The authors are grateful to Reviewer #1 for sharing his concerns with the authors of this submission. Your concerns were carefully taken into consideration for the manuscript revision, leading to a more consistent and scientifically sound manuscript. Major changes and additions were performed in the entire manuscript. In the pages below, we elaborate with the major and minor concerns that led you to the decision for major revision. A point by point reply is given below.

1. Does the paper address relevant scientific questions within the scope of HESS? YES, it explores the response of European river basins to climate change

2. Does the paper present novel concepts, ideas, tools, or data? YES, although projections for hydrological impacts of climate change in itself are not new. Projections with this particular model JULES and based on this set of CORDEX simulations are new. It uses comprehensively bias corrected data following new methods (itself described elsewhere). The focus on low flows and droughts as presented here, is also relatively unexplored.

3. Are substantial conclusions reached? YES

4. Are the scientific methods and assumptions valid and clearly outlined? YES

5. Are the results sufficient to support the interpretations and conclusions? YES minor: twice (p7281 L1 and p7292 L1) a statement is made on floods which to my opinion cannot be derived from the present analysis. I suggest to simply omit these.

REPLY: The flood related unsupported statements were removed from the manuscript.

6. Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)? YES

7. Do the authors give proper credit to related work and clearly indicate their own new/original contribution? YES, extensively

8. Does the title clearly reflect the contents of the paper? The title mentions water stress which is a function of both availability and demand. This is not directly analyzed in the paper. Also hydrological model biases are not analysed, only the effects of forcing biases. I suggest to change the title to something like: "High-end climate change impacts on European runoff and low flows: exploring the effects of forcing biases"

REPLY: Following the reviewer's indication, the title was changed to better correspond with the topic of the manuscript. The new title is "High-end climate change impact on European runoff and low flows. Exploring the effects of forcing biases".

9. Does the abstract provide a concise and complete summary? YES

10. Is the overall presentation well-structured and clear? Overall, the paper is well structured and clear. However, the introduction is too long and its structure is not always clear: â˘A ´c p 7268-7269 (para 1 and 2) are OK â˘A ´c para 3, 5 and 7 are bit long but generally OK â˘A ´c p7270-7271 (para 4): the discussion on added stresses of population growth and human activities is not relevant in present context; a single statement reflecting its significance here or in the discussions section (4.1) suffices â˘A ´c para 9, p7273, is superfluous after para 8 â˘A ´c p7273-7274 (para 10): the discussion of GHMs/LSMs is not too relevant in the present context â˘A ´c p7274-7275 (para 11): the JULES discussion can be omitted here and partially merged with section 2.2 Section 2 is OK except that I would put the present 2.5 – Bias correction directly following the present 2.1 – climate/forcing data. Section 3, 4 and 5 are OK

REPLY: The reviewer's suggestions were carefully taken into account into the revised version of the Introduction. The Introduction now is shorter and more focused on the topics of this study as redundant information have been removed. Specific changes made per paragraph are:

- Paragraph 4 was deleted as its content (added stresses of population growth and human activities on climate change impact) is not relevant with this study. A reference on the significance of this topic is made in Paragraph 1 of the Introduction.
- Paragraph 5 on multi-model assessments was significantly shortened and moved to the end of Paragraph 3
- Paragraph 9 was removed from the manuscript as it was repeating the information of Paragraph 8.
- As the content of Paragraph 10 (discussion on other LSMs) was not very relevant to the context of this study, only a short definition of the LSMs has been kept of this paragraph. This is now merged into a paragraph where the JULES model is briefly mentioned.
- The discussion about JULES in Paragraph 11 was removed from the Introduction section and was moved to the Data and Methods section (merged with Section 2.2.)
- The research objectives in the last paragraph of the Introduction have been reformulated.

In Section 2 we followed the reviewer's indication of moving Section 2.5 (Bias Correction) right after Section 2.1.

11. Is the language fluent and precise? Yes, the quality of english is generally high and precise, with a few minor exceptions

12. Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?

13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?

I miss a brief description of the forcing data: how large are biases? While the five models are similar in projected temperature change (due to the time slice strategy, centering on ïA¸Dˇ T =2 and 4K resp.) there is no indication of how their precipitation changes. What is the CV over the 5 models here before/after bias correction? Can we have 5 maps with ïA¸Dˇ P? Could be part of section 2.1 or a new starting subsection in 3 or an expansion of 3.6. For other variables biases could be presented in supplementary material.

