Impacts of using state-of-the-art multivariate bias correction methods on hydrological modeling over North America

Qiang Guo[1,2], Jie Chen[1,2*], Xunchang John Zhang[3], Chong-Yu Xu[1,4], Hua Chen[1]

[1] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, Hubei, P. R. China

[2] Hubei Provincial Key Lab of Water System Science for Sponge City Construction, Wuhan University, Wuhan 430072, Hubei, P. R. China

[3] USDA-ARS Grazinglands Research Lab., 7207 West Cheyenne St., El Reno, OK 73036, USA

[4] Department of Geosciences, University of Oslo, P. O. Box 1047, Blindern, Oslo N-0316, Norway

*corresponding author, Email: jiechen@whu.edu.cn; Phone: +86-17764063119

Key points:

- The correction of modeled precipitation-temperature correlations can improve the accuracy of hydrological simulations
- The advantages of using multivariate bias correction methods in hydrological simulations are weakened when comes to the validation period
- The benefits of correcting precipitation-temperature correlations in hydrological simulations are climate regime-dependent

Abstract: Bias correction techniques are widely used to bridge the gap between climate model outputs and input requirements of hydrological models to assess the climate change impacts on hydrology. In addition to univariate bias correction methods, several multivariate bias correction methods were proposed recently, which can not only correct the biases in marginal distributions of individual climate variables, but also properly adjust the biased inter-variable correlations simulated by climate models. Due to the diversities of climate regime and climate model bias, hydrological simulation for watersheds under different climate conditions may show various sensitivities to the correction of inter-variable correlations. Therefore, it is of great importance to investigate 1) whether the correction of inter-variable correlations has impacts on the hydrological modeling, and 2) how these impacts vary with watersheds under different climate conditions. To achieve these goals, this study evaluates behaviors and their spatial variability of multiple state-of-the-art multivariate bias correction methods in hydrological modeling over 2840 watersheds distributed in different climate regimes in North America. The results show that, compared to using a quantile mapping univariate bias correction method, applying multivariate methods can improve the simulation of snow proportion, snowmelt, evaporation, and several streamflow variables. In addition, this improvement is more clear for watersheds with arid and warm temperate climates in southern regions, while is limited for northern snow-characterized watersheds. Overall, this study demonstrates the importance of using multivariate bias correction methods instead of univariate methods in hydrological climate change impact studies, especially for

watersheds with arid and warm temperate climates.

Keywords: Multivariate bias correction methods, Hydrological modeling, Inter-variable correlation, Climate regimes, North America

1. INTRODUCTION

Global climate models (GCMs) and regional climate models (RCMs) are useful tools to provide climate change information for future climate change impact studies. However, due to the systematic biases in the climate model simulations, the GCM and RCM outputs are usually not directly applicable to environment models for impact studies (Hakala et al., 2018; Maraun, 2016). Bias correction methods, as a post-processing approach for RCM and GCM outputs, have been widely used in climate change impact studies for several years (Chen et al., 2013a; Li et al., 2019; Shen et al., 2018; Shrestha et al., 2019). Climate model bias can be reflected in several aspects, such as marginal distribution, and inter-variable correlations (Kumar et al., 2014; Mehran et al., 2014). The widely used quantile mapping methods are able to reduce the biases in marginal distribution (Cannon et al., 2015; Gutiérrez et al., 2019) and consistently perform better than other methods (Chen et al., 2013b). Those methods have become standard procedures for using climate model simulations for hydrological impact studies (Chen et al., 2018; Hakala et al., 2018). However, most of those quantile mapping-based methods operate on each climate variable independently without taking biases of inter-variable correlations into consideration.

The interdependence of key climate variables, such as precipitation (P) and temperature (T) dependence, may be crucial for modeling hydrological processes in impact studies. For example, the P-T correlation can influence the transition between rainfall and snowfall, and also the snowmelt process (Chen et al., 2018; Meyer et al., 2019). With further development of bias correction methods, recent studies have put more effort into correcting or reconstructing the inter-variable correlations of climate model outputs. For example, Li et al. (2014) proposed a joint bias correction (JBC) method to correct P and T simultaneously based on a Gaussian copula function, and found that the proposed method is able to reduce the bias of GCM-simulated P-T correlations. Vrac and Friederichs (2015) proposed an empirical copula-bias correction (ECBC) method to adjust the sequence of each GCM-simulated climate variable to match to the corresponding observed sequence, so that the corrected inter-variable, spatial as well as temporal correlations are close to observations. Cannon (2016, 2017) proposed a series of three multivariate bias correction (MBC) methods including MBCp, MBCr, and MBCn to model the inter-variable correlations for climate model outputs. Within these three methods, MBCp and MBCr are two similar methods, both of which combining univariate bias correction and a multivariate linear bias correction algorithm (Bürger et al., 2011) to correct Pearson and Spearman correlation coefficients, respectively. The MBCn method is adapted from an image processing algorithm, and it has been illustrated to effectively reduce the inter-variable correlation bias simulated by climate models, and also has been tested in many impact studies. More recently, Guo et al. (2019) proposed a two-stage quantile mapping (TSQM) method to introduce the observed correlation matrix to climate model outputs by using the distribution-free shuffle algorithm of Iman and Conover (1982). This method can efficiently reconstruct the inter-variable correlation of climate model outputs to match the observations. A comparison with other commonly used methods also showed that the TSQM method consistently performs better with respect to reproducing the observed inter-variable correlations.

However, one of the ultimate goals of using multivariate bias correction methods is for hydrological modeling and impact studies. With the development of these new multivariate bias correction techniques, some researchers started to investigate the advantages of using multivariate bias correction methods in hydrological impact studies over the last two years. For example, Chen et al. (2018) compared the hydrological simulation of the independent bias correction (IBC) method and JBC method over 12 watersheds, and found that JBC apparently outperforms IBC for 11 out of the 12 watersheds for the calibration period. As for the validation period, the advantages of using JBC are mainly reflected in arid/tropical and snowfall-rainfall-mixed watersheds. Räty et al. (2018) compared the hydrological simulation of univariate quantile mapping corrected data with two multivariate methods (JBC and MBCn) corrected data over 4 watersheds, and found that the additional benefit of using multivariate bias correction methods is not obvious, and only a slight improvement in simulating snow water equivalents is observed. Seo et al.

(2019) investigated the impacts of biased P-T correlation on hydrological variables over two watersheds, and found that the impacts of P-T correlation are more evident on low flow and sub-surface hydrological variables while less remarkable to flow variables with high variability. More recently, Meyer et al. (2019) compared univariate quantile mapping and MBCn in simulating hydrological variables over two alpine catchments. They found that the snow water equivalents, glacier volumes, and streamflow regime simulated using MBCn-corrected data are consistently better than those simulated using univariate quantile mapping corrected data.

To date, the investigation of using multivariate bias correction methods in hydrological modeling is just at its infant stage. The above existing studies are fragmented with limitations in the number of catchments and multivariate bias correction methods. The benefits of using multivariate bias correction methods in hydrological impact studies have not been documented, especially in terms of spatial variability. Due to climate diversity in the world, streamflow may show different sensitivities to variations of P, T and their correlations (Berghuijs et al., 2014; Jefferson et al., 2008; Vano et al., 2012). Considering the inadequate representations of the studied watersheds and the uncertainty related to the choice of bias correction methods, it is inappropriate to draw a general conclusion or make a recommendation for using multivariate bias correction methods for hydrological impact studies.

Accordingly, this study quantifies the impact of using multivariate bias correction methods on hydrological modeling for North America. The spatial variability is specifically investigated by using 2840 watersheds distributed in different climate regimes over North America. Six state-of-the-art multivariate bias correction methods, including JBC, MBC series, TSQM and ECBC were applied to correct 20 GCM simulations for the selected watersheds. To our knowledge, the selected methods include all newly developed multivariate bias correction methods in the literature. For comparison purposes, a univariate quantile mapping method named as daily bias correction (DBC) (Chen et al., 2013b) was also used.

The paper is structured as follows. In the next section, the study area and data are presented. Seven bias correction methods, the hydrological model and the data analysis method are introduced in section 3. The results are shown in section 4, and the discussion and conclusions are presented in section 5 and section 6, respectively.

2. STUDY AREA AND DATA

2.1 Study area

This study was conducted over 2840 watersheds located in Canada (448 watersheds) and the United States (2392 watersheds). These watersheds differ in drainage sizes and climate conditions. The surface area of these watersheds ranges from 302 to 153, 260 km^2, while the average daily discharge varies from 0.3 to 1886.7 m^3/s. In terms of climate conditions, the average annual precipitation ranges from 317 to 4396 mm, while the average daily temperature ranges from -6.2 to 22.7 ℃. According to the Koeppen-Geiger climate classification (Kottek et al., 2006), the

2840 watersheds can be divided into 11 climate regimes, covering arid, warm temperate, snow and polar climates. The detailed climate classification and the basic characteristics of each climate regime are shown in Fig. 1 and Table 1, respectively.

2.2 Data

This study used both observed (served as reference data) and climate model simulated daily P, maximum temperature (T_{max}) and minimum temperature (T_{min}). Since gridded meteorological data are used in this study, the watershed averaged data for running the hydrological model is calculated by averaging all grid points within and around a watershed. The observed P, T_{max}, T_{min} and daily discharge of 448 watersheds in Canada were taken from the Canadian Model Parameter Experiment (CANOPEX) database (Arsenault et al., 2016). For the United States, the meteorological data of 2392 watersheds were taken from the Santa Clara daily database (Livneh et al., 2013), while the daily discharge data were extracted from the United States Geological Survey (USGS) database. Considering the uncertainty related to climate models, outputs of 20 GCMs from the Coupled Model Intercomparison Project Phase 5 (CMIP5) (Taylor et al., 2012) were selected. The basic information of these GCMs is presented in Table S1. The daily meteorological data used in this study cover the 1950-2005 period with the first 28 years (1950-1977) used for calibration and the remaining 28 years (1978-2005) used for validation.

3. METHODS

3.1 Bias correction methods

This study uses 7 bias correction methods, which consist of six multivariate methods (JBC, MBCp, MBCr, MBCn, TSQM, and ECBC), and one univariate method DBC. All these 7 methods are conducted at the daily scale for each specific month. The multivariate bias correction and univariate quantile mapping methods are first calibrated at the calibration period and then applied to the validation period for daily P, T_{max} and T_{min} simulated by 20 GCMs over all 2840 watersheds. A summary of these 7 bias correction methods is shown in Table 2, and a more detailed introduction for these 7 methods is provided in the supporting information (Text S1, 3.1.1-3.1.5).

3.2 Hydrological model

The GR4J-9 hydrological model is used for streamflow simulations over 2840 watersheds. The GR4J-9 model is a 9-parameter, lumped, conceptual hydrological model, which couples GR4J (Perrin et al., 2003) (5-parameter version) rainfall-runoff model with the CemaNeige (Valéry et al., 2014) (4 parameters) snow accumulation and melt routines.

