

Reply to Comments

We want to thank all reviewers for their valuable and constructive comments!

Please find our point-by-point replies to the reviewer's comments below. The comments are taken from the original commentary. Replies are given in green below the respective comment.

Reply to comments by reviewer 1:

The authors added a new analysis (dynamics of driving mechanisms) to the manuscript and extended existing sections of the manuscript to clarify and describe the new additions (methods, results, discussion, and conclusion) as recommended by both reviewers. The supplemental material has also been updated and extended (model performance, dynamics of HF₁₀₀). Furthermore, we moved a part describing the benefit of the LE approach to the supplemental material. A point-by-point reply to the comments is given below.

1. The fourth paragraph in Introduction (Lines 73-77, Page 3) is vague. The two first sentences do not seem to be the reason for the choice of study area.

Reply: The authors rephrased this paragraph to clarify the choice of study area.

2. Can the authors explain the choices of CanESM2 and the use of RCP8.5 only?

Reply: The authors added a few explanations for the choice of a single GCM as well as a single scenario.

3. The study area has abundant in situ data. Should model parameters (i.e., soil properties) be calibrated?

Reply: The authors added a paragraph explaining the necessity for calibrating the hydrological model despite the abundance of data available for the region.

4. Since the reliability of the hydrological model affects the simulated discharge and the further analysis, the performance of the hydrological model should be presented and discussed in more detail. For example, for 16 gauges having NSE lower than 0.5 and 5 gauges having KGE lower than 0.5 (Lines 203-204, Page 7), an explanation is needed to show that the unsatisfactory is acceptable. For the other gauges with NSE and KGE higher than 0.5, how much higher than 0.5 are they? I think it is worth having maps that show the value of model performance metrics at all gauges.

Reply: The authors added further details about the model performance within the manuscript and added maps showing the spatial distribution of the model performance for each gauge within the supplementary material (S4).

5. Panels c, d, and e in Fig. 7 do not show the entire variation ranges.

Reply: The authors added the entire range of variation to the panels c, d, and e in Figure 7.

6. I think it is worth having a map that visualizes the spatial variation of the change in return period under climate change impacts, some interesting insights might be found. I am curious about the difference between the changes in return period in mainstream

and tributaries. Similar to the change in magnitude/intensity of the 100-year flood events under climate change impacts.

Reply: The authors added maps showing the spatial distribution of the changes in HF₁₀₀ frequency and magnitude to the supplementary material (S6) and refer the reader in the manuscript to this section of the supplements.

7. The authors show that using hydro-SMILE provides a more robust extreme hydrological discharge values under climate change impacts, but do not discuss on how to make use of that finding in designing flood protection infrastructures, the problem that authors state from the beginning of the paper.

Reply: The authors added the following statement to the end of the conclusion section of the manuscript to clarify that the authors are not able provide advice for the adaptation of flood protection infrastructures other than providing more robust values of future events: “Further studies are necessary focusing on flood inundation to fully analyze the extent of the increase and frequency of this event for the design of flood protection infrastructure”.

8. [Technical correction] Line 297, Page 12: “0.49 and 1.91 for 100 AM values (panel c) and 0.56 and 1.60 for 200 AM values (panel e)”. Should it be (panel b) and (panel c)?

Reply: The authors changed the references of the respective panels.

Reply to comments by reviewer 2:

The authors added a new analysis (dynamics of driving mechanisms) to the manuscript and extended existing sections of the manuscript to clarify and describe the new additions (methods, results, discussion, and conclusion) as recommended by both reviewers. The supplemental material has also been updated and extended (model performance, dynamics of HF₁₀₀). Furthermore, we moved a part describing the benefit of the LE approach to the supplemental material. A point-by-point reply to the comments is given below

1. The manuscript has a title of “...high return levels of peak flows...” but specifically focused on 100-year floods. With a large ensemble, the authors could investigate a range of high return levels of peak flows and see how the frequency, magnitude, and dynamics are projected to change in the future climate.

