Postprocessing Ensemble Weather Forecasts for Introducing Multisite and Multivariable Correlations Using Rank Shuffle and Copula Theory

JIE CHEN,[a] XIANGQUAN LI,[b] CHONG-YU XU,[c] XUNCHANG JOHN ZHANG,[d] LIHUA XIONG,[a] AND QIANG GUO[a]

[a] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, China
[b] Changjiang Institute of Survey, Planning, Design and Research, Wuhan, China
[c] Department of Geosciences, University of Oslo, Oslo, Norway
[d] USDA-ARS Grazinglands Research Laboratory, El Reno, Oklahoma

(Manuscript received 30 April 2021, in final form 5 December 2021)

ABSTRACT: Statistical methods have been widely used to postprocess ensemble weather forecasts for hydrological predictions. However, most of the statistical postprocessing methods apply to a single weather variable at a single location, thus neglecting the intersite and intervariable dependence structures of forecast variables. This study synthesized a multisite and multivariate (MSMV) postprocessing framework that extends the univariate method to the MSMV version by directly rearranging the postprocessed ensemble members (post-reordering strategy) or by rearranging the latent variables used in the univariate method (pre-reordering strategy). Based on the univariate generator-based postprocessing (GPP) method, the two reordering strategies and three dependence reconstruction methods [rank shuffle (RS), Gaussian copula (GC), and empirical copula (EC)] totaling six MSMV methods (RS-Pre, GC-Pre, EC-Pre, RS-Post, GC-Post, and EC-Post) were evaluated in postprocessing ensemble precipitation and temperature forecasts for the Xiangjiang basin in China using the 11-member ensemble forecasts from the Global Ensemble Forecasting System (GEFS). The results showed that raw GEFS forecasts tend to be biased for both the forecast ensembles and the intersite and intervariable dependencies. The univariate method can improve the univariate performance of ensemble mean and spread but misrepresent the intersite and intervariable dependence among the forecast variables. The MSMV framework can well utilize the advantages of the univariate method and also reconstruct the intersite and intervariable dependencies. Among the six methods, RS-Pre, RS-Post, GC-Post, and EC-Post perform better than the others with respect to reproducing the univariate statistics and multivariable dependences. The post-reordering strategy is recommended to combine the univariate method (i.e., GPP) and reconstruction methods.

KEYWORDS: Bias; Ensembles; Downscaling; Forecasting

1. Introduction

Ensemble weather forecasting (EWF) has become the state-of-the-art method of numerical weather prediction (NWP) since the 1990s (Gneiting and Raftery 2005). When running the NWP models, the perturbations from analysis and model errors are added to the initial state and the model physical process for generating the ensemble weather forecasts (Bauer et al. 2015). The ensemble weather forecasts are capable of predicting the flow-dependent variations and showing greater economic value than a best-guess deterministic forecast (Zhu et al. 2002; Zhu 2005; Leutbecher and Palmer 2008). Despite the above advantages, ensemble weather forecasts are still biased and typically underdispersed (Hagedorn et al. 2008; Hamill et al. 2008; Scheuerer and Hamill 2015a). Therefore, the raw ensemble weather forecasts need to be improved by postprocessing methods before used in an environmental model for environmental predictions. Various methods, like nonhomogeneous Gaussian regression [(NGR) or ensemble model output statistics (EMOS)] (Gneiting et al. 2005), Bayesian model averaging (BMA; Raftery et al. 2005; Sloughter et al. 2007), kernel dressing (Bröcker and Smith 2008), logistic regression (Wilks 2009), and generator-based postprocessing (GPP; Chen et al. 2014a), have been developed.

However, most statistical postprocessing methods apply to a single weather variable at a single location, thus neglecting the spatiotemporal dependence structures present in the real climate system. The spatiotemporal dependence indicates the physical connection of different variables in the neighboring region, thus considerably influencing the performance of EWF and its applications. For example, Keune et al. (2014) found that spatiotemporal postprocessing would enhance the predictable signal while the univariate postprocessing might undermine the predictive accuracy.

Direct extension of univariate methods to multisite and multivariate (MSMV) is feasible but restricted to cases where the considered variables are limited (low-dimensional case) or highly structured. For example, in the low-dimensional case where the forecast errors are assumed to follow a multivariable normal distribution, the multivariable dependence can be directly modeled using the covariance matrix (Pinson et al. 2009; Schuhen et al. 2012; Sloughter et al. 2013). In the highly structured case where the weather field forecasts have Gaussian error distributions, the spatial dependence structure can be modeled via a geostatistical output perturbation (GOP) method which requires estimating the spatial covariance parameters (Gel et al. 2004; Berrocal et al. 2007, 2008; Feldmann et al. 2015). The direct extension of univariate methods has a strict

Corresponding authors: Jie Chen, jiechen@whu.edu.cn; Xiangquan Li, leexiangquan@whu.edu.cn

DOI: 10.1175/MWR-D-21-0100.1

restriction of the application scenarios and requires estimating a large number of parameters even in the above special cases (Wilks 2015).

For the high-dimensional case where a large number of variables are simultaneously considered and their marginal distributions are usually assumed to be of different types, the copula methods show advantages, as they allow independently modeling the marginal distribution and multivariable dependence structure, exhibiting great benefits in many areas, for example, in the hydrological time series analysis (Xiong et al. 2014, 2015; Jiang et al. 2019). The copula methods are easily combined with the univariate methods and the reconstructed multivariable dependencies which may otherwise be misrepresented or lost if only univariate methods are applied as presented below.

The copula methods can be used before or after the implementation of univariate method, which is, respectively, defined as pre-reordering or post-reordering in this study. The post-reordering strategy combines the univariate methods and the dependence reconstruction method for devising the MSMV method (Schefzik et al. 2013; Wilks 2015; Schefzik 2016, 2017). For example, Schefzik et al. (2013) proposed a multistage procedure for high-dimensional ensemble weather forecasts postprocessing. In this study, univariate methods were first used for obtaining the calibrated and sharp predictive distributions of each variable at a single location. The rank of multivariables estimated from the raw ensemble forecasts using empirical copula (EC) was then used to rearrange the generated samples from the marginal predictive distribution.

In contrast, the pre-reordering strategy uses the rank of the past observations for rearranging the independent samples from the marginal distribution for devising the MSMV method (Schefzik 2016). For example, Möller et al. (2013) used the pre-reordering strategy that combines the univariate BMA method and the Gaussian copula (GC) method. Specifically, the GC with an estimated parameter, the multivariable residual correlation matrix, was used to model the dependencies of the latent variable. The latent variable was then used in the univariate predictive distribution for generating the postprocessed forecast ensembles.

When devising the MSMV methods, the key components are the univariate methods, the dependence reconstruction methods, and how the two are combined. Several univariate methods have been proved effective in postprocessing the ensemble weather forecasts, and extensively evaluated and compared for determining their advantages and disadvantages (Wilks 2006, 2015; Wilks and Hamill 2007; Schmeits and Kok 2010; Li et al. 2019, 2020). Choosing the dependence reconstruction methods is vital for devising the MSMV methods, but the choice is generally made according to the author's experiences and preferences. This requires an intercomparison and evaluation of these methods for documenting their advantages and disadvantages. Besides, the vital component for devising the MSMV methods is how the univariate methods and the dependence reconstruction methods are combined, which is generally less considered and discussed in the literature. More importantly, although the pre-reordering and post-reordering strategies have been individually used in different studies, they are not specifically compared in terms of combining univariate methods and the dependence reconstruction methods for EWFs.

Therefore, the main objective of this study is to formulate a general MSMV framework by extending the commonly used univariate method to MSMV version via pre-reordering or post-reordering the dependence reconstruction methods. The performance of different dependence reconstruction methods [i.e., EC, GC, and rank shuffle (RS)] and different reordering strategies (i.e., post-reordering and pre-reordering) are compared to find appropriate MSMV methods for postprocessing of EWFs. Even though the univariate method and dependence reconstruction methods used in this study are all available in the literature, this work first synthesizes them and investigates the influence of using pre-reordering and post-reordering strategies on the postprocessing of EWFs. This study considers postprocessing ensemble precipitation and air temperature forecasts over multiple stations in a Chinese watershed as an example. The ensemble weather forecasts were taken from the second version of the Global Ensemble Forecast System (GEFS) reforecasts.

2. Study area and data

a. Study area

The performances of the MSMV framework are evaluated over the Xiangjiang River basin (Fig. 1). The Xiangjiang River, with a length of 856 km and a total drainage area of 94 660 km^2, belongs to the Dongting Lake drainage system in the middle section of the Yangtze River. Mountains (the mean elevation > 200 m MSL) are distributed in the eastern and southern upstream areas, while the plains (the mean elevation < 100 m MSL) are mainly in the central, northern, and western downstream areas. The mean annual precipitation is about 1584 mm, and the mean annual temperature is about 17°C. The Xiangjiang River basin is strongly influenced by the Pacific monsoon climate, which brings about 70% of the annual total precipitation during the rainy seasons from April to September and 50% of the flooding events occur in June and July (Xu et al. 2013).

b. Data

The dataset consists of observed and EWF daily total precipitation and mean air temperature. The observed data were taken from the China Meteorological Data Sharing Service System (http://cdc.cma.gov.cn). The observed precipitation and mean air temperature are available from 1961 to 2014 and achieved in the grid of 0.5° in both latitude and longitude. This dataset was created by interpolating the station data to a 0.5° grid using a modified kriging interpolation method. The EWF data were obtained from the second version of the Global Ensemble Forecast System (GEFS) reforecasts (http://portal.nersc.gov/project/refcst/v2/). The GEFS reforecasts provide 11-member ensemble forecasts for precipitation, mean air temperature over 16 lead days available from December 1984 to the present. The forecast data are archived in a global grid of 1° on latitude and longitude.

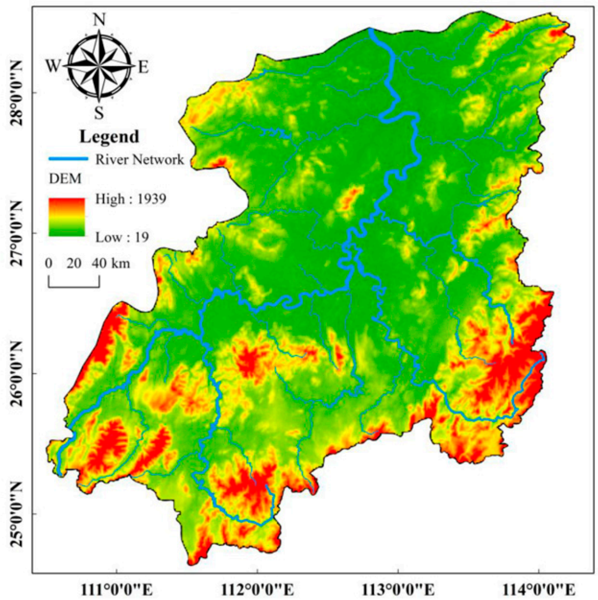

FIG. 1. The map of the Xiangjiang River basin.