The GR4J model is a soil moisture accounting model, which routes streamflow through two reservoirs and two unit hydrographs. The original version GR4J has four free parameters to be calibrated, which consist of the maximum capacity of the production store, groundwater exchange coefficient, 1-day-ahead maximum capacity of the routing store, and the time base of the unit hydrograph. This model has been tested in a large sample of catchments and shows competitive performances over more complicated models with more parameters (Edijatno et al., 1999; Kunnath-Poovakka & Eldho, 2019). In addition, a previous study (Yang et al., 2019) also showed that the performance of GR4J is more stable than other models (i.e. WASMOD, HBV and XAJ) in a changing climate. In this study, the fixed coefficient in percolation leakage is also set as a free parameter to fit the study area better and is calibrated for each watershed. The potential evaporation in this hydrological model is calculated with the Oudin method (Oudin et al., 2005).

Since there is no snow accumulation and snowmelt module in the GR4J model, it cannot be used to watersheds with significant snowmelt over North America. Therefore, a general snow accounting routine named CemaNeige is added. In the CemaNeige module, precipitation is first divided into rainfall and snowfall according to the magnitude of the daily mean temperature, and the potential snowmelt is then computed by a degree-day approach. The CemaNeige module originally has two parameters, one of which is the snowmelt factor and the other is the cold-content factor. To apply this method to calculate the actual daily snowmelt in North America, one parameter for snowpack threshold and the other for the coefficient of actual snowmelt are also required to be calibrated for each watershed.

The observed daily P, T_{max}, T_{min} and discharge were used to calibrate and validate the GR4J-9 model for all 2840 watersheds. The time periods for calibration and validation are longer than 10 years for all watersheds to obtain reliable parameters. To reduce the influence of non-stationarity of climate time series on model performances, the odd years were used for calibration of the GR4J-9 model and the even years were used for validation (Arsenault et al., 2017). The model parameters were calibrated by the shuffled complex evolution method SCE-UA (Duan et al., 1994), using the Nash-Sutcliffe efficiency (NSE) (Nash & Sutcliffe, 1970) as an objective function. The NSE (shown in Fig. 2) are above 0.5 for all the 2840 watersheds and mainly fall within the range of 0.75-0.85 for calibration and 0.7-0.8 for validation, which indicates the good performance of the GR4J-9 model in the study area.

3.3 Data analysis method

The seven bias correction methods were first evaluated in terms of correcting climate simulations. Since these methods have been extensively evaluated in terms of reproducing the observed marginal distributions (Maraun et al., 2019; Volosciuk et al., 2017), their performances were only demonstrated in terms of correcting the monthly mean value for each variable. In addition, the corrected P-T_{max} and P-T_{min} correlations of monthly time series were presented to demonstrate the performance of each method in correcting the inter-variable correlations. The Spearman correlation coefficient was used as it has no requirement for a particular distribution. The root-mean-square-error (RMSE) was calculated for each climate criterion against the climate reference data over all watersheds.

In terms of evaluating the hydrological simulations, the first two years of the simulations are regarded as a warming up period and removed from both calibration (removing 1950-1951) and validation (removing 1978-1979) periods before the evaluation. Three hydrological state variables including the winter snow proportion, the spring daily mean snowmelt and the summer wet-day potential evaporation were used as the evaluation metrics. In addition, these methods are also evaluated with respect to driving the hydrological model to simulate monthly mean flow, high flow and low flow, and time variables (e.g. time to the peak discharge, and time to the beginning and the end of the annual maximum flood). Time to the beginning and to the end of the flood is calculated based on the cumulative annual hydrograph of each watershed. Specifically, four breakpoints in the cumulative annual hydrograph are picked out and connected by straight lines, aiming at minimizing the RMSE error with the original cumulative annual hydrograph. The number of days from the beginning of the year to reach the first breakpoint is then defined as the beginning of the flood, and the number of days to reach the second breakpoint is defined as the end of the flood. The procedures for calculating the time of the beginning and end of the flood is presented in Fig. S1 in the supporting information. The same method was also used in Chen et al. (2011, 2018). To quantify the results, absolute error (AE), absolute relative error (ARE) and RMSE were also calculated for the simulated hydrological variables against the hydrological variables simulated using climate reference data. The summary and definitions of climatic and hydrological metrics are shown in Table S2.

To statistically test the impacts of using multivariate bias correction methods relative to the univariate quantile mapping method in hydrological modeling, the analysis of variance (ANOVA) and Dunnett-t test (Dunnett, 1955) were conducted based on RMSE of multivariate bias-corrected simulations and univariate bias-corrected simulations for each hydrological variable. The Dunnett-t test is a multiple comparison method designed for comparing the difference in the mean value of the control group and multiple experimental groups. In the Dunnett-t test, the RMSE derived from DBC (20 values for 20 GCMs) is regarded as the control group, and the RMSE derived from each of 6 multivariate methods were compared with the control group. If the P-value of the Dunnett-t test was smaller than 0.1, the results of the multivariate method are considered to be significantly different from the DBC method. To show the reliability of the Dunnett-t test, the statistical power was also calculated for those tests whose P-value was smaller than 0.1.

4. RESULTS

4.1 Observed P-T dependence

To show the variation of the P-T dependence in terms of climate regimes and seasons, the mean observed P-T_{max} and P-T_{min} Spearman correlation coefficients are calculated for the watersheds from 11 climate regimes for each month and for both 1950-1977 and 1978-2005 periods (Fig. S2). Results show that the P-T_{max} and P-T_{min}

correlations varied greatly with climate regimes and seasons. For arid climate BSk and warm temperate climate Csa and Csb, the P-T_{max} correlations are negative for all months, while for most of the rest climate regimes, the P-T_{max} correlations are mostly negative in summer and positive in winter. For the Dfa climate, the P-T_{max} correlation in summer is weak and the correlation coefficient is near 0. In terms of P-T_{min} correlation, almost all the climate regimes show a positive result for all months except for the negative P-T_{min} correlation of the Csa climate in April and May. Apart from the variation of P-T correlation in climate regimes and seasons, nonstationarity of the observed correlation coefficient was also observed between these two continuous time periods. For example, the negative P-T_{max} correlations for Csa climate in March and April are apparently weakened from 1950-1977 to 1978-2005 periods. All these results emphasize the necessity to analyze the spatial variability of using multivariate bias correction methods for hydrological modeling.

4.2 Climate simulations

4.2.1 The performance in representing the univariate distributional characteristics

The RMSE of the monthly mean values of corrected P, T_{max} and T_{min} are presented as boxplot in Fig. 3 for 4 typical months over 2840 watersheds for the validation period. Each box is constructed by 20 RMSE values of 20 GCMs. Results show that the univariate quantile mapping method and the six multivariate bias correction methods perform similarly in correcting the monthly mean of precipitation and temperature for all 4 months. In addition, all these 7 methods show smaller RMSE value and smaller uncertainty related to GCMs in July but show larger RMSE value and larger uncertainty in January for all three variables.

4.2.2 Performance in correcting the inter-variable correlation

Fig. 4 present the spatial distribution of corrected P-T_{max} correlation coefficients for July at the validation period. The climate model BCC-CSM1.1 (m) is used as an example to demonstrate the results. Observed P-T_{max} correlation coefficients are also plotted for comparison. Results show that the DBC method cannot reproduce the observed P-T_{max} correlation coefficients, whose results are similar to correlation coefficients of the raw climate model (results not shown). For the 6 multivariate methods, MBC series, TSQM and ECBC methods have similar performances and all properly reproduce the observed P-T_{max} correlation coefficients, though the MBCp method is slightly worse. However, the JBC method has limited capability to correct the simulated P-T_{max} correlation coefficients, and it has no apparent advantage over the DBC method. The results are also presented in Fig. S3 for January at the validation period. Generally, the performance of each method in January is similar to that in July.

The RMSEs of the inter-variable correlations of corrected time series over 2840 watersheds are shown in Fig. 5 for 20 GCMs. Results are shown as boxplots for 4 typical months and both calibration and validation periods. Similarly, each box consists of 20 values corresponding to 20 GCMs. Results show that DBC has the largest RMSE among all methods. For the 6 multivariate methods, the JBC method shows the largest RMSE, which has also been shown in the spatial distribution of the corrected P-T_{max} correlation coefficients in Figs. 4 and S3. For the other 5 multivariate methods, MBCn, TSQM, and ECBC perform similarly and better than the other two methods in terms of the RMSE for both calibration and validation periods.

4.3 Hydrological simulations

4.3.1 Performance in simulating hydrological state variables

Fig. 6 presents the mean AE of winter snow proportion calculated using climate model simulations with and without bias correction across 20 GCMs over all watersheds for the validation period. Results show that AEs of the proportional precipitation in snow as simulated by the raw GCMs are greatest in the temperate climate zone (mostly between 32 and 48N), and lowest to both north and south of the zone due to all snowfall in the former and all rainfall in the latter in winter, indicating that the bias correction of P-T correlation is essential for properly simulating winter hydrology in the temperate zones. The use of univariate DBC method can improve the winter snow proportion simulation, especially for the central United States. The multivariate methods can further reduce the AE of the winter snow proportion simulation. For example, the AE in the northeastern United States reduces from around 8-10% when using DBC to 1-3 % when using TSQM. The reduction of AE indicates that the use of multivariate methods is able to better distinguish the snowfall from the rainfall in winter.

Fig. 7 presents the mean ARE of spring daily mean snowmelt calculated using climate model simulations with and without bias correction across 20 GCMs over all watersheds for the validation period. Results show that the spring daily mean snowmelt simulated by DBC-corrected data are more accurate compared to the raw GCMs data, especially for most northern watersheds. The multivariate methods consistently perform better than the DBC method with respect to reproducing the spring daily snowmelt calculated using observed data, especially for central North America. For example, the ARE in the central continent reduced from 32-40 % when using DBC to 4-16 % when using MBCn.

Fig. 8 presents the difference between dry-day and wet-day potential evaporation (dry days minus wet days) in summer, calculated using reference data and 20 corrected GCMs simulations over all watersheds for the validation period. For reference data, the evaporation in dry days is lower than that in wet days in eastern North America as indicated by the red color in this area, while an opposite pattern is observed in western North America. This phenomenon is consistent with the observed P-T correlations of these two regions in summer. Fig. S4 shows the observed correlation between P and daily mean temperature (T_{mean}) of the validation period over the 2840 watersheds for each month. For the eastern region, P and T_{mean} are positively correlated in summer, which results in that the evaporation in dry days is lower than in wet days. However, for the western region, the correlation between P and T_{mean} in summer is mostly negative, which results in that the evaporation in dry days are higher than in wet days. Generally, all bias correction methods can represent this pattern. However, the DBC method fails to capture the positive difference in Canada and the negative difference in the southeastern United States. The JBC and MBCp methods also cannot capture the differences in these two regions. However, the rest of multivariate methods (MBCr, MBCn, TSQM and ECBC) more properly reproduce the observed pattern of the difference between dry and wet day potential evaporation.

