Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 346045, 11 pages
http://dx.doi.org/10.1155/2013/346045

Research Article

Uncertainty Analysis of Multiple Hydrologic Models Using the Bayesian Model Averaging Method

Leihua Dong,[1,2] Lihua Xiong,[1] and Kun-xia Yu[1]

[1] *State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China*
[2] *National Research Center for Sustainable Hydropower Development, China Institute of Water Resources and Hydropower Research, Beijing 100038, China*

Correspondence should be addressed to Lihua Xiong; xionglh@whu.edu.cn

Received 30 August 2013; Accepted 6 November 2013

Academic Editor: Y. P. Li

Since Bayesian Model Averaging (BMA) method can combine the forecasts of different models together to generate a new one which is expected to be better than any individual model's forecast, it has been widely used in hydrology for ensemble hydrologic prediction. Previous studies of the BMA mostly focused on the comparison of the BMA mean prediction with each individual model's prediction. As BMA has the ability to provide a statistical distribution of the quantity to be forecasted, the research focus in this study is shifted onto the comparison of the prediction uncertainty interval generated by BMA with that of each individual model under two different BMA combination schemes. In the first BMA scheme, three models under the same Nash-Sutcliffe efficiency objective function are, respectively, calibrated, thus providing three-member predictions ensemble for the BMA combination. In the second BMA scheme, all three models are, respectively, calibrated under three different objective functions other than Nash-Sutcliffe efficiency to obtain nine-member predictions ensemble. Finally, the model efficiency and the uncertainty intervals of each individual model and two BMA combination schemes are assessed and compared.

1. Introduction

To date, various hydrological models have been put forward and widely used in flood forecasting, planning, and water resources management [1, 2]. Since different models have strengths in capturing different aspects of the real world processes, combining the results from diverse models by weighting procedures can present a better performance than any individual model [3–5]. The early model combination researches in hydrologic forecasting employed such tools as neural network [6] and fuzzy system [7]. Recently, Bayesian Model Averaging (BMA), a method for averaging over different competing models, has been introduced to ensemble hydrologic predictions.

Bayesian Model Averaging came to prominence in statistics in the mid-1990s, and Madigan and Raftery [8] were the first to propose this method for combining predictions. Subsequently, Raftery [9] and Draper [10] gave more detailed discussion about BMA. It has been applied in diverse fields such as economics [11], biology [12], ecology [13], public health [14], toxicology [15], meteorology [16], and management science [17]. In many case studies, BMA produces accurate and reliable predictions and was shown to be a better scheme than other model-combining methods [18–20]. In recent years, hydrologists have also applied BMA to hydrologic modeling, such as groundwater [21] and rainfall-runoff modeling [22–24].

A prediction from a single model has been recognized to be associated with a certain degree of uncertainty, and so is the prediction from combining a number of different single models. Thus, uncertainty analysis is an indispensable element for any hydrologic modeling study. The uncertainty usually arises from errors during the calibration of parameters, the design of model structure, and measurements of input and output data [25, 26]. To account for these uncertainties, many uncertainty analysis techniques have been developed and applied to diverse catchments, such as Generalized Likelihood Uncertainty Estimation (GLUE),

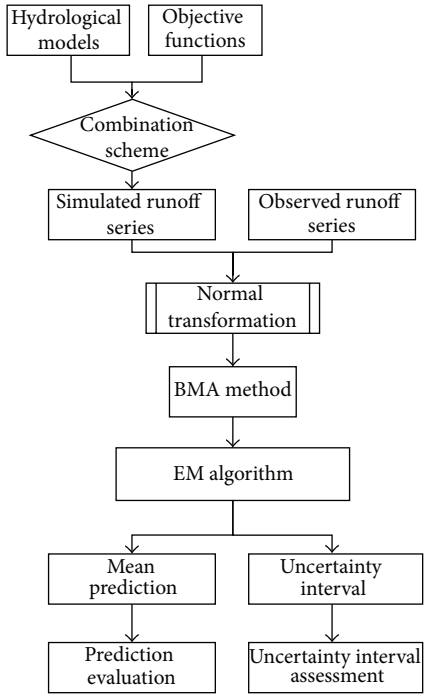

Figure 1: Flowchart of using BMA scheme for hydrological ensemble prediction as well as for prediction uncertainty analysis.

Parameter Solution (ParaSol), and Bayesian inference based on Markov chain Monte Carlo (MCMC) [27, 28]. Each of those techniques has its own advantage in uncertainty analysis. In the uncertainty analysis of BMA scheme, the composition of Monte Carlo method [29] is used to generate BMA probabilistic ensemble predictions, and then the 90% uncertainty intervals can be derived within the range of the 5% and 95% quantiles.

Previous studies of BMA in hydrology mostly focused on the comparison of the BMA mean prediction with each individual model's prediction, to prove the better performance of the prediction after weighted averaging. As BMA also has the ability to provide a statistical distribution of the quantity to be forecasted, the research focus in this study is shifted onto the comparison of the prediction uncertainty interval generated by the BMA with that of each individual model, in order to see if BMA can also improve the prediction reliability. The technical route of the research in this paper is described in Figure 1. Another purpose of this paper is that by calibrating different hydrological models under different objective functions, each of which has distinctive advantages in better modeling certain flow ranges, we can construct different sets of ensemble members for combination in order to fully explore the superiority of BMA. Therefore, two kinds of BMA combination schemes are designed, analyzed, and compared. In the first BMA scheme, we calibrate each of the three models under the same Nash-Sutcliffe efficiency objective function, thus providing three-member predictions ensemble for the BMA combination. In the second BMA scheme, three different objective functions other than Nash-Sutcliffe efficiency are adopted, each of which is supposed to have some advantage of better simulating a certain range of flows (low flow, medium flow, and high flow). All three models are, respectively, calibrated for each of three objective functions to obtain the optimized parameter sets.

2. Methods

2.1. Bayesian Model Averaging. Bayesian Model Averaging (BMA) is a statistical technique designed to infer a prediction by weighted averaging over many different competing models. This method is not only a scheme for model combination but also a coherent approach for accounting for between-model and within-model uncertainty [22]. Below is a brief description of the basic ideas of this method.

