Supplement of Hydrol. Earth Syst. Sci., 23, 4033–4050, 2019
https://doi.org/10.5194/hess-23-4033-2019-supplement

Hydrology and Earth System Sciences
Open Access
EGU

Supplement of

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

Hui-Min Wang et al.

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

Supplement

S1 Weighting methods

S1.1 Reliability ensemble averaging (REA)

The reliability ensemble averaging method of Giorgi and Mearns (2002) considers two reliability criteria for a GCM. The first one is the model performance criterion that evaluates the ability of a climate model to simulate historical observation, and the other is the model convergence criterion that examines the difference of a model to the multi-model mean in the future period. The reliability factor of a model is defined as

$$\mathrm{REA}_i = \left\{ \left[\frac{\in}{\mathrm{abs}(B_i)} \right]^m \times \left[\frac{\in}{\mathrm{abs}(D_i)} \right]^n \right\}^{1/mn} \tag{S1}$$

where $\in$ represents the natural climate variability estimated by the interval between the maximum and minimum of 20-year moving averages of yearly observation series. B_i is the bias of a simulation to the observation in terms of the climatological mean, and D_i is the distance between the change of a given model and the REA-weighted mean change. In addition, if the absolute value of bias B_i or distance D_i is smaller than climate variability $\in$, this climate simulation is regarded to be reliable in the corresponding respect (i.e. $\in / \mathrm{abs}(B_i)$ or $\in / \mathrm{abs}(D_i)$ is set to 1). The parameters m and n represent the weight assigned to performance and convergence criteria, respectively, and are both set to 1 in this study.

S1.2 Weighing scheme accounting for performance and interdependence (PI)

Since many climate models share similar modules or parts of codes, they cannot be regarded as independent of each other as in model democracy. Thus, Knutti et al. (2017) proposed a weighting scheme accounting for both performance and interdependencies (PI). The interdependence score I_i of an ith model is evaluated as

$$\mathrm{I}_i = \frac{1}{1 + \sum_{j \neq i}^{N} e^{-\frac{D_{ij}^2}{\sigma_D^2}}} \tag{S2}$$

where D_{ij} measures the distance between the ith and the jth model in terms of the climatological mean. The uniqueness radius σ_D determines how strongly the model interdependency criterion is stressed. When a model is far from all the other models, its interdependence score becomes larger but no more than 1. The performance score P_i of the ith model is evaluated as

$$\mathrm{P}_i = e^{-\frac{B_i^2}{\sigma_B^2}} \tag{S3}$$

where B_i measures the distance of the ith model to the observation in terms of the climatological mean. The skill radius σ_B determines how strongly the model performance criterion is stressed. The overall score of the ith model is calculated by multiplying its interdependence score and performance score as follows:

$$\mathrm{PI}_i = \mathrm{P}_i \times \mathrm{I}_i \tag{S4}$$

Two parameters, σ_D and σ_B, are measured by the multiples of the median distances across all model pairs, and are chosen by visual inspection based on two standards. First, the choice of σ_D should attempt to guarantee that the group of models that are known to be similar (i.e. MIROC-ESM-CHEM, MIROC-ESM and MIROC5 in this study) should gain an I_i about $1/k$ (k is the number of alike models) (Sanderson et al., 2017). Second, σ_B is sampled via perfect model tests (cross validation), in which each model is alternatively regarded as the truth model and the others are used to calculate the PI weights (Knutti et al., 2017). The determination of σ_D should attempt to guarantee that 80% of the truth models fall into the 10-90% range projected by the corresponding weighted ensemble in the future period. For the Manicougan-5 watershed, $\sigma_D = 0.35$ and $\sigma_B = 2$. For the Xiangjiang watershed, $\sigma_D = 0.25$ and $\sigma_B = 2.8$.

S1.3 Representation of the annual cycle (RAC)

The skill score of representation of the annual cycle (RAC) is developed based on the Taylor diagram, which is used to indicate the similarity between a climate simulation series and an observation series (Taylor, 2001). The RAC method can be expressed as the following 4th order formulation.

$$\mathrm{RAC}_i = \frac{4(1+r)^4}{(\sigma + 1/\sigma)^2(1+r_0)^4} \tag{S5}$$

where r is the correlation coefficient between the monthly observed and simulated series, and r_0 is the maximum correlation, which is set to 1 in this study. The parameter $\sigma = \sigma_s/\sigma_o$ is the ratio between the standard deviation of a monthly simulated series and that of a monthly observed series.

