

Evaluation of the BMA probabilistic inflow forecasts using TIGGE numeric precipitation predictions based on artificial neural network

Yixuan Zhong, Shenglian Guo, Huanhuan Ba, Feng Xiong, Fi-John Chang and Kairong Lin

ABSTRACT

Reservoir inflow forecasting is a crucial task for reservoir management. Without considering precipitation predictions, the lead time for inflow is subject to the concentration time of precipitation in the basin. With the development of numeric weather prediction (NWP) techniques, it is possible to forecast inflows with long lead times. Since larger uncertainty usually occurs during the forecasting process, much attention has been paid to probabilistic forecasts, which uses a probabilistic distribution function instead of a deterministic value to predict the future status. In this study, we aim at establishing a probabilistic inflow forecasting scheme in the Danjiangkou reservoir basin based on NWP data retrieved from the Interactive Grand Global Ensemble (TIGGE) database by using the Bayesian model averaging (BMA) method, and evaluating the skills of the probabilistic inflow forecasts. An artificial neural network (ANN) is used to implement hydrologic modelling. Results show that the corrected TIGGE NWP data can be applied sufficiently to inflow forecasting at 1–3 d lead times. Despite the fact that the raw ensemble inflow forecasts are unreliable, the BMA probabilistic inflow forecasts perform much better than the raw ensemble forecasts in terms of probabilistic style and deterministic style, indicating the established scheme can offer a useful approach to probabilistic inflow forecasting.

Key words | artificial neural network (ANN), Bayesian model averaging (BMA), Danjiangkou reservoir, probabilistic flood forecasting, TIGGE database

Yixuan Zhong

Shenglian Guo (corresponding author)

Huanhuan Ba

Feng Xiong

State Key Laboratory of Water Resources and Hydropower Engineering Science, Hubei Collaborative Innovation Center for Water Resources Security, Wuhan University, Wuhan, China
E-mail: slguo@whu.edu.cn

Fi-John Chang

Department of Bioenvironmental Systems Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, Taiwan, ROC

Kairong Lin

Department of Water Resources and Environment, Sun Yat-Sen University, Guangzhou, China

INTRODUCTION

Flood is widely regarded as the most frequent and severest natural disasters worldwide, which brings countless deaths each year (ICOLD 2006; Kussul *et al.* 2008; Uddin *et al.* 2013; Gao *et al.* 2015; Kwon & Kang 2016; Chen & Singh 2017). To mitigate the negative impacts of flood disasters, numerous structural measures (e.g. dams, levees, retention basins and detention basins) and non-structural measures (e.g. flood forecasting and alarms, reservoir operation and flood insurance) are utilized in practice, among which flood forecasting takes a vital role since it can offer flood information beforehand and hence set aside time for flood

operation decision making and evacuation activity (Guo *et al.* 2004).

Flood forecasting is an important tool for reducing vulnerabilities and flood risk and forms an important ingredient of the strategy to 'live with floods', thereby contributing to national sustainable development (WMO 2010). Traditional flood forecasting is usually implemented by simply inputting precipitation predictions and other necessary elements to drive the hydrologic model. Thus the forecasting results are deterministic and can hardly describe the uncertainty of flood events. To overcome the shortages

of deterministic flood forecasting, the probabilistic flood forecasting based on a probability distribution function to describe the future discharge has been proposed and gained great focus by many authors in recent years (Van Steenbergen *et al.* 2015; Bellier *et al.* 2016; Hardy *et al.* 2016; Liu *et al.* 2016; Li *et al.* 2017; Todini 2017). As a conclusion of previous studies, probabilistic flood forecasting outperforms deterministic flood forecasting in several aspects: (1) probabilistic flood forecasting can quantify uncertainty and enable the decision makers to hedge against the probability of forecast results; (2) probabilistic flood forecasting usually has a longer lead time and can provide more timely flood information; and (3) probabilistic flood forecasting usually has higher skills than deterministic forecasting.

The basic requirements of flood forecasting usually consist of specific precipitation predictions and calibrated hydrologic models. With flood protection and awareness continually rising on the political agenda, a strong demand has generated for achieving high quality precipitation prediction to obtain flood forecasts with sufficient lead time (Bao *et al.* 2011). Since the early 1900s, numerical weather prediction (NWP) techniques have been developed and offered a new approach to generate precipitation prediction (Bjerknes 1904). Cloke & Pappenberger (2009) reviewed the importance of NWP in flood forecasting and emphasized that NWP must be used for achieving reliable 2–15 days ahead flood forecasting. The Interactive Grand Global Ensemble (TIGGE) is part of the Hydrologic Ensemble Prediction Experiment (HEPEX) and ‘a project designed to develop, demonstrate and evaluate a multi-model, multi-analysis and multi-national ensemble prediction system’, intending to provide reliable, skillful, open source hydrologic forecast procedures by using emerging weather and climate ensemble forecast techniques (WMO 2005). By coupling the TIGGE precipitation predictions with hydrologic models, the lead times can be extended from a few hours to several days, thus holding great benefits for flood protection and decision making. Huang *et al.* (2010) conducted uncertainty assessment by probabilistic flood forecasts with the TIGGE data and applied the cost-benefit analysis based on the forecasting results. Zhao *et al.* (2011) compared the TIGGE data from different meteorological centers and found that the ECMWF performs better than the CMA and NCEP in the upper Huaihe basin. Bao *et al.* (2011)

used the TIGGE data to drive the Xiananjiang model and conducted probabilistic forecasts at Xixian station in the Huaihe basin. The results showed that by coupling the TIGGE data with hydrologic models, flood events could be efficiently identified several days ahead. Liu & Xie (2014) used Bayesian model averaging (BMA) to the TIGGE precipitation predictions in the Huaihe basin and successfully applied the probabilistic precipitation to heavy-rain warning. Fan *et al.* (2015) used ensemble flood forecasts of the TIGGE data to verify hydropower reservoir inflows and showed that ensemble flood forecasts are more consistent than deterministic ones in terms of sequential decisions. Coustau *et al.* (2015) assessed the impacts of the ECMWF NWP products on streamflow forecasts and revealed that the atmospheric forcing is especially significant for streamflow forecasts of small catchments. These studies give evidence that the TIGGE data are valuable for flood forecasting and disaster relief decision making.

Among the many types of research on the TIGGE database, few have coupled it with a hydrologic model and evaluated its ability for generating probabilistic inflow forecasts in large catchments. Thus the objective of this paper is to establish a probabilistic inflow forecasting scheme in Danjiangkou basin based on the TIGGE data and the BMA method, and evaluate the skills of the probabilistic inflow forecasts. Three TIGGE NWP models, i.e. CMA, NCEP and ECMWF, were selected for the case study. The raw NWP data were downloaded and post-processed automatically from the official website of the ECMWF (<http://apps.ecmwf.int/datasets/data/tigge>). The remainder of this paper is organized as follows. The next section gives a brief introduction of Danjiangkou basin and presents the data used in this study. The following section shows the main methodologies, and the application is described in the Results and discussion section. The final section displays the main conclusions.

STUDY AREA AND DATA

Study area

The Danjiangkou reservoir is a multipurpose reservoir in the Hanjiang River located to the west of Hubei province, China. The Danjiangkou reservoir basin (32.6°36′–33°48′N,

110°59'–111°49'E) has an area of 95,217 km², accounting for about 60% of the Hanjiang basin as shown in Figure 1. The annual mean inflow runoff at its dam site is about 36.81 billion m³, of which 61.4% concentrates during the flood season from July to October. The reservoir is also the water source of China's central line project of south-to-north water diversion, which benefits 50 million citizens suffering from water shortages (Liang et al. 2017).

The topography of the basin is very complex, of which the mountainous area accounts for 85% of the land area. The basin belongs to a sub-tropic monsoon climatic region with transitional climatic characteristics. The annual precipitation varies from 850 to 1062 mm and the mean discharge is about 1,250 m³/s (Li et al. 2009). The mean flood concentration time of the basin is about 3 days. With comprehensive benefits of flood control, hydropower generation and water supply, the 1–3 d lead time inflow forecasts have drawn special attention from the reservoir managers, which will directly influence the reservoir operation decision process.

Data

Two types of data are used in this study, including precipitation and reservoir inflow records. The daily inflow data of the Danjiangkou reservoir from 2008 to 2013 (2192

days in total) were collected from the Water Conservancy Bureau of the Hanjiang River. The daily areal precipitation from 2008 to 2013 was derived from the China Gauge-based Daily Precipitation Analysis (CGDPA) database by grid-based averaging method. The CGDPA data can be downloaded from the meteorological data center of China Meteorological Administration (CMA, <http://data.cma.cn>) and have proved to be reliable for providing high quality daily accumulated rainfall amount ending at 00:00 UTC over the mainland of China (Xie et al. 2007; Shen et al. 2010; Sun et al. 2016).

The NWP data from 2008 to 2013 in the Danjiangkou reservoir basin were downloaded from the TIGGE database retrieval website at ECMWF (<http://apps.ecmwf.int/datasets/data/tigge>). A summary of the three TIGGE models is presented in Table 1. All the NWP data were spatially interpolated from the original NWP data automatically by the website server at a spatial resolution of 0.5° × 0.5°. The original time interval was 6 h and the daily precipitation predictions were obtained by summing the four successive predictions. As shown in Table 1, the ECMWF model had the most ensemble members, which were more than double of the NCEP and three times that of the CMA. The areal precipitation predictions of the Danjiangkou basin were calculated using a grid-based averaging method.

