

# Multi-Model Ensemble Simulation of Flood Events using Bayesian Model Averaging

Ruirui Zhu<sup>1,2</sup> Hongxing Zheng<sup>2</sup> Enli Wang<sup>2</sup> Weimin Zhao<sup>3</sup>

*1. Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China*

*2. CSIRO land and water, Canberra, ACT 2601, Australia*

*3. Hydrological Bureau, Yellow River Water Resources Committee, Zhengzhou, 450004, China  
Email: [Ruirui.zhu@csiro.au](mailto:Ruirui.zhu@csiro.au)*

**Abstract:** Flood event simulation using hydrological model is subject to various uncertainties. Multi-model ensemble simulation approach has proved to improve forecasting uncertainty by reducing the systematic bias in comparison with its single model. In this paper, the Bayesian Model Averaging (BMA) approach, a statistical scheme based on multi-model ensemble, was applied for flood prediction and uncertainty estimation of flood event predictions in Qingjianhe Catchment. Five hydrological models including GR4J, HYMOD, Simhyd, XAJ and modified SCS were employed and calibrated with two objective functions NSE and  $WR^2$ . Ten ensemble simulations were then used for further BMA analysis. The results showed that the five hydrological models performed reasonably well in QJH catchment, but with significant difference in the simulated hydrographs. The modified SCS model performed the best among the five models in term of NSE and  $WR^2$ . The BMA weights for the model prediction were roughly consistent with the model performance. GR4J model weighted higher than other models. Predictions with BMA median performed less well than those from the best individual model SCS, especially for peak flows. However, BMA gave more reliable predictions. For most flood events with different recurrent periods in the study catchment, the 50% confidence interval seemed sufficient to bracket the observed flood discharge. It indicates that BMA approach is helpful in reducing uncertainties, thereby increasing the level of confidence in prediction results. The prediction uncertainty quantified via BMA can be very helpful for decision makers to develop flood management strategies.

**Keywords:** *flood event, hydrological model, ensemble simulation, BMA*

## 1 INTRODUCTION

Hydrological models have been developed and used in water resources assessment, flood management and forecasting for decades. However, quantification of the uncertainties in model predictions remains a challenge due to variations in measurement errors, equifinality and model structural inadequacies (Vrugt and Robinson, 2007). Hydrological models can be conceptual or physically-based, lumped or distributed. Unfortunately, no matter how sophisticated model is, there remain conditions and situations in which the model is unsuitable to give an accurate forecasting (Parrish *et al.*, 2012). Due to the incomplete representation of hydrology processes in the models, predictions from a single model often lead to over-confidence and significant bias of the forecast (Hoeting *et al.*, 1999; Neuman, 2003; Raftery *et al.*, 2003, 2005; Vrugt and Robinson, 2007; Parrish *et al.*, 2012).

Multi-model ensemble simulation has therefore become increasingly popular, for it has the capability to overcome the limitations of the single-model approach and improve the estimation of model uncertainties (Cane *et al.*, 2013). With the ensemble approach, results from multiple models need to be combined. In some earlier studies, the weight of each model result was either treated as equal or determined using linear regression technique. Neither of the treatments connected the weights of models to model performance. To overcome this deficiency, an alternative approach, named Bayesian Model Averaging (BMA), was proposed by Hoeting *et al.* (1999). BMA weights a model by its performance and likelihood of predicting the observation, resulting in a probabilistic forecast. Various case studies show that BMA can produce more reliable predictions than other multi-model methods (Viallefont *et al.*, 2001; Raftery *et al.*, 2003, 2005; Duan *et al.*, 2007; Vrugt and Robinson, 2007).

The objective of this paper is to explore how BMA can improve accuracy and reliability of flood predictions when multiple conceptual rainfall-runoff models are used. The performance of BMA will be investigated for forecasting flood events with different reoccurrence period in a case study in the Qingjianhe catchment, at the middle reaches of the Yellow River (China).