S1.4 Upgraded reliability ensemble averaging (UREA)

Since the REA method may artificially reduce uncertainty by its convergence criterion and only consider one metric (i.e. climatological mean), Xu et al. (2010) proposed upgraded reliability ensemble averaging (UREA) to eliminate the model convergence criterion and to introduce other statistics. Even though multiple climate variables were simultaneously evaluated by multiplying their skill scores in Xu et al. (2010), this study individually evaluated each variable as follows.

$$\mathrm{UREA}_i = \left[\frac{\epsilon_a}{\mathrm{abs}(B_{a,i})}\right]^{m_1} \times \left[\frac{\epsilon_v}{\mathrm{abs}(B_{v,i})}\right]^{m_2} \tag{S6}$$

where $B_{a,i}$ and $B_{v,i}$ are the biases of a climate simulation in the average and variance, respectively. ϵ_a and ϵ_v represent the natural climate variability in terms of annual average and inter-annual variation, respectively. The variation is measured by the standard deviation for temperature series and by the coefficient of variation for precipitation and runoff series. In addition, if

the absolute value of bias in the average $B_{a,i}$ or variance $B_{v,i}$ is smaller than climate variability $\in$, this climate simulation is regarded to be reliable in the corresponding respect (i.e. $\in_a / \mathrm{abs}(B_{a,i})$ or $\in_v / \mathrm{abs}(B_{v,i})$ is set to 1). The parameters m_1 and m_2 represent the weight assigned to two metrics and are both set to 1 in this study.

S1.5 Bayesian model averaging (BMA)

Bayesian model averaging (BMA) is a statistical inference approach to obtain probabilistic forecasts from multi-model ensemble simulations based on Bayes theory. BMA has been used to develop probabilistic predictions for ensembles of weather forecasting models, climate models or hydrological predictions (Duan et al., 2007; Min et al., 2007; Raftery et al., 2005). Denote y as the variable to be predicted, $D = [y_1^o, y_2^o, \dots, y_T^o]$ as the observed series with a length of T, and $f = [f_1, f_2, \dots, f_N]$ as the ensemble of series simulated by climate models. Based on the total probability rule, the probability density function of the prediction $p(y|D)$ can be presented as follows.

$$p(y|D) = \sum_{i=1}^{N} p(f_i|D) \cdot p_i(y|f_i, D) \tag{S7}$$

where each simulation f_i is associated with a conditional probability density function, $p_i(y|f_i, D)$, which represents the conditional distribution of y on f_i, given that f_i is regarded as the best simulation for D. The posterior probability $p(f_i|D)$ represents the likelihood that a simulation is the right simulation. It can also be seen as the weight, $w_i = p_i(y|f_i, D)$, which reflects the capability of a simulation to reproduce the observation. Then, the posterior mean is as follows.

$$E[y|D] = \sum_{i=1}^{N} p(f_i|D) \cdot E[p_i(y|f_i, D)] = \sum_{i=1}^{N} w_i f_i \tag{S8}$$

As the use of BMA in Duan et al. (2007), this study assumed that $p_i(y|f_i, D)$ consists of a Gaussian distribution; monthly data series were then adopted as model simulated series f_i. For the variables that do not follow a Gaussian distribution (i.e. precipitation and streamflow in this study), the Box-Cox transformation was used to transform the variables before the BMA algorithm. This study used the Expectation-Maximization algorithm to solve the BMA weights. More details of this algorithm can be found in Duan et al. (2007).

S1.6 Climate prediction index (CPI)

The Climate prediction index (CPI) was introduced by Murphy et al. (2004) to weight climate models based on their relative reliability to correctly simulate climate observation. Assuming that the simulated variable belongs to the Gaussian distribution, the likelihood of a simulated statistic is proportional to the following equation.

$$\mathrm{CPI}_i = \exp\left[-0.5 \frac{(s_i - o_i)^2}{\sigma_{ANN}^2}\right] \tag{S9}$$

where the climatological mean of a simulated series s_i is assumed to have a Gaussian distribution with an expectation of o_i (the observational climatological mean) and a variance simply estimated by σ_{ANN}^2 (the inter-annual variance of the simulated series).

