

Data transformation models utilized in Bayesian probabilistic forecast considering inflow forecasts

Wei Xu, Xiaoying Fu, Xia Li and Ming Wang

ABSTRACT

This paper presents a new Bayesian probabilistic forecast (BPF) model to improve the efficiency and reliability of normal distribution transformation and to describe the uncertainties of medium-range forecasting inflows with 10 days forecast horizons. In this model, the inflow data will be transformed twice to a standard normal distribution. The Box–Cox (BC) model is first used to quickly transform the inflow data with a normal distribution, and then, the transformed data are converted to a standard normal distribution by the meta-Gaussian (MG) model. Based on the transformed inflows in the standard normal distribution, the prior and likelihood density functions of the BPF are established, respectively. In this study, the newly developed model is tested on China's Huanren hydropower reservoir and is compared with BPFs using MG and BC, separately. Comparative results show that the new BPF model exhibits significantly improved data transformation efficiency and forecast accuracy.

Key words | Bayesian probabilistic forecast, forecasting inflow, medium-range, normal distribution transformation, uncertainty

Wei Xu (corresponding author)

Xia Li

Ming Wang

College of River and Ocean Engineering, National Engineering Research Center for Inland Waterway Regulation, Chongqing Jiaotong University, Chongqing, China
E-mail: xuwei19850711@163.com

Wei Xu

Xiaoying Fu

State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Sichuan, China

INTRODUCTION

There are significant hydrologic uncertainties in the simulated and forecasted processes that limited the practical application of the forecast information. Many methods have been proposed to analyze the uncertainties of the hydrological data, in which the Bayesian hydrological probability forecasting based on the Bayesian theory is the most widely applied method (Krzysztofowicz 1999; Biondi & Luca 2013; Liu *et al.* 2016; Han & Coulibaly 2017). The Bayesian probabilistic forecast (BPF) model is currently the most efficient model to address the uncertainties of the hydrologic process (Han & Coulibaly 2017). In the BPF model, the data of the hydrological process are generally assumed as a linear-normal distribution. The uncertainties of the forecasting inflows mainly come from the randomness of the runoff and the hydrology model (Krzysztofowicz 1999; Krzysztofowicz & Kelly 2000). The uncertainties of the inflow process and the hydrology model can be described by prior density

and likelihood density functions. Based on the prior density and likelihood density functions, the posterior probability density function of the forecasting inflows can be generated by the Bayesian theory (Krzysztofowicz & Kelly 2000; Krzysztofowicz & Maranzano 2004).

The BPF model gives the uncertainty distributions of the forecasting values based on the statistical historical information. The BPF model has the corrective capability for forecasting information and thus provides better information (Han & Coulibaly 2017; Liang *et al.* 2017; Ma *et al.* 2018). The BPF model adopts a linearization description to the observed and the forecasting inflows and then obtains the analytical solution of the posterior density function (Krzysztofowicz & Maranzano 2004; Krzysztofowicz 2014). Since Krzysztofowicz introduced a Bayesian forecasting system in 1999, it has been gaining in popularity worldwide, and various improved models were developed to improve

the efficiency (Han & Coulibaly 2017; Barbetta *et al.* 2018). Reggiani & Weerts (2008) used a Bayesian forecasting system to obtain the posterior probability distributions for making effective decisions under inflow uncertainties during flood periods. Biondi *et al.* (2010) applied the hydrologic uncertainty processor (HUP) proposed by Krzysztofowicz in 1999 to a small semi-arid watershed in southern Italy. Ma *et al.* (2013) established a Bayesian statistical forecasting theory for middle and long-term runoff forecasting at a hydropower reservoir to study the mathematical forecasting theory, and the precision of forecasting inflows was improved significantly. Liu *et al.* (2015) applied the quantile water inflows of the Bayesian forecasting system for dynamic control of the flood limiting water level in the Three Gorges Reservoir, and the impacts of the input uncertainty were evaluated concerning benefits, dam safety, and downstream effects during the flood season. Barbetta *et al.* (2016) constructed a model conditional processor based on the Bayesian forecasting approach to address flood forecasting and warning systems in real-time operating. Ahmadisharaf *et al.* (2018) proposed a probabilistic framework to evaluate the impact of the uncertainties on six hydrograph attributes, and the sensitives of the attributes were analyzed.