Reply: The authors added a statement to the manuscript explaining the reasons for focusing on the 100-year flood event to the manuscript.

2. The manuscript focused on proving that it is beneficial to have a large ensemble to estimate extreme peak flow events (Sections 3.1 and Figures 4 and 5), which, I think, is very obvious so I suggest moving them to the supplementary. The authors have generated a powerful dataset for extreme peak flow estimation, however, there is no analysis of the changes in flood frequency and magnitude. The authors are suggested to substantially expand the analysis on flood frequency and characteristics. See Yu et al. (2020) for an example of flood frequency analysis.

Yu, G., Wright, D. B., & Li, Z. (2020). The upper tail of precipitation in convection-permitting regional climate models and their utility in nonstationary rainfall and flood frequency analysis. *Earth's Future*, 8,

e2020EF001613. <https://doi.org/10.1029/2020EF001613>

Reply: The authors remained with their approach for a flood frequency analysis to illustrate the dynamics of the 100-year flood within a changing climate as provided within the manuscript in sections 2.2.4 and 3.2. We further added maps showing future changes in HF100 frequency and magnitude to the supplements (S6).

3. The authors only gave limited information on the evaluation results of the hydrologic model which are essential for building confidence in the following analysis. I suggest including figures and/or tables showing the evaluation results such as time series of flow events with other quantitative metrics such as correlation coefficient, % bias, root mean square error, KGE, NSE, etc. One related question would be how the level of trust (LOT) is calculated.

Reply: The authors further elaborated on the model's performance within the manuscript and referred the reader to other publications presenting the performance of the very same model. We also added a section to the supplemental material (S4) providing maps of the performance metrics.

4. The changing dynamics in the future climate (Section 3.2) are very interesting and worth digging into. The authors could dig into it with a mechanistic investigation of possible explanations for why they see such changes in future projections. For example, linking snow water equivalent and rain characteristics to the dynamical changes that are projected at the nivo-pluvial stations.

Reply: The authors added two new sub-sections (one in the method section and one in the result section) describing the methods to analyze the dynamics in driving mechanisms for extreme discharges above the HF₁₀₀ event and showing the results. We further added paragraphs to the discussion and conclusion sections regarding results of the dynamics in driving mechanisms.

Minor comments:

Lines 57-59: it is unclear how prediction is a reason for challenges in modeling and predicting high flows by Brunner et al. (2021a).

Reply: The Authors removed the prediction part from the manuscript.

Line 71: change "extraordinary" to "extreme".

Reply: The authors exchanged the term as recommended.

The paragraph starting from line 73: Needs substantial rephrasing. 1) Rephrase the sentence "This approach of high spatiotemporal resolution for climate and hydrological modeling is computationally demanding." Do the authors mean "This ensemble-based climate and hydrological modeling approach is computationally demanding because of the high spatiotemporal resolution"?

Reply: The authors rephrased the mentioned paragraph for better clarity.

2) Rephrase sentence “However, considering spatially refined catchment features (e.g., slopes, soil characteristics, land use), precise values due to higher temporal resolution, and the application of a SMILE for hydrological modeling supports an enhanced representation of extreme values within models.” Do the authors mean high spatio-temporal resolution of hydro-SMILE is particularly valuable for an enhanced representation of extreme values in models because hydro-SMILE considers spatially-refined catchment features at high temporal resolutions?

Reply: The authors rephrased this paragraph for better clarity.

3) Rephrase “Thus, this study focuses on the major Bavarian river basins (upper Danube, Main, Inn) with all their tributaries”. It is unclear to me what your study area has to do with the above two statements.

Reply: The authors rephrased this sentence to better explain that a high spatio-temporal resolution is beneficial and necessary for the heterogeneity within the study area.

Line 84: Remove “Therefore”.

Reply: The authors removed ‘Therefore’ from the manuscript.