## 2 Study area and data

The Qingjianhe catchment (QJH) is located at the middle reaches of the Yellow River in China. It has an area of 4080km<sup>2</sup> with 167.8km of river channel within the catchment (Fig.1). The catchment belongs to arid or semi-arid climate region with mean annual temperature between 9~10°C and annual precipitation amount of 477.6mm. More than 75% of annual precipitation falls during the period from June to September. The rainfall in the rainy season was characterized by heavy storm in limited area, showing high density and low duration. Geomorphologically, the QJH catchment belongs to the Loess Hilly region with numerous of gullies. The density of gully is about 4~7km/km<sup>2</sup>. The vegetation in the catchments mainly consists of grass and shrubs. The low vegetation cover has been considered as an important factor for serious soil erosion in the region. The area identified with high erosion rate amounts to 4,006km<sup>2</sup>, with erosion coefficient larger than 10,000t/km<sup>2</sup>.

In this study, hourly precipitation from 16 gauge stations and hourly runoff data collected at Yanchuan station were obtained from Hydrological Bureau of the Yellow River Conservancy Commission (Fig.1). Daily potential evapotranspiration (PET) was estimated based on routine meteorological observations from three stations near the catchment using Penman-Monteith method recommended by FAO (1998), and was then used to estimate hourly PET using a simple circular function (sin). All the precipitation, potential evapotranspiration and runoff data covered most of the period from 1980 to 2007, but with missing data in 1987~1991. In order to test BMA scheme for flood predictions, flood events with different recurrence periods or accumulated probability ( $\alpha$ ) are selected. The flood events selected are  $\alpha=1\%$ 、 $5\%$ 、 $10\%$ 、 $25\%$ 、 $50\%$ 、 $75\%$  and  $90\%$  respectively, where  $\alpha=1\%$  represents the flood event with the largest peak flow.

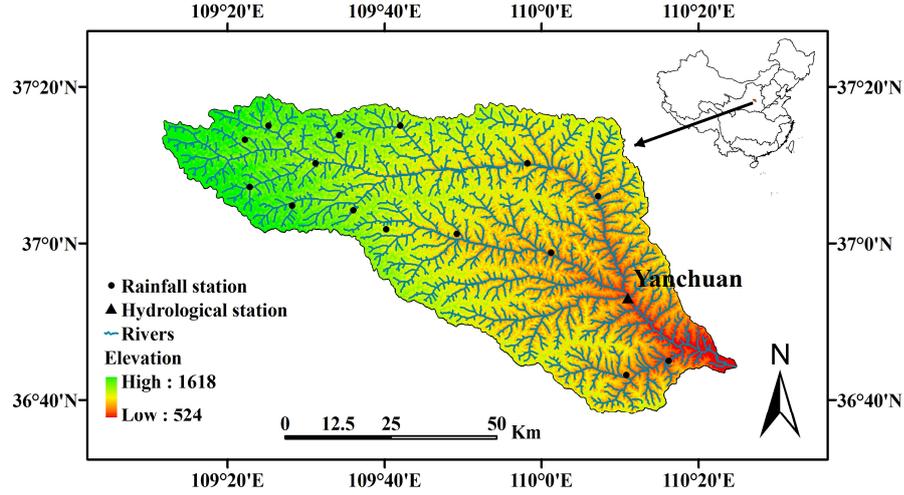


Figure 1 Sketch map of the Qingjianhe catchment

### 3 Methodology

#### 3.1 Bayesian Model Averaging

Consider a quantity  $y$  to be forecasted,  $D = \{y_1, y_2, \dots, y_T\}$  to be training data with length  $T$ , and  $f = \{f_1, f_2, \dots, f_k\}$  to be the ensemble of all considered model predictions.  $k$  is the number of ensemble members. According to the law of total probability, the probability density function of the BMA prediction for  $y$  can be expressed as:

$$p(y|D) = \sum_{i=1}^k P(f_i|D) p(y|f_i, D) \quad (1)$$

Where  $i$  is the index of each model member;  $p(y|f_i, D)$  is conditional probability distribution function of  $y$  on  $f_i$  and  $D$ ;  $P(f_i|D)$  is the posterior distribution of model prediction  $f_i$  or the relative likelihood of model  $f_i$  being correct given the training data  $D$ . If we denote  $w_i = P(f_i|D)$ , the sum of  $w_i$  is equal to 1. Correspondingly, mean and variance of BMA prediction can be expressed as:

$$E[y|D] = \sum_{i=1}^k P(f_i|D) \int_{-\infty}^{+\infty} y p(y|f_i, D) dy = \sum_{i=1}^k w_i f_i \quad (2)$$

$$Var[y|D] = \sum_{i=1}^k w_i \left( f_i - \sum_{i=1}^k w_i f_i \right)^2 + \sum_{i=1}^k w_i \sigma_i^2 \quad (3)$$

Where  $\sigma_i^2$  is variance associated with model prediction  $f_i$  with respect to training data  $D$ . Eq. (2) shows that the BMA mean is the average of predictions weighted by the likelihood  $P(f_i|D)$ . Actually, the likelihood  $P(f_i|D)$  describes how well model predictions  $f_i$  match the training data  $D$ . The higher the likelihood, the better is the model performance. Variance is a measure of the prediction uncertainty. Variance of BMA predictions defined in Eq. (3) consists of two terms, the first representing the ensemble spread error and the second representing within-ensemble prediction variance.

Usually,  $p(y|f_i, D)$  is expressed as a normal distribution  $N(f_i, \sigma_i^2)$  for computational convenience. Successful implementation of BMA approach requires specification of weights  $w_i$  and stand variances  $\sigma_i$  of each ensemble member, for which the maximum likelihood estimation method can be used (Raftery *et al.*, 2003, 2005). Let  $\theta = \{w_i, \sigma_i, i=1, 2, \dots, k\}$ , the log-likelihood function can be approximated as:

$$l(\theta) = \log \left( \sum_{i=1}^k w_i \times p(y|f_i, D) \right) \quad (4)$$

Obviously, it is impossible to obtain analytical resolution of  $\theta$ . Thereby, we performed Expectation-Maximization algorithm (EM) recommended by Raftery *et al.* (2005) to search the optimal value of  $\theta$ . It starts with an initial guess  $\theta^0$ , and then EM algorithm alternates between expectation step and maximization step to update  $\theta^{iter}$ , where *iter* is updating number. Expectation step and maximization step are repeated continually until  $|l(\theta^{iter}) - l(\theta^{iter-1})|$  is less than or equal to a pre-specified tolerance level. Detailed description about EM algorithm is presented by McLachlan and Krishnan (1997). After convergence of EM algorithm, we will have determined weights for each model. Therefore, probability density function (PDF) of the BMA predictions (i.e. Eq.1) can be derived and its mean and variance can also be estimated through Eq.2~Eq.3, respectively. With PDF, uncertainty intervals can be specified. For example, 50% confidence interval is with the range of 25% and 75% quartiles.

### 3.2 Hydrological models

To test BMA scheme for flood events prediction, an ensemble of competing predictions were produced from using multiple hydrological models, including GR4J (Perrin *et al.*, 2003), HYMOD (Boyle *et al.*, 2001), Simhyd (Chiew *et al.*, 2002), Xinanjiang (Zhao *et al.* 1980) and a modified SCS model. These five models are listed in the order of increasing complexity. The number of tunable parameter in the five models is 4, 5, 9, 14 and 14, respectively. The modified SCS model is an event-oriented rainfall-runoff model, in which runoff generation is calculated using SCS Curve number and flow concentration follows the Clark unit-hydrograph (1945). The parameters of the five hydrological models were optimized by using the Particle Swarm Optimization (PSO) approach (Eberhart and Kennedy, 1995). The Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) and  $WR^2$  were adopted as the objective functions in the procedure of calibration, which are calculated as:

$$NSE = 1 - \frac{\sum_{i=1}^n (q_{ri} - q_{ci})^2}{\sum_{i=1}^n (q_{ri} - \bar{q}_{ri})^2} \quad (5)$$

