Author’s response to the comments of reviewer 1:

In this study, the authors assess the impacts of climate change on the 100-year flood events in several rivers in Bavaria region, in which most of the flood protection infrastructures are designed based on the 100-year flood events. Specifically, the authors explore a large ensemble of 50 members of CanESM2 climate model, which are produced using different initial conditions but the same model structure, parameters, and a high emission scenario (RCP8.5). A hydrological model is then used to simulate river discharge, followed by a dataset of 1500 model years (30 years of the reference period x 50 members), which are used for extreme value analysis. Also, the authors show the benefit of using hydro-SMILE in providing a more robust extreme hydrological discharge values under climate change impacts. In overall the study is beneficial for the study region. However, there are still some concerns, which require further improvements. My detailed comments are below.

AC: The authors would like to thank the reviewer for the valuable comments to further improve the manuscript. Detailed author comments (AC) on the reviewer comments (RC) are provided for each individual comment.

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.

AC: The authors agree with the comment and will rephrase the mentioned sentences to better point out the choice of this study area.

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

AC: The CanESM2 large ensemble was chosen because of its availability of high temporal resolution outputs necessary for driving the CRCM5. Other CMIP5-era large ensembles have either not archived the necessary fields needed for dynamical downscaling, or it was too difficult to retrieve the data. Secondly, there was basically no choice of different RCP scenarios, since we were dependent on the driving GCM data. The CMIP5-era large ensembles were all ensembles of opportunity where mainly only one scenario was possible to run. Hence, this era of large ensembles has almost exclusively produced RCP8.5 simulations. In our project we also only had the opportunity to downscale one large ensemble.

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

AC: The authors agree that the study area offers comparably very good in situ data. Yet, these data are mainly available as point measurement or spatially interpolated which means that there are inherent uncertainties. Hence, the authors consider a model calibration as necessary for the selected model. Furthermore, as the model is physically based, thus, not considered entirely physical, it allows for many free parameters to be set in the various modules which not all of them are measurable. The holistic modelling approach chosen for the hydrological model which provides a generalized parameter set for the entire region as described in Willkofer et al. 2020, aimed towards a reduction of these free parameters, thus reducing parameters to be calibrated. Furthermore, soil parameters were not calibrated (except for karstic areas where it was necessary to simulate faster infiltration). However, this approach had direct impact on the model performance as catchment specific characteristics were largely disregarded. The authors can briefly elaborate on this comment within the manuscript.

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.

AC: Maps of the mentioned metrics are already published in several publications (see: Poschlod et al. 2020, Willkofer et al. 2020). Detailed discussions on the matter are also presented in Willkofer et al. 2020 and Brunner et al. 2021. The authors therefore would refer the reader to these references for further reading on the model performance. We will briefly elaborate on the analysis within the manuscript without going into exhaustive details. In addition, we can provide maps of the performance metrics within the supplement material. Appropriate references to these figures will be placed in the main manuscript. In general, the gauges of the hydrological Bavaria are in parts heavily impacted by transfer systems, hydropower dams, flood protection measures, channels, as well as by local natural characteristics (e.g., Karst, shallow catchments with little variance in mean monthly discharge) which affects model performance due to the holistic approach.

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

AC: the authors agree and will change the scale of the figures to display the entire variation ranges.

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.

AC: The authors agree and can provide maps of the spatial variation of the change in return period. The maps will simply display the values of the presented violin plots of the changes in frequency and intensity. However, the authors consider placing these maps within in the supplement material to be able to display further extensive analysis on the dynamics in change drivers as suggested by the other reviewer.

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.

AC: Unfortunately, we the authors recede from suggesting designs for flood protection infrastructures since we lack expertise in engineering. However, we aim to provide the means for engineers to use the presented results for the design of such structures by stating that CC impacts should be considered in their design. The authors will add the following to the conclusions to make it clearer that this study does not cover designs but rather provides data to build upon for the design: “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)?

AC: The authors will change the mentioned values