

Uncertainty analysis of hydrological multi-model ensembles based on CBP-BMA method

Shaokun He, Shenglian Guo, Zhangjun Liu, Jiabo Yin, Keping Chen and Xushu Wu

ABSTRACT

Quantification of the inherent uncertainty in hydrologic forecasting is essential for flood control and water resources management. The existing approaches, such as Bayesian model averaging (BMA), hydrologic uncertainty processor (HUP), copula-BMA (CBMA), aim at developing reliable probabilistic forecasts to characterize the uncertainty induced by model structures. In the probability forecast framework, these approaches either assume the probability density function (PDF) to follow a certain distribution, or are unable to reduce bias effectively for complex hydrological forecasts. To overcome these limitations, a copula Bayesian processor associated with BMA (CBP-BMA) method is proposed with ensemble lumped hydrological models. Comparing with the BMA and CBMA methods, the CBP-BMA method relaxes any assumption on the distribution of conditional PDFs. Several evaluation criteria, such as containing ratio, average bandwidth and average deviation amplitude of probabilistic application, are utilized to evaluate the model performance. The case study results demonstrate that the CBP-BMA method can improve hydrological forecasting precision with higher cover ratios more than 90%, which are increased by 4.4% and 3.2%, 2.2% and 1.7% over those of BMA and CBMA during the calibration and validation periods, respectively. The proposed CBP-BMA method provides an alternative approach for uncertainty estimation of hydrological multi-model forecasts.

Key words | CBP-BMA method, hydrologic forecasting, multi-model ensemble, probabilistic distribution, uncertainty estimation

Shaokun He
Shenglian Guo (corresponding author)
Zhangjun Liu
Jiabo Yin
Keping Chen
Xushu Wu

State Key Laboratory of Water Resources and
Hydropower Engineering Science,
Wuhan University,
Wuhan 430072,
China
E-mail: slguo@whu.edu.cn

INTRODUCTION

Hydrological forecasting is an essential endeavor in water resources management and optimal control of rivers, such as, flood control, reservoir operation, navigation, and leisure activities (Montanari & Brath 2004; Fan *et al.* 2016; Liu *et al.* 2017a, 2017b; Wu *et al.* 2017). Despite the tremendous amount of resources invested in developing more hydrologic models, no one can convincingly claim that any particular model in existence today is superior to other models for all type of applications and under all conditions (Wu *et al.* 2015; Liu *et al.* 2016; Ba *et al.* 2017). Different models are capable of capturing different aspects of the

hydrologic processes. The uncertainty of each model arises from parameters, calibration, design of model structure, and input measurements, which partially bring underlying imprecise influence (Götzinger & Bárdossy 2008; Li *et al.* 2011; Hemri *et al.* 2015). 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; Han *et al.* 2014).

In recent years, some techniques develop the probability distribution of forecast to account for different sources of

uncertainty. The Bayesian forecasting system proposed by Krzysztofowicz (1999) provides a general framework for hydrologic probability forecasting. Krzysztofowicz & Kelly (2000) developed a meta-Gaussian hydrologic uncertainty processor (HUP), which converts the meta-Gaussian distribution to Gaussian distribution by normal quantile transform. Due to the advantages of non-normal distribution assumption, the meta-Gaussian HUP has been widely implemented in hydrologic and meteorological fields (Chen *et al.* 2013).

Aside from traditional HUP, the Bayesian model average (BMA) method introduced by Raftery *et al.* (2005) follows a statistical technique to combine the advantages of different models. Different from other multi-model methods, the BMA method presents a more realistic description of predictive uncertainty, since the BMA predictive variance can be broken down into two components: between-model variability and within-model variability (Ajami *et al.* 2007). The BMA method is a statistical procedure that infers consensus predictions by weighing individual predictions based on their probabilistic likelihood measures, with the better performing predictions receiving higher weights than the worse performing ones. The method has been explored to improve both the accuracy and reliability of stream flow predictions (Vrugt & Robinson 2007; Liang *et al.* 2011). Duan *et al.* (2007) concluded that the combination of multi-model ensemble strategies using the BMA framework could quantify statements on prediction uncertainty and significantly improve verification performances. Zhou *et al.* (2016) compared the mean prediction of BMA with its individual parameter transfer method (physical similarity approach) and demonstrated that the probabilistic predictions of BMA could reduce the uncertainty with a significant degree. Nevertheless, the standard BMA method imposes a large number of pseudo-variation requirements and this influences precise understanding of data variations, which gives rise to further development on this theoretical research (Madadgar & Moradkhani 2014).

Copula functions are versatile and promising tools for characterizing the dependence structures in hydrologic processes as it allows modeling the dependence structure among random variables independently of the marginal distributions, which have been successfully applied in

probabilistic analysis in recent years (Lee & Salas 2011; Li *et al.* 2013; Chen *et al.* 2015; Liu *et al.* 2015; Liu *et al.* 2017a, 2017b). Klein *et al.* (2016) used a mixture of marginal density distribution to estimate predictive uncertainty of hydrologic multi-model ensembles by using pair-copula construction. Similar research shows that copula technique is an effective tool for reflecting the unclear and complex relationships, because it can flexibly choose an arbitrary type for the marginal distributions instead of Gaussian distribution (Carreau & Bouvier 2016; Khajehei & Moradkhani 2017). According to the promising results of using copula functions in post-processing of hydrologic forecasts, Madadgar & Moradkhani (2014) firstly integrated copula functions with BMA to estimate the posterior distribution and found that copula-BMA (CBMA) is an effective post-processor to relax any assumption on the distribution of conditional probability density function (PDF). The CBMA not only displayed better deterministic skill than BMA but also confirmed the impact of posterior distribution in calculating the weights of individual models by EM algorithm. Results indicated that the predictive distributions are more accurate and reliable. It is also shown that the post-processed forecasts have better correlation with observations after CBMA application. The CBMA method in the meteorological application does not need to assume the shape of posterior distribution and leaves out the data-transformation procedure and demonstrates that predictive distributions are less bias and more confident with small uncertainty (Möller *et al.* 2013).

Inspired by the ideas of Madadgar & Moradkhani (2014), a general framework of the combination of copula Bayesian processor with BMA (CBP-BMA) is proposed in this study, where the Bayesian theory is applied in the transformation of posterior distribution. The flowchart of different probability forecast methods based on deterministic models is described in Figure 1. The remainder of the paper describes the case study and three hydrological models in the next section. This is followed by illustrations of concepts of three different methods of probability forecast as well as the evaluation criteria applied for multi-model techniques. Results and discussions of the case study are then given, and finally a summary of the conclusions drawn.