REPLY: Following the reviewer's indication, two Figures describing the effect of bias correction on the forcing variables, have been added to the Electronic Supplementary Material. Figure S1 shows the effect of bias correcting against the WFDEI dataset and Figure S2 against the E-OBS dataset. Results are shown for precipitation and temperature as the rest of the forcing variables were not bias adjusted. The absolute differences between bias corrected and raw input (bc-raw) are shown for all the participating GCMs and for their ensemble mean. The cv between the ensemble members before and after bias correction has also been calculated. In each sub-figure, the spatial average of each illustrated map is noted in each sub-figure.

These were used to also understand the differences between the two observational datasets. A relative comment deduced from these figures has been added to Section 3.6.: "From Figures S1 and S2 of the ESM (showing the effect of bias correction on the forcing variables of precipitation and temperature) it can be deduced that that E-OBS corrected precipitation has lower values than precipitation adjusted against the WFDEI dataset. This explains the lower runoff produced by the E-OBS bias adjusted dataset, as it is reasonable for the differences in precipitation to reflect on the output of the hydrological model. "

The two figures are shown below:

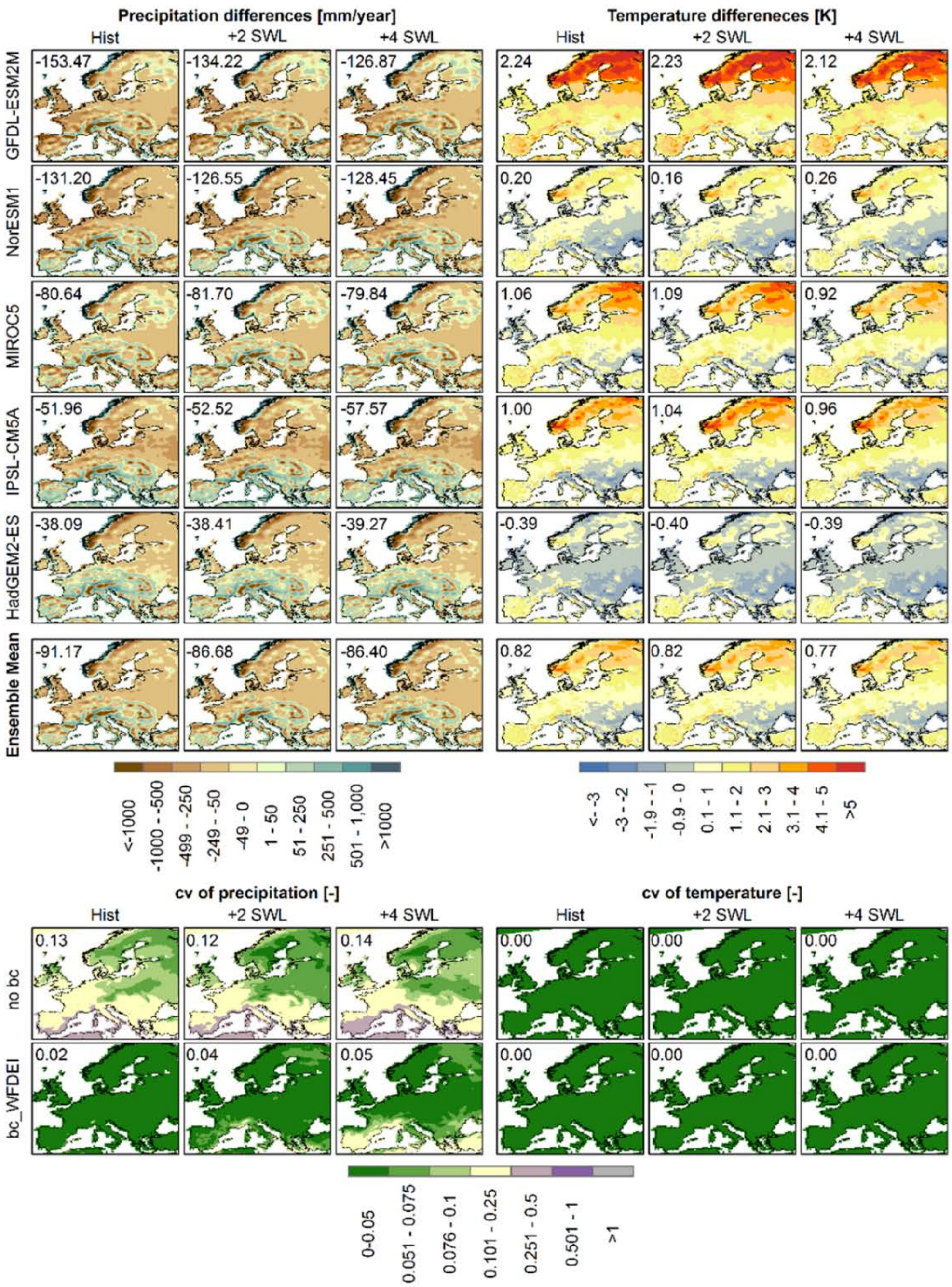

Figure S1. Absolute differences between Euro-CORDEX data bias adjusted against the WFDEI dataset and raw Euro-CORDEX data, for the variables of precipitation (right block) and temperature (left block). Differences are calculated from the historical (1976-2005), +2 SWL and +4 SWL time-slice averages, for all dynamical downscaled GCMs and their ensemble mean. Bottom block: Coefficient of variation between the ensemble members, for raw and bias corrected against the WFDEI dataset precipitation and temperature forcing variables, for the historical, +2 SWL and +4 SWL time-slices. The average value for the pan-European area is shown in each sub-figure.