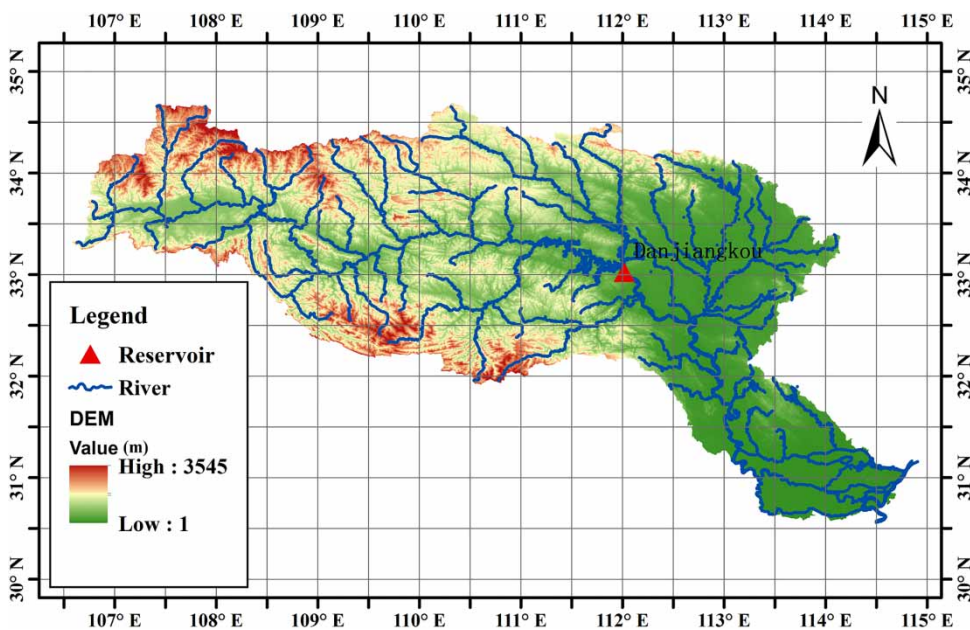


Figure 1 | Terrain map of the Hanjiang River basin and Danjiangkou reservoir.

Table 1 | Basic information of the TIGGE data used in this study

Property	ECMWF	NCEP	CMA
Administration location	Europe	United States	China
Number of ensemble members	51	21	15
Perturbation method	Singular vectors	Ensemble transform	Bred vectors
Maximum lead time	15 days	16 days	10 days

METHODOLOGY

In this study, a probabilistic forecasting scheme was constructed to derive probabilistic inflow forecasts in the Danjiangkou reservoir (see Figure 2). Three artificial neural network (ANN) models were calibrated first, the input combinations of which were optimized with one of the state-of-the-art input selection methods, i.e. the Gamma Test. After that, ensemble forecasts of the Danjiangkou reservoir with three lead times were generated using both the antecedent flood and precipitation records and also the corrected TIGGE NWP data. The BMA method was applied to generate probabilistic forecasts with the ensemble forecasts. The probabilistic forecasts were evaluated and compared with the raw ensemble forecasts under a coherent set of evaluation criteria. The methods employed in this study are briefly addressed in the following.

Gamma test

The Gamma test (GT), introduced by Agalbjorn *et al.* (1997), is a data analysis technique for assessing the extent to which a given set of M data points can be modeled by an unknown smooth non-linear function. The Gamma statistic (T) is an estimator of the model output's variance that cannot be accounted for through a smooth data model. The most attracting feature of the GT method is that it is sufficient for input selection without knowing the input-output system structure. Recent studies revealed that the GT method is an effective way of identifying ANN input and can thus reduce the input dimensions as well as produce precise output of ANNs (Moghaddammia *et al.* 2009; Noori *et al.* 2011; Chang *et al.* 2013, 2014). Therefore, the GT method is used to determine the combinations of antecedent inflow and precipitation for the ANN models in this study. A brief introduction of conducting GT follows:

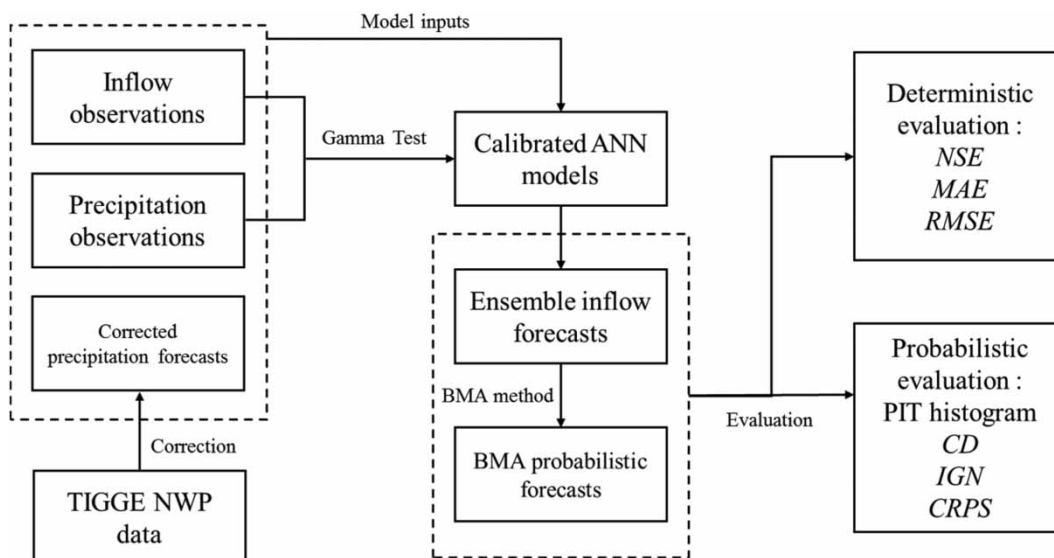


Figure 2 | Sketch diagram of study process.

Suppose the dataset is given in the form of:

$$[\mathbf{X}, Y] = \{(\mathbf{x}_i, y_i)\}, i = 1, 2, \dots, n \quad (1)$$

where \mathbf{X} and Y are the input data matrix and the output data vector with the same sample size of n , respectively; $\mathbf{x}_i = [x_{1,i}, x_{2,i}, \dots, x_{m,i}]$ is an m -dimensional input vector for the i th sample and y_i is the output scalar for the i th sample.

The underlying relationship of this input-output system for \mathbf{X} and Y is defined as:

$$y_i = f(x_{1,i}, x_{2,i}, \dots, x_{m,i}) + r \quad (2)$$

where $f(\cdot)$ is an unknown smooth function, r denotes the random error or noise. Then, the GT is assessed based on the k th nearest neighbor for each input vector $\mathbf{x}_{i,k}$ and output $y_{i,k}$ with Delta function:

$$\delta_n(k) = \frac{1}{n} \sum_{i=1}^n |\mathbf{x}_{i,k} - \mathbf{x}_i|^2, k \in [1, p] \quad (3)$$

$$\gamma_n(k) = \frac{1}{2n} \sum_{i=1}^n |y_{i,k} - y_i|^2, k \in [1, p] \quad (4)$$

where $|\cdot|$ denotes the Euclidean distance; the maximum neighboring distance p is set as 10 (Chang et al. 2013); $y_{i,k}$ denotes the y value corresponding to $\mathbf{x}_{i,k}$ in the dataset.

Finally, the Gamma statistic Γ can be estimated using linear regression with the equation:

$$\gamma = A\delta + \Gamma \quad (5)$$

where A is the regression coefficient. For more detailed information about GT and its demonstration, readers can refer to Agalbjorn et al. (1997).

Artificial neural network

ANN is an intelligent algorithm widely used in pattern classification, clustering, function approximation, forecasting, image completion and many complex problems in the real world (Hajmeer et al. 2000). With the strong power of describing complex systems, the ANN model

has been widely used to simulate various hydrologic processes and previous researchers reported that the ANN model performs quite well on flood forecasting (Chiang et al. 2007; Kan et al. 2015; Nanda et al. 2016; Ba et al. 2017).

According to Chiang et al. (2004) and Wang et al. (2015), the multilayer feed-forward ANN with one input layer, several hidden layers and one output layer can have better convergence and performance. Consider a three-layer ANN with one input layer, one hidden layer and one output layer, whose numbers of nodes are m , n and 1, respectively. Then the ANN structure for flood forecasting can be expressed by the following formula:

$$Q_{f,t+k} = f(X_t, \varpi, \theta, m, n) \\ = \theta_0 + \sum_{i=1}^n w_i \varphi \left(\sum_{j=1}^m \omega_{ij} X_{t,j} + \theta_i \right) \quad (6)$$

where $Q_{f,t}$ denotes the flood forecast at time t ; X_t denotes the input vector, which usually contains hydrometeorological quantities such as antecedent discharges and precipitation predictions; k denotes the lead time; $\varphi(\cdot)$ denotes the transfer function type, which usually takes Sigmoid function for flood forecasting; w_{ij} denotes the weight coefficient between the j th node of the input layer and the i th node of the hidden layer; θ_i are the corresponding bias of the i th hidden layer node; w_i denotes the weight coefficient between the i th hidden layer node and the output layer node; θ_0 is the bias at the output layer node.