The Bayesian forecasting system can improve the accuracy of the forecasting information; however, the characteristics of the non-linear and non-normal distribution of the hydrological processes increase the difficulty to modeling. Much effort has been made to solve this problem. Liu *et al.* (2016) investigated and compared the capability of three updating procedures, namely, autoregressive (AR) model, recursive least-squares model and HUP in the real-time flood forecasting. Liu *et al.* (2017) proposed a copula-based HUP model based on a copula function, in which the prior and likelihood density functions were explicitly expressed, and the posterior density function generated by using Monte Carlo sampling. Liang *et al.* (2017) presented a hybrid model called SVR-HUP (support vector regression) by combining the support vector regression and HUP to predict long-term runoff and quantify the uncertainties. Xie *et al.* (2019) proposed a Bayesian probabilistic forecasting framework to characterize the uncertainties of the wind forecasts in the wind power generation operation, in which the infinite Markov switching AR model was

developed to describe the non-linear data. In the above studies, the non-linear data of the hydrological processes are generally characterized by the complex non-linear mathematical models to make the residual errors closer to the normal distribution.

The Bayesian probabilistic forecasting models, combining non-linear mathematical models, are too complicated in real-time operation, and the theories are difficult for the operator to understand. In addition to the non-linear model, the accuracy of the BPF model is also affected by the transformation model (Biondi & Luca 2013; Chen *et al.* 2015). The transformation model with high efficiency and stability can reduce the complexity of the BPF model and makes the BPF model run faster. Currently, two types of transformation method are used to transform the hydrologic processes, i.e., the meta-Gaussian (MG) model and the Box-Cox (BC) transformation model. Krzysztofowicz (1999) transformed the observed and forecasted data into Gaussian or normal space using a MG model, i.e., normal quantile transform (NQT). Then, the MG model was widely used to transform the non-normal data into normal distribution space for the Bayesian forecasting model (Biondi & Luca 2013; Liu *et al.* 2015; Hao *et al.* 2016; Barbetta *et al.* 2018). Weerts *et al.* (2011) proposed a quantile regression model to estimate the predictive uncertainty, and the MG model was used to transform the residual of the rainfall-runoff. Chen *et al.* (2013) constructed a normal-linear BPF and MG BPF to convert deterministic forecasts for Bayesian probabilistic forecasting, and the performances demonstrated that the MG BPF has better applicability. In the MG model, the marginal distribution function of the data needs to be determined before transforming (Krzysztofowicz 1999; Krzysztofowicz & Maranzano 2004; Tanessong *et al.* 2017). However, in real-time operation, the marginal distribution function is difficult to identify from the sets of distributions. Since Box and Cox proposed the BC model in 1964, it has been widely used in non-normal data transformation for only one parameter needs to be determined, especially in hydrological Bayesian model averaging (BMA). In the BMA model, the hydrological data, e.g., streamflow, precipitation, and evaporation, are commonly transformed to ensure that the converted data are approximately Gaussian distribution (Zhang *et al.* 2011; Madadgar & Moradkhani 2014; Chen *et al.* 2015). Madadgar &

Moradkhani (2014) constructed a BMA model, in which the BC transformation was used to transform the inflows. Qu et al. (2017) developed the BC transformation to convert the runoff forecasts of the Fu River basin in China to satisfy the assumption of a normal distribution. Zhong et al. (2018) developed a Bayesian averaging model, and the BC transformation was performed on both the observed inflows and the ensemble forecasting members. Ma et al. (2018) proposed a general framework for blending multiple satellite precipitation data using the dynamic BMA, and the data were transformed by the BC model.

The primary purpose of this paper is to study the efficiency and stability of the normal distribution transformation models by using numerical weather forecast information to forecast 10 days inflows in the future. In the study, the MG model and the BC transformation model (BC) were first constructed as a normal distribution transformation model. Then, the new model is proposed by combining the BC and MG models, which is represented as a BC-MG model. Based on the three transformation models mentioned above, the observed and forecasted inflows of the Huanren Reservoir in China were transformed and the probabilistic inflow forecasts with 10 days horizon were simulated, respectively. Finally, the efficiency and stability of the probabilistic inflow forecasts were evaluated and compared.

