Authors' responses to comments

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 appreciate the editor's and both anonymous reviewers' valuable suggestions and comments on the manuscript. All suggestions are helpful to improve this manuscript. We have carefully studied, considered and responded to all comments point-by-point as follows. For clarity, all comments are given in italics and responses are given in plain text. The manuscript has been modified accordingly.

Responses to Reviewer #1's comments

We sincerely appreciate the referee's comments and suggestions on the manuscript. Our responses are as follows.

The manuscript is about the transferability of the climate model uncertainties, introduced by the selection of climate models, to hydrological impacts. To this end, two envelope-based selection methods, K-means clustering and the Katsavounidis-Kuo-Zhang (KKZ) method, are used to select subsets from an ensemble of 50 climate models over two watersheds with different climate characteristics. The transferability of the climate model uncertainties is evaluated by comparing uncertainty coverage between 31 climate variables and 16 hydrological variables that are simulated by the hydrological model GR4J. In addition, also the importance of choosing climate variables properly while selecting subsets is investigated by in- and excluding temperature variables. The manuscript is well structured and written. The manuscript covers a topic that is original and novel, and might interest a large amount of readers, including climate scientists and hydrologists. To my opinion, this manuscript needs some minor revisions. I have added a few comments/suggestions that need to be addressed before acceptance.

We appreciate that the reviewer is in favor of the content of this research. All the comments and suggestions have been replied to below and have been addressed in the revision.

General Comments

I would rephrase the title a bit. In the title, the authors are referring to the transferability of climate simulation (model?) uncertainty to "hydrological climate change impacts", whereas in the Abstract and other parts of the manuscript the authors write about the transferability to "hydrological impacts". I would change the "hydrological climate change impact" into *"hydrological impacts". In this way, the authors can put more emphasis on the transferability of uncertainties to "hydrological impacts" specifically, and the title has a better connection with the Abstract and manuscript or vice versa.*

Thank you for the suggestion on the title. The "hydrological climate change impacts" has been changed to "hydrological impacts" in the revised manuscript. With regard to the use of "climate simulation" instead of "climate model", although global climate models (GCMs) are the main uncertainty source considered, this research also includes uncertainty related to future emission scenarios (i.e. RCP4.5 and RCP8.5). Thus, we think that "climate simulation" is more appropriate than "climate model" in the title.

Two watersheds with different climate characteristics are selected for this study. It would be good to spend some text in the Discussion explaining what the potential effects of transferability are in other climatic regions, such as high-mountain regions.

Thanks for the comment. We agree with the reviewer that it is necessary to further discuss the transferability in other climatic regions.

For example, for high-mountain regions, precipitation may be influenced by complex topography and snowmelt often has the greatest contribution to runoff. Therefore, it is recommended that climate variables related to orographic precipitation and the evolution of snowpack be included in the selection process. Additionally, for arid regions, Hortonian overland flow may be the predominant runoff mechanism, and thus precipitation variables that are capable of describing the intensity of precipitation events may need to be stressed in the selection of subsets.

These points have been added to the Discussion section of the revised manuscript [Lines 29, Page 14-Lines 4, Page 15].

Specific Comments

1. Introduction; L4-9: the authors indicate that the selection methods inherit the potential flaws of the past-performance approach, when the emphasis is on model performance. What are the potential flaws of the past-performance approach? Combining envelope coverage criteria and past-performance would, to my opinion, be better since not only the models are selected to represent a full range of climate conditions, but also are tested in their performance to simulate regional (historical) climate characteristics, especially in those regions where, for instance, monsoon systems prevail.

Thank you for the comment. In this sentence, the potential flaws are meant as "In general, the assumption that models with good performance over the near-past provide more realistic climate change signals is questionable (Knutti et al., 2010; Reifen and Toumi, 2009), and the metrics commonly used to evaluate model performance are often manually defined based on the fields of interest, which leads to substantial subjectivity within the weighting process (Mendlik and Gobiet, 2016)". Specifically, the best performing models may not produce the most realistic climate change signal in the future. In addition, the ranking or weighting GCMs is highly dependent on the definition of metrics. Thus, the past-performance approach may lead to subjectivity in the selection of climate model simulations.

With regards to combining envelope coverage criteria and past-performance criteria, on one hand, we actually wanted to say (but failed to say it clear enough) that many approaches combining both criteria put more emphasis on the past-performance criteria. These methods may inherit the potential flaws of the past-performance approach. This information has been clarified in the revised manuscript [Lines 8-9, Page 3]. On the other hand, we agree with the reviewer that combining envelope coverage criteria and past-performance criteria may be better at selecting climate simulations for impact studies. In other words, it may be more reasonable to remove unrealistic models rather than keep models with "best performance". This point has been stated in the last paragraph of the Discussion section [Lines 30-32, Page 15]. This question deserves further investigation.

> *1. Introduction; L20-25: To my opinion, the number of variables that is chosen for a selection approach depend on what the scope is of the study. If a study has only a focus on projecting changes in water availability or changes in the water balance it would, to my opinion, not be necessary to take indices into account that represent climatic extremes, whereas a study with a focus on hydrological extremes needs to include these indices in the selection approach. Therefore, the authors need to elaborate more on why a certain number of variables should be selected or not.*

Thanks for this comment. We agree with the reviewer that it might not necessary to take so many climate extreme indices into consideration when a study does not focus on hydrological extremes. However, it is difficult to determine a one-to-one linkage between climate and hydrological variables due to the non-linearity of hydrological responses. Only a few climate variables may be not enough to describe climate conditions which have great influence on hydrological processes. The main objective of this study is to investigate the transferability of climate simulation uncertainty to hydrological impacts. The results show that including climate extreme indices improves the transferability of climate simulation uncertainty, compared with the results of Chen et al. (2016). Therefore, climate conditions described by extreme indices are found to be important for the transferability of climate uncertainty.

These issues have been discussed in the revised manuscript [Lines 18-21, Page 14].

> *2.2.1 Climate Simulations; L9: Why did the authors select 50 models? The authors mentioned before that the CMIP5 archive includes 61 models. Is there a reason why the other 11 models*

are excluded from the selection approaches? In addition, each climate model has one or more ensemble member. Did the authors select the first ensemble member or did they select random ensemble members? The authors need to include this information in the method description, for instance by adding extra information to Table 1.

Thanks for your comments. We actually employed 26 GCMs from the CMIP5 ensemble (Table 1) with simulations based on two Representative Concentration Pathways (RCP4.5 and RCP 8.5), with the exception of CMCC-CESM, which only used RCP8.5, and MRI-ESM1, which only used RCP4.5. On the whole, 50 climate simulations were used. Some GCMs in the CMIP5 ensemble were excluded due to lack of relevant variables (e.g., daily outputs that are necessary to drive the hydrological model) or lack of temporal coverage (e.g., the reference or future periods used in this study). In addition, GCMs employed in this study are mostly consistent with Chen et al. (2016) to make the two studies more comparable.

In addition, only the first ensemble member of each GCM was used; this information has been clarified in the Data section [Line 9, Page 5].

2.2.2 Observations; L17: Where are the data from the 100 rain gauges, 8 temperature gauges, and 1 streamflow gauge obtained? The authors have to include some references to the sources where they obtained the meteorological and discharge data.

Thanks for the suggestion to include references to the sources of observational data. Observations in the Xiangjiang watershed are the same as those used by Zeng et al. (2016) and Xu et al. (2013) and were provided by the Changjiang Water Resources Commission. This information has been clarified in the Data section [Line 18, Page 5].

3.2 Generation of Climate Scenarios: Why did the authors use the DS method to downscale GCMs and not a method such as the Advanced Delta Change approach or the Quantile Mapping approach that also take changes in extremes into consideration? It might be interesting to discuss potential uncertainties that are introduced by the downscaling approaches in the Discussion.

Thanks for the comment. We failed to describe clear enough that Daily Scaling (DS) method is an advanced delta change approach combining delta change and quantile mapping methods. The rationale behind using a change factor method instead of a bias correction method is that climate change signals may be modified by some forms of bias correction (e.g., see Cannon et al., 2015). Still, in order to investigate the influence of alternative bias correction/downscaling approaches, we have also used the Quantile Mapping (QM) approach to post-process the CMIP5 GCMs. Figure R1 shows the results of average PSC in this case. For the Xiangjiang watershed, temperature variables were excluded in the process of selection, while they were included for the Manicouagan 5 watershed. Compared with the results calculated by the DS method, the overall characters of the results are mostly the same. The choice of the downscaling method may have little influence on the conclusions of this study. However, to highlight the potential sensitivity of results to different downscaling methods, this information has been added to the Discussion section of the revised manuscript [Lines 33, Page 15-Lines 3, Page 16 and Fig. 10].

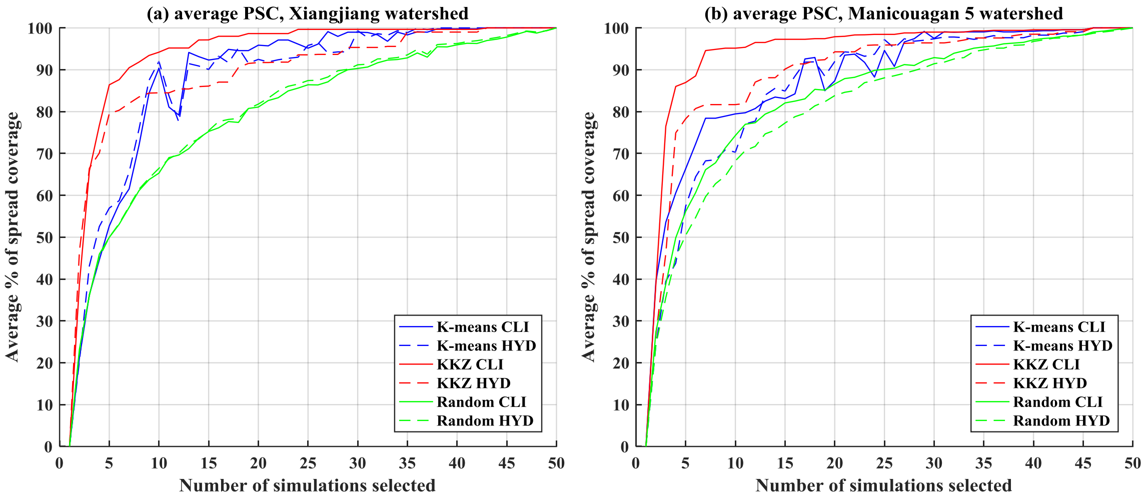

Figure R1: The average PSC for three different selection methods over two watersheds when using the QM approach (CLI = climate variables and HYD = hydrological variables).

3.3.1 Hydrological Modelling; L15-21: I would recommend replacing and to discuss this part in more detail in the Results Section, for instance under a separate subsection "Calibration and Validation"

We agree with the reviewer. The calibration and validation of the hydrological model have been presented in the Results section as follows [Lines 15-28, Page 9, Table 4 and Fig.3].

The basin-averaged daily minimum and maximum temperature and precipitation in calibration and validation periods, as shown in Table R1, were used to calibrate and validate the GR4J-6 model over the two watersheds. Model parameters were obtained by the shuffled complex evolution optimization (Duan et al., 1992) based on the objective to maximize Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970). The optimally chosen sets of parameters yield a NSE between 0.87 and 0.93 over both watersheds. The calibrated GR4J-6 also yields small relative errors of water balance with values between -0.3% and 5.4% for calibration and validation, which demonstrate the applicability of the calibrated model over both watersheds (Table R1). Figure R2 presents the mean daily hydrographs (average discharge of each calendar day across the years) for the calibration and validation periods over two watersheds. The calibrated GR4J-6 has good performance in most parts of the year, with the exception that discharge is overestimated in the winter for the Manicouagan 5 watershed. In addition, since snow accumulation and snowmelt processes are not important in the Xiangjiang watershed, the GR4J model (excluding snow module) was also calibrated in this watershed. Results showed that there was little difference between the calibrated GR4J and GR4J-6 models, and thus the presence of the CemaNeige snow module would not influence the performance of GR4J-6 in the rainfall-characterized Xiangjiang watershed (Table R1).

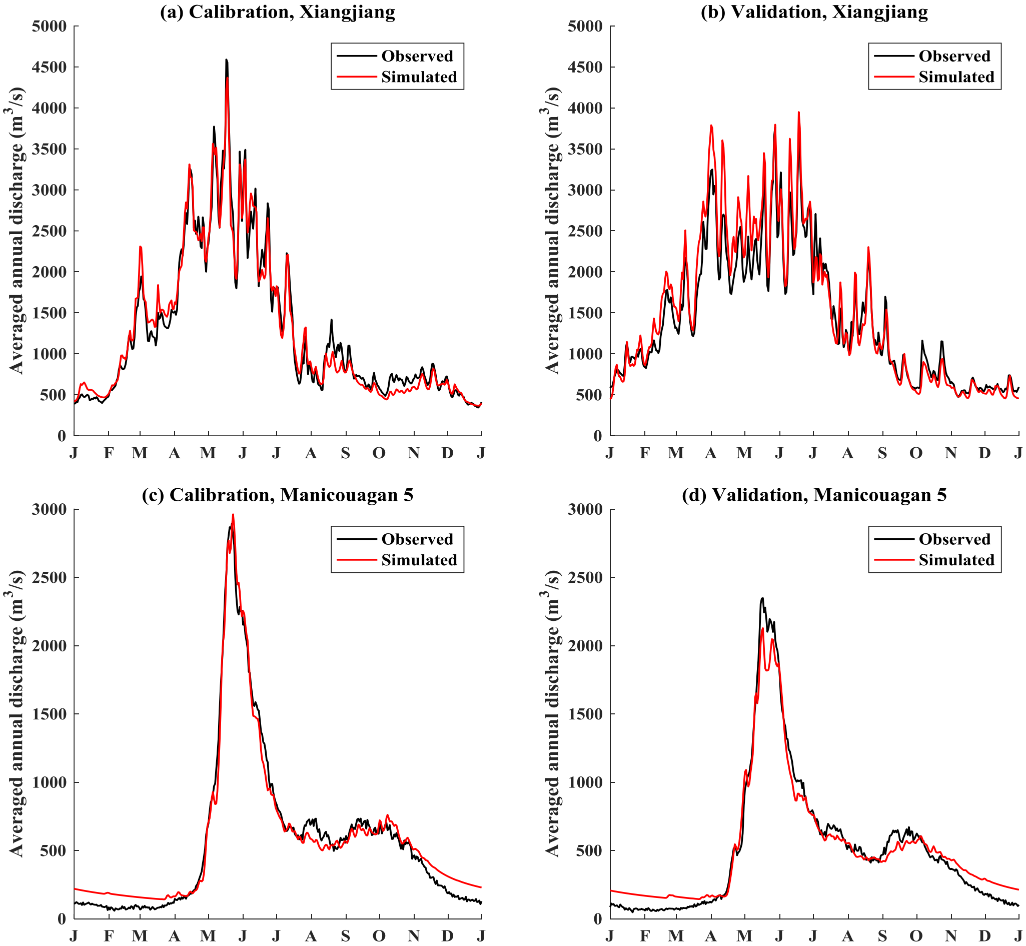

Figure R2: Observed and simulated mean hydrographs for (a, c) calibration and (b, d) validation periods over the (a, b) Xiangjiang and (c, d) Manicouagan 5 watersheds.