Let us consider a quantity Q to be predicted on the basis of input data $D = [X, Y]$ (X denotes the input forcing data, and Y stands for the observational flow data). $f = [f_1, f_2, \ldots, f_K]$ is the ensemble of the K-member predictions. The probabilistic prediction of BMA is given by

$$p(Q \mid D) = \sum_{k=1}^{K} p(f_k \mid D) \cdot p_k(Q \mid f_k, D). \quad (1)$$

The terms in (1) are explained as follows. $p(f_k \mid D)$ is the posterior probability of the prediction f_k given the input data D and reflects how well model f_k fits Y. Actually $p(f_k \mid \mathrm{D})$ is just the BMA weight w_k, and better performing predictions receive higher weights than the worse performing ones; all weights are positive and should add up to 1. $p_k(Q \mid f_k, D)$ is the conditional probability density function (PDF) of the predictand Q conditional on f_k and D. For computation convenience, $p_k(Q \mid f_k, D)$ is always assumed to be a normal PDF and is represented as $g(Q \mid f_k, \sigma_k^2) \sim N(f_k, \sigma_k^2)$, where σ_k^2 is the variance associated with model prediction f_k and observations Y. In order to make this assumption valid, some techniques such as Box-Cox transformation are needed to make the data approximately normally distributed and to narrow the data range.

The BMA mean prediction is a weighted average of the individual model's predictions, with their posterior probabilities being the weights. In the case that the observations and individual model predictions are all normally distributed, the BMA mean prediction can be expressed as

$$E[Q \mid D] = \sum_{k=1}^{K} p(f_k \mid D) \cdot E\left[g\left(Q \mid f_k, \sigma_k^2\right)\right] = \sum_{k=1}^{K} w_k f_k. \quad (2)$$

2.2. EM Algorithm for BMA Parameter Estimation. To estimate BMA weight w_k and model prediction variance σ_k^2, the Expectation-Maximization (EM) algorithm, which has proved to be an efficient technique for BMA calculation based on the assumption that K-member predictions are normally distributed, is described in this section [23].

Firstly, if we denote the set of BMA parameters to be estimated by $\theta = \{w_k, \sigma_k^2,\ k = 1, 2, \ldots, K\}$, the log form of likelihood function can be represented as

$$l(\theta) = \log\left(p(Q \mid D)\right) = \log\left(\sum_{k=1}^{K} w_k \cdot g\left(Q \mid f_k, \sigma_k^2\right)\right). \quad (3)$$

It is difficult to maximize the function (3) by analytical method. The EM algorithm is a method for finding the maximum likelihood by alternating between two steps, the expectation step and maximization step. The two steps are iterated to convergence when there is no significant change between two consecutive iterative log-likelihood estimations. In EM algorithm, a latent variable (unobserved quantity) z_k^t is used as an assistant for estimating BMA weight w_k. The procedure of EM algorithm for BMA scheme is described as follows.

(1) Initialization. Set Iter = 0.

Initialize

$$w_k^{(0)} = \frac{1}{K},$$
$$\sigma_k^{2(0)} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T}\left(Y^t - f_k^t\right)^2}{K \cdot T}, \quad (4)$$

where Iter is the number of iteration and T is the number of data in the calibration period. Y^t and f_k^t are denoted as the observation and the corresponding prediction by the kth model for the time t.

(2) Calculate the Initial Likelihood:

$$l(\theta)^{(0)} = \sum_{t=1}^{T} \log\left(\sum_{k=1}^{K}\left(w_k^{(0)} \cdot g\left(Q \mid f_k^t, \sigma_k^{2(0)}\right)\right)\right). \quad (5)$$

(3) Compute the Latent Variable. Set Iter = Iter + 1, then calculate

$$z_k^{t\,(\mathrm{Iter})} = \frac{g\left(Q \mid f_k^t, \sigma_k^{2(\mathrm{Iter}-1)}\right)}{\sum_{k=1}^{K} g\left(Q \mid f_k^t, \sigma_k^{2(\mathrm{Iter}-1)}\right)}. \quad (6)$$

(4) Update the Weight:

$$w_k^{(\mathrm{Iter})} = \frac{1}{T}\left(\sum_{t=1}^{T} z_k^{t\,(\mathrm{Iter})}\right). \quad (7)$$

(5) Update the Variance:

$$\sigma_k^{2(\mathrm{Iter})} = \frac{\sum_{t=1}^{T} z_k^{t\,(\mathrm{Iter})} \cdot \left(Y^t - f_k^t\right)^2}{\sum_{t=1}^{T} z_k^{t\,(\mathrm{Iter})}}. \quad (8)$$

(6) Update the Likelihood:

$$l(\theta)^{(\mathrm{Iter})} = \sum_{t=1}^{T} \log\left(\sum_{k=1}^{K}\left(w_k^{(\mathrm{Iter})} \cdot g\left(Q \mid f_k^t, \sigma_k^{2(\mathrm{Iter})}\right)\right)\right). \quad (9)$$

(7) Check for Convergence. If $l(\theta)^{(\mathrm{Iter})} - l(\theta)^{(\mathrm{Iter}-1)}$ is less than a prespecified tolerance level, stop the whole estimation procedure; else go back to Step (3).

2.3. Estimation of Prediction Uncertainty Interval. After BMA weight w_k and prediction variance σ_k^2 being estimated, we use the composition of Monte Carlo method to generate BMA probabilistic predictions for any time t [29]. The procedures are described as follows.

(1) Generate an integer value of k from $[1, 2, \ldots, K]$ with probability $[w_1, w_2, \ldots, w_K]$. A specific procedure is described as follows.

  (1a) Set the cumulative weight $w_0' = 0$ and compute $w_k' = w_{k-1}' + w_k$ for $k = 1, 2, \ldots, K$.

  (1b) Generate a random number u between 0 and 1.

  (1c) If $w_{k-1}' \le u < w_k'$, it indicates that we choose the kth member of the ensemble predictions.

(2) Generate a value of Q_t from the PDF of $g(Q_t \mid f_k^t, \sigma_k^2)$. Here, $g(Q_t \mid f_k^t, \sigma_k^2)$ represents the normal distribution with mean f_k^t and variance σ_k^2.

(3) Repeat the above steps (1) and (2) for M times. M is the probabilistic ensemble size. In this paper, we set $M = 100$.

After generating the BMA probabilistic ensemble predictions, sort them in the ascending order. Then the 90% uncertainty intervals can be derived within the range of the 5% and 95% quantiles.

For each individual model in the BMA scheme, the prediction uncertainty interval can also be constructed, with the Monte Carlo sampling method still being used to approximate the assumed PDF of $g(Q_t \mid f_k^t, \sigma_k^2)$.

3. Materials

3.1. Study Area and Data. The study area is Mumahe catchment, a branch of Han River. It is located in Shanxi Province of China and the total area is 1224 km^2. The basin has a subtropical climate, and the area is humid with fairly high precipitation. The mean annual rainfall for the period of 1980–1987 is 1070 mm, and the mean annual runoff is 687 mm, or roughly 64% of the annual rainfall. The hydrological data include daily runoff, rainfall, and evaporation. There are 2992 data points in total, and 1825 (the period of 1980.1.1–1985.12.31) of them are used for calibration, while the rest 1167 data points (the period of 1986.1.1–1987.12.31) are used for validation.