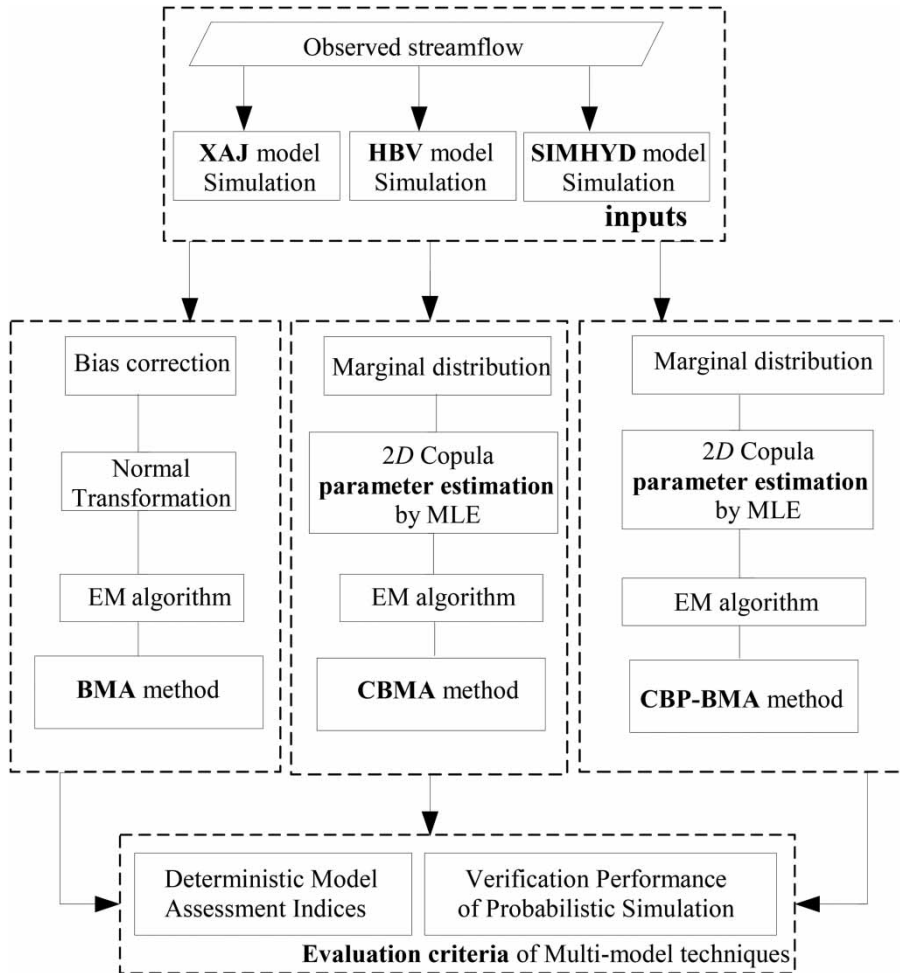


Figure 1 | Flowchart of hydrologic multi-model ensembles for uncertainty analysis.

MATERIALS AND HYDROLOGICAL MODELS

Study area and data

The proposed methods are applied to daily mean flow discharge at the Mumahu catchment (Figure 2), a sub-basin of Hanjiang River Basin in China. The catchment lies in Shanxi Province with an area of 1,224 km² and locates in the subtropical monsoon region with humid climate and fairly heavy precipitation. The annual mean precipitation and runoff is 1,070 mm and 687 mm, respectively. The available dataset contains daily precipitation, runoff and evaporation with a length of 11 years (1980–1990). The

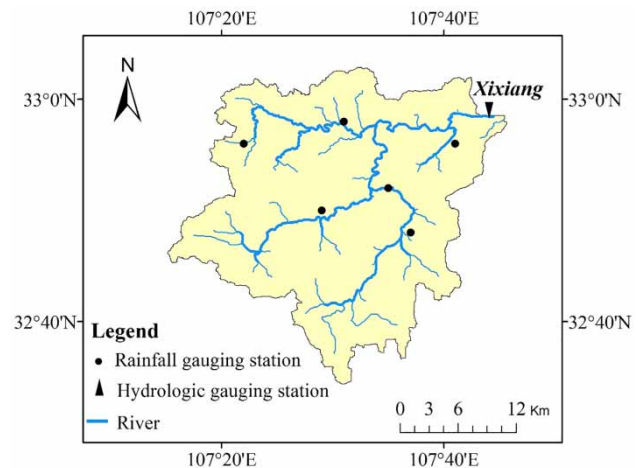


Figure 2 | Sketch map of the Mumahu catchment.

first year (1980) is used as the spin-up period for each hydrological model in order to achieve the best effective model formulation. The remaining years (1981–1990) are divided into two sub-periods, with 7 years (1981–1987) for calibration and 3 years (1988–1990) for validation.

Description of the hydrological models

Three well established conceptual hydrological models are implemented in the Mumahe catchment, including the Xinanjiang (XAJ), HBV and SIMHYD models. The XAJ model has been used in humid and semi-humid regions worldwide (Zhao 1992). It consists of a runoff generation component with seven parameters and a routing component with ten parameters. These physical parameters within the model represent abstract conceptual expressions of watershed features. The HBV model is a synthetic flow model with 13 parameters which need to be calibrated. Units of the HBV model form the routines for snowmelt accumulation, evapotranspiration and soil routine and response function. The core concept assumes runoff volume changes with soil humidity exponentially (Montero et al. 2016). The SIMHYD model is a lumped conceptual hydrological model which contains seven parameters requiring calibration. This model divides runoff into three components: surface flow, interflow and base flow. The surface flow is infiltration excess runoff, inter-flow is estimated as a linear function of the soil wetness, and base flow is simulated as a linear recession from the groundwater store (Chiew et al. 2009; Yu & Zhu 2015). The infiltration rate is a core of the model.

METHODOLOGY

Bayesian model averaging

Raftery et al. (2005) successfully extended BMA to statistical post-processing for forecast ensembles. The BMA method addresses total model uncertainty by conditioning not only on a single outstanding model, but on the entire ensemble models. The method was originally proposed as a pathway for method combination of several competing models (Duan et al. 2007; Liang et al. 2011).

The BMA method

According to BMA (Duan et al. 2007), the ensemble predictive density of the actual flow variable q , given the different hydrologic model simulations of K models $[S_1, S_2, \dots, S_K]$ and the observations during the training period, Q , can be expressed in terms of the law of total probability:

$$p(q|S_1, S_2, \dots, S_K, Q) = \sum_{i=1}^K p(S_i|Q) \cdot p_i(q|S_i, Q) \quad (1)$$

where $p(S_i|Q)$ is the posterior probability of i th model prediction. This static term can also be expressed as ω_i , reflecting how well the ensemble term fits the observation dataset, it ranges from 0 to 1 since the posterior model probabilities add up to one. Prior to the implantation of BMA algorithm, the expected value of observation and forecast for each model should be equal to zero ($E[q-S_i] = 0$). Any bias-correction method, such as linear regression, should be applied to substitute the bias-corrected forecast (f_i) for original deterministic forecast:

$$f_i = a_i + b_i \cdot S_i \quad (2)$$

where $\{a_i, b_i\}$ are the coefficients of linear regression model.

The term $p_i(q|f_i, Q)$ is the conditional pdf (probability density function) of h based on the bias-corrected simulation f_i and the observation dataset. Moreover, the power Box-Cox transformation is taken for the computational convenience of using a Gaussian distribution. The posterior distribution $p_i(q|f_i, Q)$ is mapped to a Gaussian space with mean f_i and variance s_i^2 ; i.e., $p_i(q|f_i, Q) \sim g(q|f_i, \sigma_i^2)$. The BMA predictive mean and variance of q are defined as follows (Raftery et al. 2005):

$$E(q|Q) = \sum_{i=1}^K p(f_i|Q) \cdot E[p_i(q|f_i, Q)] = \sum_{i=1}^K \omega_i f_i \quad (3)$$

$$\text{Var}(q|Q) = \sum_{i=1}^K \omega_i \left(f_i - \sum_{i=1}^K \omega_i f_i \right)^2 + \sum_{i=1}^K \omega_i \sigma_i^2 \quad (4)$$

Estimation of probabilistic prediction

Successful application of the BMA method requires estimations of the weight ω_i and variance σ_i^2 of the individual pdf. The log maximum likelihood function, rather than the likelihood function, is optimized for reasons of both numerical stability and algebraic simplicity. If the BMA parameters are estimated by $\theta = \{\omega_i, \sigma_i, i = 1, 2, \dots, K\}$, the log likelihood function of θ is mathematically denoted as:

$$l(\theta) = \log \left(\sum_{i=1}^K \omega_i \cdot p_i(q|f_i, Q) \right) \quad (5)$$

After the completion of BMA parameter estimation by the EM algorithm (Duan et al. 2007), another feature of the BMA method is to make use of the Monte Carlo method to derive BMA probabilistic ensemble prediction for any time t (Kuczera & Parent 1998). The procedures are described as follows (Zhou et al. 2016).

1. Select the probabilistic ensemble size, M ($M = 100$).
2. Randomly generate a value of k from the numbers $[1, 2 \dots K]$ with probabilities $[\omega_1, \omega_2, \dots, \omega_i]$ using:
 - (a) initial the cumulative weight $\omega'_0 = 0$ and compute $\omega'_i = \omega'_{i-1} + \omega_i$ for $i = 1, 2, \dots, K$;
 - (b) generate a random number u between 0 and 1;
 - (c) if $\omega'_{i-1} \leq u \leq \omega'_i$, then the i th member of the ensemble predictions are chosen.
3. Generate a value of q from the pdf of $p_i(q|f_i, \sigma_i^2)$.
4. Repeat steps (2) and (3) for M times.