METHODS

Transformation model

MG model

The MG model is a common normal distribution transformation method for non-normal data. In this model, the non-normal data are transformed to the standard normal distribution by the NQT (Krzysztofowicz & Maranzano 2004). In the MG model, the normal distribution, Gamma distribution, Weibull distribution, and logarithmic Weibull distribution are selected as candidate models for the marginal distribution function. Before using the MG model to transform the set of given data, the candidate model that best fits the empirical probability needs to be selected as

the marginal distribution function, and the parameters of the function need to be calibrated. Then, the theoretical probabilities of the given data can be obtained from the marginal distribution function. At last, the NQT of the given data can be generated from the standard normal distribution function according to the theoretical probabilities. The NQT of the observed and forecasted inflows for the standard normal distribution can be calculated using Equations (1) and (2), respectively.

$$W_k = Q^{-1}(\bar{\Gamma}_k(H_k)), \quad k = 0, 1, \dots, K_1 \quad (1)$$

$$X_k = Q^{-1}(\bar{\Lambda}_k(S_k)), \quad k = 0, 1, \dots, K_2 \quad (2)$$

where H_k and S_k represent the observed and forecasted inflows, respectively; K_1 and K_2 represent the number of observed and forecasted inflows, respectively; Q represents the standard normal distribution function; Q^{-1} is its inverse function; and W_k and X_k represent the normal quantiles of the H_k and S_k , respectively; $\bar{\Gamma}_k$ and $\bar{\Lambda}_k$ represent the marginal distribution functions of the H_k and S_k , respectively.

BC model

By using the MG model, the marginal distribution function should be determined at first. In real-time operation, it is difficult to identify the marginal distribution function and calibrate the parameters. However, the BC model can transform the non-normal distribution data into a normal distribution, and only one parameter needs to be calibrated.

In the BPF model, the observed and the forecasted inflows need to be transformed to the normal distribution space, and then, the distribution functions of the prior and likelihood of the Bayesian theory are established, respectively. In the BPF model, the prior and likelihood density functions must be established in the same space. This study uses a uniform parameter of the BC to transform the observed and the forecasted inflows at the same time period to ensure that the inflows are transformed to the same space. The uniform parameter is generated by averaging the parameters, which are estimated by the observed and the forecasted inflows, separately.

The BC transformation model is composed of exponential and logarithmic transformations, as shown in Equation (3).

The BC model parameter is calibrated by a maximum likelihood model (Madadgar & Moradkhani 2014; Chen et al. 2015). The BC model consists of input data, a transformation parameter, and output data. $X = \{x_1, x_2, \dots, x_n\}$ represents the input data sequence, λ represents the transformation parameter, and $Y = \{y_1, y_2, \dots, y_n\}$ is the output data sequence.

$$y_j = \begin{cases} \frac{x_j^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log x_j & \lambda = 0, \end{cases} \quad j = 1, 2, \dots, n \quad (3)$$

where x_j represents the flow information, and the value is greater than zero. If there is a number less than zero, the whole sequence needs to be shifted to make the number greater than zero, meaning $(x_j + a) > 0$. The λ can be estimated by the maximum likelihood model.

The λ estimation by the maximum likelihood model is as follows.

The distribution function of y_j is defined as $y_j \sim N(X\beta, \sigma^2 I)$. The likelihood function contains the parameters of β and σ^2 , which can be obtained by fixing the λ , as shown in Equation (4).

$$L(\beta, \sigma^2) = \frac{1}{(\sqrt{2\pi^2})^n} \times \exp\left\{-\frac{1}{2\sigma^2} (y_j - X\beta)^T (y_j - X\beta)\right\} \times J, \quad (4)$$

$$J = \prod_{j=1}^n x_j^{\lambda-1}$$

The derivative of Equation (4) is obtained by taking β and σ^2 as variables, and the estimation functions for β and σ^2 are shown in Equations (5) and (6), respectively.

$$\hat{\beta}(\lambda) = (X^T X)^{-1} X^T y_j \quad (5)$$

$$\sigma^2(\lambda) = \frac{1}{n} y_j^T (I - X^T (X^T X)^{-1} X) y_j = \frac{1}{n} \text{RSS}(\lambda, y_j) \quad (6)$$

The maximum value of the likelihood function can be expressed as follows:

$$L_{\max}(\lambda) = L(\hat{\beta}(\lambda), \sigma^2(\lambda)) = (2\pi e)^{n/2} J \left(\frac{\text{RSS}(\lambda, y_j)}{n}\right)^{n/2} \quad (7)$$

Equation (8) is obtained by transforming Equation (7) into a logarithm and omitting the constants unrelated to λ .