This study considers the period of 1985–2014 for evaluating the postprocessing methods. The lead time of ~1–7 days was used, since the previous studies have shown that the precipitation forecasts lost their forecast skill after one week (e.g., Liu and Coulibaly 2011; Chen et al. 2014a; Zhang et al. 2019). Both the observed and forecast data were bilinearly interpolated to the grids with a common resolution of 0.3° on latitude and longitude, forming 116 grids in the study basin. The interpolation makes the spatial dependence of the nearby grids is closer to the real world system. The methods are evaluated using a cross-validation approach with 30-yr observations and forecasts. Specifically, when postprocessing the forecasts of a selected year, the remaining 29-yr observations and forecast data were used for estimating the parameters for the univariate method and the dependence reconstructing methods. The 30-yr postprocessed ensemble weather forecasts with the desired MSMV dependence are obtained, when repeating the above procedure for each year from 1985 to 2014.

3. Methodology

a. MSMV methods

The theoretical basis for the MSMV framework is derived from ensemble member reordering (Flowerdew 2012). Specifically, the spatial and intervariable dependencies of the ensemble forecasts can be properly preserved by directly rearranging the ensemble members (post-reordering) or by rearranging the latent variables determining the ensemble members (pre-reordering). The temporal dependence is not specifically considered in devising the MSMV methods. The pattern for rearranging the ensemble members or the latent variables can be either obtained from the raw ensemble forecasts or the past observations (Wilks 2015). This study uses past observations because both the multivariable rank structure and the multivariable correlation matrix can be extracted/estimated from past observations. This guarantees that the dependence reconstruction methods use the same pattern for reconstructing the dependence structure of the weather variables. Besides, using the past observations as the pattern does not have the requirement of the ensemble size after postprocessing as using the raw ensemble forecasts does.

The GPP method proposed by Chen et al. (2014a) is chosen as the univariate method in this study, since previous comparison studies (Li et al. 2019, 2020) have shown that GPP performed similar to or slightly better than others. However, other univariate methods can also be used. The chosen dependence reconstruction methods consist of three commonly used methods: RS, EC, and GC. The univariate method and dependence reconstruction methods are combined by using either pre-reordering or post-reordering strategy. A brief introduction to these methods is given as follows.

1) THE UNIVARIATE METHOD

GPP is used to postprocess ensemble precipitation and temperature forecasts. The predictive distribution of precipitation and temperature estimated from the univariate methods is denoted as $F(y_m|x_{1,m}, \ldots, x_{L,m})$, where y_m is the univariate weather quantity of the variable m; $x_{1,m}, \ldots, x_{L,m}$ are the corresponding L-member raw ensemble forecasts (L is the ensemble size of the raw forecasts and equals to 11 in this study).

(i) Univariate method for precipitation

When using the GPP method, the predictive distribution of precipitation can be expressed as follows:

$$F\left(y_m \middle| x_{1,m}, \ldots, x_{L,m}\right) = \begin{matrix} 0 & \text{if} \quad P_f \le P(y_m = 0) \\ h(y_m) & \text{if} \quad P_f > P(y_m = 0) \end{matrix}, \qquad (1)$$

where $P(y_m) = 0$ is the probability of precipitation with zero amount and $h(y_m)$ is the distribution of the precipitation amounts with values being larger than zero, both estimated using observed precipitation time series. The term P_f is the forecasted precipitation probability of the corresponding member m.

Determining the two components for the predictive distribution of precipitation consists of two steps: 1) The ensemble mean forecasts and the observations pairs in the training period are first divided into different groups, according to the date in four seasons, the forecast lead time, and the precipitation classes (e.g., ~0–1, 1–2 mm, etc., and >50 mm). For example, for forecasting the 1-lead-day precipitation within the range ~ 0–1 mm in the springtime, the ensemble mean forecasts within the range ~ 0–1 mm, and the corresponding observations are selected in the whole training period. These selected ensemble-mean forecasts and observations are used to calibrate the precipitation model as indicated in Eq. (1). 2) Estimating $P(y_m) = 0$ and $h(y_m)$ is conducted for each group. Specifically, for each group, the probability of precipitation $P(y_m) = 0$ is assumed to be the observed precipitation occurrence, and the predictive distribution of the precipitation amounts $h(y_m)$ is assumed to follow a skewed two-parameter gamma distribution and fitted using the cubic root transformed nonzero observed precipitation amounts.

When obtaining the postprocessed ensemble weather forecasts, the ensemble mean forecasts out of the raw GEFS forecasts are first used to determine the forecasted precipitation class. The probability of precipitation and the predictive distribution of the precipitation amounts are selected according to the forecasted precipitation class. Two sets of standard uniform random numbers with the range of [0, 1] are used to sample the estimated predictive distribution for precipitation. One is used to generate the precipitation occurrence to determine if the corresponding member is a rainy day, and the other is used to sample the predictive distribution for generating the precipitation amount.

(ii) Univariate method for temperature

When using GPP to postprocess the air temperature forecasts, the predictive distribution is assumed to follow a two-parameter normal distribution, as specified by

$$F\left(y_m \middle| x_{1,m}, \ldots, x_{L,m}\right) = N(\mu, \sigma^2), \qquad (2)$$

where mean μ is the postprocessed (i.e., bias corrected) ensemble mean, and variance σ^2 is fixed and estimated by using an iteration approach of Chen et al. (2014a). Correcting the ensemble mean adopts a linear correction equation fitted using the temperature anomalies of both observations and ensemble mean forecasts. The temperature anomalies are obtained by subtracting the long-term daily average of observed temperature from the observed and ensemble mean temperatures, respectively. The training data for fitting the linear equation are selected using a 31-day window centered on the day of interest during the training period. Variance is estimated using an iteration approach that finds the variance value producing the smallest root mean squared error (RMSE) of rank histogram values. For air temperature, only one set of standard uniform random numbers with the range of [0, 1] is used to sample the predictive distribution for generating the discrete forecast ensemble. More details of using GPP can be found in Chen et al. (2014a).

2) Dependence reconstruction methods

Three dependence reconstruction methods are used to introduce the MSMV dependences by rearranging the ensemble before or after using the univariate method.

As mentioned in the univariate method, two set of random numbers are used sample the predictive distribution for precipitation and one set of random numbers is used to sample the predictive distribution for temperature. When combining the dependence reconstruction method, a random number matrix $\mathbf{S}_{N\times M}$ is first generated for sampling from the univariate predictive distribution, where N is the number of the generated ensemble members (can be different from the raw ensemble, it is set to 1000 in this study); M is the number of variables being considered, equaling to the product of the number of weather variables and the number of stations. The dependence reconstruction methods are then used to rearrange each column in $\mathbf{S}_{N\times M}$ for obtaining the matrix $\mathbf{S}^*_{N\times M}$ with the desired dependence structure among the M columns. The corresponding rank matrix of $\mathbf{S}^*_{N\times M}$ is denoted by $[\mathbf{S}]^*_{N\times M}$ and each element in $[\mathbf{S}]^*_{N\times M}$ represents its relative rank order in the column. The three dependence reconstructions are presented as follows.

(i) Rank shuffle

RS, also called Iman shuffle, is a distribution-free method for constructing the desired dependence structure among the variables and sites (Iman and Conover 1982). RS was proved successful in inducing desired rank correlation between precipitation amount and duration (Zhang 2005; Chen et al. 2009) and in reconstructing intersite and intervariable dependencies (Brissette et al. 2007; Li 2013; Chen et al. 2018; Li and Babovic 2019; Guo et al. 2019) in climate downscaling studies.

Using RS to obtain the correlated random number matrix $\mathbf{S}^*_{N\times M}$ consists of the following procedures:

$$\mathbf{V}_{N\times M}\left[n, m\right] = \Phi^{-1}\left(\frac{\mathrm{rank}\{\mathbf{S}_{N\times M}[n, m]\}}{N+1}\right), \qquad (3)$$

$$\mathbf{C}_{M\times M} = \mathbf{R}_{M\times M}\mathbf{R}'_{M\times M}, \qquad (4)$$

$$\mathbf{S}^*_{N\times M} = \mathbf{V}_{N\times M}\mathbf{R}'_{M\times M}. \qquad (5)$$

The van der Waerden score matrix $\mathbf{V}_{N\times M}$ is first estimated, and its element in the nth row and m-th column is denoted by $\mathbf{V}_{N\times M}[n, m]$and calculated using Eq. (3), where rank$\{\mathbf{S}_{N\times M}[n, m]\}$ is the rank order of the corresponding element $S_{N\times M}[n, m]$ along the mth column; Φ^{-1} is the inverse function of the standard normal distribution. The correlation matrix estimated from the observations of the same month in the training period is factorized using Cholesky Factorization for obtaining the lower triangular matrix $\mathbf{R}'_{M\times M}$ using Eq. (4). Finally, multiplying $\mathbf{V}_{N\times M}$ and $\mathbf{R}'_{M\times M}$ gives the desired $\mathbf{S}^*_{N\times M}$. One possible problem is that the correlation matrix estimated from the observations is not always positive definite due to data noises. In this case, the spectral decomposition method proposed by Rebonato and Jäckel (2000) is used to adjust the poorly observed correlation matrix $\mathbf{C}_{M\times M}$.

(ii) Gaussian copula

GC is widely used in atmospheric and hydrological science for representing the multivariable dependence among the variables (Hao and Singh 2016; Li et al. 2017). For the M-dimensional normal distribution $N_M(0_{M\times 1}, \mathbf{C}_{M\times M})$, the correlation matrix $\mathbf{C}_{M\times M}$ is calculated using the selected past observations. The correlated random number matrix $\mathbf{S}^*_{N\times M}$ is generated using $N_M(0_{M\times 1}, \mathbf{C}_{M\times M})$.

(iii) Empirical copula

EC is a nonparametric method for modeling the complex dependencies beyond the linear dependence revealed by the correlation matrix (Bárdossy and Pegram 2009). The copula structure is defined by the independent rank transformation of training data for each of the M-dimensional multivariable data spaces (Wilks 2015).

