Version of Record: https://www.sciencedirect.com/science/article/pii/S0022169417300276
3c9e62c1b3536571560b66c31cebe550

Towards an Improved Ensemble Precipitation Forecast: A Probabilistic Post-Processing Approach

Sepideh Khajehei and Hamid Moradkhani

Abstract

Recently, ensemble post-processing (EPP) has become a commonly used approach for reducing the uncertainty in forcing data and hence hydrologic simulation. The procedure was introduced to build ensemble precipitation forecasts based on the statistical relationship between observations and forecasts. More specifically, the approach relies on a transfer function that is developed based on a bivariate joint distribution between the observations and the simulations in the historical period. The transfer function is used to post-process the forecast. In this study, we propose a Bayesian EPP approach based on copula functions (COP-EPP) to improve the reliability of the precipitation ensemble forecast. Evaluation of the copula-based method is carried out by comparing the performance of the generated ensemble precipitation with the outputs from an existing procedure, i.e. mixed type meta-Gaussian distribution. Monthly precipitation from Climate Forecast System Reanalysis (CFS) and gridded observation from Parameter-Elevation Relationships on Independent Slopes Model (PRISM) have been employed to generate the post-processed ensemble precipitation. Deterministic and probabilistic verification frameworks are utilized in order to evaluate the outputs from the proposed technique. Distribution of seasonal precipitation for the generated ensemble from the copula-based technique is compared to the observation and raw forecasts for three sub-basins located in the

Western United States. Results show that both techniques are successful in producing reliable and unbiased ensemble forecast, however, the COP-EPP demonstrates considerable improvement in the ensemble forecast in both deterministic and probabilistic verification, in particular in characterizing the extreme events in wet seasons.

Post-Processing; Precipitation; Copulas; Climate Forecast System; Hydrologic Forecasting

1. Introduction

Uncertainty in hydrologic simulation and forecast arises from the uncertainties associated with the forcing data, parameters, initial condition, and hydrologic model structure. To achieve accurate hydrologic forecasts, each of these components should be estimated accurately. In the past few years, researchers have proposed various techniques to tackle uncertainty in hydrologic modeling from different angles. For instance, data assimilation is often proposed to deal with the uncertainty in the initial and boundary conditions (DeChant and Moradkhani, 2014; Li et al., 2009; Zehe and Blöschl, 2004). The skill of forecasts can be enhanced by post-processing through multi-modeling (Duan et al., 2007; Madadgar and Moradkhani, 2014a; ; Najafi and Moradkhani, 2015), or other statistical methods, such as quantile mapping or non-parametric procedures (Madadgar et al., 2014; Wood and Schaake, 2008; Zhao et al., 2011; Zhao et al., 2015); Generalized Linear Model Post-Processor (GLMPP) (Ye et al., 2014; Zhao et al., 2011); and the combination of data assimilation and post processing (Bourgin et al., 2014). In addition, characterizing uncertainty in forcing data has received attention in recent years (Clark and Vrugt, 2006; Kavetski et al., 2006; Raleigh et al., 2015; Steinschneider et al., 2012). Moreover, techniques have been developed for generating ensemble meteorological forecasts (Clark et al., 2004; Clark and Hay, 2004; Robertson et al., 2013; Schaake et al., 2007; Tao et al., 2014; Wu et al., 2011; Zhao et al., 2011). Ensemble precipitation forecasts provide the forecasts of the most likely events as well as the uncertainty information. (Park et al., 2008; Tao et al., 2014).

Products from the National Centers for Environmental Prediction (NCEP) models are available in a diverse range of spatial and temporal scales. For instance, Short Range Forecast System (SREF), Global Ensemble Forecast System (GEFS), and Climate Forecast System (CFS) are widely utilized in various studies (Kumar et al., 2012; Peng et al., 2013; Wang et al., 2014). Despite all of these developments in climate models, forecasts are still prone to bias in the mean and insufficient spread (Hamill and Whitaker, 2006; Hamill et al., 2006; Robertson et al., 2013; Tao et al., 2014; Wu et al., 2011). These errors are more significant in climate variables, such as precipitation, which are affected immensely by changes in spatial scale. Thus, it is recommended to not employ the ensemble products from climate models directly, due to three main reasons: 1) there is low accuracy in the ensemble climate forecasts, 2) The models are developed with various assumptions which may not necessarily hold for every regions, and 3) models are mostly developed for large scale applications, and therefore, even if the variables are free from error at their original scale, they might be biased at the catchment scale (Rayner et al., 2005; Tao et al., 2014; Wu et al., 2011) .

There is a demand for methods that are able to generate reliable ensemble forecasts for hydrologic applications. A promising approach is to apply statistical procedures to generate ensemble forecast from Numerical Weather Prediction (NWP) -generated single-value forecasts. The procedure is based on the bivariate probability distribution between the observation and the single-value precipitation forecast. In the past few years, various methods were applied to meet this objective. Kelly and Krzysztofowicz (1997) developed a bivariate meta-Gaussian distribution function based on a normal quantile transformation of two variables according to the

Gaussian law in the Bayesian Forecasting System (BFS). The method was later used by Krzysztofowicz and Herr (2001) to assess the uncertainty in the precipitation data. Clark and Hay (2004) employed Model Output Statistics (MOS) to downscale the model outputs of Global Forecast System (GFS), a medium range forecast system, developed in the National Weather Service (NWS) cooperative network. To preserve and represent space-time variability of climate variables, Clark et al. (2004) introduced a procedure, the so called Schaake Shuffle reconstructing the ensemble members according to the historical values. Schaake et al. (2007) described a full procedure, which is used at the National Weather Service River Forecasting System (NWSRFS) for developing the ensemble meteorological forecasts as the input to the ESP. It is based on the mixed distribution of the two variables and applying the normal quantile transformation for building the joint distribution between the two non-normal variables. In addition to the above mentioned methods, Linear Regression methods have been employed in various studies to post-process precipitation at different temporal and spatial scales (Roulin & Vannitsem, 2012; Sumner, Homar, & Ramis, 2001; Wilks, 2009).

Recently, Wu et al. (2011) developed the Mixed type meta-Gaussian joint distribution built upon the method of Kelly and Krzysztofowicz (1997), which models the precipitation intermittency decisively in comparison to the Schaake et al. (2007) method, which models each of the marginal distributions as a convex combination of the continuous distributions. Robertson et al. (2013) used the Bayesian Joint Probability approach (BJP) developed by Wang and Robertson (2011), and Wang et al. (2009) for generating the ensemble precipitation forecast for a sub-daily weather forecast in Australia. The number of parameters is one of the challenges in EPP; therefore, one of the merits of BJP is that it uses a lesser number of parameters. Tao et al. (2014) employed EPP in combination with multi-modeling to generate a more reliable forecast from The Observing

System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE) products.

According to the above methods, it is assumed that the joint probability distribution between the observations and the forecasts is following a multivariate normal distribution. Furthermore, using normal joint distribution leads to the necessity for a transformation of the non-Gaussian variable into the normal space. The transformation may affect the accuracy of the estimated probability distribution (Brown and Seo, 2013; Madadgar and Moradkhani, 2014a). Therefore, Brown and Seo (2013) introduced a nonparametric probability distribution. Also, Madadgar et al. (2014) demonstrated that the nonparametric probability distributions are highly dependent on the number of thresholds for observed variables. Consequently, it is beneficial to develop a procedure that can capture the uncertainty in a way that there would be no need for variable transformation, and simultaneously adopt a true probability distribution between observation and forecast. Sklar (1959) introduced the concept of constructing multivariate distributions using copulas. Copula functions are useful in capturing the dependency in most of the multivariate distributions such as, bivariate Pareto and bivariate gamma. Copula functions have the capability to draw the joint distribution regardless of the marginal distribution (Favre et al., 2004; Joe, 1997). Copula functions have been used substantially in hydrological applications including precipitation estimation and drought forecasting (Bárdossy, 2006; Bárdossy and Pegram, 2013; Bárdossy and Pegram, 2014; Dung et al., 2015; Favre et al., 2004; Madadgar and Moradkhani, 2013; Madadgar and Moradkhani, 2014b; Salvadori and De Michele, 2010). Madadgar et al. (2014) proposed copula functions to construct the joint distribution between two variables with any level of dependency because they have the potential to be applicable in post-processing the hydrological simulation based on the observation and modeled streamflow simulations. They showed that the copula-based method could provide a more accurate and skillful forecast in comparison to the quantile mapping approach, a widely-used post-processing method.

In this study, we evaluate the capability of copula functions to build the joint distribution between the observation and climatological forecast. This paper is organized in five sections. The introduction is followed by methodology in Section 2, which includes a description of the two techniques that are used in the study, as well as the forecast verification metrics that are employed to evaluate the generated forecasts. Section 3 describes the data and study area. Results and discussion are elaborated in Section 4, and finally, the summary and conclusion are provided in Section 5.

2. Methodology

The main assumptions in post-processing the forecasts are that the observation and forecast are correlated, and that future behavior of the system will remain the same. The purpose of this study is to integrate copula functions into the EPP framework. Copula functions relax these assumptions, and are able to build ensemble forecasts based on historical observations and climatological forecasts. To assess the robustness and reliability of this new technique, results are compared with the outputs from an existing approach called Mixed type Bivariate meta-Gaussian distribution (MBG), introduced by Wu et al. (2011). The theory of copula functions is provided below.

2.1. Data classification and Marginal Distribution

The dataset is divided into twelve monthly classes, and the procedure is applied to each class separately. The procedure starts with fitting different marginal distributions to the observations and forecasts. These distributions include Weibull, Exponential, Lognormal, and Gamma, which are well suited to non-negative data (i.e. precipitation in this case). Furthermore, they are powerful in representing the extreme values of precipitation data. To find the best marginal distribution that can describe the observation and forecast values, Bayesian Information Criterion (BIC) and Kolmogorov-Smirnov (K-S) tests are employed. For a detailed description of these tests, readers are referred to Aho et al. (2014), Raftery (1986), and Stephens (1974).

2.2. Bivariate Meta-Gaussian Distribution (MBG) Approach

Here we discuss generating ensembles through the MBG method (Wu et al., 2011). The goal of this approach is to build the joint cdf of two variables, the observation (O), and the forecast (Y), and then perform sampling from this cdf to generate the ensemble members at each time step.

$$F(y, o) = P\ (Y \leq y, O \leq o) \qquad (1)$$

Since the joint cdf is going to be built based on the bivariate meta-Gaussian distribution, probability distribution of observation and forecast requires to be transformed into normal space. Normal Quantile Transform (NQT) is employed to derive W and Z as the replacements of F(O) and F(Y), the cdfs of observation and forecast in the normal space, respectively.

$$Z = Q^{-1}\big(F(Y)\big) \qquad (2)$$

$$W = Q^{-1}\big(F(O)\big) \qquad (3)$$

Where, Q denotes the standard normal distribution function.