The results are sorted in ascending order and the 90% confidence interval can be derived within the range of the 5% and 95% quantiles.

The hybrid CBMA

As illustrated before, the BMA predictive distribution provides a weighted average of simulation pdf which generally complies with a parametric distribution, e.g., Gaussian distribution after the Box-Cox transformation. Madadgar & Moradkhani (2014) employed copula to estimate the posterior distribution of forecast variables for each model, i.e., $p_i(q|f_i, Q)$, and found that the

hydrological forecasts are improved after the integration of CBMA. A series of research demonstrates that the procedures of CBMA not only eliminate the prophase bias-correction and the external calculation of variance, but also simplify the calculation of the weighted average and the probability model structure by copula (Möller et al. 2013).

Copula theory

Following Sklar's theorem (Sklar 1959), the mathematical expression $F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ of the joint multivariate distribution can be uniquely determined by respective univariate marginal probability distributions $F_{X_1}(x_1)$, $F_{X_2}(x_2)$, \dots , $F_{X_n}(x_n)$ of n correlated random variables with the help of the copula function C .

$$\begin{aligned} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= C[F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)] \\ &= C(u_1, u_2, \dots, u_n) \end{aligned} \quad (6)$$

where $u_i \in [0, 1]$ corresponds to the i th uniformly distributed variable transformed from x_i using $F_{X_i}(x_i)$. If F is absolutely continuous, its pdf can be also written via the copula density c as follows:

$$\begin{aligned} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= \frac{\partial^n C(u_1, u_2, \dots, u_n)}{\partial u_1 \partial u_2 \dots \partial u_n} \prod_{i=1}^n f_{X_i}(x_i) \\ &= c(u_1, u_2, \dots, u_n) \prod_{i=1}^n f_{X_i}(x_i) \end{aligned} \quad (7)$$

Alternatively, in statistical applications, the conditional probability distribution of h given s_i ($i = 1, 2, 3$) is expressed as (Madadgar & Moradkhani 2014):

$$f(q|s_i) = \frac{f(q, s_i)}{f(s_i)} = \frac{c(u, v_i) \cdot f(q) \cdot f(s_i)}{f(s_i)} = c(u, v_i) \cdot f(q) \quad (8)$$

where $c(u, v_i)$ is computed for each pair of (u, v_i) , $f(q)$ represents the marginal distribution of actual flow. Despite different copula families having been proposed and described in current studies (Chebana & Ouarda 2007),

several families of Archimedean copulas, including Frank, Gumbel and Clayton, have been popular choices for dependence models in hydrologic analyses due to their simplicity and generation properties.

The CBMA method

The predictive distribution of CBMA is modified as follows (Madadgar & Moradkhani 2014):

$$f(q|s_1, s_2, \dots, s_K) = \sum_{i=1}^K \omega_i f(q|s_i) = \sum_{i=1}^K \omega_i \cdot c(u, v_i) \cdot f(q) \quad (9)$$

It can be seen from Equation (9) that it relaxes any assumption on the type of posterior distribution $f(q|s_i)$, whose term can be directly inferred with the help of copula functions. Once the term $f(q|s_i)$ is defined, their weights are estimated by the EM algorithm with a few adjustments, which can refer to Madadgar & Moradkhani (2014) in detail.

The hybrid CBMA model applies the idea of 'pair and ensemble'. The pair of observation q and the i th model simulation is established to acquire the probability task by the well developed copula theory, while the ensemble is to formulate a consensus probability interval.

Copula Bayesian processor associated with BMA (CBP-BMA) method

Copula Bayesian processor

Copula Bayesian processor (CBP) is developed as another component of the probabilistic forecasting system in virtue of the integration of Bayesian theory and copula functions. The CBP procedure generates probabilistic result and quantifies the hydrologic uncertainty under the assumption that input uncertainty is ignored, which refers to HUP (Krzysztofowicz & Kelly 2000). This method also has the advantage of leaving out a data transformation procedure into Gaussian space. The Bayesian procedure based on the law of total probability involves two parts for information revision of uncertainty (Zhang & Singh 2007). (1) The expected conditional density

function of deterministic simulation, S_i given $Q = q$ is expressed as:

$$\kappa(s_i|q) = \int f(s_i|q) \cdot g(q) dq \quad (10)$$

where $f(q|s_i)$ has the same conception as before, $g(q)$ represents the prior density function. (2) The posterior density function conditional on a deterministic result $S_i = s_i$ is derived via Bayes' theorem:

$$\phi(q|s_i) = \frac{f(s_i|q) \cdot g(q)}{\kappa(s_i|q)} \quad (11)$$

Equations (10) and (11) could be rewritten by means of copula functions. The CBP form of right term of Equation (12) is mathematically expressed by:

$$\phi(q|s_i) = \frac{f(s_i|q) \cdot g(q)}{\int f(s_i|q) \cdot g(q) dq} = \frac{c(u, v_i)}{\int_0^1 c(u, v_i) du} \cdot g(q) \quad (12)$$

The final CBP outputs a posterior distribution of the process, conditional upon the deterministic simulation. Since the analytical solution to the integral term $\int_0^1 c(u, v_i) du$ is very complex, the Monte Carlo technique is used to estimate the posterior density function $\phi(q|s_i)$ (Robert & Casella 2001; Kroese et al. 2011).

The CBP-BMA method

The difference between the CBP-BMA and CBMA methods is the estimation procedure of the posterior density function.

$$\phi(q|s_1, s_2, \dots, s_K) = \sum_{i=1}^K \omega_i \phi(q|s_i) = \sum_{i=1}^K \omega_i \frac{c(u, v_i)}{\int_0^1 c(u, v_i) du} \cdot g(q) \quad (13)$$

It should be rational to assign weights on account of multiple deterministic results. The calculation process of weights is conducted by the EM algorithm. The three main steps of the presented weights calculating paradigm can be

summarized as:

$$\begin{aligned}
 w_i^{Iter} &= \frac{1}{T} \sum_{t=1}^T z_{i,t}^{Iter} \\
 z_{i,t}^{Iter} &= \frac{w_i^{Iter-1} \cdot \phi(q_t | s_{i,t})}{\sum_{i=1}^K w_i^{Iter-1} \cdot \phi(q_t | s_{i,t})} \\
 &= \frac{w_i^{Iter-1} \cdot c(u_t, v_{i,t})g(q_t) / \int_0^1 c(u_i, v_{i,t})du_t}{\sum_{i=1}^K w_i^{Iter-1} \cdot c(u_t, v_{i,t})g(q_t) / \int_0^1 c(u_i, v_{i,t})du_t} \\
 l(\theta_{Iter}) &= \log \left(\sum_{i=1}^K w_i^{Iter-1} \cdot \sum_{t=1}^T c(u_t, v_{i,t})g(q_t) / \int_0^1 c(u_i, v_{i,t})du_t \right) \tag{14}
 \end{aligned}$$

where T is the length of the training period; and z is a latent variable. Compared with the standard BMA method, the calculation of variance and data transformations are eliminated in Equation (14). The posterior probability of q_t is calculated only once while it need be re-calculated every time in the standard BMA method.

Evaluation criteria for multi-model techniques

Deterministic model assessment indices

To evaluate the quality of the deterministic model, three metrics are introduced as follows.

1. Nash–Sutcliffe efficiency coefficient (*NSE*)

$$NSE = 1 - \frac{\sum_{i=1}^N (q_o^i - q_m^i)^2}{\sum_{i=1}^N (q_o^i - \bar{q}_o)^2} \tag{15}$$

where N denotes the length of data sequence. q_o^i and q_m^i are the observed and the simulated, respectively, and \bar{q}_o represents the average of the observed. *NSE* has a range of $(-\infty, 1]$ with the best *NSE* equal to 1 (Nash & Sutcliffe 1970).