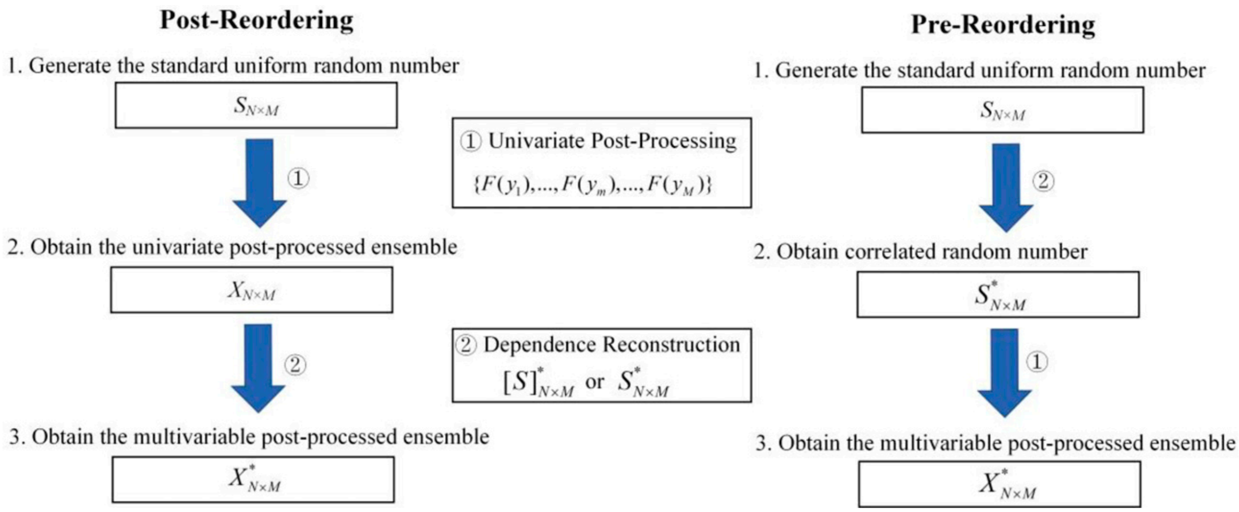

FIG. 2. The schematic diagram for post-reordering and pre-reordering strategy.

In this study, a time window around the forecast date is used for obtaining N samples from the historical observations, forming a matrix of $\mathbf{O}_{N\times M}$. Each column in $\mathbf{S}_{N\times M}$ is rearranged using the column ranks in $\mathbf{O}_{N\times M}$, producing the correlated random number matrix $\mathbf{S}^*_{N\times M}$. Expanding or reducing the selected observations is needed to let the selected observations have the same length as the number of the generated ensemble.

3) REORDERING STRATEGIES

The univariate method is combined with each of the three dependence reconstruction methods by using pre-reordering or post-reordering strategy. Figure 2 shows the schematic diagram of procedures for using the two reordering strategies.

The procedure of using the post-reordering methods for a given day consists of two steps 1) the random numbers are first used to generate the univariate ensemble forecasts $\mathbf{X}_{N\times M}$; and 2) the rank matrix $[\mathbf{S}]^*_{N\times M}$ is then used to rearrange each column for obtaining the final ensemble $\mathbf{X}^*_{N\times M}$.

The procedure of using the pre-reordering methods (RS-Pre, EC-Pre, GC-Pre) for a given day also consists of two steps: 1) two sets of correlated random number matrix $\mathbf{S}^*_{N\times M}$ are first generated by using each of the three dependence reconstruction methods with the first matrix used to represent the precipitation occurrence, and the second matrix used to represent the random numbers for sampling the predictive distribution of precipitation amounts and air temperature values; and 2) using these random numbers in GPP gives the ensemble $\mathbf{X}^*_{N\times M}$. However, when using the pre-reordering strategy, the correlation matrix $\mathbf{C}'_{M\times M}$ calculated from $\mathbf{X}'_{N\times M}$ is generally less correlated than the $\mathbf{C}_{M\times M}$ estimated from the observations (Brissette et al. 2007). Therefore, if the correlation matrix is used in the dependence reconstruction methods, an iterative approach of Brissette et al. (2007) is used to increase each value of the correlation matrix $\mathbf{C}_{M\times M}$ by 0.01 for each iteration. The adjusted correlation matrix $\mathbf{C}_{M\times M}$ is used in the dependence reconstruction methods for obtaining the new random number matrix $\mathbf{S}^*_{N\times M}$. Repeat the above two steps until $\mathbf{C}'_{M\times M}$ approximates $\mathbf{C}_{M\times M}$ within the specified tolerance. The $\mathbf{X}'_{N\times M}$ in the final iteration is the desired ensemble $\mathbf{X}^*_{N\times M}$.

Based on the univariate method GPP, three dependence reconstruction methods and two reordering strategies form six MSMV methods. These methods are denoted by RS-Post, RS-Pre, EC-Post, EC-Pre, GC-Post, and GC-Pre.

b. Verification metrics

A good ensemble forecasting needs to maximize the sharpness of the predictive distribution of forecast variables subject to calibration (Gneiting et al. 2007). Calibration refers to the statistical consistency between the forecasts and the observations, and the sharpness refers to the concentration of ensemble forecasts. The forecasts are sharp if the observations can be interpreted as random draws from the predictive distribution and the forecast uncertainty denoted by the ensemble spread is as small as possible. Besides, scoring rules are also adopted for quantitatively assessing the predictive performance of the ensemble forecasts.

Evaluating the ensemble weather forecasts in this study consists of 1) the performance of the ensemble weather forecasts measured in the single variable at the single site, and 2) the performance of the ensemble weather forecasts measured in multiple variables and sites simultaneously.

For item 1, the chosen univariate metrics include rank histogram and its associated reliability metrics (Δ) to quantify departures from a flat histogram, the deterministic metric: mean absolute error (MAE), and the scoring metric: continuous ranked probability score (CRPS). For producing the rank histogram, when many ensemble members have the exact same value, with zero precipitation for forecasts and zero precipitation for observations, the same frequency is evenly distributed to all identical intervals. For example, if n

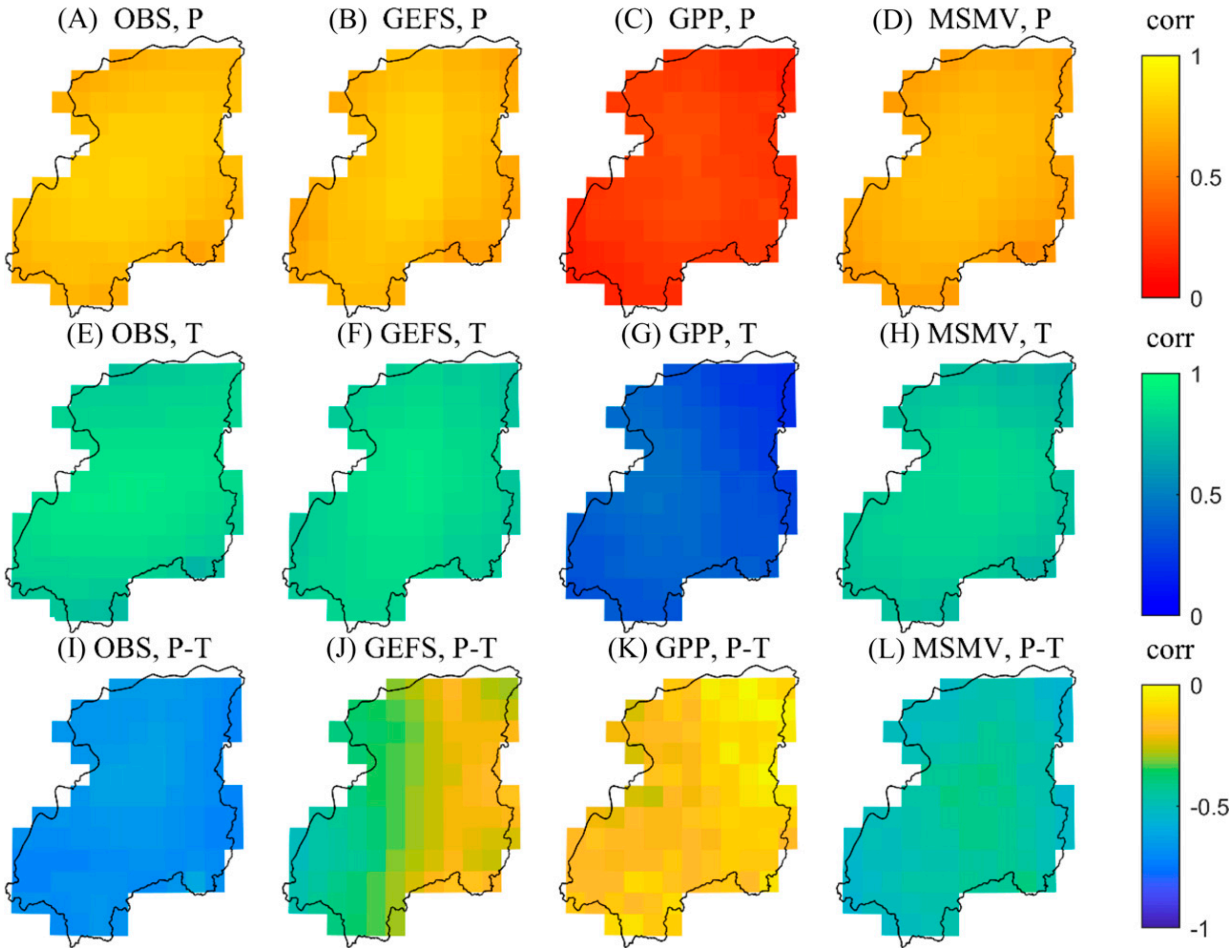

FIG. 3. The mean spatial correlation field of (top) precipitation and (middle) air temperature, and (bottom) the intervariable correlation between precipitation and temperature for the 1-day-ahead ensemble weather forecasts. The mean spatial correlation of any grid is obtained by averaging the inter-site correlation between the selected grid and all other grids. (a),(e),(i) Observations, (b),(f),(j) GEFS forecasts, (c),(g),(k) GPP forecasts, and (d),(h),(l) MSMV (e.g., RS-Post) forecasts. One randomly selected member is used to calculate the correlation matrix and the evaluation is carried out for ensemble forecasts in July.

members predict zeros precipitation for a day, $n + 1$ interval are still divided, but the first n intervals all range from 0 to 0, the identical frequency of $1/n$ is then assigned to all of the first n intervals. This method was also used in many other previous studies (Hamill and Colucci 1997; Hamill 2001). When the ensemble weather forecasts produce a flat rank histogram (Δapproximates to zero), a small MAE value (0 is best), and a lower CRPS value (0 is the best), show the ensemble weather forecasts are well calibrated, de-biased, and skillful. A detailed introduction to the above metrics can be found in Brown et al. (2010).