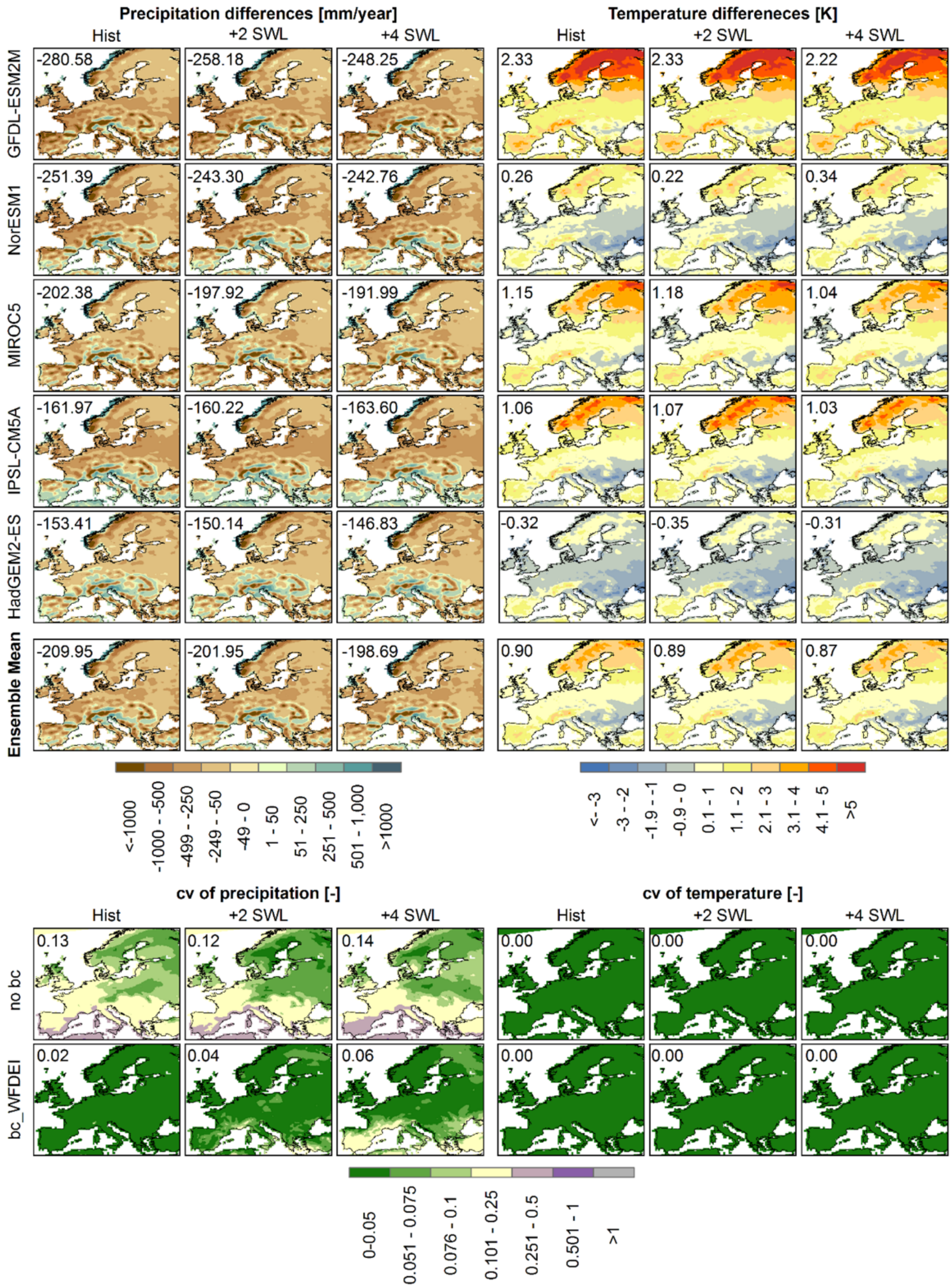

Figure S2. Absolute differences between Euro-CORDEX data bias adjusted against the E-OBS dataset and raw Euro-CORDEX data, for the variables of precipitation (right block) and temperature (left block). Differences are calculated from the historical (1976-2005), +2 SWL and +4 SWL time-slice averages, for all dynamical downscaled GCMs and their ensemble mean. Bottom block: Coefficient of variation between the ensemble members, for raw and bias corrected against the E-OBS dataset precipitation and temperature forcing variables, for the historical, +2 SWL and +4 SWL time-slices. The average value for the pan-European area is shown in each sub-figure.

In section 2.2 I miss a paragraph on the hydrological performance of JULES over Europe from previous studies. How well does it perform wrt discharge (average, high and especially low flows)? And then in the discussion 4.1 what does that imply for the results of the present paper?

REPLY: A piece on the hydrological performance of JULES over Europe has been added to the description of the model section in Section 2 (Data & Methods).
"Other studies give insight into the hydrological performance of JULES specifically. Blyth et al. (2011) extensively evaluated the JULES model for its ability to capture observed fluxes of water and carbon. Concerning discharge, their findings suggest that for the European region seasonality is captured well by the model. For temperate regions (like most of central Europe) to model exhibited a tendency towards underestimating river flows due to overestimation of evapotranspiration. Prudhomme et al. (2011) assessed JULES' ability in simulating past hydrological events over Europe. In general terms the model was found to capture the timing of major drought events and periods with no