S1.7 Evaluation of the probability density function (PDF)

Perkins et al. (2007) proposed a skill score to evaluate climate models' ability to reproduce the probability density functions (PDF) of observation. Expressed formally, the skill score of a climate simulation is given as

$$\mathrm{PDF}_i = \sum_{1}^{K} \mathrm{minimum}(Z_s, Z_o) \tag{S10}$$

where the probability density function of simulated or observed daily series is separated into K bins, and Z_s and Z_o represent the frequency in a given bin, respectively.

S2 Hydrological model: GR4J-6

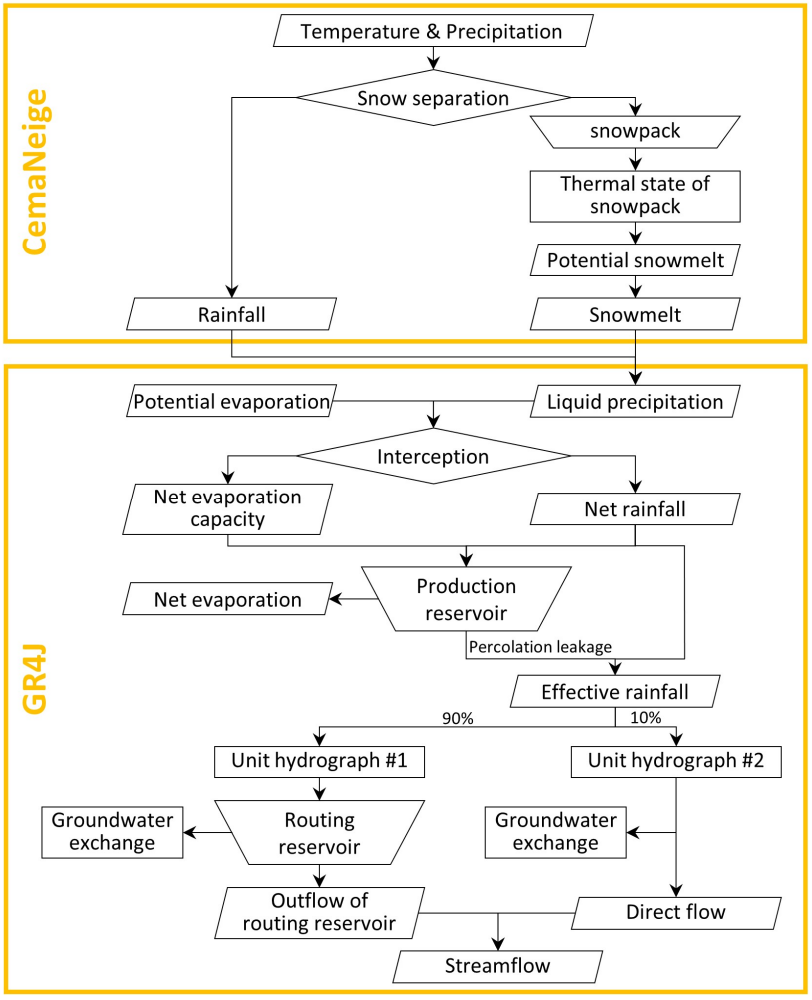

Figure S1. The flowchart of the GR4J-6 hydrological model.

