

Portland State University

PDXScholar

---

Civil and Environmental Engineering Faculty  
Publications and Presentations

Civil and Environmental Engineering

---

11-13-2014

# Improved Bayesian Multi-Modeling: Integration of Copulas and Bayesian Model Averaging

Shahrbanou Madadgar  
*Portland State University*

Hamid Moradkhani  
*Portland State University*, [hamidm@pdx.edu](mailto:hamidm@pdx.edu)

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/cengin\\_fac](https://pdxscholar.library.pdx.edu/cengin_fac)



Part of the [Civil and Environmental Engineering Commons](#)

Let us know how access to this document benefits you.

---

## Citation Details

Madadgar, Shahrbanou and Moradkhani, Hamid, "Improved Bayesian Multi-Modeling: Integration of Copulas and Bayesian Model Averaging" (2014). *Civil and Environmental Engineering Faculty Publications and Presentations*. 174.

[https://pdxscholar.library.pdx.edu/cengin\\_fac/174](https://pdxscholar.library.pdx.edu/cengin_fac/174)

This Post-Print is brought to you for free and open access. It has been accepted for inclusion in Civil and Environmental Engineering Faculty Publications and Presentations by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

# Improved Bayesian Multi-modeling: Integration of Copulas and Bayesian Model Averaging

Shahrbanou Madadgar<sup>a</sup> and Hamid Moradkhani<sup>b,c</sup>

<sup>a</sup>Postdoctoral Research Associate

<sup>b</sup> Associate Professor, Department of Civil and Environmental Engineering, Portland State University, Portland, OR 97201, USA, [hamidm@pdx.edu](mailto:hamidm@pdx.edu)

<sup>c</sup> Corresponding Author, 1930 SW 4th Avenue, suite 200, Portland, OR 97201, phone: (503) 734-6413

## Abstract

Bayesian Model Averaging (BMA) is a popular approach to combine hydrologic forecasts from individual models, and characterize the uncertainty induced by model structure. In the original form of BMA, the conditional probability density function (PDF) of each model is assumed to be a particular probability distribution (e.g. Gaussian, gamma, etc.). If the predictions of any hydrologic model do not follow certain distribution, a data transformation procedure is required prior to model averaging. Moreover, it is strongly recommended to apply BMA on unbiased forecasts, whereas it is sometimes difficult to effectively remove bias from the predictions of complex hydrologic models. To overcome these limitations, we develop an approach to integrate a group of multivariate functions, the so-called copula functions, into BMA. Here, we introduce a copula-embedded BMA (Cop-BMA) method that relaxes any assumption on the shape of conditional PDFs. Copula functions have a flexible structure and do not restrict the shape of posterior distributions. Furthermore, copulas are effective tools in removing bias from hydrologic forecasts. To compare the performance of BMA with Cop-BMA, they are applied to hydrologic forecasts from different rainfall-runoff and land-surface models. We consider the streamflow observation and simulations for ten river basins provided by the Model Parameter Estimation Experiment (MOPEX) project. Results demonstrate that the predictive distributions are more accurate and reliable, less biased, and more confident with small uncertainty after Cop-BMA application. It is also shown that the post-processed forecasts have better correlation with observation after Cop-BMA application.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/2014WR015965

## 1. Introduction

Reliability of hydrologic forecasts is affected by several sources, including the uncertain and inaccurate meteorological forcing, initial condition at forecast date (e.g. soil moisture, groundwater level, snow water equivalent, etc.), and erroneous model structure and parameters. There are several techniques to quantify the uncertainty in short- and long- term hydrologic forecasts, including Generalized Likelihood Uncertainty Estimation (GLUE; Beven and Binley, 1992), Bayesian Recursive Estimation (Thiemann et al., 2001), Bayesian Total Error Analysis (BATEA; Kavetski et al., 2002; Kuczera et al., 2006), Data Assimilation (Moradkhani et al., 2005, Moradkhani et al., 2012), Bayesian Model Averaging (Duan et al., 2007), and Bayesian hierarchical models (Huard and Mailhot, 2008; Najafi and Moradkhani, 2013).

One of the primary techniques to reflect different uncertainties in hydrological forecasts is to create an ensemble of forecast trajectories (McEnery et al., 2005; Seo et al., 2006; Olsson and Lindstrom, 2008; Moradkhani and Sorooshian, 2008; DeChant and Moradkhani, 2011; Moradkhani et al., 2012; Madadgar and Moradkhani, 2011; Pagano et al., 2013; Madadgar et al., 2014). An ensemble of streamflow forecasts, the so-called Ensemble Streamflow Prediction (ESP; Day, 1985) may be generated by forcing a hydrologic model with an ensemble of historical climate observations or the climate forecasts from numerical climate models. Some recent developments to improve the forecast skill of ESPs include integrating data assimilation to ESP to more correctly account for uncertainty in initial conditions (DeChant and Moradkhani, 2011) and weighting ESP traces (Najafi et al., 2012). Very recent study by DeChant and Moradkhani (2014) has provided an integrated approach to account for both initial condition and model structural uncertainties in seasonal streamflow forecast.

Aside from traditional ESP, some techniques develop the probability distribution of forecast to account for different sources of uncertainty. The forecast probability distribution is usually assumed to be a multivariate normal distribution joining the observed and predicted variables (Kelly and Krzystofowicz, 1997; Schaake et al., 2007; Todini, 2008; Zhao et al., 2011; Ye et al., 2014). Using the multivariate normal distribution requires the transformation of non-Gaussian variables to standard normal variates

which can affect the accuracy of estimated probability distribution (Brown and Seo, 2012; Madadgar et al., 2014). Brown and Seo (2012) discussed the difficulties in parametric estimation of the conditional probability distribution of observation given forecast and proposed a non-parametric technique to approximate the full conditional probability distribution with a discrete set of thresholds for observed variable. However, Madadgar et al. (2014) argued the accuracy of the non-parametric distribution (Brown and Seo, 2012) for being highly dependent on the number of observation thresholds and applied copula functions (Joe, 1997; Nelsen, 1999) to develop a new technique for estimating the conditional probability distribution and post-processing the forecast of hydrologic models.

Since copula functions can model the correlation structure among the variables in hydrologic processes, they have been examined in several hydrologic applications (Favre et al., 2004; Bárdossy, 2006; Shiau, 2006; Salvadori and De Michele, 2010; Kao and Govindaraju, 2010; Madadgar and Moradkhani, 2013a). Unlike other approaches for estimating the conditional probability distributions, copula functions join variables via their marginal distributions; and hence, the unknown and complex relationships in hydrological processes do not hinder modelling the joint behavior of variables. Copula functions are not restricted to any particular type of parametric functions (e.g. normal distribution) for the marginal distributions or the joint probability distribution. According to the promising results of using copula functions in post-processing of hydrologic forecasts, Madadgar and Moradkhani (2013b, 2014) applied copula functions to obtain the conditional probability of future droughts within Bayesian network of sequential events.

Uncertainties in hydrologic predictions are partially due to model structure, parameterization, and spatial discretization of physical processes. Since any hydrologic model is a simplified representation of complex physical processes in the hydrologic system, the assumptions in model conceptualization cause hydrologic predictions to become inaccurate and imprecise. To address this, several techniques have been developed to combine the prediction of multiple models to reduce model uncertainty through multi-model combination (e.g., Duan et al., 2007; Hsu et al., 2009; Najafi et al., 2011; Parrish et al., 2012). An attractive attribute of multi modeling is that the forecast skill of a multi-model ensemble is generally

better than the participating models alone. Weigel et al. (2008) discussed the overconfidence of single-model ensembles may be reduced through a multi-model approach, where the ensemble spread becomes a more appropriate representation of the uncertainty. A well-known approach to combine an ensemble of models is model averaging, which is a linear combination of different models. Some model-averaging techniques such as equal weights, Granger-Ramanathan averaging (Granger and Ramanathan, 1984), Bates-Granger averaging (Bates and Granger, 1969), AIC and BIC-based model averaging (Buckland et al. 1997; Burnham and Anderson 2002; Hansen, 2008) take the linear average of the deterministic outputs and produce a combined single-value forecast (Diks and Vrugt, 2010). Despite the promising performance of these model-averaging techniques, Hoeting et al. (1999) argued that the weights could not properly reflect the strength of single models and recommended the use of Bayesian Model Averaging (BMA). BMA combines the forecast PDF of different models and build a weighted predictive distribution out of them. Neuman (2002) discussed the computational effort and prior information required in BMA and proposed a maximum likelihood version of BMA, which later initiated the application of BMA in subsurface hydrology (Ye et al., 2004). Thereafter, several other studies have reported BMA application in groundwater hydrology and hydraulics (e.g. Tsai and Li, 2008; Rojas et al, 2010; Ye et al., 2010). In order to calculate the weight of each forecast model, Raftery et al. (2005) used Expectation-Maximization algorithm (EM) and estimated the weights based on the performance of each model during a training period. They applied BMA in developing the predictive PDF of an ensemble of meteorological forecasts, which motivated several applications of BMA in surface hydrology (Vrugt et al., 2006; Duan et al., 2007; Vrugt and Robinson, 2007; Ajami et al., 2007). In a climate change impact study, Najafi et al. (2011) used the BMA framework to incorporate the outputs of different hydrologic models forced by a group of downscaled Global Climate Models (GCMs). Recently, Möller et al. (2013) evaluated the joint behavior of weather quantities in a two-step approach using BMA and multivariate functions; where in the first step, BMA applied to post-process the forecast ensembles of several meteorological variables, and in the second step, a Gaussian copula estimated the multivariate distribution of forecast variables.

In standard BMA (Raftery et al., 2005), the conditional PDF of each individual model is assumed to follow a normal distribution, which is valid for only a limited group of forecast variables, e.g. temperature and sea-level pressure. For other variables such as precipitation or streamflow, the normal distribution might be a poor choice (Sloughter et al., 2006), and gamma distribution might be a better alternative for representing the posterior distribution of model outputs (Vrugt and Robinson, 2007; Sloughter et al., 2010). However, data transformation is usually required in a BMA application to transform the model forecasts to the space of posterior distribution. In a recent study by Parrish et al. (2012) the Gaussian assumption of the likelihood function was relaxed by combining sequential data assimilation and BMA. The proposed method showed greater skill and reliability in high ranges of hydrograph volatility.

This study seeks to integrate a group of multivariate functions, called copula functions, into BMA to estimate the posterior distribution of model forecasts. This approach removes the need to assume the form of the posterior distribution, and eliminates the data-transformation procedure. Application of copula functions in post-processing of hydrologic forecasts (Madadgar et al., 2014) and estimating the probability of future droughts (Madadgar and Moradkhani, 2013b and 2014) have indicated the capability of copula functions in estimating the conditional probability of hydrologic forecasts. In addition to the integration of these functions into BMA, copula functions can be utilized in any Bayesian formulation where a conditional PDF should be approximated.