3.2. Hydrological Models and Optimization Algorithm. In this study, three conceptual hydrological models are employed for testing the capability of BMA: the Xinanjiang Rainfall-Runoff Model (XAJ), the Soil Moisture Accounting and Routing Model (SMAR), and SIMHYD Rainfall-Runoff Model.

Xinanjiang Rainfall-Runoff Model was developed in 1970s. It is a conceptual hydrologic model, which has been widely used in humid and semihumid regions of China. And all the 15 parameters of this model have strong physical meanings. SMAR model is a lumped conceptual model with soil moisture as a central theme. The model consists of two

components in sequence: a water balance component with 5 water balance parameters and a routing component with 4 routing parameters. SIMHYD model is a daily conceptual model that estimates daily stream flow from daily rainfall and areal potential evapotranspiration data and it contains 7 parameters [30]. For calibrating these hydrological models, Shuffled Complex Evolution (SCE-UA) method is employed here for parameter optimization [31].

3.3. Objective Functions. The selection of objective function (OF) is of great importance since it will have great influence on the values of calibrated parameters and thus on simulation results of the rainfall-runoff model. Different objective functions can be adopted for different kinds of practical issues. For example, the objective function of squared model errors of squared transformed flow can be applied in high flow studies, and the objective function of squared model errors of logarithmic transformed flow can be applied in low flow studies [32]. In this study, four objective functions have been used for the parameter calibration.

(1) OF1: The Nash-Sutcliffe Coefficient of Efficiency (R^2):

$$R^2 = 1.0 - \frac{\sum_{t=1}^{T} \left(Q_{\text{obs}}^t - Q_{\text{sim}}^t\right)^2}{\sum_{t=1}^{T} \left(Q_{\text{obs}}^t - \overline{Q_{\text{obs}}}\right)^2}, \tag{10}$$

where Q_{obs}^t and Q_{sim}^t are observed and simulated data at time t and $\overline{Q_{\text{obs}}}$ is the average of observed data in the calibration period.

(2) OF2: Mean Squared Error of Squared Transformed (MSEST):

$$\text{MSEST} = \frac{\sum_{t=1}^{T} \left(Q_{\text{obs}}^{t\,2} - Q_{\text{sim}}^{t\,2}\right)^2}{T}. \tag{11}$$

Transforming the observed data in squared form puts great emphasis on fitting peak values.

(3) OF3: Mean Squared Error of Squared Root Transformed (MSESRT):

$$\text{MSESRT} = \frac{\sum_{t=1}^{T} \left(\sqrt{Q_{\text{obs}}^t} - \sqrt{Q_{\text{sim}}^t}\right)^2}{T}. \tag{12}$$

MSESRT can be employed in the medium flow simulation.

(4) OF4: Mean Squared Error of Logarithmic Transformed (MSELT):

$$\text{MSELT} = \frac{\sum_{t=1}^{T} \left(\ln Q_{\text{obs}}^t - \ln Q_{\text{sim}}^t\right)^2}{T}. \tag{13}$$

This transformation helps model parameterization to better fit the low flow values.

3.4. Construction of BMA(3) and BMA(9) Schemes. When the prediction data are highly non-Gaussian, we should firstly transform the data to be normally distributed by Box-Cox transformation before using EM algorithm. OF1 is the most widely used objective function for parameter optimization and is used in calibrating each of three hydrological models mentioned above to generate three different predictions. We combine these three different predictions by BMA to construct a three-member predictions ensemble; thus, we denote the first BMA scheme as BMA(3). Figure 2 shows the procedure of BMA(3) combination scheme. The other three objective functions, that is, OF2, OF3, and OF4 are, respectively, fit for high, medium, and low flow simulation. All three hydrological models are, respectively, calibrated for each of these three objective functions to obtain the optimized parameter sets. As the same model with different parameter sets will give rise to different outcomes, nine different predictions are generated. We can use BMA method to combine these nine different predictions to construct a nine-member predictions ensemble, which is just the second BMA scheme denoted as BMA(9). The procedure of BMA(9) combination scheme is described in Figure 3.

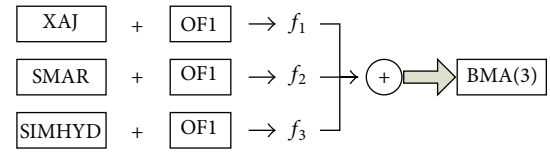

Figure 2: Diagram of BMA(3) combination scheme.

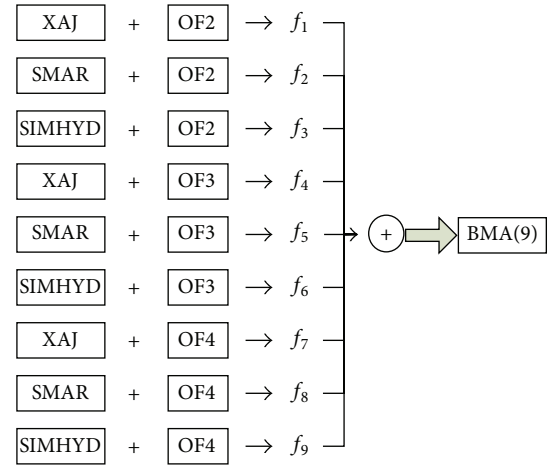

Figure 3: Diagram of BMA(9) combination scheme.

Let E denote the uncertainty of the forecast, and it can be written as $E = [E_h, E_m, E_l]$, including three components, that is, the high flow simulation uncertainty E_h, the medium flow simulation uncertainty E_m, and the low flow simulation uncertainty E_l. In BMA(9), the forecasts which are generated under OF2 have relatively small E_h, so they can get higher weights than other forecasts in high flow simulation. Similarly, the forecasts generated under OF3 have relatively high weights in medium flow simulation, while the ones generated under OF4 have higher weights than others in low flow simulation. By averaging the forecasts from a set of different combinations of hydrological model and objective function, the advantage of BMA(9) is its ability to reduce the simulation

errors by giving weights to each of the nine-member forecasts according to their performance in different flow ranges.

3.5. Performance Criteria for Evaluating the Mean Prediction. There are three indices for evaluating the mean prediction.

(1) The Nash-Sutcliffe Coefficient of Efficiency (R^2). The definition of R^2 has expressed in (10). R^2 is not only an objective function but also a widely used performance criterion. It ranges from minus infinity to 1.0, with higher values indicating better agreement. It is difficult to evaluate the performance of the model with R^2 in all flow ranges, since the value of R^2 is always negative in the medium flow range.

(2) Daily Root Mean Square Error (DRMS):

$$\mathrm{DRMS} = \sqrt{\frac{\sum_{t=1}^{T}\left(Q_{\mathrm{obs}}^{t} - Q_{\mathrm{sim}}^{t}\right)^2}{T}}, \tag{14}$$

where Q_{obs}^{t} and Q_{sim}^{t} are observed and simulated data at time t. *DRMS* is sensitive to the differences between the observations and simulations. The lower the DRMS value is, the better the prediction performance is.