2. Daily root mean square error (*DRMS*)

$$DRMS = \sqrt{\frac{\sum_{i=1}^N (q_o^i - q_m^i)^2}{T}} \tag{16}$$

As the second tool employed is sensitive to the differences between observations and simulations, the values of *DRMS* approaching 0 stand for better performance.

3. Kling-Gupta efficiency (*KGE*)

$$\begin{aligned}
 KGE &= 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \\
 \beta &= \bar{q}_m / \bar{q}_o \\
 \gamma &= CV_m / CV_o = \frac{\sigma_m / \bar{q}_m}{\sigma_o / \bar{q}_o} \tag{17}
 \end{aligned}$$

where r is the Pearson correlation between the observation and simulation, β is the bias ratio indicator; γ is the variability ratio (Kling et al. 2012). All calculative variables are replaced by the expected values of the estimate predictive distributions.

Verification of probabilistic simulations

With regard to assessment of assessing the uncertainty analysis of simulation interval, Xiong et al. (2009) and Dong et al. (2013) presented multiple verification indices and applied in hydrologic practice. Three main metrics are selected to evaluate the simulation uncertainty intervals generated by the BMA, CBMA and CBP-BMA methods.

1. Containing ratio (*CR*)

The containing ratio is utilized as a significant index for assessing the goodness of the uncertainty interval. It is defined as the percentage of observed data points that fall between the prediction bounds, directly reflecting the interval performance.

$$CR = \frac{\sum_{i=1}^N C(q_l^i \leq q_o^i \leq q_u^i)}{N} \times 100\% \tag{18}$$

where q_l^i is denoted as the lower bound corresponding to 5% quantile at time t , q_u^i is denoted as the upper bound corresponding to 95% of the quantile. And $\sum_{i=1}^N C$ is the number of the observed data points q_o^i that satisfy the inequality conditions.

2. Average bandwidth (*B*)

$$B = \frac{1}{N} \sum_{i=1}^N (q_u^i - q_l^i) \tag{19}$$

where B is also an index measuring the average width of estimated uncertainty interval just as the definition name

indicates. Smaller values of B show a greater precision. Consider two forecasts with the same containing ratio; the situation with smaller B is preferred, because it has less uncertainty or greater precision.

3. Average deviation amplitude (D)

The average deviation amplitude D is an index to quantify the average deflection of the curve of the middle points of the prediction bounds from the observed streamflow hydrograph. It is defined as

$$D = \frac{1}{N} \sum_{i=1}^N \left| \frac{1}{2}(q_u^i + q_l^i) - q_o^i \right| \quad (20)$$

where the notations are defined previously.

RESULTS AND DISCUSSION

Different multi-model techniques, i.e., BMA, CBMA and CBP-BMA, are applied to combine the ensemble streamflow simulation. The structures of three hydrologic models ought to be determined as the deterministic results are crucial to the final uncertainty analysis. As mentioned above, the calibration parameters of the first BMA method are ω_k and σ_k^2 . In the CBMA method, they are the parameters of marginal distributions, weights ω_k and the parameters of the pdf of copula. In the CBP-BMA method, the Monte Carlo sampling technique is also used to obtain the integral item.

Deterministic hydrologic model simulations

The genetic and simplex algorithms are used for model calibration on account of their flexibility and good convergence. The genetic algorithm can acquire the global optimal value independent of initial parameter values. The simplex algorithm is of high accuracy with low convergence rate. With the merit of two methods integrated, the approximately optimal values of model parameters are obtained. Three deterministic assessment indices, NSE , $DRMS$ and KGE scores over the calibration period (1981–1987) and the validation period (1988–1990) are calculated for XAJ, HBV and SIMHYD models. Table 1 indicates that the XAJ model has the best results, the HBV model takes second place, and the

Table 1 | Deterministic accuracy assessment of different hydrological models

Model	Calibration			Validation		
	NSE (%)	$DRMS$	KGE (%)	NSE (%)	$DRMS$	KGE (%)
XAJ	88.25	30.06	90.59	84.85	24.06	87.08
HBV	84.81	34.16	52.99	82.24	26.06	42.89
SIMHYD	86.25	32.50	82.51	84.98	23.97	85.01
BMA	88.87	33.79	89.14	85.72	24.81	90.62
CBMA	88.93	26.54	90.06	86.07	21.25	88.45
CBP-BMA	89.76	27.63	90.96	86.69	23.39	89.23

Note: values in bold represent the optimal result.

SIMHYD model behaves worst among the three. The reason can be attributed to the dissimilar process for the calibration of each model (Nasonova et al. 2009). In practice, it might partially refer to inaccurate estimation of model parameters as one of error sources of the model structure, abstract formulation of physical processes, and different sources of forcing dataset for each model. In general, these simulation results can be used as the input data of multi-model ensemble in terms of the values of NSE and KGE , which are 85% and higher than 0.82, respectively, apart from the poor value of KGE of HBV.

Determination of the marginal distributions

The marginal distributions of the random variables of H and S_i ($i = 1, 2, 3$) need to be determined (Li et al. 2017). Five common candidate distributions (in Table 2) have been fitted to the daily mean streamflow values as well as to the XAJ, HBV and SIMHYD model simulations.

With regard to random variable H , the parameters of five candidate distributions were estimated by the method of L -moment (Jonathan 1990; Chen et al. 2017; Li et al. 2017) and the parameter values are listed in Table 3. The K-S tests are used to verify the null hypothesis and the corresponding statistic D_{KS} values are also listed in Table 3. It is shown that the null hypothesis could not be rejected at the 95% confidence level (threshold value $D_n = 1.36/\sqrt{N}$, N is the number of sampling points) for Log Normal distributions with providing the minimum D_{KS} value. Meanwhile, Figure 3 indicates the Log Normal is satisfactory on visual inspection and that the cumulative

Table 2 | Candidate univariate distributions adopted in this study

Distribution	PDF	Range
Gumbel	$f(x) = \alpha \exp[-\alpha(x - \mu) - e^{-\alpha(x - \mu)}]$	$-\infty < x < +\infty$
Gamma	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$	$x > 0$
Pearson type III	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} (x - \mu)^{\alpha-1} \exp[-\beta(x - \mu)]$	$x > \mu$
Log Normal	$f(x) = \frac{1}{x\sqrt{2\pi}\sigma_Y} \exp\left[-\frac{(\ln x - \mu_Y)^2}{2\sigma_Y^2}\right], Y = \ln X$	$x > 0$
Normal	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right]$	$-\infty < x < +\infty$

Table 3 | Parameter estimations and K-S statistic D_{KS} test of five candidate marginal distributions for variables

Variable	H	S ₁	S ₂	S ₃	
Gumbel	α	40.3	40.6	39.5	37.2
	μ	17.8	14.9	15.1	14.5
	D_{KS}	0.025	0.024	0.026	0.026
Gamma	α	0.4	0.3	0.3	0.3
	β	103.4	129.7	153.2	106.5
	D_{KS}	0.148	0.122	0.141	0.153
Pearson type III	α	0.20	0.24	0.20	0.19
	β	0.0056	0.0067	0.0050	0.0063
	μ	6.02	4.43	5.27	4.59
	D_{KS}	0.034	0.028	0.029	0.031
Log Normal	μ_Y	2.83	2.11	2.43	2.57
	σ_Y	1.17	1.46	1.28	1.26
	D_{KS}	0.016	0.013	0.017	0.015
Normal	μ	41.08	38.27	38.43	34.69
	σ	49.57	49.85	41.19	43.03
	D_{KS}	0.074	0.062	0.075	0.064

Note: values in bold denote that the distribution model passes the goodness-of-fit test at 0.05 significance level.

distribution function (cdf) plots of the theoretical Log Normal distributions fitted the empirical cdf values obtained from the Gringorten plotting-position formula relatively well. The estimation of marginal distributions for S_i had similar procedures. The Kolmogorov-Smirnov statistics D_{KS} indicate that the Log Normal distribution also gives the best fit in this study.