The main contribution of copula-integrated BMA is the estimation of posterior distribution with the help of copula functions. Some limitations of standard BMA method such as independency of prediction models remains unchanged in the new method. It should be noted that a key requirement for reliable performance of BMA is to select independent models (mutually exclusive). Without independent models, the uncertainty of predictive distribution will be overestimated. On the other hands, the models should be collectively exhaustive to assure capturing the observation variability within the ensemble. With the limitations in model selections, however, it is not practically possible to have a mutually exclusive and collectively exhausted (MECE) ensemble of models (Refsgaard et al., 2012). One approach to assure the collectively exhaustive criterion is to construct a large ensemble of models, which may violate the

mutually exclusive criterion. Therefore, it is important to establish a balance in constructing the ensemble so that the MECE criterion is met. The same challenge is expected for Cop-BMA method, since the overall structure of both methods is similar.

This paper is organized as follows. Section 2 describes the standard BMA, and Section 3 introduces the new integrated copula-BMA model. Section 4 explains the hydrologic models and the study river basins, and Section 5 describes the verification measures employed to compare the performance of model averaging techniques. Section 6 discusses the results, and finally, Section 7 summarizes the major remarks of the study.

## 2. Bayesian Model Averaging (BMA)

Bayesian Model Averaging (BMA) is an approach to combine the forecast densities predicted by different models, producing a new forecast PDF. According to BMA, the predictive distribution of a forecast variable  $y$ , given the independent predictions of  $k$  models,  $[M_1, M_2, \dots, M_k]$ , and the observations during the training period,  $Y$ , can be expressed by the law of total probability as:

$$p(y|M_1, M_2, \dots, M_k, Y) = \sum_{i=1}^k p(M_i|Y) p(y|M_i, Y) \quad (1)$$

In Eq. 1,  $p(y|M_i, Y)$  is the posterior distribution of  $y$  given the model prediction,  $M_i$ , and training data,  $Y$ . More simply,  $p(y|M_i, Y)$  is the forecast PDF of  $y$  given model  $i$ .  $p(M_i|Y)$  is the likelihood of model prediction being correct, given the observations,  $Y$ , during the training period, which reflects the performance of model  $i$  in predicting the forecast variable during the training period, such that  $\sum_{i=1}^k p(M_i|Y) = 1$ . Since the posterior probabilities of model predictions sum to unity,  $p(M_i|Y)$  is the weight of model  $i$ ; and therefore, the BMA approach returns the weighted average of forecast PDF generated by each model. While Eq. 1 shows the general form of the BMA forecast density, model predictions are time-variant, where  $y$  and  $M_i$  in Eq. 1 are replaced by  $y^t$  and  $M_i^t$ , respectively. Hence, Eq. 1 may be re-written as follows:

$$p(y^t | M_1^t, M_2^t, \dots, M_k^t, Y) = \sum_{i=1}^k w_i p(y^t | M_i^t, Y) \quad (2)$$

Since  $w$  is estimated from the model performance during the training period, it is static for each model.

Application of BMA requires the model forecasts,  $M_i$ , to be bias-corrected; that is, the expected value of observation should be equal to the expected value of forecasts for each model ( $E[Y - M_i] = 0$ ). If model forecasts are biased, any bias-correction methods such as linear regression should be applied prior to BMA implementation and the bias-corrected forecasts ( $f_i^t$ ) should replace the original model predictions ( $M_i^t$ ) (Raftery et al., 2005):

$$f_i^t = a_i + b_i M_i^t \quad (3)$$

Where,  $f_i^t$  is the bias-corrected forecast; and  $\{a_i, b_i\}$  are the coefficients of linear regression model. Defining the posterior distribution of Eq. 2 is a significant challenge in deterministic modeling of hydrologic variables, where the forecast probability is not basically estimated by the model. In order to approximate the forecast probability, it has become a common practice to assume the posterior distribution as a Gaussian function, with mean  $f_i^t$  and variance  $\sigma_i^2$ ; i.e.  $p(y^t | f_i^t, Y) \sim g(y^t | f_i^t, \sigma_i^2)$ . The variance  $\sigma_i^2$  reflects the uncertainty of the  $i^{th}$  model in respect with the mean. Despite the computational convenience of using a Gaussian distribution, the Gaussian assumption is not valid for all types of forecast variable. For non-Gaussian variables, a power transformation is needed to map the variables from their original space to a Gaussian space. Box and Cox (1964) proposed Box-Cox transformation as a general form of power transformation. The one-parameter Box-Cox transformation is defined as follows:

$$f_{i,t}^{(\lambda)} = \begin{cases} \frac{f_{i,t}^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(f_{i,t}) & \lambda = 0 \end{cases} \quad (4)$$

Where,  $f_{i,t}$  is the bias-corrected forecast of model  $i$  at time  $t$ ;  $\lambda$  is the Box-Cox coefficient; and  $f_{i,t}^{(\lambda)}$  is the transformed, bias-corrected forecast of model  $i$  at time  $t$ . The Box-Cox transformation is applied to



the observation and unbiased forecast of each model. Since several optimization algorithms may involve in estimating the Box-Cox coefficient ( $\lambda$ ) in Eq. 4, this study sets the value of  $\lambda$  equal to the average amount obtained from seven different methods as presented in Asar et al. (2014), including Shapiro-Wilk and Shapiro-Francia tests (Shapiro and Wilk, 1965), Cramer-von Mises test (Cramer, 1928), Pearson Chi-square test (Pearson, 1900), Lilliefors (Kolmogorov-Smirnov) test (Lilliefors, 1967), Jarque-Bera test (Jarque and Bera, 1987), and artificial covariate method (Dag et al., 2013). Finally, the K-S test statistics is utilized to prove the Gaussianity of transferred data.

The variance ( $\sigma_i^2$ ) and weight ( $w_i$ ) of each forecast model can be estimated by the log-likelihood function. Setting the vector of parameters as  $\theta = \{w_i, \sigma_i^2, i = 1..k\}$ , the log-likelihood function of  $\theta$  can be approximated as follows:

$$l(\theta) = \log \left( \sum_{i=1}^k w_i \cdot p(y|f_i, Y) \right) \quad (5)$$

Since the analytical solution to maximize the log-likelihood function is complex, Raftery et al. (2005) suggested a procedure called the Expectation-Maximization (EM) algorithm. This is an iterative algorithm with an embedded optimization component to update the weights and variances of posterior distributions until a termination criterion ( $|l(\theta_{Iter}) - l(\theta_{Iter-1})| < \varepsilon$ ) is achieved. In the first iteration (Iter = 1), the initial weight and variance for each model are set to  $w_{i,Iter} = \frac{1}{k}$  and  $\sigma_{i,Iter}^2 =$

$\frac{1}{k} \sum_{t=1}^T \frac{\sum_{i=1}^k (y^t - f_i^t)^2}{T}$ . As the EM algorithm proceeds,  $w_i$  and  $\sigma_i^2$  are updated as follows:

$$w_{i,Iter} = \frac{1}{T} \sum_{t=1}^T z_{i,Iter}^t$$

$$\sigma_{i,Iter}^2 = \frac{\sum_{t=1}^T z_{i,Iter}^t \cdot (y^t - f_i^t)^2}{\sum_{t=1}^T z_{i,Iter}^t} \quad (6)$$

$$z_{i,Iter}^t = \frac{w_{i,Iter-1} \cdot g(y^t | f_i^t, \sigma_{i,Iter-1}^2)}{\sum_{i=1}^k w_{i,Iter-1} \cdot g(y^t | f_i^t, \sigma_{i,Iter-1}^2)}$$

$$l(\theta_{Iter}) = \log \left( \sum_{i=1}^k w_{i,Iter} \sum_{t=1}^T g(y^t | f_i^t, \sigma_{i,Iter}^2) \right)$$

Where, T is the length of the training period; and z is a latent variable. All other variables are defined the same as in the previous equations.

In assigning the weights, Refsgaard et al. (2012) and Rojas et al. (2010) discussed that model rankings are not necessarily similar for both training and testing periods. They argued that how much the ranking would be stable where the conditions or even variables used in validation period are different from those used for calibration of the multi-modelling method. In other words, the optimal weights during the calibration period may not remain optimal for the validation period. However, if the training period can reasonably capture the overall behavior of each model, the rankings are probably similar for both calibration and validation period.

### 3. Copula-Embedded Bayesian Model Averaging (Cop-BMA)

As described in section 2, the BMA predictive distribution is a weighted average of forecast PDFs which are generally assumed to be a parametric distribution, e.g. Gaussian distribution. Here, we explain a procedure that when combined with BMA, it relaxes the assumption on posterior parametric distribution. This can modify the BMA predictive distribution and increase the multi-modeling reliability. To find the posterior distribution, we replace the parametric distribution  $g(y|f_i, \sigma_i^2)$  with a group of multivariate functions called copula, which have been reported as promising tools in hydrological forecasts. Madadgar et al. (2014) used copula functions to post-process the hydrologic forecasts and determined the streamflow forecast probability distribution. Following the similar statistical approach, this study employs copula functions to estimate the posterior distribution of forecast variables for each model, i.e.  $p(y^t | f_i^t, Y)$ , .

Copulas are multivariate distribution functions on the  $n$ -dimensional unit cube. The variables of copula function are uniformly distributed on the interval  $[0, 1]$ , i.e.  $C: [0, 1]^n \rightarrow [0, 1]$  (Joe, 1997; Nelsen,

1999). Using Sklar's theorem (Sklar, 1959), a multivariate distribution,  $P(x_1..x_n)$ , can be expressed by copula functions as follows:

$$P(x_1, \dots, x_i, \dots, x_n) = C[P(x_1), \dots, P(x_i), \dots, P(x_n)] = C(u_1, \dots, u_i, \dots, u_n) \quad (7)$$

where,  $C$  is the Cumulative Distribution Function (CDF) of the copula; and  $P(x_i)$  is the marginal distribution of  $x_i$  being uniform on the interval  $[0, 1]$ , which is also denoted by  $u_i$ . If a copula's CDF is absolutely continuous, its PDF can be expressed as:

$$c(u_1, \dots, u_n) = \frac{\partial^n C(u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n} \quad (8)$$

Using the PDF of copula (Eq. 8), the joint probability density function of  $(x_1..x_n)$  can be defined as follows:

$$p(x_1, \dots, x_n) = c(u_1, \dots, u_n) \prod_{i=1}^n p(x_i) \quad (9)$$

Alternatively, in statistical applications, the conditional probability distribution of  $x_1$  given  $x_2$  is expressed as:

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)} \quad (10)$$

Replacing the joint probability distribution of  $p(x_1, x_2)$  in Eq. 10 with Eq. 9, the conditional probability distribution of Eq. 10 can be revised as:

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)} = \frac{c(u_1, u_2) \cdot p(x_1) \cdot p(x_2)}{p(x_2)} = c(u_1, u_2) \cdot p(x_1) \quad (11)$$

