Authors point-by-point response (in blue) to reviewer 1 comments (in black):

1. Overall, the writing is mediocre and needs improvements. This is not only about English but the way the story is told.

   We have made several changes in the manuscript that make the flow easier to follow. We hope that our main goals and conclusions are now clearer to the reader.

2. Abstract needs to be revised to better discuss the methods and major results. Some quantitative results (numbers) are needed; as of now, everything has been discussed subjectively.

   The abstract has also been updated accordingly, by discussion the methods and the results. Also, numbers are given to discuss the finding more objectively.

3. In the Introduction section, it is not clear which approaches and methods you are addressing in different references. Please clarify. My comments in the PDF can clarify what I specifically refer to.

   We have tackled them in the new version of the manuscript. Some relevant responses from the pdf comments are listed below:

   3.1 Line 36: Moderate rain on snow also makes steep risings, but it is not related to the storage. Please clarify this with appropriate references.

   We have added a short explanation in the introduction about moderate rain on snow that can cause steep rising.

   3.2 Line 62: Catchment land cover, season (time) as well as the rainfall type affect this ratio. How did you separate them?

   Depending on the combination of these factor; catchment land cover, rainfall type, soil type and initial conditions etc., the catchment will respond differently and show a distinctive flood hydrograph. We are calculating the p/V ratio based on the mean daily flows (MDF) for each event recognized in the monthly instantaneous peak, so their effects are already reflected (and included) in the p/V ratio.

   3.3 At line 68: Why MHQ?

   We include the mean maximum flow because it is easier to find the best linear model setting that estimates adequately all flood quantiles. Since we have a variety of predictors (not all are included in this paper), it is easier to find the best setting based on the MHQ that will represent both low and high flood quantiles. If our aim was three quantiles, then the selection of the predictors may change for each case. In case one quantile is used as a target, then most probably the other quantiles will be under/over – estimated. Another important point to keep in mind is that other studies comparing IPF estimation method also analyse the annual maximum flow on average, and do not compare the quantiles. Thus, for a fairer comparison, the MHQ is used as a target variable.

3.4 Line 90: Why? Sometimes IPF can be bigger than the annual maximum peak even if it did not happen in the annual maximum peak day.

Yes indeed. However, correcting each event as in application 1, can lead to very high correction values for the MDF if smaller events are not properly distinguished. Thus, it might be more robust to correct the annual maximum daily peak, even though they may not occur at the same day at the IPF. That is why we compare these two applications with one another, to see which of them performs better in our catchments.

3.5 Line 209: Is there any snowmelt in May and June from high elevation areas? Line 220: Snowmelt in the winter or spring?

The main assumption is that the snowmelt occurs in winter (November to April). Of course, there might still be some snowmelt in May and June in the alpine catchments, and these events are classified as summer events. However, this affects only a very small portion of our catchment. We have discussed this shortly in the updated manuscript.

3.6 Line 228-230: But it is not only related to the size. What does the increasing annual maxima mean? Annual maxima is only one value per year

We mean here that mostly in small catchments, the annual maxima are wrongly chosen in winter season by the mean daily flows, rather than in summer season as observed in the instantaneous peak flows.

3.7 Line 402: In the methodology section, this was mentioned as 1,000 times. Please check and be consistent

Yes, in the methodology section we mean 1,000 realisations, and 1,000 realisations are used for the uncertainty analysis. However as explained in Line 401, this is just an example with a reduced number of 100 realisations to explain the uncertainty sources visually.

4. Please explicitly state the objective and research gap in the Introduction section.

   The objective and the research gap in the introduction section are better explained.

5. In addition to drainage area and topography, what other factors affect the peak/mean ratio? Why do you solely study the two factors?

   Prior to our study we have included a total of 56 descriptors (climate, soil, topography and hydro-geological descriptors) that were investigated in a stepwise regression to determine their importance on the peak ratio. As it turns out, area and elevation proved to be more important and that’s why they are our only focus presented here.

   We have included a short explanation in the updated version of the manuscript to make it clear why we have chosen these descriptors.

6. Instead of using the three return periods, why not using the historical events for the analyses between flows and drainage area-elevation?

   The focus of the paper is the flood frequency analysis with the aim to improve the extreme flood events for design purposes. That is why the focus is mainly on discharge levels with different return periods.

7. Methodology section lacks proper references for the assumptions, methods and equations.

   We have included more explanation in the methodology section.

8. There is a mention of the availability of IPF monthly maximum flows, but monthly is a too large timescale for instantaneous peak flow analyses. What is your justification for using that?

   I think there is a misunderstanding here. IPF series is provided as a monthly time series, that means for each month we have the highest instantaneous peak recorded (in m^3/s). This means for each year we have 12 maxima instantaneous peak recorded, and we can select the maximum one to continue with the annual maximum series. We have stated this clearly in the updated version of the manuscript.