Archimedes copula selection and estimation

In the application of the CBMA and CBP-BMA methods, a copula function to link the cdf of observation and model

simulations needs to be defined. Different families of copula have been proposed and described, such as Archimedes copula, T-copula, pair-copula. The Archimedes copula family is more desirable for hydrologic analyses, because most of the family can be easily constructed and some of the family can be applied to describe the positive correlation of variables exactly reflecting the relationship among observation and model simulations (Madadgar & Moradkhani 2014; Chen et al. 2015). In consequence, the Gumbel, Clayton and Frank copula belonging to Archimedes family are chosen to test for flexibility and universality.

For Archimedes copula, the Kendall correlation coefficient τ_i ($i = 1, 2, 3$) between observed and different simulated flows is firstly derived. The higher τ_i indicator reflects the stronger correlation of observation and model simulation. The corresponding copula parameter θ_i is calculated by the method based on the inversion of τ_i in Table 4. The parameter estimators and goodness-of-fit test (*RMSE* and *AIC*) are used to determine the best-fit copula for integrating the streamflow properties. The results illustrate that copulas have the good performance in exploring the associations of observed and simulated flows. All variables passed the null hypothesis for Gumbel and Frank copulas. Gumbel copula performs with the lowest *RMSE* and *AIC* values.

Deterministic assessment of three ensemble methods

Figure 4 displays the weight estimates of individual models. These weights range from 0.31 to 0.38, which reveals the three approximately occupy an equal proportion. We check the mean simulation of hydrologic multi-model

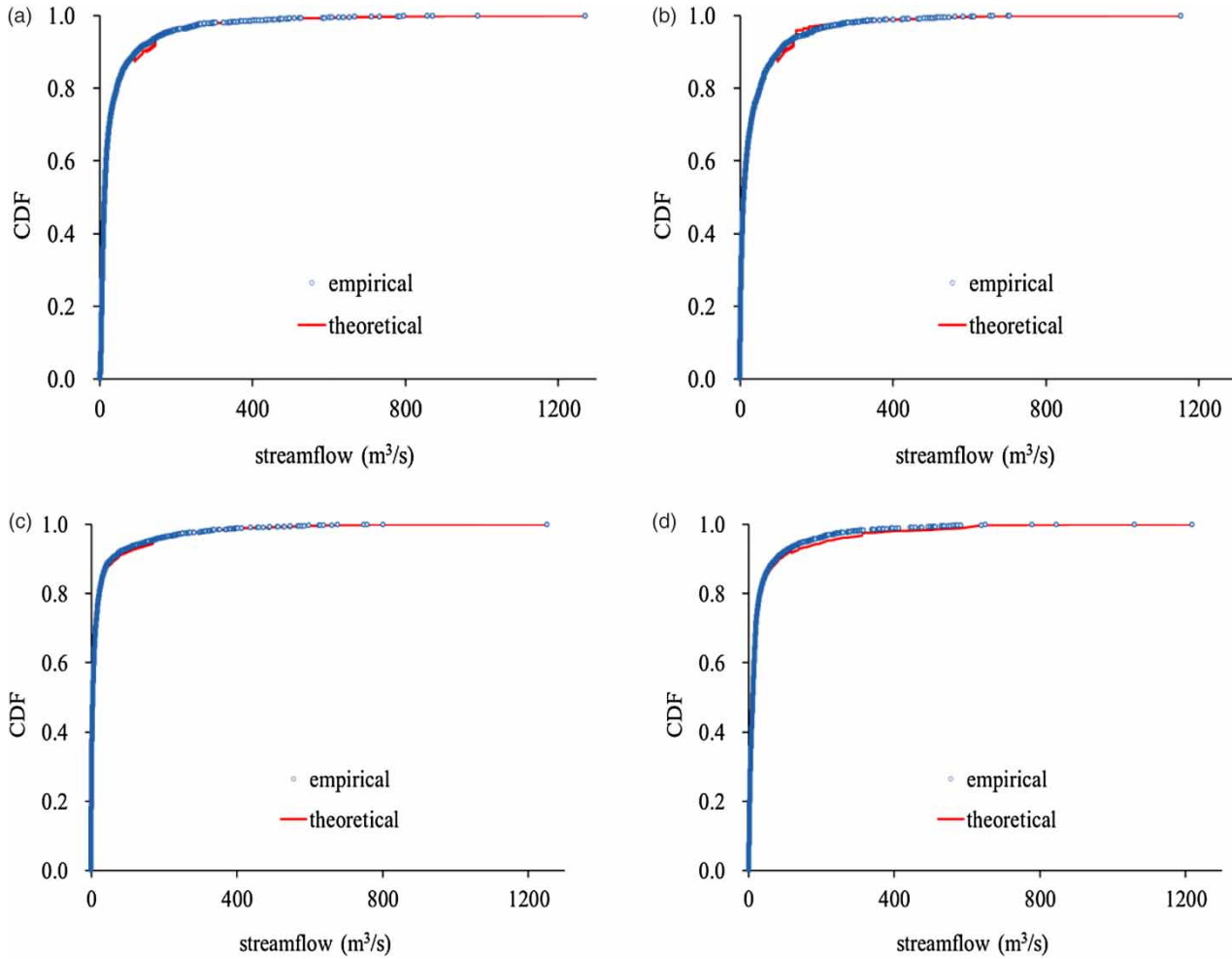


Figure 3 | Comparison of the empirical and theoretical cumulative distribution function of different variables (a) H , (b) S_1 , (c) S_2 , and (d) S_3 .

Table 4 | Summary of the three candidate bivariate Archimedean copulas and the relationship between their parameter and Kendall correlation coefficient

Copula	$C_\theta(u, v)$	$\theta \in$	τ
G-H	$\exp\left\{-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right\}$	$[1, \infty)$	$1 - \theta^{-1}$
Frank	$-\frac{1}{\theta} \ln \left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right]$	$\mathbb{R} \setminus \{0\}$	$\tau = 1 + \frac{4}{\theta} \left[\frac{1}{\theta} \int_0^\theta \frac{t}{\exp(t) - 1} dt - 1 \right]$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$(0, \infty)$	$\theta/(\theta + 2)$

ensembles using three criteria as shown in Equations (15)–(17). The effectiveness results of BMA, CBMA and CBP-BMA methods are listed in Table 1. The performances of different multi-models are better than that of the individual XAJ model in terms of *NSE*. The BMA method outperforms the reference model at the cost of *DRMS* and *KGE*

indicators. The CBMA and CBP-BMA methods slightly improve in all aspects during the calibration period, which have excellent properties in the validation period. The reason of the CBMA and CBP-BMA methods enhancing the performance can be attributed to that copula functions are efficient tools to remove bias instead of a simple bias

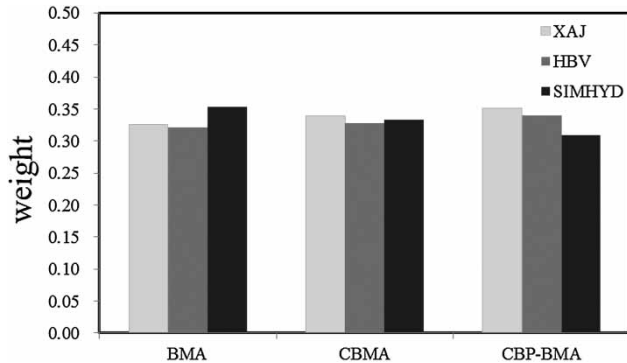


Figure 4 | Histogram of weights of individual model simulations in three methods.