Madadgar et al. (2014) applied the joint probability distribution of Eq. 11 in post-processing of hydrologic forecasts, finding that the raw forecast of hydrologic models is improved significantly after using copula functions. The post-processed forecast was shown to increase reliability while reducing uncertainty comparing to the raw forecast. Given the forecast  $x_2$ , the conditional PDF of observation  $x_1$  might be a multi-modal distribution, since the conditional PDF is defined as the product of observation PDF and the PDF of copula function. To build the conditional PDF of Eq. 11, Madadgar et al. (2014) recommended Monte Carlo sampling from the copula density function,  $c(u_1, u_2)$ . In the Monte Carlo

sampling,  $u_2$  is fixed as the CDF of forecast  $x_2$ , and  $u_1$  is calculated for each sample from the observation space. Then,  $c(u_1, u_2)$ , is computed for each pair of  $(u_1, u_2)$ , and  $p(x_1|x_2)$  is obtained from Eq. 11. In this study, we consider  $x_1$  as the quantity to be forecasted ( $y^t$ ), and  $x_2$  as the  $i^{th}$  model prediction ( $f_i^t$ ). Replacing the posterior distribution in Eq. 2 by the conditional probability distribution in Eq. 11, the predictive distribution of BMA is updated as follows:

$$p(y^t|f_1^t, f_2^t, \dots, f_k^t, Y) = \sum_{i=1}^k w_i p(y^t|f_i^t, Y) = \sum_{i=1}^k w_i c(u_{y^t}, u_{f_i^t}) p(y^t) \quad (12)$$

Using Eq. 12 relaxes any assumption on the type of posterior distribution. Fig. 1 shows the flowchart of BMA and Cop-BMA and compares the different steps of each method. As seen in Cop-BMA, the posterior distribution,  $p(y|f_i, Y)$ , is directly obtained from Eq. 12 without a need to use EM algorithm. Another advantage of using copula functions is their strong ability to remove bias from model predictions (Madadgar et al., 2014). Therefore, no external bias-correction method is required to be involved in Cop-BMA.

After the posterior distribution is defined, their weights are estimated via the EM algorithm (Eq. 6) with a few adjustments in some equations:

$$w_{i,Iter} = \frac{1}{T} \sum_{t=1}^T z_{i,Iter}^t$$

$$z_{i,Iter}^t = \frac{w_{i,Iter-1} \cdot p(y^t|f_i^t)}{\sum_{i=1}^k w_{i,Iter-1} \cdot p(y^t|f_i^t)} = \frac{w_{i,Iter-1} \cdot c(u_{y^t}, u_{f_i^t}) \cdot p(y^t)}{\sum_{i=1}^k w_{i,Iter-1} \cdot c(u_{y^t}, u_{f_i^t}) \cdot p(y^t)} \quad (13)$$

$$l(\theta_{Iter}) = \log \left( \sum_{i=1}^k w_{i,Iter} \sum_{t=1}^T c(u_{y^t}, u_{f_i^t}) \cdot p(y^t) \right)$$

As seen, the calculation of variance in Eq. 6 is not appeared in Eq. 13. Moreover, the posterior probability of  $y^t$  is calculated only once in Eq. 13 and that remains the same for all the iterations. In contrast, the posterior probability in standard BMA should be re-calculated every time that the variance is updated. In addition, Cop-BMA does not need any data transformations; while in standard BMA, it is required to

transform data to comply with the distributional property of the variable of interest, such as normal distribution in case of streamflow forecast.

#### **4. Hydrologic Modeling of Study Basins**

To compare the performance of the BMA and Cop-BMA methods in streamflow forecasting, observed and simulated streamflow of 10 unregulated river basins in the United States were used. The data was provided by the Model Parameter Estimation Experiment (MOPEX) project, which is a collaborative endeavor supported by several international organizations since 1996. A full description of the MOPEX project can be found in Duan et al. (2006). The study basins are located in different climate regions of the eight States across the Southeastern quadrant of the United States; i.e. IN, MD, VA, WV, IA, NC, LA, and MO. Figure 2 shows the location of river basins, and Table 1 summarizes the specifications of each basin. The basin and station IDs are obtained from the United States Geological Survey (USGS). The drainage area of the river basins ranges from  $\sim 600$  to  $\sim 2600 \text{mi}^2$ , and the elevation of the outlet station varies between 0 and 1950 feet in the North America Vertical Datum of 1988 (NAVD88) system. The last four columns of Table 1 summarize the climatic characteristics of river basins. The river basins are selected from a variety of climate regimes, as indicated by the ratios of P/PE (mean annual precipitation, P, to the mean annual potential evapotranspiration, PE) and E/P (mean annual evaporation (E) to P). For each basin, the ratio P/PE is estimated for the gridded values of P from the Parameter-elevation Relationships on Independent Slopes Model (PRISM) dataset (Daly et al., 1994) and gridded values of PE from the National Oceanic and Atmospheric Administration's (NOAA) Freewater Evaporation Atlas (Farnsworth and Peck, 1982). Wet climate regions are expected to have high values of P/PE, while dry climate regions should result in low values (Dooge, 1997). Similar to Risley et al., (2011) and Najafi et al., (2011), we use the aridity index ( $\alpha$ ) (Budyko, 1974; Milly and Dunne, 2002) to characterize the climate regime of each basin:

$$\frac{E}{P} = \left[ \alpha \left( \tanh \frac{1}{\alpha} \right) (1 - \cosh \alpha + \sinh \alpha) \right]^2 \quad (14)$$

In Eq. 14,  $\alpha$  is the aridity index, and E and P are the mean annual evaporation and precipitation, respectively. Basins with  $\alpha \geq 1$  are located in water-limited regions, where the evaporation is constrained by water supply, implying that the region is dry. In contrary, basins with  $\alpha < 1$  are energy-limited, where the evaporation is constrained by radiation and temperature (i.e. the region is wet). Given E/P in the left-hand side of Eq. 14, the value of  $\alpha$  can be obtained for each river basin. According to the values of P/PE and  $\alpha$  in Table 1, basins #7 and #10 are located in dry regions, and basin #8 is located in the wettest region, as it has the highest value of P/PE and the lowest value of  $\alpha$ .

Streamflow observations were reported at USGS gage stations as listed in Table 1, and streamflow simulations are available via MOPEX dataset. Seven different hydrologic models, as listed in Table 2, are used to estimate the streamflow at the outlet of each river basin. The first three models in Table 2 (SAC, GR4J, SWB) are conceptual, lumped-parameter, rainfall-runoff hydrologic models and the rest (ISBA, NOAH, SWAP, VIC) are land-surface models. Interested readers are encouraged to study the references of each model as provided in Table 2.

## 5. Performance Assessment of Multi-modeling Techniques

In this section, the performance of Cop-BMA and BMA are compared using different verification measures. Accuracy, reliability, sharpness, and overall skill of the forecast predictive distribution are evaluated for each method.

*Accuracy:* Forecast accuracy is evaluated by *Mean Relative Absolute Error* (MRAE) and *Kling-Gupta Efficiency* (KGE; Kling et al., 2012) metrics. MRAE varies  $[0, \infty)$  with perfect forecast at MRAE = 0, and KGE varies  $(-\infty, 1]$  with perfect forecast at KGE = 1. In deterministic forecasts, these metrics evaluate the agreement between the simulation ( $y_s$ ) and observation ( $y_o$ ) as follows:

$$\text{MRAE} = \frac{1}{T} \sum_{t=1}^T \frac{|y_o^t - y_s^t|}{y_o^t} \quad (15)$$

$$\text{KGE} = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2}$$

Where;

$$\beta = \frac{\mu_s}{\mu_o} \tag{16}$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o}$$

Where  $\mu_o$  and  $\mu_s$  are the mean of observed and simulated variable, respectively;  $\beta$  is an indicator of bias;  $r$  is the correlation coefficient between the observation and simulation;  $CV$  is the coefficient of variation;  $\sigma_o$  and  $\sigma_s$  are the standard deviation of observation and simulation; and  $\gamma$  is variability ratio. In perfect forecast, where KGE equals unity, the  $r = \beta = \gamma = 1$ . According to Eq. 16,  $\beta < 1$  corresponds with an overall negative bias, in which  $\mu_s < \mu_o$ . Similarly,  $\beta > 1$  indicates a positive bias in forecast. While MRAE and KGE are clearly defined for deterministic forecasts, they cannot directly apply to probabilistic forecasts where the final forecast product is in the form of a predictive distribution. In such applications,  $y_s^t$  in Eq. 15 is replaced by the expected value of the estimated predictive distribution,  $E(y^t | M_1^t \dots M_k^t)$ . The same replacement occurs in the calculation of  $r$ ,  $\mu_s$ , and  $CV_s$  in Eq. 16.

*Reliability:* Forecast reliability is indicated by the supportive quantitative scores of predictive quantile-quantile (Q-Q) plot (Laio and Tamea, 2007; Thyer et al., 2009). In the predictive Q-Q plot, the quantiles in which the observations fall within the forecast distribution are compared to the cumulative uniform distribution,  $U[0,1]$ . A guide to interpreting the predictive Q-Q plot is provided in Fig. 3. While the Q-Q plot of a perfect forecast follows the uniform line, it falls below or above the uniform line in biased forecasts (Fig. 3a). The PDF of biased forecast is located behind/ahead of the actual PDF of observation, indicating a negative/positive bias, respectively (Fig. 3b). Such definition of negative/positive bias is consistent with the definition of bias in Eq. 16 ( $\beta$ ). For example, the PDF of forecast with negative bias (Fig. 3b) demonstrates the overall smaller value of forecasts comparing with observations ( $\mu_s < \mu_o \Rightarrow \beta < 1$ ), and vice versa.

If the Q-Q plot crosses the uniform line and has a small slope around the quantile 0.4-0.6, then the predictive uncertainty is overestimated or the forecast is underconfident. Conversely, a Q-Q plot with high slope around the midrange quantile (0.4-0.6) demonstrates the underestimation of uncertainty and overconfidence of forecast. The Q-Q plots associated with overconfident and underconfident forecasts in Fig. 3a cross the uniform line at quantile 0.5, indicating the median of forecast and observation distributions are superimposed as shown in Fig. 3b. Underestimation or overestimation of uncertainty indicates too little or too large spread of predictive distribution, respectively. Note that in Fig. 3, the illustrated forecasts are either purely biased or purely under-/overconfident, which might be rare in real applications. Operational forecasts are usually both positively/negatively biased and under-/overconfident to some extent.