For item 2, the chosen multivariable metrics include the metric that assesses calibration [band-depth histogram (BDH); Thorarinsdottir et al. 2016], and proper scoring rules [variogram-based score (VS); Scheuerer and Hamill 2015b]. A brief introduction to the above metrics is as follows. The formation of BDH needs to first transform the multivariable properties into a single dimension based on the band depth function (a "pre-rank" function) which assesses the centrality of observations within the forecast ensemble, and then obtain the histograms of the ranks of the observation's "pre-ranks." Using BDH to assess the calibration includes the following cases: overdispersed ensembles give a skewed histogram with too many high ranks; underdispersed or biased ensembles give a skewed histogram with too many low ranks; slightly correlated or highly correlated ensembles give a U-shape or hump-shaped histogram, respectively; and well-calibrated ensembles have a flat histogram. VS calculates the weighted squared variogram difference for the pairwise components in the multivariable quantity. VS is proved to be more sensitive to consider the incorrect dependence structure in the forecast ensemble compared to the existing multivariable scoring metrics, i.e., energy score (Gneiting et al. 2008). When using VS, the power order of the variogram for calculating VS is set to be the recommended value of 0.5 (Scheuerer and Hamill 2015b), and the weighting scheme is based on the equal weight scheme. VS is negatively orientated and 0 is the best.

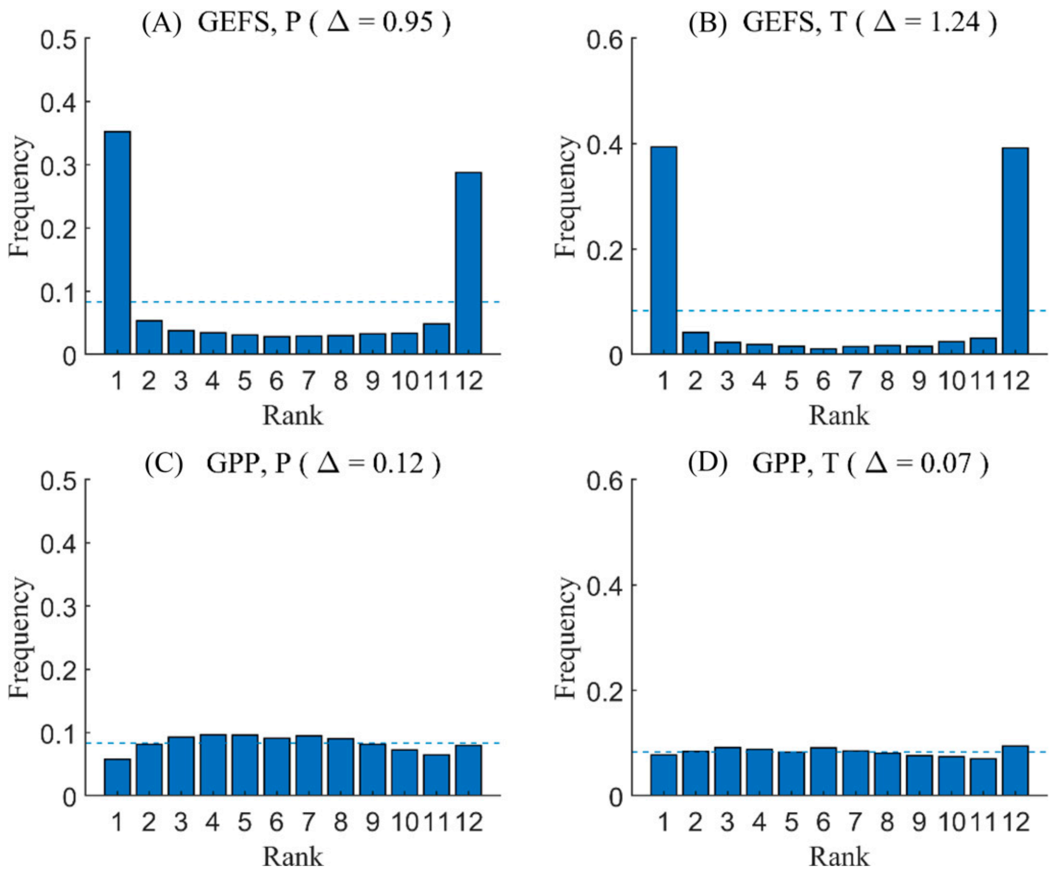

FIG. 4. Calibration checks using rank histogram for 1-day-ahead ensemble forecasts of (left) precipitation and (right) air temperature. For each rank, the rank histogram values are calculated by averaging the corresponding rank histogram values of over 116 grids. (a),(b) GEFS ensemble forecasts (11 members). (c),(d) GPP post-processed ensemble forecasts (11 out of 100 members are randomly selected).

4. Results

a. The correlation performance

Figure 3 illustrates the mean spatial correlation field of precipitation and air temperature, and the intervariable correlation field between precipitation and air temperature. The intersite or intervariable correlation is calculated using the Spearman correlation coefficient. Different color bars with different ranges are used since the value range for the intersite or intervariable correlation is different. It is found that GEFS forecasts (second column) have a similar spatial correlation field as the observations, while the intervariable correlation is biased. Similar to findings in Wilks (2015), it is improper to directly use the multivariable rank structure derived from the raw GEFS forecasts. For the univariate GPP method (third column), the forecasts are generated independently for each grid and variable, as shown by a nearly zero spatial and intervariable correlation field. Thus, the use of the univariate method loses the inherited spatial and intervariable dependencies, and the neglected dependence structure influences the forecast performance of the ensemble forecasts in return (see section 4d). For the MSMV method (the fourth column), the generated ensemble forecast produces the closest mean spatial correlation coefficient field to the observations.

b. The calibration performance

The calibration performances were evaluated using the univariate rank histogram and multivariate band-depth histogram. Figure 4 shows the univariate rank histogram of 1-day-ahead ensemble forecasts of precipitation and air temperature. For GEFS (top row), a large number of the observations are falling into the lowest and highest rank to form a U-shaped rank histogram, indicating that the GEFS forecasts are underdispersive (the forecast uncertainty is highly underestimated). A nearly uniform histogram is obtained by the univariate post-processed ensemble forecasts using GPP (bottom row). This shows that GPP is effective to adjust the calibration in terms of a single variable. The reliability metric value decreases from 0.95 (GEFS) to 0.12 (GPP) for precipitation, and from 1.24 (GEFS) to 0.07 (GPP) for air temperature. It shows that the calibration of air temperature is easier compared to the calibration of precipitation.

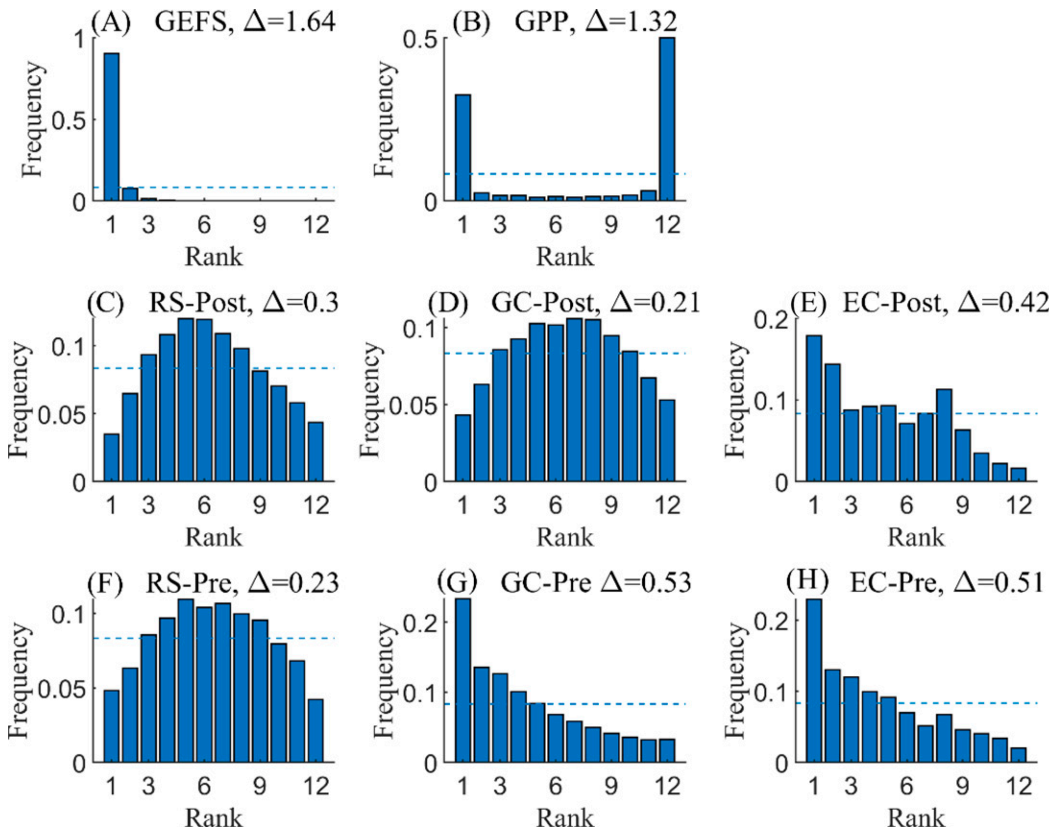

FIG. 5. Band-depth histogram of the 1-day-ahead ensemble weather forecasts for (a) GEFS, (b) GPP, (c)–(e) post-reordering methods, and (f)–(h) the pre-reordering methods. The variables considered to build BDH consist of precipitation and air temperature over 116 grids, a total of 232 variables.

Figure 5 presents the band-depth histogram of the ensemble weather forecasts for GEFS, GPP, and six MSMV methods. Band-depth histogram of GEFS (Fig. 5a) is skewed to the lowest ranks and a majority of observations are falling into the first two ranks, indicating the ensemble weather forecasts are highly underdispersive. The band-depth histogram of GPP (Fig. 5b) is U-shaped, a sign of variables being less correlated. This is reasonable because the use of the univariate method loses the intrinsic correlations among variables and sites. The calibration performance has been substantially improved when using the MSMV methods. For RS-Post (Fig. 5c), GC-Post (Fig. 5d), and RS-Pre (Fig. 5f), the shape of the band-depth histogram is slightly concave and this shows the multivariable calibration is well adjusted for RS-Post, GC-Post, and RS-Pre. While for EC-Post (Fig. 5e), GC-Pre (Fig. 5g), and EC-Pre (Fig. 5h), band-depth histogram is slightly inclined toward the high rank, indicating the ensembles are still underdispersive.

c. The univariate performance

The univariate performances were evaluated using the deterministic metric MAE and the probabilistic metric CRPS. It is noted that for the ensemble generated using GPP, the post-reordering methods only rearrange the member sequence while not altering the member value; therefore, they have the same univariate performance.

