Replies to Referee #2

Transferability of climate simulation uncertainty to hydrological climate change impacts

Hui-Min Wang, Jie Chen, Alex J. Cannon, Chong-Yu Xu, Hua Chen

We would like to thank the reviewer for the time taken in reviewing this paper. All comments will be incorporated into the revised manuscript. Please find the point-by-point responses below. For clarity, comments are given in italics, and our responses are given in plain text. We will make the revisions to the manuscript as suggested.

The authors do a good job in their attempt to shed light on the important problem that impact modelers face in efficiently and effectively capturing the range of uncertainty in climate model simulations. Furthermore, they investigate whether covering this range in climate model output variables translates to capturing the uncertainty range of hydrological variables. The paper is well written and clearly presented. Though, in the end, I was not convinced that impact modelers can actually save much time and effort using this methodology. I would recommend that the manuscript needs minor revisions. Importantly, the authors need to make it clearer how an end user can avoid downloading all 50 simulations in order to prove which subset of 10 are most appropriate to cover the uncertainty range in their study.

Thanks for your positive evaluation in general and for your professional comments. Please find our responses in next page.

I would begin by asking this. What do end users or impact modelers gain by this paper? You have shown that for your two different watersheds, a subset of approximately 10 model simulations are needed to reasonably capture the spread of the model uncertainty for both climate and hydrological variables. Additionally, you point out that not using the temperature variables to obtain the subset affects the hydrology of the two watersheds differently. As a result, you illustrate that the selection of the 10 climate models is unique to each impact assessment study. Furthermore, you needed all 50 simulations to test which subset was optimal for your two cases. I do not see how an impact modeler would not need to repeat precisely what you have done. In order to replicate your method, but specific to their study interest or area, they would need to "extract, store, and compute" (page 2, line 14) all 50 model simulations themselves. Then, couldn't they just as easily implement the entire set of simulations instead of a smaller subset? To ask it more directly: How can an impact modeler know which 10 model simulations to use, for their unique case, without testing the ensemble ranges of each possible subset with respect to the entire set of simulations? And to do this, would they not need to run *all 50 model simulations?*

Sorry for the lack of clarity in the manuscript. Depending on the choice of climate variables and climate model ensemble, it may not be necessary to "extract, store, and compute" climate indices from all climate model simulations in the ensemble of interest. For example, pre-computed ETCCDI climate extreme indices for GCMs participating in CMIP3 and CMIP5 are publically available from http://climate-modelling.canada.ca/climatemodeldata/climdex/. In addition, the end-user may be able to refer the results of this study for watersheds with similar climate and hydrological characteristics, although it is likely that some level of site and study-specific analysis will be required.

However, the main objective of this study is to investigate the transferability of climate simulation uncertainty to hydrological impacts. If the climate simulation uncertainty is transferable in the hydrological impacts, the selected 10 climate simulations can be directly used to drive a hydrological model for impacts studies instead of using all climate simulations. This is crucial for hydrological modelers as they usually spend a lot of computational costs in running a large number of climate simulations with a complicated hydrological model (e.g. SWAT). The conclusion of this study shows that the climate simulation uncertainty is transferable in the envelope-based selection based on multiple climate variables, and the subset of around 10 climate simulations can cover the majority of uncertainty. Therefore, end-users can choose the group of climate variables according to their knowledge to the climate and hydrological characteristics of watershed of interest and then select the representative subset of climate simulations to save computational costs in the hydrology world.

All above information will be discussed in the Discussion section of the revised manuscript.

Some more specific comments and questions are as follows:

In section "2.2.1 Climate Simulations": Does it make sense to lump the uncertainty ranges of both RCP4.5 and RCP8.5? These are two different concentration pathways that represent very different conditions. It is true that we currently can't know which is more likely. I would recommend either treating each pathway independently with different ranges of uncertainty, or I would recommend also including simulations from pathways RCP2.6 and RCP6.

Thanks for your comment. RCP4.5 is the medium stabilization scenario and RCP8.5 represents the very high radiative forcing scenario. The mitigation scenario, RCP2.6, was not used because recent analyses suggest that this RCP will be very difficult to achieve with current emission trajectories (Arora et al., 2011; Rozenberg et al., 2015). RCP6.0 is a scenario with radiative forcing that is bracketed by RCP4.5 and RCP8.5 and was not simulated by as many modeling centers as RCP4.5 and RCP8.5. Thus, we used RCP4.5 and RCP8.5 to include a range of realistic projections (Lutz et al., 2016). Due to unknown future emission scenarios, two concentration pathways were used in an undifferentiated manner to cover uncertainty resulting from emission scenarios in our study.