Table R1: Nash-Sutcliffe Efficiencies (NSE) and relative errors of water balance (RE) of hydrological models in the calibration and validation over two watersheds

Country	Watershed name	Area (km^2)	Hydrological Model	Calibration period	NSE calibration	RE calibration	Validation period	NSE validation	RE validation
China	Xiangjiang	52150	GR4J-6	1975-1987	0.912	-0.3%	1988-2000	0.871	5.4%
			GR4J	1975-1987	0.912	-0.2%	1988-2000	0.872	5.5%
Canada	Manicouagan 5	24610	GR4J-6	1970-1979	0.926	3.8%	1980-1989	0.881	2.7%

3.4 Data Analysis: It would be good to an additional sentence on what it means when a PSC is high or low. For the reader, it might be more difficult to image the meaning of a high or low PSC.

Thanks for your suggestion. A higher PSC means that the selected subset covers a larger uncertainty range. This has been clarified in the revised manuscript [Line 4, Page 9].

Figure 1: The longitude axes are given, but the latitude axes are missing. Further I think the figure does not contain a lot of information. I would add a digital elevation model or another topographic/geographic info to give the reader more valuable information on the

characteristics of the catchments. In addition, I would add the positions of the discharge gauging stations used for the calibration/validation. Finally, I recommend making inlets, including the catchment maps, larger.

We agree with the reviewer. Figure R3 has been updated in the revised manuscript [Fig.1].

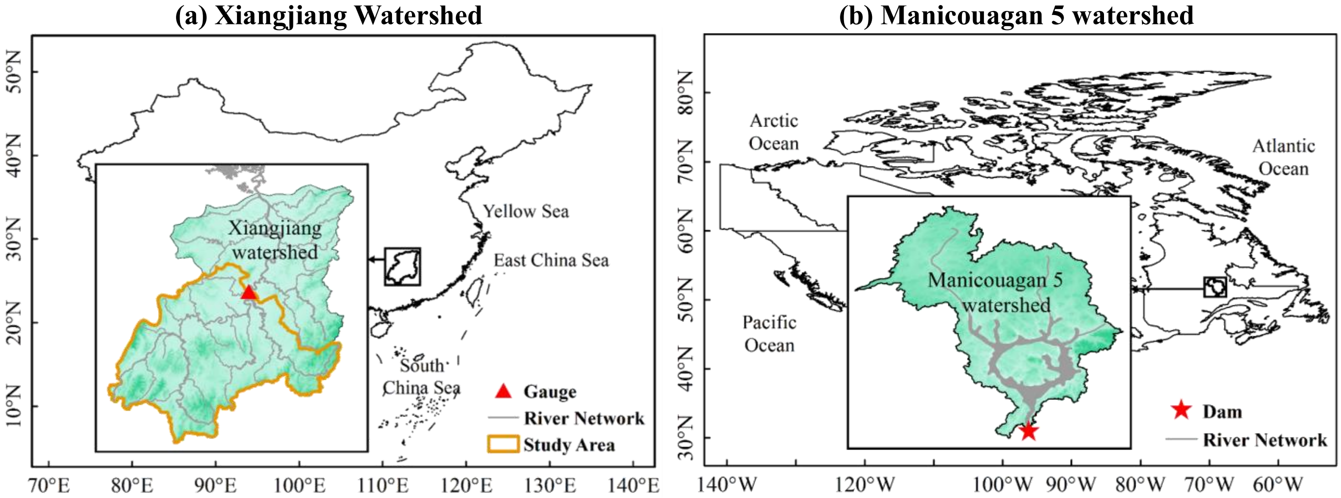

Figure R3: Location maps of the (a) Xiangjiang and (b) Manicouagan 5 watersheds (The study area in the Xiangjiang watershed is one of its sub-basins as the orange boundary shows).

Technical Comments

Abstract; L16: ". . .the importance of choosing climate variables properly while selecting subsets. . ." instead of ". . .the importance of properly choosing climate variables in selecting subsets. . ."

This sentence has been revised as suggested [Line 18, Page 1].

Responses to Reviewer #2's comments

We would like to thank the reviewer for the time taken in reviewing this paper. Please find the point-by-point responses below. We have made 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 below.

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.

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

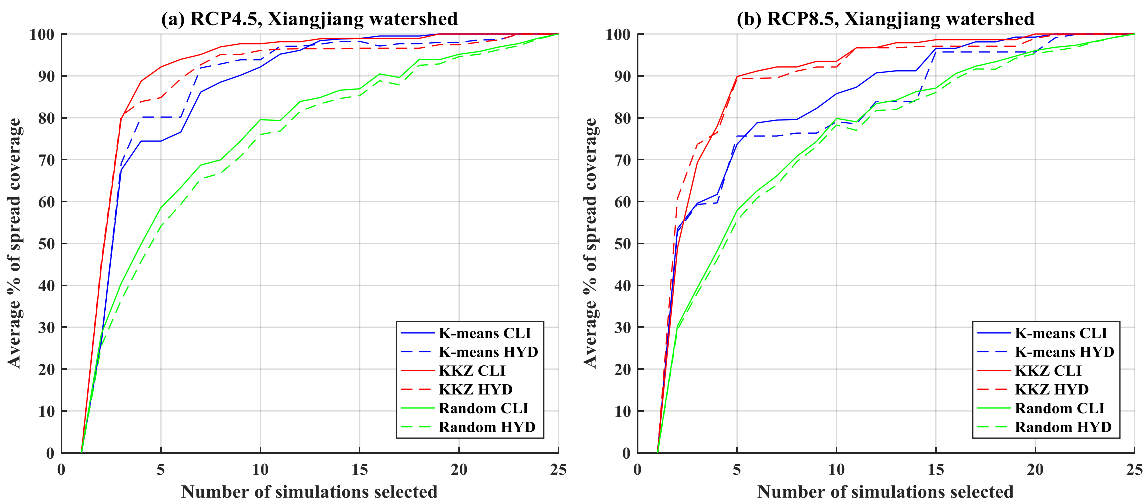

Figure R1: The average PSC for three different selection methods over the Xiangjiang watersheds when only one emission scenario is considered.

own needs.

In order to stress this comment, an explanation on the choice of emission scenarios has been added in the Discussion section [Lines 3-14, Page 16 and Fig.11].

> *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 have been clarified in the revised manuscript [Lines 26-29, Page 7].

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

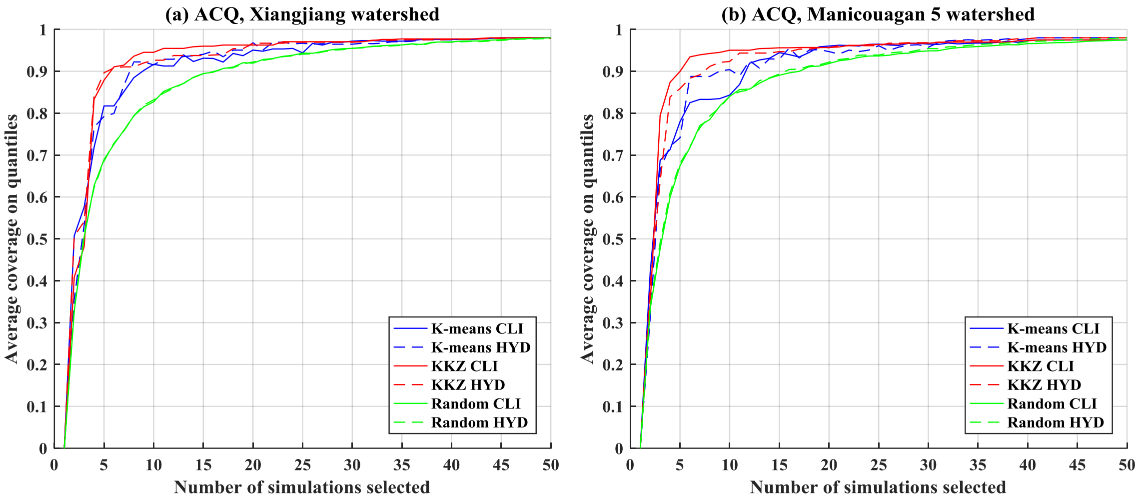

Figure R2: The average coverage on quantiles (ACQ) for three different selection methods over two watersheds (CLI = climate variables and HYD = hydrological variables)

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 has been clarified in the revised manuscript [Line 22, Page 9].

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.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes?, Journal of Climate, 28, 6938-6959, https://doi.org/10.1175/JCLI-D-14-00754.1, 2015.

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.

Chen, J., Brissette, F. P., and Lucas-Picher, P.: Transferability of optimally-selected climate models in the quantification of climate change impacts on hydrology, Climate Dynamics, 47, 3359-3372, https://doi.org/10.1007/s00382-016-3030-x, 2016.

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resources Research, 28, 1015-1031, https://doi.org/10.1029/91WR02985, 1992.

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.
Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A.: Challenges in Combining Projections from Multiple Climate Models, Journal of Climate, 23, 2739-2758, https://doi.org/10.1175/2009jcli3361.1, 2010.
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.
Mendlik, T., and Gobiet, A.: Selecting climate simulations for impact studies based on multivariate patterns of climate change, Climatic Change, 135, 381-393, https://doi.org/10.1007/s10584-015-1582-0, 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.
Nash, J. E., and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282-290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.
Reifen, C., and Toumi, R.: Climate projections: Past performance no guarantee of future skill?, Geophysical Research Letters, 36, https://doi.org/10.1029/2009gl038082, 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.
Xu, H., Xu, C.-Y., Chen, H., Zhang, Z., and Li, L.: Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of China, Journal of Hydrology, 505, 1-12, https://doi.org/10.1016/j.jhydrol.2013.09.004, 2013.
Zeng, Q., Chen, H., Xu, C.-Y., Jie, M.-X., and Hou, Y.-K.: Feasibility and uncertainty of using conceptual rainfall-runoff models in design flood estimation, Hydrology Research, 47, 701-717, https://doi.org/10.2166/nh.2015.069, 2016.

Transferability of climate simulation uncertainty to hydrological ~~climate change~~ impacts

Hui-Min Wang[1], Jie Chen[1], Alex J. Cannon[2], Chong-Yu Xu[1,3], Hua Chen[1]

[1]State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, 430072, China
[2]Climate Research Division, Environment and Climate Change Canada, Victoria BC, Canada
[3]Department of Geosciences, University of Oslo, Oslo, Norway

Correspondence to: Jie Chen (jiechen@whu.edu.cn)

Abstract: Considering rapid increases in the number of climate model simulations being produced by modelling centres, it is often ~~Increasing number of climate models are being produced to cover the uncertainty, which makes it~~ infeasible to use all of them in climate change impact studies. In order to thoughtfully select subsets of climate simulations from a large ensemble, several envelope-based methods have been proposed. The subsets are expected to cover a similar uncertainty envelope as the full ensemble in terms of climate variables. However, it is not a given that the uncertainty in hydrological impacts will be similarly well represented. Therefore, this study investigates the transferability of climate uncertainty related to the choice of climate simulations to hydrological impacts. Two envelope-based selection methods, K-means clustering and Katsavounidis-Kuo-Zhang (KKZ) method, are used to select subsets from an ensemble of 50 climate simulations over two watersheds with very different climates using 31 precipitation and temperature variables. Transferability is evaluated by comparing uncertainty coverage between climate variables and 17 hydrological variables simulated by a hydrological model. The importance of ~~properly~~ choosing climate variables properly when ~~in~~ selecting subsets is investigated by including and excluding temperature variables. Results show that KKZ performs better than K-means at selecting subsets of climate simulations for hydrological impacts, and the uncertainty coverage of climate variables is similar to that of hydrological variables. The subset of first 10 simulations covers over 85% of total uncertainty. As expected, temperature variables are important for the snow-related watershed, but less important for the rainfall-driven watershed. Overall, envelope-based selection of around 10 climate simulations, based on climate variables that characterize the physical processes controlling hydrology of the watershed, is recommended for hydrological impact studies.

1 Introduction

In studies of climate change impacts on hydrology, multi-model ensembles (MMEs) formed by multiple Global Climate Models (GCMs) and multiple emission scenarios have been widely used to drive hydrological models (Minville et al., 2008; Vaze and Teng, 2011; Mehran et al., 2014; Chen et al., 2011b). There are two strengths of using MMEs: (1) the MME mean typically performs better than any individual model in representing the mean of historical climate observations (Gleckler et al., 2008; Pierce et al., 2009; Pincus et al., 2008; Mehran et al., 2014); and (2) the spread of a MME can be used to estimate climate change impact uncertainties, for example those related to GCM structure, future greenhouse gas concentrations, and internal climate variability (Mendlik and Gobiet, 2016; Knutti et al., 2010; Chen et al., 2011b; Tebaldi and Knutti, 2007). While climate projection uncertainty and spread or coverage of a MME are not equivalent, the latter does provide an imperfect estimate of uncertainty and, for sake of simplicity, we use the terms interchangeably in the remainder of this study.

The number of GCM simulations available for impact studies is increasing rapidly. For instance, the Coupled Model Intercomparison Project Phase 3 (CMIP3) contains outputs from 25 different GCMs, whereas CMIP5 contains outputs from 61 GCMs (https://pcmdi.llnl.gov), with each GCM contributing one or more simulation runs (Taylor et al., 2012). Although it is usually advised that as many climate simulations as possible be used in impact studies, the extraction, storage, and computational costs associated with a large MME may be prohibitive. In practice, it is not uncommon for impact studies to instead rely on a small subset of climate simulations, the members of which are often selected manually, relying on expert judgement.

Several studies have considered more objective means of selecting subsets of climate simulations for impact studies based on different criteria. Generally, there are two main types of selection approaches. The past-performance approach selects a subset by weighting or selecting simulations according to their ability to represent the observed near-past climate (Gleckler et al., 2008; Perkins et al., 2007; Pincus et al., 2008). Climate model performance is generally evaluated based on agreement with observed climate conditions, which is often defined by a suite of climate metrics. For example, Perkins et al. (2007) ranked climate models based on probability density functions of observed temperature and precipitation. Similarly, Gleckler et al. (2008) evaluated the performances of 22 GCMs according to relative errors of some climatological fields, but stressed that a wider range of metrics might give more robust results. In general, the assumption that models with good performance over the near-past provide more realistic climate change signals is questionable (Knutti et al., 2010; Reifen and Toumi, 2009), although recent work on emergent constraints suggests that it may be possible to remove models that fail to represent certain key physical processes that dictate the evolution of long-term climate projections (Klein and Hall, 2015). In practice, however, the metrics commonly used to evaluate model performance are often manually defined based on the fields of interest, which leads to substantial subjectivity within the weighting process.

Another means of selecting climate simulations is the envelope-based approach, which tries to select a representative subset of climate simulations that covers as much of the full ensemble's range of future climate change signals as possible (Warszawski et al., 2014; Cannon, 2015; Logan et al., 2011). For instance, Cannon (2015) used two automated multivariate statistical algorithms, K-means clustering and Katsavounidis-Kuo-Zhang (KKZ) method (Katsavounidis et al., 1994), to select subsets of CMIP5 GCMs that bracket the overall range of changes in a suite of 27 climate extreme indices. The goal of the envelope-based approach coincides with the motivation behind the usage of a MME, namely to account for different sources of projection uncertainty, including structural uncertainty (Wilcke and Bärring, 2016; Tebaldi and Knutti, 2007).

Some studies have proposed selection methods that combine both near-past performance and climate change envelope coverage criteria (~~Mendlik and Gobiet, 2016;~~ McSweeney et al., 2012; Lutz et al., 2016a; Giorgi and Mearns, 2002). For example, Lutz et al. (2016a) took both model historical skill and the range of projected changes in means and extremes into consideration through a three-step sequential selection procedure. Since many examples of such selection method emphasize on ranking or weighting climate model performance, they may inherit the potential flaws of the past-performance approach~~With an emphasis on model performance, these selection methods inherit the potential flaws of the past-performance approach~~.