(3) Relative Error of Total Runoff (RE):

$$\mathrm{RE} = 1.0 - \frac{\sum_{t=1}^{T} Q_{\mathrm{sim}}^{t}}{\sum_{t=1}^{T} Q_{\mathrm{obs}}^{t}}. \tag{15}$$

It reflects the performance in the simulation of the total runoff amount. Lower values of RE indicate better agreement of total surface runoff.

3.6. Performance Criteria for Assessing the Prediction Uncertainty Interval. Xiong et al. [33] have presented a set of indices for assessing the prediction uncertainty intervals generated by the uncertainty analysis methods. Three main indices are selected here to assess the prediction uncertainty intervals produced by BMA schemes as well as from each individual hydrological model.

(1) Containing Ratio (CR). The containing ratio is used for assessing the goodness of the uncertainty interval. It is defined as the percentage of observed data points that are covered in the prediction bounds.

(2) Average Band-Width (B). Consider

$$B = \frac{1}{T}\sum_{t=1}^{T}\left(q_u^t - q_l^t\right), \tag{16}$$

where q_u^t and q_l^t are denoted as upper and lower prediction bounds at time t. The average band-width B is also an index for measuring the performance of estimated uncertainty interval.

(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

$$D = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{1}{2}\left(q_u^t + q_l^t\right) - Q_{\mathrm{obs}}^{t}\right|, \tag{17}$$

where Q_{obs}^{t} is the observed discharge at time t.

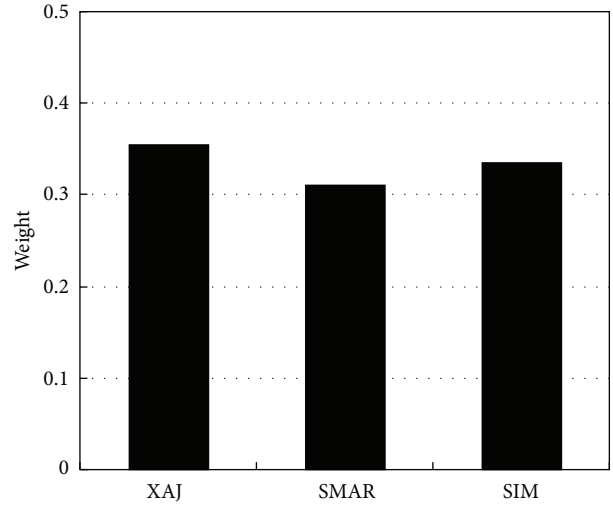

Figure 4: Histogram of weights of individual model predictions in BMA(3) scheme.

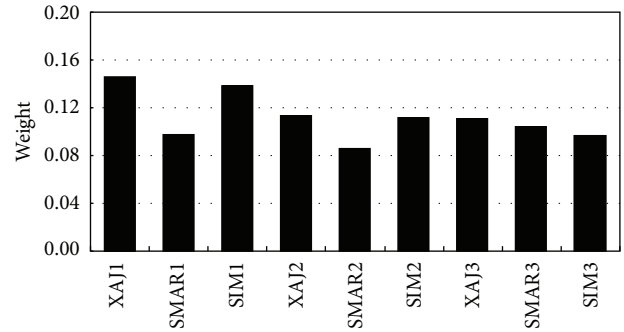

Figure 5: Histogram of weights of the individual model predictions in BMA(9) scheme.

4. Results and Discussion

The weights of individual models in BMA(3) scheme are displayed in Figure 4, while the weights in BMA(9) are showed in Figure 5. Moreover, in order to compare the performance of two BMA schemes in different flow ranges, according to the characteristics of the streamflow values of Mumahe catchment, data are broken into three flow ranges: high flow (top 10%), medium flow (middle 50%), and low flow (bottom 40%).

4.1. BMA(3) Results. We check the mean prediction of BMA(3) using three criteria illustrated in Section 4.1. Results of BMA(3) and its 3 individual models in the mean prediction

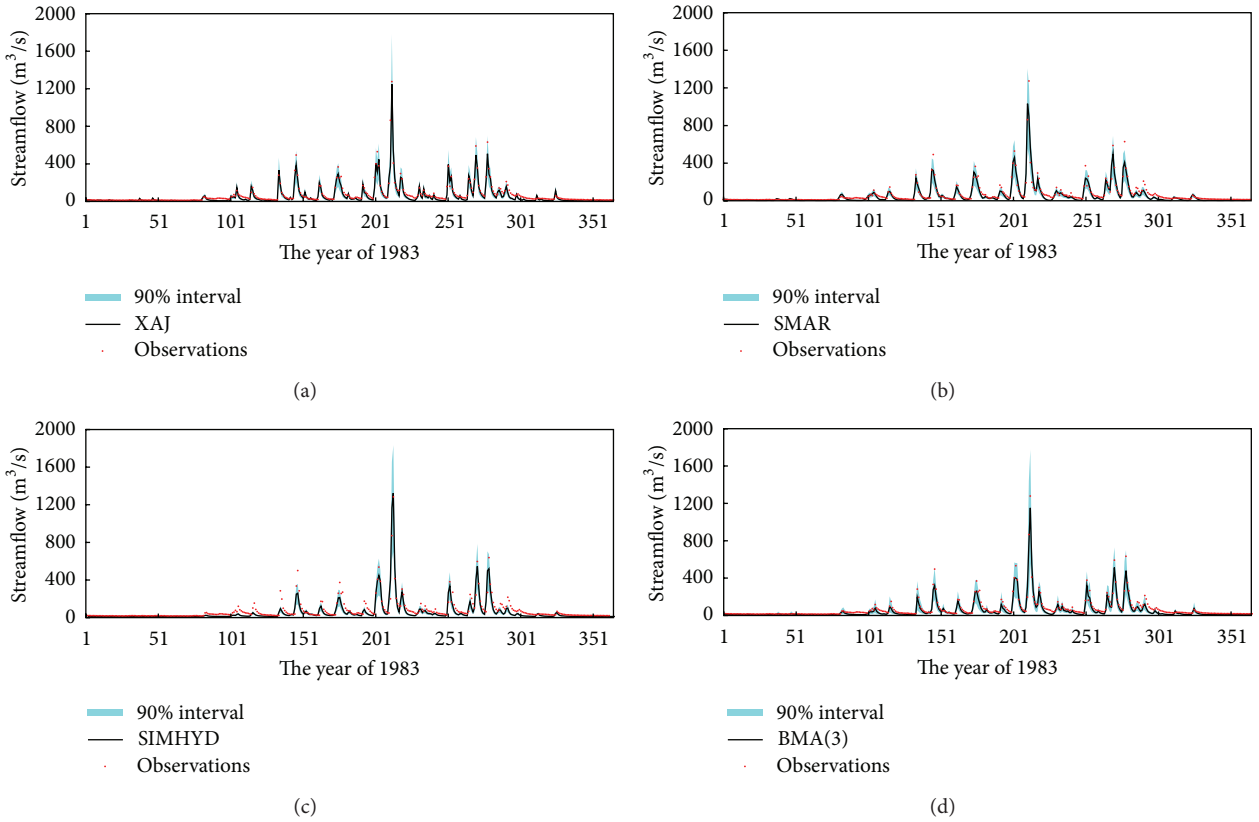

FIGURE 6: The mean prediction and 90% uncertainty interval of both BMA(3) and 3 individual models for the Mumahe catchment in 1983 during the calibration period.