The paragraph that starts from Line 84: add section numbers throughout the paragraph.

Reply: The authors added section numbers as recommended.

Line 85: Confusing sentence. Remove “...to meet the requirements for the hydrological modeling.”

Reply: The authors removed this part of the sentence as recommended.

Line 86: “...hydro-SMILE along...” should be “...along with...”.

Reply: The authors added ‘with’ to this sentence as recommended.

Line 95: Remove “As a result”

Reply: The authors removed this phrase.

Line 102: “...(up to 1100 mm precipitation sums in the north, 2500 mm in the south; an average temperature of 10 °C in the north, down to 5 °C (-8 °C on alpine summits)...”. Are the authors referring to annual total precipitation and annual mean temperature? Be clear on that. Also, be sure to mention the data sources for these numbers - are they from Poschlod et al. (2020) as well?

Reply: The authors rephrased this sentence to include the actual meaning of the values and added the reference ‘Poschlod et al., (2020)’ to the values within the brackets to better indicate their source.

Line 108: “The major river catchments were divided into a total of 98 smaller sub-catchments based on a common interest in flood protection and a more detailed variation in catchment characteristics, using a selection of gauges (Willkofer et al., 2020).” Rephrase this sentence. It is unclear to me whether the 98 sub-catchments were divided based on the spatial distribution of the 98 selected gauges or whether the gauges were selected because of the division of the 98 sub-catchments.

Reply: The authors rephrased the sentence to clarify the reason for selecting the 98 catchments.

Line 125: Figure caption of Figure 2: Also introduce what are SDCLIREF and WaSiM in the Figure caption.

Reply: The authors added the explanation for SDCLIREF and WaSiM to the Figure caption.

Line 142: What is “T63”?

Reply: T63 is a term describing the original grid resolution of the climate model. As it does not contribute to further understanding of the data, the authors removed this term.

Line 149: “Furthermore, the individual members of the CRCM5-LE are considered independent for the hydrological evaluation period from 1981 to 2099, as the analysis of variations in temperature and precipitation over land and ocean shows (Leduc et al., 2019).” This is the first time the authors mention “hydrological evaluation period”. This is a confusing sentence. Aren’t the CRCM5-LE individual members independent no matter what time period?

Reply: The authors combined this comment and the comment in line 214 and added the recommendation provided in the comment in line 214 here to clarify on the independence of the members of the CRCM5-LE.

Line 152: “...showing regional and seasonal variations in magnitude over Europe (Leduc et al., 2019).” What variables do the authors mean?

Reply: The authors will add the variables (temperature and precipitation) to this sentence in the manuscript.

Line 156: add “match” after “were adjusted to”.

Reply: The authors added ‘match’ to this sentence.

Line 157: Change “RCM scale” to “RCM grid”. Did the authors do the interpolation onto the RCM grid? If yes, be clear on what interpolation scheme is used.

Reply: The authors changed “RCM scale” to “RCM grid” and added the interpolation approach to the supplements (S3).

Line 160: “...3-hourly correction factors for every quantile and month”. Unclear how the 3-hourly correction factor is applied.

Reply: The authors added a paragraph to the supplements explaining the application of the 3-hourly correction factors (S3).

Sentence starting on Line 162: I think the authors want to stress that bias correction is inevitable. Rephrase it to “Despite the benefits (increasing reliability of climate change projections of the hydrological impact model, reducing bias in mean annual discharge) and shortcomings (disrupting feedbacks between fluxes, modification of change signals, assumption of a stationary bias) of bias correction are highly debatable (e.g., Teutschbein and Seibert, 2012; Maraun, 2016; Ehret et al., 2012; Dettinger et al., 2004; Chen et al., 2021; Huang et al., 2014), bias correction is often inevitable for climate change impact studies (Gampe et al., 2019).”

Reply: The authors rephrased this sentence accordingly.