S3 Supplementary results

S3.1 Weights of GCMs

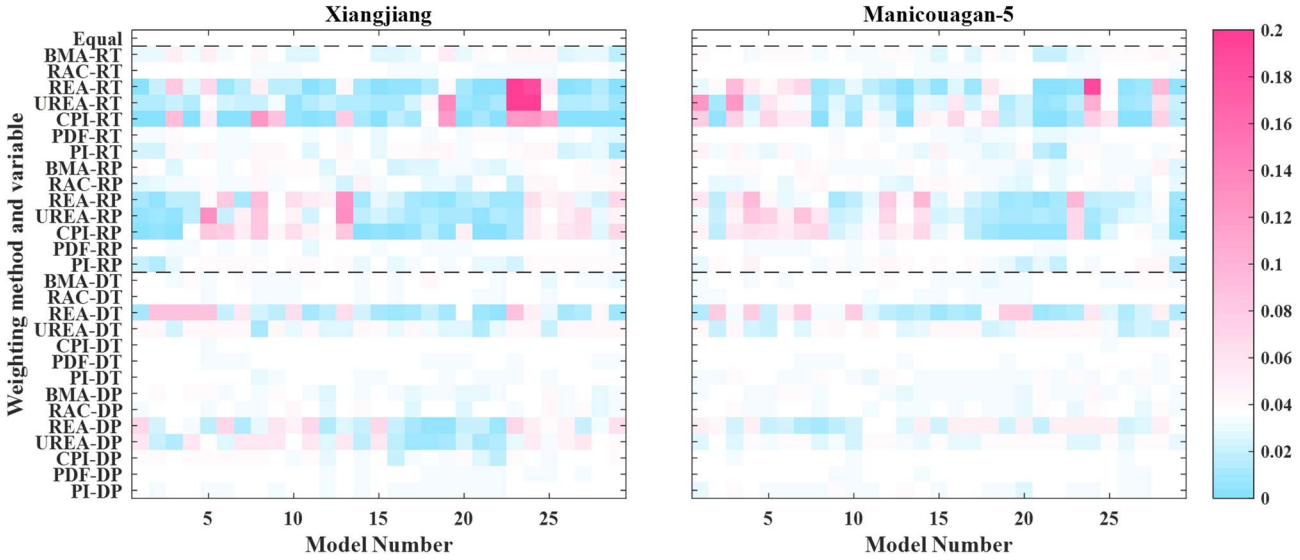

Figure S2. Weights assigned by 8 weighting methods based on raw temperature (*RT*) and precipitation (*RP*) of GCM outputs and bias-corrected temperature (*DT*) and precipitation (*DP*) of GCM outputs for two watersheds. (Equal weight is presented in white. Weights greater than equal are presented in red, and weights less than equal in blue.)