9. How were IPF_{stat} and MDF_{stat} Please clarify.

   We have clarified this in the updated version of the manuscript. The *stat* refers shortly for statistics. Statistics on annual maximum series derived from both IDF and MDF series are used for the correction: like for instance the mean maximum discharge (MHQ), or the L-moments and so on.

10. Equations 2-3 need references. Have they been developed by the authors or others (need references)?

    Equation 3 are developed by the authors, while Equation 2 is motivated by the results of Fischer et al. (2016) and Fischer (2018). We have clarified this in the updated version of the manuscript.

11. Any rational for using the GEV distribution?

    According to Maidment 1993, three distributions are reasonable choices for describing flood flows namely: Generalized extreme value (GEV), lognormal and Pearson type 3 distributions. The previous study performed by Ding et al. 2015 in Lower-Saxony Germany, showed that the GEV had highest p-value (and acceptance of the distributions) in comparison to the other two types. Moreover, other studies like Villarini et al. (2011) or Haktanuer and Horlacher (1993) have also used GEV in several catchments in Germany. In our catchments we have performed the Cramer-von-Mises test for both original and corrected series and GEV was accepted in all of them (p-value = 5%). Since, we have relatively long observations (up to 148 years), we have employed GEV due to its flexibility to be fit different tails, as the shape parameter is adjusted independently for each gauge.

A short explanation on why we have chosen the GEV is included in the updated version of the manuscript.

Lastly, we would like to point out that we choose the p/V-Lmoment as the best method (correcting the L-moments of the mean daily flows) in order to have a more universal method which is not affected by the assumption of a probability distribution.

12. Why did you use a linear regression model in Equation 3?

    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 or climate 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.

13. What is Q_{suc} in Equation 4?

    Equation 4 is the model proposed by Chen et al. (2017) and is our reference for improvement. Their methodology (also called the slope method) corrects the IPF series only based of the MDF information. They calculated a slope by using the mean daily peak discharge (Q_{peak}), mean daily discharge one day before the peak (Q_{pre}) and the mean daily discharge one day after the peak (Q_{suc}). We have stated this clearly in the updated version of the manuscript.

14. In line 92, what does “IPF and MDF do not necessarily overlap” mean? Do you mean in terms of their timing or magnitude? Please clarify.

    One of the main assumptions for estimating IPF base on MDF series, is that the peaks of both data occur at the same dates. With the expression “IPF and MDF do not necessarily overlap” we mean exactly this, that the peaks of the two series do not necessarily happen on the same day. We have clarified this in the updated version of the manuscript.

15. Line 143: What does “annual maxima from monthly maximum” mean?

    As mentioned before the IPF series are maximum instantaneous peak given for each month. To compute the annual maxima, we compute the maximum from 12 values (one maximum instantaneous peak for each month) for each year- Like this we can built the annual maximum series (for the flood frequency analysis).

    On the other hand, for the MDF series are daily average flow data. Here we select for each year the maximum average daily discharge observed and so we obtain the annual maxima series.

16. Line 148: To what extent, does the sample size change the uncertainty? Is 1,000 a sufficient sample size?

    When comparing methods in terms of their uncertainty, the number for random resampling will influence all methods similarly. When comparing sources of uncertainty, 1,000 realisations are enough to shed light if the linear model has higher uncertainty than the sample/parameter uncertainty. For the overall uncertainty, there were in total 1,000,000 realisations. Of course, the uncertainty ranges might change slightly if more realisations are included, but previous test that we have conducted have shown that 1,000,000 realisations are enough to capture the overall trends of the uncertainty.

17. Lines 165 and 189 are inconsistent in terms of the number of discharge stations (648 vs 653). Which one is right?

    We are sorry for the confusion; we will clarify this better in the updated version of the manuscript. The right number of discharge stations is 648.

18. Overall, there is a lack of connectivity between the subsections under Results section. This section needs a better flow.

    We have restructured a bit the section of the results in the updated manuscript and hopefully the flow within the results section is now easier to follow. The result section is now separated in two three parts: Part 1 – estimating the mean maximum flow (MHQ), Part II – estimating the GEV parameters and flood quantiles, and Part III – assessing the main source of uncertainty and the overall uncertainty range of the best selected model. For each Part I and II, we first start with an analysis on the MDF series, to see how well they match with their respective IPF (also the influence of area and elevation is discussed) and then we assess the performance of the models. In Part I we focus on the MHQ to find the most suitable predictors and linear models, in Part II we use the predictors of Part I to assess the performance of the models in terms of GEV parameters and flood quantiles. In Part II the best model is selected. Lastly in part III the uncertainty of the best model is analysed.

19. Section 4.3: It is expected that because the two databases (IPF and MDF) are different, their distribution parameters are different too. What is the main reason for comparing the parameters of distributions? A more proper comparison should be on quantiles (different return periods).