large-scale droughts present were also well reproduced. The model showed a positive drought duration bias, more profoundly present in northwest Spain and East Germany-Czech Repuplic. Prudhomme et al. (2011) argue that this feature is related to overestimation of evaporation by the model. For regions where droughts tend to last longer, JULES exhibited a better ability of reproducing the drought events' characteristics. Gudmundsson et al. (2012) compared nine large scale hydrological models, and their ensemble mean, based on their skill in simulating the interannual variability of observed runoff percentiles in Europe. According to the overall performance (accounting for all examined percentiles and evaluation metrics), JULES was ranked third best out of the 10 models, after the multi-model ensemble mean and the GWAVA model. For low and moderately low flows, expressed as 5th and 25th percentile respectively, JULES is also in the top three models regarding the representation of interannual variability in runoff. In the study of Gudmundsson et al. (2012b), where an ensemble of hydrological models is evaluated for their ability to capture seasonal runoff climatology in three different hydroclimatic regime classes in Europe, JULES exhibits a good performance, comparable to that of the best performing multi-model ensemble mean. In other studies employing multi-model ensembles, focusing on the whole European region (Gudmundsson and Seneviratne, 2015) or a single basin in Europe (Harding et al., 2014; Weedon et al., 2015) JULES' simulations also correspond with these of the other models."

There is redundancy between fig 2 and 4, and 3 and 5 respectively. Can be reduced in discussion with the technical editor perhaps? E.g., adding perhaps one column each in figures 2 and 3 with the CV of absolute change only.