2.2.1. Building and Sampling from the Conditional Joint Distribution

In this step, the conditional distribution between the observation and forecast is formed. By replacing the y and o by z and w respectively, it is assumed that the joint distribution between observation and forecast would be equal to the bivariate normal joint distribution between z and w ($B(z, w; \rho)$):

$$H(y, o; \rho) \equiv B(z, w; \rho) \qquad (4)$$

where H (y, o; ρ) is the bivariate meta-Gaussian distribution of Y and O, as introduced by Kelly and Krzysztofowicz (1997), and ρ is the Pearson product-moment correlation coefficient between Z and W. It can be assumed that:

$$F(y, o) \approx H(y, o; \rho) \qquad (5)$$

According to the meta- Gaussian distribution of O and Y, the conditional distribution of O given Y=y can be written as follows:

$$H(O|Y = y) = Q\left(\frac{w - \rho z}{(1 - \rho^2)^{\frac{1}{2}}}\right) \qquad (6)$$

To create the ensemble forecast at each time step, Equation (6) is being solved employing stratified sampling of the observation given forecast, and therefore we have $p = H\left(O_{samples}\middle|y\right)$. The ensemble members will be generated by following the Equation (7) at each p-probability.

$$Member_p = F_O^{-1}(Q\left(\rho z + (1 - \rho^2)^{\frac{1}{2}} Q^{-1}(p)\right)) \qquad (7)$$

2.2.2. The Schaake Shuffle

To represent temporal dependence, the Schaake Shuffle is used to shuffle the ensemble members at each time step according to the historical observation. In this technique, the ensemble members are being ranked and matched with the historical data for the same month in the past. Then, the members will be reordered to follow the same order as the one for the historical data (Clark et al., 2004).

2.3. Ensemble Post-Processing Based on Copula Functions (COP-EPP)

2.3.1. The Theory of Copula Functions

Copula functions were introduced by Sklar (1959) as functions in the unit cube, which can link multi-dimensional distributions to their one-dimensional marginals. Mathematically, the n-dimensional copula C is represented as: $C: [0,1]^n \rightarrow [0,1]$.

In the bivariate case, a copula satisfies the following property:

$$C(u_1, u_2) + C(v_1, v_2) - C(u_1, v_2) - C(v_1, u_2) \geq 0 \; if \; u_1 \geq v_1 \; and \; u_2 \geq v_2 \tag{8}$$

Where u and v are the marginal distributions of two random variables. In n-dimensional space, the original cumulative distribution can be written as:

$$F(x_1, x_2, \dots, x_n) = C\left[F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)\right] = C(u_1, u_2, \dots, u_n) \tag{9}$$

$$C(u_1, u_2, \dots, u_n) = Pr\{U_1 \leq u_1, \dots, U_2 \leq u_2\} \tag{10}$$

Based on the derivative of cumulative density function (cdf) of the copula, the probability distribution function (pdf) of copula is obtained:

$$c(u_1, \dots., u_n) = \frac{\boldsymbol{\partial}^{\boldsymbol{n}} C(u_1, \dots,, u_n)}{\boldsymbol{\partial} u_1 \dots \boldsymbol{\partial} u_1} \tag{11}$$

The joint density function can then be written as follows:

$$f(x_1, \dots, x_n) = c(u_1, \dots., u_n) \prod_{i=1}^{n} f_{X_i}(x_i) \tag{12}$$

There are several types of copula functions (Nelsen, 1999). Two of these families are applicable in hydrology, Archimedean and Elliptical (Nelsen, 1999; Sklar, 1959). In this study, five copula functions have been used: Gaussian and “t” from Elliptical, and Frank, Clayton, and Gumbel from Archimedean categories.

To find the best copula function that describes the relationship between observations and forecasts, we use Goodness of Fit (GoF) test statistics based on the distance between the empirical and parametric copula described by Genest et al. (2009). Cramer-von Mises and Kolmogorov-Smirnov statistics were used for measuring the distance (Anderson, 1962):

$$S_n = \int_u \Delta C_n\,(n)^2 dC_n(u) \qquad (13)$$

$$\Delta C_n = \sqrt{n}\left(C_n - C_{\theta_n}\right) \qquad (14)$$

Where C_n is the empirical copula with sample size of n, and $C_{\theta n}$ is the theoretical copula estimated for a sample size of n. The null hypothesis assumes that the theoretical copula fits the data. Therefore, the copula function with the greatest p-value (p>0.05) and smallest S_n is desired. To calculate the p-value, parametric bootstrap procedure with 1000 iterations and α=0.01 is used. In this manner, the p-value is calculated as number of times that $S_{n\text{-bootstrap}}$ is greater than S_n divided by the number of iterations. In this study, we have used Inference Functions for Margins (IFM) to estimate the parameters of copula functions (Dupuis, 2007; Joe, 1997).

2.3.2. Ensemble Construction Based on Copula Functions

Ensemble members are generated by sampling from the conditional pdf of the observation given forecast at each time step. In a Bayesian network, the joint distribution of observation and forecast can be simplified as follows:

$$f(y,o) = f(y).f(o|y) \qquad (15)$$

From Equation (15), the conditional probabilities can be derived as:

$$f(o|y) = \frac{f(o,y)}{f(y)} \qquad (16)$$

By using copula functions in estimating the joint pdf (Equation (12)), the conditional pdf (Equation (16)) can be decomposed to a simpler form as:

$$f(s_t|y_t) = \frac{c(U_s = u_s, U_Y = u_y)f(y_t)f(s_t)}{f(y_t)} = c(U_s = u_s, U_Y = u_y)f(s_t) \qquad (17a)$$

where $f(s_t|y_t)$ is the conditional pdf of the sample and forecast at time t, $f(y_t)$ and $f(s_t)$ are the marginal pdf of the forecast, and the sample from the observation at time t. $c(.,.)$ is the copula pdf (Madadgar et al., 2014).

The conditional pdf ($f(S|y_t)$) is built by Monte Carlo sampling from the copula density function ($c(U_s, U_Y = u_y)$), in the following steps:1) The forecast is fixed at time t ($U_Y = u_y$ at time t), 2) Generate U_s by sampling from the historical observation (500 samples in this study). 3) compute the value of $c(U_s, U_Y = u_y)$. Now Equation (17a) is modified as follows:

$$f(S|y_t) = c(U_s, U_Y = u_y)f(s) \qquad (17b)$$

Following Schaake et al. (2007), to sample from the conditional pdf and to generate the ensemble of size *n* (in this study, 20) the probabilities at equal intervals (0.05, 0.1, 0.15, …, 0.95) are considered. Then, the ensemble members associated with above probabilities are obtained. To preserve the temporal variability, the Schaake Shuffle is used to shuffle the ensemble members at each time step according to the historical observation. Figure 1 demonstrates the schematic of COP-EPP procedure.

Figure 1. Schematic of the Copula-based ensemble post-processing (COP-EPP). The ensemble members are generated by sampling from conditional pdf and reconstructed according to Schaake shuffle.

2.4. Forecast Verification

To examine the robustness, reliability, and effectiveness of the proposed approach, results of both approaches, i.e. MBG and COP-EPP are compared through deterministic and probabilistic verification metrics. Accordingly, one deterministic and two probabilistic measures are chosen to evaluate the two generated ensembles. To have a clear comparison between results, normalized measures are employed. For sufficient examination of results, the generated ensembles are analyzed during both the calibration and verification periods.

2.4.1. Deterministic Measures

To inspect the errors in the mean ensemble precipitation forecast, deterministic measures are used. Also, it would be beneficial to study the relationship between the observation and raw forecast in a deterministic framework.

Absolute Percent bias evaluates the deviation of the ensemble mean from the observation. The optimal value of Absolute percent bias is zero (Gupta, Sorooshian, & Yapo, 1999; Moriasi et al., 2007).

$$\text{Absolute } Percent\ Bias = \frac{\sum_{i=1}^{n}|y_i - o_i|}{\sum_{i=1}^{n}(o_i)} \times 100 \qquad (18)$$

To assess the accuracy of the forecast versus observation, the Kling-Gupta Efficiency (KGE) measure is utilized. KGE was introduced as the modified version of Nash-Sutcliffe Efficiency (NSE) by Gupta et al. (2009). This measure captures the correlation, bias, and variability in the forecast data versus the observation.

$$KGE = 1 - ED \qquad (19)$$

$$ED = \sqrt{(r-1)^2 - (\alpha-1)^2 - (\beta-1)^2} \qquad (120)$$

ED is defined as the Euclidean distance between two variables, r represents the linear correlation coefficient between the observation and forecast, α is the ratio of variance of forecast to variance in observation (relative variability in the forecast and observation), and β represents the ratio bias. ED is always non-negative, and thus KGE will have a value from -∞ to 1, with optimal value of 1.

2.4.2. Probabilistic Measures

Since the deterministic measures are affected by over or under -confident forecasts, it is essential to examine the generated forecast through probabilistic measures as well. Probabilistic measures are beneficial in assessing the reliability of ensemble forecast (DeChant and Moradkhani, 2014b; DeChant and Moradkhani, 2015). To assess the forecast skill of the generated ensembles, Continuous Ranked Probability Skill Score (CRPSS) is employed. This measure is the normalized version of Continuous Ranked Probability Score (CRPS) that has been introduced as the extension of Brier Score over all the possible thresholds (Hersbach, 2000).

$$CRPS = \int_{x=-\infty}^{x=\infty} \left(F^{y}(x) - F_{o}\ (x)\right)^{2} dx \qquad (21)$$

and

$$F_{O}\,(x) = \begin{cases} 0, x < observed\ value \\ 1, x \geq observed\ value \end{cases} \qquad (22)$$

Where $F^{fo}(x)$ is the forecast probability, i.e. cdf of the forecast . In the deterministic case CRPS will be the same as MAE with the optimal value of zero.

$$CRPSS = \frac{\overline{CRPS_{ref}} - \overline{CRPS_{forecast}}}{\overline{CRPS_{ref}}} \qquad (23)$$

$CRPS_{ref}$ refers to the Continues Ranked Probability Score for the observation. The range for CRPSS is -∞ to one with optimal value of one.

$$CRPS_{ref} = \frac{1}{T}\sum_{t=1}^{T}\sum\left|o\ (t) - \bar{o}\right| \qquad (24)$$

One of the techniques used to evaluate the forecast skill is to employ the Relative (or Receiver) Operating Characteristic (ROC) curve in which the hit rate and false- alarm are compared (Mason, 1982). The area under the curve represents the ROC score ranging between 0 and 1 with the perfect score of 1. In this study, the ROC score is used to examine the resolution of the generated forecast for winter precipitation (December, January, and February) during the verification period (2000-2014).