    A comparison between the quantiles is already given in Figure 10, Figure 11 and Table 6. Nevertheless, one must keep in mind that the quantiles are estimated from the fitted GEV distribution, and to understand why some quantiles are not properly represented, one should look at the GEV parameters. For instance, in Figure 10, you will see the mean error over all quantiles for MDF is negative. This is explained by Figure 8 and 9. It is clear the GEV location parameter from MDF series is underestimating the IPF-location parameter. This will affect all quantiles as the distribution is shifted to the left, hence all quantiles are underestimated.

20. Figure 13: What is the main reason for similarity among IPF-bs, MDF-bs, LM-bs-full and LM-bs-bs in different HQ years?

    The reason for this similarity is that the main source of uncertainty is the sample and parameter uncertainty (MDF-bs). LM-bs-mean, which illustrates the uncertainty only due to the linear model fitting (here the sites are resampled 1000 times in space before fitting the linear model), is considerably lower than the MDF-bs. This explains why the LM-bs-bs is very similar to MDF-bs. The LM-bs-full is just propagating the MDF-bs uncertainty through the existing model, thus the uncertainty in this case will still originate (and be similar) to the MDF-bs. On the other hand, the IPF-bs and MDF-bs are slightly different from each other, where MDF-bs is slightly less uncertain with fewer outliers than IPF-bs.

We have extended the explanation in the updated version of the manuscripts.

21. Figure 14: What is the difference between the median of the three HQ-years in each part of the confidence interval? They seem similar in the boxplot median.

For each of the 486 validation sites, the median error over all realisations is computed and shown in the boxplot-median (for each flood quantile). The median error over all realisation is reflecting the same behaviour as the actual error as obtained in Figure 11. In Figure 11 the boxplots of three quantiles were also similar to one another, with clear differences in the outliers and slightly larger error spread for the higher quantiles. The same is true for the median error of all the realisations.

For each site, apart from the median error over all realisations, the 2.5 % and 97.5% error (here referred to as the lower CI and upper CI) quantiles are as well computed. The boxplot lower CI and upper CI – show the error spread among all sites. Important to see here is as the flood quantiles are increasing, the lower CI median will get lower (so higher underestimation), while the upper CI median will get higher (so higher overestimation). This means that the higher the flood quantile, the higher the error and the uncertainty. However, we also see here that for HQ100, the median of lower CI is not symmetrically mirrored in the upper CI (as is the case for HQ10 and HQ50). This means that the errors are positively skewed.

We have extended the explanation in the updated version of the manuscripts.

22. Are all methods and approaches sensitive to the database type? Can those be generalized to other catchments? If so, what are some considerations?

This is indeed an interesting question. The method proposed here is more sensitive to the flood typology of other catchments, which indirectly is mirrored in the database. In theory the predictor p/V, as it as a normalized predictor, should work well for other catchments as well. However, if the dominant floods in a catchment have a timescale less than a day (say flash floods with durations short than a day) then the daily measurements of the flow will not capture adequately the flood dynamics. Hence the linear models based on the p/V predictor may not yield good results. This was also the case in our catchments with areas lower than 100km^2. In other cases, when the flood timescale is larger than a day, then the p/V predictor should be able to capture the flood dynamics. Still, attention must be paid to the baseflow separation, to make sure that the calculated p/V predictor is representative of the catchment behaviour.

Another thing to keep in mind, is what gauges and most important how many gauges one should group together for the fitting of the linear model. In the optimal case that the p/V predictor describes the flood dynamics correctly at each catchment, the question becomes how good one linear model can represent the whole group of catchments. Although L-moments are considered more robust than parameters or quantiles, it might be that L-moments are considerably different within the catchment group, then it makes sense to break the group down in more subgroups to better capture the L-moments. In this case we suggest the flood index clustering as suggested in Howking and Wallis (1997).

23. All acronyms and abbreviations should be spelled out in the keywords, figures, tables and headings.

    We will make sure that all acronyms and abbreviations are spelled out in the new version of the manuscript.

24. Please italicize all parameters and coefficients throughout the text.

    We have italicized all parameters and coefficient in the updated version of the manuscript.

Authors point-by-point response (in blue) to reviewer 2 comments (in black):

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.

   We have decided to use the term "p/V approach /method". We have made the necessary changes in the updated version of the manuscript. This includes changes in the text, tables and figures.

3. 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.

   We have updated it as "sample and parameter uncertainty" in the new version of the manuscript.

4. 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.

5. 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.

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

   We have changed it as you proposed in the updated version of the manuscript.

7. 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.

8. 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.

9. 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 have provided a short discussion about this at our reply to the reviewer.

10. 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.

11. 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.

12. 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).

13. 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.

14. 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 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).

15. 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.

16. 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 made this clearer in the updated version of the manuscript.

17. 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.