Figures 6 and 7 show the MAE of the ensemble forecasts for precipitation and air temperature, respectively. RS-Pre, GPP, and three post-reordering methods have a similar MAE performance, all consistently better than GC-Pre for air temperature but worse than EC-Pre for precipitation. It is found that the MAE performance of GC-Pre is consistently worse than GEFS for all lead days. When looking at the spatial MAE performance for 1-lead-day ensemble forecasts, the following results can be found. 1) All methods share a similar spatial pattern of MAE performance. 2) For precipitation, the MAE in the central and southeastern areas is relatively smaller than the MAE in the rest of the areas. But this relatively better-performing area is shifted to the central and northwestern areas after postprocessing. 3) For air temperature, the relatively worse-performing area is generally distributed in the northern area after postprocessing.

Figures 8 and 9 show the results of CRPS of ensemble forecasts for precipitation and air temperature, respectively. The performances of RS-Pre, GPP, and three post-reordering methods are similar and they all have a consistently better CRPS performance compared to the other methods. Similar to results of MAE, GC-Pre is not recommended, since its performances are only slightly better than GEFS within 5 lead days for precipitation and 6 days for air temperature. The relatively better performance is distributed in the central and

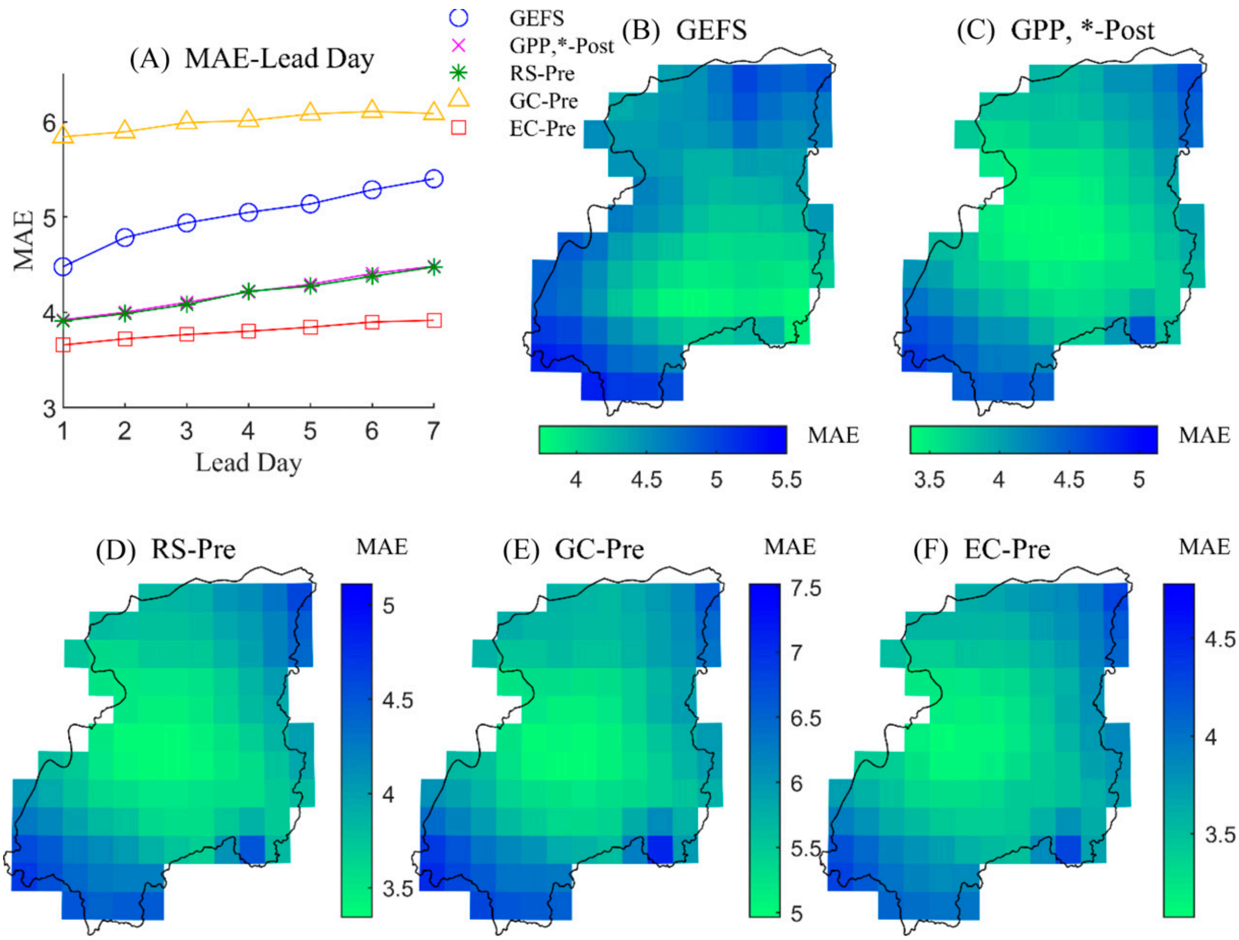

FIG. 6. Univariate evaluation of the ensemble forecasts for precipitation using the deterministic metric MAE. (a) The line plot of the averaged MAE over 116 grids against the lead time and (b)–(f) the spatial distribution of the MAE value for 1 lead day.

northern parts for precipitation. While for air temperature, the northern and southeastern parts exhibit a higher proportion of lower CRPS grids.

d. The multivariable performance

Figures 10, 11, and 12 show the multivariable predictive performance measured in VS for precipitation–air temperature, precipitation, and air temperature, respectively. As expected, the predictive performance of GEFS ensemble forecasts is poor, as shown by a high VS value. The ensemble forecasts using univariate method produce a worse VS performance compared to GEFS after 3 lead days. This highlights that the lost intersite and intervariable dependence strongly influences the predictive performance of the postprocessed forecasts. When precipitation is considered in multivariable postprocessing, the six MSMV methods, except GC-Pre, all outperform GEFS and GPP in terms of VS. When multivariable postprocessing air temperature forecasts, all the multivariable methods outperform GEFS and the univariate GPP method, and the differences between the multivariable methods are not obvious.

5. Discussion and conclusions

Statistical methods are usually used to postprocess ensemble weather forecasts for hydrological predictions. However, most previous methods apply to a single variable at a single station and do not well preserve the intrinsic dependence structure among climate variables over multiple locations as existed in the real climate system. Recently, some MSMV methods with different reordering strategies have been developed. There is a need to synthesize a general framework for combining the univariate methods and the dependence reconstruction methods to generate MSMV ensemble weather forecasts. The framework synthesized in this study is conducive to gather and employ abundant knowledge in univariate postprocessing and dependence reconstruction, which helps to evaluate the existing strategy in MSMV postprocessing and to provide some insight for further method development.

a. The univariate method

The choice of the univariate methods is important for devising the MSMV methods, as they are used to correct the biases

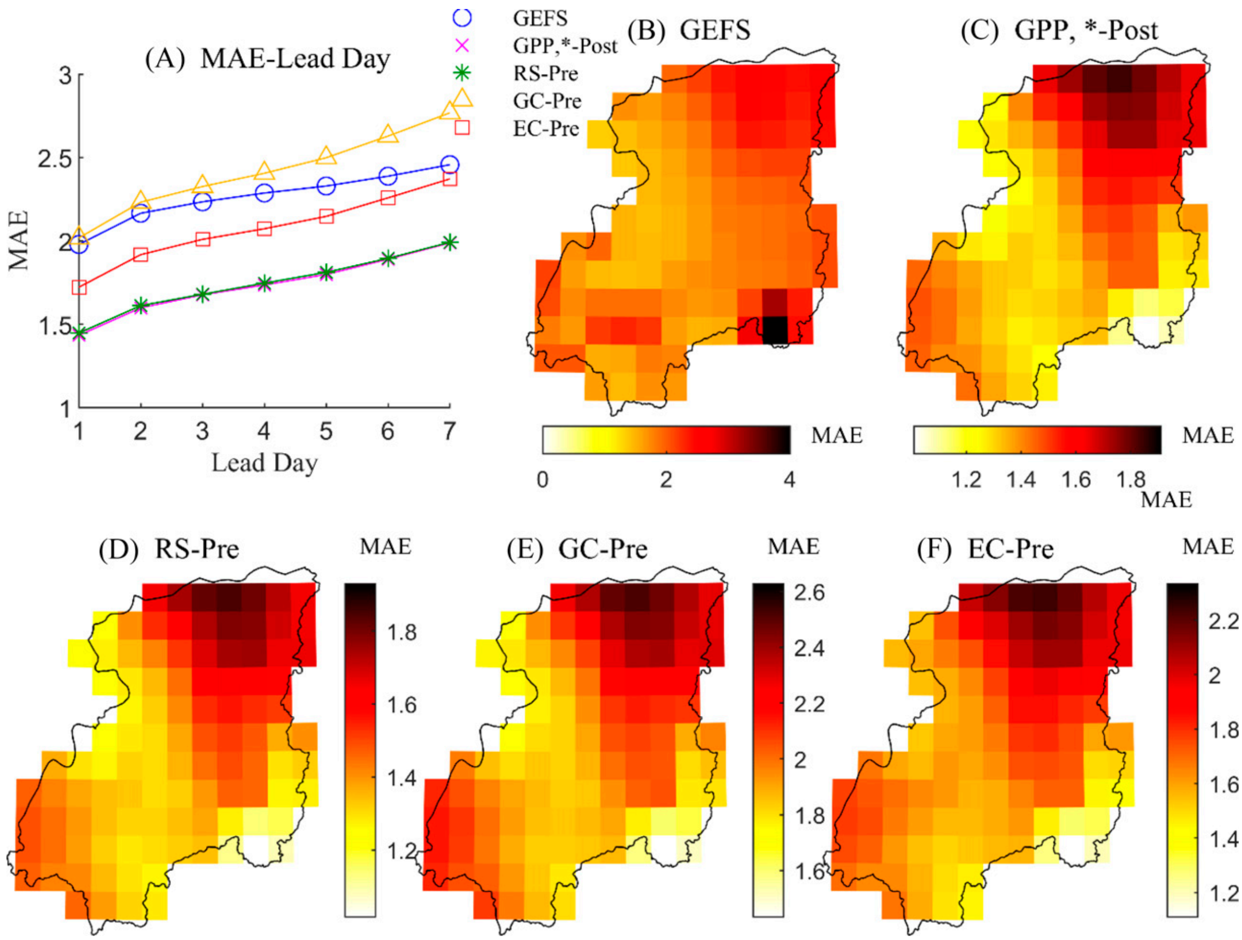

FIG. 7. As in Fig. 6, but for air temperature.

and dispersions existing in the raw ensemble weather forecasts and determine the univariate performances of the proposed MSMV methods. There are a large number of univariate methods available for devising the MSMV methods, including the mostly used BMA and EMOS methods (Möller et al. 2013; Schefzik et al. 2013; Wilks 2015; Schefzik 2017). This study chose the univariate GPP method as an example of evaluating the proposed MSMV framework because GPP has comparable or even competitive performances compared to the existing methods according to previous studies about comparing different univariate methods (Chen et al. 2014a; Chen and Brissette 2015; Li et al. 2019, 2020).