Fig. 9 presents the Dunnett-t test of RMSE between the DBC and multivariate methods for simulating three hydrological state variables. The blue color represents the fact that the multivariate methods perform significantly better than the DBC at the significance level of 10%, while the red color represents the opposite result, and the white color represents there is no significant difference between these two methods. For the winter snowfall proportion (Fig. 9(a) and (b)), the multivariate methods perform significantly better than the DBC method for almost all climate regimes and for both calibration and validation periods. There are only two cases (the Dfa in the calibration period and the Csa in the validation period) that the TSQM method performs significantly worse than the DBC method. For the spring daily mean snowmelt (Fig. 9(c) and (d)), the multivariate methods consistently outperform the DBC method over almost all climate regimes for the calibration period, but for the validation period, the results are dependent on climate regimes. Specifically, the multivariate methods perform significantly better for arid climate (BSk) and warm temperate climate (Cfa, Cfb, and Csb). However, for most snow climate (Dfb, Dfc, Dsb and Dsc) and polar climate (ET), the multivariate methods perform similarly to or even worse than the DBC method. For the mean wet-day evaporation in summer (Fig. 9(e) and (f)), five out of six multivariate methods show significantly better performance than the DBC method for all climate regimes in calibration period, and the advantages of multivariate methods are weakened when comes to the validation period. However, an exception is observed when using the JBC method, which demonstrates the incapability of this method in correcting the P-T correlation in summer as indicated in Fig. 4.

4.3.2 Performance in simulating streamflow variables

Fig. 10 presents the Dunnett-t test for RMSE of the mean streamflow over 11 climate regimes for 4 seasons of the calibration and validation periods. Prior to using the Dunnett-t test, the 12 monthly RMSEs of the mean streamflow were averaged to obtain 4 seasonal values. For the calibration period, the multivariate methods consistently perform better than or comparable to the DBC method in simulating the mean streamflow over all climate regimes and 4 seasons, though a few of disadvantages were also observed. For the validation period, the advantages of using multivariate methods are generally not significant, even though the performances of these methods are climate-and season-dependent. For the arid climate and warm temperate climate, the multivariate methods do not show significant differences with the DBC method for almost all 4 seasons, and only two cases that multivariate methods perform better than the univariate counterpart in winter. However, for snow-characterized climate, the performance of multivariate methods differs between spring-summer and autumn-winter. Multivariate methods show advantages in simulating the mean streamflow in autumn and winter for snow climate, while showing comparable or even worse performances in spring and summer. For polar climate, the performances of multivariate and univariate bias correction methods are generally comparable for all 4 seasons.

Fig. 11 presents the Dunnett-t test results for RMSE of high flow and low flow over all 11 climate regimes for both calibration and validation periods. In simulating the high flow, multivariate methods perform better than the DBC method for most climate regimes, while several worse cases are exhibited for JBC and TSQM methods. However, in the validation period, the multivariate methods do not show advantages over the univariate DBC method, and on the contrary, these methods even perform worse than the univariate DBC method for specific climate regimes (e.g. ECBC in the Csb climate and JBC in the Dfc climate). In simulating the low flow, multivariate methods significantly perform better than DBC over most climate regimes (except for arid climate BSk) for the calibration period, with the exception of the TSQM method. In the validation period, the advantages of using multivariate methods are not apparent, and the multivariate methods show comparable performances with the DBC method with both advantages and disadvantages were observed.

The Dunnett-t test for RMSE of time variables is shown in Fig. 12 over all 11 climate regimes for both calibration and validation periods. Similar to other hydrological variables, the multivariate methods significantly outperform the DBC over most climate regimes for the calibration period, especially for simulating the time to the peak discharge. In simulating the time to the beginning and end of the flood, the advantages of multivariate methods were mostly observed in snow-characterized and polar climates. For the validation period, the advantages of using multivariate methods are not very significant. Significant advantages are observed for simulating the time to the peak discharge and time to the beginning of flood for warm temperate Csb, and simulating time to the end of flood for polar climate ET.

To show the reliability of the hypothesis testing results, the statistical power is calculated for the Dunnett-t test that shows significant differences between multivariate and univariate methods for hydrological variables. The statistical power values are shown as boxplots in Fig. S5. For the calibration period, the mean values of statistical power are greater than 0.9 for all six types of hydrological variables, and the values for three state variables are greater than those of streamflow variables. For the validation period, the mean values of statistical power are greater than 0.8 except for the mean flow whose mean value ranges between 0.6 and 0.7. These high values of statistical power prove the rationality of the hypothesis testing used in this study.

The hypothesis testing shows that multivariate bias correction methods can significantly improve the simulation of the streamflow for the calibration period, but do not show many significant advantages for the validation period. To investigate the effects of using multivariate methods more explicitly in simulating the streamflow for the validation period, the mean values of the RMSE (20 values of 20 GCMs) of each multivariate method are also compared to the corresponding mean RMSE values of DBC for each streamflow variable (Fig. S6). The ratio (%) that the mean values of RMSE derived from the multivariate methods being smaller than those derived from DBC in simulating each hydrological variable for each climate regime is also calculated and shown in Table 3. The results show that the multivariate methods have smaller RMSE than DBC in simulating most streamflow variables for the validation period with the ratio ranging from 55-82%, except for summer mean flow with the ratio of 27%. In terms of climate regimes, the ratio that multivariate methods showing advantages is higher for arid and warm temperature climates whose value is 70% and 71%, respectively. However, for snow and polar climates, the performances of multivariate methods are generally comparable to those of DBC with the advantage ratios of 54% and 55%, respectively. Overall, these comparisons illustrate that the multivariate methods generally perform better than the univariate counterparts at simulating the streamflow variables for the validation period, even though these advantages do not reach a significant level when using the hypothesis testing. In addition, the multivariate methods perform better than the univariate counterpart in streamflow simulations for arid and warm temperate climate regimes located in southern regions.

5. DISCUSSION

This study quantifies the impacts of using multivariate bias correction methods on hydrological modeling and investigates their spatial variability by using 2840 watersheds distributed in different climate regimes over North America. Results show that the multivariate bias correction methods significantly outperform the univariate bias correction method in simulating hydrological variables for the calibration period. As for the validation period, the advantages of multivariate methods are not as profound as for the calibration period. But they are still significant for the hydrological state variables, while statistically insignificant for streamflow variables for most climate regimes based on the Dunnett-t test. However, the direct comparisons using RMSE show the multivariate methods still perform better than the univariate method in streamflow simulations in general.

Compared to the commonly used univariate quantile mapping method, the multivariate bias correction methods are able to adjust the biased inter-variable correlations simulated by climate models. Therefore, the necessity of using multivariate bias correction methods for hydrological modeling depends on the biases of inter-variable correlations simulated by climate models for a specific region. The performance of GCM in simulating the inter-variable correlations of climate variables is regionally dependent. Fig. S7 presents the mean absolute relative error (MARE) of monthly P-T_{max} and P-T_{min} correlation coefficient simulated by 20 GCMs over 11 climate regimes for both calibration and validation periods. Results show that GCM-simulated P-T_{max} correlations are mostly biased for warm temperate climate regime in summer while for snow and polar climate regime in winter. For the P-T_{min} correlations, biases are mostly observed for arid climate and polar climate regimes. It is easy to find that multivariate methods show apparent advantages in simulating the mean streamflow for regions (as shown in Fig. 10) where climate models are more biased in simulating inter-variables correlations (as shown in Fig. S7). Specifically, the multivariate methods show apparent advantages for the warm temperate climate in summer as well as for snow climate in winter in the calibration period. This example further proves that the correction of biased P-T correlation is able to improve the simulation of mean streamflow. Due to the complex terrain or inadequate representation of basic physical processes for some regions, the climate model may show a low capability to simulate the inter-variable correlations and thus results in large biases. For these regions, it is clearly necessary to use multivariate methods for hydrological impact studies.

This study shows that the multivariate bias correction methods significantly outperform the univariate method in simulating most hydrological variables for the calibration period, but these advantages are weakened for most climate regimes and hydrological variables when coming to the validation period, especially for streamflow variables. This may be because the bias of inter-variable correlation simulated by climate models is not stationary, and the observed inter-variable correlation itself is also not invariable. Previous studies (Chen et al., 2017; Hui et al., 2018; Maraun, 2012) have shown that bias correction methods can deteriorate the original climate simulations when bias directions are different between future and historical periods (or calibration and validation periods) or when future biases reduce to less than half the calibration biases. The nonstationarity of inter-variable correlation bias of climate models can be observed in Fig. S7. For example, larger P-T_{max} correlation bias is observed for the Dfa climate in summer for the 1978-2005 period compared to the 1950-1977 period. Besides the nonstationarity of model bias, the variation of observed inter-variable correlations also results in the weakened performance of multivariate methods. To further explore this problem, Fig. S8 presents the differences of observed monthly P-T_{max} correlation coefficient between 1950-1977 and 1978-2005 period for all 2840 watersheds in North America. Results show that the observed P-T_{max} correlation changes considerably for most watersheds and months. For some watersheds, these changes can be larger than 0.3 in either positive or negative. However, all existing multivariate bias correction methods only introduce the correlation coefficients at the calibration period to the validation period or future period. In other words, the nonstationarity of inter-variable correlations is not considered. This partly explains why the advantages of using multivariate bias correction methods in the calibration period may even reverse when comes to the validation period. Berg et al. (2015) found that due to global warming, large parts of the land surface show more significantly negative summer P-T correlations for the 2071-2100 period than for the 1971-2000 period. However, in some other areas, the P-T correlations in summer may also become significantly more positive. Mahony and Cannon (2018) also found that the P-T correlation may change more obviously in the future due to natural variability and climate sensitivity. More recently, Hao et al. (2019) found that the P-T correlation would be influenced by global warming and thus may result in new compound extreme events in the future. However, most GCMs show a limited capability to simulate the changes in P-T correlations. With the continuous change of temperature in the foreseeable future, the P-T correlation may vary with time, which may challenge the multivariate bias correction methods that reproducing the historical correlation for the future. To deal with this problem, multivariate bias correction methods taking into account the nonstationarities of model biases as well as inter-variable correlations need to be developed. This may be an avenue for future studies. In addition, the different performances of the hydrological model between the calibration and validation periods (Fig. 2) may also contribute to the weakened performances of the multivariate methods in the validation period.

