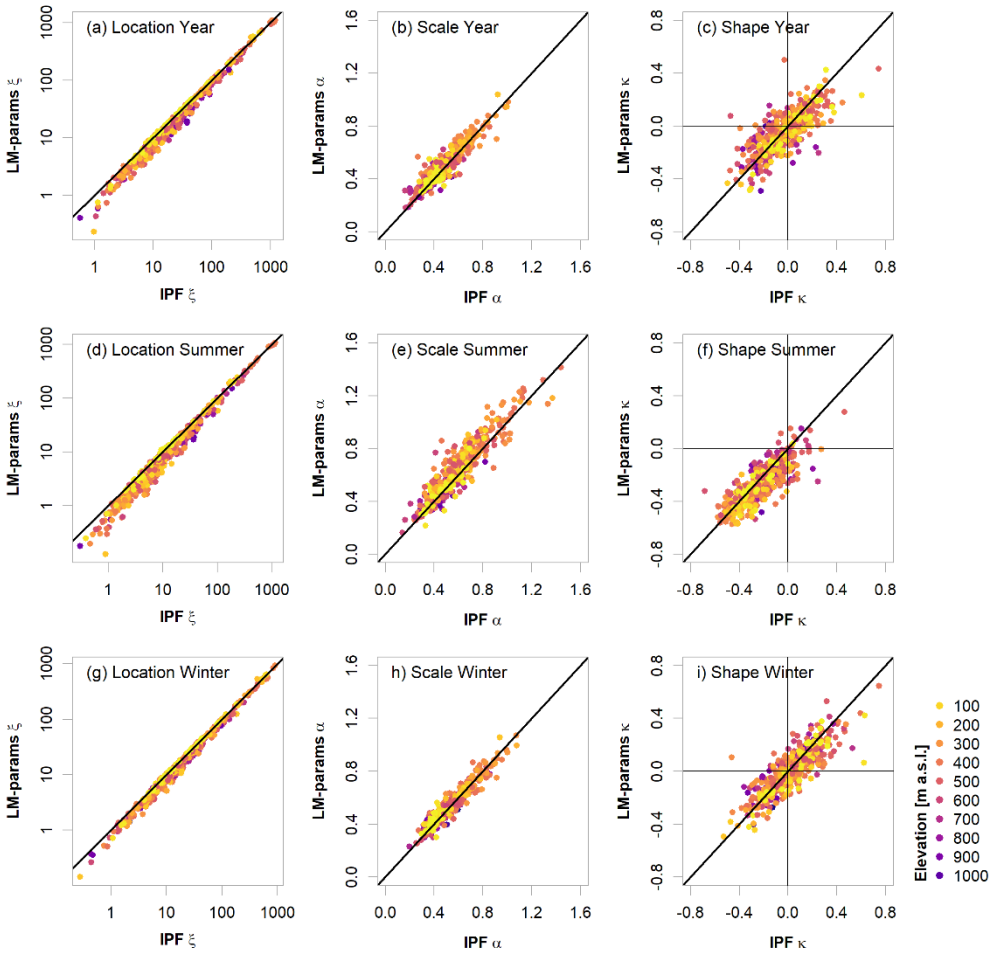

Figure1 Estimated GEV parameters from the slope-event method in comparison with the actual GEV parameters from the IPFs.

