Replies to Referee #3

Does the weighting of climate simulations result in a more reasonable quantification of hydrological impacts?

Hui-Min Wang, Jie Chen, Chong-Yu Xu, Hua Chen, Shenglian Guo, Ping Xie, Xiangquan Li

We sincerely appreciate the referee's comments and suggestions on the manuscript. All suggestions are helpful to improve this manuscript. We have carefully studied, considered and responded to all comments point-by-point. For clarity, all comments are given in *italics* and responses are given in plain text. The manuscript will be modified accordingly.

> *This study applies different combinations of bias-correction (BC) and model weighting (MW) to post-process climate and hydrological projections in two catchments. Both BC and MW are receiving sustained attention in the community, and so far only few studies combine both. What is important to stress, is that although the underpinnings of these two approaches are quite different, their aim is arguably quite similar: close the gap between simulations and observations. This leads me to comment on the two main findings of the study:*

We would like to thank the referee for the time taken in reviewing this manuscript. All comments have been replied to below and will be addressed in the revision.

> *Finding 1: "when using raw GCM outputs with no bias correction, streamflow-based weights better represent the mean hydrograph and reduce the bias of annual streamflow" P1L19-20: in my view, this is a natural consequence of applying MW, and in a way, it means that MW is used to correct for/mitigate climate model biases.*

Thanks for the comment. We agree with the referee that MW is used to mitigate biases, but this is not the specific focus of this study and we failed to state the conclusion clear enough. Actually, in this sentence, we intended to emphasize the advantages of streamflow-based weights over the weights calculated using climate variables (i.e. temperature and precipitation in this study). As stated in P12, L9-14, when dealing with the raw GCM-simulated streamflows, biases in multi-model mean of annual streamflow are reduced more by the weights based on the impact variable (streamflow), comparing with the weights based on climate variables. Herein, we will modify the expression of Finding 1 to make this point clearer.

> *Finding 2: "when applying bias correction to GCM simulations before driving the hydrological model, the climate simulations become rather close to the observed climate, so that compared to equal weighting, the streamflow-based weights do not bring significant differences in the multi-model ensemble mean" P1L21-23: my interpretation is that employing*

successively two techniques with the same purpose makes the second technique redundant. Reducing the biases in the climate simulations, and then applying MW, makes it extremely difficult for the MW to discriminate between good and poor models. I recognise that BC is applied to the climate simulations and MW to the hydrological simulations, but since all the climate simulations are run through the same hydrological model, calibrated presumably with the forcing dataset also used to perform the BC, the differences in the streamflow simulations are minimal (as shown in Figure 3c and especially 4c). This lack of differences explains why the different weighting methods lead to similar results under current climate (the simulations are almost the same, so how they are combined makes little difference).

We agree with the referee that in this study, MW loses the ability to discriminate the performances of climate simulations after the bias correction. This is also a finding of this study, which was mentioned in P12, L18-20. We will modify this sentence to make this point clearer.

In fact, MW is not designed for dealing with hydrological simulations but a necessary process to handle the ensemble of multiple climate simulations. Even after bias correction, there still exist some differences between climate simulations. In order to obtain evaluation of climate change impacts, it is unavoidable to choose a MW method to synthesize the simulation results from the ensemble (whether or not bias correction is done). Thus, MW is an indispensable process. Actually, both BC and MW are common procedures in regional impact studies. Although it is common to use equal weighting for bias-corrected ensembles, whether unequal weighting is the best choice remains to be investigated (Alder and Hostetler, 2019). The results of this study show that when the bias correction is done in impact studies, unequal weighting does not bring much difference to the impact evaluation. This supports the usage of equal weighting for bias-corrected ensembles so far. Nonetheless, we still think that with further development of weighting methods (e.g., more aggressive or multi-objective weighting methods), unequal weighting maybe helps to bring different or more reasonable consequences. The discussion on the weighting methods for the bias-corrected ensembles will be added in the revised manuscript.