In addition to the visual interpretations, a few quantitative scores can be computed from a Q-Q plot for probabilistic verification of forecasts. There are two reliability measures described by Q-Q plot,  $\alpha$  and  $\varepsilon$ :

$$\alpha = 1 - 2 \left[ \frac{1}{T} \sum_{t=1}^T |P_t(y_o^t) - U(y_o^t)| \right] \tag{17}$$

$$\varepsilon = 1 - \frac{1}{T} \sum_{t=1}^T I[P_t(y_o^t) = 1 \text{ or } P_t(y_o^t) = 0]$$

In Eq. 17,  $P_t(y_o^t)$  is the non-exceedance probability of observation using the forecast CDF,  $U(y_o^t)$  is the non-exceedance uniform probability of observation, and  $I$  is the indicator function. In Eq. 17,  $\alpha$  is a measure of the uniformity of the Q-Q plot, and it varies between 0 (worst reliability) and 1 (perfect reliability), and  $\varepsilon$  measures the portion of observations that occurs inside the predictive distribution, and it varies between 0 (worst reliability) and 1 (perfect reliability).

*Sharpness*: Sharpness or precision of the predictive distribution is measured by the third score derived from the Q-Q plot,  $\pi$ , defined as follows:



$$\pi = \frac{1}{T} \sum_{t=1}^T \frac{E(y^t | M_1^t \dots M_k^t)}{\sigma[y^t | M_1^t \dots M_k^t]} \quad (18)$$

$E(y^t | M_1^t \dots M_k^t)$  and  $\sigma[y^t | M_1^t \dots M_k^t]$  are the expected value and the standard deviation of the predictive distribution at time  $t$ . Larger values of  $\pi$  are the result of smaller standard deviations of the forecast distribution, indicating a greater sharpness or precision. Given two forecasts with same reliability, the forecast with larger  $\pi$  is preferred, because it has greater sharpness or less uncertainty.

Another measure indicating the effects of sharpness on the reliability of predictive distribution is called confidence, as introduced by Moradkhani et al. (2012). The confidence score indicates that if the forecast is overconfident (too little spread) or underconfident (too large spread). The score is originally defined for verification of forecast ensembles, but is considered here for predictive distributions instead. To calculate the confidence score in this study, we build an ensemble of forecast values by extracting the  $\left[\frac{1}{100}, \frac{2}{100}, \dots, \frac{99}{100}\right]$  quantiles from the forecast probability distribution. Then, the relative location of observation at time  $t$  ( $y^t$ ) from the  $i^{th}$  upper and lower quantiles ( $P_{1,i}$  and  $P_{2,i}$ ) is obtained, and the confidence score is expressed as follows:

$$z_t = \frac{1}{N} \sum_{i=1}^N x_i \quad , \quad x_i = \begin{cases} 1 & y_o^t > y_i^t \\ 0 & otherwise \end{cases}$$

$$P_{1,i} = \frac{i}{N} \quad , \quad P_{2,i} = 1 - \frac{i}{N}$$

$$W_i = \frac{1}{T} \sum_{t=1}^T \lambda_t \quad , \quad \lambda_t = \begin{cases} 1 & P_{1,i} < z_t < P_{2,i} \\ 0 & otherwise \end{cases} \quad (19)$$

$$C = \frac{2}{N} \sum_{i=1}^{N/2} [(P_{2,i} - P_{1,i}) - W_i]$$

$z_t$  is the quantile of the predictive distribution in which the observation at time  $t$  is located,  $N$  is the size of forecast ensemble (here  $N = 100$ ),  $W_i$  is the frequency that the observation falls between the  $i^{th}$  predictive bounds, and  $C$  is the confidence value. The  $C > 0$  indicates overconfidence and  $C < 0$  indicates underconfidence of predictive distribution.

## 6. Results and Discussion

BMA and Cop-BMA are employed to combine the streamflow forecast of an ensemble of seven hydrologic models. Daily forecast of streamflow is available for all river basins over the time period of 1960-1997. However, to avoid missing values in daily forecasts, we used 5-day forecasts (accumulated daily forecasts over five days) in our analyses. The first 3-year period (1960-1962) is used as the spin-up period for each hydrologic model, and the rest (1962-1997) is used for training and validating each method. In training phase, the parameters of each method are calibrated over the period of 1962-1987 at the USGS gage stations as listed in Table 1, and in validation phase, multi-modeling methods are applied to the remaining dataset (1988-1997). To consider the seasonality effects, multi-modeling is repeated for each particular month of year; that is, the 5-day forecast of the seven hydrologic models are treated separately for Jan, Feb, etc. At the end, the entire timeseries of 5-day forecast is built including all 12 months to check the performance of each multi-modeling method. In BMA, the calibration parameters are  $w_i$  and  $\sigma_i^2$ ; while in Cop-BMA, they are  $w_i$  and those associated with copula function (Eq. 11) including the parameters of marginal distributions,  $u_y$  and  $u_{M_i}$ , and the parameters of the PDF of copula,  $c$ .

In Cop-BMA application for a certain month of year, it is required to find the best probability distribution fitting the associated observations and model forecasts. Seven different probability distributions, including Gamma, Gaussian, Lognormal, Generalized Extreme Value (GEV), Exponential, Weibull, and Gumbel distributions are tried in this study. The method of Maximum Likelihood Estimation (MLE) is used for parameter estimation of each distribution, and the Kolmogorov-Smirnov (K-S) and the Akaike Information Criterion (AIC; Akaike, 1974) tests are then applied to find the best marginal distribution. The goodness of a certain distribution given a significance level ( $\alpha=0.05$  in this study) is indicated by the p-value of K-S test, and its superiority over other competing distributions is evaluated by the AIC test. After finding the best marginal distributions, a copula function is required to link the CDF of model forecasts and observations. Among different alternative copulas, the most desirable is the one giving the best connection between variables. In this study, five copula functions are

tested: Gaussian and t from Elliptical copulas, and Gumbel, Clayton, and Frank from Archimedean copulas. Then, the parametric bootstrap procedure (Genest and Rémillard, 2008) was utilized to estimate the Cramér–von Mises statistic ( $S_n$ ) and the associated p-value in order to choose the best copula function. The  $S_n$  and p-values indicated that the Gumbel copula gives the best fit for the Cop-BMA application in this study.

As explained earlier, the streamflow forecast of seven hydrologic models are available for each river basin of Table 1. The MRAE scores over the validation period (1988-1997) are calculated for each hydrologic model (Table 3) before and after the application of multi-modeling methods. The MRAE scores also indicate that the performance of land-surface models (ISBA, NOAH, SWAP, VIC) is not as good as the rainfall-runoff models (SAC, GR4J, SWB). This is mainly due to the different procedures used for the calibration of each model (Nasonova et al., 2009). It can also partially refer to the model structure, formulation of physical processes, and dissimilar sources of forcing dataset for each model. A general observation of Table 3 is that the hydrologic forecast of dry basins has lower quality than wet basins (compare the results of basins #7 and #10 with basin #8), which can be explained by various challenges in low-flow modeling such as imperfect forcing data and inaccurate estimation of model parameters in dry basins (Nasonova et al., 2009).

Table 3 also reports the results of combining hydrologic forecasts and examines the effectiveness of BMA and Cop-BMA methods for each river basin. For all river basins, the performance of Cop-BMA is better than standard BMA. The MRAE of Cop-BMA is smaller than the best individual forecast for each river basin. On the other hand, BMA could constantly outperform the best individual forecast except for three river basins (#1, #8, #9). One reason of smaller MRAE after Cop-BMA comparing with BMA is that the copula functions are very efficient at removing bias from model forecasts (Madadgar et al., 2014). They consider the dependencies and correlations between forecasts and observations, leading to significant reduction of errors and biases. Therefore, application of copula functions in BMA would not need a simple bias correction such as linear regression prior model averaging procedure anymore. Moreover, the superior performance of Cop-BMA might be partially due to the weight of each individual

model, which is directly affected by the estimation of posterior distributions. This issue is illustrated in Fig. 7 and will be discussed later.

Figure 4 shows the bar plots of KGE score and its components  $r$ ,  $\beta$ , and  $\gamma$  (Eq. 16) after model averaging. For each river basin, the KGE score is significantly greater after Cop-BMA application in comparison with BMA. The percentage of KGE improvement after Cop-BMA is from 10% in basin #4 to 200% in basin #9. The correlation coefficient between forecast and observation is quite smaller after BMA application (comparing with Cop-BMA) in basin #8 and #10. It varies between 0.53 for basin #10 and 0.87 for basin #4 after BMA multi modeling, while it remains constantly above 0.8 after Cop-BMA application for all river basins except basin #10 ( $R=0.73$ ). Regarding the bias score, both models reach almost similar results except for three river basins (#1, #8, #9). According to the definition of bias in Eq. 16, both methods generally underestimate the mean flow (negative bias); however, Cop-BMA is more promising in locating forecasts close to observations. Unlike bias, the variability of forecast is significantly different between the two methods (Fig. 4d). Results indicate that the variability of forecast is larger after Cop-BMA but still less than the sufficient variability (i.e.  $\gamma = 1$ ). Except for three cases (basins #1, #5, #10), Cop-BMA could perfectly capture the variability of observed flow ( $\gamma \approx 1$ ), while BMA could only capture the 80 percent of variability in most of the river basins.

For probabilistic verification of forecasts, the predictive Q-Q plots are illustrated for all ten river basins in Figure 5. Overall, the Q-Q plot indicates the higher reliability and smaller bias of Cop-BMA forecasts as compared with BMA. In basin #9, the bias is quite large after BMA application which has been already expressed in Fig. 4c. There is an obvious negative bias in basins #1 and #8, and a clearly positive bias in basin #7 after the application of each method. Regarding forecast reliability, the area between the Q-Q plot and the uniform line for a few river basins (i.e. #2, #3, #7) is approximately the same for the two methods. Then, it is expected that the  $\alpha$  scores (Eq. 17) be similar for those river basins. In terms of predictive uncertainty, it is seen that the BMA forecasts are underconfident (i.e. overestimated uncertainty) for most of the river basins, whereas Cop-BMA forecast is underconfident just in few basins, mainly #6, #9, #10. However, in cases with small reliability and large bias (i.e. basins #1, #7, #8, #9 after

the application of either method), a visual inspection of the Q-Q plot cannot give sufficient information on the under-/overconfidence of predictive distribution; hence, the quantitative measures such as sharpness ( $\pi$ ) can help verify the predictive uncertainty. Note that, in hydrological applications, improving the forecast reliability and accuracy is the first priority, whereas sharpness (precision) is at the second place. However, for similarly reliable forecasts, the one with higher sharpness is desirable.