Apart from the stationary assumption of the inter-variable correlations, most multivariate bias correction methods involve modifying the time sequence of the simulated variable to induce the desired correlation matrix. Fig. S9 in the supporting information presents the Spearman correlation coefficients between the corrected and raw climate model-simulated variables for July over both calibration and validation periods, using BCC-CSM1.1(m) as an example. Results show that, DBC-and JBC-corrected data have the highest correlation coefficient with the raw data, as these two methods do not alter the time sequence of the raw climate model data. Only the correlation of precipitation between the corrected and raw data is slightly reduced due to the correction of the wet-day frequency and the different correction factors in each quantile. The three MBC methods adjust the time sequence of the raw model data when inducing the desired correlation matrix. The impacts on correlations are dependent on the adjustment of the temporal sequence. MBC-corrected simulations have high correlation coefficients with the raw data for all three variables. The TSQM method keeps the original simulated time sequence of precipitation while re-ranks the temperature sequence, so it has high correlation coefficients for precipitation but low correlation coefficients for temperature with raw model data. The ECBC method reorders the original time sequence of each simulated variable to match the sequence of historical observations, which loses the sequence information of the model outputs, thus the correlation coefficients between ECBC-corrected and raw model data are nearly 0. By definition, all MBC, TSQM and ECBC modify the temporal sequence of the climate model outputs to induce the desired inter-variable correlations. However, ECBC assumes the temporal sequence of climate simulations is identical to historical observations. This assumption may not be valid in a changing climate, as the temporal sequence likely changes for the future period. In contrast, the MBC method does not rely on this assumption and allows the rank orders of climate data to evolve over time, and the TSQM method has a similar feature to allow the rank order of a key climate variable to evolve with time. For the JBC method, though it does not involve modifying the time sequence and maintains the original rank orders equally to the univariate DBC method, it does not show significant advantages over the DBC method for the validation period. The similar performances of DBC and JBC partially reflect the fact that the additional parameters and processes in JBC may not help much in improving the hydrological simulations for the validation period.

There is also a limitation in this study that needs to be acknowledged. Due to the computational load for a large domain over North America, only one lumped hydrological model was used. The use of a more physical-based land surface or hydrological model can further take advantages of the spatial dependence of cross-correlated multiple climate variables. This can be an avenue for future studies.

6. CONCLUSIONS

In this study, the impacts of using state-of-the-art multivariate bias correction methods on hydrological modeling were investigated over 2840 watersheds in North America. The main conclusions are summarized as follows:

(1) Most multivariate bias correction methods can effectively reproduce the observed inter-variable correlation coefficient for the watersheds in North America, while the univariate bias correction method shows limited ability in this aspect.

(2) In terms of hydrological modeling, the use of multivariate bias correction methods can significantly improve the simulation of hydrological variables for most climate regimes in the calibration period. However, the advantages of using multivariate methods in the validation period are not as profound as for the calibration period, especially for simulating streamflow variables, because of the non-stationarity of the inter-variable correlations.

(3) The advantage of using multivariate bias correction methods for hydrological modeling is region-dependent. In general, the multivariate methods show more advantages in arid and warm temperate climate regimes mainly in the southern regions, while showing fewer advantages in snow-characterized and polar climate regimes mainly in the northern regions. This regional difference is more obvious for the validation period.

(4) In terms of the performances of 6 multivariate bias correction methods, the MBC series and ECBC method show more advantages over univariate DBC method in simulating hydrological variables in North America, while JBC and TSQM show limited advantages, especially in simulating streamflow variables.

Acknowledgments:

This work was partially supported by the National Natural Science Foundation of China (Grant no. 51779176, 51539009), the National Key Research and Development Program of China (No. 2017YFA0603704), the Overseas Expertise Introduction

Project for Discipline Innovation (111 Project) funded by Ministry of Education and State Administration of Foreign Experts Affairs P.R. China (Grant No. B18037), and the Thousand Youth Talents Plan from the Organization Department of CCP Central Committee (Wuhan University, China).

The authors would like to thank Dr. Chao Li at the University of Victoria for providing scripts of the JBC method and Dr. Alex Cannon at the Environment and Climate Change Canada (Climate Research Division) for publishing scripts of MBCp, MBCr and MBCn methods. The authors would also like to thank Dr. Arsenault at the University of Quebec, and Dr. Livneh at the University of Colorado Boulder and Dr. Maurer at Santa Clara University for making their data (CANOPEX and Santa Clara database, respectively) available, and these data can be accessed as guided in their articles. The authors would like to acknowledge the contribution of the USGS, and the World Climate Research Program Working Group on Coupled Modelling, and to thank the climate modeling groups listed in Table S1 for producing and making their model outputs available. The URIs of these data are as follows:

CANOPEX: (http://canopex.etsmtl.net)

Santa Clara database:

(ftp://livnehpublicstorage.colorado.edu/public/Livneh.2013.CONUS.Dataset/)

USGS database: (https://www.usgs.gov/)

CMIP5 database: (https://www.wcrp-climate.org/wgcm-cmip/wgcm-cmip5)

References

Arsenault, R., Bazile, R., Ouellet Dallaire, C., & Brissette, F. (2016). CANOPEX: A Canadian hydrometeorological watershed database. *Hydrological Processes, 30*(15), 2734-2736. doi:10.1002/hyp.10880

Arsenault, R., Essou, G. R. C., & Brissette, F. P. (2017). Improving Hydrological Model Simulations with Combined Multi-Input and Multimodel Averaging Frameworks. *Journal of Hydrologic Engineering, 22*(4), 04016066. doi:10.1061/(asce)he.1943-5584.0001489

Bürger, G., Schulla, J., & Werner, A. T. (2011). Estimates of future flow, including extremes, of the Columbia River headwaters. *Water Resources Research, 47*(10). doi:10.1029/2010wr009716

Berg, A., Lintner, B. R., Findell, K., Seneviratne, S. I., van den Hurk, B., Ducharne, A., Chéruy, F., Hagemann, S., Lawrence, D. M., Malyshev, S., Meier, A., & Gentine, P. (2015). Interannual Coupling between Summertime Surface Temperature and Precipitation over Land: Processes and Implications for Climate Change*. *Journal of Climate, 28*(3), 1308-1328. doi:10.1175/jcli-d-14-00324.1

Berghuijs, W. R., Woods, R. A., & Hrachowitz, M. (2014). A precipitation shift from snow towards rain leads to a decrease in streamflow. *Nature Climate Change, 4*(7), 583-586. doi:10.1038/nclimate2246

Cannon, A. J. (2016). Multivariate Bias Correction of Climate Model Output: Matching

Marginal Distributions and Intervariable Dependence Structure. *Journal of Climate, 29*(19), 7045-7064. doi:10.1175/jcli-d-15-0679.1

Cannon, A. J. (2017). Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dynamics, 50*(1-2), 31-49. doi:10.1007/s00382-017-3580-6

Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *Journal of Climate, 28*(17), 6938-6959. doi:10.1175/jcli-d-14-00754.1

Chen, J., Brissette, F. P., Poulin, A., & Leconte, R. (2011). Overall uncertainty study of the hydroogical impacts of climate change for a Canadian watershed. *Water Resources Research, 47,* W12509, doi: :10.1029/2011WR010602

Chen, J., Brissette, F. P., Chaumont, D., & Braun, M. (2013a). Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resources Research, 49*(7), 4187-4205. doi:10.1002/wrcr.20331

Chen, J., Brissette, F. P., Chaumont, D., & Braun, M. (2013b). 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. doi:10.1016/j.jhydrol.2012.11.062

Chen, J., Brissette, F. P., Liu, P., & Xia, J. (2017). Using raw regional climate model outputs for quantifying climate change impacts on hydrology. *Hydrological Processes*, *31*(24), 4398-4413. doi:10.1002/hyp.11368

Chen, J., Li, C., Brissette, F. P., Chen, H., Wang, M., & Essou, G. R. C. (2018). Impacts of correcting the inter-variable correlation of climate model outputs on hydrological modeling. *Journal of Hydrology, 560*, 326-341. doi:10.1016/j.jhydrol.2018.03.040

Duan, Q., Sorooshian, S., & Gupta, V. K. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology, 158*(3-4), 265-284. doi:10.1016/0022-1694(94)90057-4

Dunnett C.W. (1955) A multiple comparison procedure for comparing several treatments with a control. J Am Stat Assoc 50(272):1096–1121

Edijatno, De Oliveira Nascimento, N., Yang, X., Makhlouf, Z., & Michel, C. (1999). GR3J: a daily watershed model with three free parameters. *Hydrological Sciences Journal*, *44*(2), 263-277. doi:10.1080/02626669909492221

Guo, Q., Chen, J., Zhang, X., Shen, M., Chen, H., & Guo, S. (2019). A new two-stage multivariate quantile mapping method for bias correcting climate model outputs. *Climate Dynamics*, *53*(5-6), 3603-3623. doi:10.1007/s00382-019-04729-w

Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., Roessler, O., Wibig, J., Wilcke, R., Kotlarski, S., San Martín, D., Herrera, S., Bedia, J., Casanueva, A., Manzanas, R., Iturbide, M., Vrac, M., Dubrovsky, M., Ribalaygua, J., Pórtoles, J., Räty, O., Räisänen, J., Hingray, B., Raynaud, D., Casado, M. J., Ramos, P., Zerenner, T., Turco, M., Bosshard, T., Štěpánek, P., Bartholy, J., Pongracz, R., Keller, D. E., Fischer, A. M., Cardoso, R. M., Soares, P. M. M., Czernecki, B., & Pagé, C. (2019). An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. *International Journal of Climatology, 39*(9), 3750-3785. doi:10.1002/joc.5462

Hakala, K., Addor, N., & Seibert, J. (2018). Hydrological Modeling to Evaluate Climate Model Simulations and Their Bias Correction. *Journal of Hydrometeorology, 19*(8), 1321-1337. doi:10.1175/jhm-d-17-0189.1

Hao, Z., Phillips, T. J., Hao, F., & Wu, X. (2019). Changes in the dependence between global precipitation and temperature from observations and model simulations. *International Journal of Climatology*. doi:10.1002/joc.6111

Hui, Y., Chen, J., Xu, C. Y., Xiong, L., & Chen, H. (2018). Bias nonstationarity of global climate model outputs: The role of internal climate variability and climate model sensitivity. *International Journal of Climatology, 39*(4), 2278-2294. doi:10.1002/joc.5950

Iman, R. L., & Conover, W. J. (1982). A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics - Simulation and Computation, 11*(3), 311-334.