In this study, two widely used input variables, the antecedent inflow and precipitation, are selected as inputs of the ANN models for the Danjiangkou reservoir inflow forecasting. The ANN model parameters are optimized by the genetic algorithm (GA) with the following objective function:

$$\min \sum_{i=1}^n (Q_{o,i} - Q_{f,i})^2 \quad (7)$$

where $Q_{o,i}$ denotes the observed inflow at time i , respectively; n denotes the data length.

NWP data correction

Previous research literature shows that it is essential to correct the TIGGE data in order to use them more effectively

due to the following reasons: (1) the accuracy of the raw TIGGE data are still limited nowadays and thus unsuitable for direct application to hydrological forecasting, despite the fact that the NWP techniques have gained significant improvements; (2) the spread of the raw TIGGE data may be unreliable since the uncertainty range obtained from the ensemble spread may not contain statistically consistent observations; and (3) the raw spatiotemporal resolution of the TIGGE data may not satisfy the forecasting models (Tao et al. 2014). A simple while effective correction method is applied here to preprocess the TIGGE ensemble NWP (Peng et al. 2015):

1. When the precipitation prediction result does not exceed the heavy rain threshold, the prediction value is adjusted by the following equation:

$$P_{crt} = \frac{\overline{P_o}}{\overline{P_f}} \cdot P_f \tag{8}$$

where P_f indicates the precipitation prediction, $\overline{P_f}$ and $\overline{P_o}$ indicates the mean value of the historic predictions and observations, respectively, and P_{crt} indicates the corrected prediction.

2. When the precipitation prediction indicates an intense rain event, the prediction value is adjusted by the following equation:

$$P_{crt} = (P_f + P_{mid})/2 \tag{9}$$

where P_{mid} indicates the middle value of the historic observed intense rain event, which is 40.4 mm/d in this study basin. The method is used to correct the systemic bias of the TIGGE NWP and meanwhile give sufficient attention to severe rain events (Peng et al. 2015).

Bayesian model averaging

BMA is a statistical post-processing method for deriving more skillful and reliable probabilistic forecasts than the original ensemble forecasts (Raftery et al. 2005; Duan et al. 2007). The BMA method has been broadly used in hydrometeorological fields. By determining the weights of the probability distribution functions (PDF) of the ensemble

forecasting members, the BMA can give a combined PDF of the ensemble forecasts, i.e. the BMA-PDF, $p(y)$, given by:

$$p(y|M) = \sum_{i=1}^N p(f_i|M)p_i(y|f_i, M) \tag{10}$$

where y denotes the variable to be forecasted; $M = [y_{obs,1}, y_{obs,2}, \dots, y_{obs,m}]$ denotes the observed y for training with a length of m ; f_i denotes the i th ensemble member value; $p(f_i|M)$ denotes the posterior probability of f_i , which actually reflects the skill of the i th member during the training period and can be denoted as a weight parameter, $\varpi_i = p(f_i|M)$. Obviously, $\sum_{i=1}^n \varpi_i = 1$, where n is the number of ensemble members. A higher ϖ_i value usually indicates the i th ensemble member has better forecasting performance during the training period.

The conditioned distribution has different types for different forecasting variables. As for inflow discharge, it is reasonable to choose skewed distribution, such as Pearson type 3 distributions, which will make the parameter estimation of the BMA-PDF very difficult. In order to simplify the parameter estimation phase and ensure the results are correct, a Box-Cox transformation is performed on both the observed inflow discharges and the ensemble forecasting members to make their distributions approximate to Gaussian distribution:

$$z = \begin{cases} \log(y), & \lambda = 0 \\ \frac{y^\lambda - 1}{\lambda}, & \lambda \neq 0 \end{cases} \tag{11}$$

where z is the transformed variable by the Box-Cox transformation; λ is the Box-Cox coefficient, which can be optimized via goodness-of-fit tests (Asar et al. 2017). In this study, λ is estimated by the observed inflow discharges and the ensemble forecasts are transformed with the same λ of y .

Supposing that the conditioned distribution is fellow Gaussian distribution, then Equation (10) can be transformed into:

$$p(y|M) = \sum_{i=1}^N \varpi_i N(a_i + b_i f_i, \sigma_i^2) \tag{12}$$

where $N(a_i + b_i f_i, \sigma_i^2)$ represents a Gaussian distribution with a mean value $a_i + b_i f_i$ and variance σ_i^2 . There are

some methods to estimate the parameters in Equation (12). Raftery *et al.* (2005) chose the Expectation–Maximization (EM) method for solving this problem in their seminal paper. However, the EM method cannot guarantee the global convergence of the optimized parameters. Vrugt *et al.* (2008) proposed the Differential Evolution Adaptive Metropolis (DREAM) Markov Chain Monte Carlo (MCMC) method to optimize the parameters of the BMA model. The DREAM-MCMC method has the following advantages over the EM method (Vrugt *et al.* 2008). First, it can conveniently be used with distributions other than Gaussian distribution. Second, it can provide a full view of the posterior distribution of the BMA weights and variances. Finally, it is proved that the DREAM-MCMC is sufficient for handling a relatively high number of BMA parameters. Thus the DREAM-MCMC was chosen in this study to optimize the BMA parameters.

Once the BMA-PDF is known, the BMA quantile forecast ζ given the non-exceedance probability $r \in [0, 1]$ can be derived by solving the below equation:

$$\int_0^{\zeta} p(\zeta|M)d\zeta = r \quad (13)$$

The deterministic forecasting results of the BMA-PDF can be expressed as:

$$f_{BMA} = \sum_{i=1}^N \omega_i f_i \quad (14)$$

which is naturally a weighted combination of the ensemble forecasts.

Evaluation criteria

A coherent set of evaluation criteria is utilized to evaluate the flood forecasting results. As for the deterministic forecasts, three familiar metrics, namely the Nash–Sutcliffe efficiency (*NSE*), mean average error (*MAE*) and root mean square error (*RMSE*), are selected and expressed by:

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - F_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (15)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |F_i - O_i| \quad (16)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (F_i - O_i)^2}{n}} \quad (17)$$

where O_i and F_i denotes the observations and the forecasts, respectively, and n denotes the length of the data series.

The *NSE* ranges from $-\infty$ to 1 and is positive oriented (Nash & Sutcliffe 1970). The *MAE* and the *RMSE* are widely regarded as standard metrics for forecasting errors in the fields of hydrology, for which low values indicate a good forecasting skill (Gneiting *et al.* 2007; Chai & Draxler 2014).

Both visual plots in conjunction with statistical metrics are used to evaluate the probabilistic forecasts, including the percentage integrated transform (PIT) histogram, the calibration deviation (*CD*), the ignorance score (*IGN*) and the continuous ranked probability score (*CRPS*). The PIT histogram is widely used to assess the probabilistic calibration or reliability (Gneiting *et al.* 2007; Bourdin *et al.* 2014), which should be approximately flat for perfectly reliable forecasts. The PIT values are given by:

$$PIT_t = G_t(O_t) \quad (18)$$

where O_t is the verifying observation at time t , and G_t is the corresponding forecast cumulative distribution function (CDF), which is the integration of the BMA-PDF function in this study. The number of PIT bins is usually arbitrarily selected from 10 to 20. When the PIT histogram is not flat, its shape can reflect the problems with the probabilistic results. As for the U-shape histogram, it usually indicates the forecast PDF has inadequate spread or underdispersion. On the contrary, the humpback-shape histogram indicates the forecast PDF has overdispersion. However, it should be noted that a flat PIT histogram is not a sufficient condition for the reliability, since a combination of negatively and positively biased forecast distributions can also yield a flat PIT histogram while being unreliable (Hamill 2009).

The metric *CD* is more objective compared with the PIT histogram, which can measure the degree of deviation from a flat PIT histogram (Nipen & Stull 2011). The metric *CD* is

calculated as follows:

$$CD = \sqrt{\frac{1}{k} \sum_{i=1}^k \left(bin_i - \frac{1}{k} \right)^2} \quad (19)$$

where bin_i is the bin frequency of the i th bin, and k is the number of the bins. The small CD values are preferred, which means the deviation from a flat PIT histogram is small.

When evaluating a probabilistic forecast, we also expect to know whether the derived PDF concentrates in the correct area, which can be realized by the dimensionless metric *IGN* (Roulston & Smith 2002) defined as:

$$IGN = -\frac{1}{n} \sum_{i=1}^n \log_2(g_t(O_t)) \quad (20)$$

where g_t denotes the forecast PDF at time t , and n denotes the data length. It is obvious that low *IGN* values are preferred since this shows that high probability is placed in the vicinity of the observations.

The metric *CRPS* is an index that can address both the reliability and sharpness of probabilistic forecasts (Gneiting et al. 2007), which is calculated by:

$$CRPS = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} [G_t(O_t) - H(x - O_t)]^2 dx \quad (21)$$

where x denotes the inflow discharge; H is the Heaviside function of a given real number s given by:

$$\begin{aligned} H(s) &= 1 & s \geq 0 \\ H(s) &= 0 & s < 0 \end{aligned} \quad (22)$$

When the forecast G is deterministic, the *CRPS* will reduce to the *MAE* (Gneiting et al. 2007). This makes it convenient for making comparisons between deterministic and probabilistic forecasts.