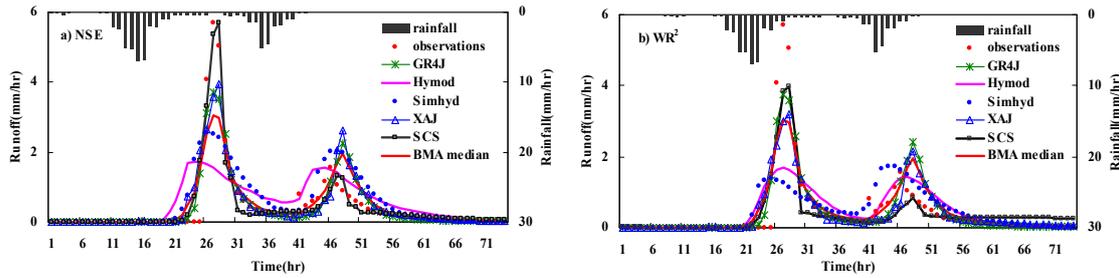
$$WR^2 = R^2 - \frac{\left| \sum_{i=1}^N q_{ci} - \sum_{i=1}^N q_{ri} \right|}{\sum_{i=1}^N q_{ri}} \quad (6)$$

Where  $q_{ri}$  and  $q_{ci}$  are observed and simulated streamflow respectively,  $\bar{q}_{ri}$  and  $\bar{q}_{ci}$  are the means of  $q_{ri}$  and  $q_{ci}$ . The larger *NSE* or  $WR^2$  value, the better is the model performance. The combination of five models and two objective functions produces a 10 ensemble simulations of the specific flood event in QJH catchment. This ensemble simulation is the basis for implementing BMA described in the previous section.

## 4 Results

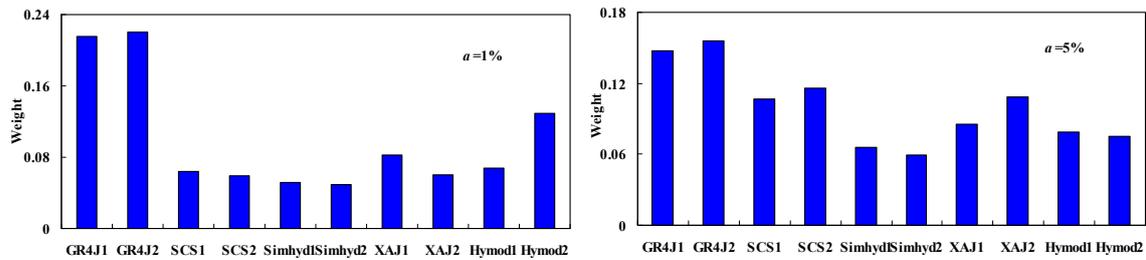
### 4.1 BMA weights of individual models

Fig. 2 presents the performance of the five models calibrated using either *NSE* or  $WR^2$ . The BMA median is also plotted for comparison purpose. Each model has strength to capture some aspects of hydrology process. Even though one hydrological model was calibrated using different objective functions (*NSE* or  $WR^2$ ) separately, it gave significantly different results (Fig.2). In general, hydrographs generated from models calibrated using *NSE* fit better with the observed ones than those from models calibrated using  $WR^2$ , especially for the peak flow. For example, for  $a=1\%$  flood event, SCS model obtained *NSE* value of 0.91 while  $WR^2$  of 0.85. Furthermore, peak flow predicted from models calibrated using *NSE* was more close to the observed value than that from models calibrated with  $WR^2$ . *NSE* and  $WR^2$  statistics further confirmed difference among different models. Mean *NSE* and  $WR^2$  values of SCS model were highest ( $>0.67$ ) among the five models, which indicates SCS model was the best out of the five models in flood simulation in QJH.