REPLY: Following the reviewer's indication Figures 2 -5 have been restructured. Figure 2 was merged with Figure 4 and Figure 3 with Figure 5. Moreover, the CV for the projected period was substituted with the CV of the absolute differences. Finally, a sub-figure showing model agreement towards a wetter change in the projected time-slice has been added to the two new Figures. The new version of the Figures is shown below.

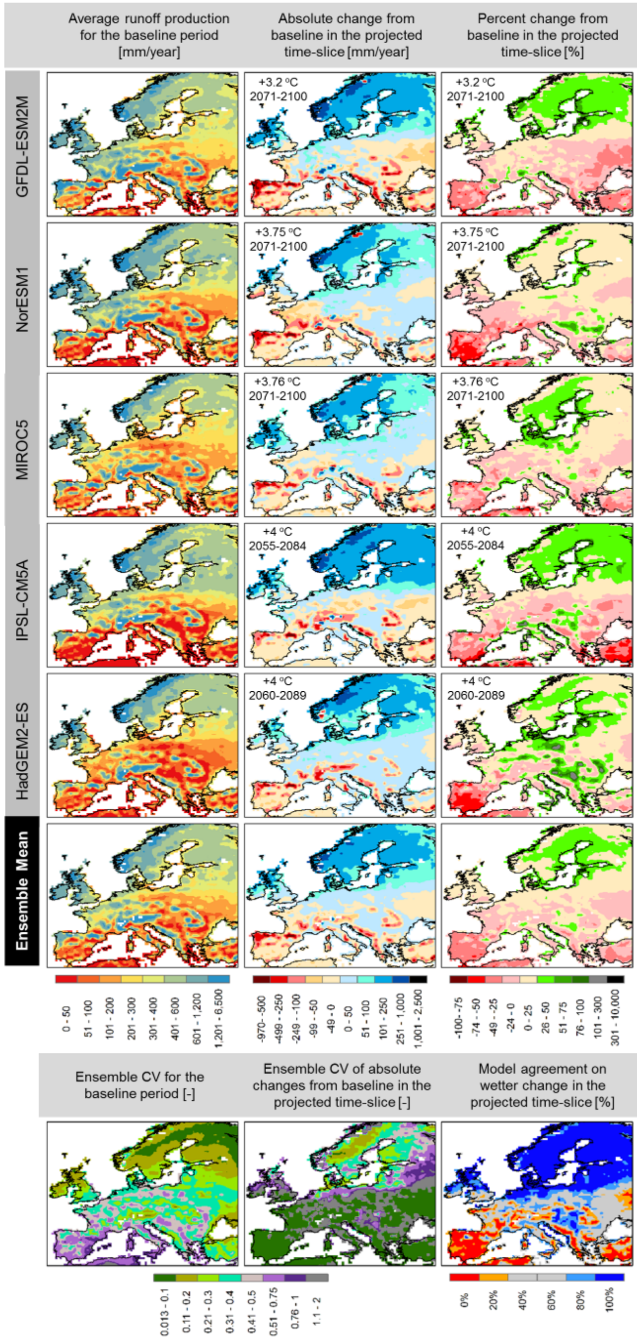

Figure 1. Average runoff production from raw Euro-CORDEX data for all dynamical downscaled GCMs and their ensemble mean. Runoff production averaged over the baseline period (1976-2005) (left column), absolute change in runoff in the +4 SWL projected time-slice (middle column) and percent change in the +4 SWL projected time-slice (right column). Bottom row: coefficient of variation of the ensemble members for the baseline period (left column), coefficient of variation of the projected absolute changes in the +4SWL projected time-slice (middle column) and model agreement towards a wetter change in the +4 SWL projected time-slice.