The ROC score is not sensitive to bias in the forecast (Wilks, 2011); therefore, it can be considered as a measure for assessing the potential usefulness of a certain forecast. On the contrary, bias can affect the reliability of forecast (Murphy, 1993). Accordingly, the reliability of winter precipitation is evaluated based on the decomposition of the Brier score (Murphy, 1993) and calculating the reliability at threshold of a 95 percentile of the observation as follows:

$$\mathrm{Re}\,liability = \frac{1}{T}\sum_{b=1}^{B} n_b\left(\overline{p_b} - \overline{O_b}\right)^2 \qquad (25)$$

To calculate the reliability, it is important to group the forecasts into *B* forecast bins. Each bin has a population of n_b, with an average forecast probability of $\overline{p_b}$ and an observed frequency of $\overline{O_b}$. For more details about reliability, readers are encouraged to refer to (Wilks, 2011).

3. Data and Study Area

The applicability and usefulness of the methods are evaluated over three basins located in the Western United States, as shown in Figure 2. The characteristics of each of these three basins are summarized in Table 1.

Figure 2. The location of three study basins in the Western USA.

Table 1. Summary characteristics of the study basins.

This study is conducted on monthly precipitation over the historical period (1979-2014). The first twenty-one years are used to calibrate the model, followed by fifteen years of validation period. The observation dataset is extracted from the Parameter-Elevation Relationships on Independent Slopes Model (PRISM) Climate Group, http://prism.oregonstate.edu. The observed precipitation includes rainfall and melted snow over 4km grid cells.

Climate Forecast System Reanalysis (CFSR) and Climate Forecast System version 2 (CFSv2) developed by the National Centers for Environmental Prediction (NCEP) are utilized as the climatologic forecast (Saha et al., 2010; Saha et al., 2014). The six-hourly precipitation at 0.5° spatial resolution is chosen and accumulated to the monthly values for evaluation of forecast.

To prepare the datasets for this study, the observation dataset is re-gridded to $0.5^{\circ} \times 0.5^{\circ}$ resolution.

4. Results and Discussion

4.1. Raw Forecast Validation

We begin by evaluating the relationship between the raw forecast and the observation. In order to achieve this objective, KGE is calculated for the raw forecast in relation to the observation for calibration period. Results are shown in Figure 3.

Figure 3. Kling-Gupta efficiency measure for the raw forecast during calibration period. White grid cells in Rouge River Basin present missing observations.

The acceptable range for KGE is considered to be above 0.6; therefore, results in Figure 3 show that the raw forecast does not have good skill as compared with the observation. More specifically, results show that the KGE is less than 0.6 for the major portion of all three basins, indicating that the observation mean is better in describing the precipitation event in comparison to the raw forecast. The above inspections lead to the conclusion that it is not appropriate to use the raw forecast.

4.2. Selection of Copula Functions

The experiment starts with finding the best copula function that describes the joint distribution between the observation and the forecast during the calibration period. As an example, Figure 4 presents copula functions that successfully map the joint distribution for each month in each grid cell in the upper Colorado River Basin. The figure demonstrates that among the copula functions used in this basin, the Frank copula provides the best fit in describing the joint distribution between the observation and the raw forecast in most months. Further inspection reveals that this pattern is not followed for some of the months. As seen in Figure 4, the Clayton and T Copulas are the most chosen distributions in March. In June, Gumbel and T-Copula are the dominant copula functions chosen.

Figure 4. Selection of suitable copula functions for each grid cell across the Upper Colorado River Basin.

According to Figure 4, copula functions will be fitted, and ensemble members will be built from the conditional pdf of sample variable given the raw forecast at each time step. Here, the ensemble precipitation forecast with twenty members will be generated.

4.3. Deterministic Verification

The Absolute percent bias calculated for the study period (1979-2014) is presented in Figure 5. This Figure demonstrates the calculated Absolute percent bias distribution over the study basins for raw forecast and the bias-corrected ensemble means (i.e. MBG and COP-EPP). The highest bias for the raw forecast is found in the Upper Colorado basin where Absolute percent bias ranges from 50% to 300%. Both techniques improve the accuracy of the raw forecast by more than 50%. The improvement is more considerable in Rogue basin, where the average of the Absolute percent bias over the basin is reduced from 150% to 80% and 60% for MBG and COP-EPP methods, respectively. COP-EPP method shows to be more effective in reducing the bias as compared to MBG method. Overall, there is a 20% difference between the two techniques across the study basins.

Figure 5. Absolute percent bias distribution over the three study basins. The bias is calculated for the raw forecast (blue line), and the ensemble mean from the MGB and COP-EPP outputs (red and green lines, respectively).

Results are evaluated with the KGE metric, which is a deterministic measure. A deterministic comparison has been done on the ensemble means of both methods in calibration and validation periods.

Figure 6 shows the KGE results in the calibration and the validation periods for each study basin. Figures 3 and 6 show a significant improvement in generated forecast ensemble as compared to the raw precipitation forecast. In addition, COP-EPP shows about 10% higher KGE values than that of MBG during the calibration period. This is more significant for the validation period where the range of calculated KGE for COP-EPP is always larger than that of MBG. Furthermore, the discrepancy between the validation and the calibration period is less significant in the COP-EPP method than in the MBG method. For instance, in the Rogue River Basin, the calculated KGE for the MBG method shows a decrease of 20% when compared with its calibration results; while this is not the case for COP-EPP. In the Upper Colorado Basin, the MBG method results in KGE values are less than 0.6 in some grids, which may not be considered accurate enough in some specific studies (Bisselink, Zambrano-Bigiarini, Burek, & de Roo, 2016; Thiemig, Rojas, Zambrano-Bigiarini, & De Roo, 2013).

Figure 6. KGE measure calculated over the study basins after post-processing using the COP-EPP and MBG methods during calibration (left) and verification (right) periods.

4.4. Probabilistic Verification

Figure 7 shows the results for the probabilistic verification of the ensemble forecast generated by each method in each basin. This figure presents the CRPSS analysis for both the calibration and the validation periods. The results of the CRPSS analysis demonstrate a significant improvement of the raw forecast after EPP in both techniques, with a more significant improvement based on COP-EPP. During the calibration period, the copula method shows superior results by almost 10% in comparison to the MBG's results in all basins, whereas in Rogue River Basin, disparity in the MBG results is slightly higher, indicating that in basins with higher precipitation, COP-EPP is more successful than MBG in generating more skillful ensemble forecasts.

Figure 7 indicates that both methods result in higher values of CRPSS in the western parts of Upper Colorado River Basin. For central parts of Southern Snake Basin, CRPSS shows the lowest values for the generated ensemble precipitation compared to the observation. Significantly, the spatial pattern is maintained in the validation period for both techniques in all the basins.

Comparing the results for both techniques in the validation period indicates that copula outputs are more successful in preserving the accuracy of the generated ensemble in this period. For instance, in the case of Rogue River Basin, CRPSS is reduced by almost 2% in the validation period compared to the calibration for COP-EPP outputs. Meanwhile, this difference is about 10% for MBG results. This schema is visible for the other two basins with less noticeable difference.

Figure 7. CRPSS measure calculated for 3 basins after implementing two post-processing methods for the calibration (left) and verification (right) periods.

The ROC scores are presented in Figure 8. Both post-processing methods maintain a ROC score above 0.5, which is an encouraging outcome for the generated ensemble forecast. Comparing the results from the Southern Snake and Upper Colorado basins shows that COP-EPP generates the ensemble forecast with higher resolution than that of the MBG method. The difference between the two methods is more noticeable in Rogue River Basin, which receives relatively higher precipitation than the other basins.

Figure 8. Assessment of forecast resolution based on ROC score for winter (Dec, Jan, and Feb) precipitation during verification period (2001-2014).

Since the focus of this study is to minimize the bias in the forecast, in addition to resolution, reliability of the generated forecast is evaluated for each basin. According to the formulation of reliability (eq. 21), the lower the score the more reliable the ensemble forecast will be. Figure 9 illustrates that COP-EPP has a reliability score of less than 0.05 for approximately all the grid cells in the three basins. Although, in general, the MBG method, provides a reliable ensemble forecast across the basins, the COP-EPP shows noticeably better performance than the MBG . Since the reliability of the forecasts is examined for the ninety-fifth percentile of the observation, this can also indicate the reliability of the generated forecasts on the extremes. Although both methods are shown capable of detecting the extremes, the COP-EPP demonstrates more accuracy on extreme values.

Figure 9. Reliability measure for winter precipitation (Dec, Jan, and Feb) calculated at 95th percentile of observation during the verification period (2001-2014). This measure ranges from 0 to 1 with the optimal value of 0.

4.5. Seasonal Evaluation of the Generated Ensemble Forecasts

To study the performance of the generated ensemble precipitation based on the COP-EPP, seasonal precipitation is evaluated for each basin. Accordingly, the most probable ensemble member (the mode of the conditional pdf), which can substitute the raw forecast as a better predictor, is chosen for this purpose. For a clear inspection, the spatial average of the most probable ensemble member is utilized over the study basin.

Results of seasonal assessment are summarized in Figure 10. In the figure, the observation is shown by the green line, the raw forecast is shown by the red line, the blue line displays the most probable ensemble member from the generated ensemble forecast by COP-EPP, and the black line represents the indicated ensemble member from MBG's generated forecast. Overall, the post-processed forecast shows a significant improvement in describing seasonal precipitation distribution in comparison to the raw forecast. For an instance, in the spring, the precipitation distribution from the raw forecast (red line) overestimates the observation (green line), whereas the COP-EPP forecast almost follows the same distribution as the observation. In summer, the raw forecast distribution shows better performance compared to other seasons. On the other hand, in a severe condition such as Rogue River Basin, which is a coastal region, the raw forecast is not able to follow the observation’s distribution. This is more noticeable in winter when basins receive more precipitation.