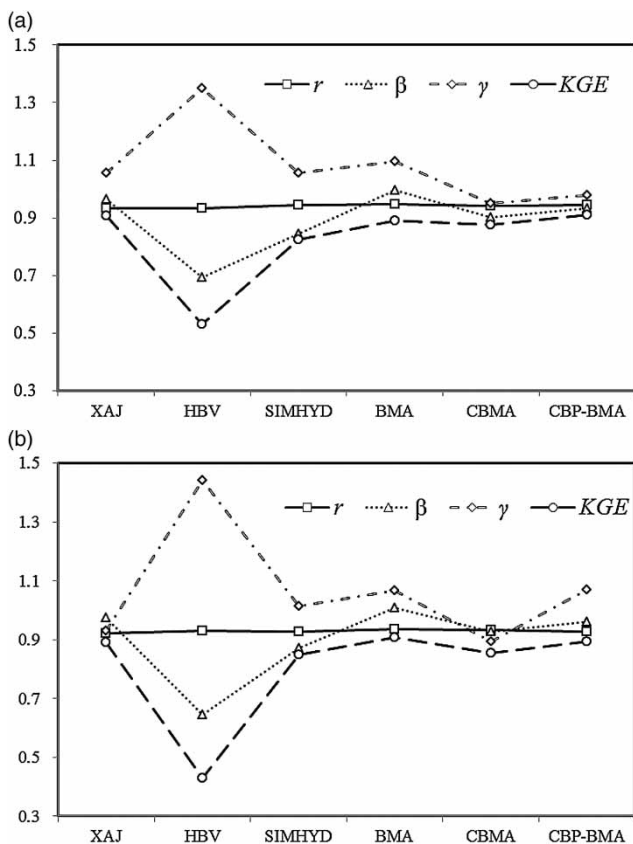


Figure 5 | The simulation results of KGE score and its components for (a) the calibration period and (b) the validation period.

correction such as linear regression in the BMA method (Madadgar & Moradkhani 2014). Especially, copula has reliable parameter estimation prior model average procedure. Another reason might be owed to the weight of each individual model, which is directly influenced by the estimation of posterior distributions.

Figure 5 illustrates the bar plots of KGE score and its components. The KGE score might be a little descending through BMA or CBMA application, a little incremental through CBP-BMA application in comparison with the best XAJ model. The correlation coefficients between observation and simulation of individual models are up to 0.93 in the calibration period and 0.92 in the verification period, which represent stronger correlation for the values are more than 0.9. However, β indicator of deterministic models varies from 0.64 for HBV model to 0.97 for XAJ model. The value less than 1 indicates total amount of streamflow simulation in any individual model is less than that of observation. It might cause the general underestimation of the mean streamflow (negative bias) in hydrological multi-model ensemble applications. The BMA method is such a promising method in locating simulation to observation for its term β closer to 1. Regarding the variability ratio, all methods except for HBV could perfectly perform, but no particular method is superior to others with all $\gamma \approx 1$.

Probabilistic verification of three ensemble methods

For probabilistic verification of simulation, Figures 6 and 7 describe the uncertainty bands of different methods for the representative year during calibration and verification periods with a visual inspection. These two plots indicate that the observed values approximately fall within the 5%–95% uncertainty range and fit the mean flow hydrograph for all multi-model ensembles. In this case, the 90% confidence interval could capture the flood peaks but miss more low flow values.

Three probabilistic verification measurements (containing ratio (CR), uncertainty (B), average deviation amplitude (D)) are presented in Table 5. It can be seen from these quantitative indices that they have a good performance in terms of containing ratio, which correspond to the confidence interval. The probability of observed value falling in the range should be in accord with the percentage of confidence interval containing points through many independent statistical experiments. The CBP-BMA method performs better than the CBMA method in terms of CR index, because it covers roughly 91% of the sample points, which is more than CBMA. A combination of CR

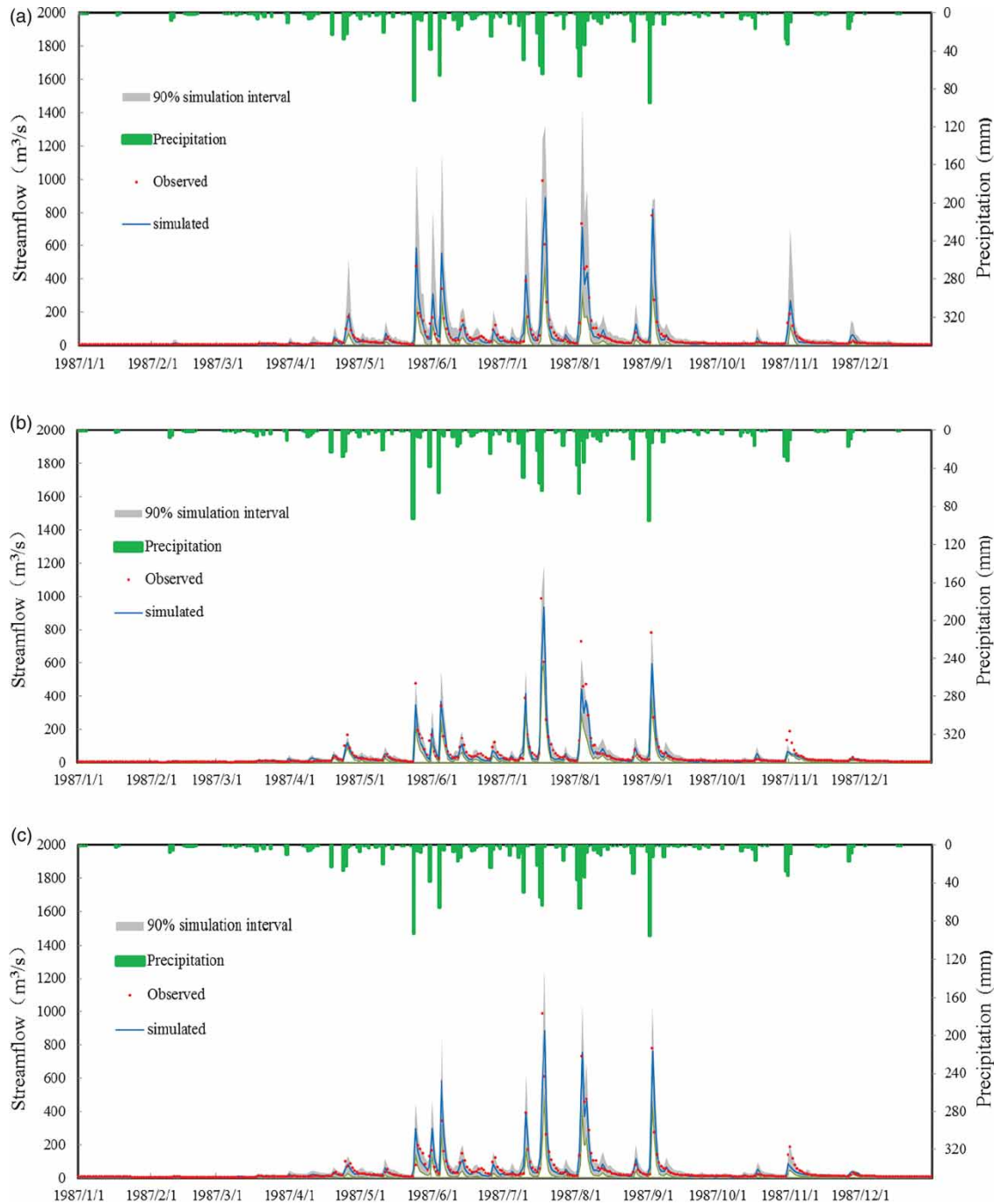


Figure 6 | The 90% uncertainty interval, observed, mean simulation for the Mumahe catchment in 1987 during the calibration period for (a) BMA, (b) CBMA, and (c) CBP-BMA.

and B possess the power to form a decision on model probabilistic performance. The comparison between the CBMA (and CBP-BMA) and BMA methods exactly illustrates that the CBMA method outperforms the BMA method, for either CR , B or D . In particular, the containing ratios of

the CBP-BMA method in different periods are up to 91.17% and 91.33%, respectively. Referring to the smaller B result in the CBMA and CBP-BMA methods, the total predictive variance is reduced by relaxing the pdf generated by copula functions rather than the Gaussian posterior