RESULTS AND DISCUSSION

ANN model calibration and evaluation

In advance of calibrating the ANN models for the Danjiangkou reservoir inflow forecasting, the Gamma test is applied

to identify the suitable combination of antecedent inflow flood Q_o and precipitation P_o . For both Q_o and P_o , the first five lags are considered as the potential ANN inputs since the mean flood concentration time of the Danjiangkou reservoir is about 3 days and the larger lags are physically independent with the inflow flood. Thus a total of 1023 ($2^{(5+5)}-1$) Γ values are computed. The produced Γ values are sorted in an ascending order, in which Γ values smaller than the 10th percentile (Γ_{10}) are classified as the best group ($F_{\Gamma \leq \Gamma_{10}}$) and those larger than the 90th percentile (Γ_{90}) are classified as the worst group ($F_{\Gamma \geq \Gamma_{90}}$). The factor score (fs) are calculated by the following equation, where the range of fs is $[-\infty, 1]$:

$$fs = 1 - \frac{F_{\Gamma \geq \Gamma_{90}}}{F_{\Gamma \leq \Gamma_{10}}} \quad (23)$$

Following Chang et al. (2014), the threshold of fs is set as 0.5 and a higher fs value means the specific time lag is effective. The Gamma test results of the antecedent inflow and the precipitation are displayed in Figure 3. It can be seen from Figure 3 that the effective time lags for both antecedent inflow and precipitation are 1, 2 and 3. It is also noted that for the precipitation, the 2 d time lag has the highest fs value, which is approximate to the mean concentration time of the Danjiangkou basin. The input combinations of the ANN models with different lead times optimized by the Gamma test are summarized in Table 2. Considering the data availability, the three ANN models have different antecedent inflow inputs while the precipitation inputs are the same. Also, when making inflow forecasts, the NWP data are used instead of the precipitation observations as shown in Table 2.

The evaluation metrics for the calibrated ANN models are listed in Table 3 and the simulation results during calibration and validation periods are displayed in Figure 4. It can be seen that the ANN models have quite good performances and can fit the observed inflow series very well. The ANN simulated inflows with 1 and 2 d lead times are relatively better than that of 3 d lead time, which are obviously overestimated during the calibration period in Figure 4(c). All the *NSE* values during calibration period are larger than 0.85 and the *MAE* values and *RMSE* values are very low. The validation periods appear less satisfied than the calibration periods under the criteria of *NSE*, *MAE* and *RMSE*.

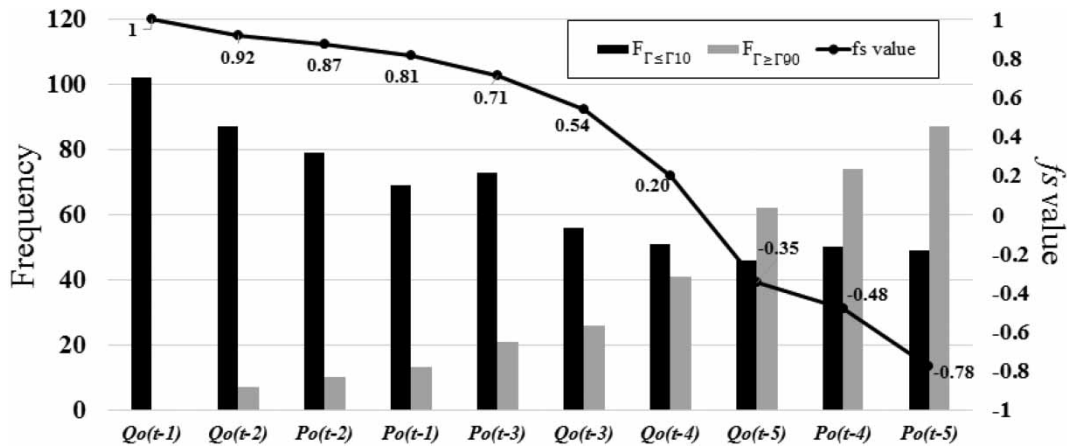


Figure 3 | Gamma test results for determining the input combinations of ANN models.

Table 2 | ANN model input combinations selected by gamma test

Lead time	Output	Simulation	Forecast
1 d	$Q_{f,t+1}$	$Q_{o,t-2}, Q_{o,t-1}, Q_{o,t}, P_{o,t-2}, P_{o,t-1}, P_{o,t}$	$Q_{o,t-2}, Q_{o,t-1}, Q_{o,t}, P_{o,t-2}, P_{o,t-1}, P_{f,t}$
2 d	$Q_{f,t+2}$	$Q_{o,t-1}, Q_{o,t}, P_{o,t-1}, P_{o,t}, P_{o,t+1}$	$Q_{o,t-1}, Q_{o,t}, P_{o,t-1}, P_{f,t}, P_{f,t+1}$
3 d	$Q_{f,t+3}$	$Q_{o,t}, P_{o,t}, P_{o,t+1}, P_{o,t+2}$	$Q_{o,t}, P_{f,t}, P_{f,t+1}, P_{f,t+2}$

* $Q_{f,t}$: Reservoir inflow forecast at time t ; $Q_{o,t}$: Observed reservoir inflow at time t ; $P_{o,t}$: Gauged precipitation during the t th time interval; $P_{f,t}$: Precipitation prediction during the t th time interval.

Hence, the three ANN models are used to generate ensemble inflow forecasts.

Performance of the ensemble inflow forecasts

As mentioned above, the raw TIGGE NWP must be corrected before applying to hydrometeorological purposes.

Figure 5 shows the comparisons of *NSE*, *MAE* and *RMSE* values of the raw TIGGE data and the corrected TIGGE data. For the raw TIGGE data, it can be seen from Figure 5 that the accuracy of precipitation prediction declines with lead time increase. Among the three TIGGE models, the ECMWF performs the best. The TIGGE data from the CMA generally performs worst with the smallest *NSE*

Table 3 | Evaluation of the deterministic results of the raw ensemble inflow forecasts and the BMA probabilistic inflow forecasts with 1–3 d lead times

Method		Lead time	<i>NSE</i>	<i>MAE</i> /(m ³ /s)	<i>RMSE</i> /(m ³ /s)
Ensemble	Calibration (2008–2011)	1 d	0.917	236	550
		2 d	0.846	338	757
		3 d	0.731	394	1033
	Validation (2012–2013)	1 d	0.778	225	466
		2 d	0.682	284	560
		3 d	0.665	310	561
BMA probabilistic	Calibration (2008–2011)	1 d	0.933	227	522
		2 d	0.861	304	726
		3 d	0.774	366	927
	Validation (2012–2013)	1 d	0.782	201	417
		2 d	0.709	247	524
		3 d	0.704	298	543