for the whole flow series are presented in Table 1. In terms of R^2, the mean prediction of BMA(3) can achieve 90.68% in calibration period and 86.98% in validation period, which is better than its best individual model prediction (XAJ). However, in terms of RE, the mean prediction of BMA(3) performs much worse than its best individual model prediction.

Three indices illustrated in Section 4.2 are used for assessing the prediction uncertainty intervals of both BMA(3) and its three individual models. The results for the whole flow series are also showed in Table 1. It is clear that BMA(3) uncertainty interval has the largest values of CR and B, and almost the smallest D, in both calibration and validation periods. In other words, BMA(3) uncertainty interval has better properties than any individual model's uncertainty interval in terms of CR and D, but worse in terms of B. Then we compare the differences between BMA(3) and its individual model in uncertainty interval by the graph. Figure 6 displays the mean prediction and 90% uncertainty interval of both BMA(3) and its 3 individual models for Mumahe catchment in the year of 1983 during the calibration period. The observations of 1983 are shown as dots, and the BMA(3) mean prediction and its individual models' predictions are represented by solid curve. As the statistical results showed in Table 1, the uncertainty intervals of the individual models have low containing ratio and large deviation amplitude. But the uncertainty interval of BMA(3) is much broader than that of any of its individuals. It can be found from Figure 7 that the results of validation period are similar to that of the calibration period. In general, the uncertainty interval of BMA(3) has better performance than its individual models for the whole flow series.

4.2. BMA(9) Results. Table 2 lists the results of BMA(9) and its 9 individual models in the mean prediction for the whole flow series. And from it we can easily find that in calibration period, the mean prediction of BMA(9) performs better than its best individual prediction according to the value of R^2 and DRMS, though the mean prediction of BMA(9) does not have any advantage in comparison to its individual model predictions in terms of RE.

The results of the uncertainty intervals of BMA(9) and its 9 individual models are also listed in Table 2. The containing ratio of BMA(9) uncertainty interval reaches 91.11% in calibration period and 90.23% in validation period, which are much higher than those of the uncertainty intervals of any individual model. The average deviation amplitude of the BMA(9) uncertainty interval is smaller than that of the uncertainty intervals of most of its nine individual models. From Figures 8 and 9, the similar conclusion can be concluded both in calibration and validation periods.

4.3. Comparison of BMA(3) and BMA(9). The results of both BMA(3) and BMA(9) in terms of the mean prediction and 90% uncertainty interval for the whole flow series are listed

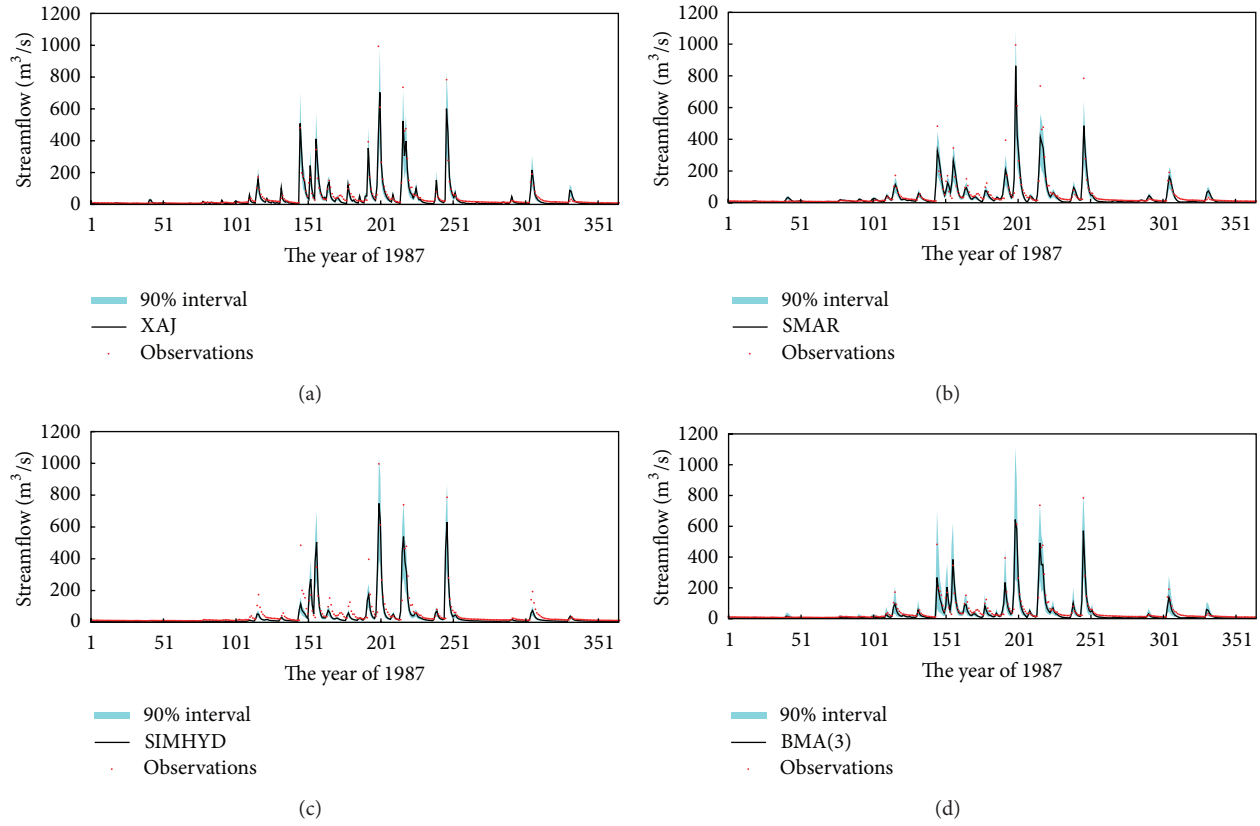

Figure 7: The mean prediction and 90% confidence interval of both BMA(3) and 3 individual models for the Mumahe catchment in 1987 during the validation period.