Dear Editor,

below are the listed changes that we have introduced to the updated version of the manuscript following the comments of the two reviewers. In blue we are showing the actual changes in the manuscript. Please note that due the amount of revision made, not all of the changes are mentioned point-by-point. Instead we have tried to group the changes made.

1. As already discussed we have added a new author (as second author) and changed the corresponding author:

   Anne Bartens[1], Bora Shehu[2], Uwe Haberlandt[1]

   [1]Institute of Hydrology and Water Resources Management, Leibniz University of Hannover, Germany
   [2]Institute for Environmental Sciences and Geography, University of Potsdam, Germany

   Correspondence to: Uwe Haberlandt (haberlandt@iww.uni-hannover.de)

2. The abstract has been adapted accordingly to make it clear the need for and the novelty of the study, together with the main findings. As wished by the reviewer 1 now the statements in the abstract are more objective (with number that support our statements). The updated version of the manuscript is given below:
   Lines 1-16: "In many cases flood frequency analysis (FFA) needs to be carried out on mean daily flow (MDF) instead of instantaneous peak flow (IPF), which can lead to underestimation of design flows. Typically, correction methods are applied on the MDF data to account for such underestimation. In this study, we first analyse the error distribution of MDF derived flood quantiles over 648 catchments in Germany. The results show that using MDF instead of IPF data can lead to underestimation of the mean annual maximum flow (MHQ) up to 80% and is mainly dependable on the catchment area but appears to be influenced as well from gauge elevation. This relationship is shown to be different for summer vs winter floods. To correct such underestimation, different linear models based on predictors derived from MDF hydrograph and catchment characteristics are investigated. Apart from the catchment area, a key predictor in these models is the event-based ratio of flood peak and flood volume (p/V ratio) obtained by MDF data. The p/V models applied either on MDF derived events or statistics, seem to outperform other reference correction methods. Moreover, they require a minimum of data input, are easily applied and valid for the entire study area. Best results are achieved when the L-moments of the MDF annual maximum series are corrected with the proposed model, which reduces the flood quantile errors up to 60%. This approach behaves particularly well in smaller catchments (<500km2), where reference methods fall short. However, the limit of the proposed approach is reached for catchment sizes below 100 km², where the hydrograph information from the daily series is no longer capable of approximating instantaneous flood dynamics, and gauge elevation below 100m, where the difference between MDF and IPF floods is very small."
3. The introduction to the topic has been updated to improve the flow of the text following mainly the comment of reviewer 1.
   - More explanations are given to tackle the comments from both reviewers.
     Lines 21-23: "For FFA to be as accurate as possible, two criteria need to be met; first a large number of observed peak flows is necessary to ensure an adequate selection and fitting of the probability distribution, and secondly it is important that the peak flows are measured with high precision in order to account for the best description of maximum flood magnitude and dynamics."
     Lines 25-27: "Such data is rarely available, or at the best case only available for short periods, which is insufficient for flood frequency analysis. Typically, long observation of floods are available as mean daily flows records and oftentimes FFA needs to be carried out on these records instead."
     Lines 28-30: "Particularly for small basins, there is a considerable underestimation of flood peak by the mean daily flows (Fill and Steiner, 2003). Hence it becomes essential to develop

new methods based on easily accessible data to correct the mean daily flows for a better representation of the flood peaks."

Lines 35-36: "For instance, there is a visible trend of the peak ratio to decrease with the basin area, which is expected as larger basins have higher baseflows (Ellis and Grey, 1966)."

Lines 38: "This means that the peak ratio of rainfall and snowmelt events are different from one another."

Lines 39-42: "…can be generally classified as methods based on the catchment characteristics as in Fuller (1914), Ellis & Gray (1966), Canuti & Moisello (1982), Ding et al. (2015) or including also climate characteristics as in Taguas et al. (2008), Muñoz et al. (2012) and Gaal et al. (2015). Mostly, these methods are in the form of linear models based on maximum MDFs and the selected catchment or climate predictors"

Lines 47: "…and/or high intensity rainfall or due to moderate intensity rainfall on snow"

Lines 65-67: "Indeed, the estimation methods based on the catchment or hydrograph characteristics remain still more desirable due to their simplicity, as they are based on easily accessible data and popular methods (i.e. linear models)."

- Additionally, we have added two new paragraph to emphasize the novelty that the study is bringing as well to give a summary of how the manuscript is organized:

Lines 68-94:"So far, the two main IPF estimation methods are developed distinctively from one another with no combination of both catchment and hydrograph information. In this study we propose linear models that facilitate IPF estimation using a combination of daily event hydrographs and functional dependencies with geomorphic catchment descriptors, while keeping the data input to a minimum. Key predictor in this approach is the ratio of direct event peak runoff and direct event volume. This ratio is expected to effectually describe the shape of a flood event, which in turn gives an idea about the expected instantaneous peak: the larger the daily peak and the smaller the event volume, the larger the expected difference between IPF and MDF and vice versa. We assume that the peak-volume ratio (*p/V*) holds important information on the general behavior of flood events (Tan et al., 2006; Gaál et al. 2015; Fischer, 2018) and thus on the expected magnitude of the IPF peaks as well. Moreover, the *p/V* of individual events can describe the internal variability at a site by reflecting different types of floods caused by different rainfall and/or snowmelt inputs. At the same time the *p/V* accounts for the variability between sites caused by local flood generating processes governed by general physiographic and climatic conditions.