Jefferson, A., Nolin, A., Lewis, S., & Tague, C. (2008). Hydrogeologic controls on streamflow sensitivity to climate variation. *Hydrological Processes, 22*(22), 4371-4385. doi:10.1002/hyp.7041

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift, 15*(3), 259-263. doi:10.1127/0941-2948/2006/0130

Kumar, D., Kodra, E., & Ganguly, A. R. (2014). Regional and seasonal intercomparison of CMIP3 and CMIP5 climate model ensembles for temperature and precipitation. *Climate Dynamics, 43*(9-10), 2491-2518. doi:10.1007/s00382-014-2070-3

Kunnath-Poovakka, A., & Eldho, T. I. (2019). A comparative study of conceptual rainfall-runoff models GR4J, AWBM and Sacramento at catchments in the upper Godavari river basin, India. *Journal of Earth System Science, 128*(2). doi:10.1007/s12040-018-1055-8

Li, C., Sinha, E., Horton, D. E., Diffenbaugh, N. S., & Michalak, A. M. (2014). Joint bias correction of temperature and precipitation in climate model simulations. *Journal of Geophysical Research: Atmospheres, 119*(23), 13,153-113,162. doi:10.1002/2014jd022514

Li, W., Chen, J., Li, L., Chen, H., Liu, B., Xu, C.-Y., & Li, X. (2019). Evaluation and bias correction of S2S precipitation for hydrological extremes. *Journal of Hydrometeorology*. doi:10.1175/jhm-d-19-0042.1

Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., Maurer, E. P., & Lettenmaier, D. P. (2013). A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions. *Journal of Climate, 26*(23), 9384-9392. doi:10.1175/jcli-d-12-00508.1

Mahony, C. R., & Cannon, A. J. (2018). Wetter summers can intensify departures from natural variability in a warming climate. *Nat Commun, 9*(1), 783. doi:10.1038/s41467-018-03132-z

Maraun, D. (2012). Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. *Geophysical Research Letters, 39*(6). doi:10.1029/2012gl051210

Maraun, D. (2016). Bias Correcting Climate Change Simulations - a Critical Review. *Current Climate Change Reports, 2*(4), 211-220. doi:10.1007/s40641-016-0050-x

Maraun, D., Widmann, M., & Gutiérrez, J. M. (2019). Statistical downscaling skill under present climate conditions: A synthesis of the VALUE perfect predictor experiment. *International Journal of Climatology, 39*(9), 3692-3703. doi:10.1002/joc.5877

Mehran, A., AghaKouchak, A., & Phillips, T. J. (2014). Evaluation of CMIP5 continental precipitation simulations relative to satellite-based gauge-adjusted observations. *Journal of Geophysical Research: Atmospheres, 119*(4), 1695-1707. doi:10.1002/2013jd021152

Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J., & Cannon, A. J. (2019). Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments. *Hydrology and Earth System Sciences, 23*(3), 1339-1354. doi:10.5194/hess-23-1339-2019

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology, 10*(3), 282-290. doi:10.1016/0022-1694(70)90255-6

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model? *Journal of Hydrology, 303*(1-4), 290-306. doi:10.1016/j.jhydrol.2004.08.026

Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology, 279*(1-4), 275-289. doi:10.1016/s0022-1694(03)00225-7

Räty, O., Räisänen, J., Bosshard, T., & Donnelly, C. (2018). Intercomparison of Univariate and Joint Bias Correction Methods in Changing Climate From a Hydrological Perspective. *Climate*, *6*(2), 33. doi:10.3390/cli6020033

Seo, S. B., Das Bhowmik, R., Sankarasubramanian, A., Mahinthakumar, G., & Kumar, M. (2019). The role of cross-correlation between precipitation and temperature in basin-scale simulations of hydrologic variables. *Journal of Hydrology, 570*, 304-314. doi:10.1016/j.jhydrol.2018.12.076

Shen, M., Chen, J., Zhuan, M., Chen, H., Xu, C.-Y., & Xiong, L. (2018). Estimating uncertainty and its temporal variation related to global climate models in quantifying climate change impacts on hydrology. *Journal of Hydrology, 556*, 10-24. doi:10.1016/j.jhydrol.2017.11.004

Shrestha, R. R., Cannon, A. J., Schnorbus, M. A., & Alford, H. (2019). Climatic Controls on Future Hydrologic Changes in a Subarctic River Basin in Canada. *Journal of Hydrometeorology*. doi:10.1175/jhm-d-18-0262.1

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society, 93*(4), 485-498. doi:10.1175/bams-d-11-00094.1

Valéry, A., Andréassian, V., & Perrin, C. (2014). '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. doi:10.1016/j.jhydrol.2014.04.058

Vano, J. A., Das, T., & Lettenmaier, D. P. (2012). Hydrologic Sensitivities of Colorado River Runoff to Changes in Precipitation and Temperature*. *Journal of Hydrometeorology, 13*(3), 932-949. doi:10.1175/jhm-d-11-069.1

Volosciuk, C., Maraun, D., Vrac, M., & Widmann, M. (2017). A combined statistical bias correction and stochastic downscaling method for precipitation. *Hydrology and Earth System Sciences, 21*(3), 1693-1719. doi:10.5194/hess-21-1693-2017

Vrac, M., & Friederichs, P. (2015). Multivariate—Intervariable, Spatial, and Temporal—Bias Correction*. *Journal of Climate*, *28*(1), 218-237. doi:10.1175/jcli-d-14-00059.1

Yang, X., Magnusson, J., Huang, S., Beldring, S., & Xu, C.-Y. (2019). Dependence of regionalization methods on complexity of hydrological models in multiple climate regions. *Journal of Hydrology*, 582: 124357. doi: 10.1016/j.jhydrol.2019.124357

Tables

Table 1 Koeppen-Geiger climate classification of 2840watersheds and the basin number of each climate regime

Climate (Basin Number)	**Characteristics**	**Climate (Basin Number)**	**Characteristics**	**Climate (Basin Number)**	**Characteristics**
BSk (47)	Arid, Steppe, Cold Arid	Csb (215)	Warm Temperate, Summer Dry, Warm Summer	Dsb (49)	Snow, Summer Dry, Warm Summer
Cfa (981)	Warm Temperate, Fully Humid, Hot Summer	Dfa (350)	Snow, Fully Humid, Hot Summer	Dsc (18)	Snow, Summer Dry, Cool Summer
Cfb (93)	Warm Temperate, Fully Humid, Warm Summer	Dfb (675)	Snow, Fully Humid, Warm Summer	ET (30)	Polar, Polar Tundra
Csa (20)	Warm Temperate, Summer Dry, Hot Summer	Dfc (362)	Snow, Fully Humid, Cool Summer		

Table 2 The summary of the seven bias correction methods

Name	Type	Description	Citation
Daily bias correction (DBC)	Univariate	Quantile mapping-based, correcting the biases in the cumulative distribution function of each variable	Chen et al., 2013b
Joint bias correction (JBC)	Multivariate	Copula function-based, establishing the copula function for multiple variable first and then correcting one variable conditionally upon the other one	Li et al., 2014
Multivariate bias correction: Pearson version (MBCp)	Multivariate	Shuffle-based, inducing the desired Pearson correlation matrix by iteration algorithm	Cannon, 2016
Multivariate bias correction: Spearman version (MBCr)	Multivariate	Shuffle-based, inducing the desired Spearman correlation matrix by iteratively adjusting the time sequence of simulated variable	Cannon, 2016
Multivariate bias correction: N-dimensional probability density function transform (MBCn)	Multivariate	Rotation-based, inducing the desired multivariate distribution function by iteratively rotation and correction	Cannon, 2017
Two-stage quantile mapping (TSQM)	Multivariate	Shuffle-based, inducing the desired Spearman correlation matrix by distribution-free shuffle algorithm	Guo et al., 2019
Empirical Copula bias correction (ECBC)	Multivariate	Shuffle-based, reproducing the observed correlation matrix by reordering the time sequence of simulated variables to corresponding observations	Vrac & Friederichs, 2015

Table 3 The ratio (%) that the mean values of RMSE of multivariate methods are smaller than DBC in simulating the 9 streamflow variables for the validation period

	Spring mean flow	**Summer mean flow**	**Autumn mean flow**	**Winter mean flow**	**High flow**	**Low flow**	**Peak time**	**Flood begin time**	**Flood end time**	**In general**
Arid	67	17	67	83	67	67	100	83	83	70
Warm temperate	88	42	67	79	71	50	83	79	79	71
Snow	53	10	67	73	53	57	57	40	77	54
Polar	33	0	33	50	67	33	83	100	100	55
In general	67	27	65	74	59	55	73	65	82	63

Figure captions

Fig. 1 The 11 climate regimes of the 2840 watersheds in North America based on Koeppen-Geiger climate classification

Fig. 2 The Nash-Sutcliffe efficiency for the 2840 watersheds in both calibration and validation periods using GR4J-9 hydrological model

Fig. 3 The root-mean-square-error of variable means for P, T_{max}, and T_{min} for 4 typical months over the 2840 watersheds in the validation period. The x-axis label from left to right is DBC, JBC, MBCp, MBCr, MBCn, TSQM and ECBC, respectively, representing seven bias correction methods. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the red '+' symbol

Fig. 4 The spatial distribution of observed and corrected simulated (BCC-CSM1.1 (m) as an example) P-T_{max} correlation of July for the 2840 watersheds in the validation period. OBS represents P-T_{max} correlation of reference data, and DBC, JBC, MBCp, MBCr, MBCn, TSQM and ECBC represent the corrected P-T_{max} correlation by the corresponding bias correction method, respectively

Fig. 5 The root-mean-square-error for the 2840 watersheds between the observed and corrected P-T_{max} and P-T_{min} correlation for 4 typical months in both calibration ((a) and (b)) and validation ((c) and (d)) periods. The x-axis label from left to right is DBC, JBC, MBCp, MBCr, MBCn, TSQM and ECBC, respectively, representing seven bias correction methods. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the red '+' symbol

Fig. 6 The spatial distribution of the absolute error in % of winter snowfall proportion between the reference data and the model outputs (including the raw data and corrected data) for the 2840 watersheds in the validation period. RAW represents the result of climate model outputs without bias correction, and DBC, JBC, MBCp, MBCr, MBCn, TSQM and ECBC represent the result of the corresponding bias correction method, respectively

Fig. 7 The spatial distribution of the absolute relative error (%) of spring daily mean snowmelt between the reference data and the model outputs (including the raw data and corrected data) for the 2840 watersheds in the validation period. RAW represents the result of climate model outputs without bias correction, and DBC, JBC, MBCp, MBCr, MBCn, TSQM and ECBC represent the result of the corresponding bias correction method, respectively

Fig. 8 The spatial distribution of the daily potential evaporation difference between dry day and wet day (dry days minus wet days) in summer between the reference data and the corrected model outputs for the 2840 watersheds in the validation period. OBS represents the result of reference data, and DBC, JBC, MBCp, MBCr, MBCn, TSQM and ECBC represent the result of the corresponding bias correction method, respectively

Fig. 9 The Dunnett-t test of root-mean-square-error in simulating winter snowfall proportion, spring daily mean snowmelt and summer wet day evaporation in both calibration and validation periods. The x-axis label represents 11 climate regimes, and the y-axis label represents 6 multivariate bias correction methods. The blue color represents the multivariate method is significantly better than DBC, while the red color represents the multivariate method is significantly worse than DBC, and white color represents there is no significant difference between these two methods. This is also applicable to the Figures hereafter

Fig. 10 The Dunnett-t test of root-mean-square-error in simulating mean streamflow for 4 seasons in both calibration and validation periods. The x-axis label represents 11 climate regimes, and y-axis label represents 6 multivariate bias correction methods

Fig. 11 The Dunnett-t test of root-mean-square-error in simulating high and low flow in both calibration and validation periods. The x-axis label represents 11 climate regimes, and y-axis label represents 6 multivariate bias correction methods

Fig. 12 The Dunnett-t test of root-mean-square-error in simulating time variables in both calibration and validation periods. The x-axis label represents 11 climate regimes, and y-axis label represents 6 multivariate bias correction methods

Figure 1.