**Figure 2** Performance of five individual hydrological models as compared to BMA median predictions. (a) and (b) show the results from models calibrated using different objective functions (NSE vs  $WR^2$ ).

One of the assumptions in the BMA approach is that BMA weights should reflect relative model skill. This means the highest BMA weight is for a model which produces the lowest quadratic forecast errors over the complete hydrograph. Fig.3 summaries the BMA weights for the ten ensemble members in two flood event cases ( $\alpha=1\%$ ,  $\alpha=5\%$ ). Indeed, the weights of the models were significantly different, and the GR4J model weighted highly compared to other models, even higher than the “best-fit” SCS model. This result shows that the rank order of BMA weights does not completely track order of the NSE or inverse order of HMLE. This behavior could be explained by the complementary correlations between individual forecast in the ensemble. In the optimization of BMA weights, best single forecast may be less than combination of highly complementary forecasts, which therefore lead to de-weighting of the best single forecaster and over-weighting of the poorer single forecasters. This phenomenon is also consistent with other BMA studies (Vrugt and Robinson, 2007; Wöhling *et al.*, 2008).



**Figure 3** BMA weights of each individual model. 1 and 2 after the model name denote model calibration using objective function of NSE and  $WR^2$  respectively

#### 4.2 BMA predictions

BMA median as the weighted-average of predictions based on individual models also shows good agreement with the observations (Fig.2). However, it was not as good as the best individual model (SCS model), especially for peak flow prediction. However, performance of BMA median was much better than other models. It implies that BMA approach does not necessarily improve predictive capabilities, which is consistent with previous study (Vrugt and Robinson, 2007; Duan *et al.*, 2007), but, it can provide deterministic predictions comparable to best individual model.

One important feature of the BMA approach is that it can produce probabilistic distribution of the quantity to be forecasted from the competing individual deterministic simulations. As shown in Eq.3, uncertainty intervals can be derived with posterior probability distribution function  $p(y|D)$ . For most flood events the 50% confidence interval seems sufficient to bracket the observed flood flow very well except for the flood event with  $\alpha=90\%$  (Fig.4). In particular, for flood event with  $\alpha=50\%$ , a 20% confidence interval was sufficient to fully bracket the observations. This implies that estimated probability density function (PDF) of forecasted streamflow in QJH was quite consistent with the observations.

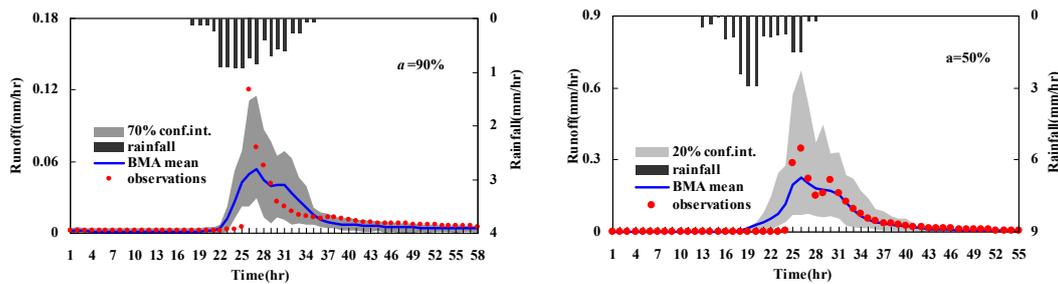


Figure 4 Median and 50% confidence interval of BMA predictions for the two flood events.

## 5 Conclusions

Flood event simulation using hydrological model is subject to various uncertainties. Multi-model ensemble simulation approach has proved to improve forecasting uncertainty by reducing the systematic bias as compared to a single model approach. The results in this paper, derived using the Bayesian Model Averaging (BMA) approach together with predictions from five hydrological models, indicated that the BMA weights for the model predictions were roughly consistent with the model performance. Although the predictions with BMA median performed less well than those from the best individual model, the ensemble results from BMA gave more reliable predictions. For most flood events with different recurrent periods in the study catchment, the 50% confidence interval seemed sufficient to bracket the observed flood discharge. It indicates that BMA approach is helpful in reducing uncertainties, thereby increasing the level of confidence in prediction results. The prediction uncertainty quantified via BMA can be very helpful for decision makers to develop flood management strategies.