As presented in Figs. 3c,g,k and Fig. 5b, the univariate method introduces errors in the intervariable and intersite dependence field, since the spatial and intervariable dependence is generally ignored. From the view of multivariable evaluation, the "seemingly improved" univariate postprocessing forecasts measured in univariate metrics are even worse than the raw GEFS forecasts. In other words, if not considering the dependence reconstruction, the univariate methods may do more harm than good when used in a physical-based environmental model where the dependence structure among the variables and sites is important.

b. The dependence reconstruction methods

The dependence reconstruction methods are used to amend the lost dependence information among the variables and sites for the univariate postprocessed ensemble weather forecasts. This study compared three widely used dependence reconstruction methods: EC, GC, and RS.

It is found that all three methods are proved effective in reconstructing the intersite and intervariable dependence, even though the concept of each method is different. For GC and RS, the introduced parameter is the correlation matrix estimated from the historical observations. The size of the correlation matrix tends to increase with the number of variables (i.e., the number of climate variables and locations) considered. When calculating the correlation matrix for a large number of variables, the calculated correlation matrix may not be positive definite, due to the excessive noise or outliers in the observed time series. The non-positive-definite correlation matrix would result in a computation problem when using Cholesky factorization. This study used a spectral

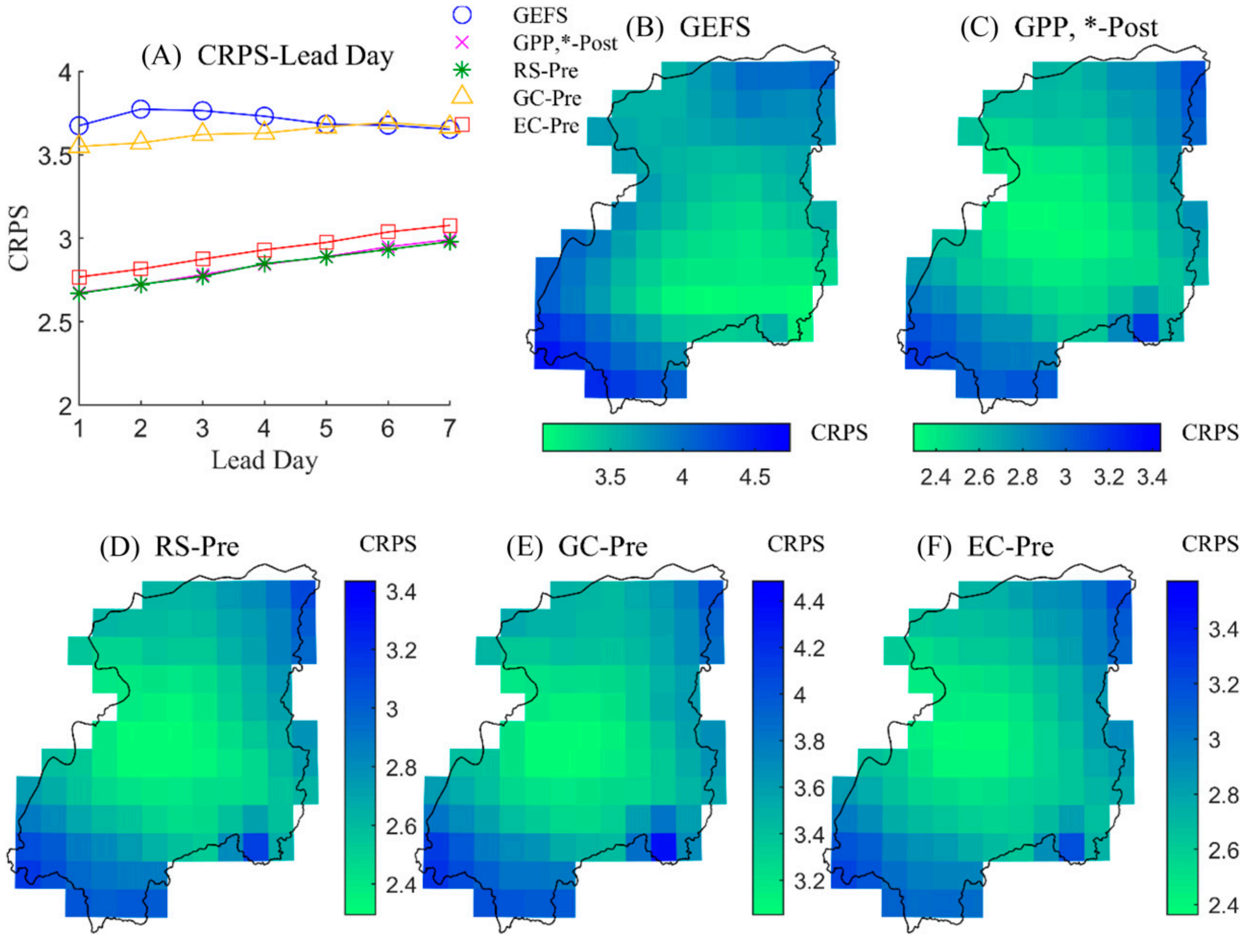

FIG. 8. Univariate checks of the ensemble forecasts for precipitation using CRPS. (a) The line plot of the averaged CRPS over 116 grids against the lead time and (b)–(f) the spatial distribution of the CRPS score for 1 lead day.

decomposition method proposed by Rebonato and Jäckel (2000) to solve this problem. For real applications where the correlation of two distant stations is weak or none, we can also set it as a very small value close to zero to avoid using the improper correlation value due to data error. Besides, a stable estimate of the correlation matrix requires a long observation period. Short time series may generate a biased correlation matrix. EC used in this study is a Schaake shuffle method (Clark et al. 2004), for it uses the multivariable ranks obtained from the historical observations. The nonpositive definition is not a problem when using this method.

c. The reordering strategy

Two reordering strategies was used to combine the univariate method and the dependence reconstruction method. The pre-reordering strategy uses the dependence reconstruction methods in the process of generating the forecast ensemble. One problem for the pre-reordering strategy is that the correlation of the generated forecast ensemble is generally smaller than the correlation of the used correlated random matrix, especially for precipitation. In other words, the correlation of the random matrix must be adjusted higher to achieve the target of the observed precipitation matrix. The iterative scheme proposed by Brissette et al. (2007) was used in this study to overcome this problem. However, the iterative scheme may fail to converge when generating the correlation matrix for precipitation for small watersheds with many stations (Chen et al. 2014b). The study of Chen et al. (2014b) showed that the correlation of precipitation generated using two identical random number series between two nearby stations is still less than that of the observed data for some cases. Moreover, the iterative scheme may not be effective for the GC, as seen by the worse univariate performances compared to the GPP method. This is because GC is sensitive to the adjustment of the correlation matrix. However, these are not problems for the post-reordering strategy, since the introduction of the observed correlation matrix does not affect the marginal distribution of univariate postprocessed variables. Also, the dependence reconstruction method can be used with any univariate method, because the first step is independent of the second one. Thus, the post-reordering strategy can make the best use of the existing univariate postprocessed results. Due to the above reasons, the post-reordering strategy is

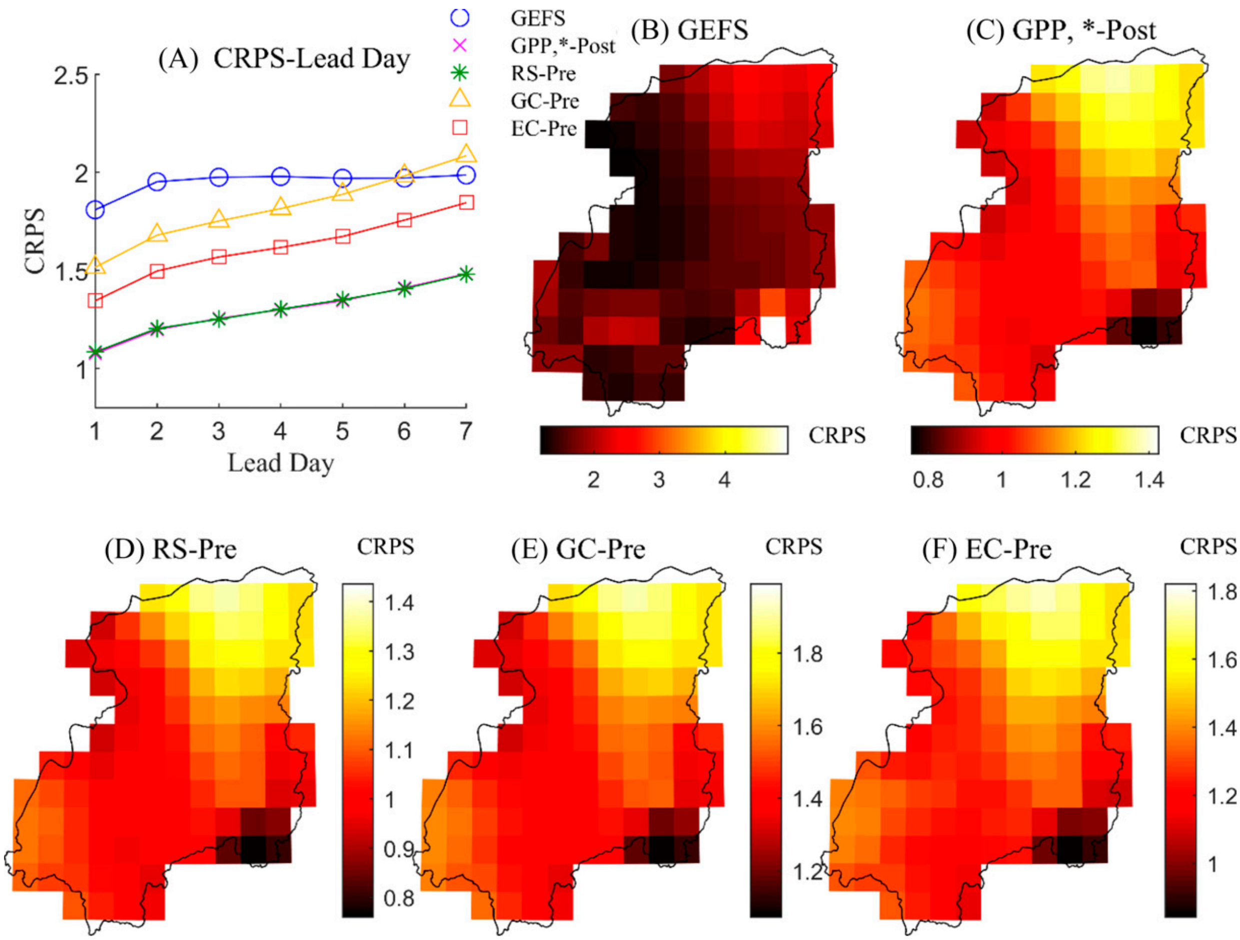

FIG. 9. As in Fig. 8, but for air temperature.

more flexible than the pre-reordering strategy. The results also showed that the post-reordering strategy performs comparably to or even better than the pre-reordering strategy.