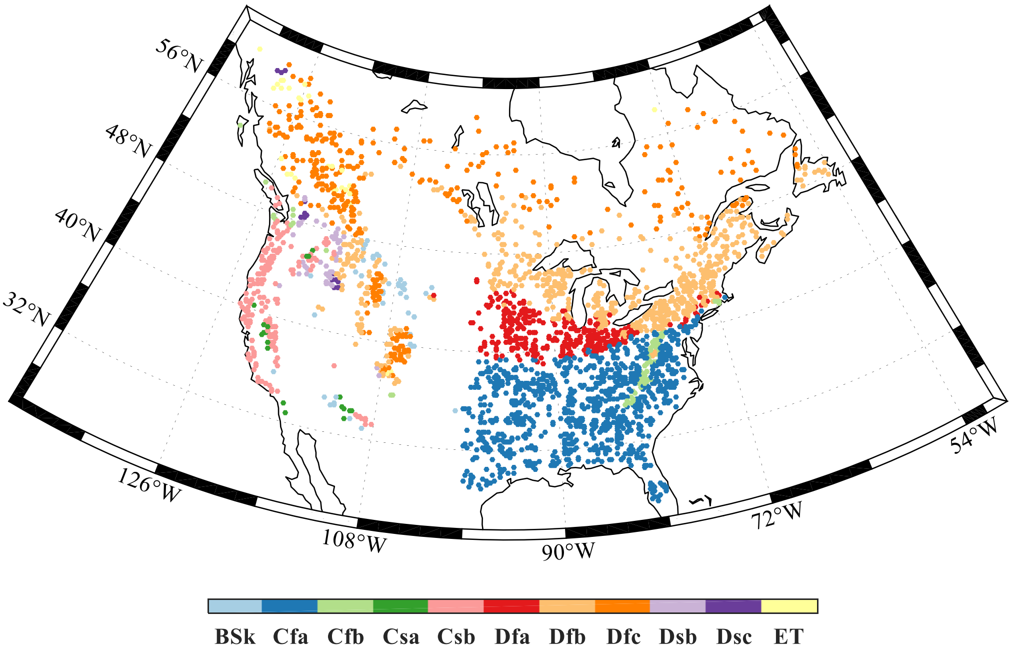
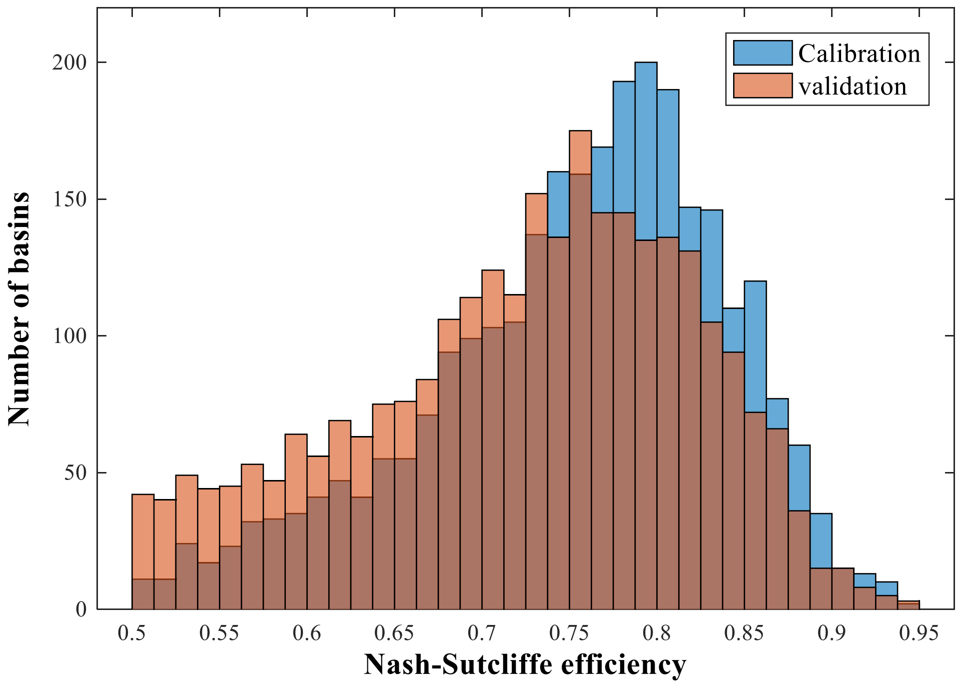

56°N
48°N
40°N
32°N
126°W
108°W
90°W
72°W
54°W
BSk
Cfa
Cfb
Csa
Csb
Dfa
Dfb
Dfc
Dsb
Dsc
ET

Figure 2.



Calibration
validation
Number of basins
Nash-Sutcliffe efficiency
0
50
100
150
200
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95

Figure 3.

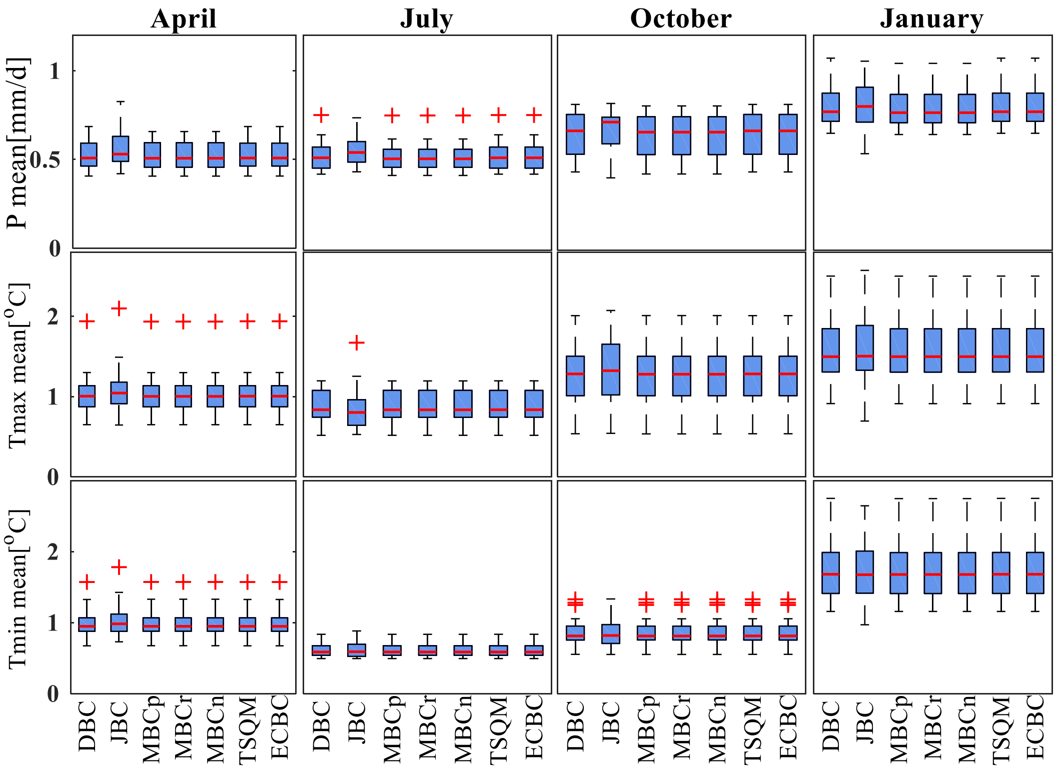
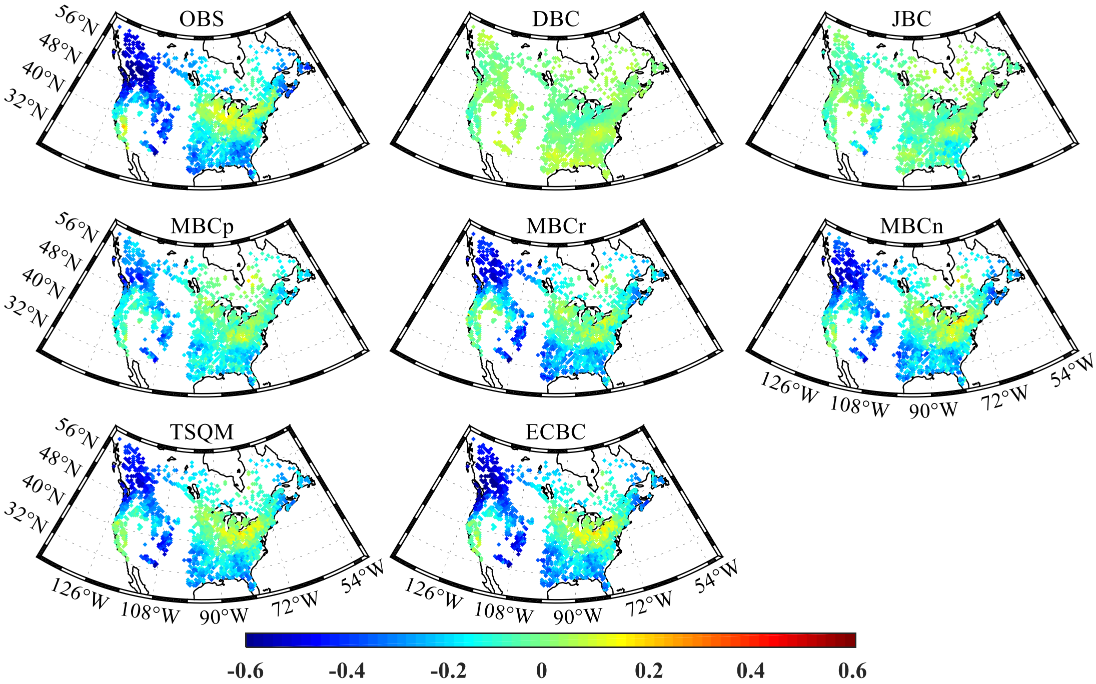

April
July
October
January
P mean[mm/d]
Tmax mean[oC]
Tmin mean[oC]
0
0.5
1
2
DBC
JBC
MBCp
MBCr
MBCn
TSQM
ECBC

Figure 4.



OBS
DBC
JBC
MBCp
MBCr
MBCn
TSQM
ECBC
56°N
48°N
40°N
32°N
126°W
108°W
90°W
72°W
54°W
-0.6
-0.4
-0.2
0
0.2
0.4
0.6

Figure 5.