S3.2 Impacts on the hydrological regime

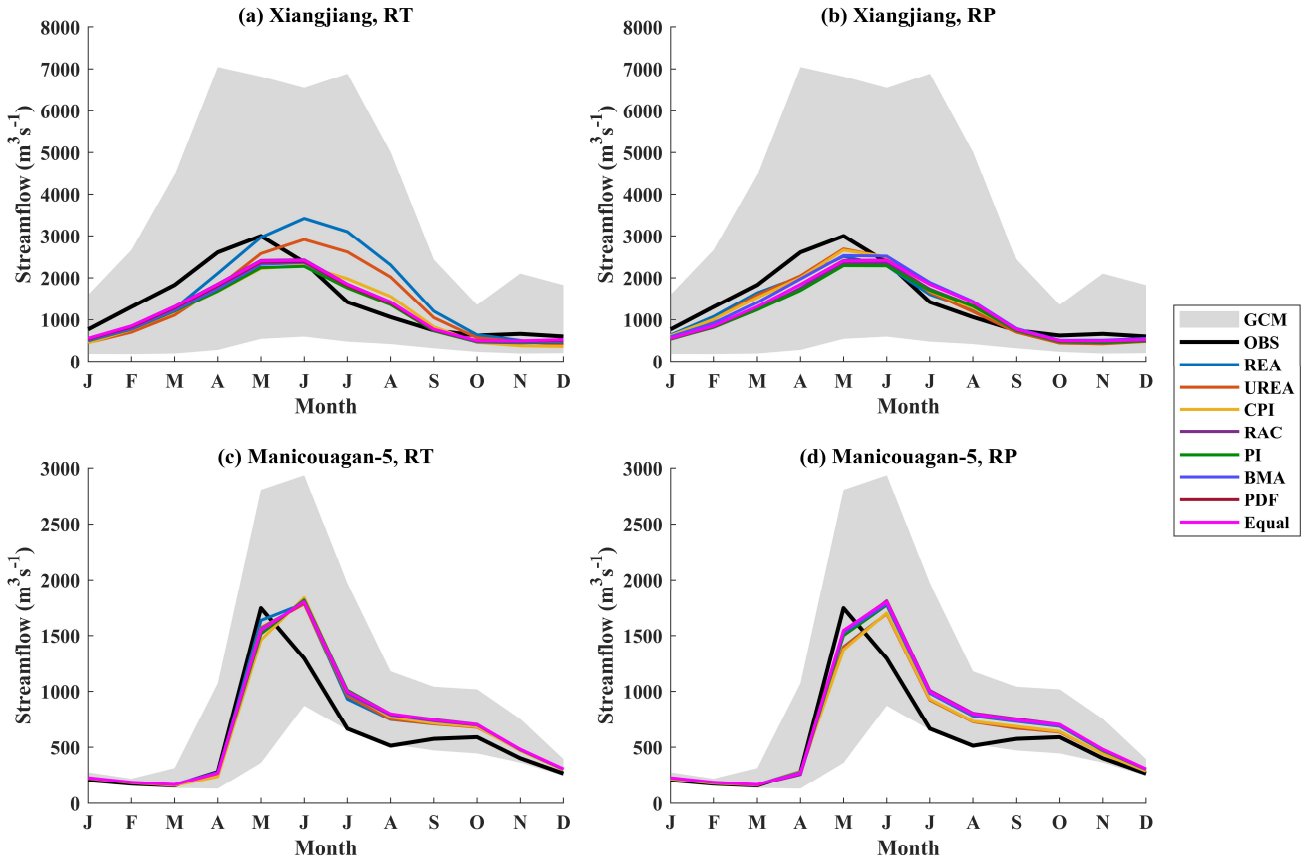

Figure S3. The envelope of monthly mean streamflows simulated by 29 raw and bias-corrected GCM outputs and the multi-model ensemble means of monthly mean streamflows weighted by 8 weighting methods based on raw temperature (*RT*) and precipitation (*RP*) of GCM outputs in both watersheds for the reference period (OBS is the hydrograph simulated from meteorological observation).

S3.3 Bias in multi-model mean

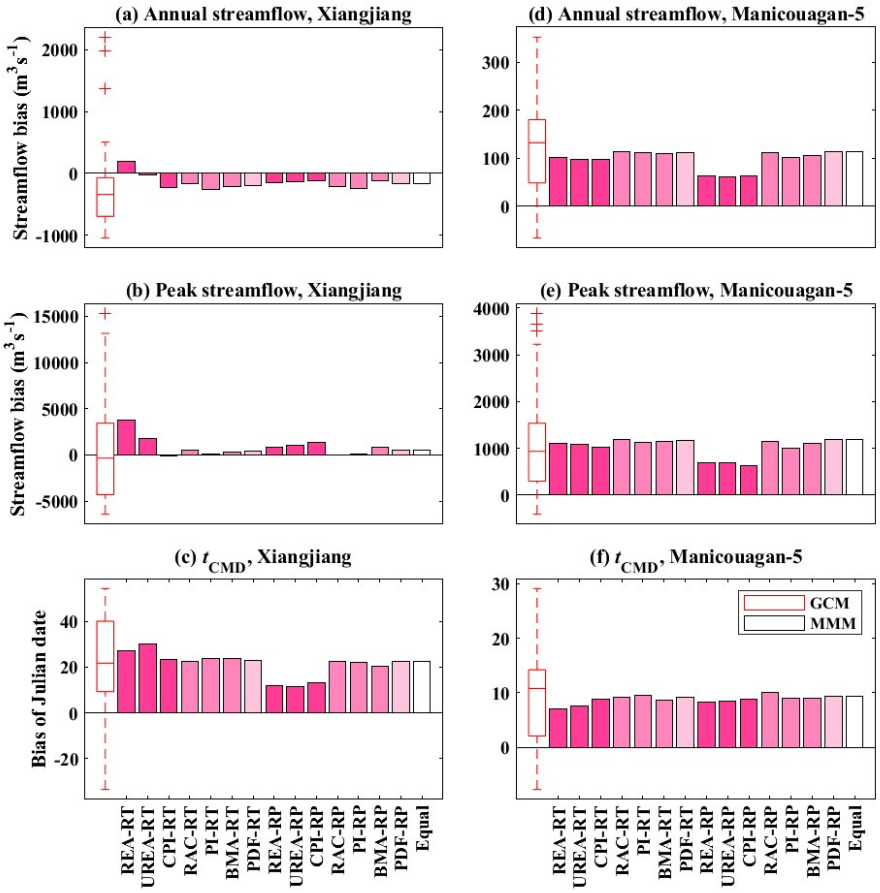

Figure S4. Bias in mean annual streamflow, mean peak streamflow and mean center of timing of annual flow (t_{CMD}) simulated using 29 raw or bias-corrected GCM outputs and the multi-model means (MMM) combined by weights based on raw temperature (*RT*) and raw precipitation (*RP*) in both watershed for the reference period. (The depth of pink in bars of MMM represents the level of inequality of weights as indicated in Table 3.)