Another important point to be considered is that most of the studies mentioned before investigate the performance on IPF maximum series and pay little attention to how these methods estimate the design flows with specific return periods. The general assumption is that, if the IPF maximum series are estimated well enough on average, so are the IPF quantiles. However, a well estimated average IPF maximum may still lead to underestimation of design flows with a high return period (say 100years). It makes sense to investigate as well if linear models based on MDF- moments, parameters or quantiles are more favorable for the estimation of the IPF quantiles. Accordingly, p/V models are employed here to correct MDF information at different levels; correction of individual flood events from MDF, correction of MDF annual or seasonal maximum series, and the direct correction of MDF-derived statistics (like mean maximum flow, L-moments, distribution parameters or even flood quantiles).

In this study, the linear models based on the *p/V* as key predictor (referred here as p/V models) are developed and assessed based on flow data from 648 catchments in Germany (as described in Section 2). The description of methods and models used here for the estimation of the IPF from MDF information is given in Section 3.2. We then analyze the performance of the models in two main parts: their ability to estimate the mean maximum flow (MHQ) (see Section 4.1) and their ability to estimate probability distribution and the respective design floods (see Section 4.2). For the best model achieved, an uncertainty estimation is tackled by means of spatio-temporal resampling (see Section 4.3). Finally, the range and limitations of the proposed methodology and conclusions are given in Sections 4.4 and 5."

4. We have change the section of the "Study area and data" as the second section after the introduction.
    - Here we explain as well what the monthly instantaneous peaks are (following the comments of the first reviewer):
      Lines 96-97: "For the analyses, continuous average daily flow (MDF) and instantaneous peak flows provided for each month as monthly peaks (IPF) are available"
    - Line96: The total number of gauges used for the study was 648, and it was updated throughout the text accordingly.
5. The "Method" section has been re-arranged to introduce a better flow within the section (following the comments of both reviewers).

5.1 We start now we an introduction to the flood frequency analysis (section 3.1 Flood frequency analysis).

- An explanation is added on how the flood frequency analysis is performed
  Lines 130-133: "First the maximum series are extracted from each dataset either on an annual basis (annual maximum series – AMS) or for each season summer vs winter (seasonal maximum series). For extrapolation of the maximum series and estimation of floods with specific return periods, distributions were fitted to the annual and seasonal samples of both IPF and MDF datasets."
- At Line 136 - Equation 1 for the GEV distribution was updated accordingly to Maidment (1993):

$$F(x) = \exp\left\{-\left[1 - k \cdot \frac{(x-\xi)}{\alpha}\right]^{\frac{1}{k}}\right\}$$

- Additionally, we have added short text that explains why we have chosen the GEV distribution for the analysis:
  Lines 138-140: "The GEV has been proven before to be a suitable distribution for different catchments in Germany as indicated by Haktanir and Horlacher (1993), Villarini et al. (2011), Ding et al. (2015 ,2016), Ding and Haberlandt (2017) and therefore has been chosen in our study as well"
- We have discussed shortly the need for the seasonal flood frequency analysis:
  Lines 142-146: "When extracting annual maximum series (AMS) different flood peaks from different genesis (i.e. from convective/stratiform rainfall, from snowmelt and so on) are mixed together and described by a single GEV distribution. However, if a certain flood type is dominating the annual maxima sample but is not typical for extremely large floods, then the fitted GEV distribution becomes misleading. To consider the different genesis in the flood peaks, maximum series are derived here for two seasons; summer (May-October) and winter (November-April). Then a mixed model is applied, which combines two GEV distributions fitted to each of these subsamples of the data, like summer and winter floods."
- We have added a short explanation of the variables in equation 2:
  Lines 150-151: "with fi(x) as the annual non-exceedance probability calculated for each sub-sample (summer and winter) and Fmix(x) as the mixed-model annual non-exceedance probability for a flood value x."

5.2 In section 3.2 we describe the methods employed to estimate the instantaneous peak flow statistics. Thus this section has been named to "3.2 Analysis and estimation of instantaneous peak flows (IPF)".