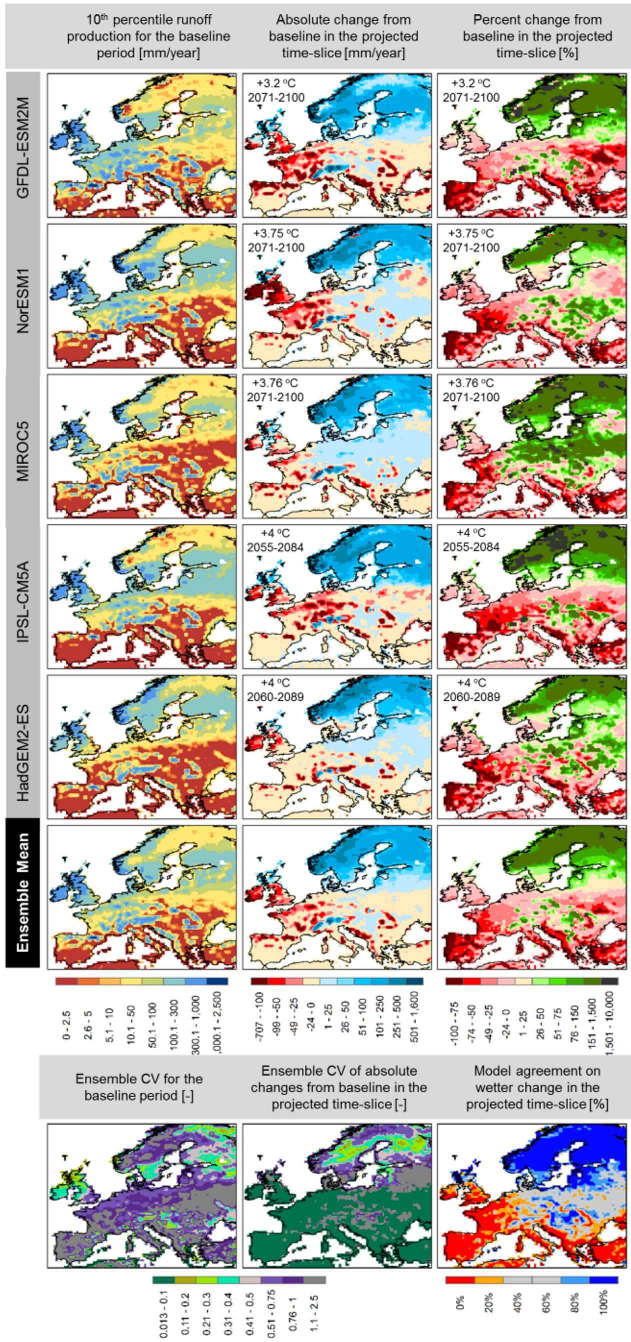

Figure 2. 10th percentile of runoff production from raw Euro-CORDEX data for all dynamical downscaled GCMs and their ensemble mean. 10th percentile runoff production derived on an annual basis and averaged over the baseline period (1976–2005), absolute change in 10th percentile runoff in the +4 SWL projected time-slice (middle column) and percent change in the +4 SWL projected time-slice (right column). Bottom row: coefficient of variation of the ensemble members for the baseline period (left column), coefficient of variation of the projected absolute changes in the +4SWL projected time-slice (middle column) and model agreement towards a wetter change in the +4 SWL projected time-slice.

Figure 9 second block is wrong: should be Guadiana and Elbe instead of a repetition of Rhine and Danube

REPLY: This mistake has been eliminated in this revised version of the manuscript.

14. Are the number and quality of references appropriate? YES, though the number of refs is on the high side

15. Is the amount and quality of supplementary material appropriate? Since the paper has a (perhaps secondary) focus on the effect of bias correction and even the use of different reference sets in these, I would like to see more information on initial biases of the 5 models with respect to the 2 ref sets. For precip in the paper itself, for the other variables in the supplementary material. Also the change signal for the forcing data, at least for precip, should be presented, e.g. in maps

REPLY: The added Figures S1 and S2 tackle the issue of the differences between the observational datasets and initial biases. Apart from these two figures, more additions have been made to the ESM to support our findings.

In the supplement pdf the reviewer asks: "why span ECS range if you only look at 2 and 4 degree warming periods?"

REPLY: The following was added to Section 2.1 (description of climate data):

"Using the SWL concept constitutes the results independent of the timing that the warming occurs. Although by definition of the SWL, the models reach the same level of warming in their time-slices, the different model sensitivity reflects on the evolution of temperature in the time-slice, as more sensitive models are expected to have higher rates of changes in the period before and after a specific SWL is achieved compared to the less sensitive models. Moreover, considering models of different ECS is important to express the range of other than temperature forcing variables produced by the GCMs (eg. radiation). "