S3.4 Impacts on uncertainty

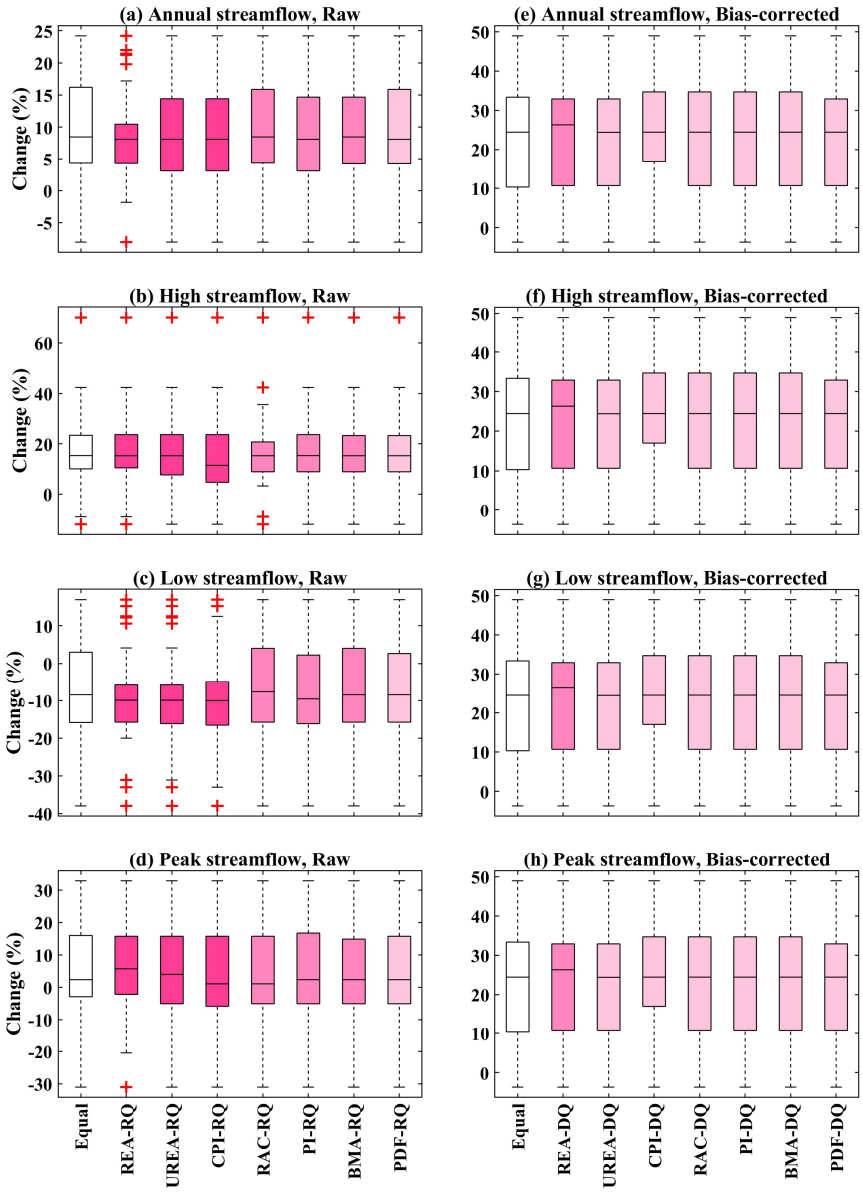

Figure S5. Box plot of changes in four hydrological indices calculated by raw or bias-corrected GCM-simulated streamflows in the Manicouagan-5 watershed. The changes of hydrological variables were sampled through the Monte Carlo approach based on the weights calculated using raw (*RQ*) or bias-corrected (*DQ*) GCM-simulated streamflows. (The depth of pink represents the level of inequality of the weights.)