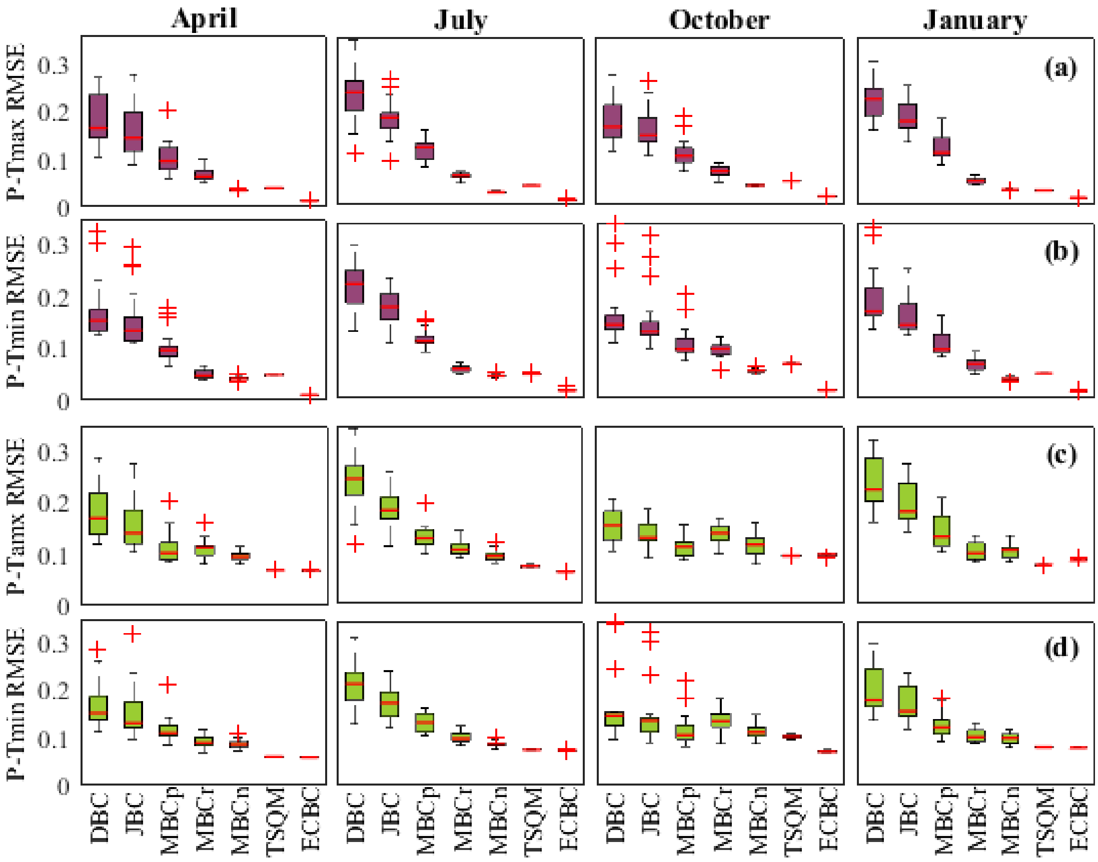
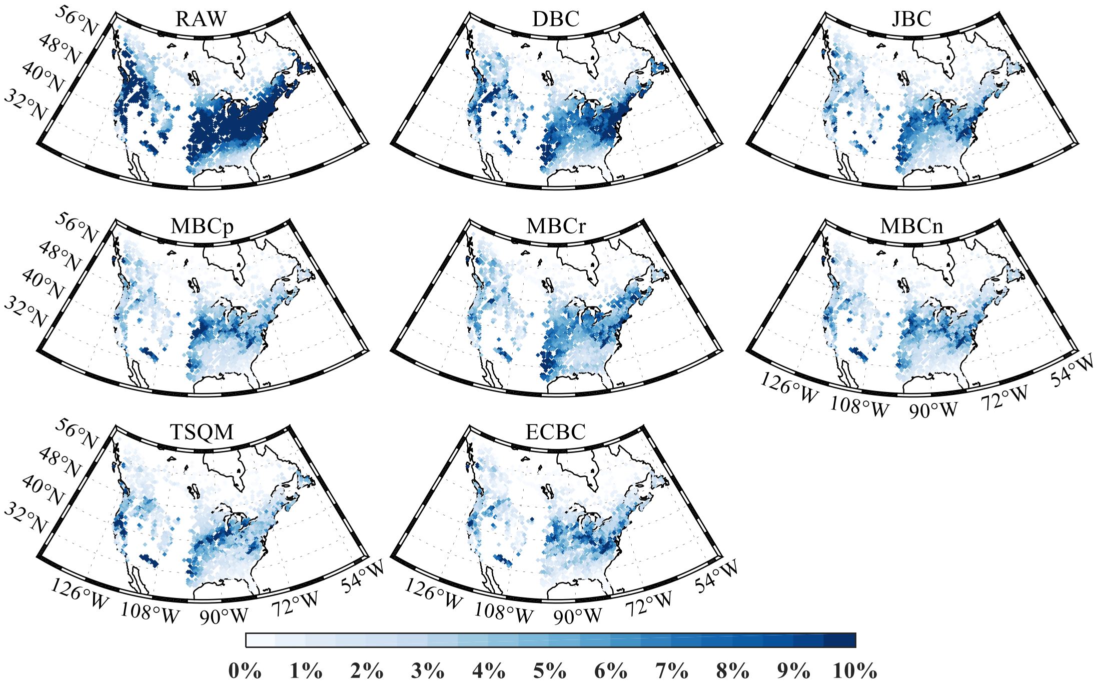

April
July
October
January
(a)
(b)
(c)
(d)
P-Tmax RMSE
P-Tmin RMSE
P-Tamx RMSE
P-Tmin RMSE
0
0.1
0.2
0.3
DBC
JBC
MBCp
MBCr
MBCn
TSQM
ECBC

Figure 6.



RAW
DBC
JBC
MBCp
MBCr
MBCn
TSQM
ECBC
56°N
48°N
40°N
32°N
126°W
108°W
90°W
72°W
54°W
0%
1%
2%
3%
4%
5%
6%
7%
8%
9%
10%

Figure 7.

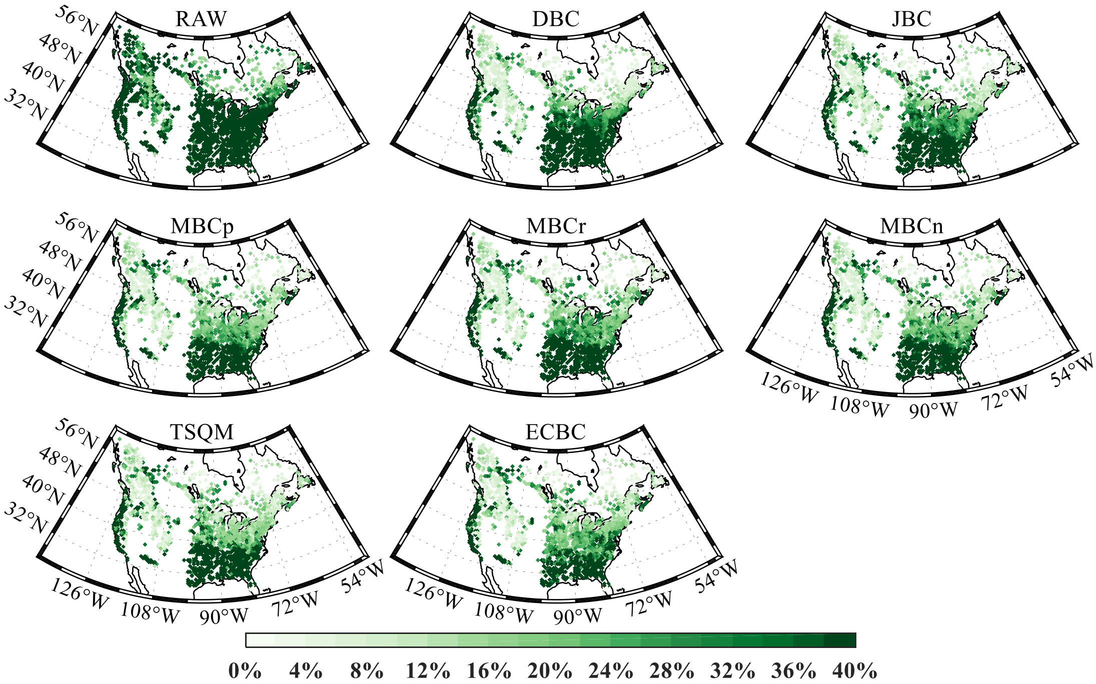

RAW
DBC
JBC
MBCp
MBCr
MBCn
TSQM
ECBC
56°N
48°N
40°N
32°N
126°W
108°W
90°W
72°W
54°W
0%
4%
8%
12%
16%
20%
24%
28%
32%
36%
40%

Figure 8.