We agree with the reviewer that two emission scenarios generate respective climate simulations. It may be more proper to separately use RCPs in the practical applications. Thus, we have also used each pathway separately to be input into the subset selection. Figure R1 shows the example results where only the one scenario (RCP4.5 or RCP8.5) was used in the Xiangjiang watershed (temperature variables were not included in the selection process). The main characters of the results are roughly the same as the original research where 2 RCPs were considered. In this case, the selection of 5 or 6 climate simulations by the KKZ method can cover adequate uncertainty range. Therefore, the specific choice of emission scenarios can be decided by end-users according to their own needs. The results for the Manicouagan 5 watershed will be further explored in the future.

In order to stress this comment, an explanation on the choice of emission scenarios will be added in the Discussion section.

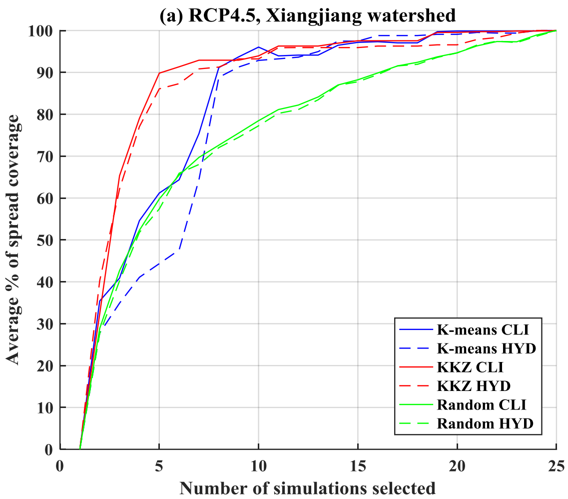

Figure R1: The average PSC for three different selection methods over the Xiangjiang watersheds when only one emission scenario is considered (CLI = climate variables and HYD = hydrological variables)

Page 7, line 17: What was the reason to use 100 quantiles instead of the total number of days in the periods (e.g., 1975-2004 for January = 30 years times 31 days = 930 days or quantiles)?

The use of 100 quantiles is to smooth the distribution of simulated daily precipitation or temperature. The smoothing process eliminates sharp scaling factors that may occur due to outliers, especially for extreme values. On the other hand, Lafon et al. (2013) found that the division of 100 quantiles in the empirical quantile mapping generates more accurate downscaling results than that of 25, 50 or 75 quantiles. The use of 100 quantiles in the Daily Scaling (DS) method is also the same as many other studies (Harrold and Jones, 2003; Mpelasoka and Chiew, 2009; Chen et al., 2013). All of these points will be clarified in the revised manuscript.

Page 9, line 5: I do not anticipate for it to change your results that much, but perhaps it is better to use something like standard deviation as a measure of the uncertainty coverage. The

Percentage of Spread Coverage (PSC) is only sensitive to the range of the minimum and maximum values. You could end up having many of the models grouped close together, and as a result, your measure would overestimate your actual uncertainty coverage.

We agree with the reviewer that the spread of a MME provides an imperfect estimate of uncertainty, and the PSC is sensitivity to the maximum and minimum values. We considered and found that it is improper to use standard deviation, in our case, as the measure of uncertainty in the evaluation on the uncertainty coverage of subsets. To be specific, the selected simulations in impact studies are often considered to be representative of specific uncertainty range instead of individual samples in the calculation of deviation. For example, selected simulations are regarded as the 10th and 90th quantiles in the range of temperature or precipitation in many impact studies (Lutz et al., 2016; Immerzeel et al., 2013; Sorg et al., 2014).

In consideration of reviewer's concern, i.e. to lower the influence from the maximum and minimum values, the average coverage on quantiles (ACQ) of the subsets have been used as an evaluation criterion. The ACQ is calculated using Eq.(R1)

$$\mathrm{ACQ} = \frac{1}{N}\sum_{i=1}^{N}\left(\max_{s} Q_{s,i} - \min_{s} Q_{s,i}\right) \tag{R1}$$

where N is the total number of climate or hydrological variables, and $Q_{s,i}$ represents the rank of quantile of the ith variable for the sth selected simulation (e.g. if the change in a variable of one selected simulation is the 80th quantile of the changes of all simulations, then $Q_{s,i} = 0.8$). ACQ evaluates the range of quantiles covered by selected simulations. Due to the use of quantiles instead of values, the ACQ is less influenced by the maximum or minimum values. Figure R2 presents the ACQ for climate variables and hydrological variables (temperature variables were excluded in the selection for the Xiangjiang watershed, while they were included for the Manicouagan 5 watershed). Compared with average PSC, ACQ results show similar characteristics but less sensitivity, and PSC