Figure 6 shows forecast reliability, sharpness, and confidence of BMA and Cop-BMA forecasts for each basin. According to Eq. 17, the value of 0 for  $\alpha$  and  $\varepsilon$  indicates a poor reliability and the value of 1 indicates a perfect reliability. As seen, Cop-BMA forecast is generally more reliable than BMA forecast in respect with  $\alpha$ . In particular, the results of basin #9 show a significant improvement from  $\alpha = 0.64$  for BMA forecast to  $\alpha = 0.89$  for Cop-BMA forecasts (i.e. 40% increase in reliability). As a reminder,  $\alpha$  is a measure of closeness of the Q-Q plot to the line 1:1 and the results in Fig. 6a have been already expected from the visual inspection of the Q-Q plots in Fig. 5. With respect to  $\varepsilon$ , BMA and Cop-BMA could both capture the observations within the predictive uncertainty; however, BMA performs better than Cop-BMA in basins #2 and #3. However, sharpness ( $\pi$ ) is much greater after Cop-BMA application (Fig. 6c). A large value of  $\pi$  implies a small standard deviation of the predictive distribution (Eq. 18), indicating large sharpness and small predictive uncertainty. Small uncertainty of the predictive distribution in Cop-BMA forecasts implies that the proposed Cop-BMA is a more precise approach. It should be noted that the total variance of predictive distribution is a combination of two terms: between-model variance and within-model variance (Duan et al., 2007). The first term represents the ensemble spread, and the second term represents the within-ensemble spread and is proportional to the variance of posterior distribution. According to the sharpness results (Fig. 6c), the total predictive variance is reduced by replacing the Gaussian posterior distribution with the PDF generated by copula functions. Since the between-model variance remains the same in both methods, it appears that the within-model variance is reduced after copula application. However, the C values in Fig. 6d indicate that the predictive distributions of both methods are rather underconfident in most of the study basins. Generally, the Cop-BMA forecasts are

more confident with smaller uncertainty. In basin #8, the predictive distribution of Cop-BMA is overconfident and highly precise as indicated by  $\pi$  and C values. In basin #1, the BMA forecast is overconfident according to its C value; however, it is less sharp than Cop-BMA forecast. It should be noted that more sharpness does not necessarily imply more confidence. According to the definition of  $\pi$  and C (Eq. 18 and 19), the value of  $\pi$  shows the sharpness or precision of the forecast distributions, while the value of C shows the confidence of forecast, which is related to how the observed values are distributed within the predictive distribution. In other words, a sharp distribution might be underconfident according to the relative location of the observation within the forecast distribution. Then, a sharp distribution might still be underconfident.

For better understanding the reasons behind the different performance of BMA and Cop-BMA, the weights assigned to each forecast model are plotted in Fig. 7. Each data point shows the weight obtained in BMA against the weight obtained in Cop-BMA for each individual model for a certain month of year; that is, there are 84 data points ( $7$  (models)  $\times$   $12$  (months) =  $84$ ) in each subplot. Except in a few river basins, the correlation of weights obtained from the two multi-modeling methods is very small, which can clearly explain the different performance of BMA and Cop-BMA. As seen in Eq. 6 and 13, the weight of each model is defined as a function of latent variable ( $z$ ), which is directly calculated from the posterior probability of training dataset. While the EM algorithm may have possibly converged to local optimums in either multi-modeling method, the considerable role of posterior distribution in EM results is certainly approved. Even using global optimization methods is not likely to change the final weights obtained by the EM algorithm (Vrugt et al., 2008). Therefore, the influence of Cop-BMA is not only limited to the definition of posterior distributions, it also affects the performance of EM algorithm and the weights obtained for each forecast model.

## **7. Summary and Conclusion**

Inaccurate and unreliable forecasts in hydro-meteorology are the artifacts of different sources of uncertainties propagated into the system. These include the forcing data, model structure, model

parameters and system initial condition. A well-known approach to characterize and reduce the model structural uncertainty in hydrologic predictions is to combine an ensemble of model predictions in a multi-model framework. Among several multi-modeling approaches, Bayesian Model Averaging has found more interest in research and operational settings. BMA returns the predictive distribution of forecast variables as a weighted average of posterior distributions. The weights of participating models are defined proportional to the performance of each model during a training period. In BMA, the EM algorithm is traditionally used to approximate the parameters of posterior distributions in an iterative procedure, which repeats until a pre-defined termination criterion is achieved. Since deterministic models do not create any probability distribution for their forecast, the posterior probability is assumed to follow a particular distribution such as Gaussian or gamma distribution. However, if the forecast variable is not following a particular distribution, data transformation might be needed before and after multi modeling, which had been shown to possibly damage the forecast skill of the predictive distribution. Furthermore, a bias-correction procedure is usually required in BMA applications.

Given the abovementioned limitations, this study employed a group of multivariate functions (copulas) to modify the structure of the BMA technique. Since copula functions have shown success in different hydrologic forecasting applications, this study utilized them in model averaging to find the posterior distribution of data given model predictions. The new method, Cop-BMA, is more flexible in defining the posterior distribution and does not impose any restriction on the type of distribution. Although BMA is not either theoretically limited to a certain type of posterior distribution, the uni-modal distributions such as Gaussian or gamma distribution are commonly used as posterior distributions. In contrast, Cop-BMA has a flexible structure which allows the posterior distribution to have any uni-modal or multi-modal shape depending on the copula function. By relaxing the assumptions on the type of posterior distribution, data transformation would not be required. Furthermore, Cop-BMA can effectively remove the bias of initial forecasts by itself and do not need any external bias-correction method.

This study applied BMA and Cop-BMA in streamflow forecasting of 10 unregulated river basins with a large range of drainage area located in different climate regimes of the Southeastern United States.

Seven hydrologic models with different complexities, including rainfall-runoff and land-surface models, were used for streamflow forecasts of each river basin. The streamflow forecast, observation, and hydrologic characteristics of each river basin are available via the MOPEX project. Using different verification measures, BMA and Cop-BMA were compared in estimating the predictive distribution of streamflow.

Cop-BMA displayed better deterministic skill than BMA in this study. According to the MRAE and KGE scores, forecast accuracy increased significantly after applying Cop-BMA in all river basins. Results of KGE scores showed higher correlation coefficient between observation and forecasts, less bias, and better variability ratio after Cop-BMA application. The probabilistic skill of predictive distributions was evaluated by four different scores, including three scores derived from the predictive QQ plot ( $\alpha, \varepsilon, \pi$ ), and Confidence score (C). The results of QQ plot showed more reliable, less biased, and more confident forecasts after applying Cop-BMA. Comparing with BMA, the predictive distribution estimated by Cop-BMA was more precise with small uncertainty. Moreover, our results confirmed the impact of posterior distribution in calculating the weights of individual models by EM algorithm. The small correlation between models' weights obtained in each multi-modeling method clearly explained the significant role of posterior distribution in the performance of EM algorithm in finding the optimal weights.

The results of this study are encouraging for further integration of copula functions into hydro-meteorological applications where unknown conditional distributions are required to be estimated. The flexible structure of copula functions allows Cop-BMA to be applied to a large number of variables in hydrology, meteorology, and climatology (e.g. precipitation, temperature, wind speed, sea level pressure). In the application of Cop-BMA, the key requirement is to find an appropriate marginal distribution for each variable, and with the availability of wide range of parametric distributions, it is always likely to find a marginal distribution with a reasonable fit.



**Acknowledgements:** Partial financial support for this work was provided by NOAA grant no. NA11NWS4680002. The dataset used in this study was provided by NOAA's National Weather Service (NWS) Model Parameter Estimation Experiment (MOPEX) [http://www.nws.noaa.gov/ohd/mopex/mo\\_datasets.htm](http://www.nws.noaa.gov/ohd/mopex/mo_datasets.htm).

## References

- Ajami, N. K., Q. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resour. Res.*, 43, W01403, doi:10.1029/2005WR004745.
- Akaike, H., 1974. A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19 (6), 716–723.
- Asar, O., Ilk, O., and O. Dag (2014), Estimating Box-Cox power transformation parameter via goodness of fit tests. arXiv preprint arXiv:1401.3812.
- Bates JM, and C. W. J. Granger (1969), The combination of forecasts, *Oper. Res.*, 451–468.
- Bárdossy, A. (2006), Copula-based geostatistical models for groundwater quality parameters, *Water Resour. Res.*, 42, W11416, doi:10.1029/2005WR004754.
- Beven, K. J., and A. M. Binley (1992), The future of distributed hydrological models: Model calibration and uncertainty prediction, *Hydrol. Process.*, 6, 279-298, doi:10.1002/hyp.3360060305.
- Brown, J. D., and D. J. Seo (2012), Evaluation of a nonparametric post-processor for bias correction and uncertainty estimation of hydrologic predictions, *Hydrol. Process.*, accepted article, doi: 10.1002/hyp.9263.
- Buckland, S. T., Burnham, K. P., and N. H. Augustin (1997), Model selection: an integral part of inference, *Biometrics*, 53, 603–618
- Budyko MI. (1974), *Climate and life*, Academic Press: New York, London.

- Burnash, R. J. C., Ferral, R. L., and R.A. McGuire (1973), A Generalized Streamflow Simulation System, Conceptual Modeling for Digital Computers, Sacramento, CA: National Weather Service and California Department of Water Resources.
- Burnham, K. P., and D. R. Anderson (2002), Model Selection and Multimodel Inference: a practical information-theoretic approach, 2nd edn. Springer, New York.
- Chen, F., K. Mitchell, J. Schaake, Y. Xue, H. Pan, V. Koren, Y. Duan, M. Ek, and A. Betts (1996), Modeling of land-surface evaporation by four schemes and comparison with FIFE observations, *J. Geophys. Res.*, 101 (D3), 7251-7268.
- Cramer, H. (1928), On the composition of elementary errors, *Scandinavian Actuarial Journal*, 11:141-180.
- Dag, O., Asar, O., and O. Ilk (2014), A methodology to implement Box-Cox transformation when no covariate is available, *Communications in Statistics-Simulation and Computation*, 43(7), 1740-1759.
- Daly, C., Neilson, R. P., and D. L. Phillips (1994), A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J Applied Meteorology*, 33, 140-158
- Day, G.N. (1985), Extended streamflow forecasting using NWSRFS, *J. Water Resources Planning and Management*, 111(2), 157–170.
- DeChant, C., and H. Moradkhani (2011), Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation, *Hydrol. Earth Syst. Sci.*, 15, 3399-3410, doi: 10.5194/hess-15-3399.
- DeChant C.M., and H. Moradkhani (2014), Toward a Reliable Prediction of Seasonal Forecast Uncertainty: Addressing Model and Initial Condition Uncertainty with Ensemble Data Assimilation and Sequential Bayesian Combination, *J. Hydrology*, special issue on Ensemble Forecasting and data assimilation, doi: 10.1016/j.jhydrol.2014.05.045.
- Diks, C. G. H., and J. A. Vrugt (2010), Comparison of point forecast accuracy of model averaging methods in hydrologic applications, *Stoch. Environ. Res. Risk Assess.*, 24, 809–820, doi:10.1007/s00477-010-0378-z.