5.2.1 First in section 3.2.1. we introduce the main predictor for out estimation methods – the p/V ratio. This section is renamed to "3.2.1 Calculation of the p/V predictor from mean daily flows". The following changes were done in this subsection:

- Here we explain our main motivation for this predictor:

  Lines 159-162: "Motivated by the recent findings of Fischer et al. (2016) and Fischer (2018) regarding different flood types, here the flood peak-volume ratio p/V extracted from mean daily flows (MDF) is considered an important predictor that can help to estimate more accurately the IPF series from the MDF ones."
- Since the calculation of the p/V ratio is based on the baseflow separation, we have introduced the baseflow separation in this section.

5.2.2 In section 3.2.2 we introduce all methods employed to estimate the instantaneous peak flow statistics. This section is renamed to "3.2.2 Estimation of instantaneous peak flows". Changes are made in the text to ensure a proper reading flow and to better explain what methods are investigated and why:

- We have included text that explains why we focus on the two additional predictors (elevation and area) and not on other ones:

  Lines 179-181: "Several catchment descriptors describing land use, soil type, average climate variables, geographic information and catchment morphology were investigated prior to the study. Two main descriptors, namely basin area and gauge elevation, were found to be more important for the linear model and hence are included in the study shown here."
- Following the comments of reviewer 1, we have added the following explanation to equation 5:

  Lines 195-196: "The slope-method estimates an instantaneous event peak flow IPFevent based on the slopes of the daily peak Qpeak to its preceding Qpre and following daily flows Qsuc as shown in Eqn. 5:"
- Following the comments of reviewer 1, the following short explanation was adapted:

  Lines 205-206: "Procedure 1) is theoretically more accurate, since maxima in IPF and MDF do not necessarily occur at the same time (no temporal overlap)."
- More information on the given methods are given.

  Lines 225-230: "An overview of all the methods employed here together with their description is given in Table 1. All methods consisting of the linear models based on the p/V ratio as a main predictor (p/V-method) have been optimized based on the calibration set only for the period 1972-2012. For the selection of the best model, the coefficient of determination (R2) and the significance of model parameters (based on the p-value) are considered. For validation all sites with their respective observed period are used. Through the validation we compare and assess the ability of the proposed models to capture the mean maximum flow (MHQ) and the probability distribution and the respective design floods.
  ."
- We have added a table with the description of all methods in order to make it clearer for the reader what we investigate:

Table 1: Description of all the methods and their applications employed here for the estimation of the IPF and their statistics.

Application	*Name*	*Description*
Reference	MDF	IPFs are taken directly without correction from average daily flows MDF
Event-based analysis	Slope-events	Estimate IPF for all flood events from MDF according to Eqn. 5
	LM-events	Estimate IPF for all flood events from MDF according to Eqn. 4

AMS-based analysis	Slope-AMS	Estimate IPF as per Eqn. 5 only for events that corresponds to MDF annual/seasonal maxima
	p/V-AMS	Estimate IPF as per Eqn. 4 only for events that corresponds to MDF annual/seasonal maxima
Statistics-based analysis	p/V-Lmoms	Estimate IPF L-moments as per Eqn. 6 based on the MDF L-moments derived from the annual/seasonal maximum series
	p/V-params	Estimate the IPF GEV parameters as per Eqn. 6 based on the MEF GEV parameters derived from annual/seasonal maximum series
	p/V-quants	Estimate IPF quantiles as per Eqn. 6 based on MDF quantiles derived from annual/seasonal maximum series
	p/V-MHQ	IPF mean maxima (MHQ) estimated as per Eqn. 6 based on MHQ extracted from MDF annual/seasonal maximum series

- Following the comments of the review 2 we have changed the name of the proposed method to "p/V method". This has been changed throughout the manuscript.
- We have as well updated equation 6 following the comments of the reviewer 2 (since it was a typo):

$$\mathrm{IPF_{stat}} = \frac{\mathrm{MDF_{stat}}}{(a + b_1 * \mathrm{p/V_{mean}} + b_2 * \mathrm{CD}_1 + \cdots + b_{n+1} * \mathrm{CD}_n)}$$

5.2.3 In section 3.2.3 we give the description of how the methods applied are tested. This subsection is now named "3.2.3 Analysis of instantaneous peak flows".

Following the comments of reviewer 1, we have added an explanation why a direct comparison for each event is not possible:

Line 235: Since the IPF data are not continuous rather a maximum for each month (see Section 2), a direct comparison for each flood event is not possible.

- Following the comments of reviewer 1, we have added the following explanation to equations 7 to 9:

Line 237: "For this purpose the general difference between IPF statistic IPFstat and MDF-estimated flood statistics MDFstat are calculated as following:"

Line 245-247: "where N is the number of the validation sites, MDFstati and IPFstati are the respective statistics from MDF and IPF series, and sdIPFstat is the standard deviation of IPF statistics over all considered sites. These criteria are computed for each of the methods described in Table 1."