Seasonal results for the Upper Colorado Basin, a semi-arid region, demonstrate no significant difference between precipitations in different seasons whereas in the Southern Snake Basin, spring and winter receive more precipitation than summer and fall. These changes become more significant when a coastal basin, such as the Rogue River Basin, is studied. In inspecting the results for the Upper Colorado Basin, it can be seen that the raw forecast is over-predicting the observation, especially in spring and winter. This pattern is approximately the same for the Southern Snake River Basin. However, in the Rogue River Basin, summers receive much less precipitation compared to other seasons. Observations for this basin show the same distribution pattern for a high level of precipitation in fall and spring. Results for winter in this basin show much higher precipitation with an average of about 600 mm. In comparing the raw forecast with the three other datasets, i.e. observation and two generated forecasts, results indicate that the raw forecast has higher seasonal precipitation values in all seasons with a significant difference in winter. In summer, the forecast is following the same distribution as the observation with minor under-prediction for the Rogue River Basin. Furthermore, in wet seasons, the generated forecasts successfully follow the observation distribution, whereas MBG is showing minor under-prediction. In extreme cases, such as winter in the Rogue River Basin, the raw forecast shows a large bias while COP-EPP was successful in reducing the bias, similar to spring and fall.

Overall, COP-EPP has shown significant potential in generating more reliable and accurate ensemble forecasts. Compared to the benchmark technique, COP-EPP shows more consistency in the validation period as compared to that of MBG. In the cases with extreme precipitation, COP-EPP still shows superior results. Lastly, COP-EPP is robust in conserving the spatial pattern of calculated measures. Generally, results show higher accuracy in wet seasons; a specifically generated ensemble forecast is showing promising results in the extreme cases such as winter for the Rogue River Basin.

Brown and Seo (2013) argue that back and forth transformation from Gaussian space can invalidate the optimality of estimated parameters of the conditional probability distribution. However, the improvement in the COP-EPP compared to the MBG can be attributed to the procedure used by COP-EPP to model the dependence structure between observation and forecast. COP-EPP joins the variables via their marginal distribution. Therefore, the complexity in the hydro-meteorological processes does not constrain modeling the joint behavior of observation and forecast (Dupuis, 2007; Madadgar & Moradkhani, 2014).

Figure 10. Probability density functions of seasonal precipitation from the observation, raw forecast, and most probable forecast based on the COP-EPP and MBG approach for three study basins. Seasonal precipitation is spatially averaged over all grid cells of each basin. Seasons are categorized in the following order: Spring (Mar, Apr, and May), summer (Jun, Jul, and Aug), fall (Sep, Oct, and Nov), and winter (Dec, Jan, and Feb).

5. Conclusion and Summary

The purpose of this study is to examine the accuracy and reliability of ensemble precipitation prediction utilizing copula functions. The technique is based on the relationship between the single value forecast and historical observation. Therefore, the assessment is done by comparing results from the new copula method with that of a well-known procedure, MBG.

Comparison is undertaken by employing three different basins with semi-arid to coastal climates to study the performance of the techniques in different climate regimes. Deterministic verification indicates promising improvement in the mean ensemble using the COP-EPP for generating ensemble precipitation forecast. In order to assess the forecast skill, probabilistic measures including CRPSS, reliability, and the ROC score are employed. The results of CRPSS indicate that the generated ensemble forecast from COP-EPP is more reliable and accurate in comparison to the meta-Gaussian one. Furthermore, through analysis of reliability, it is noticed that the copula- based method is more successful in generating the ensemble forecasts that represent extremes. The ROC score indicated that both techniques are capable of generating potentially useful ensemble forecasts with high resolution; however, in the basin with higher precipitation (i.e., Rogue River Basin), COP-EPP proves to be even more superior.

Overall, both techniques show promising results, and existing procedures generated ensembles with acceptable reliability. However, using copula functions will help improve the quality of ensemble forecasting. COP-EPP is shown to be more precise in building the ensemble precipitation forecast. In other words, results demonstrate that the copula procedure is approximately independent of spatial and temporal changes in the data.

It is worth mentioning that incorporating copula functions into EPP helps overcome the assumption of normal distribution for the observation and forecast. It is therefore possible to eliminate the transformation step in EPP procedure. Moreover, copula functions are capable of building joint distribution between two datasets with any level of dependency, and for any marginal distributions. These characteristics of copula functions help us generate more accurate ensemble forecast.

Acknowledgement

The authors would like to acknowledge the financial support for this study by NOAA Grants No. NA11NWS4680002 and NA140AR4310234.

References

Aho, K., Derryberry, D. and Peterson, T., 2014. Model selection for ecologists: the worldviews of AIC and BIC. Ecology, 95(3): 631-636.

Anderson, T.W., 1962. On the distribution of the two-sample Cramer-von Mises criterion. The Annals of Mathematical Statistics: 1148-1159.

Bisselink, B., Zambrano-Bigiarini, M., Burek, P., de Roo, A., 2016. Assessing the role of uncertain precipitation estimates on the robustness of hydrological model parameters under highly variable climate conditions. J. Hydrol. Reg. Stud. 8, 112–129. doi:http://dx.doi.org/10.1016/j.ejrh.2016.09.003

Bourgin, F., Ramos, M.H., Thirel, G. and Andréassian, V., 2014. Investigating the interactions between data assimilation and post-processing in hydrological ensemble forecasting. Journal of Hydrology, 519, Part D: 2775-2784.

Brown, J.D. and Seo, D.J., 2013. Evaluation of a nonparametric post-processor for bias correction and uncertainty estimation of hydrologic predictions. Hydrological Processes, 27(1): 83-105.

Bárdossy, A., 2006. Copula-based geostatistical models for groundwater quality parameters. Water Resources Research, 42(11).

Bárdossy, A. and Pegram, G., 2013. Interpolation of precipitation under topographic influence at different time scales. Water Resources Research, 49(8): 4545-4565.

Bárdossy, A. and Pegram, G., 2014. Infilling missing precipitation records – A comparison of a new copula-based method with other techniques. Journal of Hydrology, 519, Part A: 1162-1170.

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B. and Wilby, R., 2004. The Schaake shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields. Journal of Hydrometeorology, 5(1): 243-262.

Clark, M.P. and Hay, L.E., 2004. Use of medium-range numerical weather prediction model output to produce forecasts of streamflow. Journal of Hydrometeorology, 5(1): 15-32.

Clark, M.P. and Vrugt, J.A., 2006. Unraveling uncertainties in hydrologic model calibration: Addressing the problem of compensatory parameters. Geophysical Research Letters, 33(6).

DeChant, C.M. and Moradkhani, H., 2014b. Toward a reliable prediction of seasonal forecast uncertainty: Addressing model and initial condition uncertainty with ensemble data assimilatio n and Sequential Bayesian Combination. Journal of Hydrology, 519, Part D(0): 2967-2977.

DeChant, C.M. and Moradkhani, H., 2015. On the assessment of reliability in probabilistic hydrometeorological event forecasting. Water Resources Research.

Duan, Q., Ajami, N.K., Gao, X. and Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. Advances in Water Resources, 30(5): 1371-1386.

Dung, N.V., Merz, B., Bárdossy, A. and Apel, H., 2015. Handling uncertainty in bivariate quantile estimation – An application to flood hazard analysis in the Mekong Delta. Journal of Hydrology, 527: 704-717.

Dupuis, D. J. (2007). Using copulas in hydrology: Benefits, cautions, and issues. *Journal of Hydrologic Engineering*, *12*(4), 381–393.

Favre, A.C., El Adlouni, S., Perreault, L., Thiémonge, N. and Bobée, B., 2004. Multivariate

hydrological frequency analysis using copulas. Water resources research, 40(1).
Genest, C., Rémillard, B. and Beaudoin, D., 2009. Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and economics, 44(2): 199-213.
Gupta, H.V., Kling, H., Yilmaz, K.K. and Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377(1–2): 80-91.
Gupta, H.V., Sorooshian, S., Yapo, P.O., 1999. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. J. Hydrol. Eng. 4, 135–143.
Hamill, T.M. and Whitaker, J.S., 2006. Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. Monthly Weather Review, 134(11): 3209-3229.
Hamill, T.M., Whitaker, J.S. and Mullen, S.L., 2006. Reforecasts: An Important Dataset for Improving Weather Predictions. Bulletin of the American Meteorological Society, 87(1): 33-46.
Hersbach, H., 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. Weather and Forecasting, 15(5): 559-570.
Joe, H., 1997. Multivariate models and multivariate dependence concepts, 73. CRC Press.
Kavetski, D., Kuczera, G. and Franks, S.W., 2006. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. Water Resources Research, 42(3).
Kelly, K.S. and Krzysztofowicz, R., 1997. A bivariate meta-Gaussian density for use in hydrology. Stochastic Hydrology and Hydraulics, 11(1): 17-31.
Krzysztofowicz, R. and Herr, H.D., 2001. Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model. Journal of Hydrology, 249(1): 46-68.
Kumar, A., Chen, M., Zhang, L., Wang, W., Xue, Y., Wen, C., Marx, L. and Huang, B., 2012. An analysis of the nonstationarity in the bias of sea surface temperature forecasts for the NCEP Climate Forecast System (CFS) version 2. Monthly Weather Review, 140(9): 3003-3016.
Li, H., Luo, L., Wood, E.F. and Schaake, J., 2009. The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. Journal of Geophysical Research: Atmospheres (1984–2012), 114(D4).
Madadgar, S. and Moradkhani, H., 2013. A Bayesian Framework for Probabilistic Seasonal Drought Forecasting. Journal of Hydrometeorology, 14(6): 1685-1705.
Madadgar, S. and Moradkhani, H., 2014a. Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging. Water Resources Research: n/a-n/a.
Madadgar, S. and Moradkhani, H., 2014b. Spatio-temporal drought forecasting within Bayesian networks. Journal of Hydrology, 512: 134-146.
Madadgar, S., Moradkhani, H. and Garen, D., 2014. Towards improved post-processing of hydrologic forecast ensembles. Hydrological Processes, 28(1): 104-122.
Mason, I., 1982. A model for assessment of weather forecasts. Aust. Meteor. Mag, 30(4): 291-303.
Murphy, A.H., 1993. What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather and forecasting, 8(2): 281-293.
Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. Asabe 50, 885–900.

Najafi, M.R. and Moradkhani, H., 2015. Ensemble combination of seasonal streamflow forecasts. Journal of Hydrologic Engineering, 21(1): 04015043.

Nelsen, R.B., 1999. An introduction to copulas, volume 139 of Lecture Notes in Statistics. Springer-Verlag, New York.