- Dooge, J.C. (1997), Scale problems in hydrology. In: Buras, N. (Ed.), *Reflections on Hydrology: Science and Practice*, American Geophysical Union, Washington, DC, pp. 84–143.
- Duan, Q. , Schaake, J., Andreassian, V., Franks, S., Goteti, G., H.V. Gupta et al.(2006), Model parameter estimation experiment: overview of science strategy and major results of the second and third workshops, *J Hydrology*, 320, 3-17
- Duan, Q., N. K. Ajami., X. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using Bayesian model averaging, *Adv. Water Resour.*, 30(5), 1371–1386.
- Farnsworth, R. K., and E. L. Peck (1982), *Evaporation atlas for the contiguous 48 United States*, NOAA Tech. Rep. NWS 33, US Dept. of Commerce, Washington DC.
- Favre, A. C., Adlouni, S. E., Perreault, L., Thiéonge, N., and B. Bobée (2004), Multivariate hydrological frequency analysis using copulas, *Water Resour. Res.*, 40, W01101.
- Genest, C., and B. Rémillard (2008), Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models, *Annales de l’Institut Henri Poincaré. Probabilités et Statistiques*, 44(6), 1096–1127.
- Granger CWJ, and R. Ramanathan (1984), Improved methods of combining forecasts, *J Forecast*, 3, 197–204
- Gusev, Ye M., and O. N. Nasonova (1998), The land surface parameterization scheme SWAP: Description and partial validation, *Global Planet. Change*, 19, 63–86.
- Hansen, B. E. (2008), Least-squares forecast averaging, *J Econom*, 146, 342–350.
- Hoeting, J. A., Madigan, D., Raftery, A. E., and C. T. Volinsky (1999), Bayesian model averaging: a tutorial, *Stat. Sci.*, 14, 382–417.
- Huard, D., and A. Mailhot (2008), Calibration of hydrological model GR2M using Bayesian uncertainty analysis, *Water Resour. Res.*, 44, W02424, doi:10.1029/2007WR005949.
- Hsu K, Moradkhani H, and S. Sorooshian (2009), A sequential Bayesian approach for hydrologic model selection and prediction, *Water Resour. Res.*, 45:W00B12, 15 pp, DOI:10.1029/2008WR006824.

- Jarque, C.M., and A. K. Bera (1987), A test for normality of observations and regression residuals, *International Statistical Review*, 55(2):163-172.
- Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.
- Kao, S., and R. S. Govindaraju (2010), A copula-based joint deficit index for droughts, *J. Hydrology*, 380, 121–134.
- Kavetski, D., Franks, S., and G. Kuczera (2002), Confronting input uncertainty in environmental modelling in calibration of watershed models, in *Water Sci. Appl. Ser.*, vol. 6, edited by Q. Y. Duan, et al., pp. 49-68, AGU, Washington, D. C.
- Kelly, K.S., and R. Krzysztofowicz (1997), A bivariate meta-Gaussian density for use in hydrology, *Stochastic Hydrology and hydraulics*, 11, 17–31.
- Kling, H., Fuchs, M., and M. Paulin (2012), Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, *Journal of Hydrology*, 424, 264-277.
- Koren, V., J. Schaake, K. Mitchell, Q. Y. Duan, F. Chen, and J. M. Baker (1999), A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *J. Geophys. Res.*, 104, 19569-19585.
- Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Towards a Bayesian total error analysis of conceptual rainfall - runoff models: Characterising model error using storm-dependent parameters, *J. Hydrology*, 331(1-2), 161-177, doi:10.1016/j.jhydrol.2006.05.010.
- Laio, F. and S. Tamea (2007), Verification tools for probabilistic forecasts of continuous hydrological variables, *Hydrology and Earth System Sciences*, 11, 1267-1277.
- Li, H., Luo, L., Wood, E.F., and J. Schaake (2009), The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, *J. Geophysical Research*, 114(D4): D04114.
- Li, X., and F. T.-C. Tsai (2009), Bayesian model averaging for groundwater head prediction and uncertainty analysis using multimodel and multimethod, *Water Resour. Res.*, 45, W09403, doi:10.1029/2008WR007488.

- Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A Simple Hydrologically Based Model of Land Surface Water and Energy Fluxes for GSMs, *J. Geophys. Res.*, 99(D7), 14,415-14,428.
- Lilliefors, H.W. (1967), On the Kolmogorov-Smirnov test for normality with mean and variance unknown, *Journal of the American Statistical Association*, 62:399-402.
- Lohmann, D., et al. (1996), A large scale horizontal routing model to be coupled to land surface parameterization schemes, *Tellus* (48A), 708-721.
- Pearson, K. (1900), On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philosophical Magazine Series*, 5 50(302):157–175.
- Madadgar, S. and H. Moradkhani (2011), Improving the Ensemble Streamflow Prediction by Adjusting Hydrologic Ensemble Traces, In the proceedings of ASCE World Environmental and Water Resources Congress 2011, Palm Spring, CA.
- Madadgar, S., and H. Moradkhani (2013a), Drought Analysis under Climate Change using Copula, *J. Hydrologic Eng.*, 18(7), 746–759.
- Madadgar, S., and H. Moradkhani (2013b), A Bayesian Framework for Probabilistic Seasonal Drought Forecasting, *J. Hydrometeorology*, special issue of Advances in Drought Monitoring, 14, 1685–1705, doi: <http://dx.doi.org/10.1175/JHM-D-13-010.1>.
- Madadgar, S., Moradkhani, H., and D. Garen (2014), Towards Improved Post-processing of Hydrologic Forecast Ensembles, *Hydrological Processes*, 28: 104–122. doi: 10.1002/hyp.9562.
- Madadgar, S., and H. Moradkhani (2014), Spatio-Temporal Drought Forecasting within Bayesian Networks, *J. Hydrology*, vol. 512, 134-146.
- McEnery, J., Ingram, J., Duan, Q. Y., Adams, T., and L. Anderson (2005), NOAA’s advanced hydrologic prediction service, Building pathways for better science in water forecasting, *Bulletin of American Meteorological Society*, 86, 375–385.

- Milly, P.C.D., and K. A. Dunne (2002), Macroscale water fluxes 2. Water and energy supply control of their interannual variability, *Water Resources Research*, 38, 1206.
- Möller, A., Lenkoski, A. and T. L. Thorarinsdottir (2013), Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas, *Q.J.R. Meteorol. Soc.*, 139: 982–991.  
doi: 10.1002/qj.2009
- Moradkhani, H. and S. Sorooshian (2008), General Review of Rainfall-Runoff Modeling: Model Calibration, Data Assimilation, and Uncertainty Analysis, in Hydrological Modeling and Water Cycle, Coupling of the Atmospheric and Hydrological Models, Springer, *Water Science and Technology Library*, Volume 63, pp 1-23.
- Moradkhani, H., DeChant, C. M., and S. Sorooshian (2012), Evolution of Ensemble Data Assimilation for Uncertainty Quantification using the Particle Filter-Markov Chain Monte Carlo Method, *Water Resources Research*, VOL. 48, W12520, 14 PP., doi:10.1029/2012WR012144.
- Najafi, M., H. Moradkhani, and I. Jung (2011), Assessing the uncertainties of hydrologic model selection in climate change impact studies, *Hydrol. Processes*, 25(18), 2814–2826.
- Najafi, M.R., Moradkhani, H., and T. Piechota (2012), Ensemble Streamflow Prediction: Climate Signal Weighting vs. Climate Forecast System Reanalysis, *J. Hydrology*, 442–443 (p105–116),  
<http://dx.doi.org/10.1016/j.jhydrol.2012.04.003>.
- Najafi, M. R., and H. Moradkhani (2013), Analysis of Runoff Extremes using Spatial Hierarchical Bayesian Modeling, *Water Resources Research*, 49, 1–15, DOI:10.1002/wrcr.20381.
- Nasonova, O. N., Yeugeny, M. G., and Y. E. Kovalev (2009), Investigating the Ability of a Land Surface Model to Simulate Streamflow with the Accuracy of Hydrological Models: A Case Study Using MOPEX Materials, *J. Hydrometeorology*, 10, 1128–1150.
- Nelsen, R. B. (1999), *An Introduction to Copulas*, Springer, New York.
- Neuman, S. P. (2002), Accounting for conceptual model uncertainty via maximum likelihood model averaging, in Proceedings of 4th International Conference on Calibration and Reliability in

- Groundwater Modelling (ModelCARE 2002), edited by K. Kovar, and Z. Hrkal, pp. 529–534, Charles Univ., Prague, Czech Republic.
- Noilhan, J. and S. Planton (1989), A simple parameterization of land surface processes for meteorological models, *Mon. Wea. Rev.*, 117, 536-549.
- Noilhan, J., and J. -F. Mahfouf (1996), The ISBA land surface parameterization scheme, *Global and Plan. Change*, 13, 145-159.
- Olsson, J., and G. Lindström (2008), Evaluation and calibration of operational hydrological ensemble forecasts in Sweden, *J. Hydrology*, 350, 14–24.
- Pagano, T.C., Shrestha, D. L., Wang, Q.J., Robertson, D., and P. Hapuarachchi (2013), Ensemble dressing for hydrological applications, *Hydrological Processes*, 27: 106-116
- Parrish, M., Moradkhani, H., and C. M. DeChant (2012), Towards reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation, *Water Resources Research*, 48, W03519, doi:10.1029/2011WR011116.
- Perrin, C., Michel, C., and V. Andréassian (2003), Improvement of a parsimonious model for streamflow simulation, *J. Hydrology*, 279(1-4), 275-289.
- Raftery, A. E., Gneiting, T., Balabdaoui, F., and M. Polakowski (2005), Using bayesian model averaging to calibrate forecast ensembles, *Mon. Weather Rev.*, 113:1155–74.
- Refsgaard, J. C., Christensen, S., Sonnenborg, T. O., Seifert, D., Højberg, A. L., and L. Troldborg (2012), Review of strategies for handling geological uncertainty in groundwater flow and transport modeling, *Advances in Water Resources*, 36, 36-50.
- Risley, J., Moradkhani, H., Hay, L., and S. Markstrom (2011), Statistical Trends in Watershed Scale Response to Climate Change in Selected Basins Across the United States, *AMS Earth Interaction*, 15 (14) 1-26.
- Rojas, R., Batelaan, O., Feyen, L., and A. Dassargues (2010), Assessment of conceptual model uncertainty for the regional aquifer Pampa del Tamarugal–North Chile, *Hydrology and Earth System Sciences*, 14(2), 171-192.