Table 1: Results of BMA(3) and its 3 individual models in the mean prediction as well as 90% uncertainty interval for the whole flow series.

Models	Mean prediction			90% uncertainty interval		
	R^2 (%)	DRMS	RE (%)	CR (%)	B (m^3/s)	D (m^3/s)
			Calibration period:			
XAJ	88.69	30.77	21.04	24.83	31.41	16.69
SMAR	87.69	32.11	**16.21**	32.83	32.80	17.21
SIM	80.73	40.17	31.51	14.83	**27.38**	22.33
BMA(3)	**90.68**	**27.92**	27.87	**40.72**	43.76	**16.06**
			Validation period:			
XAJ	85.77	29.22	17.79	24.28	24.66	14.09
SMAR	85.30	29.70	**14.19**	31.91	25.52	14.56
SIM	69.81	42.56	39.48	14.33	**18.40**	20.07
BMA(3)	**86.98**	**27.95**	30.72	**40.65**	36.71	14.13

Note: bolded values represent the best results.

in Table 3 for comparison. BMA(3) mean prediction has slightly better performance than BMA(9) mean prediction in terms of R^2 and DRMS in both calibration and validation periods, while BMA(3) mean prediction is slightly worse than BMA(9) mean prediction in terms of RE. For the uncertainty intervals, some findings are listed as follows: (1) in terms of CR, BMA(9) uncertainty interval is much higher than BMA(3) uncertainty interval in both calibration and validation periods; (2) in terms of B, BMA(9) uncertainty interval is obviously larger than BMA(3) uncertainty interval in both calibration and validation periods; (3) in terms of D, BMA(9) uncertainty interval performs slightly better than BMA(3) uncertainty interval in both calibration and validation periods.

Further, we compare the BMA(3) and BMA(9) mean predictions with respect to three flow ranges in Table 4. According to the values of three indices for mean prediction, BMA(3) mean prediction has better performance than BMA(9) mean

Table 2: Results of BMA(9) and its 9 individual models in the mean prediction and 90% uncertainty interval for the whole flow series.

Objective function	Models	Mean prediction			90% uncertainty interval		
		R^2 (%)	DRMS	RE (%)	CR (%)	B (m^3/s)	D (m^3/s)
Calibration period							
OF2 (MSEST)	XAJ	85.45	34.89	30.24	17.89	29.43	21.46
	SMAR	84.61	35.89	6.96	31.67	36.51	19.30
	SIM	80.73	40.17	31.51	15.39	**28.47**	22.67
OF3 (MSESRT)	XAJ	89.78	29.25	10.44	68.06	33.37	**11.75**
	SMAR	80.25	40.66	10.13	44.17	35.37	17.39
	SIM	72.42	48.05	**−5.82**	47.72	42.57	21.26
OF4 (MSELT)	XAJ	79.99	40.93	12.39	63.94	33.92	14.75
	SMAR	58.01	59.29	−9.22	42.28	43.45	28.32
	SIM	52.71	62.92	−41.07	38.89	55.51	26.93
BMA(9)		**90.49**	**28.22**	21.40	**91.11**	70.98	14.54
Validation period							
OF2 (MSEST)	XAJ	82.70	32.21	31.92	14.79	**21.56**	18.20
	SMAR	80.05	34.59	**0.66**	30.23	29.52	16.64
	SIM	69.81	42.56	39.48	20.84	24.43	22.32
OF3 (MSESRT)	XAJ	**88.52**	**26.25**	4.54	68.56	26.95	**9.62**
	SMAR	78.26	36.11	7.48	44.56	27.59	14.53
	SIM	71.09	41.64	8.98	53.86	27.69	16.47
OF4 (MSELT)	XAJ	77.25	36.94	8.74	63.07	26.68	11.85
	SMAR	43.43	58.25	−18.79	35.53	35.76	27.36
	SIM	72.27	40.79	−21.69	34.05	36.22	18.96
BMA(9)		84.54	30.46	25.42	**90.23**	55.91	13.20

Note: bolded values represent the best results.

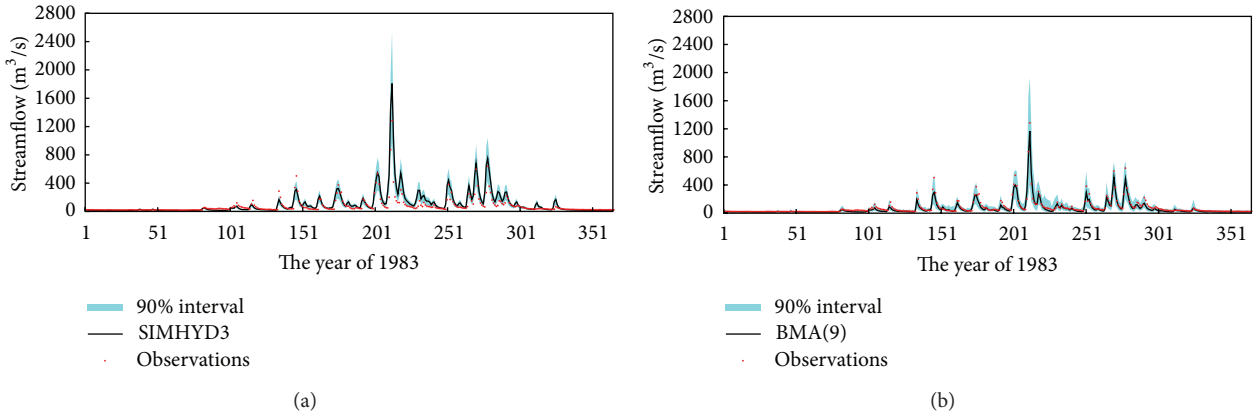

Figure 8: The mean prediction and 90% uncertainty interval of both BMA(9) and SIMHYD3 model (the SIMHYD with the objective function OF3) for the Mumahe catchment in 1983 during the calibration period.

prediction in high flow range, but has worse performance in medium and low flow ranges, during both calibration and validation periods. Then we compare the uncertainty intervals of BMA(3) and BMA(9) in three different flow ranges and have some findings as follows: (1) the CR value of BMA(9) uncertainty interval has absolute predominance in comparison with that of BMA(3) uncertainty interval for each of three flow ranges in both calibration and validation periods; (2) the B value of BMA(9) uncertainty interval is larger than that of BMA(3) uncertainty interval for all three