Park, Y.Y., Buizza, R. and Leutbecher, M., 2008. TIGGE: Preliminary results on comparing and combining ensembles. Quarterly Journal of the Royal Meteorological Society, 134(637): 2029-2050.

Parrish, M.A., Moradkhani, H. and DeChant, C.M., 2012. Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation. Water Resources Research, 48(3): n/a-n/a.

Peng, P., Barnston, A.G. and Kumar, A., 2013. A comparison of skill between two versions of the NCEP climate forecast system (CFS) and CPC's operational short-lead seasonal outlooks. Weather and Forecasting, 28(2): 445-462.

Raftery, A.E., 1986. Choosing models for cross-classifications. American Sociological Review, 51(1): 145-146.

Raleigh, M.S., Lundquist, J.D. and Clark, M.P., 2015. Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. Hydrol. Earth Syst. Sci., 19(7): 3153-3179.

Rayner, S., Lach, D. and Ingram, H., 2005. Weather Forecasts are for Wimps: Why Water Resource Managers Do Not Use Climate Forecasts. Climatic Change, 69(2-3): 197-227.

Robertson, D.E., Shrestha, D.L. and Wang, Q.J., 2013. Post processing rainfall forecasts from numerical weather prediction models for short term streamflow forecasting. Hydrology and Earth System Sciences Discussions, 10(5): 6765-6806.

Roulin, E., & Vannitsem, S. (2012). Postprocessing of ensemble precipitation predictions with extended logistic regression based on hindcasts. *Monthly Weather Review*, *140*(3), 874–888.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J. and Behringer, D., 2010. The NCEP climate forecast system reanalysis. Bulletin of the American Meteorological Society, 91(8): 1015-1057.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y. and Iredell, M., 2014. The NCEP climate forecast system version 2. Journal of Climate, 27(6): 2185-2208.

Salvadori, G. and De Michele, C., 2010. Multivariate multiparameter extreme value models and return periods: A copula approach. Water resources research, 46(10).

Schaake, J., Demargne, J., Hartman, R., Mullusky, M., Welles, E., Wu, L., Herr, H., Fan, X. and Seo, D.J., 2007. Precipitation and temperature ensemble forecasts from single-value forecasts. Hydrology and Earth System Sciences Discussions Discussions, 4(2): 655-717.

Sklar, M., 1959. Fonctions de répartition à n dimensions et leurs marges. Université Paris 8.

Steinschneider, S., Polebitski, A., Brown, C. and Letcher, B.H., 2012. Toward a statistical framework to quantify the uncertainties of hydrologic response under climate change. Water Resources Research, 48(11).

Stephens, M.A., 1974. EDF statistics for goodness of fit and some comparisons. Journal of the American statistical Association, 69(347): 730-737.

Sumner, G., Homar, V., & Ramis, C. (2001). Precipitation seasonality in eastern and southern coastal Spain. *International Journal of Climatology*, *21*(2), 219–247. http://doi.org/10.1002/joc.600.

Tao, Y., Duan, Q., Ye, A., Gong, W., Di, Z., Xiao, M. and Hsu, K., 2014. An evaluation of post-processed TIGGE multimodel ensemble precipitation forecast in the Huai river basin. Journal of Hydrology, 519, Part D(0): 2890-2905.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., & De Roo, A. (2013). Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin. *Journal of Hydrology*, *499*, 324–338.

Wang, Q.J. and Robertson, D.E., 2011. Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. Water Resources Research, 47(2): W02546.

Wang, Q.J., Robertson, D.E. and Chiew, F.H.S., 2009. A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. Water Resources Research, 45(5): W05407.

Wang, W., Hung, M.-P., Weaver, S.J., Kumar, A. and Fu, X., 2014. MJO prediction in the NCEP Climate Forecast System version 2. Climate dynamics, 42(9-10): 2509-2520.

Wilks, D. S. 2009. Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteorological Applications*, *16*(3), 361–368.

Wilks, D.S., 2011. Statistical methods in the atmospheric sciences, 100. Academic press.

Wood, A.W. and Schaake, J.C., 2008. Correcting errors in streamflow forecast ensemble mean and spread. Journal of Hydrometeorology, 9(1): 132-148.

Wu, L., Seo, D.-J., Demargne, J., Brown, J.D., Cong, S. and Schaake, J., 2011. Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast for hydrologic ensemble prediction. Journal of hydrology, 399(3): 281-298.

Yan, H., DeChant, C.M. and Moradkhani, H., Improving Hydrologic Data Assimilation by a Multivariate Particle Filter-Markov Chain Monte Carlo, pp. 0773.

Ye, A., Duan, Q., Yuan, X., Wood, E.F. and Schaake, J., 2014. Hydrologic post-processing of MOPEX streamflow simulations. Journal of Hydrology, 508: 147-156.

Zehe, E. and Blöschl, G., 2004. Predictability of hydrologic response at the plot and catchment scales: Role of initial conditions. Water Resources Research, 40(10).

Zhao, L., Duan, Q., Schaake, J., Ye, A. and Xia, J., 2011. A hydrologic post-processor for ensemble streamflow predictions. Advances in Geosciences, 29(29): 51-59.

Zhao, T., Wang, Q.J., Bennett, J.C., Robertson, D.E., Shao, Q. and Zhao, J., 2015. Quantifying predictive uncertainty of streamflow forecasts based on a Bayesian joint probability model. Journal of Hydrology, 528: 329-340.

Figure

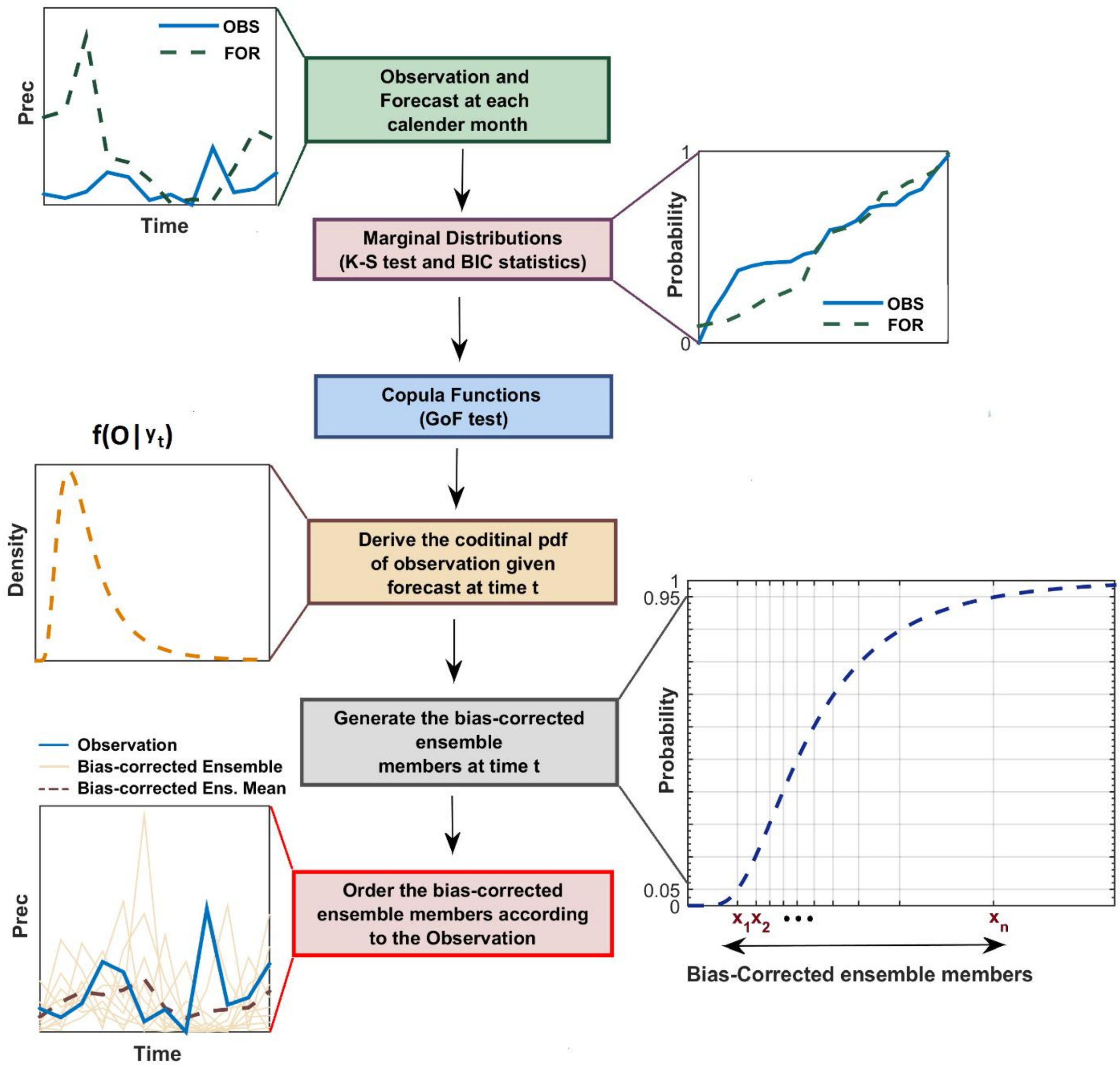


Figure 1. Schematic of the Copula-based ensemble post-processing (COP-EPP). The ensemble members are generated by sampling from conditional pdf and reconstructed according to Schaake shuffle.