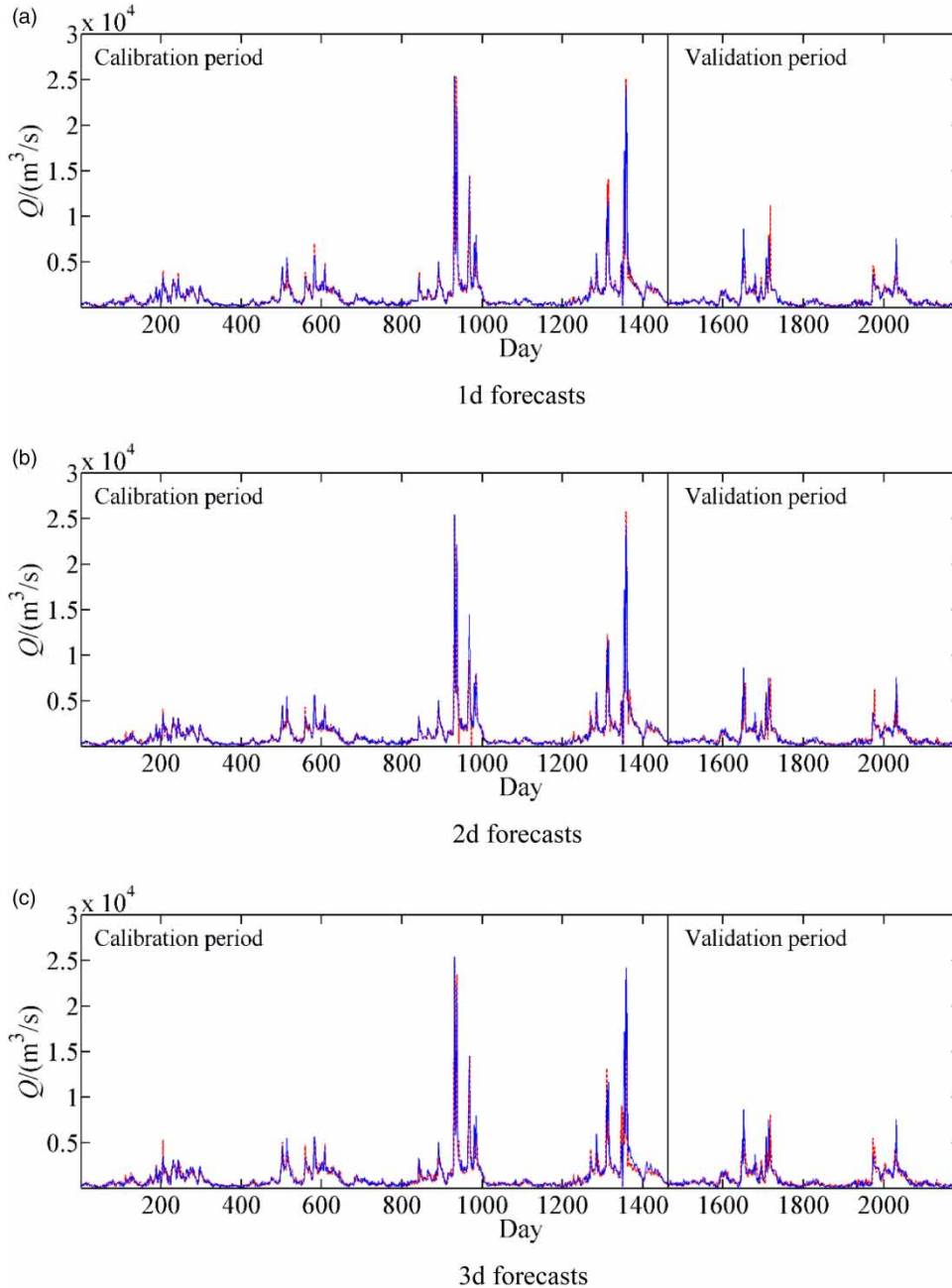


Figure 4 | The observed (solid line) and simulated (dash line) inflows of the Danjiangkou reservoir with 1–3 d lead times during calibration and validation periods.

values and the largest *MAE* and *RMSE* values. By the correction method introduced above under ‘NWP data correction’, significant improvements are observed for these three metrics above. The systematic errors are sufficiently removed since the corrected TIGGE data are more approximate to the precipitation observations. The

corrected TIGGE data are used as ANN inputs to generate ensemble flood forecasts together with the antecedent flood discharge of the Danjiangkou reservoir.

By feeding the corrected TIGGE data into ANN models with the antecedent precipitation and discharge records, a total of 87 ensemble inflow forecasts (the raw ensemble

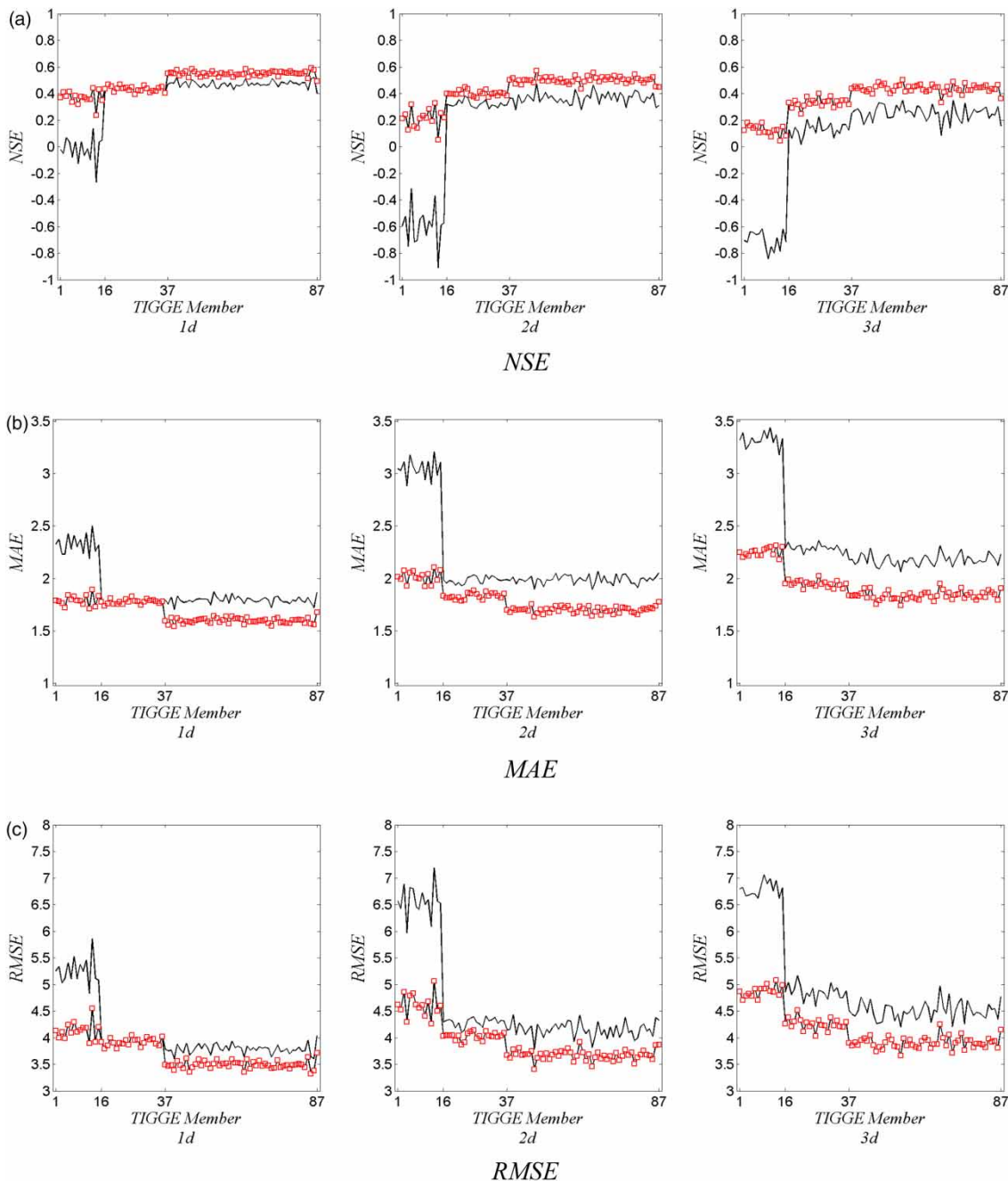


Figure 5 | Comparison between the evaluation metrics for the raw TIGGE data (diamond line) and the corrected TIGGE data (dotted line).

inflow forecasts) are obtained in accordance with the TIGGE NWP member size for each lead time. Figure 6 displays the evaluation results of the raw ensemble inflow forecasts with box-whisker plots. The upper and lower ends of the whiskers are the estimated 95% and 5% quantiles; the upper and lower edges of the boxes are the

estimated 75% and 25% quantiles, respectively. The center lines are the median and the crosses outside the whiskers are outliers. It can be seen from Figure 6 that the widths of the evaluation metric intervals apparently increase with the lead time, which reveals that the discrepancy between the inflow forecasts is more significant as the lead time

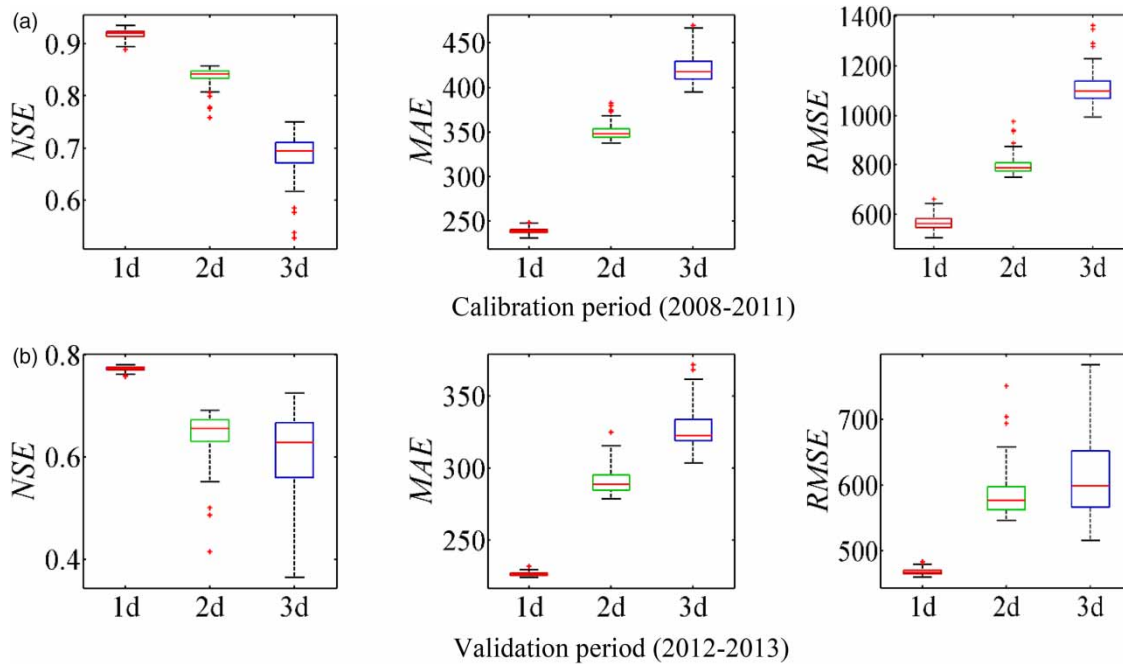


Figure 6 | Box-whisker plots of the ANN model evaluation metrics with 1–3 d lead times during calibration and validation periods.

increases. This phenomenon indicates that the forecasting uncertainty cannot be neglected for middle- and long-term hydrologic forecasting. The reservoir managers and decision makers appreciate forecasting results with more accuracy and less uncertainty. It is very difficult to make decisions with the ensemble inflow forecasts that are quite different from each other. This difficulty for choosing a ‘best’ forecasting member seems more difficult for longer lead times since the uncertainty is more significant (the results are more dispersive), and in fact no single forecasting member can always have the best performance, which is widely addressed by the previous research studies (Cloke & Pappenberger 2009; He *et al.* 2009; He *et al.* 2010; Alfieri *et al.* 2013). To better support the decision making, a useful approach is to generate probabilistic inflow forecasts based on the ensemble results using post-processing methods (Cloke & Pappenberger 2009; Gneiting & Katzfuss 2014).