Overall, I suggest shifting the focus from current climatic conditions (for which no climate model and hence MW or BC is necessary) to future conditions (which rely on climate model simulations, which may need BC/MW). In my view, the focus is currently too much on the current conditions. For instance, in the abstract, the authors write "when applying bias correction to GCM simulations before driving the hydrological model, the climate simulations become rather close to the observed climate". This is true because of the nature of bias-correction, and was shown in previous studies (e.g., Hakala et al., 2018). What the grey area in Figures 3d and 4d tells us, however, is that under future conditions, there is substantial spread among the hydrological simulations, although the driving GCM simulations have been bias-corrected (likely because of the different sensitivities of the climate models).

Thanks for the comment. We agree with the referee that more attention should be paid to the future projections. In this version of manuscript, future simulations are only evaluated in the form of uncertainty (Section 4.4), since there is no observation in the future period to be compared with. In order to partly overcome this problem, we have added the out-of-sample testing for this study following the suggestion of referee #2. In out-of-sample testing, the output of one climate model was regarded as the "truth" and the outputs of the remaining 28 climate models were used as simulations to this "truth" model. Then the weights were re-calculated for the remaining models. Since there is a "truth" result for the future period in this case, the performances of weighting methods in reproducing the future "truth" can be evaluated. In the out-of-sample testing, each climate model was regarded as truth in turn. In general, the results of out-of-sample testing are similar to the results using historical observations, which supports the conclusion of this study. The detailed results and analyses of out-of-sample testing will be added and discussed in the revised manuscript.

In addition, it is true that the differences between ensemble members have been greatly reduced during the reference period while there are still considerable differences in the future period (which had been stated in P9, L15-19). This may be because the bias of climate models is nonstationary (Hui et al., 2018). However, the sentence in the abstract is only an explanation to the results of Finding 2 instead of a focus of this study. But we failed to state this logic clear enough. Therefore, this sentence will be modified to make the focus of this study clearer.

> *Is there any way to apply MW based on these projected changes, and not based on the streamflow simulations under current climate? In other words, are some of these projections more reliable than others and/or are some projections interdependent, and should be downweighted?*

We thank the referee for this suggestion. Actually, the REA method in this study concludes projected values when assigning weights. The REA considers both smaller differences to the observation in the reference period and more concentrated projections in the future period. Although the weights calculated by the REA method are most differentiated for the bias-corrected ensemble (as Fig. 2 shows), they still bring little impacts on the final results of multi-model means. In addition, the PI method considers independency between climate simulations when determining weights, but it only relies on reference values which have been tuned by the bias-correction methods. The ability of independent criterion may fail because of the bias correction. Therefore, in the case of bias-corrected ensembles, some modifications may be needed for these MW methods to include future values. This point will be further discussed in the revised manuscript.

> *In summary, my impression is that Finding 1 is relevant but quite foreseeable. I think that Finding 2 is to a great extent due to the experimental design, in particular to the decision to apply BC and MW successively. I encourage the authors to rethink how to best combine MW*

and BC, for instance by using different periods and/or criteria for the MW.

We appreciate the comments from the referee. As presented in the last response, the out-of-sample testing will be added in the discussion as a complement. In addition, we will better state that the main focus of this study is to investigate the influences of MW methods on the evaluation of climate change impacts (when the bias correction is or is not done), and to study whether the weighting determined based on the impact variable (streamflow) can induce more reasonable results. This investigation is necessary because MW is a procedure to generalize the results of ensembles and the best way to do it remains questionable. This explanation to the usage of MW and BC will be added in the Discussion section.

References

Alder, J. R., and Hostetler, S. W.: The Dependence of Hydroclimate Projections in Snow‐Dominated Regions of the Western United States on the Choice of Statistically Downscaled Climate Data, Water Resources Research, 55, 2279-2300, https://doi.org/10.1029/2018wr023458, 2019.

Hui, Y., Chen, J., Xu, C. Y., Xiong, L., and Chen, H.: Bias nonstationarity of global climate model outputs: The role of internal climate variability and climate model sensitivity, International Journal of Climatology, 39, 2278-2294, https://doi.org/10.1002/joc.5950, 2018.