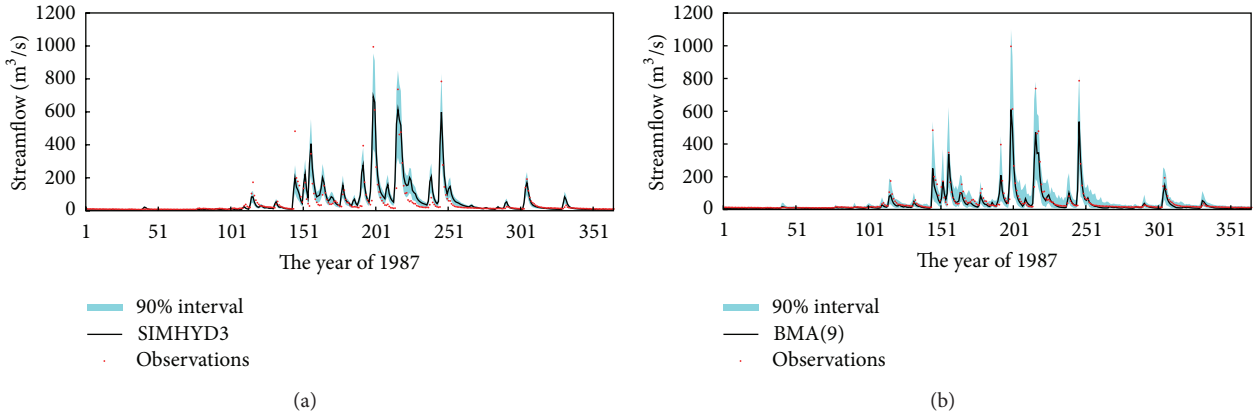

Figure 9: The mean prediction and 90% confidence interval of BMA(9) and SIMHYD3 model (the SIMHYD with the objective function OF3) for the Mumahe catchment in 1987 during the validation period.

Table 3: The comparison of BMA(3) and BMA(9) in the mean prediction and 90% uncertainty interval for the whole flow series.

Indices	Calibration		Validation	
	BMA(3)	BMA(9)	BMA(3)	BMA(9)
Mean Prediction				
R^2 (%)	90.68	90.49	86.98	84.54
DRMS (m^3/s)	27.92	28.22	27.95	30.46
RE (%)	27.87	21.40	30.72	25.42
90% uncertainty interval				
CR (%)	40.72	91.11	40.65	90.23
B (m^3/s)	43.76	70.98	36.71	55.91
D (m^3/s)	16.06	14.54	14.13	13.20

Table 4: The comparison of BMA(3) and BMA(9) in the mean prediction and 90% uncertainty interval for three flow ranges.

Indices	High flow		Medium flow		Low flow	
	BMA(3)	BMA(9)	BMA(3)	BMA(9)	BMA(3)	BMA(9)
Calibration period						
Mean prediction						
R^2 (%)	93.01	91.74	32.28	52.76	95.83	96.39
DRMS (m^3/s)	78.15	84.90	23.24	19.41	7.81	7.27
RE (%)	15.48	17.44	35.66	21.51	69.29	46.73
90% uncertainty interval						
CR (%)	88.74	92.05	45.91	91.32	27.40	90.75
B (m^3/s)	273.17	342.61	40.34	74.97	6.39	19.23
D (m^3/s)	59.78	63.66	18.33	15.44	6.21	5.02
Validation period						
Mean prediction						
R^2 (%)	89.00	85.47	22.03	41.82	93.66	94.94
DRMS (m^3/s)	92.51	106.35	19.01	16.42	6.87	6.14
RE (%)	22.49	27.68	31.35	17.66	67.48	45.11
90% uncertainty interval						
CR (%)	85.33	88.00	46.76	90.81	28.60	90.02
B (m^3/s)	252.88	282.17	34.97	61.22	7.19	18.45
D (m^3/s)	65.67	66.12	14.90	14.03	5.99	4.82

flow ranges in both calibration and validation periods; (3) the D value of BMA(9) uncertainty interval is slightly larger than that of BMA(3) in high flow range but smaller in medium and low flow ranges in both calibration and validation periods.

5. Conclusions

In this paper, the Bayesian Model Averaging (BMA) method is employed to construct a three-member predictions ensemble, denoted by BMA(3), and a nine-member predictions ensemble, denoted by BMA(9), for ensemble prediction as well as for prediction uncertainty analysis. There are three kinds of comparisons made in terms of both mean prediction and prediction uncertainty interval in this study: BMA(3) with its three individual models, BMA(9) with its nine individual models, and BMA(3) with BMA(9). In particular, we break observational flows into three different ranges for detailed comparison and analysis. The performance of two BMA schemes can be summarized as follows.

(1) In terms of mean predictions, BMA(3) performs generally better than any of its individual models. And BMA(9) mean prediction has generally higher accuracy than each of its individual model predictions. The comparison between BMA(3) and BMA(9) in mean predictions indicates that BMA(9) does not have any advantage compared to BMA(3) as far as the entire flow series is concerned. The performance of BMA(9) mean prediction is better than that of BMA(3) in both medium and low flow ranges, however, worse in the high flow range.

(2) In terms of the containing ratio for assessing the uncertainty intervals, the BMA(3) has a larger CR value than any of its individual models. And the containing ratio of BMA(9) uncertainty interval is also markedly larger than that of all its individual models when the CR value is calculated for the whole flow series. When the CR value is compared for different flow ranges, BMA(9) uncertainty interval performs better than its individual models in high, medium, and low flow ranges. In comparison with BMA(3), BMA(9) uncertainty interval also has absolute predominance in terms of CR.

(3) The average band-width B of BMA(3) uncertainty interval is larger than that of all its individuals. And the average band-width of BMA(9) uncertainty interval is even larger than that of BMA(3). It is found that, for uncertainty intervals, the increase of containing ratio is accompanied by the increase of band-width, which has already been pointed out by Xiong et al. [33].

(4) The average deviation amplitude D of BMA(3) uncertainty interval is generally smaller than the best individual in the ensemble. In terms of D, BMA(9) uncertainty interval also has a better performance than the best individual among its nine-member ensemble, especially in high flow range. Moreover, in terms of D, BMA(9) uncertainty interval performs better than BMA(3) uncertainty interval in medium and low flow ranges, but worse in the high flow range.

Based on this study, it is found that BMA is a particularly useful method for dealing with two issues. Firstly, when there are two or more competing models or methods available for the same problem, BMA can assess the relative performances of all models by assigning weights to each model or method and then produce more accurate mean prediction by weighted averaging of all predictions from those models or methods. Secondly, BMA can be used when there is uncertainty over control variables. The uncertainty intervals for both individual predictions and the BMA prediction can be derived when the distribution of the data is known or assumed.

Two issues from this study of BMA also need to be pointed out. The first is about the data transformation process. It is obvious that the daily flow data do not strictly obey the normal distribution even after the Box-Cox transformation. In fact, it is impossible to make every prediction from every model be normally distributed by using only a uniform transformation coefficient. Another problem is about the quality of the hydrological models chosen for combination. In this paper, the models employed here are all conceptual hydrological models. If better models are chosen as the ensemble members, then it is expected that the better results will come out of the BMA combination.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Grant nos. 51190094, 51079098), which is greatly appreciated. The comments and suggestions from the editor and the reviewers are very helpful in the improvement of the paper and are greatly appreciated.