Regardless of underlying approach, most selection methods are only conducted on climate variables that can be calculated directly from the MME simulation outputs. Even subsets of simulations that account for most of the ensemble spread in climate variables can be identified, it is not guaranteed that the same level of spread coverage extends to hydrological impacts variables due to the complexity and nonlinearity of hydrological processes. For example, small perturbations in the frequency or intensity of temperature and precipitation regimes may have noticeable impacts on streamflow patterns and flood magnitudes (Muzik, 2001; Whitfield and Cannon, 2000). Consequently, whether the adequate coverage of climate simulation uncertainty is transferable to hydrological impacts should be evaluated before applying envelope-based selection methods in hydrological impacts studies.

Chen et al. (2016) investigated the transferability of optimally-selected climate simulations in the quantification of hydrological impacts by using two automatic selection methods (K-means clustering and KKZ method) over a Canadian watershed. They concluded that selected subsets of climate simulations do not remain optimal for hydrological variables. However, the selection methods used in their study were applied to just two climate variables, mean annual temperature and mean annual precipitation, which is a common strategy employed by practitioners who employ envelope-based approaches (Immerzeel et al., 2013; Warszawski et al., 2014). Hydrological responses are driven both by annual climate conditions and intra-annual climate processes, which may not be described by a small number of climate variables. For example, both the magnitude and intensity of a rainfall event can affect the flood discharge in a rainfall-dominated watershed. The transferability of climate uncertainty may be diminished due to insufficient climate variables.

Following Cannon (2015), who considered a larger suite of climate indices, the aim of this study is to investigate the transferability of climate simulation uncertainty to the assessment of hydrological climate change impacts by using a large pool of climate variables, including seasonal means, annual means, and climate extremes. The case study is conducted over two watersheds with very different climate conditions, one of which is seasonally snow-covered and the other driven by summer monsoon rainfall with little winter snowfall. Two envelope-based approaches (K-means clustering and KKZ method) are used to select subsets of climate simulations based on a different sets of climate variables. Transferability is evaluated by comparing the uncertainty coverage between the climate variables and 17 hydrological variables simulated by a hydrological model.

2 Study Area and Data

2.1 Study Area

This study was conducted over two watersheds (the Xiangjiang and Manicouagan 5 watersheds) with different climate and hydrological characteristics (Fig. 1). The Xiangjiang watershed is a monsoon-climate and rainfall-dominated watershed located in south-central of China, whereas Manicouagan 5 is a temperate-climate and seasonally snow-covered watershed located in central Quebec, Canada.

2.1.1 Xiangjiang Watershed

The Xiangjiang watershed is one of the largest sub-basins of the Yangtze River watershed (Fig. 1a). The Xiangjiang River originates from the Haiyang Mountain in Guangxi Autonomous Region and flows north to the Dongting Lake in Hunan Province, which connects to the Yangtze River. The Xiangjiang River consists of several tributaries with a surface area of approximately 94,660 km^2, but only the watershed with an area of 52,150 km^2 above the Hengyang gauging station was used in this study. The watershed has a hilly topography ranging from a maximum elevation of 2042 m above sea level to a minimum elevation of 58 m above sea level at the Hengyang station. The Xiangjiang watershed is heavily influenced by a subtropical monsoon climate with hot and humid summers and mild and dry winters. The average annual precipitation over the catchment is about 1570 mm almost entirely in the form of rainfall. Around 61% precipitation occurs from April to August, resulting in high flows during this period. The average daily maximum and minimum temperatures are around 22 ℃ and 15 ℃, respectively. The average daily discharge at the Hengyang station is around 1400 m^3/s. The peak discharge of the averaged daily hydrograph is about 4420 m^3/s, mainly resulting from high intensity rainfall.

2.1.2 Manicouagan 5 Watershed

The Manicouagan 5 watershed, the largest sub-basin of the Manicouagan River watershed, is located in the center of the province of Quebec, Canada (Fig. 1a). The Manicouagan 5 River discharges into the Manicouagan reservoir, an annular reservoir within the remnant of an ancient eroded impact crater, and ends at the Daniel Johnson Dam, which is the largest buttressed multiple arc dam in the world. The drainage area of the Manicouagan 5 River is about 24,610 km^2, which is mostly covered by forest and has a moderately hilly topography ranging from a maximum elevation of 952 m to a minimum elevation of 350 m above sea level (Chen et al., 2016). The Manicouagan 5 watershed has a continental subarctic climate dominated by long and cold winters. The annual precipitation is fairly evenly distributed within the year and averages about 912 mm, around 45% of which is snowfall. The average daily maximum and minimum temperatures are around 2.4 ℃ and -7.8 ℃, respectively.

The average discharge of the Manicouagan 5 River is about 530 m^3/s. The peak discharge of averaged daily hydrograph is around 2200 m^3/s, mainly resulting from snowmelt.

2.2 Data

Both observed and simulated daily meteorological (maximum and minimum temperatures and precipitation) data over both watersheds were used in this study. All the climate data from multiple stations or grids were averaged over the watersheds.

2.2.1 Climate Simulations

Climate model simulation data used in this study were extracted from the CMIP5 archive (Taylor et al., 2012) for both the historical reference (1975-2004) and future (2070-2099) projection periods. Twenty-six GCMs from 15 institutions were ~~selected~~ employed to represent climate modeling uncertainty (Table 1). Two Representative Concentration Pathways (RCP4.5 and RCP 8.5) were used for each GCM to represent forcing scenario uncertainty, with the exception of CMCC-CESM, which only used RCP8.5, and MRI-ESM1, which only used RCP4.5. Only the first run of each GCM was used. On the whole, an ensemble of 50 climate simulations was used in this study.

2.2.2 Observations

Observed daily meteorological data used to downscale the GCM outputs and calibrate the hydrological model cover the 1975-2004 period for both watersheds. Meteorological data for the Manicouagan 5 watershed were obtained from the 10-km gridded dataset of Hutchinson et al. (2009), which was created by fitting spatially continuous functions of longitude, latitude and elevation to daily station data using a trivariate thin plate smoothing spline interpolation algorithm. Discharge data at the outlet of the Manicouagan 5 River were based on mass balance calculations at the Daniel Johnson Dam. Meteorological and discharge data for the Xiangjiang watershed were observed at 97 rain gauges, 8 temperature gauges, and 1 streamflow gauge in the catchment above the Hengyang station, which are the same as those used in Zeng et al. (2016) and Xu et al. (2013).

3 Methodology

3.1 Subset Selection of GCM Simulations

Two automated envelope-based methods were used to select subsets of climate simulations. One is the K-means clustering which finds cluster centroids that best characterize high-density regions of a multivariate space, the other is the KKZ method which recursively selects simulations that best span the spread of an ensemble (Cannon, 2015). Both selection methods operate on multivariate data, which means that they are sensitive to the choice and scaling of climate variables.

3.1.1 Climate Variables

Since the hydrological response of a watershed depends not only on annual mean temperature and precipitation but is also sensitive to intra-annual climate variability (e.g. seasonal means or extremes), subset selection should be based on a set of climate variables that includes annual and seasonal averages as well as extremes. The World Meteorological Organization's Expert Team on Climate Change Detection and Indices (ETCCDI) has recommended a set of core climate indices that can be easily derived from daily meteorological data series (http://etccdi.pacificclimate.org/list_27_indices.shtml). The ETCCDI indices are designed to monitor changes in the frequency and intensity of climate extreme events and characterize the variability of extremes (Zhang et al., 2011). Here, we assume that the ETCCDI indices are sufficient to characterize climate extremes that lead to hydrological impacts.

Specifically, this study used a set of 31 climate variables as shown in Table 2 (21 ETCCDI extreme indices and 10 seasonal or annual mean indices), including 16 temperature variables and 15 precipitation variables. Since the focus of this study is on the capability of selected GCM subsets to cover uncertainty of climate change signals, changes in climate variables (relative change for precipitation and absolute change for temperature and duration) between the historical reference period (1975-2004) and the future projection period (2070-2099) were calculated for 50 climate simulations over the two study watersheds. Changes in each climate variable were standardized to zero mean and unit standard deviation to eliminate influences from different magnitudes and units between variables. These changes in climate variables are referred to as simulated climate change signals. Once changes were calculated, subsets could be selected based on the multivariate space formed by the climate variables.

3.1.2 K-means Clustering

The K-means clustering algorithm divided the ensemble of 50 climate simulations into a user-specified number of clusters based on the objective of minimizing within-cluster sums of squared errors (SSE) (Hartigan and Wong, 1979). Each cluster is represented by its centroid. The SSE is characterized by the Euclidean distances from simulations to their corresponding cluster centroids. Some studies have applied this method to select subsets of climate simulations (Logan et al., 2011; Cannon, 2015; Houle et al., 2012). The climate simulations closest to the centroids were selected as the subsets. Due to sensitivity of the K-means clustering to initial cluster centroid positions, it was run 10000 times with different initializations and the best solution with lowest SSE was kept. A disadvantage of the K-means clustering is that the selected climate simulations are not ordered. In other words, the optimally selected simulations in a small subset may not be optimal in a larger subset, which means that it is inconvenient for end-users to change the subset size for different applications.

3.1.3 KKZ Method

The KKZ method was originally designed by Katsavounidis et al. (1994) to identify a set of optimal seed cases as initial centroids in the K-means clustering, and was introduced by Cannon (2015) in the selection of climate simulations. This method prefers the peripheral simulations in the multivariate space. The specific procedure is as follows:

1. The climate simulation closest to the centroid of whole ensemble is selected as the first simulation;

2. The simulation farthest from the first selected simulation is selected as the second representative simulation;

3. Subsequent simulations are selected as follows:

(1) The distances from each remaining simulation to every previously-selected simulation are calculated;

(2) Each remaining simulation is designated with the minimum distance among all distances calculated in step 3(1); and

(3) The simulation with the largest minimum distance, which is designated in step 3(2), is selected as the next selected simulation.

Compared to the K-means clustering, the KKZ method is deterministic and ordered. However, it is more susceptible to selecting outliers than K-means clustering. In addition, a random selection, repeated 100 times to minimize the influence of its stochastic nature, was conducted as a baseline to evaluate the K-means clustering and KKZ method.

3.2 Generation of Climate Scenarios

GCM outputs are typically on a coarse spatial grid and contain systematic biases that preclude their direct use in hydrological modeling (Mpelasoka and Chiew, 2009; Chen et al., 2011a; Chen et al., 2011b; Minville et al., 2008; Vaze and Teng, 2011). It is thus necessary to bias correct and downscale GCM outputs before running the hydrological model. The main objective of this study is to investigate the transferability of climate simulation uncertainty; hence, there is no need to use a complicated downscaling method. A commonly used change factor method, namely the Daily Scaling (DS) method proposed by Harrold and Jones (2003), was used in this study. This method assumes that climate change signals simulated by GCMs are credible and can be used to perturb observations to obtain future daily series. The DS method adjusts the observed daily series using the differences in distributions of simulated temperature/precipitation between the future period and the reference period. The specific steps are:

1. Distributions (represented by 100 quantiles in this study) of daily temperature and precipitation simulated by GCMs are calculated for both reference and future periods in each calendar month (e.g., January, February, etc.);

2. Scaling factors are estimated as the differences (for temperatures) or ratios (for precipitation) in distributions of precipitation or temperature between reference and future periods for each calendar month; and

3. Scaling factors are added (for temperatures) or multiplied (for precipitation) to corresponding distributions of observed daily temperature or precipitation for each calendar month.

The use of the DS method preserves the simulated climate change signal. It is based on differences in probability distributions between the reference and future periods, which are only caused by climate change signals. In addition, the consideration of quantile-dependent changes in the precipitation distribution is important in hydrological impact studies, because more runoff is generated in high-intensity precipitation events (Harrold and Jones, 2003; Chiew et al., 2009). The use of 100 quantiles in the DS method is the same as many other studies (Mpelasoka and Chiew, 2009; Chen et al., 2013). Lafon et al. (2013) showed that the empirical quantile mapping method based on 100 quantiles is more accurate than that based on 25, 50 or 75 quantiles. However, temporal sequencing in the future period is assumed to be the same as in the observed data. Changes in, for example, wet/dry spell lengths are not informed by the GCM simulations.

3.3 Hydrological Response Simulation

3.3.1 Hydrological Modeling

The hydrological impacts were simulated by a 6-parameter, lumped, conceptual hydrological model, GR4J-6. The GR4J-6 model consists of the GR4J rainfall-runoff model and the CemaNeige snow accumulation and melt routines (Arsenault et al., 2015). The GR4J model is a reservoir-based 4-parameter model developed on the basis of the GR3J model (Edijatno et al., 1999; Perrin et al., 2003). This model routes runoff through a production reservoir, two linear unit hydrographs and a non-linear routing reservoir. Four parameters have to be calibrated for this model. They are maximum capacity of the production reservoir, groundwater exchange coefficient, one-day-ahead maximum capacity of the routine reservoir and time base of unit hydrograph. In an evaluation of the model, Perrin et al. (2003) found that GR4J outperformed 19 models over a large sample of catchments.

Due to its lack of snow accumulation and snowmelt algorithms, the GR4J model cannot be directly used in snow-related watersheds. Thus, the general snow accounting routine proposed by Valéry et al. (2014), CemaNeige, was added. In this routine, precipitation is divided into rainfall and snowfall depending on the daily range of temperatures, and the updating of snowpack and snowmelt is based on a degree-day approach that is controlled by two free parameters (cold content factor and snowmelt factor). In addition, the Oudin formulation (Oudin et al., 2005) was used to pre-process evapotranspiration for GR4J-6.

~~The minimal required daily basin-averaged input data for GR4J-6 includes maximum and minimum air temperature and precipitation. Model parameters were calibrated using shuffled complex evolution optimization (Duan et al., 1992) to maximize Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970). The periods of daily discharge data used for model calibration and validation are shown in Table 3. The optimally chosen sets of parameters yield a NSE greater than 0.87 over both watersheds. Figure 2 presents the observed and simulated mean hydrographs for the calibration and validation periods over two watersheds. In addition, the GR4J model was also calibrated in the Xiangjiang watershed. Results showed that absence of the CemaNeige snow module would not influence the performance of GR4J-6 in the rainfall-dominated Xiangjiang watershed (Table 3).~~

3.3.2 Hydrological Variables

To examine the performance of subset selection in terms of hydrological response uncertainty, this study used a set of 17 hydrological variables based on Water Resources Indicators (WRIs), Indicators of Hydrologic Alteration (IHAs) and quantiles of daily flow series (Table ~~4~~3). WRIs have been used in many hydrological impact studies to assess streamflow alteration due to natural and anthropogenic climate change (Eum et al., 2017; Shrestha et al., 2014; Chen et al., 2011b). IHAs are used to examine the temporal variation of key streamflow hydrograph components (Eum et al., 2017; Richter et al., 1996; Shrestha et al., 2014). Quantiles of daily flow series have been used to describe the characteristics of flow regimes (Mu et al., 2007; Wilby, 2005).

Similar to climate variables, changes in hydrological variables between the reference (1975-2004) and future (2070-2099) period were calculated. To remove the influence of systematic biases between the observations and simulations, simulated runoff values instead of gauge observations were used as flow data in the reference period. The first year of each period was used to spin-up the hydrological model and was excluded when calculating the hydrological variables. Once the projected changes in hydrological variables were calculated, the uncertainty coverage of subsets could be compared between climate variables and hydrological variables to evaluate the transferability of climate simulation uncertainty.