- We have added two equations that explain the two performance criteria RMSE and BIAS and a description on how they are calculated:

Lines 241-247: "Apart from the Error (%) at each site, two additional performance criteria are calculated over all sites: the normalized root mean square error nRMSE (%) as per Eqn. 8 and the percent pBIAS (%) as per Eqn. 9.

$$nRMSE\ (\%) = \frac{\sqrt{\frac{\sum_{i=1}^{N}\left(\mathrm{MDF_{stat,i}} - \mathrm{IPF_{stat,i}}\right)^2}{N}}}{\mathrm{sd(IPF_{stat})}} * 100\ \%,$$

$$pBIAS\ \ (\%) = \frac{\sum_{i=1}^{N}\mathrm{MDF_{stat,i}} - \mathrm{IPF_{stat,i}}}{\sum_{i=1}^{N}\mathrm{IPF_{stat,i}}} * 100\ \%,$$

where N is the number of the validation sites, MDFstat,i and IPFstat,i are the respective statistics at site i from MDF and IPF series, and sd(IPFstat) is the standard deviation of IPF statistics over all considered sites. These criteria are computed for each of the methods described in Table 1."

5.3 In section 3.3 we have kept the uncertainty analysis section and renamed to "3.3 Uncertainty analysis".

- A short text was added to clarify that the analysis is performed only on the best method for the estimation of the instantaneous peak flow statistics:

Lines 250-251: "As it will be later shown in Section 4.2 the best linear model is chosen to be the p/V-Lmoms - the model correcting directly the L-moments of the MDF series."

- Additional lines were added to explain how the uncertainty experiments are performed. Following the suggestion of reviewer 2 we now use the term "sample and parameter uncertainty" instead of "uncertainty due to distribution fitting". Respectively we have changed the name of the uncertainty types and made it clear for the reader. These updated text is as follows:

Lines 251-266: "This is done by using simple resampling with replacement procedures; resampling in time when selecting the maximum series for FFA, resampling in space when selecting the sites for the p/V model (either calibration or validation of the models) and resampling both in space and time. In a first step, the series of annual/seasonal maximum from both MDF and IPF dataset are analogously resampled 1000 times with replacement (temporal sample and parameter uncertainty). For each resampling the desired flood quantiles are estimated using L-moments. The range of these estimates provides the baseline level of uncertainty due to sample and parameter uncertainty. The temporal uncertainty is calculated at each site for the original MDF and IPF series (respectively MDF-bs and IPF-bs) and are considered a benchmark for comparison.

In a second step, p/V models are fitted to each pairing of temporally resampled MDF and IPF series while considering all sites in the study area that have more than 30 years of observations. This means that the temporal sample uncertainty is propagated through the p/V model (p/V-full). To assess the uncertainty of the selected p/V model, another resampling is carried out, this time shuffling the set of considered sites, where original MDF site specific L-moments are resampled again 1000 times with replacement before fitting the p/V model (p/V-bs). Lastly the total uncertainty both in space and time is assessed by combining the temporal sample and parameter uncertainty with the uncertainty of the fitted models. This means that the maximum series are resampled 1000 times, and for each of these sets, the sites are resampled 1000 times as well before fitting the p/V model. So, the total uncertainty will be derived by 1000 x 1000 quantile estimates (p/V-bs-bs)."

6. The result and discussion section has been restructured slightly to make the results flow more understandable.
   - We have introduced two subsections 4.1 and 4.2 focusing on the two target variables; Section 4.1 focusing on the mean annual maximum peak flow and Section 4.2 focusing of the peak flow distribution and the respective quantiles. Each subsection if further divided in two parts: part 1 focusing on the differences between mean daily and instantaneous peak flows, followed by part 2 how the proposed methodology and the reference methods perform in estimating the instantaneous peak flows from mean daily flows. The best method declared in Section 4.2, is then used as a basis for the uncertainty analysis illustrated in Section 4.3. The results and discussion section is then concluded with a discussion on the ranges of applications and limitations of the proposed

methodology. The reconstructed subsections of the result and discussion are as follows:

"4.1 Mean maximum flows (MHQ)
4.1.1 Comparison of mean daily and instantaneous peak flows
4.1.2 Estimation of mean maximum flow (MHQ)
4.2 Probability distribution and derived design flows
4.2.1 Comparison of mean daily (MDF) with instantaneous peak (IPF) flow distributions
4.2.2 Estimation of instantaneous peak flow (IPF) distributions and quantiles
4.3 Uncertainty Analysis
4.6 Ranges of applications and limitations."

- The original Figure 4 and 5, have been combined together in a single Figure 4. Note that nothing has changed in the plots, only that now they are in the same Figure. Following the comments of review 1, the explanation of Figure 4-e (previously as Figure 5) is now changed to:

Lines 234-238: "Another general issue highlighted by this analysis, independent of seasonality, is that the peaks of both IPF and MDF dataset do not necessarily occur at the same day (there is no temporal overlap). In their study, Chen et al. (2017) illustrated that only on 82% of the events investigated, the peaks of both IPF and MDF series occurred on the same day. This suggests that instantaneous maxima are not always identifiable in the mean daily flows, i.e. the maxima obtained from the daily series are inevitably found in other places. The temporal 320 overlap of IPF and MDF derived peaks for our catchments is shown in 4 (e)."