- Salvadori, G., and C. De Michele (2010), Multivariate multiparameter extreme value models and return periods: A copula approach, *Water Resour. Res.*, 46, W10501, doi:10.1029/2009WR009040.
- Schaake, J. C., Koren, V. I., Duan, Q.-Y., Mitchell, K., and F. Chen (1996), Simple water balance model for estimating runoff at different spatial and temporal scales, *J. Geophys. Res.*, 101(D3), 7461-7475, doi:10.1029/95JD02892.
- Schaake, J. C., Demargne, J., Hartman, R., Mullusky, M., Welles, E., Wu, L., Herr, H., Fan, X., and D. J. Seo (2007), Precipitation and temperature ensemble forecasts from single-value forecasts, *Hydrology and Earth System Sciences Discussions*, 4, 655–717.
- Seo, D. J., Herr, H. D., and J. C. Schaake (2006), A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction, *Hydrology and Earth System Sciences Discussions*, 3(4), 1987-2035.
- Shapiro, S.S., and M. B. Wilk (1965), An analysis of variance test for normality (complete samples), *Biometrika*, 52(3/4): 591-611.
- Shiau, J. T. (2006), Fitting drought duration and severity with two-dimensional copulas, *Water Resour. Manag.*, 20(5), 795-815.
- Sklar, K. (1959), Fonctions de repartition à n Dimensions et LeuraMarges, *Publ. Inst. Stat. Univ. Paris*, 8, 229–231
- Slougher, J.M., Raftery, A. E., and T. Gneiting (2006), Probabilistic quantitative precipitation forecasting using Bayesian model averaging, University of Washington, Department of Statistics, Technical Report 496, Seattle, WA, 20 pp
- Slougher, J. M., Gneiting, T., and A. E. Raftery (2010), Probabilistic wind speed forecasting using ensembles and Bayesian model averaging, *J. Am. Stat. Assoc.*, 105, 25–35, doi :10.1198/jasa.2009.ap08615.
- Thiemann, M., Trosset, H., Gupta, H., and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, 37(10), 2521-2535, doi:10.1029/2000WR900405.



- Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W., and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, 45, W00B14, doi:10.1029/2008WR006825.
- Tsai, F. T.-C., and X. Li (2008), Inverse groundwater modeling for hydraulic conductivity estimation using Bayesian model averaging and variance window, *Water Resour. Res.*, 44, W09434, doi:10.1029/2007WR006576.
- Todini, E. (2008), A model conditional processor to assess predictive uncertainty in flood forecasting, *International Journal of River Basin Management*, 6(2), 123–137.
- Vrugt, J.A., M.P. Clark, C.G.H. Diks, Q. Duan, and B.A. Robinson (2006), Multi-objective calibration of forecast ensembles using Bayesian Model Averaging, *Geophysical Research Letters*, 33, L19817, doi:10.1029/2006GL027126.
- Vrugt, J. A., and B. A. Robinson (2007), Treatment of uncertainty using ensemble methods: comparison of sequential data assimilation and Bayesian model averaging, *Water Resour. Res.*, 43, 1–15.
- Vrugt, J. A., Diks, C. G., and M. P. Clark (2008), Ensemble Bayesian model averaging using Markov chain Monte Carlo sampling, *Environmental fluid mechanics*, 8(5-6), 579-595.
- Weigel, A. P., Liniger, M. A. and C. Appenzeller (2008), Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts?, *Q.J.R. Meteorol. Soc.*, 134, 241–260. doi: 10.1002/qj.210
- Wood, A. W., and D. P. Lettenmaier (2008), An ensemble approach for attribution of hydrologic prediction uncertainty, *Geophysical Research Letters*, vol. 35, L14401, doi:10.1029/2008GL034648
- Ye, M., Neuman, S. P., and P. D. Meyer (2004), Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff, *Water Resour. Res.*, 40, W05113, doi:10.1029/2003WR002557.

Ye, M., Pohlmann, K.F., Chapman, J.B., Pohl, G.M., and D.M. Reeves (2010), A model-averaging method for assessing groundwater conceptual model uncertainty, *Ground Water* doi:10.1111/j.1745-6584.2009.00633.x

Ye, A., Q. Duan, X. Yuan, E. F. Wood, and J. Schaake (2014), Hydrologic post-processing of MOPEX streamflow simulations, *J. Hydrology*, doi:10.1016/j.jhydrol.2013.10.055.

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

### **Table captions**

**Table 1:** Specifications of ten study basins located in the Eastern United States

**Table 2:** List of the hydrologic models used in this study

**Table 3:** Forecast accuracy indicated by Mean Relative Absolute Error (MRAE) for the time period of 1988-1997 before and after the application of either model averaging technique

## Figure captions

**Figure 1:** Flowchart of BMA vs Cop-BMA.

**Figure 2:** Location of ten study basins in different climate regions across the United States.

**Figure 3:** Schematic of a) the predictive Q-Q plot (adapted from Laio and Tamea, 2007) and b) the corresponding pdfs to different interpretations. The solid, thick pdf is the actual pdf of observation.

**Figure 4:** Comparing the performance of BMA and Cop-BMA indicated by a) KGE (Eq. 16), and its components: b) correlation coefficient, c) bias, and d) variability ratio.

**Figure 5:** Comparison of predictive QQ plot produced by BMA vs Cop-BMA for all river basins.

**Figure 6:** Comparing the performance of BMA and Cop-BMA indicated by a-b) reliability (Eq. 17), c) sharpness (Eq. 18), and d) confidence (Eq. 19).

**Figure 7:** Comparing the weights of 7 hydrologic models for 12 months after the application of BMA and Cop-BMA for each basin.

Table 1: Specifications of ten study basins located in the Eastern United States

Basin #	USGS ID	Area (mi <sup>2</sup> )	Station Name	Station ID	Lon. °	Lat. °	Elev. (ft)	P/PE	E/PE	E/P	$\alpha^*$
1	5120205	598	East Fork White River at Columbus, IN	03364000	-85.93	39.20	603	1.21	0.77	0.64	0.85
2	2070009	970	Monocacy River at Jug Bridge near Frederick, MD	01643000	-77.37	39.40	231	1.15	0.76	0.66	0.91
3	2080104	1156	Rappahannock River near Fredericksburg, VA	01668000	-77.53	38.31	70	1.2	0.77	0.64	0.87
4	5020001	1375	Tygart Valley River at Philippi, WV	03054500	-80.04	39.15	1280	1.76	0.87	0.49	0.59
5	2070001	1480	South Branch Potomac River near Springfield, WV	01608500	-78.65	39.45	561	1.64	0.86	0.52	0.64
6	5050002	1680	Bluestone River near Pipestem, WV	03179000	-81.01	37.54	1527	1.5	0.83	0.55	0.69
7	7080209	1686	English River at Kalona, IA	05455500	-91.71	41.47	633	0.89	0.67	0.75	1.19
8	6010105	1879	French Broad River at Asheville, NC	03451500	-82.58	35.61	1950	2.34	0.96	0.41	0.47
9	8070202	1884	Amite River near Denham Springs, LA	07378500	-90.99	30.46	0	1.46	0.83	0.57	0.71
10	11070207	2590	Spring River near Waco, MO	07186000	-94.57	37.25	834	0.96	0.69	0.72	1.07

\* Aridity index, Eq. 13.

Table 2: List of the hydrologic models used in this study

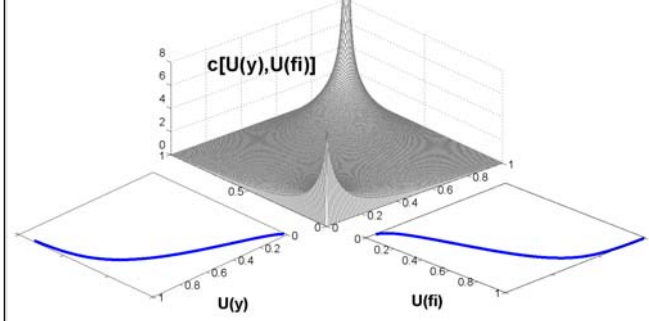
Model ID	Full Name/Developer	Developer
SAC (Burnash et al., 1973)	Sacramento Soil Moisture Accounting	National Weather Service (NWS), USA
GR4J (Perrin et al., 2003)	Génie Rural à 4 paramètres Journalier	Cemagref, France
SWB (Schaake et al., 1996)	Simple Water Balance	National Weather Service (NWS), USA
ISBA (Noilhan and Planton, 1989)	Interactions between Soil-Biosphere-Atmosphere	Météo France, France
NOAH (Chen et al., 1996)	NOAH model	National Weather Service (NWS), USA
SWAP (Gusev and Nasonova, 1998)	Soil Water Atmosphere Plant	Russian Academy of Sciences, Russia
VIC (Liang et al., 1994)	Variable Infiltration Capacity	University of Washington/Princeton University, USA

Table 3: Forecast accuracy indicated by Mean Relative Absolute Error (MRAE) for the time period of 1988-1997 before and after the application of either model averaging technique.

Basin #	Single Model							Model Averaging	
	<i>SAC</i>	<i>GR4J</i>	<i>SWB</i>	<i>ISBA</i>	<i>NOAH</i>	<i>SWAP</i>	<i>VIC</i>	<i>BMA</i>	<i>Cop-BMA</i>
1	0.67	0.46	0.71	0.58	0.54	0.74	0.74	0.56	<b>0.32</b>
2	0.58	0.49	0.47	0.54	0.56	0.83	0.63	0.42	<b>0.32</b>
3	0.52	0.45	0.50	0.46	0.65	0.63	0.66	0.40	<b>0.31</b>
4	0.93	1.87	1.18	1.82	1.83	3.06	2.00	0.64	<b>0.42</b>
5	0.84	0.73	0.66	0.65	0.70	0.73	0.83	0.51	<b>0.36</b>
6	0.65	1.10	0.62	0.76	0.57	1.17	1.09	0.52	<b>0.32</b>
7	2.03	3.79	2.61	1.92	2.30	3.92	2.38	1.49	<b>1.02</b>
8	0.32	0.16	0.32	0.27	0.20	0.34	0.26	0.36	<b>0.15</b>
9	0.61	0.48	0.43	0.46	0.40	0.63	0.88	0.51	<b>0.24</b>
10	1.13	1.18	1.09	0.96	1.11	1.17	1.32	0.81	<b>0.47</b>

### Cop-BMA

Joining the marginal distribution of biased/unbiased forecast of the  $i^{\text{th}}$  model ( $f_i$ ) and the marginal distribution of forecast variable ( $y$ ) via the PDF of copula.