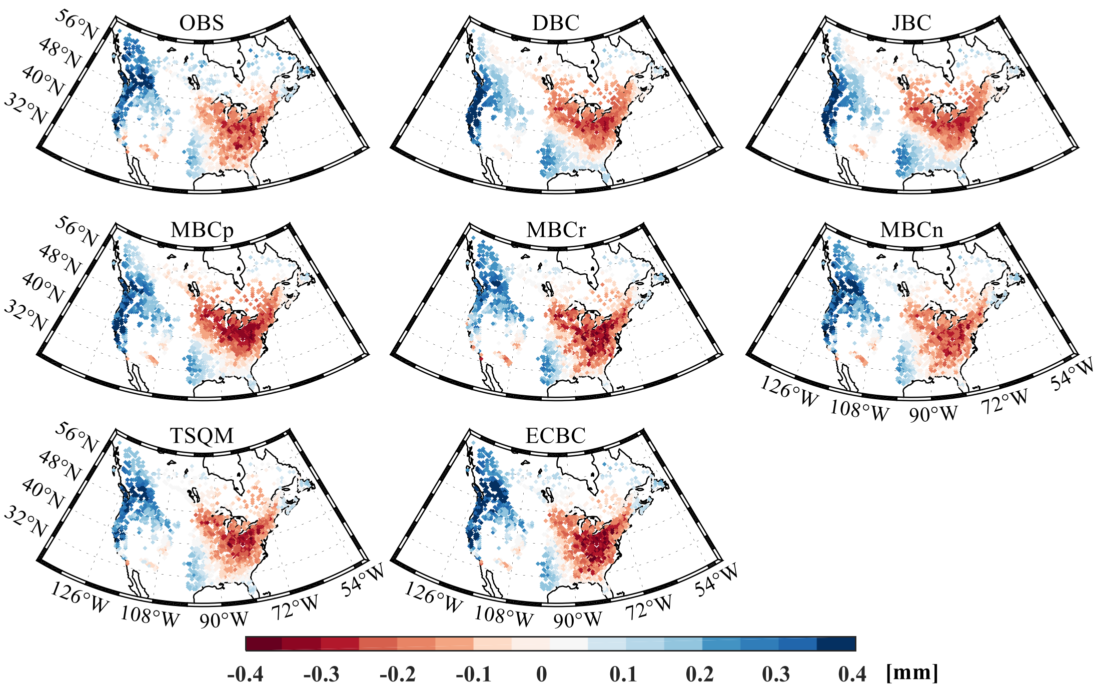
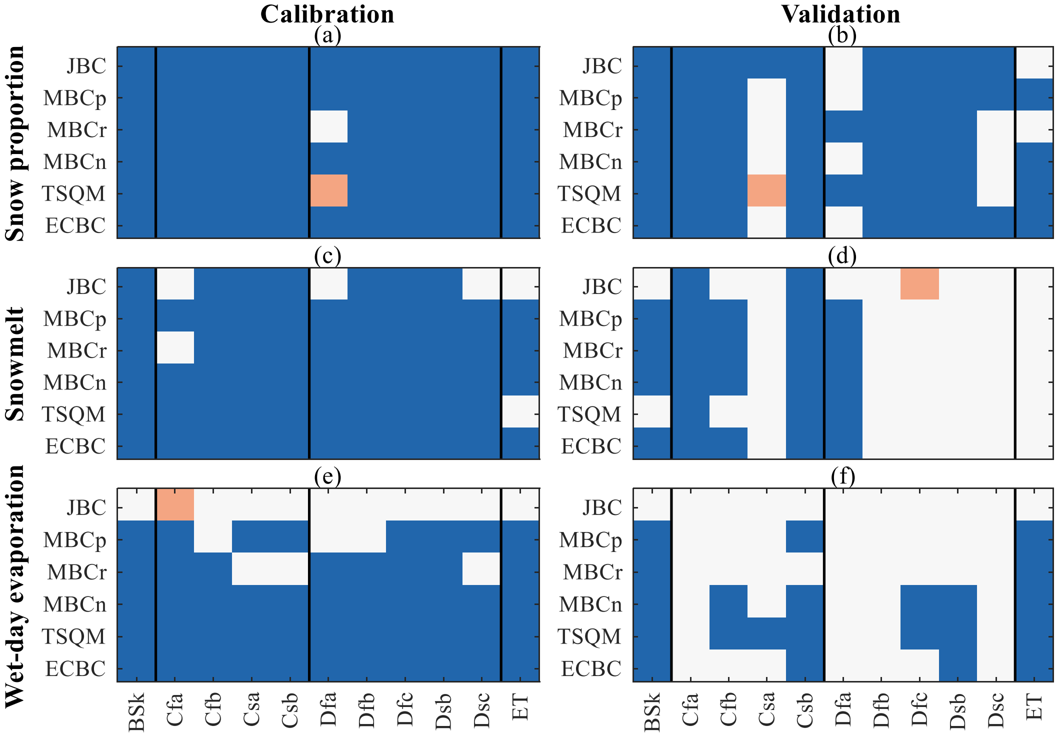
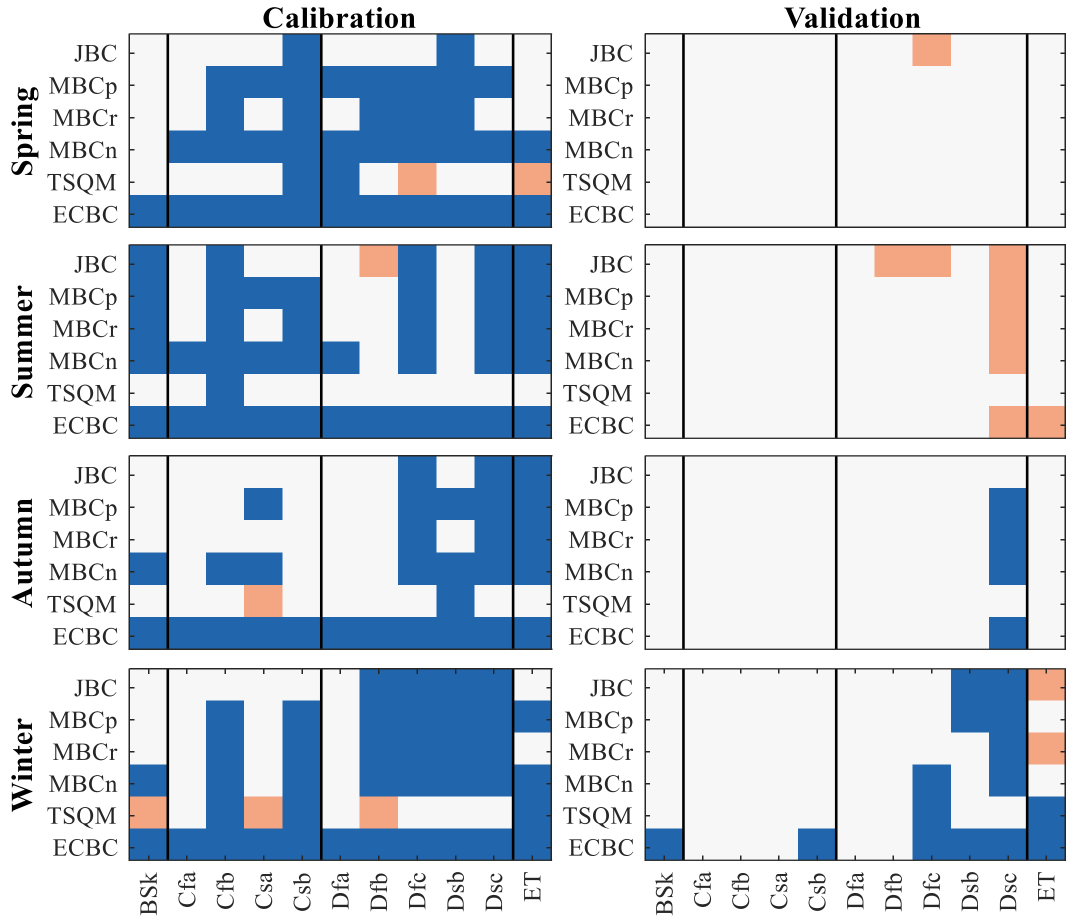

OBS
DBC
JBC
MBCp
MBCr
MBCn
TSQM
ECBC
56°N
48°N
40°N
32°N
126°W
108°W
90°W
72°W
54°W
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
[mm]

Figure 9.



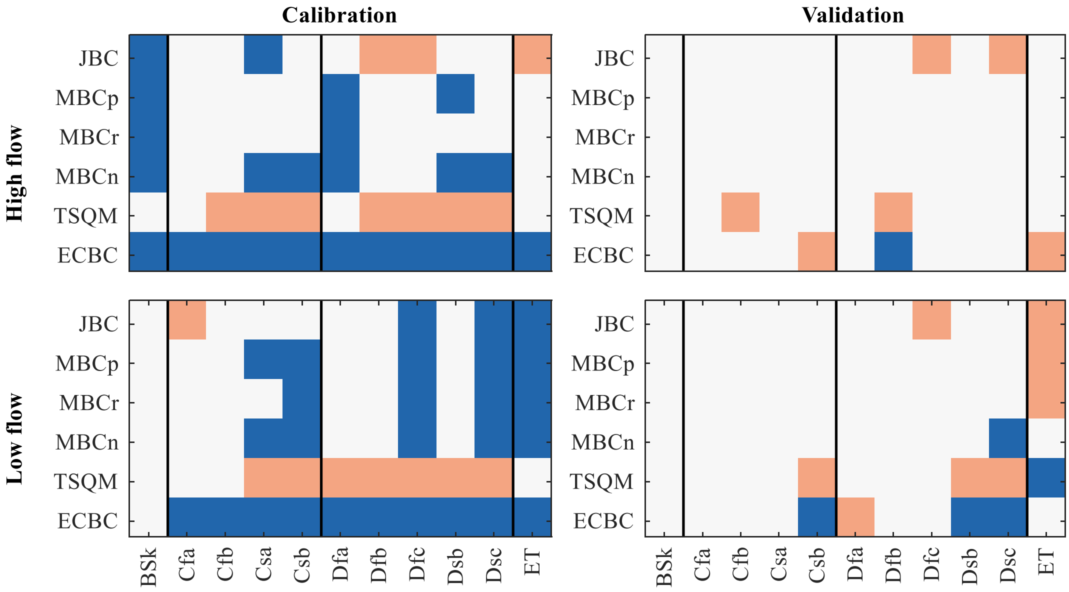
Calibration
Validation
(a)
(b)
(c)
(d)
(e)
(f)
Snow proportion
Snowmelt
Wet-day evaporation
JBC
MBCp
MBCr
MBCn
TSQM
ECBC
BSk
Cfa
Cfb
Csa
Csb
Dfa
Dfb
Dfc
Dsb
Dsc
ET

Figure 10.



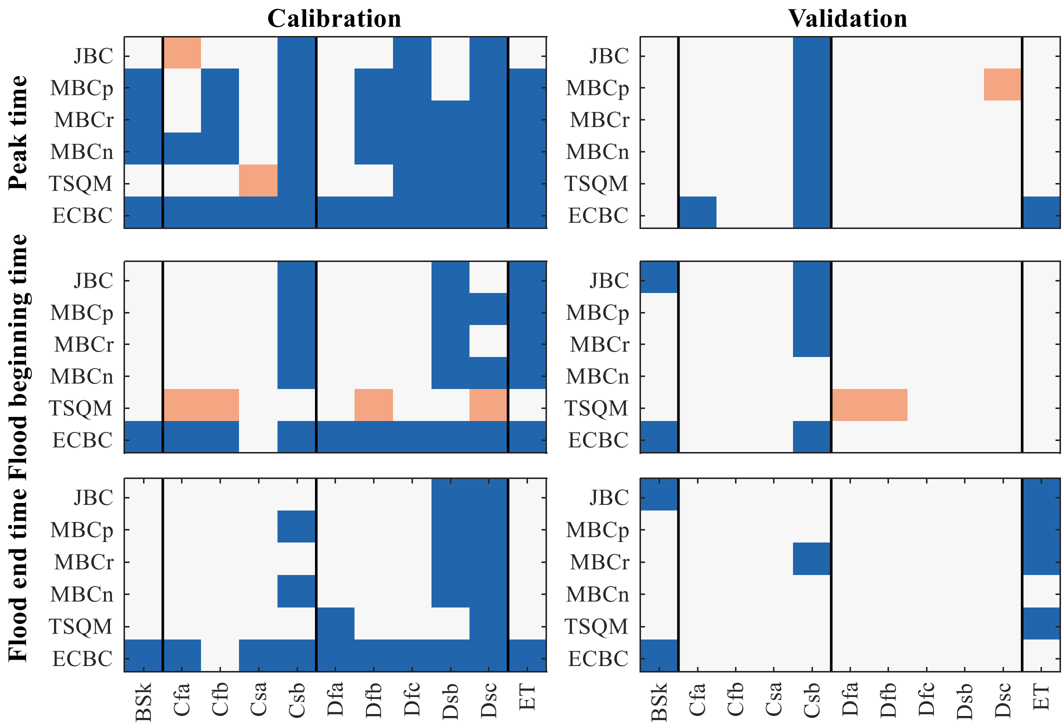
Calibration
Validation
Spring
Summer
Autumn
Winter
JBC
MBCp
MBCr
MBCn
TSQM
ECBC
BSk
Cfa
Cfb
Csa
Csb
Dfa
Dfb
Dfc
Dsb
Dsc
ET

Figure 11.



Calibration
Validation
High flow
Low flow
JBC
MBCp
MBCr
MBCn
TSQM
ECBC
BSk
Cfa
Cfb
Csa
Csb
Dfa
Dfb
Dfc
Dsb
Dsc
ET

Figure 12.



Calibration
Validation
Peak time
Flood beginning time
Flood end time
JBC
MBCp
MBCr
MBCn
TSQM
ECBC
BSk
Cfa
Cfb
Csa
Csb
Dfa
Dfb
Dfc
Dsb
Dsc
ET