Evaluation of the BMA probabilistic forecasts

We use the BMA method to generate probabilistic inflow forecasts from the raw ensemble inflow forecasts, the calibration period (2008–2011) and the validation period

(2012–2013) are consistent with the ANN models. For calculation convenience, the conditional distribution of the ensemble members is supposed to follow Gaussian distribution (Duan *et al.* 2007). The Box-Cox transformation is performed on the observed inflow series and the ensemble inflow forecasts to make them more Gaussian distributed. The transformed and original series are plotted in Figure 7. Since the ensemble member number is large, only one inflow forecast member is chosen arbitrarily to display the Box-Cox transformation performance. It can be seen clearly that the original inflow series are non-Gaussian distributed. After the Box-Cox transformation, the observed and ensemble forecasting inflow series are very close to the theoretical Gaussian distribution probability curve, which proves that the transformed series can satisfy the Gaussian assumption well.

Figure 8 shows the PIT histogram of the raw ensemble forecasts and the BMA probabilistic forecasts during the validation period. The *CD* values are calculated by Equation (19) and shown in the figures. The reference line of a perfectly flatted histogram is also shown on each plot, which is a function of the number of the bins and the sample size (Bourdin *et al.* 2014). Figure 8 clearly illustrates that the

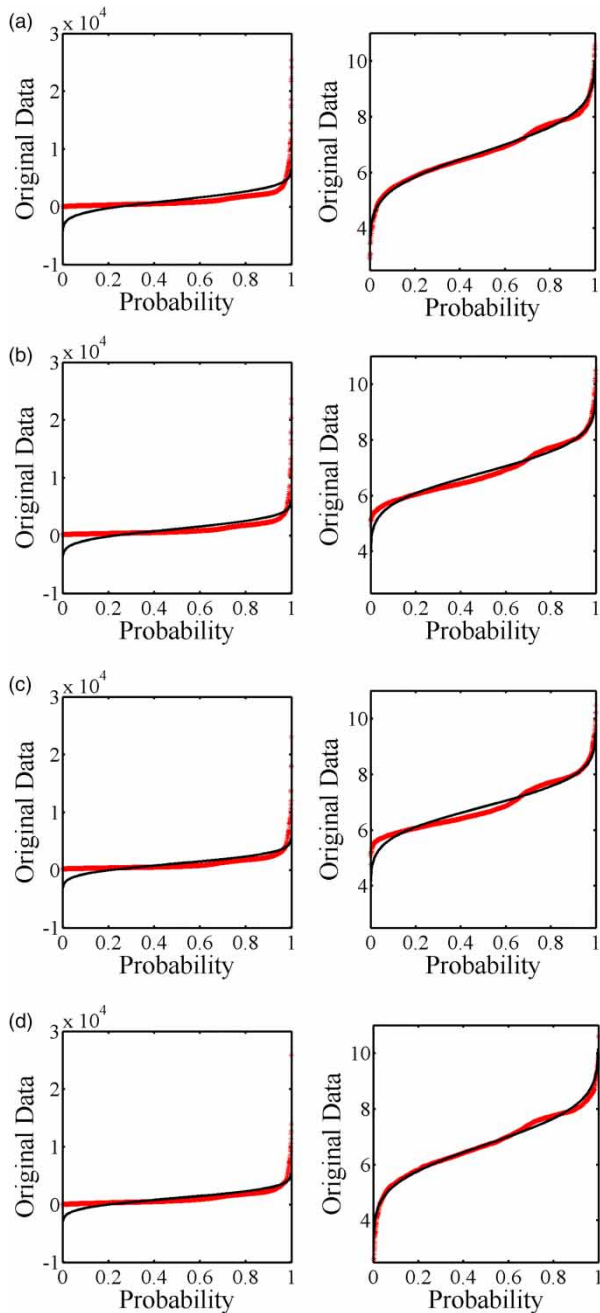


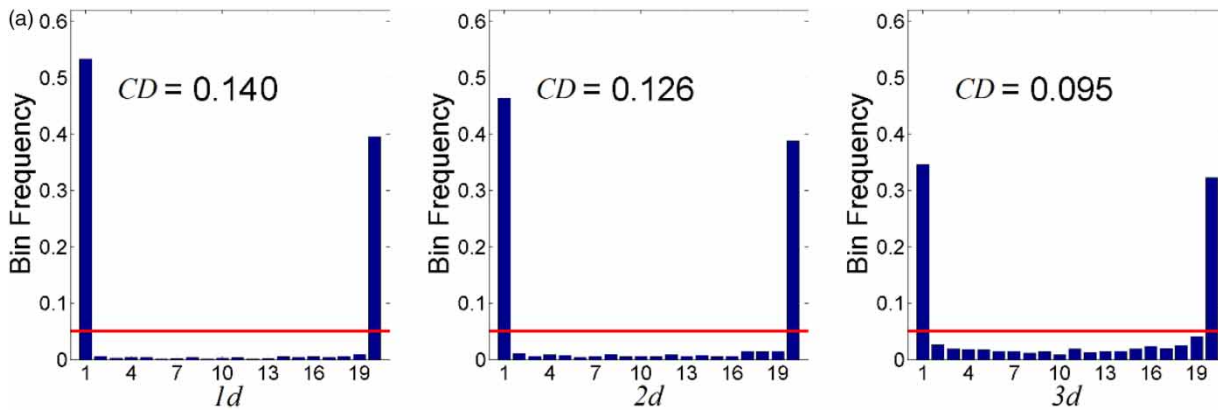
Figure 7 | Comparisons between original (solid line) and the Box-Cox transformed inflows (asterisks).

raw ensemble forecasts are highly unreliable. The PIT values are seldom distributed at the center of the probabilistic density, while the first and last bins almost accounted for the whole ensemble. After probabilistic calibration using the BMA method, the reliability of the probabilistic forecasts

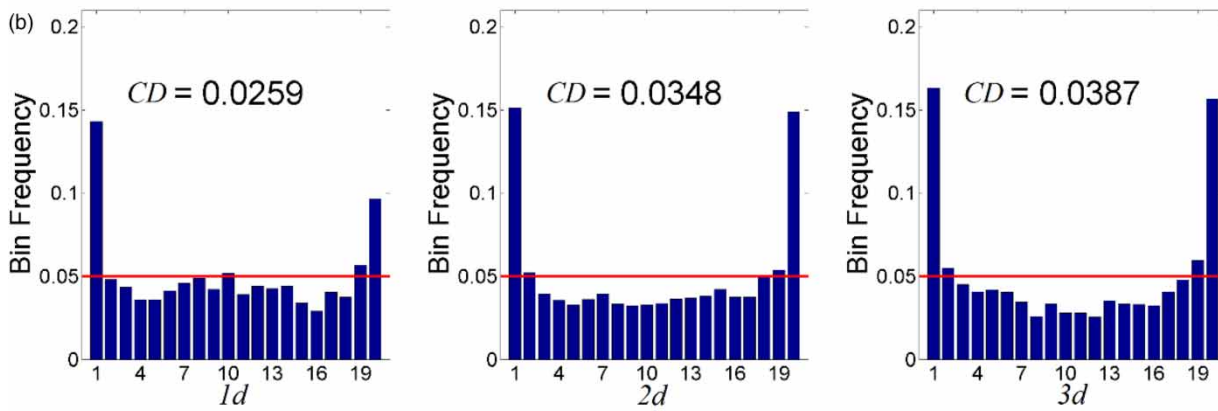
is effectively improved. Many more PIT values are distributed at the center of the probabilistic density. The PIT histograms of the BMA probabilistic forecasts are more flat compared with the raw ensemble ones, although the first and last bins still have more PITs. The metric *CD* is more intuitional than the PIT histogram. For each lead time, the *CD* value of the BMA probabilistic forecasts is much smaller than that of the raw ensemble forecasts. Results indicate that the BMA probabilistic inflow forecasts are more reliable than the raw ensemble ones.