Theoretically, the use of either pre-reordering or post-reordering would not affect the performance of MSMV methods, if they are linear processes. This is because the GPP method uses only the raw forecasting ensemble mean for both precipitation and temperature generation, the ensemble reordering by dependence reconstruction methods does not impact the univariate predictive distributions. In addition, the explicit values of random number matrices for precipitation occurrence and amount that sample the univariate predictive distributions are also not impacted by the reordering strategies. Therefore, the difference in pre-reordering and post-reordering strategies appears to only impact the rank of precipitation and temperature.

However, the process from random number fields to MSMV precipitation and temperature ensembles is nonlinear. When using the GPP method to generate a precipitation ensemble, two sets of random numbers are required to sample the estimated predictive distribution with one used to generate the precipitation occurrence and the other is used to sample the predictive distribution for generating the precipitation amount. Thus, when using the pre-reordering strategy, two MSMV correlated random number series have to be generated individually by using one of the three dependence reconstruction methods for precipitation occurrence and wet-member precipitation amounts. However, the post-reordering strategy directly rearranges precipitation values of ensemble members. In other words, when using the pre-reordering strategy, the dependence reconstruction methods are used twice for generating two sets of correlated random numbers. However, when using the post-reordering strategy, the dependence reconstruction methods are used only once for rearranging the univariate method-generated precipitation values.

Moreover, when using the pre-reordering strategy, the same correlation matrix calculated based on observed precipitation time series was used to produce both correlated random number series: one for generating precipitation occurrence and the other for generating precipitation amounts. It might make more sense to use two different correlation matrices: one calculated using observed precipitation occurrence (rain/no-rain) and the other calculated using observed precipitation amounts (excluding dry days). However, for a large watershed, like Xiangjiang River basin in this study, the common wet days are limited, especially for the dry season. More importantly, the observed correlation coefficient matrix was calculated from a

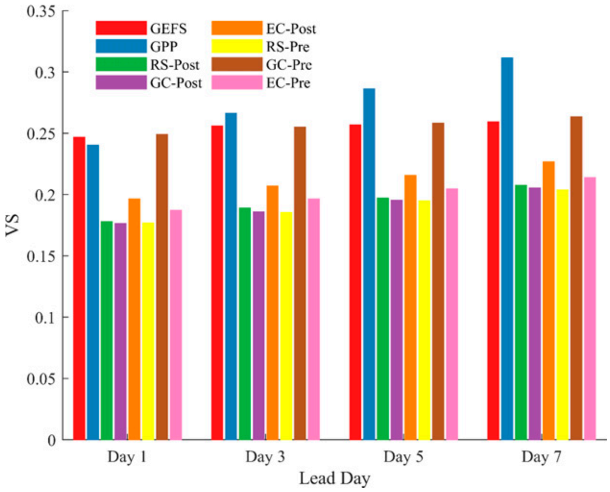

FIG. 10. VS of the ensemble weather forecasts for different methods against lead time. The variables considered calculating VS consist of precipitation and air temperature over 116 grids. The variogram order used in this study is 0.5.

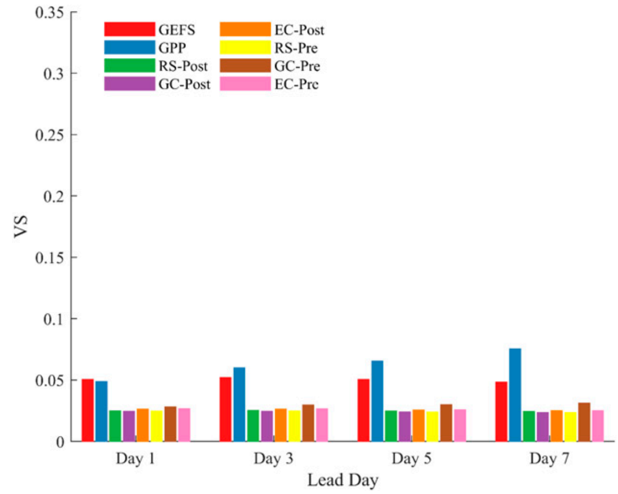

FIG. 12. As in Fig. 10, but for air temperature.

large matrix including both precipitation and temperature time series. When calculating correlation coefficient matrix based on only wet-day precipitation, a large number of temperature values has to be discarded, which might result in biased temperature dependence. From this point of view, the pre-reordering strategy may not be recommended, as it is complicated in concept and implement. Even though other univariate methods can also be used instead of GPP, the generation of precipitation ensemble usually still includes two stages: one for occurrence and the other for amounts. In other words, the use of two sets of random numbers is not unique for GPP, but also for most of other methods, if the discrete precipitation events are required to be generated.

In addition, the iterative scheme used in the pre-reordering strategy may be the other reason leading to different results between pre-reordering and post-reordering. When using the pre-reordering strategy, the generated precipitation is less correlated than the observed values when using the observed correlation matrix to produce the correlated random number field. This has been pointed out in many studies (e.g., Wilks 1998; Brissette et al. 2007; Chen et al. 2014b). Thus, the iterative scheme or other methods to inflate the observed correlation matrix is required, when using the pre-reordering strategy. However, the inflation of observed correlation matrix is not necessary when using the post-reordering strategy, since the post-reordering strategy directly applies the observed correlation matrix to precipitation members, rather than the random number field. The observed MSMV dependence is well preserved.

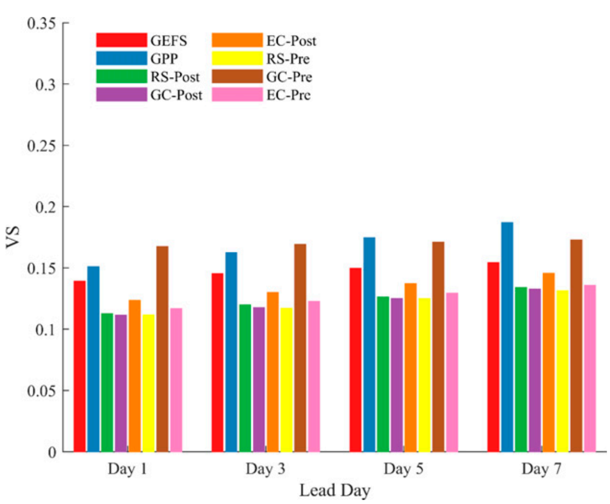

FIG. 11. As in Fig. 10, but for precipitation.

d. Summary

In conclusion, GEFS forecasts are typically biased and highly underdispersed, as concluded in many studies. Besides, it is found that the dependence structure among the variables and sites can also be biased and cannot be directly used in dependence reconstruction. Univariate postprocessing can improve the univariate performance of both ensemble mean and spread, but misrepresent the intersite and intervariable dependence among the forecast variables. The MSMV framework can well utilize the advantages of the univariate method and also reconstruct the intersite and intervariable dependencies. Among the six methods, RS-Pre, RS-Post, GC-Post, and EC-Post perform better than the other two methods in terms of their univariate and multivariable performances. The outperformers include three dependence reconstruction methods with the post-reordering strategy and one dependence reconstruction method with the pre-reordering strategy. Overall, the combination of univariate methods and dependence reconstruction methods by using the post-reordering strategy is recommended in future studies.

Acknowledgments. This work was partially supported by the National Natural Science Foundation of China (Grant 52079093), the Hubei Provincial Natural Science Foundation of China (Grant 2020CFA100), and 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 B18037). The authors would like to acknowledge the National Oceanic and Atmospheric Administration (NOAA), Boulder, Colorado, for providing GEFS ensemble precipitation and air temperature reforecasts. The authors wish to thank the China Meteorological Data Sharing Service System and the Hydrology and Water Resources Bureau of Hunan Province, China for providing the daily meteorological data in the Xiangjiang basin. In addition, the authors declare that they have no conflict of interest.

REFERENCES

Bárdossy, A., and G. G. S. Pegram, 2009: Copula based multisite model for daily precipitation simulation. *Hydrol. Earth Syst. Sci.*, **13**, 2299–2314, https://doi.org/10.5194/hess-13-2299-2009.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, https://doi.org/10.1038/nature14956.

Berrocal, V. J., A. E. Raftery, and T. Gneiting, 2007: Combining spatial statistical and ensemble information in probabilistic weather forecasts. *Mon. Wea. Rev.*, **135**, 1386–1402, https://doi.org/10.1175/MWR3341.1.

——, ——, and ——, 2008: Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. *Ann. Appl. Stat.*, **2**, 1170–1193, https://doi.org/10.1214/08-AOAS203.

Brissette, F. P., M. Khalili, and R. Leconte, 2007: Efficient stochastic generation of multi-site synthetic precipitation data. *J. Hydrol.*, **345**, 121–133, https://doi.org/10.1016/j.jhydrol.2007.06.035.

Bröcker, J., and L. A. Smith, 2008: From ensemble forecasts to predictive distribution functions. *Tellus*, **60A**, 663–678, https://doi.org/10.1111/j.1600-0870.2008.00333.x.

Brown, J. D., J. Demargne, D.-J. Seo, and Y. Liu, 2010: The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. *Environ. Modell. Software*, **25**, 854–872, https://doi.org/10.1016/j.envsoft.2010.01.009.

Chen, J., and F. P. Brissette, 2015: Combining stochastic weather generation and ensemble weather forecasts for short-term streamflow prediction. *Water Resour. Manage.*, **29**, 3329–3342, https://doi.org/10.1007/s11269-015-1001-3.

——, X.-C. Zhang, W.-Z. Liu, and Z. Li, 2009: Evaluating and extending CLIGEN precipitation generation for the loess plateau of China. *J. Amer. Water Resour. Assoc.*, **45**, 378–396, https://doi.org/10.1111/j.1752-1688.2008.00296.x.

——, F. P. Brissette, and Z. Li, 2014a: Postprocessing of ensemble weather forecasts using a stochastic weather generator. *Mon. Wea. Rev.*, **142**, 1106–1124, https://doi.org/10.1175/MWR-D-13-00180.1.

——, ——, and X. C. Zhang, 2014b: A multi-site stochastic weather generator for daily precipitation and temperature. *Trans. ASABE*, **57**, 1375–1391, https://doi.org/10.13031/trans.57.10685.