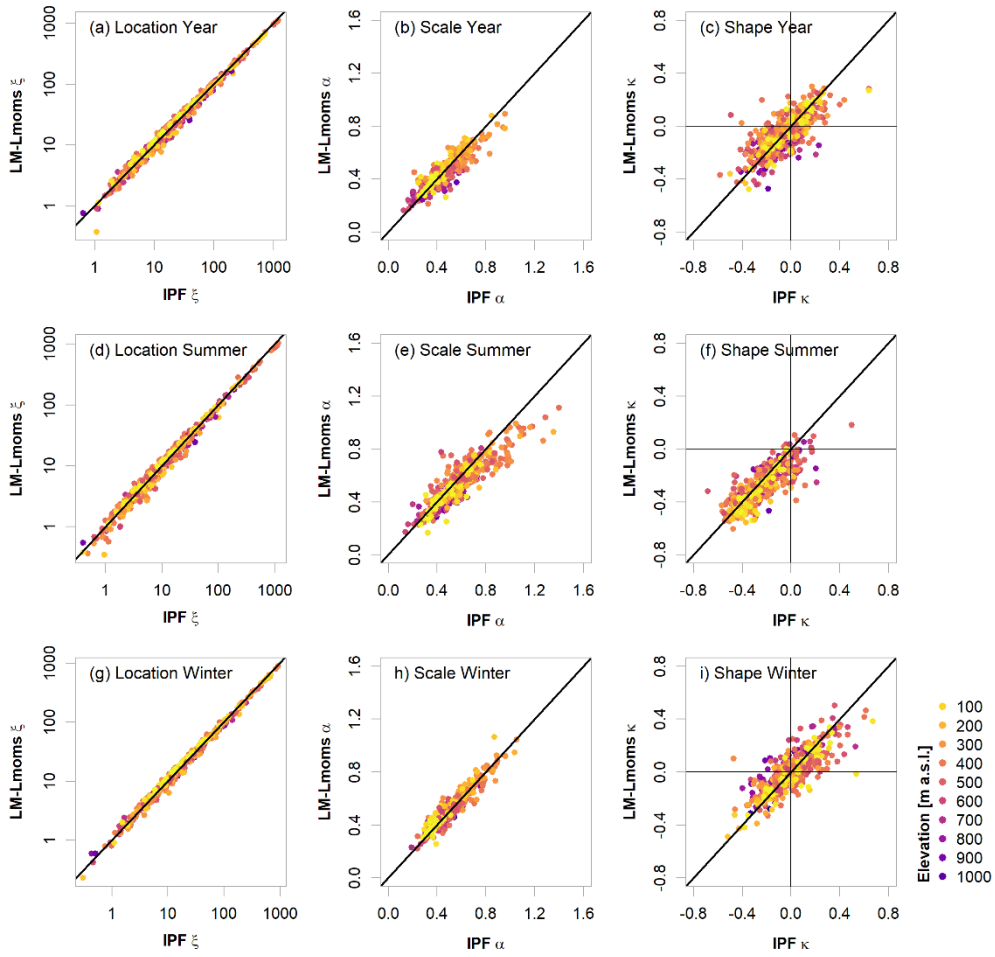

Figure 2 Estimated GEV parameters from the LM-Lmoms method in comparison with the actual GEV parameters from the IPFs.

- Figure 9: please specify in the figure caption the type of error shown (i.e. error in the parameters of the distribution).

  Yes, this the error in the corrected-MDF GEV parameters compared to the actual GEV-IPF parameters. We have added this to the caption of the figure.

- Lines 390-391: it is unclear to me.

  Mixed-Models combine the GEV distribution fitted for summer and winter floods independently. This means that the L-moment are derived separately for summer and winter maximum series, the linear models are then fitted independently for summer and winter, and finally the probabilities of both summer and winter are combined to calculate the annual non-exceeding probability. So, in comparison to the annual extremes, in mixed-models there are 2 times more linear models fitted (one for summer and one for winter) and thus more parameters to be fitted. But even in this case, with more models, the proposed methodology seems to work fine.

  To avoid confusion, we have added a short explanation about the mixed-models in the updated version of the manuscript.

- Table 8: it is unclear what “mixed-models” stands for. Does it refer to all year?

  It this case the extremes of each year are extracted independently for winter and summer maximum series. A GEV is fitted at each of these two series (summer and winter), and the probability of an extreme flood to happen annually is the multiplication of two independent non-exceedance probability.

  We have added a short explanation about the mixed-models in the updated version of the manuscript. The full description of the mixed-model is given at Fischer et al. (2016).

- Line 406: “light blue points represent 100 resampled model estimates”. It is not clear what is resampled exactly. Do the authors resample the original MDF and then they apply the correction or the other way around? Also in the following line “permutation in the linear models” and in figure 15 is not clear what you resample exactly.

  The MDF-bs and IPF-bs are uncertainties calculated by resampling the annual maximum series 1000 times and calculate the quantiles; the LM-bs-full are uncertainties calculated from fitting a linear models to the 1000 series produced from MDF-bs and IPF-bs and then calculating the quantiles; the LM-bs-mean is using the original MBFs and IPFs and sample them 1000 times in space and then fit the linear model and calculate the corresponding quantiles, and lastly, the LM-bs-bs-full is combining the local 1000 times resampling of AMS with the 1000 times of spatial resampling and then fitting the linear model and calculating the quantiles. We will clarify this in the updated version of the manuscript.

  The term “permutation in the linear models” is also updated accordingly.

- Terminology: resampling, permutation and bootstrapping are used as synonyms (as far as I understand) but they are not exactly the same. Please clarify and homogenize the terminology throughout the entire manuscript.

  We are sorry for the confusion; we have removed the terms permutation and bootstrapping, and we kept only the term „resampling“. In the new updated manuscript, we differentiate between:

  - Temporal resampling – resampling of the annual maximum series at each site to account for the sample and parameter uncertainty (MDF-bs, IPF-bs).
  - Spatial resampling – resampling the L-moments in space before fiting the linear model to account for the linear model uncertainty (LM-bs-mean).
  - Spatio-Temporal resampling – resampling annual maximum series at each site, and for each dataset we resample the L-moments again in space before fitting the linear model. This is a combination of sample and parameter uncertainty and linear model uncertainty and accounts for the overall uncertainty of our estimates (LM-bs-bs).

  There is another source of uncertainty, which is the propagation the sample and parameter uncertainty through the existing linear model (LM-bs-full).

- Lines 424-425: “At many stations there is a significant overestimation of the true IPF quantile..”. I am not sure what the authors refer to exactly. Instead, I see in figure 14 that the median is rather centred on 0.

  This is referred to the boxplot upper-CI for the LM-bs-bs, here we see that the median is of a positive value and is not symmetrical to the lower CI and it is higher compared to HQ10 and HQ50. We mean here „because of the sample and linear model uncertainty, at many stations there might be a significant overestimation of the true quantile “. We have made this clearer in the updated version of the manuscript.

- Line 425: what does “linear model transpositions” mean?

  The LM-bs-bs uncertainty illustrates the overall uncertainty by combining the sample and linear model uncertainty. With “linear model transposition” we meant originally the spatial resampling prior to the fitting of the linear model (so the linear model uncertainty). We will make this clearer in the updated version of the manuscript.

- Line 482: “even when equalizing the other factors catchment size and elevation”. What does it mean?

  Here we mean the following: Dividing the area in quadrants and considering similar catchment and elevation, still there were no considerable improvement on the results compared to pooling all gauges together. We have added a short explanation in the updated version of the manuscript.

References:

Fischer, S., Schumann, A., and Schulte, M.: Characterisation of seasonal flood types according to timescales in mixed probability distributions, J Hydrol, 539, 38-56, doi:10.1016/j.jhydrol.2016.05.005, 2016.