Lines 240-243: "This problem needs to be kept in mind when attempting to estimate instantaneous peaks from daily peaks, since the two may belong to significantly different events (different genesis) and thus to different populations."

- Certain text that was not understood by the reviewers has been rewritten to make it clear for the reader. Additionally, more information was given that explains what results are showing, to compare the results of this study with other studies from the literature and to explain the errors introduced due to generalisations in two seasons.

Lines 247-249: "We first test the suitability of various predictors to predict MHQ of IPF by fitting the p/V models to the individual events of MDFs (p/V-event), to the MDF maximum series (p/V-AMS) or lastly directly to the MDF mean maximum flow (p/V-MHQ)."

Lines 426-428: "So far the proposed p/V models were analyzed in their ability to estimate better the mean maximum flow (MHQ) from MDF data. In this subsection the focus is shifted to the ability of the methods to estimate parameter distribution of the IPF and the derived flood quantiles."

Lines 443-444: "Overall, due to the underestimation of the location parameter leads, underestimation of both the lower and higher flood quantiles by the MDF sample is expected."

- All captions of the figures have been updated with more information (mainly what data set is used to get these results), as well the names of the methods in tables and figures have been updated accordingly.
- Figure 8 (originally Figure 9) has been re-plotted with 2 rows and 3 columns to better fit the graphic in the page. Also following the comments of reviewer 2 we have added the name on the y-axis "GEV parameter error (%)".
- Following the comments of reviewer 2, in the result section of the uncertainty analysis, we have added more text to explain better what experiments were run and what they represent. As well the terms permutation and bootstrapping were removed, and we have used only the term resampling. The updated text is now as below:

Lines 534-544: "Figure 11 (b) shows the resampled IPF flood quantiles (IPF-bs) vs. the quantiles estimated using the p/V-Lmoms model by considering different sources of uncertainty; p/V-bs illustrates the uncertainty only due to the fitting of the p/V-Lmoms model, p/V-full indicates the sample and parameter uncertainty (MDF-bs) propagated through the p/V-Lmoms model, and p/V-bs-bs combines the sample and parameter uncertainty (MDF-bs) with the p/V-Lmoms model uncertainty (p/V-bs) to tackle the total uncertainty. In this example, it becomes obvious that uncertainty from the p/V model (p/V-bs) is significantly smaller than the sample and parameter uncertainty (MDF-bs or even IPF-bs). This is valid for the majority of sites and is hardly affected from the number of realisations."

- Following the comments of reviewer 1 we have adapted the explanation of Figure 12 to the following:

Lines 551-555: "Figure 12 shows the relative widths of the 95% confidence intervals for all types of uncertainty estimated. The average widths of the IPF-bs, MDF-bs and p/V-full seem to be similar with each other, where the IPF sample and parameter uncertainty shows a larger variability. The width of the average range of the p/V-Lmoms model uncertainty (p/V-bs) is very small at all sites and therefore contributes little to the overall level of uncertainty (p/V-bs-bs). Thus the overall uncertainty of the p/V-Lmoms model is mainly influenced by the sample and parameter uncertainty of the original MDF series."

- Following the comments of reviewer 1 and 2 we have adapted the explanation of Figure 13 to the following:

Lines 564-573: "Figure 13 shows the median deviations of theMDF-bs and p/V-bs-bs quantiles from the respective IPF-bs quantiles, as well as the lower and upper limits of the 95% confidence intervals of the errors for the 10-, 50- and 100-year flood quantiles. The median errors from p/V-bs-bs are very similar over the three quantiles, but the higher quantiles HQ100 exhibits higher outliers. This is in agreement with the performance of the p/V-Lmoms model illustrated in Figure 9. This means that the median errors obtained over the 1,000,000 realizations are very similar with the actual model errors at each site. Moreover, it is obvious that the overall uncertainty gets larger with increasing return period, as can be seen by the increasing distance between lower and upper confidence limits. The p/V-Lmoms estimates appear to be slightly positively skewed, which is especially noticeable in the 95% confidence interval for the HQ100. At many sites there is a significant overestimation of the true IPF quantile when combining sample and parameter uncertainty with p/V-Lmoms model uncertainty. The MDF estimates on the other hand exhibit the expected persistent underestimation."

- Following the comments of reviewer 2 we have added the following explanation:

Lines 634-636: "Dividing the study area into quadrants does not result in any differences between the subsets, even when considering similar catchment size and elevation."

7. Following the comments of the reviewer 1, we have made sure to italicize all parameters and coefficients, that all acronyms and abbreviations are spelled out in section names and figures or tables captions.