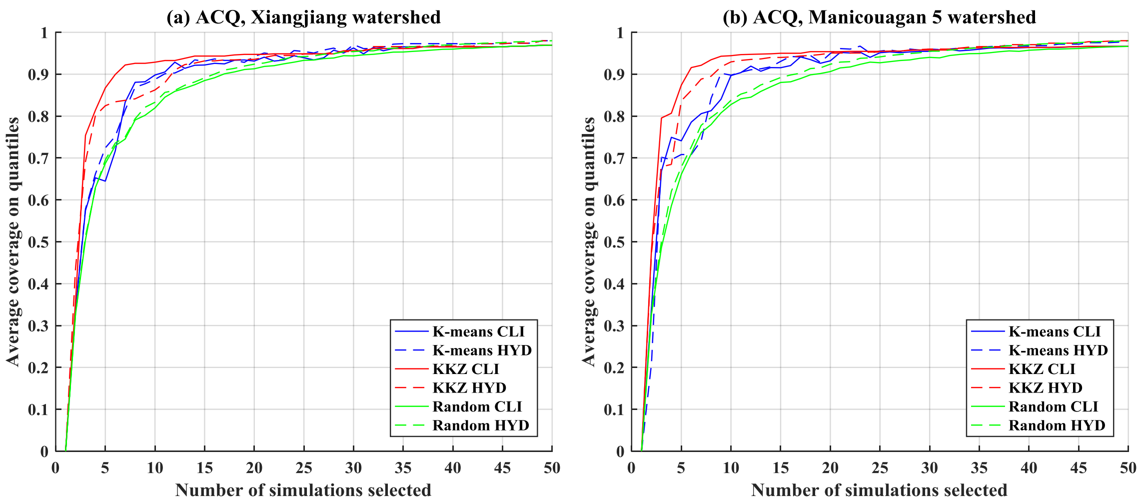

Figure R2: The average coverage on quantiles (ACQ) for three different selection methods over two watersheds (CLI = climate variables and HYD = hydrological variables)

can more intuitively provide an evaluation of the ability of subsets to cover uncertainty. Therefore, PSC will be still used as the evaluation criterion in the revised manuscript.

This will be discussed in the revised manuscript.

> *Figure 2: Are you showing the observed and simulated values for the calibration and validation for 1 year? Or is each day the average of that day across the years (e.g., for Xiangjiang: all January 1 values are averaged over the time period 1975-1987, then January 2 values are averaged over the same years, . . .)?*

The mean hydrographs showed in Fig.2 were calculated as the average of each calendar day across the years. This will be clarified in the revised manuscript.

References

Arora, V. K., Scinocca, J. F., Boer, G. J., Christian, J. R., Denman, K. L., Flato, G. M., Kharin, V. V., Lee, W. G., and Merryfield, W. J.: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases, Geophysical Research Letters, 38, https://doi.org/10.1029/2010GL046270, 2011.

Chen, J., Brissette, F. P., Chaumont, D., and Braun, M.: Performance and uncertainty evaluation of empirical downscaling methods in quantifying the climate change impacts on hydrology over two North American river basins, Journal of Hydrology, 479, 200-214, https://doi.org/10.1016/j.jhydrol.2012.11.062, 2013.

Harrold, T. I., and Jones, R. N.: Generation of rainfall scenarios using daily patterns of change from GCMs, in: Water Resources Systems - Water Availability and Global Change, edited by: Franks, S., Blöschl, G., Kumagai, M., Musiake, K., and Rosbjerg, D., 280, IAHS Press, 165-172, 2003.

Immerzeel, W. W., Pellicciotti, F., and Bierkens, M. F. P.: Rising river flows throughout the twenty-first century in two Himalayan glacierized watersheds, Nature Geoscience, 6, 742-745, https://doi.org/10.1038/ngeo1896, 2013.

Lafon, T., Dadson, S., Buys, G., and Prudhomme, C.: Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods, International Journal of Climatology, 33, 1367-1381, https://doi.org/10.1002/joc.3518, 2013.

Lutz, A. F., Immerzeel, W. W., Kraaijenbrink, P. D., Shrestha, A. B., and Bierkens, M. F.: Climate Change Impacts on the Upper Indus Hydrology: Sources, Shifts and Extremes, PLoS One, 11, e0165630, https://doi.org/10.1371/journal.pone.0165630, 2016.

Mpelasoka, F. S., and Chiew, F. H. S.: Influence of Rainfall Scenario Construction Methods on Runoff Projections, Journal of Hydrometeorology, 10, 1168-1183, https://doi.org/10.1175/2009jhm1045.1, 2009.

Rozenberg, J., Davis, S. J., Narloch, U., and Hallegatte, S.: Climate constraints on the carbon intensity of economic growth, Environmental Research Letters, 10, 95006, https://doi.org/10.1088/1748-9326/10/9/095006, 2015.

Sorg, A., Huss, M., Rohrer, M., and Stoffel, M.: The days of plenty might soon be over in glacierized Central Asian catchments, Environmental Research Letters, 9, 104018, https://doi.org/10.1088/1748-9326/9/10/104018, 2014.