References

[1] WMO, "Intercomparison of conceptual models used in hydrological forecasting," Operational Hydrology Report 7, WMO, Geneva, Switzerland, 1975.

[2] V. P. Singh, *Computer Models of Watershed Hydrology*, Water Resources Publications, 1995.

[3] D. J. Reid, "Combing three estimates of gross domestic products," *Economica*, vol. 35, pp. 431–444, 1968.

[4] J. M. Bates and C. W. J. Granger, "The combination of forecasts," *Operational Research Quarterly*, vol. 20, pp. 451–468, 1969.

[5] J. P. Dickinson, "Some statistical results in the combination of forecasts," *Operational Research Quarterly*, vol. 24, no. 2, pp. 253–260, 1973.

[6] A. Y. Shamseldin, K. M. O'Connor, and G. C. Liang, "Methods for combining the outputs of different rainfall-runoff models," *Journal of Hydrology*, vol. 197, no. 1–4, pp. 203–229, 1997.

[7] L. Xiong, A. Y. Shamseldin, and K. M. O'Connor, "A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi-Sugeno fuzzy system," *Journal of Hydrology*, vol. 245, no. 1–4, pp. 196–217, 2001.

[8] D. Madigan and A. E. Raftery, "Model selection and accounting for model uncertainty in graphical models using Occam's

window," *Journal of the American Statistical Association*, vol. 89, pp. 1535–1546, 1994.

[9] A. E. Raftery, "Bayesian model selection in social research," *Sociological Methodology*, vol. 25, pp. 111–163, 1995.

[10] D. Draper, "Assessment and propagation of model uncertainty," *Journal of the Royal Statistical Society B*, vol. 57, no. 1, pp. 45–97, 1995.

[11] C. Fernández, E. Ley, and M. Steel, "Benchmark priors for Bayesian model averaging," *Journal of Econometrics*, vol. 100, no. 2, pp. 381–427, 2001.

[12] K. Y. Yeung, R. E. Bumgarner, and A. E. Raftery, "Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data," *Bioinformatics*, vol. 21, no. 10, pp. 2394–2402, 2005.

[13] B. A. Wintle, M. A. McCarthy, C. T. Volinsky, and R. P. Kavanagh, "The use of bayesian model averaging to better represent uncertainty in ecological models," *Conservation Biology*, vol. 17, no. 12, pp. 1579–1590, 2003.

[14] K. H. Morales, J. G. Ibrahim, C.-J. Chen, and L. M. Ryan, "Bayesian model averaging with applications to benchmark dose estimation for arsenic in drinking water," *Journal of the American Statistical Association*, vol. 101, no. 473, pp. 9–17, 2006.

[15] G. Koop and L. Tole, "Measuring the health effects of air pollution: to what extent can we really say that people are dying from bad air?" *Journal of Environmental Economics and Management*, vol. 47, no. 1, pp. 30–54, 2004.

[16] A. E. Raftery, F. Balabdaoui, T. Gneiting, and M. Polakowski, "Using bayesian model averaging to calibrate forecast ensembles," Technical Report 440, Department of Statistics, University of Washington, 2003.

[17] V. Viallefont, A. E. Raftery, and S. Richardson, "Variable selection and Bayesian model averaging in case-control studies," *Statistics in Medicine*, vol. 20, no. 21, pp. 3215–3230, 2001.

[18] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging: a tutorial," *Statistical Science*, vol. 14, no. 4, pp. 382–417, 1999.

[19] M. A. Clyde, "Bayesian model averaging and model search strategies," in *Bayesian Statistics*, J. M. Bernardo, A. P. Dawid, J. O. Berger, and A. F. M. Smith, Eds., vol. 6, pp. 157–185, Oxford University Press, 1999.

[20] A. E. Raftery and Y. Zheng, "Discussion: performance of bayesian model averaging," *Journal of the American Statistical Association*, vol. 98, no. 464, pp. 931–938, 2003.

[21] S. P. Neuman, "Maximum likelihood Bayesian averaging of uncertain model predictions," *Stochastic Environmental Research and Risk Assessment*, vol. 17, no. 5, pp. 291–305, 2003.

[22] N. K. Ajami, Q. Duan, and S. Sorooshian, "An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction," *Water Resources Research*, vol. 43, no. 1, Article ID W01403, 2007.

[23] Q. Duan, N. K. Ajami, X. Gao, and S. Sorooshian, "Multi-model ensemble hydrologic prediction using Bayesian model averaging," *Advances in Water Resources*, vol. 30, no. 5, pp. 1371–1386, 2007.

[24] X. Zhang, R. Srinivasan, and D. Bosch, "Calibration and uncertainty analysis of the SWAT model using genetic algorithms and Bayesian model averaging," *Journal of Hydrology*, vol. 374, no. 3-4, pp. 307–317, 2009.

[25] K. Beven and A. Binley, "The future of distributed models: model calibration and uncertainty prediction," *Hydrological Processes*, vol. 6, no. 3, pp. 279–298, 1992.

[26] H. V. Gupta, K. J. Beven, and T. Wagener, "Calibration and uncertainty estimation," in *Encyclopedia of Hydrological Sciences*, John Wiley and Sons, Chichester, UK, 2003.

[27] L. Marshall, D. Nott, and A. Sharma, "A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling," *Water Resources Research*, vol. 40, no. 2, Article ID W02501, 2004.

[28] J. A. Vrugt, H. V. Gupta, W. Bouten, and S. Sorooshian, "A Sshuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters," *Water Resources Research*, vol. 39, no. 8, 2003.

[29] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*, Methuen, London, UK, 1975.

[30] F. H. S. Chiew, M. C. Peel, and A. W. Western, "Application and testing of the simple rainfall runoff model SIMHYD," in *Mathematical Models of Small Watershed Hydrology and Applications*, V. P. Singh and D. Frevert, Eds., pp. 335–366, Water Resources Publications, 2002.

[31] Q. Duan, S. Sorooshian, and V. Gupta, "Effective and efficient global optimization for conceptual rainfall-runoff models," *Water Resources Research*, vol. 28, no. 4, pp. 1015–1031, 1992.

[32] L. Oudin, V. Andréassian, T. Mathevet, C. Perrin, and C. Michel, "Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations," *Water Resources Research*, vol. 42, no. 7, Article ID W07410, 2006.

[33] L. Xiong, M. Wan, X. Wei, and K. M. O'Connor, "Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation," *Hydrological Sciences Journal*, vol. 54, no. 5, pp. 852–871, 2009.