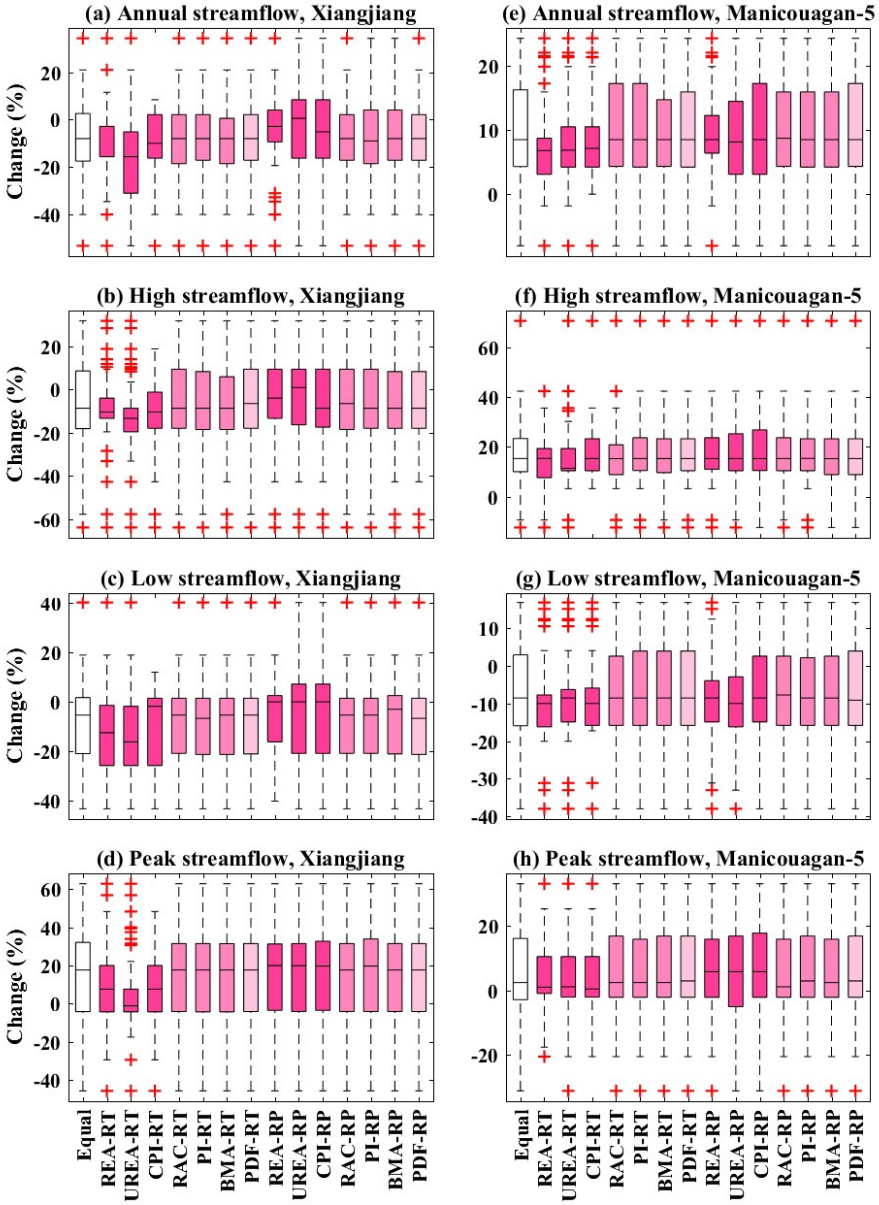

Figure S6. Box plot of changes in four hydrological indices simulated by raw GCM-simulated streamflows over both watersheds. The changes of hydrological variables were sampled through the Monte Carlo approach based on the weights calculated using raw temperature (*RT*) and precipitation (*RP*) of GCM outputs. (The depth of pink represents the level of inequality of weights.)