3.4 Data Analysis

A criterion called the Percentage of Spread Coverage (PSC) is used to measure the uncertainty coverage of a subset relative to the coverage of all simulations. For a given variable and subset, PSC was calculated by dividing the variable's range in the subset by the variable's range in all simulations. A higher PSC means that the selected subset covers a larger uncertainty range. Figure ~~3~~ 2 shows examples of PSC when 5 climate simulations are selected using the KKZ method. Since it is difficult to illustrate results in more than 3 dimensions, examples are limited to 1, 2 and 3 variables. In Fig. ~~3a~~2a, points represent the changes in 'WiT' (seasonal average temperature in winter) for 50 GCM simulations. The larger squares represent the same variable for a subset of 5 climate simulations selected by KKZ. The PSC is calculated by dividing the temperature range of the selected subset, 4.15 ℃, by that of the whole ensemble, 6.49 ℃. Therefore, for this specific variable the PSC (uncertainty coverage) of the subset is 64.01%. Similarly, every variable has a corresponding PSC associated with a subset of a given size; examples for 'WiR' (seasonal total precipitation in winter) and 'Rx1day' (annual maximum 1-day precipitation) are shown in Fig. ~~3b~~2b-c. For the random subset selection method, the reported PSC is the mean value of 100 PSCs, each calculated for a different random subset of the specified size.

4 Results

4.1 Calibration and Validation of Hydrological Model

The basin-averaged daily minimum and maximum temperatures and precipitation, as shown in Table 4, were used to calibrate and validate the GR4J-6 model over the two watersheds. Model parameters were obtained by the shuffled complex evolution optimization (Duan et al., 1992) based on the objective to maximize Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970). The optimally chosen sets of parameters yield NSE values between 0.87 and 0.93 over both watersheds for calibration and validation. The calibrated GR4J-6 also yields small relative errors of water balance between -0.3% and 5.4% for calibration and validation. These indicate the reasonable performance of GR4J-6 for both watersheds (Table 4). Figure 3 presents the mean daily hydrographs (average discharge of each calendar day across all years) for the calibration and validation periods over the two watersheds. The calibrated GR4J-6 has good performance in most parts of the year, with the exception of overestimation of winter discharge for the Manicouagan 5 watershed. In addition, since snow accumulation and snowmelt processes are not important in the Xiangjiang watershed, the GR4J model (excluding snow module) was also calibrated in this watershed. Results showed that there was little difference between the calibrated GR4J and GR4J-6 models, and thus the presence of the CemaNeige snow module would not influence the performance of GR4J-6 in the rainfall-characterized Xiangjiang watershed (Table 4).

4.2~~1~~ Transferability of Climate Uncertainty

As an illustrative example, the uncertainty transferability from one climate variable to one hydrological variable in the Xiangjiang watershed is shown in Fig. 4. The larger squares represent the 5 and 10 climate simulation subsets selected by the KKZ method. The subfigures on the top display the PSC for ‘Rx5day’ (maximum consecutive 5-day precipitation), whereas those on the bottom display the PSC for ‘Qx7day’ (7-day maximum flow). The reason for choosing these two variables is that there is a generally accepted linkage between high-intensity precipitation and high flow in a rainfall-driven watershed. Although this particular choice of climate and hydrological variables is, in some ways, unfair because the overall selection process is based on a high-dimensional multivariate climate space, these subfigures still illustrate the process of uncertainty transferability from climate simulations to hydrological impacts. Here, the PSC of the climate variable increases from 66.45% to ~~92~~87.37% as the number of selected simulations goes from 5 to 10; at the same time, the PSC of the hydrological variable increases from ~~80.53~~65.42% to 90.32~~94.59~~%. In this case, the uncertainty coverage of the subsets in terms of the climate variable is well translated to uncertainty coverage of the hydrological variable.

Figure 5 expands the example above from 1 to 2 dimensions. In this case, the subfigures on the top show a two-dimensional space formed by the changes in two climate variables, ‘ARav’ (annual total precipitation) and ‘Rx1day’ (maximum 1-day precipitation), whereas those on the bottom show changes in two hydrological variables, ‘MD’ (annual mean flow) and ‘HPD’ (mean duration of high pulses). It should be noted that the subsets of climate simulations are the same as in the 1-dimensional example above. As the number of selected simulations increases from 5 to 10, the mean PSC for the two climate variables increases from ~~67.24~~50.83% to ~~91.98~~88.72%, while the mean PSC for the two hydrological variables increases from ~~76.33~~59.46% to ~~98.06~~94.05%. The increases are mostly due to selection of outlying simulations in the top right corner of the plots (the ~~7th~~ 6th, 8th and 9~~10~~th selected simulations). There is strong consistency between locations of selected simulations in 2-dimensional climate space and hydrology space. For example, the 1~~5~~st~~th~~, 4~~7~~th and 10th selected simulations are close to each other in both climate space (Fig. 5b) and hydrology space (Fig. 5d). Accordingly, the uncertainty coverage tends to translate well from climate variables to hydrological variables in this 2-dimensional example. However, PSC increases are not consistent in all cases. For example, selection of the simulation on the left edge of Fig. 5b (the ~~8th~~ 7th selected simulation) substantially improves the PSC of 'Rx1day', but does not lie on the edge of Fig. 5d and hence does not contribute to improvements in PSC of either hydrology variable. This may be due to the nonlinearity of the hydrological model or an imperfect explanatory relationship between the climate and hydrological variables.

The discussion above is limited to results for 5 and 10 simulation subsets for one watershed selected using the KKZ method. In the study as a whole, subset sizes from 1 to 50 simulations were evaluated in terms of transferability for two watersheds and two envelope-based methods (K-means and KKZ). PSCs for all 31 climate variables and 17 hydrological variables were calculated for both selection methods and watersheds. Figure 6 shows distributions of climate and hydrological PSCs for 5, 10, 20, 30 and 40 simulation subsets. For the Xiangjiang watershed (Fig. 6a~~,~~-b~~c~~), PSCs for the climate variables are similar to those for the hydrological variables. For the Manicouagan 5 watershed (Fig. 6c~~b,~~-d), PSCs of the hydrological variables are ~~consistently~~ slightly smaller than those for the climate variables. Overall, the tendency of the hydrological PSCs to increase with subset size is comparable to that for the climate PSCs in both watersheds. In other words, as the size of subset becomes larger, the improvement in PSCs of the hydrological variables is similar to that of the climate variables. When comparing the two envelope-based methods, KKZ tends to outperform K-means clustering.

Given the large number of climate and hydrological variables under consideration and the challenges inherent in communicating information about multi-dimensional data, two summary criteria are used to generalize subset coverage results in this study. The first criterion is the average PSC for all climate or hydrological variables. Following Cannon (2015), the second criterion is the percentage of variables that reach a 90% PSC threshold (PSC90p).

Figure 7 presents the average PSC and PSC90p for climate variables (solid lines) and hydrological variables (dashed lines) when selected subsets contain K simulations (K = 1 to 50) over the two watersheds. Generally, the KKZ method performs better than K-means clustering for both evaluation criteria and both watersheds, and the two automated envelope-based methods perform better than random selection. In the case of the Xiangjiang watershed (Fig. 7a), the ~~9~~ 10 simulation KKZ subset reaches an average PSC of 90% for climate variables, while K-means and random selection require ~~19~~ 23 and ~~27~~ 29 simulations, respectively, to reach this threshold. For hydrological variables, the KKZ method still shows the best performance. To reach an average PSC of 90%, KKZ and K-means clustering require at least ~~10~~ 13 and ~~38~~ 26 simulations, respectively. In contrast to results for the climate variables, K-means clustering ~~only~~ even performs worse~~better~~ than random selection for the hydrological variables ~~when the subset size does not exceed 29~~(e.g. K = 27 to 32). In the case of the Manicouagan 5 watershed (Fig. 7c), the KKZ method again outperforms K-means clustering and random selection.

In addition, as more simulations are selected, the average PSC increases rapidly when the size of selected simulations is smaller than 10 for both watersheds, while the rate of increase slows when the number is larger than 10. For the KKZ method, a subset of 10 simulations covers more than 85% of uncertainty for climate and hydrology variables in both watersheds; selecting more than 10 climate simulations leads to little change in uncertainty coverage. For ~~these two~~both watersheds, a subset of 10 simulations selected using KKZ appears to be optimal for reducing computational costs while incurring the smallest possible loss of uncertainty information. In addition, the performance of the KKZ method is maintained for larger subsets, while the performance of K-means clustering fluctuates. In other words, a larger subset selected by the K-means clustering may not have a greater uncertainty coverage than a smaller subset. The recursive nature of the KKZ method effectively guarantees that average PSC increases monotonically with subset size.

The focus of this study is the transferability of climate simulation uncertainty to uncertainty in hydrological impacts. For a given method this can be inferred from the difference in average PSC and PSC90p between climate and hydrological variables. For Xiangjiang watershed, the average PSC of climate variables is close to that of hydrological variables for all selection methods (Fig. 7a). Especially for the KKZ method, differences in average PSC are less than 5% (with the exception of K = 2, 11 and 12 ~~and 9~~). The differences in climate and hydrology uncertainty coverage are slightly larger when using the K-means clustering and random selection methods. For PSC90p (Fig. 7b), transferability is somewhat less apparent due to the more rigorous 90% PSC threshold. Although differences in PSC90p between climate and hydrological variables are sometimes large, especially for the K-means clustering, the PSC90p of hydrological variables still exhibits similar overall tendency and behavior as the climate variables. In general, subsets of climate simulations that are selected based on a large number of relevant climate variables are effective at transferring uncertainty coverage into the realm of hydrological impacts. However, this transferability is method dependent; results are less variable and more consistent for KKZ than K-Means clustering.

Figure 7c-d presents results for average PSC and PSC90p in the Manicouagan 5 watershed. On the whole, the selection methods behave similarly in terms of transferability as in the Xiangjiang watershed, but the uncertainty coverage of the subsets for the hydrological variables is reduced slightly. Degraded transferability is most apparent in larger differences in PSC90p between the climate and hydrological variables. As noted above, however, this criterion is much more stringent than average PSC.

4.~~2~~ 3 Impact of Temperature Variables

The climate variables used in this study can be classified into two groups: temperature variables and precipitation variables. Each variable is given equal weight in the subset selection, regardless of inter-variable correlations, and all variables are assumed to exert the same influence on the hydrological variables. However, the impacts of climate variables on flow regimes may not be the same in watersheds with different hydroclimatic characteristics. For example, warmer temperatures lead to earlier spring floods in northern seasonally snow-covered watersheds (such as the Manicouagan 5 watershed) (Whitfield and Cannon, 2000; Chen et al., 2011b; Minville et al., 2008), whereas changes in temperature have little impact on the timing of floods in rainfall-dominated watersheds (such as the Xiangjiang watershed). Since the importance of temperature is different for the two study watersheds, a question is raised: Can the transferability of climate uncertainty in Xiangjiang watershed be improved if irrelevant temperature variables are removed? To answer this question, temperature variables (the first 16 variables in Table 2) were removed and subset selection was conducted again using the 15 precipitation variables. The average PSC and PSC90p were then calculated to compare with original results that includes temperature variables. Results from the precipitation analysis are shown in Fig. 8.

For Xiangjiang watershed (Fig. 8a-b), removing temperature variables from the subset selection leads to improved uncertainty coverage for the hydrological variables, especially for K-means clustering. The K-means clustering now performs better than random selection in most cases. For KKZ, average PSC for the hydrological variables ~~exceeds~~ reaches 90% with a subset of only ~~4~~ 6 simulations, whereas the same level of coverage required ~~9~~ 13 simulations when considering both temperature and precipitation. However, the effect of removing temperature variables is the opposite for the Manicouagan 5 watershed (Fig. 8c-d). Here, coverage performance for the hydrological variables is reduced ~~substantially~~ when temperature variables are not considered. The contrasting effects are consistent with the processes that generate runoff in the two watersheds. As mentioned above, the Manicouagan 5 watershed is seasonally snow-covered – snow accumulation and snowmelt are the dominant processes that contribute to runoff generation – and hence it is sensitive to changes in temperature. However, temperature variables are not relevant in the rainfall-dominated Xiangjiang watershed. The different impacts of temperature variables in the two watersheds highlights the necessity of carefully choosing climate variables for subset selection based on physical process knowledge.

4.~~3~~ 4 Transferability of Multi-model Mean

In addition to the overall spread in the projected climate change signal, policymakers are also concerned with the MME mean when communicating hydrological climate change impacts. Therefore, the selection methods are also evaluated in terms of their ability to preserve the multi-model mean of the full MME. It bears noting that the CMIP5 MME considered in this study is an ensemble of opportunity. Models are not statistically independent, for example due to shared physical parameterizations, and multiple simulations may be contributed by the same model. Also, the two envelope-based methods make very different assumptions about the underlying nature of the statistical distribution of the ensemble. The KKZ method is not biased towards high-density regions of the multivariate space, preferring uniform coverage, whereas the K-means method, which assumes a mixture of multivariate normal clusters with equal variance, will tend to select simulations that lie in regions populated by a large number of simulations. These characteristics will have implications for preservation of the MME mean.

In order to generalize the MME mean over multiple variables, standardized changes in each variable are averaged across variables and selected simulations to obtain a dimensionless criterion (referred to as averaged standardized mean change). For different sized subsets selected by the three selection methods, corresponding climate and hydrological averaged standardized mean changes were calculated and compared with values for the whole ensemble. Because projected changes are preprocessed by standardizing to zero mean and unit standard deviation, the averaged standardized mean change of the whole ensemble is zero by construction. Therefore, if the averaged standardized mean change of a subset is close to zero, the MME mean change simulated by that subset is similar to that simulated by the entire ensemble. Figure 9 shows the averaged standardized mean changes in climate and hydrological variables when K simulations (K = 1 to 50) are selected for the two watersheds. When averaged over a large number of random trials, mean values will, by definition, lie close to zero for the random selection method; thus, the envelope of results across all 100 random selections are presented as blue and pink shaded areas in each subfigure for climate and hydrological variables, respectively. Figure 9a-b presents results for subsets when temperature variables are included in the selection process, whereas Fig. 9c-d presents results when temperature variables are excluded. Overall, when gauged against the range of variability in the 100 random selections, subsets selected by both statistical methods perform well in reproducing the MME mean of the entire ensemble, with K-means clustering performing slightly better than the KKZ method. When looked at in more detail, in the Xiangjiang watershed, the averaged standardized mean changes of subsets in climate variables tend to differ slightly from those in hydrological variables when temperature variables are included (Fig. 9a). For example, when 5 simulations are selected using the KKZ method, the averaged standardized mean change for climate variables is 0.18 whereas it is -0.63 for hydrological variables. Subsets selected by the KKZ method often have higher means than the whole ensemble for climate variables, while they have lower values for hydrological variables. In other words, a subset with positive changes in climate variables gives negative changes in hydrological variables, which means that selected subsets have poor transferability in terms of MME mean. However, when temperature variables are not included in the selection process, ~~performance and~~ the transferability of multi-model mean is~~are both~~ improved (Fig. ~~9b~~9c). In the Manicouagan 5 watershed, by contrast, differences between average changes in climate variables and hydrological variables are slightly smaller when temperature variables are included (Fig. ~~9c~~9b,d). Again, this highlights the importance of selecting the appropriate climate variables when performing ensemble subset selection.