Equation (8) established the relationship between Z and λ , and the λ value can be obtained based on this relationship. In this study, the value range of λ is set to $[-0.8, 0.8]$.

$$\ln L_{\max}(\lambda) = -\frac{n}{2} \ln \text{RSS}(\lambda, y_j) + \ln J = -\frac{n}{2} \ln \text{RSS}(\lambda, Z) \quad (8)$$

$$Z = \begin{cases} \frac{y_j}{\left(\prod_{j=1}^n x_j\right)^{(\lambda-1)/n}}, & \lambda \neq 0 \\ (\ln x_j) \left(\prod_{j=1}^n x_j\right)^{1/n}, & \lambda = 0 \end{cases} \quad (9)$$

BC-MG model

In the BPF model, the observed and forecasted inflows need to be transformed, respectively. When using the BC model with the uniform parameter to transform the observed and forecasted inflows, the inflows will be transformed into the same distribution space. In this way, the new uncertainties will be generated in the BPF model for the two inflows obeying different marginal distributions. When using different parameters, the observed and forecasted inflows will be transformed into different distribution spaces. Using the transformed inflows in different spaces as inputs for the BPF model, the results are unreasonable.

To solve the problems as mentioned above, the BC-MG model was proposed to define the marginal distribution function in real-time operation as quickly as possible. In the BC-MG model, the BC model is used first to transform the observed and forecasted inflows in different normal distribution spaces, respectively. Then, the parameters of the normal distribution functions for the transformed observed and forecasted inflows can be calibrated directly. Next, the transformed inflows are transformed again from the normal distribution to the standard normal distribution in a friendly way using the MG model. In the transformation processes, when the inflows cannot be transformed into normal distribution using the BC parameter during $[-0.8, 0.8]$, then the MG needs to test different candidate distributions for the marginal distribution function. Based on the observed and forecasted inflows in the standard normal distribution space, the parameters of the prior and likelihood density functions are determined.

BPF model

Due to the randomness of the inflow processes and the uncertainty of the forecast information, the efficiency of the forecasting inflows in practical applications is reduced. The BPF model is the best methodology to couple the randomness of the inflows and the uncertainty of forecasting inflows, and to provide more inflow information for practical applications, which is beneficial to improve the application efficiency and reliability of the forecasting inflows. The BPF model was established first by Krzysztofiwicz (1999) and applied to short-range flood forecasting. This study establishes the BPF model for the medium-range inflow forecasting with 10 days forecast horizon. In this study, the prior and likelihood density functions of the BPF are established based on the transformed observed and forecasted inflows.

In the BPF, the H_0 represents the observed flow at the current time period, and H_k ($k = 1, 2, \dots, K$) represents the observed flows in the future. The K represents the forecast horizon. Similarly, S_k ($k = 1, 2, \dots, K$) denotes the deterministic forecasting inflows with k forecast horizon. The symbols of h_0 , h_k , and s_k represent the actual values of the H_0 , H_k , and S_k , respectively. When the observed inflow is h_0 at the current time period and the forecasting inflow with k forecast horizon is s_k , the posterior density function of the actual inflow at the forecast horizon k is represented as follows:

$$\phi_h(h_k|s_k, h_0) = \frac{f_k(s_k|h_k)g_k(h_k|h_0)}{\int_{-\infty}^{+\infty} f_k(s_k|h_k)g_k(h_k|h_0)dh_k} \quad (10)$$

where $\phi_h(\cdot)$ represents the posterior density function, $f_k(\cdot)$ represents the likelihood density function, and g_k represents the prior density distribution.