Line 168: For such a topographically complex region as described in the “Study area” section, I’m concerned the statistical downscaling between grids that are so different (from 12 km in RCMs to 500 m in hydrologic models) will lose important spatial heterogeneity across the domain. Does the mass-preserving approach address this problem?

Reply: The mass-preserving approach ensures that through the downscaling no additional precipitation is added or lost. Meaning, when the downscaled result is upscaled to the RCM grid the mass of the RCM grid is matched. The downscaling basically tries to distribute the coarse RCM information to a higher sub-grid scale. We don’t think that the downscaling will alter the spatial heterogeneity. If so, then rather the bias correction will alter the spatial heterogeneity. As with all interpolation schemes, the obtained spatial result is only one of many possible spatial distributions.

Line 171: “The interpolation result was then applied to the SDCLIREF reference fields (Brunner et al., 2021b)”. Unclear. Do the authors mean the SDCLIREF reference fields are also interpolated using the same method?

Reply: The authors added further information about the method to the manuscript.

Line 194: “minimizing a weighted combination of performance metrics, including Nash and Sutcliffe efficiency (NSE; Nash 195 and Sutcliffe, 1970), Kling-Gupta efficiency (KGE; Gupta et al., 2009), the logarithmic NSE and the ratio of root mean squared error to standard deviation (RSR; Moriasi et al. (2007)) (Eq. (1))”. Introducing the overall metric (OM) equation first and then describe what is in the equation. Then, give a threshold - what OM value is considered “good” or “bad”?

Reply: The authors introduced the overall metric first and then described it. We further provided an optimal value for the overall metric and clarified the meaning of deviations from that optimum.

Line 203: “(NSE: 16; KGE: 5)” Unclear. What does this mean?

Reply: The authors rephrased this sentence to better clarify the meaning of these numbers.

Line 208: “Consequently, the level of trust (LOT) for peak flows of return periods of 5, 10, and 20 years of flood events, introduced in Willkofer et al. (2020) showed a moderate to high confidence for most catchments, with gauges of poor simulated performance yielding a

lower LOT with increasing return levels.” Do the authors mean gauges with good performance have higher LOT for peak flows with return periods of 5, 10, and 20 years, whereas gauges with poor performance have lower LOT, especially for peak flows at longer return periods?

Reply: The authors rephrased this sentence to clarify the meaning of the LOT and added a description to the supplemental material as well.

Line 214: “The entire modeling period is shortened by ten years to account for the time span it takes the RCM to produce fully independent realizations due to the inertia of the ocean model (Leduc et al., 2019).” I have two comments: 1) I saw that the authors partially address my question for Line 149 here. It would be better to rearrange this part and the sentence on Line 149 such that the 10-year spin-up period and the choice of the evaluation time period are more clearly lined up and explained. 2) Rephrase this sentence to “We focus on 1961—2099 as opposed to 1950 – 2099 to account for the time it takes for the RCM to produce fully independent realizations due to the inertia of the ocean model (Leduc et al., 2019).”

Reply: As mentioned in the reply to comment of line 149, the authors rephrased the sentence and moved it from line 214 to 149 to introduce the independence of the different members earlier in the manuscript.

Figure 3: What is “HF T,BM” on the far right?

Reply: The authors added the explanation for this term within the figure caption of the manuscript.

Line 256 and thereafter: Change “intensity” to magnitude throughout the manuscript.

Reply: The authors changed the term throughout the manuscript.

Line 293: “...as indicated by the spread of the blue markers around the black benchmark line.” I think the authors mean “... as indicated by the decreasing spread of the blue markers around the benchmark line with increasing sample size.”

Reply: The authors changed this phrase accordingly.

Lines 296-297: What are panels a, c, and e?

Reply: This is a clear mistake in referencing the different parts of the Figure. The authors changed the references to Fig. 4a, 4b, and 4c.

Figures 1&6: What do “balanced” and “unbalanced” pluvial mean?

Reply: The authors added a short explanation of these terms to the respective location within the manuscript.