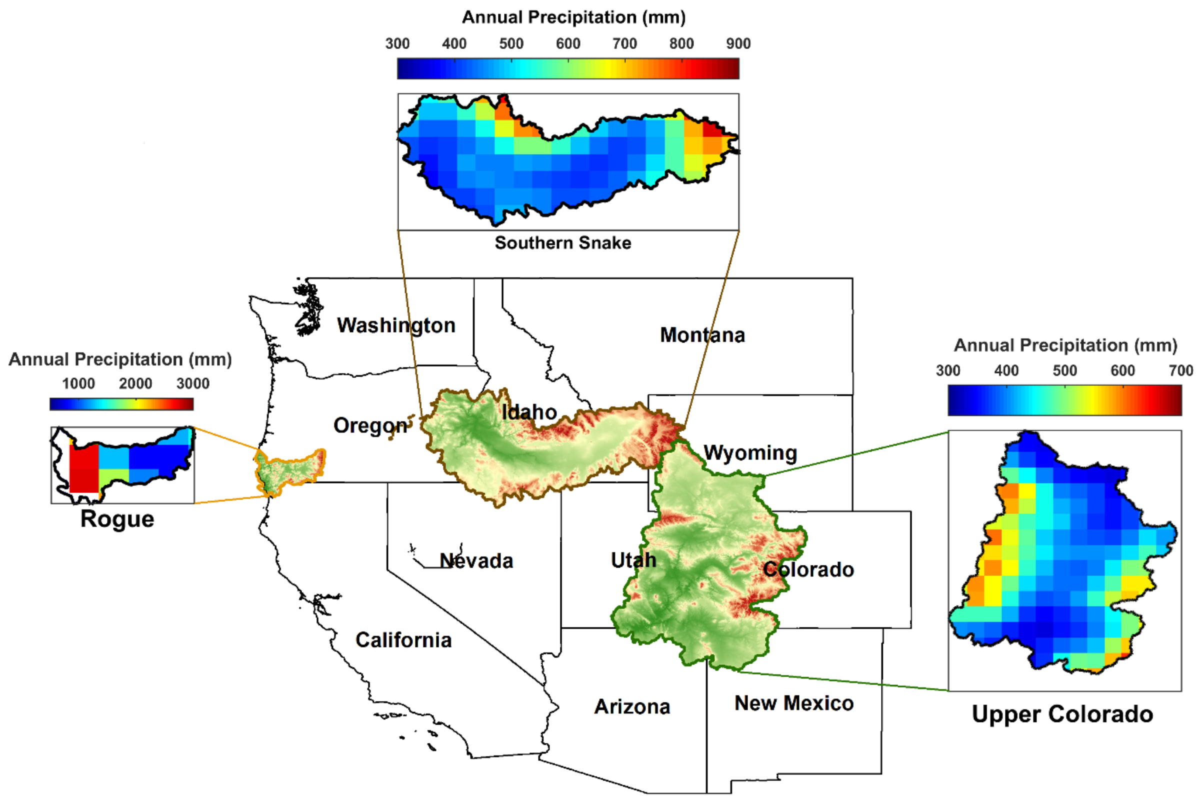


Figure 2. The location of three study basins in the Western USA.

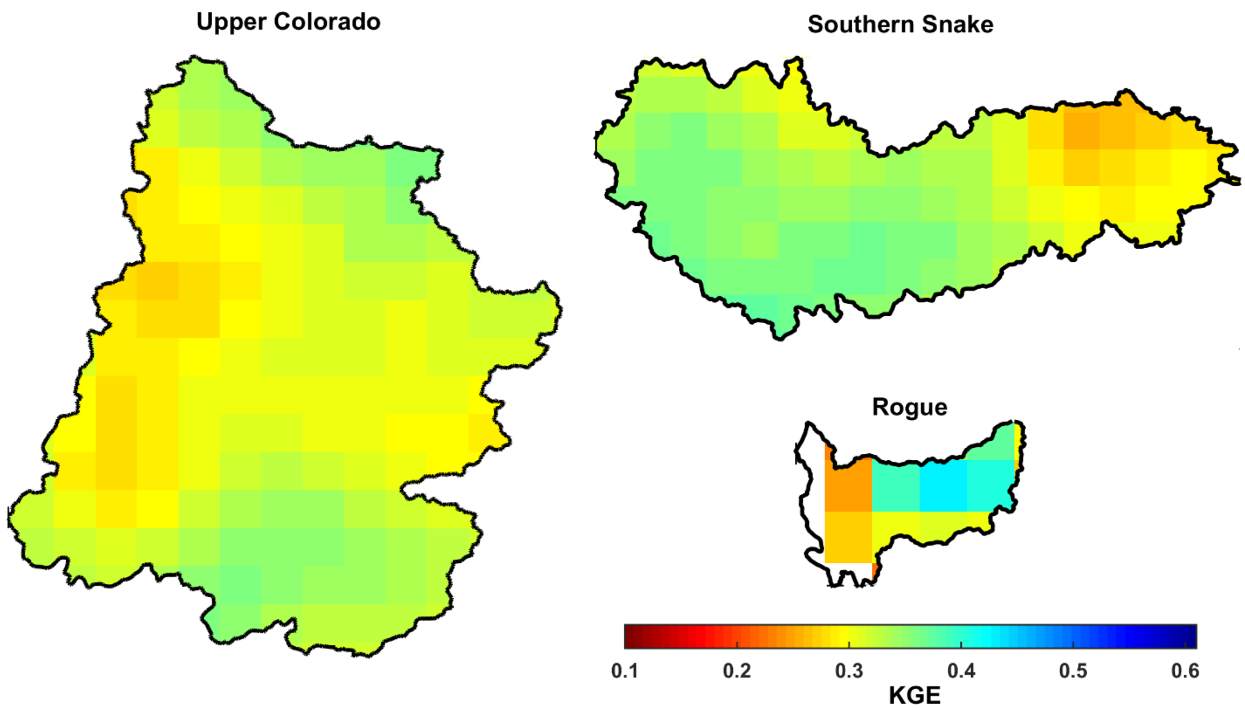
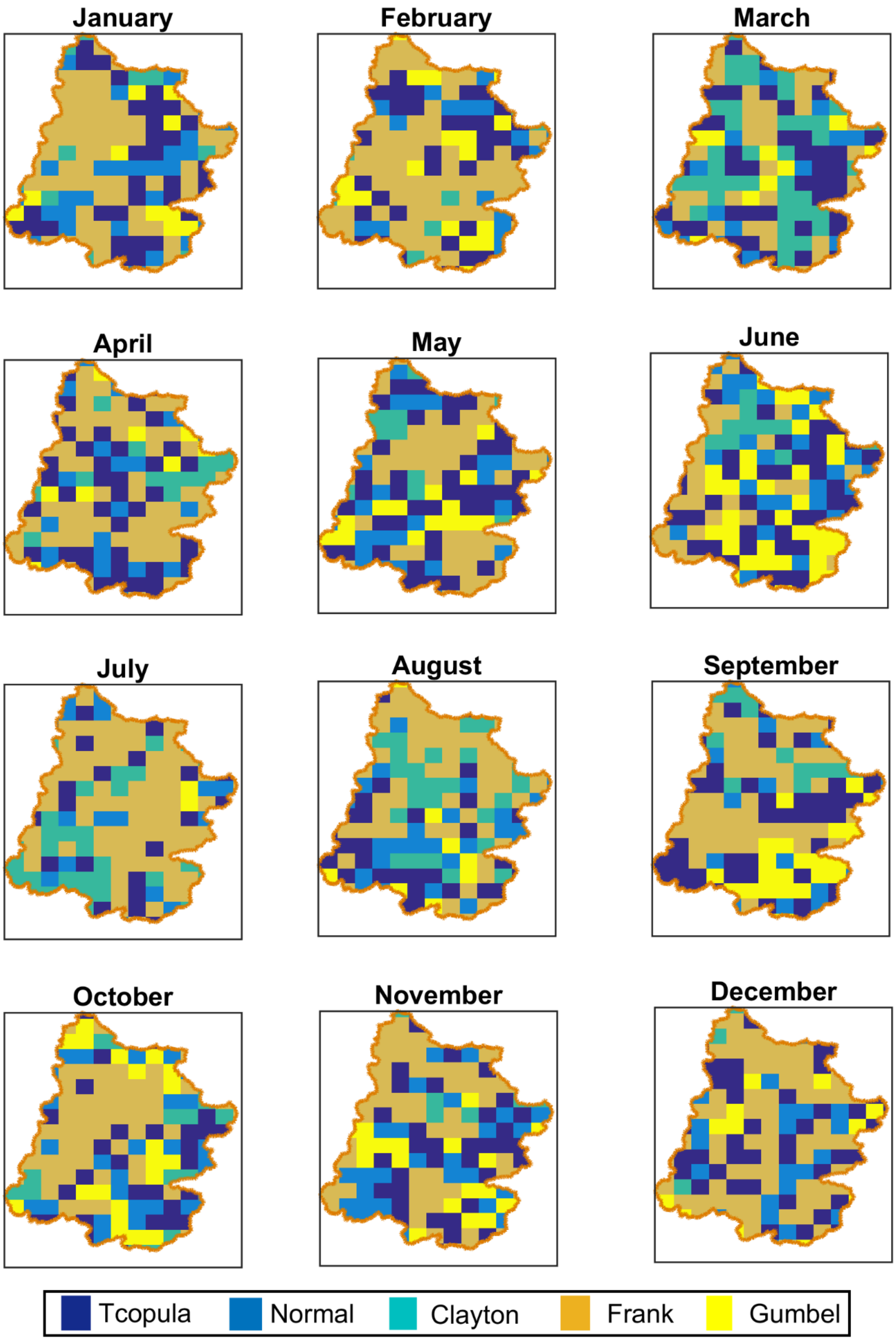


Figure 3. Kling-Gupta efficiency measure for the raw forecast during calibration period. White grid cells in Rouge River Basin present missing observations.



January
February
March
April
May
June
July
August
September
October
November
December
Tcopula
Normal
Clayton
Frank
Gumbel

Figure 4. Selection of suitable copula functions for each grid cell across the Upper Colorado River Basin.

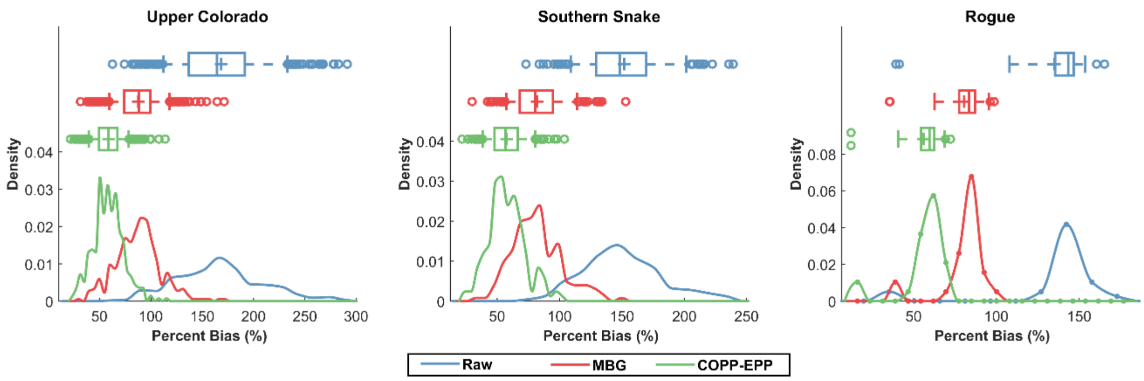


Figure 5. Absolute percent bias distribution over the three study basins. The bias is calculated for the raw forecast (blue line), and the ensemble mean from the MGB and COP-EPP outputs (red and green lines, respectively).