In this study, the one-dimensional linear regression model was used to construct the relationship between forecasting inflow s_k and actual inflow h_k , shown as follows:

$$s_k = a_k h_k + b_k + \theta_k \quad (11)$$

where a_k and b_k represent parameters of the linear regression model, θ_k represents the residual value, which can be calculated as follows:

$$\theta_k = s_k - (a_k h_k + b_k) \quad (12)$$

According to the observed inflows and the forecasted inflows with different forecast horizons, the sequences of the residual values with different forecast horizons can be calculated by Equation (12). It is assumed that the sequences of the residual values are independent, and the normal distributions represented as $\theta_k \sim N(0, \sigma_k^2)$ are used to describe the residual values.

Based on Equation (11) and the distribution of residual values, the likelihood probability density function can be constructed, as follows:

$$f_k(s_k|h_k) \sim N(a_k h_k + b_k, \sigma_k^2) \quad (13)$$

Similarly, the one-dimensional linear regression model is also used to represent the relationship between observed inflow h_0 at the current time period and actual inflow h_k at forecast horizon k , as follows:

$$h_k = c_k h_0 + d_k + v_k \quad (14)$$

where c_k and d_k represent parameters and v_k represents the residual value.

$$v_k = h_k - (c_k h_0 + d_k) \quad (15)$$

Assuming that the sequences of the residual values, calculated by Equation (15), are independent, and described by a normal distribution represented as $v_k \sim N(0, \tau_k^2)$. Then, the prior probability density function can be constructed, as follows:

$$g_k(h_k|h_0) \sim N(c_k h_0 + d_k, \tau_k^2) \quad (16)$$

Substituting Equations (13) and (16) into Equation (10), the posterior probability distribution can be represented as follows:

$$\phi_h(h_k|s_k, h_0) \sim N(C_k h_0 + D_k s_k + F_k, t_k^2) \quad (17)$$

where $C_k = c_k \sigma_k^2 / (a_k^2 \tau_k^2 + \sigma_k^2)$, $D_k = a_k \tau_k^2 / (a_k^2 \tau_k^2 + \sigma_k^2)$, $F_k = (d_k \sigma_k^2 - a_k b_k \tau_k^2) / (a_k^2 \tau_k^2 + \sigma_k^2)$, $t_k^2 = \tau_k^2 \sigma_k^2 / (a_k^2 \tau_k^2 + \sigma_k^2)$.

Verification metrics

In this study, these performances of the BPF models are evaluated using the single value and confidence interval.

The single value is evaluated by root mean square error (RMSE). The interval performances are evaluated using the indicators of average relative interval length (ARIL), percentage of observations bracketed by the confidence interval (PCI), and percentage of observations bracketed by the unit confidence interval (PUCI) (Li *et al.* 2011).

The RMSE as shown in Equation (18) is taken as an indicator to evaluate the performances of the single values, i.e., deterministic forecasts from the global forecasting system (GFS) and 50% quantile values of the three BPF models.

$$DRMS = \sqrt{\frac{\sum_{t=1}^{NT} (Q_{obs}^t - Q_{sim}^t)^2}{NT}} \quad (18)$$

where Q_{obs}^t represents the observed values at time step t , Q_{sim}^t represents the forecasting values, and NT represents the length of the entire simulation period.

The indicator values of the ARIL represent the uncertainty range of the forecast inflows. The lower value represents the better performance, which can be calculated by Equation (19). The indicator of the PCI represents the percentage of observed inflows occurring in the forecasting intervals, which can be calculated by Equation (20). The larger range of the forecast interval and the more observed inflows will occur within the interval. This study uses the PUCI as a comprehensive indicator to evaluate the uncertainty range of the inflows and the ability of the forecast interval to include the observed values by combing the ARIL and PCI, as shown in Equation (21).

$$ARIL = \frac{1}{NT} \sum_{t=1}^{NT} \left(\frac{Q_{up}^t - Q_{low}^t}{Q_{obs}^t} \right) \quad (19)$$

$$PCI = \frac{n}{NT} \quad (20)$$

$$PUCI = \frac{PCI}{ARIL} \quad (21)$$

where Q_{up}^t represents the upper boundary of the forecast interval, Q_{low}^t represents the lower boundary of the forecast interval, and n is the number of the observed inflows that occur in the forecast intervals.