5

S3.5 Out-of-sample Testing

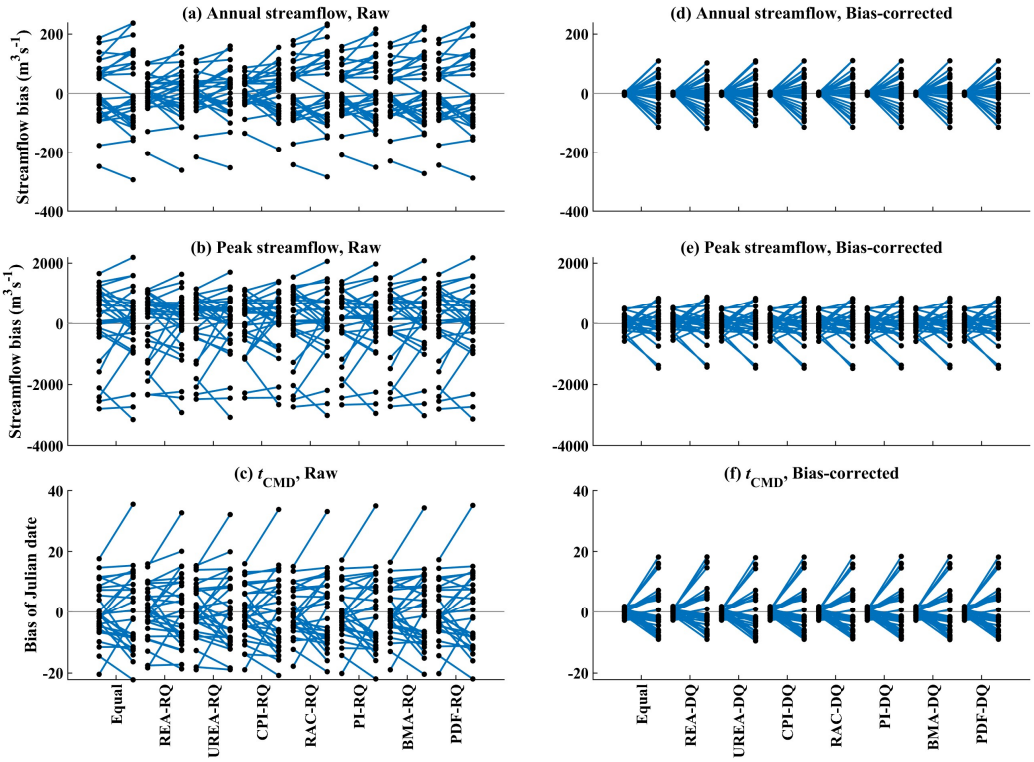

Figure S7. Bias in mean annual streamflow, mean peak streamflow and mean center of timing of annual flow (t_{CMD}) of weighted multi-model ensemble mean in the out-of-sample testing over the Manicouagan-5 watershed; 29 lines of each weighting method represent the results when each of 29 climate models was regarded as the "truth" in turn, and the left and right points in each stick represent the biases for the reference and future periods, respectively.

References

Duan, Q., Ajami, N. K., Gao, X., and Sorooshian, S.: Multi-model ensemble hydrologic prediction using Bayesian model averaging, Advances in Water Resources, 30, 1371-1386, https://doi.org/10.1016/j.advwatres.2006.11.014, 2007.

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.

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., and Eyring, V.: A climate model projection weighting scheme accounting for performance and interdependence, Geophysical Research Letters, https://doi.org/10.1002/2016gl072012, 2017.

Min, S. K., Simonis, D., and Hense, A.: Probabilistic climate change predictions applying Bayesian model averaging, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365, 2103-2116, https://doi.org/10.1098/rsta.2007.2070, 2007.

Murphy, J. M., Sexton, D. M., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., and Stainforth, D. A.: Quantification of modelling uncertainties in a large ensemble of climate change simulations, Nature, 430, 768-772, https://doi.org/10.1038/nature02771, 2004.

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.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M.: Using Bayesian Model Averaging to Calibrate Forecast Ensembles, Monthly Weather Review, 133, 1155-1174, https://doi.org/10.1175/mwr2906.1, 2005.

Sanderson, B. M., Wehner, M., and Knutti, R.: Skill and independence weighting for multi-model assessments, Geoscientific Model Development, 10, 2379-2395, https://doi.org/10.5194/gmd-10-2379-2017, 2017.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106, 7183-7192, https://doi.org/10.1029/2000jd900719, 2001.

Xu, Y., Gao, X., and Giorgi, F.: Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections, Climate Research, 41, 61-81, https://doi.org/10.3354/cr00835, 2010.