5 Discussion

In order to recommend a practical subset of climate simulations for end-users who deal with the assessment of climate change impacts on hydrology, various selection methods have ~~be~~been proposed based on different criteria (Mendlik and Gobiet, 2016; Cannon, 2015; Gleckler et al., 2008; Lutz et al., 2016a; McSweeney et al., 2012; Warszawski et al., 2014; Perkins et al., 2007). Even though these methods usually perform well in terms of the climate variables to which they are applied, their performance in terms of hydrological impacts needs to be verified. In normal usage, for example, envelope-based methods may only consider changes in mean temperature and annual precipitation (Immerzeel et al., 2013; Murdock and Spittlehouse, 2011; Warszawski et al., 2014), which will have a strong influence both on the overall measurement of climate uncertainty and subset selection results in terms of hydrological impacts. By not considering relevant climate variables, there may be a loss of information when transferring climate uncertainty to hydrological uncertainty (Chen et al., 2016). When one considers the fact that it is often hard to determine a one-to-one correspondence between climate and hydrological variables, it may be reasonable to use a large suite of climate variables. ~~However, this may result in the inclusion of irrelevant and redundant variables that could compromise performance.~~

Therefore, this study investigated the transferability of climate simulation uncertainty to the hydrological world by K-means clustering and KKZ methods using a large number of climate and hydrological variables, including both seasonal and annual means and extremes. Multiple variables, when selected carefully, can improve the transferability of climate simulation uncertainty to hydrology impacts. Although the introduction of multiple climate variables may lead to redundant information, and it may be unnecessary for impact studies with different aims (e.g. water balance or hydrological drought) to consider so many climate extreme indices, this general approach can nonetheless give a more useful and reasonable selection for the purpose of covering an overall range of future climate change and its hydrological impacts.

This study also evaluated the impact of variable selection by comparing uncertainty transferability in a rainfall-dominated watershed and a seasonally snow-covered watershed when including and excluding temperature variables. The different impacts of temperature variables over two watersheds indicate that climate variables, if not chosen with consideration of runoff generating processes, can affect the performance of the subset selection algorithms. In the rainfall-dominated Xiangjiang watershed, inclusion of temperature variables, which play little role in generating runoff, leads to a small loss of performance, whereas in the snow-related Manicouagan 5 watershed, exclusion of temperature variables results~~ed~~ in a large loss of performance. This is reflected in results both for ensemble spread and MME mean. Thus, it is important to choose proper climate variables that characterize the physical processes controlling hydrology of the watershed for subset selection. Both watersheds in this study are located in humid regions with moderately hilly topography. The proper climate variables for watersheds in other climate regions may need to be modified accordingly. For example, for high-mountain regions, precipitation may be influenced by complex topography and snowmelt often has the greatest contribution to runoff, and thus it is recommended that climate variables related to orographic precipitation and the evolution of snowpack be included in the selection process (Cannon et al., 2017; Immerzeel et al., 2012). Additionally, for arid regions, Hortonian overland flow may be the predominant runoff mechanism, and thus precipitation variables that are capable of describing the intensity of precipitation events may need to be stressed (Pilgrim et al., 1988). ~~Although the results emphasize the impacts of temperature variables~~However, the judgement on relevant climate variables in this study is somewhat subjective. Some automated variable selection procedure may provide a more objective selection on relevant climate variables, such as redundancy analysis or multivariate sparse group lasso (Li et al., 2015).

In terms of ~~methodology~~selection methods, the results of this study reveal two strengths of the KKZ method over K-means clustering. First, the KKZ method selects simulations on the boundaries of the climate simulation ensemble and, as a result, it is better able to cover overall climate uncertainty, as measured by average PSC and PSC90p, of the ensemble than K-means clustering. Second, uncertainty coverage of the KKZ method for climate variables increases monotonically as more climate simulations are selected, whereas the K-means clustering is unstable. This is because climate simulations are added incrementally, in a recursive fashion, by the KKZ method as subset size increases, whereas K-means clustering needs to be run independently for each subset. Consequently, K-means clustering produces a disordered sequence of solutions. The results of this study show that these two strengths of the KKZ method are retained for hydrological impacts. Therefore, in the aspect of overall uncertainty coverage, the KKZ method outperforms K-means clustering. Performance in terms of MME mean were also evaluated in this study. Results show that the subsets selected by K-mean clustering produce a more similar MME mean to the whole ensemble, although differences between the two methods are small. This result is expected because K-means clustering selects representative simulations for each cluster according to their closeness to the cluster centroid, which is the multivariate mean.

The two envelope-based methods in this study are from a single branch of selection methods whose purpose is to cover the spread (uncertainty) in projected changes of an ensemble. The model ranking approach is another common way to select model simulations, usually based on historical model performance, measures of statistical independence, and other evaluation metrics. Some studies have investigated the impact of weighting GCMs on the projection of climate conditions or hydrological impacts (Chen et al., 2017; Christensen et al., 2010). They concluded that weighting methods have little influence on the ensemble mean and uncertainty, and it is more appropriate to consider GCMs as being equiprobable.

Some studies have argued that certain GCMs may not be independent from one another because of shared code or parameterization schemes (Evans et al., 2013; Knutti et al., 2010; Mendlik and Gobiet, 2016). In an ensemble of opportunity like CMIP5, this dependence may lead to high density regions of climate variable space and hence influence the selection of models by methods like K-means clustering. On the other hand, the KKZ method is designed to select simulations that lie on the edges of the ensemble. If these simulations are outliers because their projections are not credible, for example due to poor process representation, and then their selection may not be warranted. Therefore, previously removing any obviously dependent or ill-behaving GCMs through model weighting methods may improve the rationality of these two equal-weighting selection methods in regional impact studies.

In this study, only one downscaling method was used to generate climate scenarios in the scale of watershed. In order to consider different downscaling methods, the Quantile Mapping (QM) approach (Maurer and Pierce, 2014; Piani et al., 2010) was additionally examined. Figure 10 presents the results of average PSC in this case. Compared with the results of DS method, the overall character of the results is roughly the same. Thus, the choice of a downscaling method may have little influence on the conclusions of this study. In addition, only two emission scenarios (RCP4.5 and RCP8.5) were considered. 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 suggested 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 centres as RCP4.5 and RCP8.5. Thus, RCP4.5 and RCP8.5 were used to include a range of realistic projections (Lutz et al., 2016b). Due to unknown future emission scenarios, two concentration pathways were used in an undifferentiated manner to cover uncertainty resulting from emission scenarios. However, the two emission scenarios do generate different climate simulations. Thus, each pathway was also separately input into the subset selection to examine the transferability when only one scenario was considered. Figure 11 shows the average PSC for either emission scenario over the Xiangjiang watershed. The main character of the results is the same as the original research where 2 RCPs were considered and selection of 5 or 6 climate simulations by the KKZ method can cover an adequate uncertainty range. Therefore, the specific choice of emission scenarios can be decided by end-users according to their own needs.

6 Conclusions

In this study, the transferability of climate simulation uncertainty to climate change impacts on hydrology was investigated over two watersheds with different climate and hydrological regimes based on multiple climate variables. Main conclusions are summarized as follows:

(1) In terms of uncertainty coverage, both the KKZ method and K-means clustering are effective at selecting subsets of climate simulations that represent the range of the climate change signal. However, when it comes to hydrological impacts, the KKZ method always performed better than random selection, while K-means clustering sometimes performed worse than random selection.

(2) Both K-means clustering and the KKZ method are capable of reproducing the MME mean of the whole ensemble, although K-means clustering performed slightly better than the KKZ method in some cases.

(3) The uncertainty of climate simulations based on multiple climate variables can be transferred to the assessment of hydrological impacts uncertainty. In other words, selected subsets can generate similar uncertainty coverage in terms of both climate simulation and hydrological impacts.

(4) In order to cover an adequate range of climate simulation and hydrological impacts uncertainty with less computational costs, selection of about 10 simulations from the ensemble of 50 simulations is required. Little improvement is gained when the number of simulations is increased beyond 10.

(5) The choice of climate variables affects the transferability of climate uncertainty to hydrological uncertainty. Thus, the climate and hydrological regimes of a watershed should be considered when choosing variables used to subset climate model simulations for hydrological impact studies.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (Grant No. 51779176, 51339004, 51539009), ~~and~~ the Thousand Youth Talents Plan from the Organization Department of CCP Central Committee (Wuhan University, China) and the Research Council of Norway (FRINATEK Project 274310). The authors would like to acknowledge the contribution of the World Climate Research Program Working Group on Coupled Modelling, and all climate modeling groups listed in Table 1 for making available their respective model outputs. The authors would also like to acknowledge Hydro-Québec and the Changjiang Water Resources Commission for providing observation data in the Manicouagan 5 and Xiangjiang watersheds, respectively.

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.

Arsenault, R., Gatien, P., Renaud, B., Brissette, F., and Martel, J.-L.: A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation, Journal of Hydrology, 529, 754-767, https://doi.org/10.1016/j.jhydrol.2015.09.001, 2015.

Cannon, A. J.: Selecting GCM Scenarios that Span the Range of Changes in a Multimodel Ensemble: Application to CMIP5 Climate Extremes Indices*, Journal of Climate, 28, 1260-1267, https://doi.org/10.1175/jcli-d-14-00636.1, 2015.

Cannon, F., Carvalho, L. M. V., Jones, C., Norris, J., Bookhagen, B., and Kiladis, G. N.: Effects of topographic smoothing on the simulation of winter precipitation in High Mountain Asia, Journal of Geophysical Research: Atmospheres, 122, 1456-1474, https://doi.org/10.1002/2016jd026038, 2017.

Chen, J., Brissette, F. P., and Leconte, R.: Uncertainty of downscaling method in quantifying the impact of climate change on hydrology, Journal of Hydrology, 401, 190-202, https://doi.org/10.1016/j.jhydrol.2011.02.020, 2011a.

Chen, J., Brissette, F. P., Poulin, A., and Leconte, R.: Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed, Water Resources Research, 47, W12509, https://doi.org/10.1029/2011wr010602, 2011b.

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.

Chen, J., Brissette, F. P., and Lucas-Picher, P.: Transferability of optimally-selected climate models in the quantification of climate change impacts on hydrology, Climate Dynamics, 47, 3359-3372, https://doi.org/10.1007/s00382-016-3030-x, 2016.

Chen, J., Brissette, F. P., Lucas-Picher, P., and Caya, D.: Impacts of weighting climate models for hydro-meteorological climate change studies, Journal of Hydrology, 549, 534-546, https://doi.org/10.1016/j.jhydrol.2017.04.025, 2017.

Chiew, F. H. S., Teng, J., Vaze, J., Post, D. A., Perraud, J. M., Kirono, D. G. C., and Viney, N. R.: Estimating climate change impact on runoff across southeast Australia: Method, results, and implications of the modeling method, Water Resources Research, 45, https://doi.org/10.1029/2008wr007338, 2009.

Christensen, J. H., Kjellström, E., Giorgi, F., Lenderink, G., and Rummukainen, M.: Weight assignment in regional climate models, Climate Research, 44, 179-194, https://doi.org/10.3354/cr00916, 2010.

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resources Research, 28, 1015-1031, https://doi.org/10.1029/91WR02985, 1992.

Edijatno, De Oliveira Nascimento, N., Yang, X., Makhlouf, Z., and Michel, C.: GR3J: a daily watershed model with three free parameters, Hydrological Sciences Journal, 44, 263-277, https://doi.org/10.1080/02626669909492221, 1999.

Eum, H.-I., Dibike, Y., and Prowse, T.: Climate-induced alteration of hydrologic indicators in the Athabasca River Basin, Alberta, Canada, Journal of Hydrology, 544, 327-342, https://doi.org/10.1016/j.jhydrol.2016.11.034, 2017.

Evans, J. P., Ji, F., Abramowitz, G., and Ekstrom, M.: Optimally choosing small ensemble members to produce robust climate simulations, Environmental Research Letters, 8, 044050, https://doi.org/10.1088/1748-9326/8/4/044050, 2013.

Giorgi, F., and Mearns, L. O.: Calculation of Average, Uncertainty Range, and Reliability of Regional Climate Changes from AOGCM Simulations via the "Reliability Ensemble Averaging" (REA) Method, Journal of Climate, 15, 1141-1158, https://doi.org/10.1175/1520-0442(2002)015<1141:coaura>2.0.co;2, 2002.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, Journal of Geophysical Research, 113, D06104, https://doi.org/10.1029/2007jd008972, 2008.

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.

Hartigan, J. A., and Wong, M. A.: Algorithm AS 136: A K-Means Clustering Algorithm, Journal of the Royal Statistical Society. Series C (Applied Statistics), 28, 100-108, https://doi.org/10.2307/2346830, 1979.

Houle, D., Bouffard, A., Duchesne, L., Logan, T., and Harvey, R.: Projections of Future Soil Temperature and Water Content for Three Southern Quebec Forested Sites, Journal of Climate, 25, 7690-7701, https://doi.org/10.1175/jcli-d-11-00440.1, 2012.

Hutchinson, M. F., McKenney, D. W., Lawrence, K., Pedlar, J. H., Hopkinson, R. F., Milewska, E., and Papadopol, P.: Development and Testing of Canada-Wide Interpolated Spatial Models of Daily Minimum–Maximum Temperature and Precipitation for 1961–2003, Journal of Applied Meteorology and Climatology, 48, 725-741, https://doi.org/10.1175/2008jamc1979.1, 2009.

Immerzeel, W. W., van Beek, L. P., Konz, M., Shrestha, A. B., and Bierkens, M. F.: Hydrological response to climate change in a glacierized catchment in the Himalayas, Clim Change, 110, 721-736, https://doi.org/10.1007/s10584-011-0143-4, 2012.

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.

Katsavounidis, I., Jay Kuo, C. C., and Zhen, Z.: A new initialization technique for generalized Lloyd iteration, IEEE Signal Processing Letters, 1, 144-146, https://doi.org/10.1109/97.329844, 1994.

Klein, S. A., and Hall, A.: Emergent Constraints for Cloud Feedbacks, Current Climate Change Reports, 1, 276-287, https://doi.org/10.1007/s40641-015-0027-1, 2015.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A.: Challenges in Combining Projections from Multiple Climate Models, Journal of Climate, 23, 2739-2758, https://doi.org/10.1175/2009jcli3361.1, 2010.

Li, Y., Nan, B., and Zhu, J.: Multivariate sparse group lasso for the multivariate multiple linear regression with an arbitrary group structure, Biometrics, 71, 354-363, https://doi.org/10.1111/biom.12292, 2015.

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.

Logan, T., Charron, I., Chaumont, D., and Houle, D.: Atlas of climate scenarios for Québec forests, uranos for Ministère des Res-sources naturelles and de la Faune du Québec Techical Report ISBN 978-2-923292-12-0, 1-132, 2011.

Lutz, A. F., ter Maat, H. W., Biemans, H., Shrestha, A. B., Wester, P., and Immerzeel, W. W.: Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach, International Journal of Climatology, 36, 3988-4005, https://doi.org/10.1002/joc.4608, 2016a.

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, 2016b.

Maurer, E. P., and Pierce, D. W.: Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean, Hydrology and Earth System Sciences, 18, 915-925, https://doi.org/10.5194/hess-18-915-2014, 2014.

McSweeney, C. F., Jones, R. G., and Booth, B. B. B.: Selecting Ensemble Members to Provide Regional Climate Change Information, Journal of Climate, 25, 7100-7121, https://doi.org/10.1175/jcli-d-11-00526.1, 2012.