Figure 9 shows the *CRPS* and *IGN* values of the raw ensemble forecasts and the BMA probabilistic forecasts for each lead time during the validation period. The *CRPS* can represent both the reliability and the sharpness of the probabilistic forecasts and the *IGN* is a metric that illustrates the ability of the probabilistic forecasts' concentration at the right area. It can be seen from Figure 9 that the *CRPS* values of the BMA probabilistic inflow forecasts are lower than the raw ensemble ones. This indicates that the BMA probabilistic inflow forecasts have better performance than the raw ensemble inflow forecasts for all three lead times. It is also noted that this advantage becomes less significant when the lead time increases. The results of the *IGN* values in Figure 9 are in agreement with the *CRPS* values.

Another finding in Figure 9 is that the *IGN* values of the raw ensemble inflow forecasts show a descending trend with the lead time, indicating the raw ensemble inflow forecasts with longer lead time can better contain the correct areas. This is caused by the ensemble generation method which couples TIGGE NWP with ANN models. With the lead time increases, more NWP data are used as model inputs and additional uncertainty are introduced into the ensemble forecasting results. As shown in Figure 6, the metrics with a 3 d lead time have the widest spreads, which indicates that the 3 d members are less sharp or consistent than the 1 and 2 d results. The 1 d raw ensemble inflow forecasts are the sharpest and have the highest *IGN*, which indicates that the 1 d raw ensemble forecasts are sharp but less reliable. Results show the raw ensemble forecasts fail to balance the sharpness and the reliability. After applying the BMA method, it is observed that the trends of the two metrics (*CRPS* and *IGN*) are consistent, i.e. the probabilistic inflow forecasts with longer lead time are less reliable and sharp.



PIT histograms for the raw ensemble inflow forecasts



PIT histograms for the BMA probabilistic inflow forecasts

Figure 8 | Comparisons of PIT histograms for the raw ensemble inflow forecasts and the BMA probabilistic inflow forecasts with 1–3 d lead times.

The deterministic inflow forecasts derived from the raw ensemble forecasts and the BMA probabilistic

forecasts are also evaluated since the deterministic results occupy an important position in practice. The results in Table 3 show that the BMA deterministic forecasts have better performances than the raw ensemble ones for all three lead times under the criteria of *NSE*, *MAE* and *RMSE*. The main reason for these improvements is that the BMA is essentially a bias correction method, which gives larger weights to the better ensemble members and vice versa (Duan et al. 2007). Given an overall conclusion, both the evaluation results of probabilistic forecasts and deterministic forecasts indicate that the scheme of post-processing the ensemble forecasts generated using TIGGE NWP data by the BMA method offers a useful approach for probabilistic forecasts in the Danjiangkou reservoir basin.

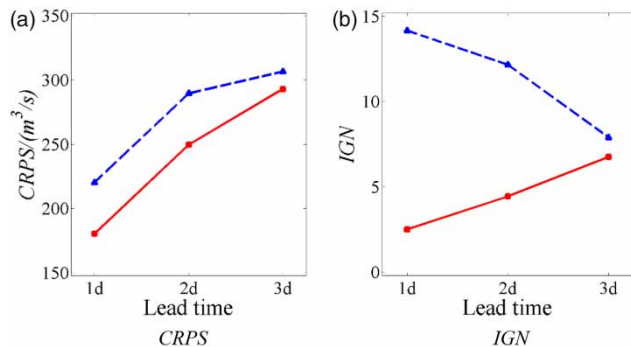


Figure 9 | Comparison between the evaluation metrics of the raw ensemble inflow forecasts (dashed line) and the BMA probabilistic inflow forecasts (solid line).

CONCLUSIONS

Nowadays, more and more attention is paid to obtaining timely and accurate flood forecasts. The development of numerical weather prediction (NWP) techniques has contributed to this goal by providing ensemble precipitation predictions, which take longer lead times and can describe the uncertainty of the meteorological system. In this study, the NWP data retrieved from three TIGGE models were used as the inputs of ANN models to forecast reservoir inflows. The ensemble inflow forecasts were then post-processed to generate probabilistic forecasts using the BMA method. The main conclusions of this study are summarized as follows:

1. The raw TIGGE NWP data are biased and must be corrected before implementing the hydrologic model. The correction method applied in this study can effectively correct the TIGGE NWP under the criteria of *NSE*, *MAE* and *RMSE* in the Danjiangkou reservoir basin.
2. The raw ensemble inflow forecasts are unreliable. The ensemble spread is very narrow for 1 d results while much wider for the 2 and 3 d results. The PIT histograms of the raw ensembles show significant U-shapes with almost all the PIT values being located at the first and the last bins.
3. The probabilistic forecasts derived from the raw ensemble forecasts by using the BMA method are significantly improved. The PIT histograms of the probabilistic forecasts are flatter as compared with the raw ensemble ones, and the *CD*, *IGN* and *CRPS* values of the probabilistic forecasts are also lower than those of the raw ensemble forecasts for all three lead times. Results show that the BMA can provide a useful approach to generate probabilistic forecasts from the TIGGE ensemble inflow forecasts.

There remain some limitations in this study. The BMA parameters were optimized with the data of the calibration period but did not change over time. This may be improved in future by using a sliding window calibration method to optimize the parameters adaptively (Bourdin *et al.* 2014). Another limitation is that the Box-Cox transformation may be replaced by the Normal Quantile Transformation (NQT) to make the data more Gaussian distributed, which

can better fulfill the Gaussian assumption. Besides, as discussed by many studies (Xiong *et al.* 2009; Li *et al.* 2010; Kasiviswanathan *et al.* 2013), the reliability and the sharpness of probabilistic forecasting are competitive, how to balance the two targets is an interesting question to be investigated in future.

ACKNOWLEDGEMENT

This study was supported by the National Key Research and Development Plan of China (Grant No. 2016YFC0402206) and the National Natural Science Foundation of China (Grant No. 51539009, 51779279 and 91647106). Thanks are also given to the three anonymous reviewers for their great efforts to help improve this manuscript.

REFERENCES

- Agalbjorn, S., Koncar, N. & Jones, A. J. 1997 A note on the gamma test. *Neuralcomput. Appl.* **5**, 131–133.
- Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D. & Thielen, J. 2013 GloFAS-global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.* **17** (3), 1161–1175.
- Asar, Ö., Ozlem, I. & Osman, D. 2017 Estimating Box-Cox power transformation parameter via goodness-of-fit tests. *Commun. Stat-Simul. C* **46** (1), 91–105.
- Ba, H. H., Guo, S. L., Wang, Y., Hong, X. J., Zhong, Y. X. & Liu, Z. J. 2017 Improving ANN model performance in runoff forecasting by adding soil moisture input and using data preprocessing techniques. *Hydrol. Res.* nh2017048. DOI: 10.2166/nh.2017.048.
- Bao, H. J., Zhao, L. N., He, Y., Li, Z. J., Wetterhall, F., Cloke, H. L., Pappenberger, F. & Manful, D. 2011 Coupling ensemble weather predictions based on TIGGE database with grid-xinjiang model for flood forecast. *Adv. Geosci.* **29**, 61–67.
- Bellier, J., Zin, I., Siblot, S. & Bontron, G. 2016 Probabilistic flood forecasting on the Rhone River: evaluation with ensemble and analogue-based precipitation forecasts. *E3S Web of Conferences. EDP Sciences* **7**, 18011.
- Bjerknes, V. 1904 Problem of weather prediction from the viewpoints of mechanics and physics. *Meteorology* **21**, 1–7.
- Bourdin, D. R., Nipen, T. N. & Stull, R. B. 2014 Reliable probabilistic forecasts from an ensemble reservoir inflow forecasting system. *Water Resour. Res.* **50**, 3108–3130, doi:10.1002/2014WR015462.