——, H. Chen, and S. L. Guo, 2018: Multi-site precipitation downscaling using a stochastic weather generator. *Climate Dyn.*, **50**, 1975–1992, https://doi.org/10.1007/s00382-017-3731-9.

Clark, M., S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, 2004: The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *J. Hydrometeor.*, **5**, 243–262, https://doi.org/10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2.

Feldmann, K., M. Scheuerer, and T. L. Thorarinsdottir, 2015: Spatial postprocessing of ensemble forecasts for temperature using nonhomogeneous Gaussian regression. *Mon. Wea. Rev.*, **143**, 955–971, https://doi.org/10.1175/MWR-D-14-00210.1.

Flowerdew, J., 2012: Calibration and combination of medium range ensemble precipitation forecasts. Tech. Rep. 567, United Kingdom Met Office, 21 pp.

Gel, Y., A. E. Raftery, and T. Gneiting, 2004: Calibrated probabilistic mesoscale weather field forecasting. *J. Amer. Stat. Assoc.*, **99**, 575–583, https://doi.org/10.1198/016214504000000872.

Gneiting, T., and A. E. Raftery, 2005: Weather forecasting with ensemble methods. *Science*, **310**, 248–249, https://doi.org/10.1126/science.1115255.

——, ——, A. H. Westveld, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, **133**, 1098–1118, https://doi.org/10.1175/MWR2904.1.

——, F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration and sharpness. *J. Roy. Stat. Soc. Ser. B Stat. Methodol.*, **69**, 243–268, https://doi.org/10.1111/j.1467-9868.2007.00587.x.

——, L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson, 2008: Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *TEST*, **17**, 211–235, https://doi.org/10.1007/s11749-008-0114-x.

Guo, Q., J. Chen, X. Zhang, M. Shen, H. Chen, and S. Guo, 2019: A new two-stage multivariate quantile mapping method for bias correcting climate model outputs. *Climate Dyn.*, **53**, 3603–3623, https://doi.org/10.1007/s00382-019-04729-w.

Hagedorn, R., T. M. Hamill, and J. S. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts Part I: Two-meter temperatures. *Mon. Wea. Rev.*, **136**, 2608–2619, https://doi.org/10.1175/2007MWR2410.1.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.

——, and S. J. Colucci, 1997: Verification of Eta–RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1312–1327, https://doi.org/10.1175/1520-0493(1997)125<1312:VOERSR>2.0.CO;2.

——, R. Hagedorn, and J. S. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts Part II: Precipitation. *Mon. Wea. Rev.*, **136**, 2620–2632, https://doi.org/10.1175/2007MWR2411.1.

Hao, Z., and V. P. Singh, 2016: Review of dependence modeling in hydrology and water resources. *Prog. Phys. Geogr.*, **40**, 549–578, https://doi.org/10.1177/0309133316632460.

Iman, R. L., and W. J. Conover, 1982: A distribution-free approach to inducing rank correlation among input variables. *Commun. Stat. Part B*, **11**, 311–334, https://doi.org/10.1080/03610918208812265.

Jiang, C., L. Xiong, L. Yan, J. Dong, and C.-Y. Xu, 2019: Multivariate hydrologic design methods under nonstationary conditions and application to engineering practice. *Hydrol. Earth Syst. Sci.*, **23**, 1683–1704, https://doi.org/10.5194/hess-23-1683-2019.

Keune, J., C. Ohlwein, and A. Hense, 2014: Multivariable probabilistic analysis and predictability of medium-range ensemble weather forecasts. *Mon. Wea. Rev.*, **142**, 4074–4090, https://doi.org/10.1175/MWR-D-14-00015.1.

Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539, https://doi.org/10.1016/j.jcp.2007.02.014.
Liu, X., and P. Coulibaly, 2011: Downscaling ensemble weather predictions for improved week-2 hydrologic forecasting. *J. Hydrometeor.*, **12**, 1564–1580, https://doi.org/10.1175/2011JHM1366.1.
Li, W., Q. Duan, C. Miao, A. Ye, W. Gong, and Z. Di, 2017: A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. *Wiley Interdiscip. Rev.: Water*, **4**, e1246, https://doi.org/10.1002/wat2.1246.
Li, X., and V. Babovic, 2019: A new scheme for multivariable, multisite weather generator with inter-variable, inter-site dependence and inter-annual variability based on empirical copula approach. *Climate Dyn.*, **52**, 2247–2267, https://doi.org/10.1007/s00382-018-4249-5.
——, J. Chen, C.-Y. Xu, L. Li, and H. Chen, 2019: Performance of post-processed methods in hydrological predictions evaluated by deterministic and probabilistic criteria. *Water Resour. Manage.*, **33**, 3289–3302, https://doi.org/10.1007/s11269-019-02302-y.
——, ——, ——, H. Chen, and S. Guo, 2020: Intercomparison of multiple statistical methods in post-processing ensemble precipitation and temperature forecasts. *Meteor. Appl.*, **27**, e1935, https://doi.org/10.1002/met.1935.
Li, Z., 2013: A new framework for multi-site weather generator: A two-stage model combining a parametric method with a distribution-free shuffle procedure. *Climate Dyn.*, **43**, 657–669, https://doi.org/10.1007/s00382-013-1979-2.
Möller, A., A. Lenkoski, and T. L. Thorarinsdottir, 2013: Multivariable probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Quart. J. Roy. Meteor. Soc.*, **139**, 982–991, https://doi.org/10.1002/qj.2009.
Pinson, P., H. Madsen, H. A. Nielsen, G. Papaefthymiou, and B. Kloeckl, 2009: From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, **12**, 51–62, https://doi.org/10.1002/we.284.
Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174, https://doi.org/10.1175/MWR2906.1.
Rebonato, R., and P. Jäckel, 2000: The most general methodology to create valid correlation matrix for risk management and option pricing purposes. *J. Risk*, **2**, 17–27, https://doi.org/10.21314/JOR.2000.023.
Schefzik, R., 2016: A similarity-based implementation of the Schaake shuffle. *Mon. Wea. Rev.*, **144**, 1909–1921, https://doi.org/10.1175/MWR-D-15-0227.1.
——, 2017: Ensemble calibration with preserved correlations: Unifying and comparing ensemble copula coupling and member-by-member postprocessing. *Quart. J. Roy. Meteor. Soc.*, **143**, 999–1008, https://doi.org/10.1002/qj.2984.
——, T. L. Thorarinsdottir, and T. Gneiting, 2013: Uncertainty quantification in complex simulation models using ensemble Copula coupling. *Stat. Sci.*, **28**, 616–640, https://doi.org/10.1214/13-STS443.
Scheuerer, M., and T. M. Hamill, 2015a: Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions. *Mon. Wea. Rev.*, **143**, 4578–4596, https://doi.org/10.1175/MWR-D-15-0061.1.
——, and ——, 2015b: Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Mon. Wea. Rev.*, **143**, 1321–1334, https://doi.org/10.1175/MWR-D-14-00269.1.
Schmeits, M. J., and K. J. Kok, 2010: A comparison between raw ensemble output, (modified) Bayesian model averaging, and extended logistic regression using ECMWF ensemble precipitation reforecasts. *Mon. Wea. Rev.*, **138**, 4199–4211, https://doi.org/10.1175/2010MWR3285.1.
Schuhen, N., T. L. Thorarinsdottir, and T. Gneiting, 2012: Ensemble model output statistics for wind vectors. *Mon. Wea. Rev.*, **140**, 3204–3219, https://doi.org/10.1175/MWR-D-12-00028.1.
Sloughter, J. M. L., A. E. Raftery, T. Gneiting, and C. Fraley, 2007: Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Wea. Rev.*, **135**, 3209–3220, https://doi.org/10.1175/MWR3441.1.
——, T. Gneiting, and A. E. Raftery, 2013: Probabilistic wind vector forecasting using ensembles and Bayesian model averaging. *Mon. Wea. Rev.*, **141**, 2107–2119, https://doi.org/10.1175/MWR-D-12-00002.1.
Thorarinsdottir, T. L., M. Scheuerer, and C. Heinz, 2016: Assessing the calibration of high-dimensional ensemble forecasts using rank histograms. *J. Comput. Graph. Stat.*, **25**, 105–122, https://doi.org/10.1080/10618600.2014.977447.
Wilks, D. S., 1998: Multisite generalization of a daily stochastic precipitation generation model. *J. Hydrol.*, **210**, 178–191, https://doi.org/10.1016/S0022-1694(98)00186-3.
——, 2006: Comparison of ensemble-MOS methods in the Lorenz '96 setting. *Meteor. Appl.*, **13**, 243–256, https://doi.org/10.1017/S1350482706002192.
——, 2009: Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteor. Appl.*, **16**, 361–368, https://doi.org/10.1002/met.134.
——, 2015: Multivariable ensemble model output statistics using empirical copulas. *Quart. J. Roy. Meteor. Soc.*, **141**, 945–952, https://doi.org/10.1002/qj.2414.
——, and T. M. Hamill, 2007: Comparison of ensemble-MOS methods using GFS reforecasts. *Mon. Wea. Rev.*, **13**, 2379–2390, https://doi.org/10.1175/MWR3402.1.
Xiong, L., K.-X. Yu, and L. Gottschalk, 2014: Estimation of the distribution of annual runoff from climatic variables using copulas. *Water Resour. Res.*, **50**, 7134–7152, https://doi.org/10.1002/2013WR015159.
——, C. Jiang, C.-Y. Xu, K.-X. Yu, and S. Guo, 2015: A framework of change-point detection for multivariate hydrological series. *Water Resour. Res.*, **51**, 8198–8217, https://doi.org/10.1002/2015WR017677.
Xu, H., C.-Y. Xu, H. Chen, Z. Zhang, and L. Li, 2013: Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of China. *J. Hydrol.*, **505**, 1–12, https://doi.org/10.1016/j.jhydrol.2013.09.004.
Zhang, J., J. Chen, X. Li, H. Chen, P. Xie, and W. Li, 2019: Combining postprocessed ensemble weather forecasts and multiple hydrological models for ensemble streamflow predictions. *J. Hydrol. Eng.*, **25**, 04019060, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001871.
Zhang, X.-C., 2005: Generating correlative storm variables for CLIGEN using a distribution-free approach. *Trans. ASAE*, **48**, 567–575, https://doi.org/10.13031/2013.18331.
Zhu, Y., 2005: Ensemble forecast: A new approach to uncertainty and predictability. *Adv. Atmos. Sci.*, **22**, 781–788, https://doi.org/10.1007/BF02918678.
——, Z. Toth, R. Wobus, D. Richardson, and K. Mylne, 2002: The economic value of ensemble-based weather forecasts. *Bull. Amer. Meteor. Soc.*, **83**, 73–83, https://doi.org/10.1175/1520-0477(2002)083<0073:TEVOEB>2.3.CO;2.