Mehran, A., AghaKouchak, A., and Phillips, T. J.: Evaluation of CMIP5 continental precipitation simulations relative to satellite-based gauge-adjusted observations, Journal of Geophysical Research: Atmospheres, 119, 1695-1707, https://doi.org/10.1002/2013jd021152, 2014.

Mendlik, T., and Gobiet, A.: Selecting climate simulations for impact studies based on multivariate patterns of climate change, Climatic Change, 135, 381-393, https://doi.org/10.1007/s10584-015-1582-0, 2016.

Minville, M., Brissette, F., and Leconte, R.: Uncertainty of the impact of climate change on the hydrology of a nordic watershed, Journal of Hydrology, 358, 70-83, https://doi.org/10.1016/j.jhydrol.2008.05.033, 2008.

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.

Mu, X., Zhang, L., McVicar, T. R., Chille, B., and Gau, P.: Analysis of the impact of conservation measures on stream flow regime in catchments of the Loess Plateau, China, Hydrological Processes, 21, 2124-2134, https://doi.org/10.1002/hyp.6391, 2007.

Murdock, T., and Spittlehouse, D.: Selecting and using climate change scenarios for British Columbia, Pacific Climate Impacts Consortium, University of Victoria, Victoria, BC, 1-39, 2011.

Muzik, I.: Sensitivity of Hydrologic Systems to Climate Change, Canadian Water Resources Journal, 26, 233-252, https://doi.org/10.4296/cwrj2602233, 2001.

Nash, J. E., and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282-290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model?, Journal of Hydrology, 303, 290-306, https://doi.org/10.1016/j.jhydrol.2004.08.026, 2005.

Perkins, S. E., Pitman, A. J., Holbrook, N. J., and McAneney, J.: Evaluation of the AR4 Climate Models' Simulated Daily Maximum Temperature, Minimum Temperature, and Precipitation over Australia Using Probability Density Functions, Journal of Climate, 20, 4356-4376, https://doi.org/10.1175/jcli4253.1, 2007.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology, 279, 275-289, https://doi.org/10.1016/s0022-1694(03)00225-7, 2003.

Piani, C., Weedon, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S., and Haerter, J. O.: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models, Journal of Hydrology, 395, 199-215, https://doi.org/10.1016/j.jhydrol.2010.10.024, 2010.

Pierce, D. W., Barnett, T. P., Santer, B. D., and Gleckler, P. J.: Selecting global climate models for regional climate change studies, Proceedings of the National Academy of Sciences, 106, 8441-8446, 2009.

Pilgrim, D. H., Chapman, T. G., and Doran, D. G.: Problems of rainfall-runoff modelling in arid and semiarid regions, Hydrological Sciences Journal, 33, 379-400, https://doi.org/10.1080/02626668809491261, 1988.

Pincus, R., Batstone, C. P., Hofmann, R. J. P., Taylor, K. E., and Glecker, P. J.: Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models, Journal of Geophysical Research, 113, https://doi.org/10.1029/2007jd009334, 2008.

Reifen, C., and Toumi, R.: Climate projections: Past performance no guarantee of future skill?, Geophysical Research Letters, 36, https://doi.org/10.1029/2009gl038082, 2009.

Richter, B. D., Baumgartner, J. V., Powell, J., and Braun, D. P.: A Method for Assessing Hydrologic Alteration within Ecosystems, Conservation Biology, 10, 1163-1174, https://doi.org/10.1046/j.1523-1739.1996.10041163.x, 1996.

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.

Shrestha, R. R., Peters, D. L., and Schnorbus, M. A.: Evaluating the ability of a hydrologic model to replicate hydro-ecologically relevant indicators, Hydrological Processes, 28, 4294-4310, https://doi.org/10.1002/hyp.9997, 2014.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93, 485-498, https://doi.org/10.1175/bams-d-11-00094.1, 2012.

Tebaldi, C., and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365, 2053-2075, https://doi.org/10.1098/rsta.2007.2076, 2007.

Valéry, A., Andréassian, V., and Perrin, C.: 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2 – Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments, Journal of Hydrology, 517, 1176-1187, https://doi.org/10.1016/j.jhydrol.2014.04.058, 2014.

Vaze, J., and Teng, J.: Future climate and runoff projections across New South Wales, Australia: results and practical applications, Hydrological Processes, 25, 18-35, https://doi.org/10.1002/hyp.7812, 2011.

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., and Schewe, J.: The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): project framework, Proceedings of the National Academy of Sciences, 111, 3228-3232, https://doi.org/10.1073/pnas.1312330110, 2014.

Whitfield, P. H., and Cannon, A. J.: Recent Variations in Climate and Hydrology in Canada, Canadian Water Resources Journal, 25, 19-65, https://doi.org/10.4296/cwrj2501019, 2000.

Wilby, R. L.: Uncertainty in water resource model parameters used for climate change impact assessment, Hydrological Processes, 19, 3201-3219, https://doi.org/10.1002/hyp.5819, 2005.

Wilcke, R. A. I., and Bärring, L.: Selecting regional climate scenarios for impact modelling studies, Environmental Modelling & Software, 78, 191-201, https://doi.org/10.1016/j.envsoft.2016.01.002, 2016.

Xu, H., Xu, C.-Y., Chen, H., Zhang, Z., and Li, L.: Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of China, Journal of Hydrology, 505, 1-12, https://doi.org/10.1016/j.jhydrol.2013.09.004, 2013.

Zeng, Q., Chen, H., Xu, C.-Y., Jie, M.-X., and Hou, Y.-K.: Feasibility and uncertainty of using conceptual rainfall-runoff models in design flood estimation, Hydrology Research, 47, 701-717, https://doi.org/10.2166/nh.2015.069, 2016.

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., Trewin, B., and Zwiers, F. W.: Indices for monitoring changes in extremes based on daily temperature and precipitation data, Wiley Interdisciplinary Reviews: Climate Change, 2, 851-870, https://doi.org/10.1002/wcc.147, 2011.

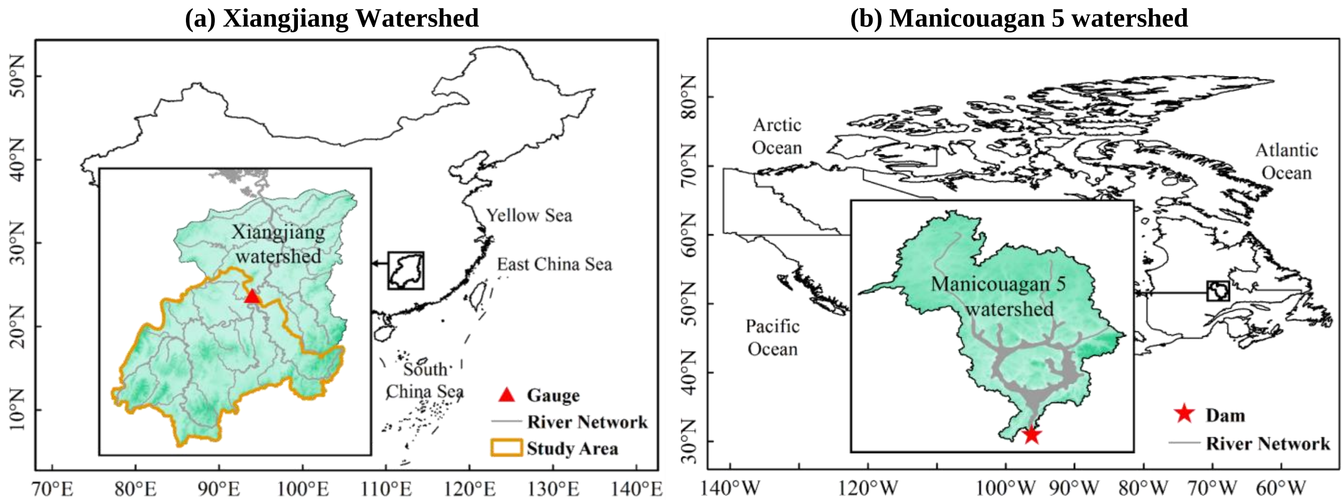

Figure 1: Location maps of the (a) Xiangjiang and (b) Manicouagan 5 watersheds **(The study area in the Xiangjiang watershed is one of its sub-basins as the orange boundary shows).**

Table 1: Basic information about the CMIP5 models

Institution	Model name	Resolution (Lon. × Lat.)
Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia	ACCESS1.0	1.875 × 1.25
	ACCESS1.3	1.875 × 1.25
Beijing Climate Center, China Meteorological Administration	BCC-CSM1.1	2.8 ×2.8
	BCC-CSM1.1(m)	1.125 ×1.125
College of Global Change and Earth System Science, Beijing Normal University	BNU-ESM	2.8 × 2.8
Canadian Centre for Climate Modelling and Analysis	CanESM2	2.8 × 2.8
Centro Euro-Mediterraneo per I Cambiamenti Climatici	CMCC-CMS	1.875 × 1.875
	CMCC-CM	0.75 × 0.75
	CMCC-CESM	3.75 × 3.7
Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancée en Calcul Scientifique	CNRM-CM5	1.4 × 1.4
Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence	CSIRO-Mk3.6.0	1.8 ×1.8
LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences; and CESS, Tsinghua University	FGOALS-g2	1.875 × 1.25
NOAA Geophysical Fluid Dynamics Laboratory	GFDL-CM3	2.5 × 2.0
	GFDL-ESM2G	2.5 × 2.0
	GFDL-ESM2M	2.5 × 2.0
Institute for Numerical Mathematics	INM-CM4	2.0 × 1.5
Institut Pierre-Simon Laplace	IPSL-CM5A-LR	3.75 × 1.9
	IPSL-CM5A-MR	2.5 × 1.25
	IPSL-CM5B-LR	3.75 × 1.9
Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	MIROC-ESM-CHEM	2.8 × 2.8
	MIROC-ESM	2.8 × 2.8
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	MIROC5	1.4 × 1.4
Max Planck Institute for Meteorology (MPI-M)	MPI-ESM-LR	1.875 × 1.875
	MPI-ESM-MR	1.875 × 1.875
Meteorological Research Institute	MRI-ESM1	1.125 × 1.125
	MRI-CGCM3	1.1 × 1.1

Table 2: Definitions of 31 climate variables. The final column indicates whether the change in a given variable is expressed in the form of relative difference (CT = change type)

Category	Index	Description	CT
ETCCDI extreme indices	TXx	Annual maximum value of daily maximum temperature	
	TXn	Annual minimum value of daily maximum temperature	
	TNx	Annual maximum value of daily minimum temperature	
	TNn	Annual minimum value of daily minimum temperature	
	TX10p	Percentage of days when daily max temperature < 10th percentile	
	TX90p	Percentage of days when daily max temperature > 90th percentile	
	TN10p	Percentage of days when daily min temperature < 10th percentile	
	TN90p	Percentage of days when daily min temperature > 90th percentile	
	WSDI	Warm spell duration index: Annual count of days with at least 6 consecutive days when TX>90th percentile	
	CSDI	Cold spell duration index: Annual count of days with at least 6 consecutive days when TN<10th percentile	
	DTR	Daily temperature range: Monthly mean difference between daily max and min temperature	
Seasonal or annual mean indices	Tav	Annual average temperature	
	SpT	Seasonal average temperature in spring	
	SuT	Seasonal average temperature in summer	
	AuT	Seasonal average temperature in autumn	
	WiT	Seasonal average temperature in winter	
ETCCDI extreme indices	R1mm	Annual count of days when precipitation ≥ 1mm	
	R10mm	Annual count of days when precipitation ≥ 10mm	
	R20mm	Annual count of days when precipitation ≥ 20mm	
	CDD	~~Maximum length of dry spell~~Consecutive dry days~~,~~: maximum number of consecutive days with daily precipitation < 1mm	
	CWD	Consecutive wet days~~Maximum length of wet spell,~~: maximum number of consecutive days with daily precipitation ≥ 1mm	
	Rx1day	Annual maximum 1-day precipitation	%
	Rx5day	Annual maximum consecutive 5-day precipitation	%
	SDII	Simple precipitation intensity index	%
	R95pTOT	Annual total precipitation when daily precipitation > 95th quantile	%
	R99pTOT	Annual total precipitation when daily precipitation > 99th quantile	%
Seasonal or annual mean indices	ARav	Annual total precipitation	%
	SpR	Seasonal total precipitation in spring	%
	SuR	Seasonal total precipitation in summer	%
	AuR	Seasonal total precipitation in autumn	%
	WiR	Seasonal total precipitation in winter	%

Table 43: Definitions of 17 hydrological variables. The final column indicates whether the change in a given variable is expressed in the form of relative difference (CT = change type)

Category	Index	Description	CT
Water Resources Indicators (WRIs)	MD	Annual mean flow	%
	SpMD	Seasonal mean flow in spring	%
	SuMD	Seasonal mean flow in summer	%
	AuMD	Seasonal mean flow in autumn	%
	WiMD	Seasonal mean flow in winter	%
	tCMD	Centre of timing of annual flow	
Quantiles of daily flow	Q5	5th quantile of daily flow series	%
	Q50	50th quantile of daily flow series	%
	Q95	95th quantile of daily flow series	%
Indicators of Hydrological Alteration (IHAs)	Qx1day	Annual mean 1-day maximum flow	%
	Qx3day	Annual mean 3-day maximum flow	%
	Qx7day	Annual mean 7-day maximum flow	%
	tQx	Julian date of annual 1-day maximum	
	LPC	Number of low pulses (annual median -25th percentile) in a year	
	HPC	Number of high pulses (annual median +25th percentile) in a year	
	LPD	Mean duration of low pulses in a year	%
	HPD	Mean duration of high pulses in a year	%

Table 4: Nash-Sutcliffe Efficiency (NSE) of hydrological models in the calibration and validation over two watersheds

Country	Watershed name	Area (km^2)	Hydrological Model	Calibration period	NSE calibration	RE calibration	Validation period	NSE validation	RE validation
China	Xiangjiang	52150	GR4J-6	1975-1987	0.912	-0.3%	1988-2000	0.871	5.4%
			GR4J	1975-1987	0.912	-0.2%	1988-2000	0.872	5.5%
Canada	Manicouagan 5	24610	GR4J-6	1970-1979	0.926	3.8%	1980-1989	0.881	2.7%

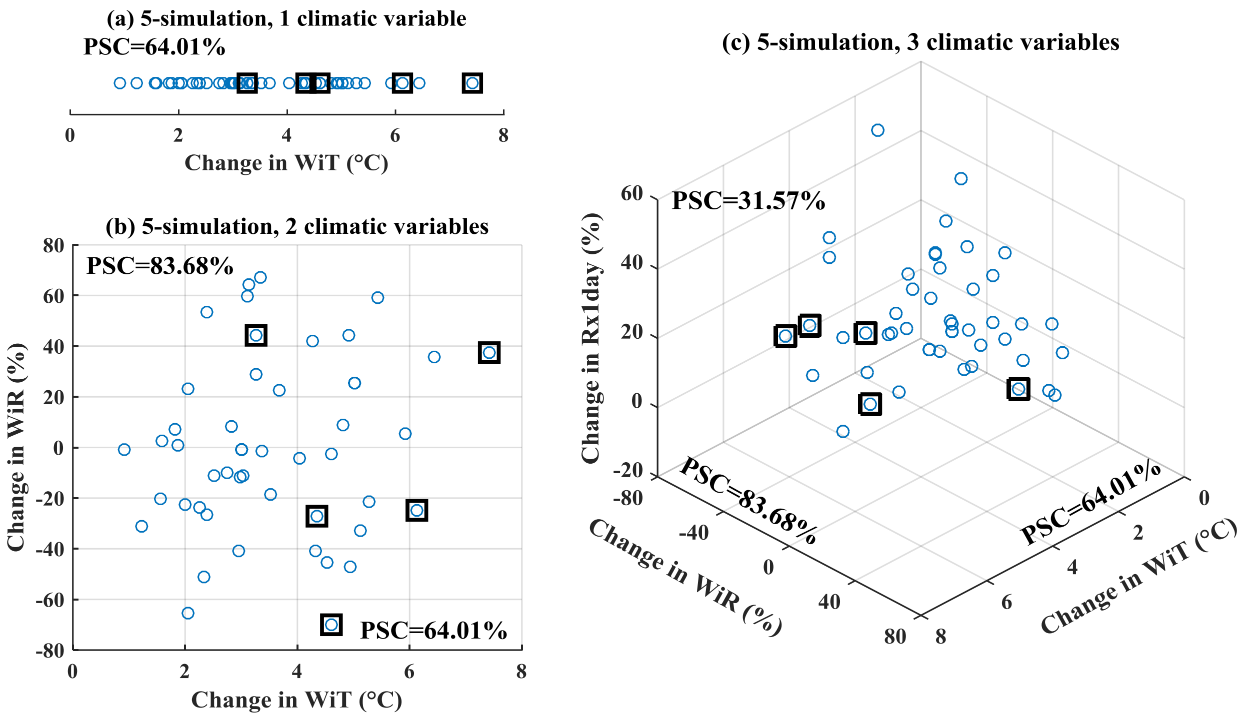

Figure ~~3~~2: Examples of PSCs when selecting 5 climate simulations over the Xiangjiang watershed using the KKZ method. The PSCs of each variable are presented beside the corresponding axes.