$$p(y | f_i, Y) = c(u_y, u_{f_i}) \cdot p(y)$$

EM algorithm

$$p(y, f_1, \dots, f_k, Y) = \sum_{i=1}^k w_i \cdot p(y | f_i, Y)$$

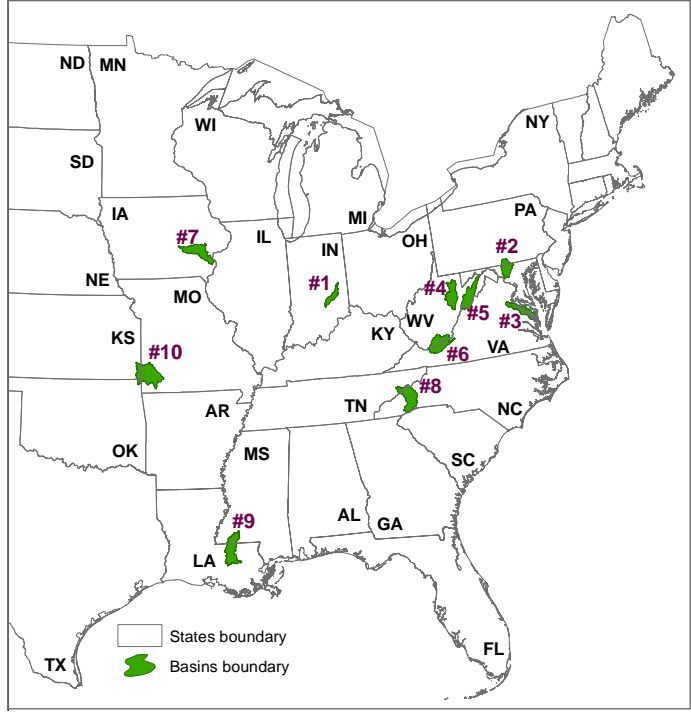
### BMA

Bias correction  
of model predictions

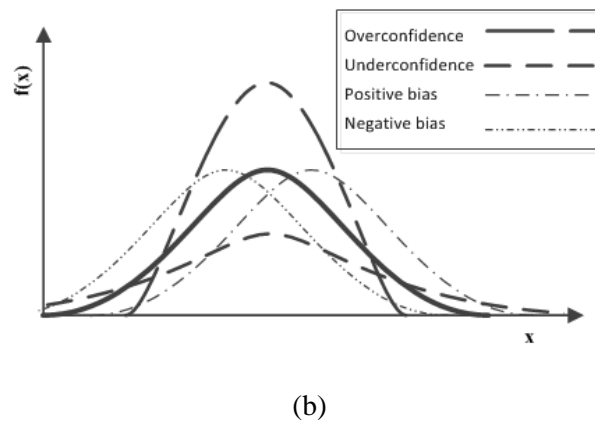
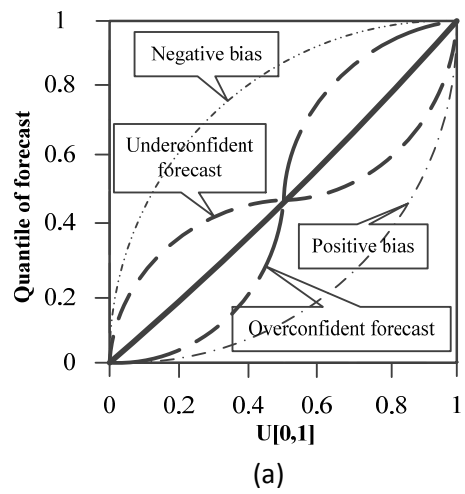
Data transformation  
from the original space to the particular  
space of posterior distribution.

EM algorithm

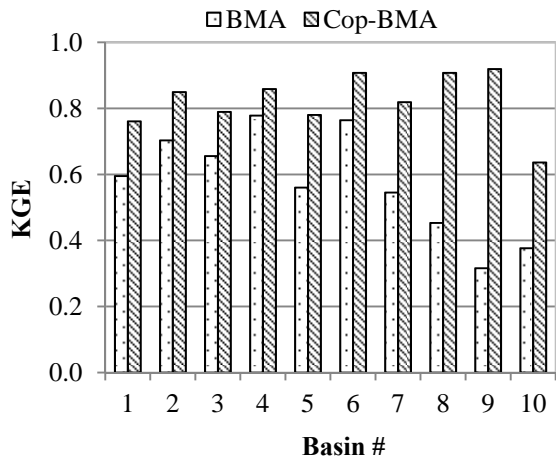
$$p(y | f_i, Y) = g(y | f_i, \sigma_i)$$



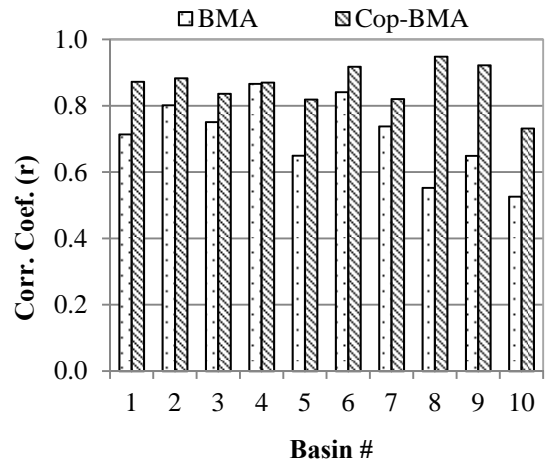




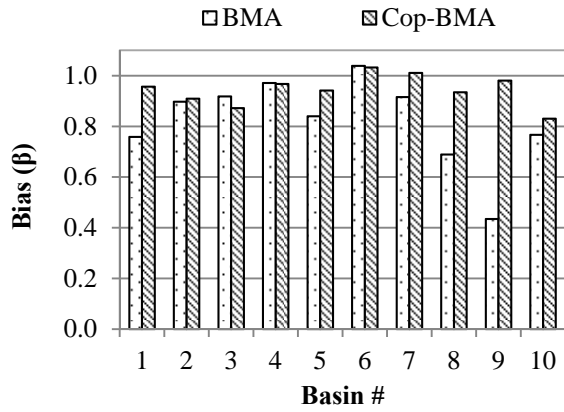
(a)



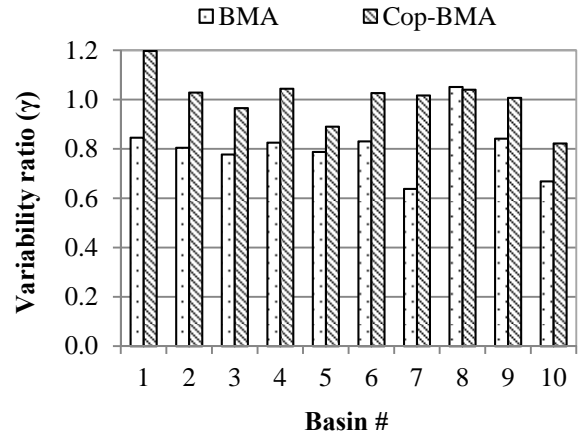
(b)

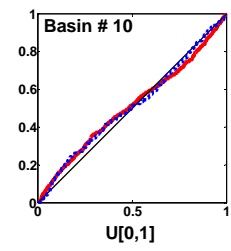
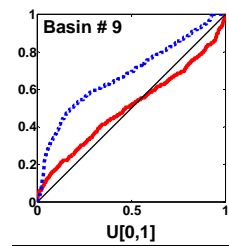
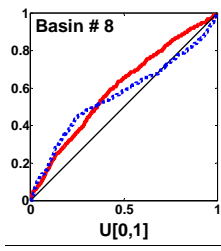
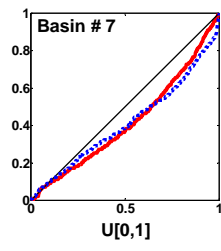
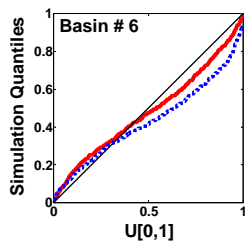
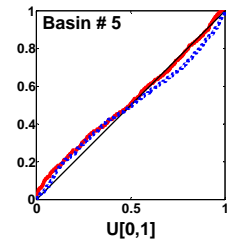
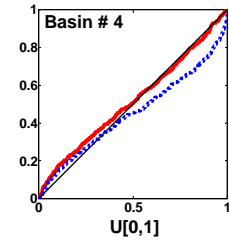
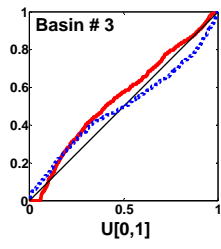
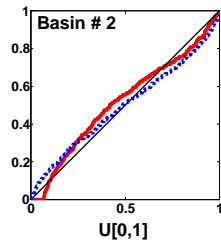
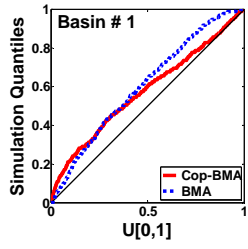


(c)

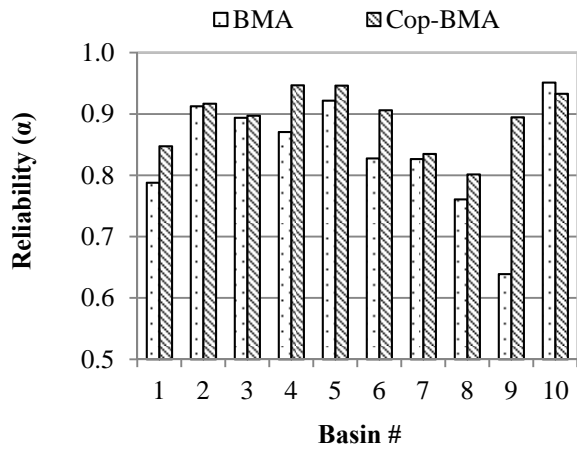


(d)

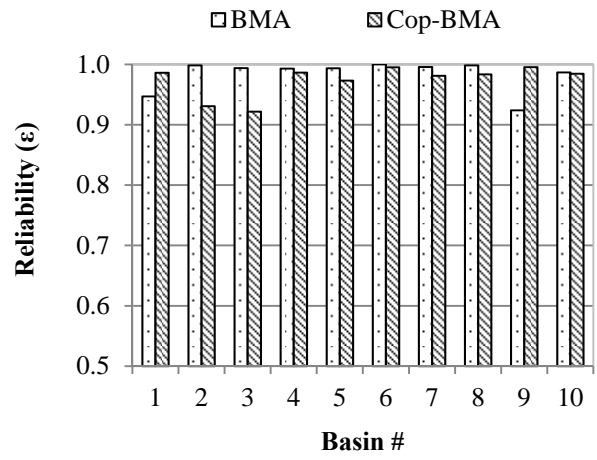




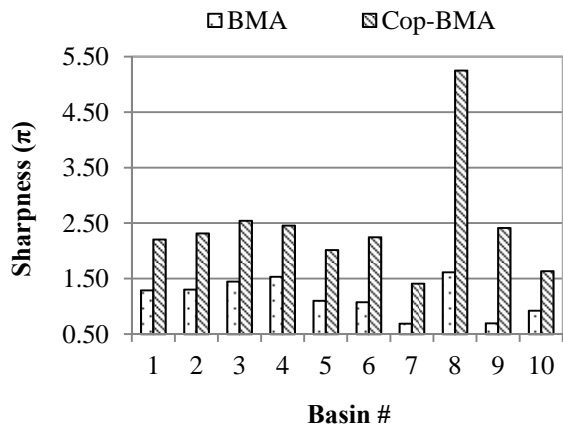
(a)



(b)



(c)



(d)