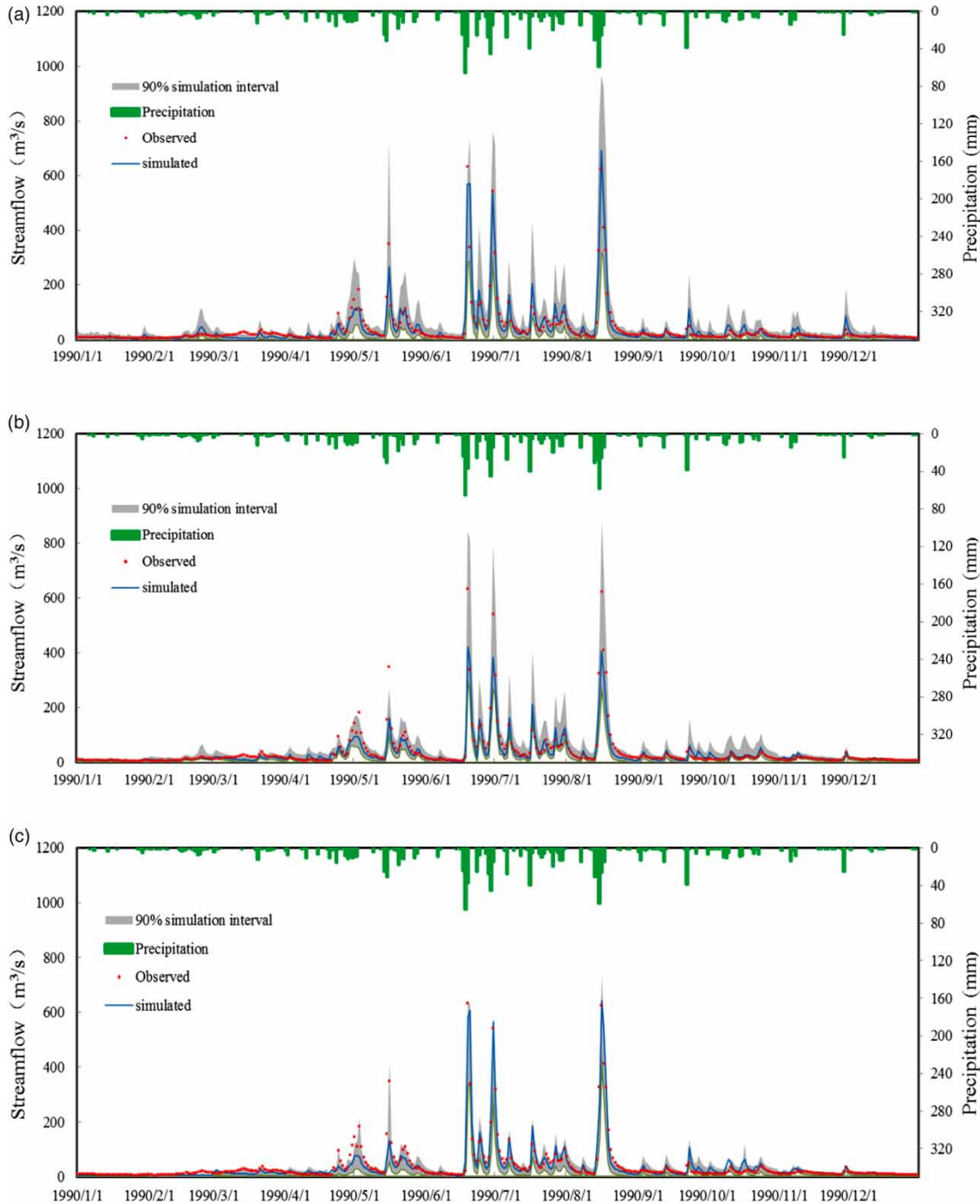


Figure 7 | The 90% uncertainty interval, observed, mean simulation for the Mumahe catchment in 1990 during the validation period for (a) BMA, (b) CBMA, and (c) CBP-BMA.

distribution via Box-Cox transformation. Since the between-model variance remains identical after using the same EM algorithm in all three methods, it is inferred that the reduction of within-model variance works.

The CBMA and CBP-BMA methods are two flexible and robust approaches to estimate uncertainty in terms of the optimal bandwidth and average deviation amplitude. They have an intuitive and simple structure conditional on several

Table 5 | Uncertainty assessment of different hydrological multi-model ensembles

Model	Calibration			Validation		
	CR (%)	B	D	CR (%)	B	D
BMA	87.34	56.79	21.26	88.53	58.14	18.79
CBMA	89.23	38.75	16.52	89.76	40.28	12.26
CBP-BMA	91.17	45.28	12.35	91.33	42.35	10.99

Note: values in bold represent the optimal result.

model simulations by the integration of BMA and copula tools, which makes this method promising to derive uncertainty. The difference between them reflected in the procedure of processing posterior distribution. Further improvement might be realized through the weight allocation for each model or the nonparametric posterior distribution.

CONCLUSIONS

Quantification of uncertainty in hydro-meteorological prediction is necessary for flood control and water resources management decision making. The proposed CBP-BMA method employed a Bayesian theory to modify the structure of the CBMA technique. The BMA, CBMA and CBP-BMA methods are implemented to evaluate uncertainties of hydrologic multi-model ensembles. The main findings are summarized as follows.

Deterministic results of different multi-model ensembles outperform those of the individual model. The CBMA and CBP-BMA methods slightly outperform BMA method in terms of *NSE*, *DRMS* and *KGE*. When the CBMA method is used as a reference, the CBP-BMA method can improve the *NSE* and *KGE* and enlarge *DRMS* values. Underestimation of all individual models may cause negative bias of ensemble multi-model.

The combination of containing ratio and bandwidth index demonstrates the probabilistic model performance with the auxiliary index-average deviation amplitude. It is found that containing ratio is approximately equal with the percentage of confidence interval. The CBMA or CBP-BMA methods outperform the BMA method in terms of evaluation criteria with a high containing ratio, small uncertainty, *B*, and average deviation amplitude, *D*.

The BMA method returns the predictive distribution of forecast variables as a weighted average of posterior distributions. The weights of the models involved are roughly defined proportional to the performance of each model. Integration of copula and BMA methods not only relaxes the assumption of posterior distribution, but simplifies the parameter estimation procedure. The results of this study are encouraging further integration of copula functions and Bayesian thought into hydro-meteorological applications where unknown conditional distributions require estimation.

ACKNOWLEDGEMENTS

This study is financially supported by the National Natural Science Foundation of China (51539009) and National Key Research and Development Project of China (2016YFC0402206).

REFERENCES

- Ajami, N. K., Duan, Q. Y. & Sorooshian, S. 2007 *An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction*. *Water Resour. Res.* **43**, W01403.
- Ba, H. H., Guo, S. L., Wang, Y., Hong, X. J. & Zhong, Y. X. 2017 *Improving ANN model performance in runoff forecasting by adding soil moisture input and using data preprocessing techniques*. *Hydrology Research* **49**. doi:10.2166/nh.2017.048.
- Carreau, J. & Bouvier, C. 2016 *Multivariate density model comparison for multi-site flood-risk rainfall in the French Mediterranean area*. *Stoch. Env. Res. Risk A* **30**, 1591–1612.
- Chebana, F. & Ouarda, T. B. M. J. 2007 *Multivariate L-moment homogeneity test*. *Water Resour. Res.* **43**, 199–212.
- Chen, F., Jiao, M. & Chen, J. 2013 *The meta-Gaussian Bayesian Processor of forecasts and associated preliminary experiments*. *Acta Meteorol. Sin.* **27**, 199–210.
- Chen, L., Singh, V. P., Guo, S. L., Zhou, J. Z. & Zhang, J. H. 2015 *Copula-based method for multisite monthly and daily streamflow simulation*. *J. Hydrol.* **528**, 369–384.
- Chen, L., Singh, V. P. & Xiong, F. 2017 *An entropy-based generalized gamma distribution for flood frequency analysis*. *Entropy* **19**, 239.
- Chiew, F. H. S., Teng, J., Vaze, J., Post, D. A., Perraud, J. M., Kirono, D. G. C. & Viney, N. R. 2009 *Estimating climate change impact on runoff across southeast Australia: method,*