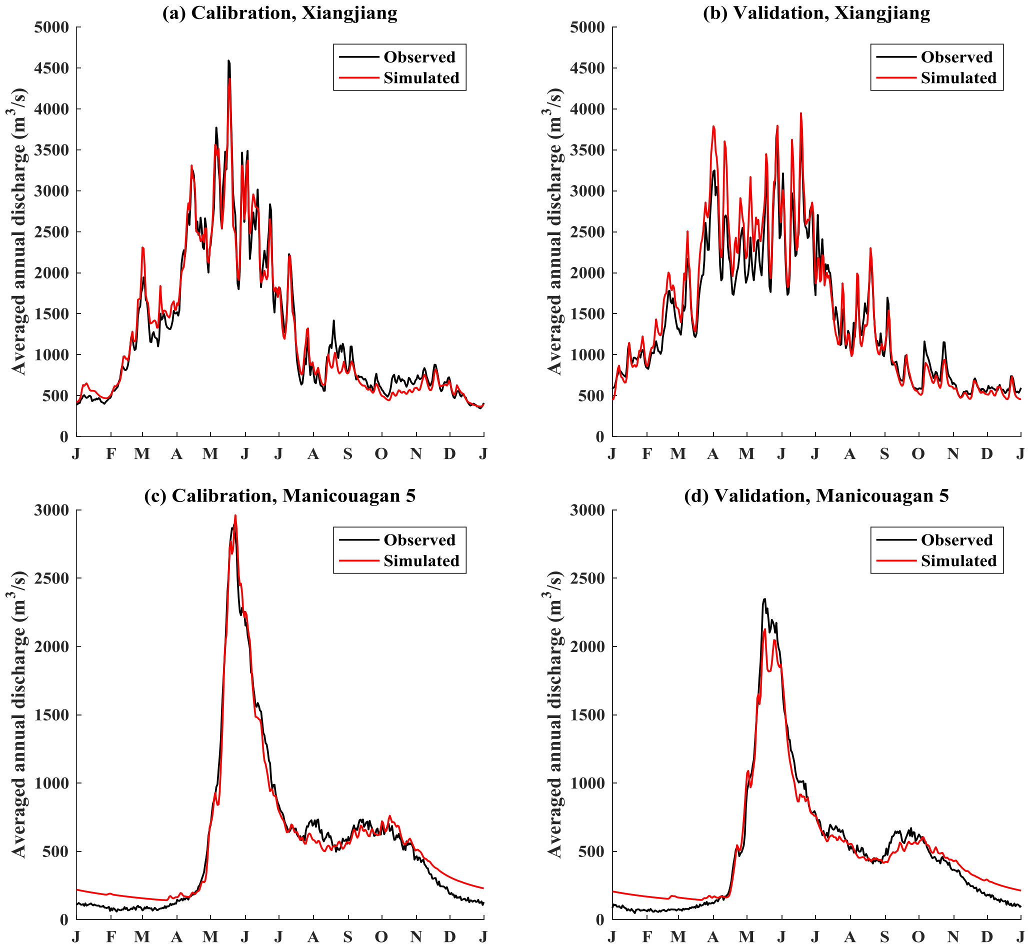

Figure ~~2~~3: Observed and simulated mean hydrographs for (a, c) calibration and (b, d) validation periods over the (a, b) Xiangjiang and (c, d) Manicouagan 5 watersheds.

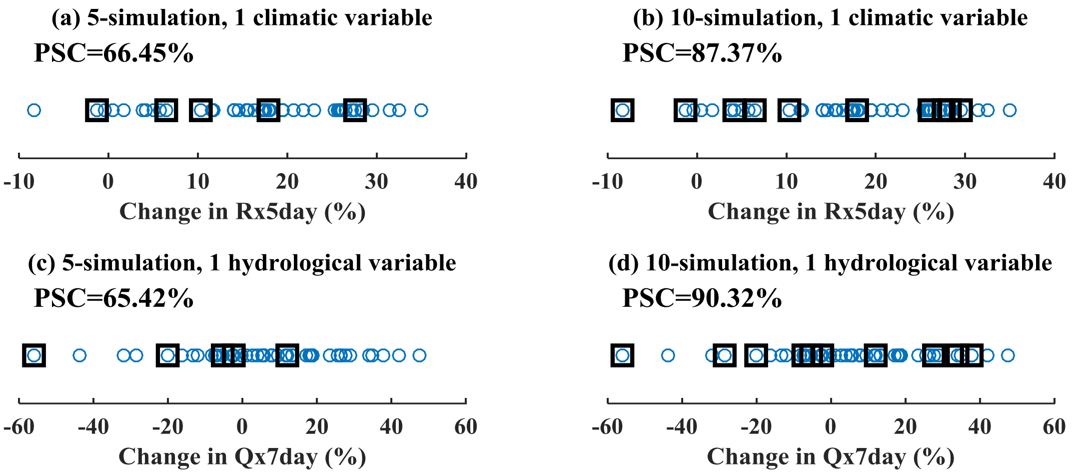

Figure 4: Examples of the transferability of climate uncertainty to hydrological impacts based on 1 variable when selecting (a, c) 5 and (b, d) 10 climate simulations over the Xiangjiang watershed using the KKZ method. The PSCs of each variable are presented in the top left corner.

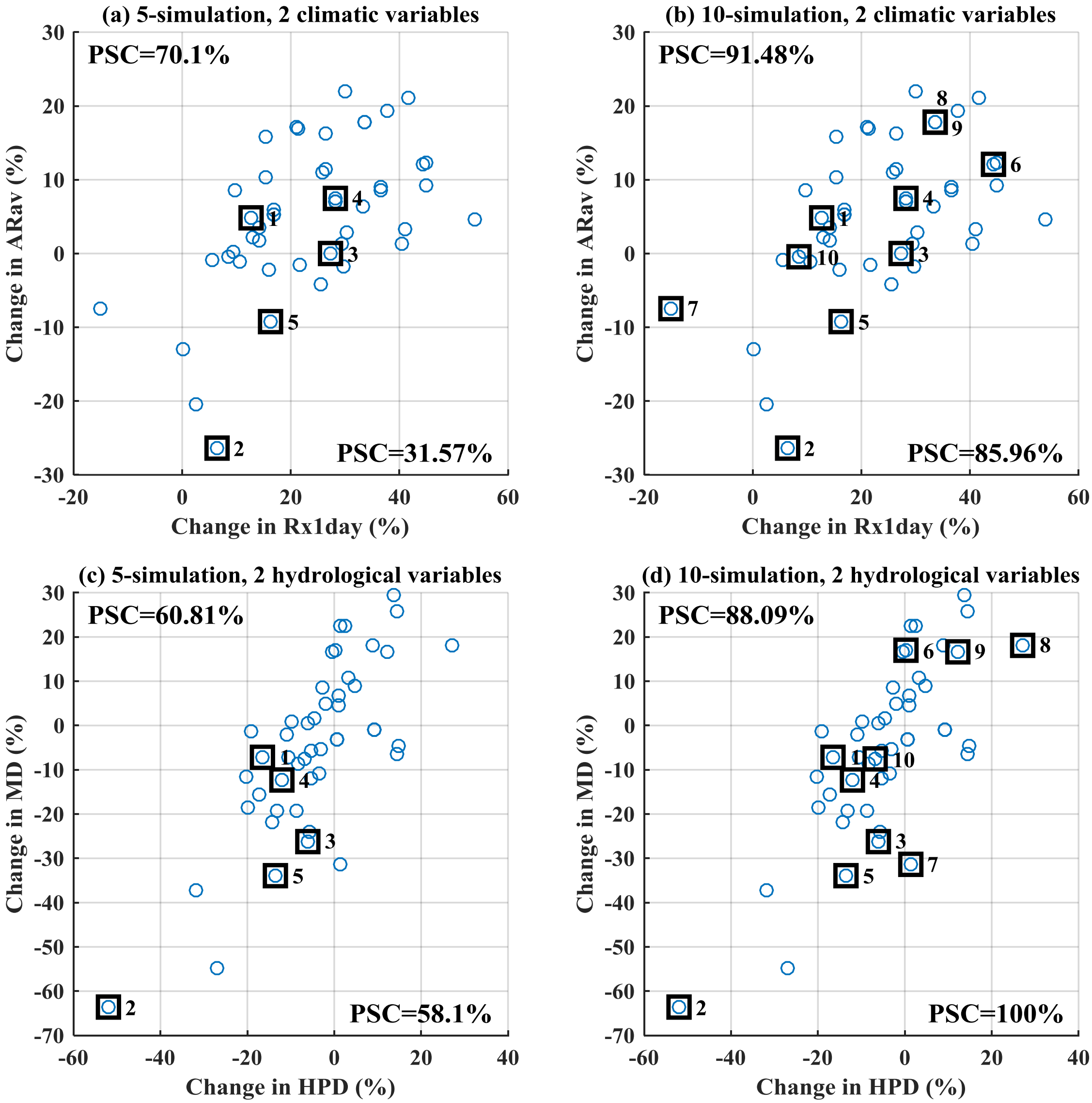

Figure 5: Examples of the transferability of climate uncertainty to hydrological impacts based on 2 variables when selecting (a, c) 5 and (b, d) 10 climate simulations over the Xiangjiang watershed using the KKZ method. The PSCs of each variable are presented beside the corresponding axes.

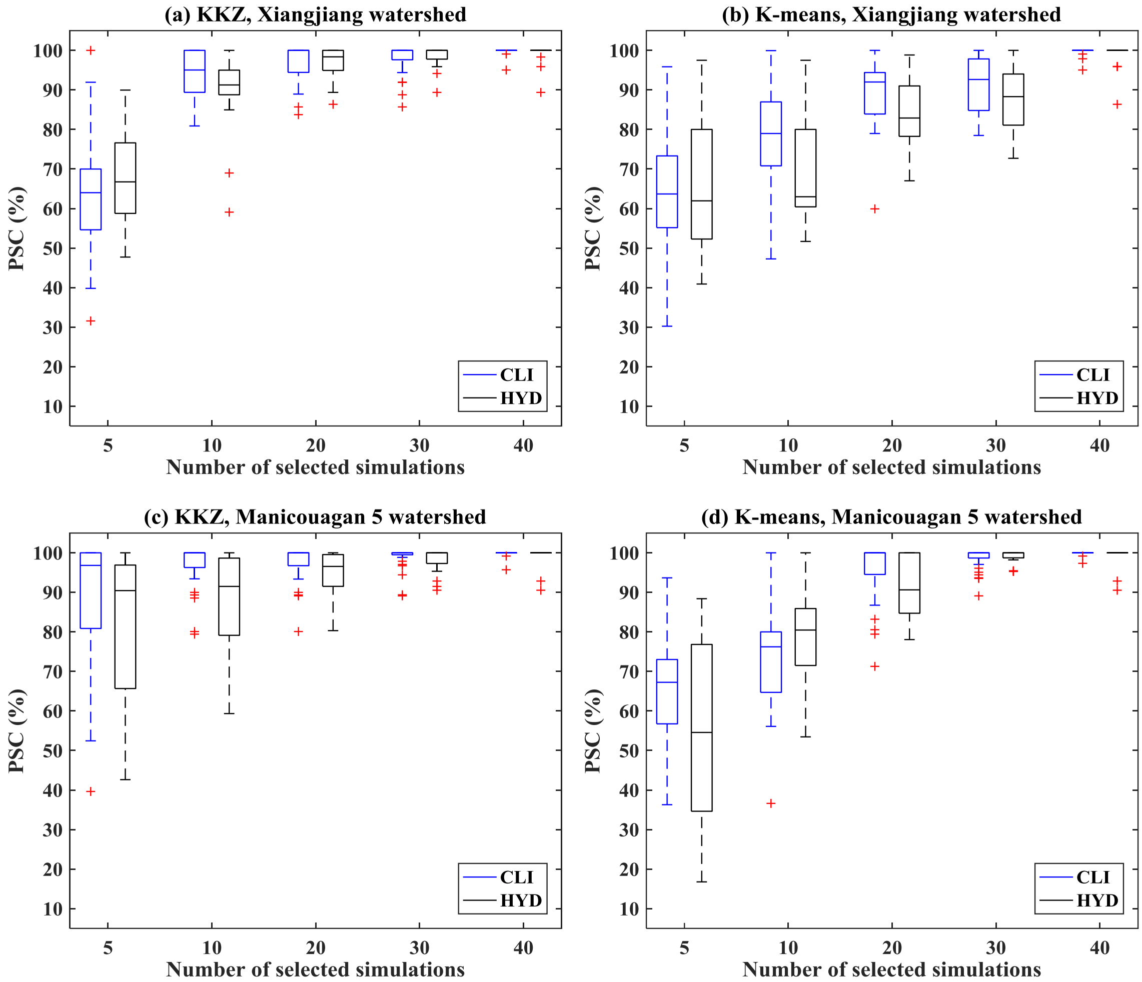

Figure 6: Boxplots of the PSCs of 31 climate variables (CLI) and 17 hydrological variables (HYD) when selecting different numbers of climate simulations over two watersheds using KKZ method and K-means clustering.