- Chai, T. & Draxler, R. R. 2014 Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **7** (3), 1247–1250.
- Chang, F. J., Chen, P. A., Liu, C. W., Liao, H. C. & Liao, C. M. 2013 Regional estimation of groundwater arsenic concentrations through systematical dynamic-neural modeling. *J. Hydrol.* **499**, 265–274.
- Chang, F. J., Chen, P. A., Lu, Y. R., Huang, E. & Chang, K. Y. 2014 Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. *J. Hydrol.* **517**, 836–846.
- Chen, L. & Singh, V. P. 2017 Generalized beta distribution of the second kind for flood frequency analysis. *Entropy* **19**, 254.
- Chiang, Y. M., Chang, L. C. & Chang, F. J. 2004 Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling. *J. Hydrol.* **290** (3), 297–311.
- Chiang, Y. M., Hsu, K. L. & Chang, F. J. 2007 Merging multiple precipitation sources for flash flood forecasting. *J. Hydrol.* **340** (3), 183–196.
- Cloke, H. L. & Pappenberger, F. 2009 Ensemble flood forecasting: a review. *J. Hydrol.* **375** (3), 613–626.
- Coustau, M., Rousset-Regimbeau, F., Thirel, G., Habets, F., Janet, B., Martin, E. C., Saint-Aubine, C. & Soubeyrouxa, J. M. 2015 Impact of improved meteorological forcing, profile of soil hydraulic conductivity and data assimilation on an operational hydrological ensemble forecast system over France. *J. Hydrol.* **525**, 781–792.
- Duan, Q. Y., Ajami, N. K., Gao, X. G. & Sorooshian, S. 2007 Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.* **30** (5), 1371–1386.
- Fan, F. M., Schwanenber, D., Colischonn, W. & Weerts, A. 2015 Verification of inflow into hydropower reservoirs using ensemble forecasts of the TIGGE database for large scale basins in Brazil. *J. Hydro Reg. Stud.* **4** (Part B), 196–227.
- Gao, C., Zhang, Z. T., Zhai, J. Q., Liu, Q. & Yao, M. 2015 Research on meteorological thresholds of drought and flood disaster: a case study in the Huai River Basin, China. *Stoch. Env. Res. Risk A* **29** (1), 157–167.
- Gneiting, T. & Katzfuss, M. 2014 Probabilistic forecasting. *J. R. Stat. Soc.* **1** (1), 125–151.
- Gneiting, T., Balabdaoui, F. & Raftery, A. E. 2007 Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. B* **69** (2), 243–268.
- Guo, S. L., Zhang, H. G., Chen, H., Peng, D. Z., Liu, P. & Pang, B. 2004 A reservoir flood forecasting and control system for China. *Hydrolog. Sci. J.* **49** (6), 959–972.
- Hajmeer, M. N., Basheer, I. A., Marsden, J. L. & Fung, D. Y. C. 2000 New approach for modeling generalized microbial growth curves using artificial neural networks. *J. Rapid Methods & Autom. Microbiol.* **8** (4), 265–283.
- Hamill, T. M. 2001 Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Weather Rev.* **129** (3), 550–560.
- Hardy, J., Gourley, J. J., Kirstetter, P. E., Hong, Y., Kong, F. & Flamig, Z. L. 2016 A method for probabilistic flash flood forecasting. *J. Hydrol.* **541**, 480–494.
- He, Y., Wetterhall, F., Cloke, H. L., Pappenberger, F., Wilson, M., Freer, J. & McGregor, G. 2009 Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions. *Meteorol. Appl.* **16** (1), 91–101.
- He, Y., Wetterhall, F., Bao, H. J., Cloke, H., Li, Z., Pappenberger, F., Hu, Y. Z., Manful, D. & Huang, Y. C. 2010 Ensemble forecasting using TIGGE for the July–September 2008 floods in the Upper Huai catchment: a case study. *Atmo. Sci. Lett.* **11** (2), 132–138.
- Huang, Y. C., Li, Z. J., He, Y., Wetterhall, F., Manful, D., Cloke, H. & Pappenberger, F. 2010 Uncertainty assessment of early flood warning driven by the TIGGE ensemble weather predictions. *EGU General Assembly Conference Abstracts* **12**, 15497.
- International Commission on Large Dams (ICOLD) 2006 *Roles of Dams in Flood Mitigation – A Review*. ICOLD Bulletin 131, Paris.
- Kan, G., Yao, C., Li, Q. L., Li, Z., Yu, Z., Liu, Z., He, X. Y. & Liang, K. 2015 Improving event-based rainfall-runoff simulation using an ensemble artificial neural network based hybrid data-driven model. *Stoch. Env. Res. Risk A* **29** (5), 1345–1370.
- Kasiviswanathan, K. S., Cibin, R., Sudheer, K. P. & Chaubey, I. 2013 Constructing prediction interval for artificial neural network rainfall runoff models based on ensemble simulations. *J. Hydrol.* **499**, 275–288.
- Kussul, N., Shelestov, A. & Skakun, S. 2008 Grid system for flood extent extraction from satellite images. *Earth Sci. Inform.* **1** (3), 105.
- Kwon, H. Y. & Kang, Y. O. 2016 Risk analysis and visualization for detecting signs of flood disaster in twitter. *Spat. Inform. Res.* **24** (2), 127–139.
- Li, S., Cheng, X. L., Xu, Z. F., Han, H. Y. & Zhang, Q. F. 2009 Spatial and temporal patterns of the water quality in the Danjiangkou Reservoir, China. *Hydrol. Sci. J.* **54** (1), 124–134.
- Li, L., Xia, J., Xu, C. Y. & Singh, V. P. 2010 Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. *J. Hydrol.* **390**, 210–221.
- Li, W., Zhou, J. Z., Sun, H. W., Feng, K., Zhang, H. & Tayyab, M. 2017 Impact of distribution type in Bayes probability flood forecasting. *Water Resour. Manag.* **31** (3), 961–977.
- Liang, Z. M., Tang, T. T., Li, B. Q., Liu, T., Wang, J. & Hu, Y. 2017 Long-term streamflow forecasting using SWAT through the integration of the random forests precipitation generator: case study of Danjiangkou Reservoir. *Hydrol. Res.* nh2017085. DOI: 10.2166/nh.2017.085.
- Liu, J. & Xie, Z. 2014 BMA probabilistic quantitative precipitation forecasting over the Huaihe basin using TIGGE multimodel ensemble forecasts. *Mon. Weather Rev.* **142** (4), 1542–1555.
- 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** (7), 2111–2126.
- Moghaddamnia, A., GhafariGousheh, M., Piri, J., Amin, S. & Han, D. 2009 Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. *Adv. Water Resour.* **32**, 88–97.

- Nanda, T., Sahoo, B., Beria, H. & Chatterjee, C. 2016 A wavelet-based non-linear autoregressive with exogenous inputs (WNARX) dynamic neural network model for real-time flood forecasting using satellite-based rainfall products. *J. Hydrol.* **539**, 57–75.
- Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models: part 1: a discussion of principles. *J. Hydrol.* **10**, 282–290.
- Nipen, T. & Stull, R. 2011 Calibrating probabilistic forecasts from an NWP ensemble. *Tellus Ser. A* **63** (5), 858–875.
- Noori, R., Karbassi, A. R., Moghaddamnia, A., Han, D., Zokaei-Ashtiani, M. H., Farokhnia, A. & Ghafari Gousheh, M. 2011 Assessment of input variables determination on the SVM model performance using PCA, gamma test and forward selection techniques for monthly stream flow prediction. *J. Hydrol.* **401**, 177–189.
- Peng, Y., Xu, W., Wang, P. & You, F. 2015 Flood forecasting coupled with TIGGE ensemble precipitation forecasts. *J. Tianjin Univ. (Sci. Technol.)* **48** (2), 177–184.
- Raftery, A. E., Gneiting, T., Balabdaoui, F. & Palokowski, M. 2005 Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* **133** (5), 1155–1174.
- Roulston, M. S. & Smith, L. A. 2002 Evaluating probabilistic forecasts using information theory. *Mon. Weather Rev.* **130** (6), 1653–1660.
- Shen, Y., Xiong, A. Y., Wang, Y. & Xie, P. 2010 Performance of high-resolution satellite precipitation products over China. *J. Geophys. Res. Atmo.* **115** (D2), 355–365.
- Sun, R., Yuan, H., Liu, X. & Jiang, X. 2016 Evaluation of the latest satellite-gauge precipitation products and their hydrologic applications over the Huaihe River basin. *J. Hydrol.* **536** (2016), 302–319.
- Tao, Y. M., Duan, Q. Y., Ye, A. Z., Gong, W., Di, Z., Xiao, M. & Hsu, K. 2014 An evaluation of post-processed TIGGE multimodel ensemble precipitation forecast in the Huai river basin. *J. Hydrol.* **519**, 2890–2905.
- Todini, E. 2017 Flood forecasting and decision making in the new millennium. Where are we? *Water Resour. Manage.* **31**, 3111–3129.
- Uddin, K., Gurung, D. R., Giriraj, A. & Shrestha, B. 2013 Application of remote sensing and GIS for flood hazard management: a case study from Sindh province, Pakistan. *Am. J. Geogr. Inform. Syst.* **2**, 1–5.
- Van Steenbergen, N. & Willems, P. 2015 Uncertainty decomposition and reduction in river flood forecasting: Belgian case study. *J. Flood Risk Manag.* **8** (3), 263–275.
- Vrugt, J. A., Diks, C. G. H. & Clark, M. P. 2008 Ensemble Bayesian model averaging using Markov Chain Monte Carlo sampling. *Environ. Fluid Mech.* **8** (5), 579–595.
- Wang, Y., Guo, S. L., Xiong, L. H., Liu, P. & Liu, D. D. 2015 Daily runoff forecasting model based on ANN and data preprocessing techniques. *Water* **7** (8), 4144–4160.
- World Meteorological Organization (WMO) 2005 *First Workshop on the THORPEX Interactive Grand Global Ensemble (TIGGE)*, Final Report. WMO, Geneva, Switzerland.
- World Meteorological Organization (WMO) 2010 *Workshop on the Strategy and Action Plan of the WMO Flood Forecasting Initiative*, Final Report. WMO, Geneva, Switzerland.
- Xie, P. P., Chen, M. Y. & Yang, S. 2007 A gauge-based analysis of daily precipitation over east Asia. *J. Hydrometeorol.* **8** (3), 607.
- Xiong, L. H., Wan, M., Wei, X. J. & O'Connor, K. 2009 Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation. *Hydrolog. Sci. J.* **54** (5), 852–871.
- Zhao, L. N., Tian, F. Y., Wu, H., Qi, D., Di, J. Y. & Wang, Z. 2011 Verification and comparison of probabilistic precipitation forecasts using the TIGGE data in the upriver of Huaihe basin. *Adv. Geosci.* **29**, 95–102.

First received 12 November 2017; accepted in revised form 26 January 2018. Available online 21 February 2018