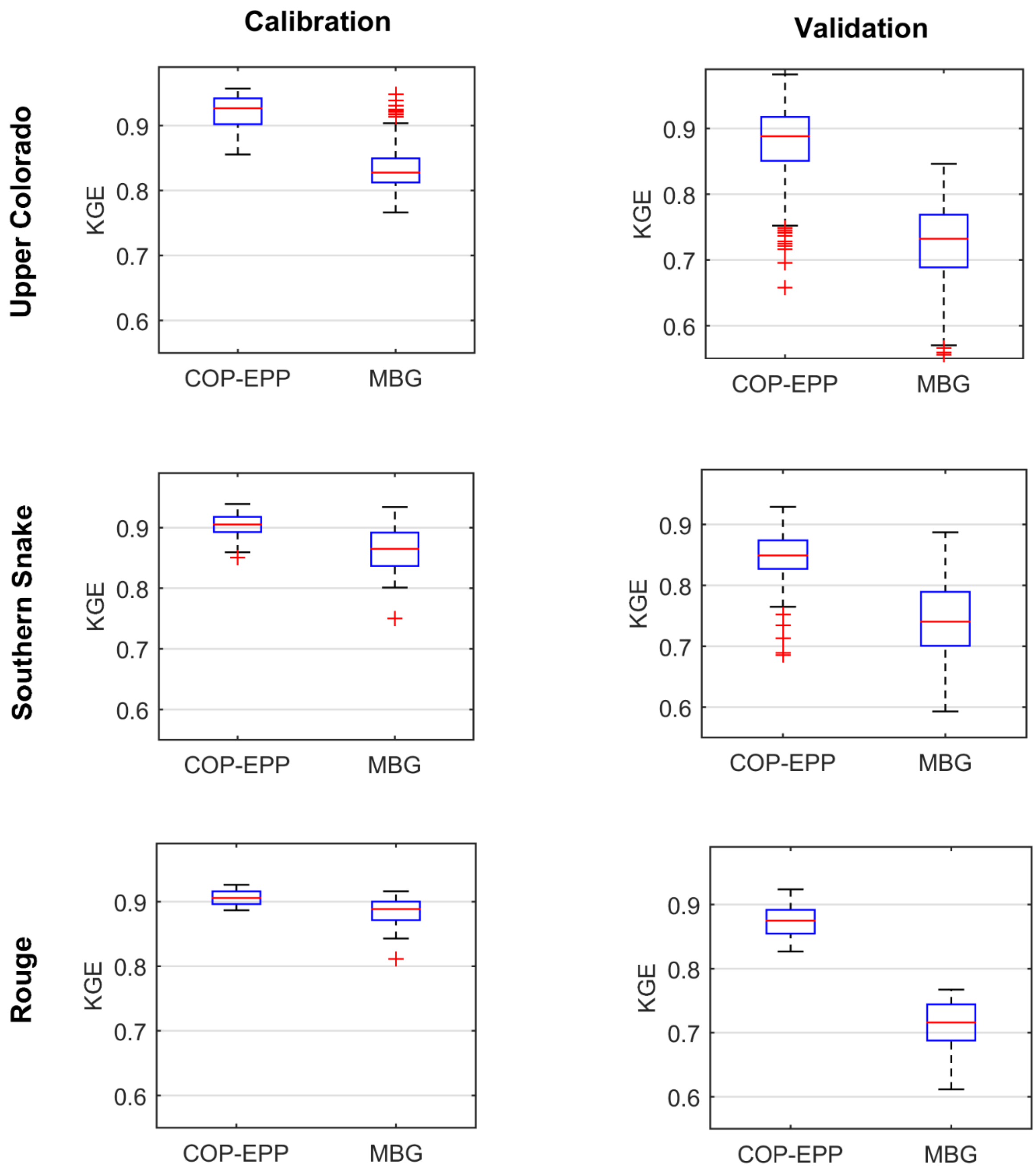
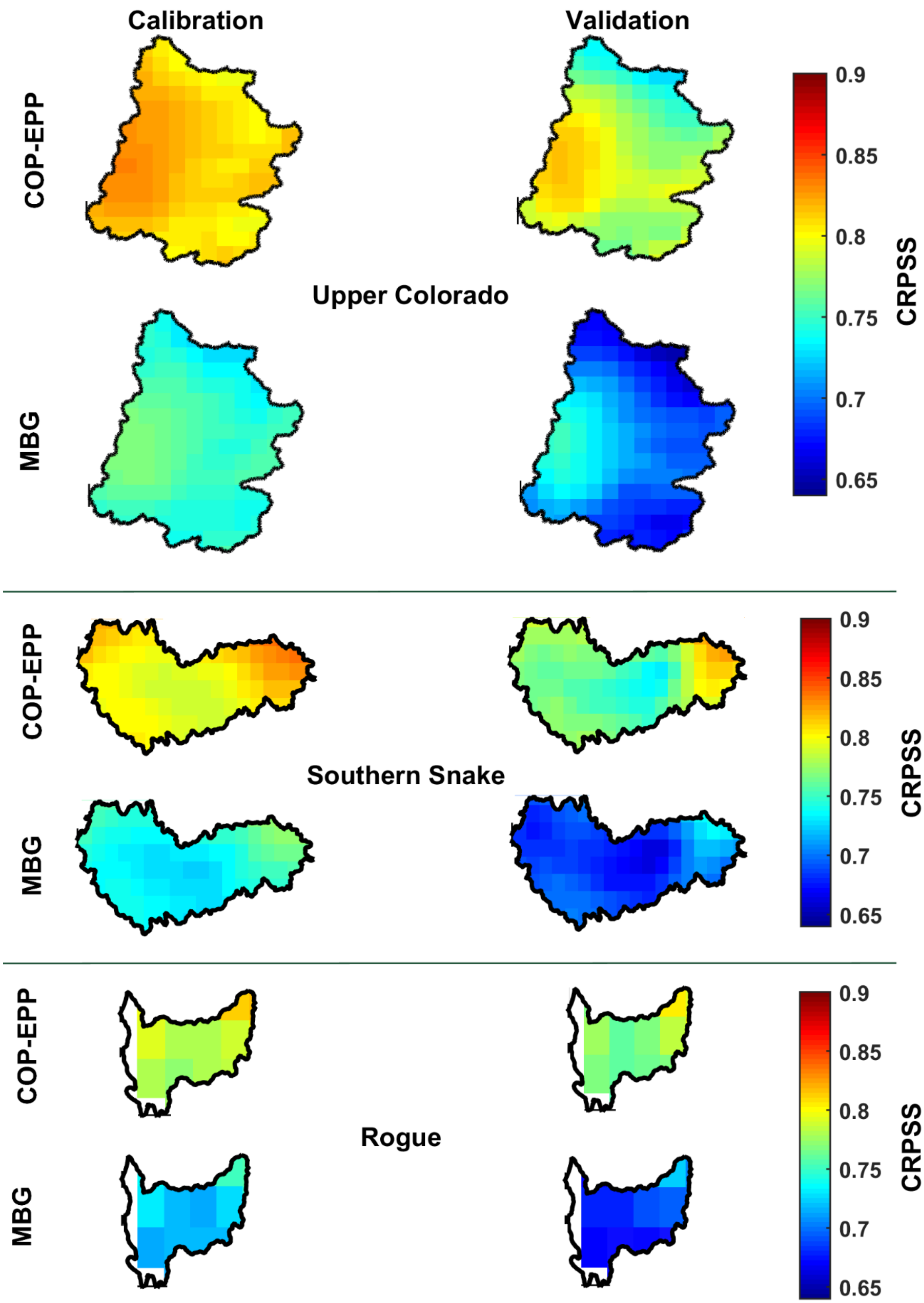


Figure 6. KGE measure calculated over the study basins after post-processing using the COP-EPP and MBG methods during calibration (left) and verification (right) periods.



Calibration
Validation
COP-EPP
MBG
Upper Colorado
0.9
0.85
0.8
0.75
0.7
0.65
CRPSS
COP-EPP
MBG
Southern Snake
0.9
0.85
0.8
0.75
0.7
0.65
CRPSS
COP-EPP
MBG
Rogue
0.9
0.85
0.8
0.75
0.7
0.65
CRPSS

Figure 7. CRPSS measure calculated for 3 basins after implementing two post-processing methods for the calibration (left) and verification (right) periods.