## ACKNOWLEDGMENTS

This study is financially supported by the Public Welfare Projects of China Ministry of Water Resources (No: 200901016) and China National Natural Science Foundation (No: 41101032).

## REFERENCES

- Boyle, DP., Gupta, HV., Sorooshian, S., Koren V., Zhang, Z., and Smith M.(2001). Toward improved streamflow forecast: value of semidistributed modeling. *Water Resources Research*, 37(11):2749-59.
- Chiew, F.H.S., Peel, M.C. and Western, A.W. (2002), Application and testing of the simple rainfall-runoff model SIMHYD, In: *Mathematical Models of Small Watershed Hydrology and Applications* (Editors: V.P.Singh and D.K. Frevert), Water Resources Publication, Littleton, Colorado, USA, pp.335-367.
- Clark, C.O. (1945). Storage and the Unit Hydrograph. *Transactions of the American Society of Civil Engineers*, 110, 1419-1446.
- D. Cane, S. Ghigo, D. Rabuffetti, and M. Milelli (2013). Real-time flood forecasting coupling different postprocessing techniques of precipitation forecast ensembles with a distributed hydrological model. The case study of May 2008 flood in western Piemonte, Italy. *Nat. Hazards Earth Syst. Sci.*, 13, 211-220.
- Duan, Q., Ajami, N. K., X. Gao, and S. Sorooshian (2007). Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.*, 30(5), 1371-1386.
- Eberhart, R. C., and Kennedy, J.(1995). A new optimizer using particle swarm theory. *Proceeding of the Sixth International Symposium on Micro Machine and Human Science*. Nagoya, Japan, 39-43. Piscataway, NJ: IEEE Service Center.
- Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian modeling averaging: A tutorial. *Stat. Sci.*, 14(4), 382-417.
- McLachlan, G.J., and Krishnan, T. (1997). *The EM algorithm and extensions*. New York: Wiley.
- Nash, J. E., and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models. Part I: A discussion of principles. *Journal of Hydrology*, 27(3), 282-290.
- Neuman, S. P. (2003). Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochastic Environ. Res. Risk Assess.*, 17, 291-305, doi:10.1007/800477-003-0151-7.

- Parrish, M. A., H. Moradkhani, and C. M. DeChant, (2012). Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation, *Water Resources Research.*, 48, W03519, doi:10.1029/2011WR011116.
- Perrin, C., Michel, C., and Andre'assian, V. (2003). Improvement of a parsimonious model for stream flow simulation, *Journal of Hydrology*, 279(1-4):275-289
- Raftery, A.E., Balabdaoui, F., Gneiting, T., and Polakowski, M. (2003). Using Bayesian Model Averaging to calibrate forecast ensembles. Technical Report No.440, Department of Statistics, University of Washington.
- Raftery, A.E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 113, 1155-1174.
- Viallefont, V., Raftery, A.E., Richardson, S. (2001). Variable selection and Bayesian model averaging in epidemiological case-control studies. *Statist Med.*, 20, 3215-3230.
- Vrugt, J. A., and B. A. Robinson (2007). Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian Model Averaging. *Water Resource Research*, 43, W01411, doi: 10.1029/2005WR004838.
- Wöhling, T., and Vrugt, J. A. (2008). Combining multiobjective optimization and Bayesian model averaging to calibrate forecast ensembles of soil hydraulic models. *Water resources research*, 44, W12432, doi:10.1029/2008WR007154.
- Zhao, R.J., Zhang, Y.L., and Fang, L.R. (1980). The Xinanjiang model. *Hydrological Forecasting Proceeding Oxford Symposium*, IASH-AISH Publication.no.129, Washington, DC, 351-356.