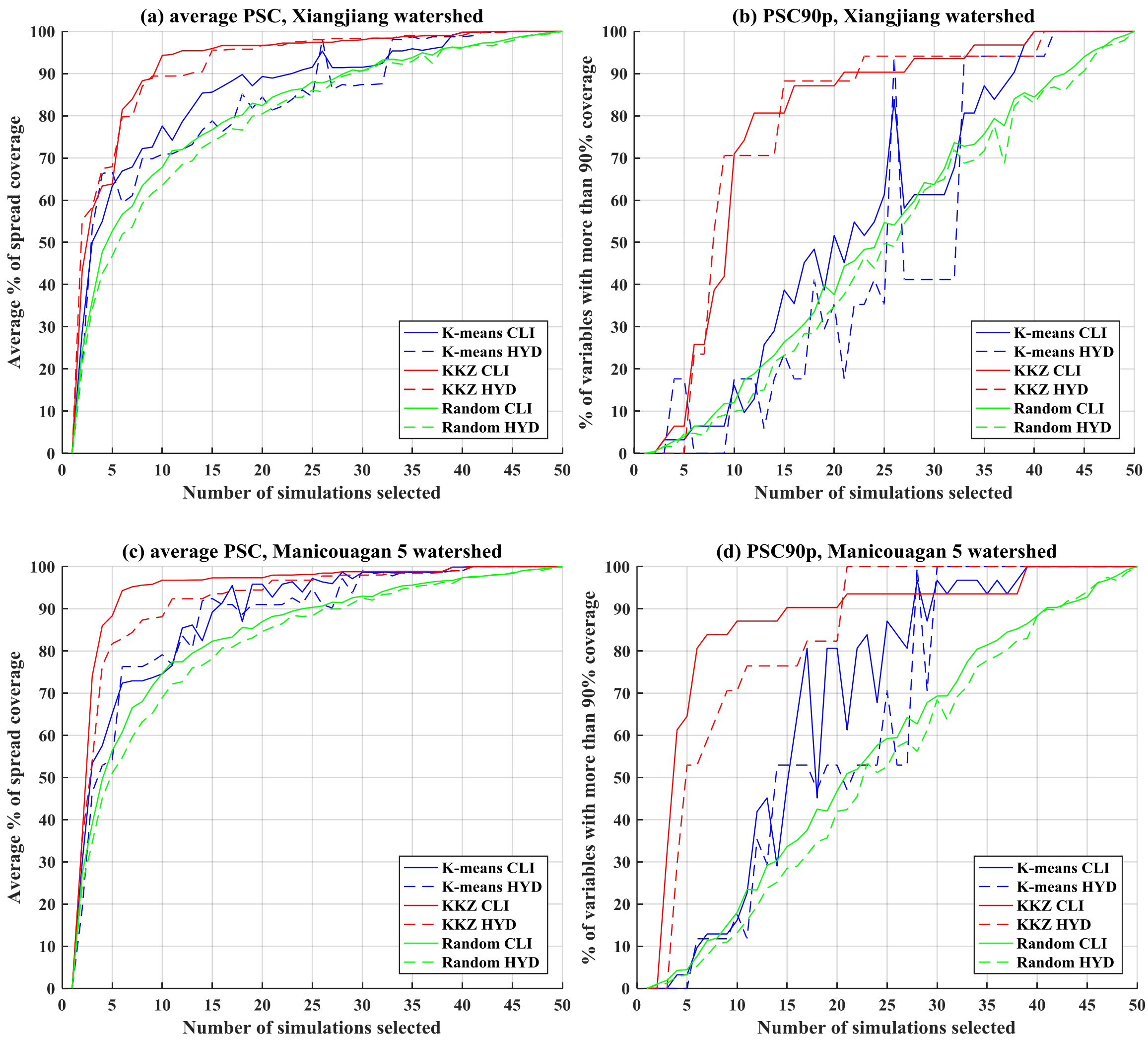

Figure 7: The (a, c) average PSC and (b, d) PSC90p for three different selection methods (K-means, KKZ and random selection) over the (a, b) Xiangjiang watershed and the (b, d) Manicouagan 5 watershed (CLI = climate variables and HYD = hydrological variables).

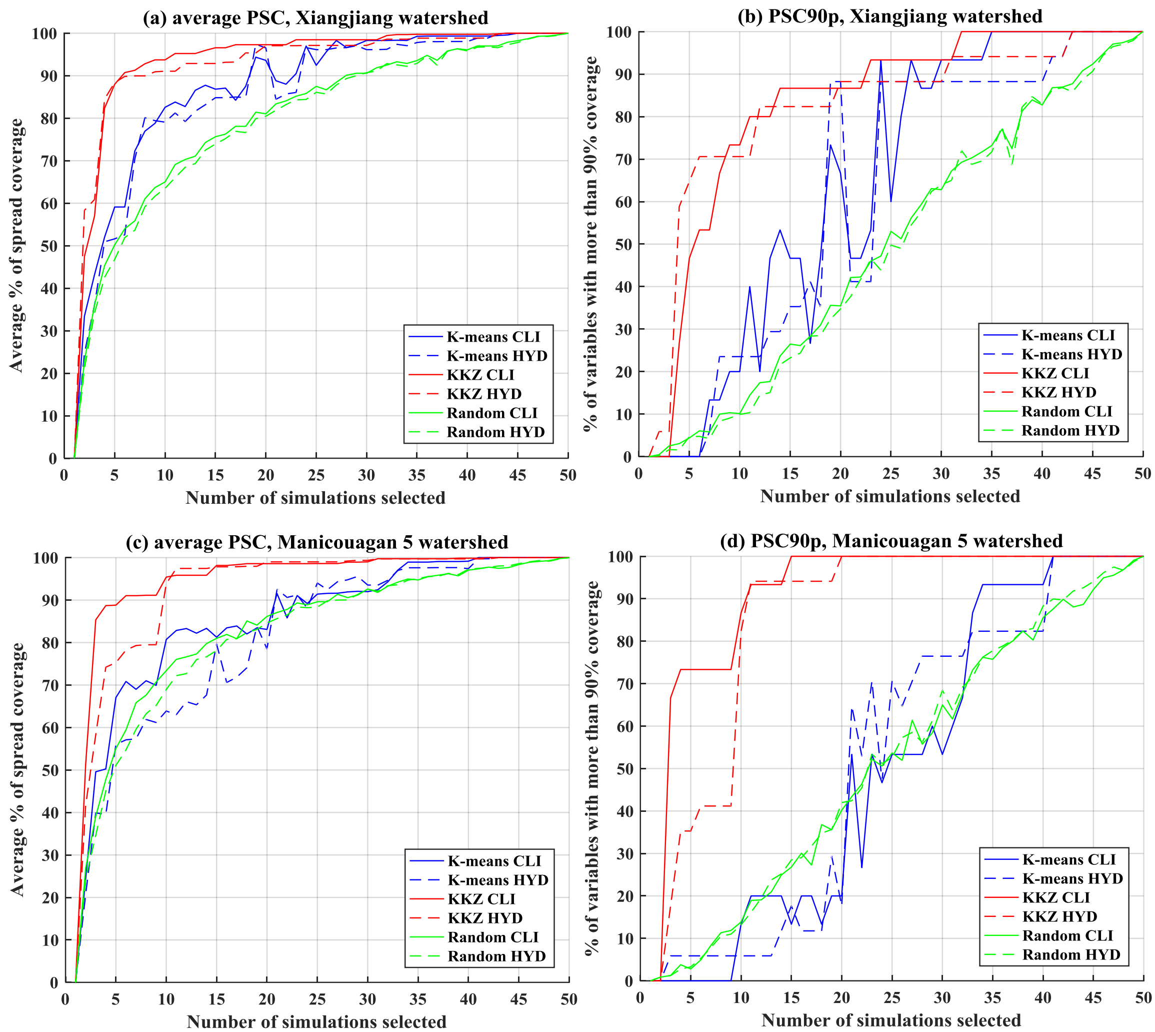

Figure 8: The (a, c) average PSC and (b, d) PSC90p for three different selection methods (K-means, KKZ and random selection) over two watersheds when temperature variables are excluded in the process of simulation selection (CLI = climate variables and HYD = hydrological variables).

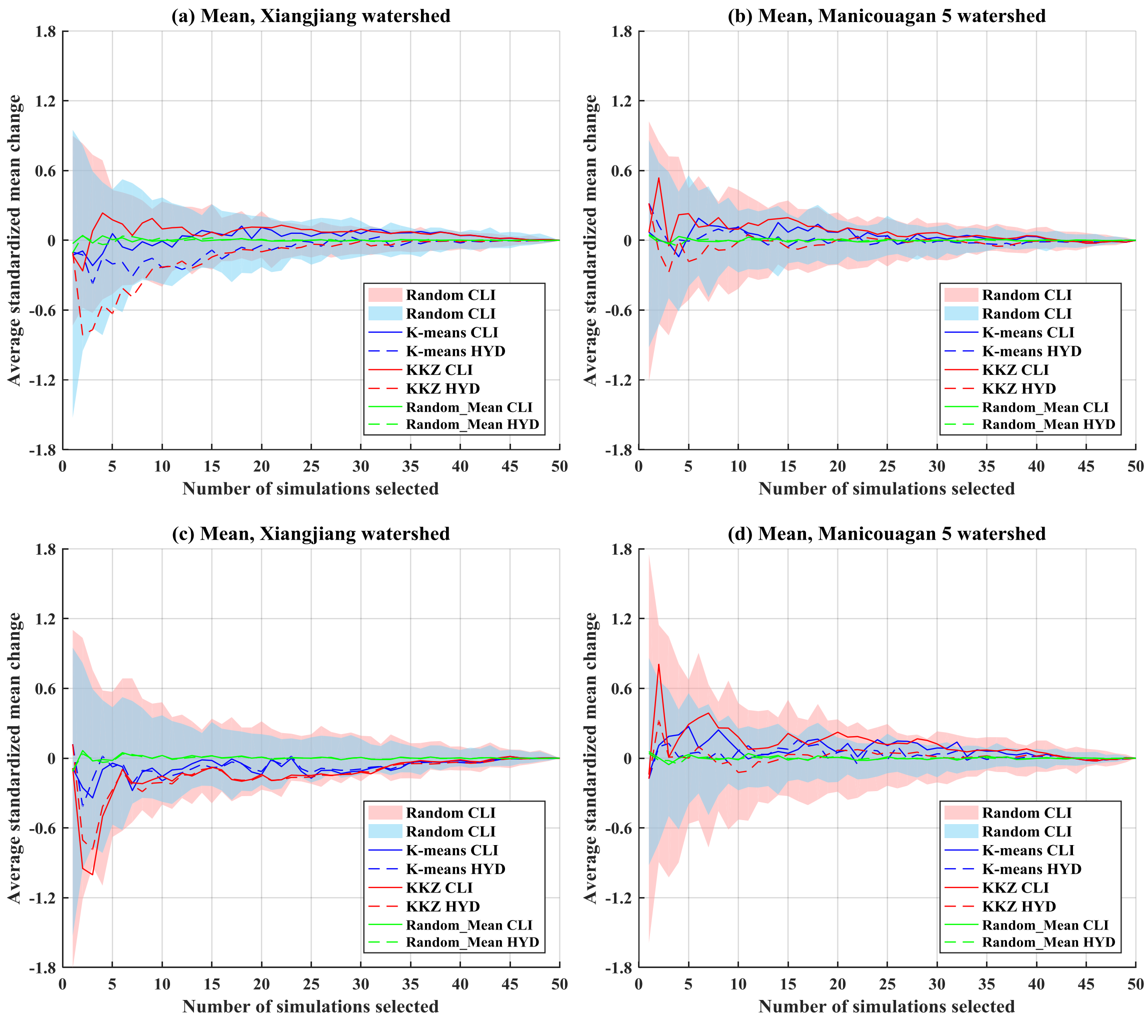

Figure 9: Averaged standardized mean changes in climate (CLI) and hydrological (HYD) variables of subsets selected by three selection methods (K-means, KKZ and random selection) over the Xiangjiang and Manicouagan 5 watersheds when temperature variables are (a, b) included or (c, d) excluded in the process of selection. The pink and blue panels are the envelopes resulting from 100 random selections.

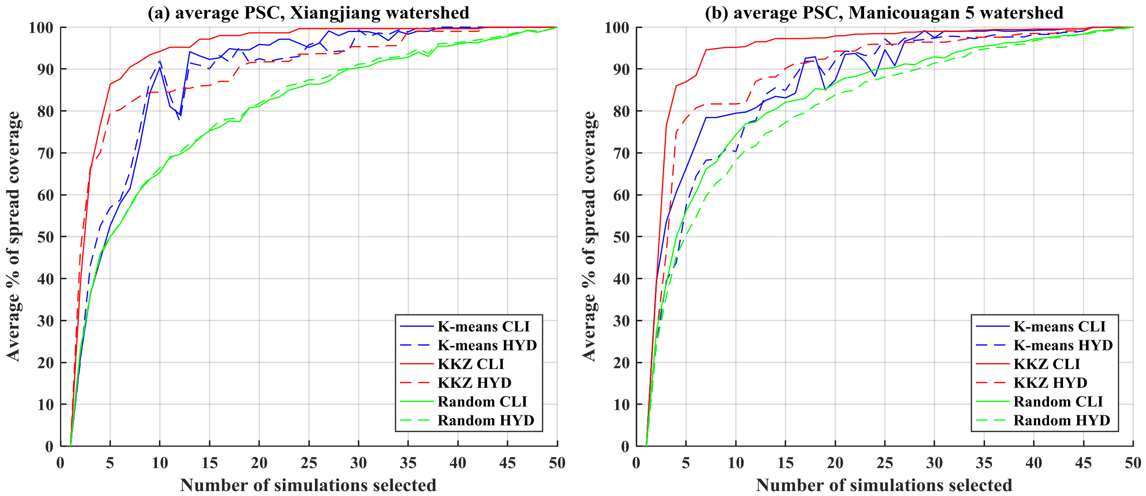

Figure 10: The average PSC for three different selection methods (K-means, KKZ and random selection) over two watersheds when using QM methods (temperature variables are excluded in the Xiangjiang watershed while they are included in the Manicouagan 5 watershed).

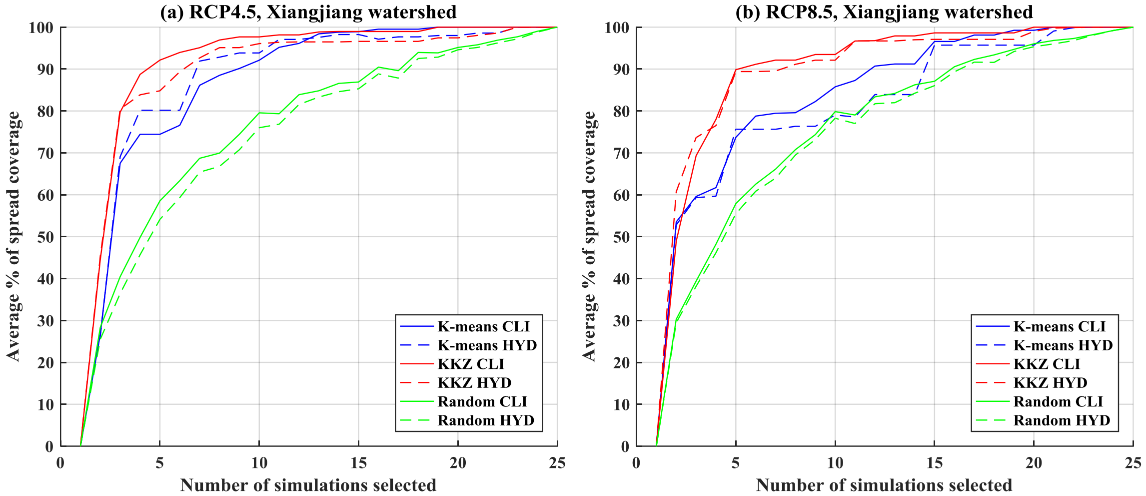

Figure 11: The average PSC for three different selection methods (K-means, KKZ and random selection) over the Xiangjiang watershed when only one emission scenario (RCP4.5 or RCP8.5) is considered.