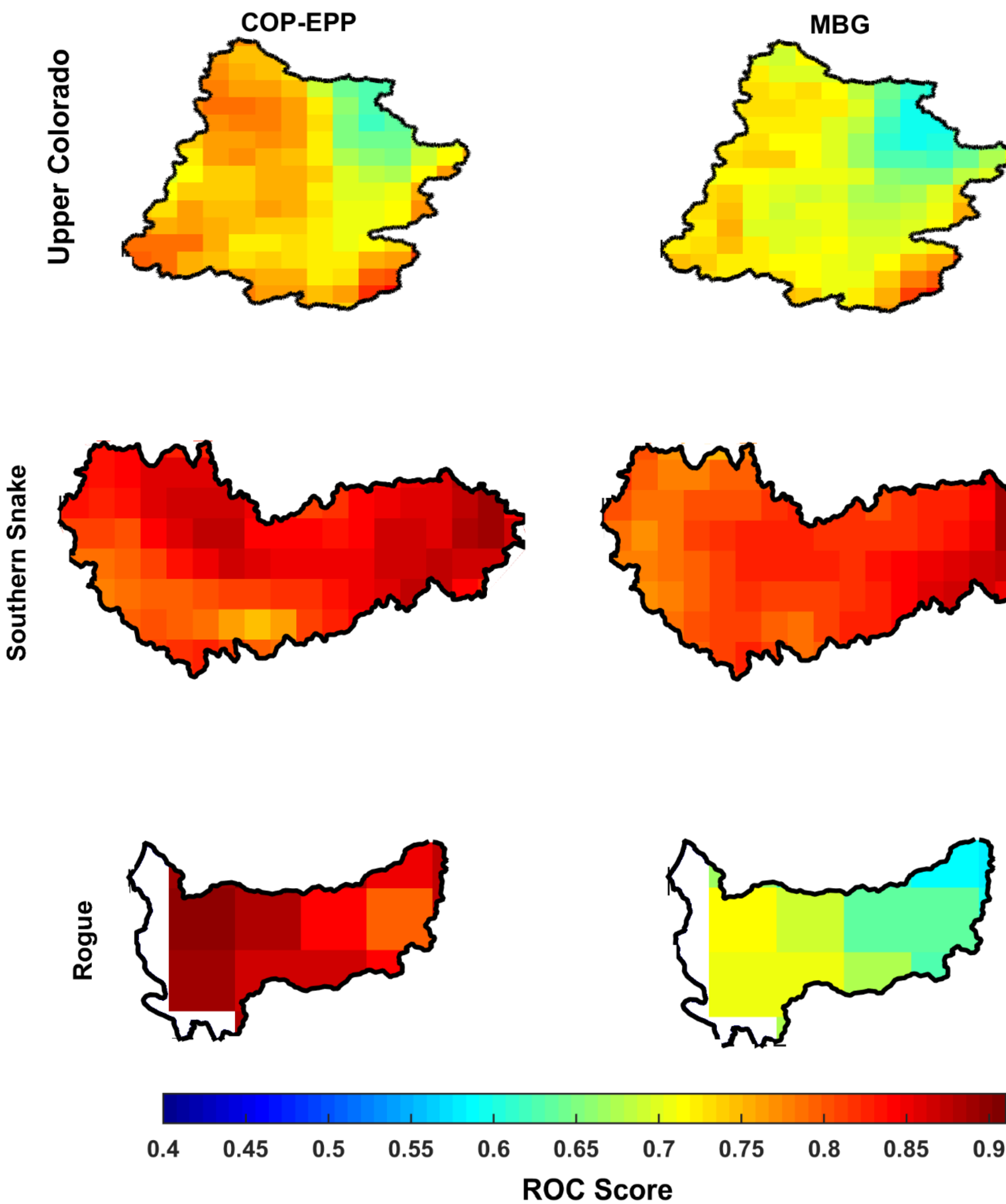


Figure 8. Assessment of forecast resolution based on ROC score for winter (Dec, Jan, and Feb) precipitation during verification period (2001-2014).

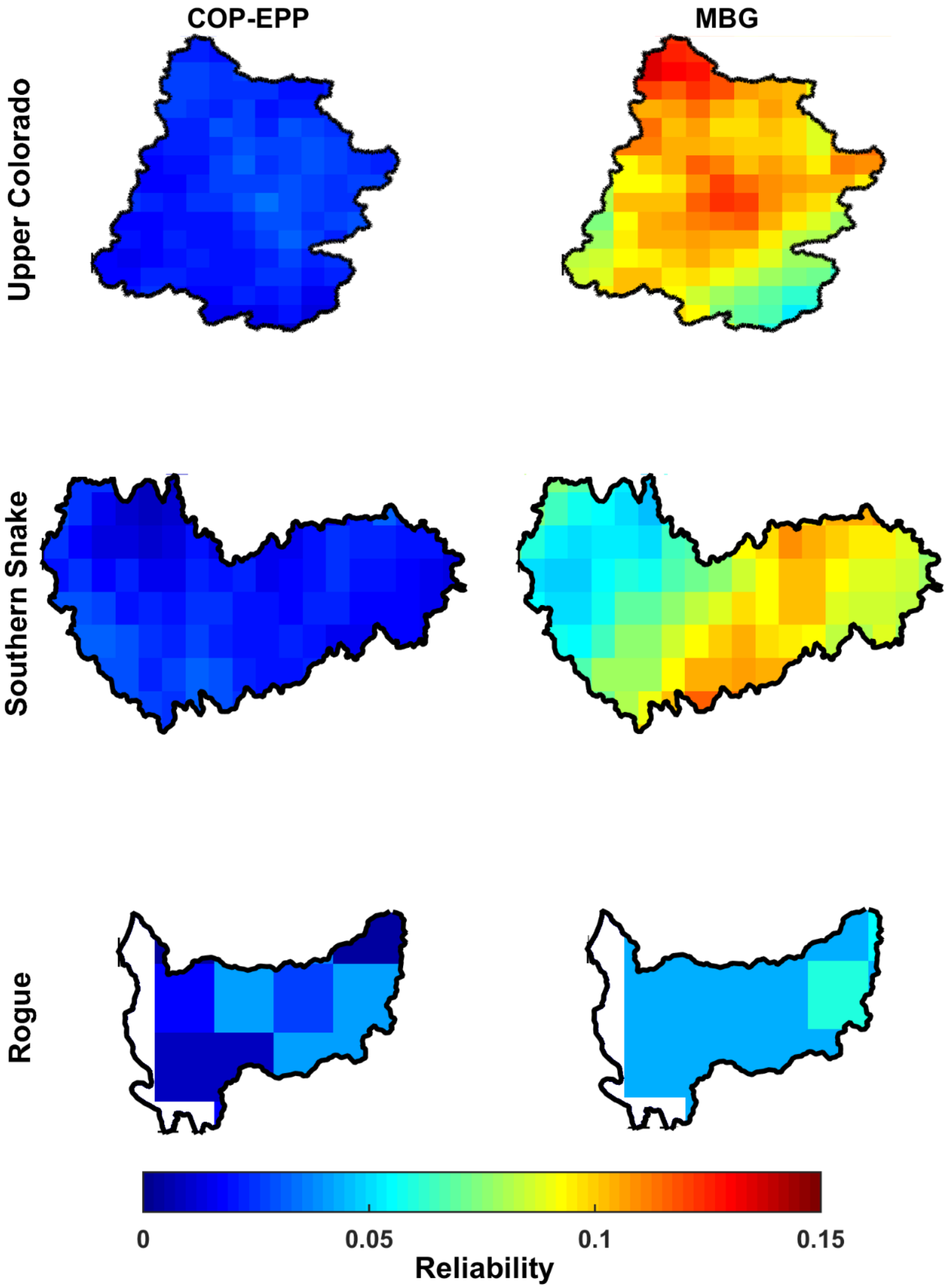

COP-EPP
MBG
Upper Colorado
Southern Snake
Rogue
0
0.05
0.1
0.15
Reliability

Figure 9. Reliability measure for winter precipitation (Dec, Jan, and Feb) calculated at 95th percentile of observation during the verification period (2001-2014). This measure ranges from 0 to 1 with the optimal value of 0.

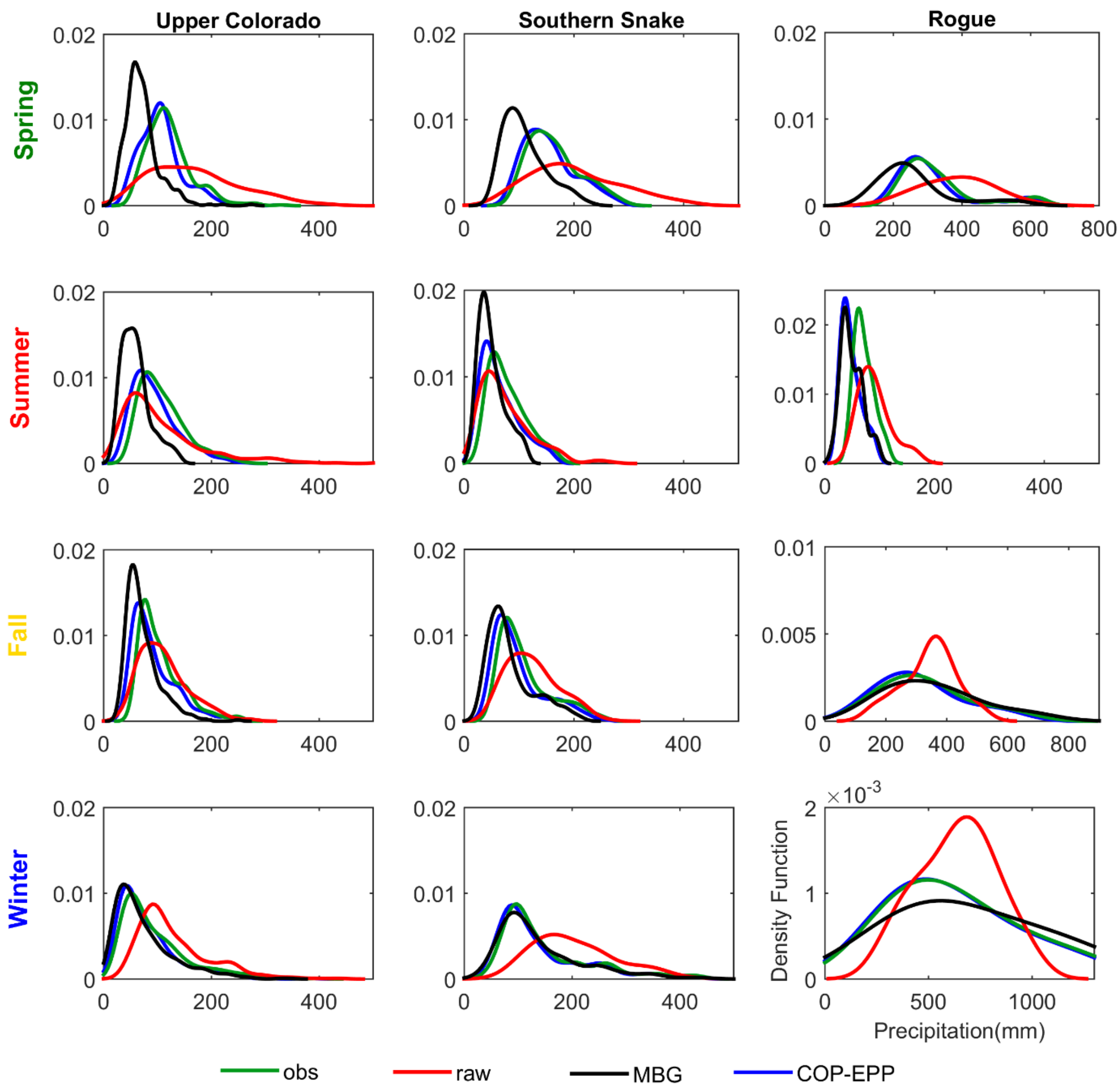


Figure 10. Probability density functions of seasonal precipitation from the observation, raw forecast, and most probable forecast based on the COP-EPP and MBG approach for three study basins. Seasonal precipitation is spatially averaged over all grid cells of each basin. Seasons are

categorized in the following order: Spring (Mar, Apr, and May), summer (Jun, Jul, and Aug), fall (Sep, Oct, and Nov), and winter (Dec, Jan, and Feb).

Table

Table 1. Summary characteristics of the study basins

Basin Name	Drainage Area (km^2)	Annual Precipitation (mm)
Upper Colorado	280000	164
Southern Snake	180000	493
Rogue	5318	970