- results, and implications of the modeling method. *Water Resour. Res.* **45**, 82–90.
- Dong, L. H., Xiong, L. H. & Yu, K. X. 2013 Uncertainty analysis of multiple hydrologic models using the Bayesian model averaging method. *J. Appl. Math.* **2013**, 1–11.
- Duan, Q. Y., Ajami, N. K., Gao, X. & Sorooshian, S. 2007 Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.* **30**, 1371–1386.
- Fan, Y. R., Huang, G. H., Li, Y. P., Wang, X. Q. & Li, Z. 2016 Probabilistic prediction for monthly streamflow through coupling stepwise cluster analysis and quantile regression methods. *Water Resour. Manag.* **30**, 5313–5331.
- Götzinger, J. & Bárdossy, A. 2008 Generic error model for calibration and uncertainty estimation of hydrological models. *Water Resour. Res.* **44**, 1393–1442.
- Han, J. C., Huang, G. H., Zhang, H., Li, Z. & Li, Y. P. 2014 Bayesian uncertainty analysis in hydrological modeling associated with watershed subdivision level: a case study of SLURP model applied to the Xiangxi River watershed, China. *Stoch. Env. Res. Risk A* **28**, 973–989.
- Hemri, S., Lisniak, D. & Klein, B. 2015 Multivariate post-processing techniques for probabilistic hydrological forecasting. *Water Resour. Res.* **51**, 7436–7451.
- Jonathan, R. M. 1990 L-moment-analysis and estimation of distributions using linear-combinations of order-statistics. *J. Roy. Stat. Soc.* **52**, 105–124.
- Khajehi, S. & Moradkhani, H. 2017 Towards an improved ensemble precipitation forecast: a probabilistic post-processing approach. *J. Hydrol.* **546**, 476–489.
- Klein, B., Meissner, D., Kobiakka, H. U. & Reggiani, P. 2016 Predictive uncertainty estimation of hydrological multi-model ensembles using pair-copula construction. *Water-Sui.* **8**, 1–22.
- Kling, H., Fuchs, M. & Paulin, M. 2012 Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.* **424–425**, 264–277.
- Kroes, D. P., Taimre, T. & Botev, Z. I. 2011 *Handbook of Monte Carlo Methods*. John Wiley, New York, pp. 772.
- Krzysztofowicz, R. 1999 Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.* **35**, 2739–2750.
- Krzysztofowicz, R. & Kelly, K. S. 2000 Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.* **36**, 3265–3277.
- Kuczera, G. & Parent, E. 1998 Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the metropolis algorithm. *J. Hydrol.* **211**, 69–85.
- Lee, T. & Salas, J. 2011 Copula-based stochastic streamflow generation of hydrological data applied to Nile River flows. *Hydrol. Res.* **42**, 318–330.
- Li, L., Xu, C. Y., Xia, J., Engeland, K. & Reggiani, P. 2011 Uncertainty estimates by Bayesian method with likelihood of AR (1) plus normal model and AR (1) plus Multi-Normal model in different time-scales hydrological models. *J. Hydrol.* **406**, 54–65.
- Li, C., Singh, V. P. & Mishra, A. K. 2013 Monthly river flow simulation with a joint conditional density estimation network. *Water Resour. Res.* **49**, 3229–3242.
- Li, W., Zhou, J. Z., Sun, H. W., Feng, K. L., Zhang, H. & Tayyab, M. 2017 Impact of distribution type in Bayes probability flood forecasting. *Water Resour. Manag.* **31**, 961–977.
- Liang, Z., Wang, D., Guo, Y., Zhang, Y. & Dai, R. 2011 Application of Bayesian model averaging approach to multimodel ensemble hydrologic forecasting. *J. Hydrol. Eng.* **18**, 1426–1436.
- Liu, Z., Zhou, P., Chen, X. & Guan, Y. 2015 A multivariate conditional model for streamflow prediction and spatial precipitation refinement. *J. Geophys. Res. Atmos.* **120**, 10116–10129.
- Liu, Z. J., Guo, S. L., Zhang, H. G., Liu, D. D. & Yang, G. 2016 Comparative study of three updating procedures for real-time flood forecasting. *Water Resour. Manag.* **30**, 2111–2126.
- Liu, X. Y., Chen, Lu., Zhu, Y. h., Singh, V. P., Qu, G. & Cuo, X. h. 2017a Multi-objective reservoir operation during flood season considering spillway optimization. *J. Hydrol.* **552**, 554–563.
- Liu, Z. J., Guo, S. L., Xiong, L. H. & Xu, C.-Y. 2017b Hydrological uncertainty processor based on a copula function. *Hydrolog. Sci. J.* **63**, 74–86. doi:10.1080/02626667.2017.1410278.
- Madadgar, S. & Moradkhani, H. 2014 Improved Bayesian multimodeling: integration of copulas and Bayesian model averaging. *Water Resour. Res.* **50**, 9586–9603.
- McEnery, J., Ingram, J., Duan, Q. Y., Adams, T. & Anderson, L. 2005 NOAA's advanced hydrologic prediction service: building pathways for better science in water forecasting. *B. Am. Meteorol. Soc.* **86**, 375–385.
- Möller, A., Lenkoski, A. & Thorarindottir, T. L. 2013 Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Q. J. Roy. Meteor. Soc.* **139**, 982–991.
- Montanari, A. & Brath, A. 2004 A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resour. Res.* **40**, 75–88.
- Montero, R. A., Schwanenberg, D., Krahe, P., Lisniak, D., Sensoy, A., Sorman, A. A. & Akkol, B. 2016 Moving horizon estimation for assimilating H-SAF remote sensing data into the HBV hydrological model. *Adv. Water Resour.* **92**, 248–257.
- Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I – A discussion of principles. *J. Hydrol.* **10**, 282–290.
- Nasonova, O. N., Gusev, Y. M. & Kovalev, Y. E. 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. Hydrometeorol.* **10**, 1128–1150.
- Raftery, A. E., Gneiting, T., Balabdaoui, F. & Polakowski, M. 2005 Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* **133**, 1155–1174.
- Robert, C. P. & Casella, G. 2011 Monte Carlo statistical methods. *Technometrics* **35**, 430–431.

- Seo, D. J., Herr, H. D. & Schaake, J. C. 2006 A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. *Hydrol. Earth Syst. SC* **3**, 1987–2035.
- Sklar, M. 1959 *Fonctions de répartition à n dimensions et leurs marges*. Publ. Inst. Statist. Univ. Paris, Paris, pp. 229–231 (in French).
- Vrugt, J. A. & Robinson, B. A. 2007 Treatment of uncertainty using ensemble methods: comparison of sequential data assimilation and Bayesian model averaging. *Water Resour. Res.* **43**, 223–228.
- Wu, J., Zhou, J. Z., Chen, L. & Ye, L. 2015 Coupling forecast methods of multiple rainfall-runoff models for improving the precision of hydrological forecasting. *Water Resour. Manag.* **29**, 5091–5108.
- Wu, X. S., Wang, Z. L., Guo, S. L., Liao, W., Zeng, Z. & Chen, X. 2017 Scenario-based projections of future urban inundation within a coupled hydrodynamic model framework: a case study in Dongguan City, China. *J. Hydrol.* **547**, 428–442.
- Xiong, L. H., Min, W., Wei, X. J. & O'Connor, K. M. 2009 Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation. *Hydrolog. Sci. J.* **54**, 852–871.
- Yu, B. & Zhu, Z. 2015 A comparative assessment of AWBM and SimHyd for forested watersheds. *Hydrolog. Sci. J.* **60**, 1–13.
- Zhang, L. & Singh, V. P. 2007 Bivariate rainfall frequency distributions using Archimedean copulas. *J. Hydrol.* **332**, 93–109.
- Zhao, R. J. 1992 The Xinanjiang model applied in China. *J. Hydrol.* **135**, 371–381.
- Zhou, Y. L., Guo, S. L., Xu, C. Y., Chen, H., Guo, J. & Lin, K. 2016 Probabilistic prediction in ungauged basins (PUB) based on regional parameter estimation and Bayesian model averaging. *Hydrol. Res.* **47**, 1087–1103.

First received 28 September 2017; accepted in revised form 29 December 2017. Available online 1 March 2018