To compare the transforming efficiency of the three models on the inflows, the average relative error of quantiles (AREQs) is used to evaluate the ability of the three models to transform data, as shown in Equation (22). Equation (22) is evaluated by comparing the observed quantile and the theoretical quantile, which can represent the deviation between the transformed data and theoretical distribution.

$$AREQ = \frac{1}{NT} \sum_t \left(\frac{1}{W} \sum_{w=1}^W \frac{|q_{obs}^w - q_e^w|}{q_{obs}^w} \right) \quad (22)$$

where the q_{obs}^w and q_e^w represent the quantiles of observed and theoretical, and W represents the number of the transformed data.

CASE STUDY

Huanren hydropower reservoir

In this study, the Huanren hydropower reservoir is taken as an example, which is located in the middle and lower reaches of the Hunjiang River in the northeast of China, as shown in Figure 1. There is approximately 10,400 km² basin area, and the annual average precipitation is about 860 mm. Precipitation in dry and wet seasons varies significantly, and about 75% of the precipitation occurs in the wet seasons (from May to September).

Forecasted and observed data

In this study, the observed precipitation and inflow data of the Huanren Reservoir are obtained from 1967 to 2010. The observed data are used to calibrate and verify the forecasting model. The medium-term numerical weather prediction data of the GFS from 2001 to 2010 is used to forecast the inflows by using the hydrology model. The GFS is developed by the U.S. National Centers for Environmental Prediction. The hydrology model was constructed for the study basin by combining the multiple linear regression model and the Xin'anjiang model (Xu *et al.* 2013). The observed inflows from 1968 to 2007 and the deterministic forecasted inflows from 2001 to 2007 were used to calibrate

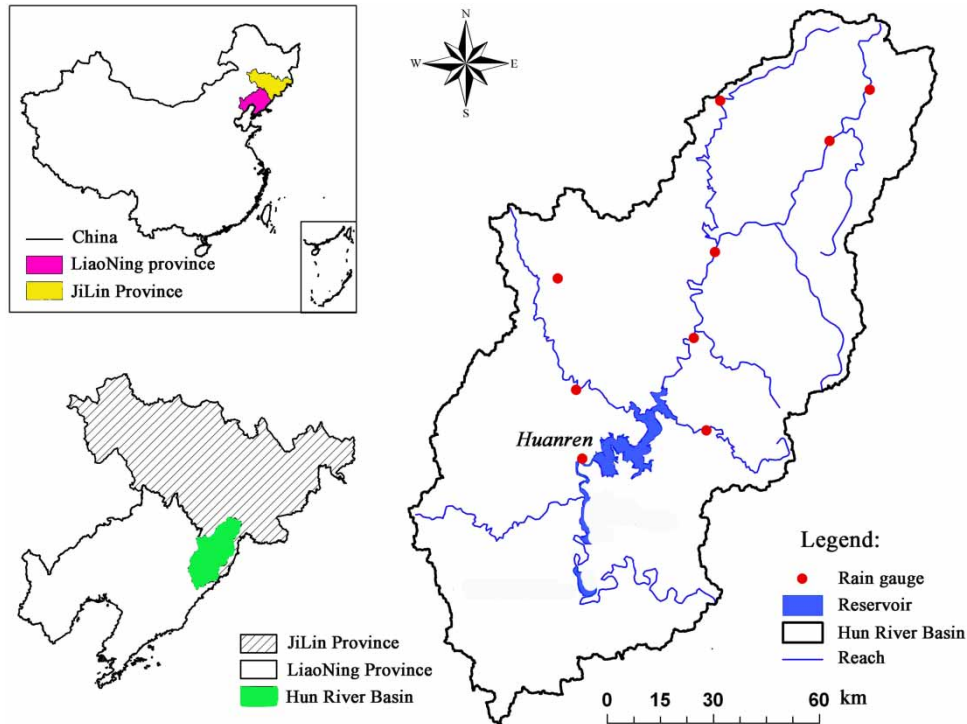


Figure 1 | Location of the Huanren hydropower reservoir.

the model parameters, and the observed and forecasted data from 2008 to 2010 were used to verify the models.

Based on the transformation models, the observed and forecasted data were transformed first by the MG, BC, and BC-MG, respectively. Then, the BPF-MG, BPF-BC, and BPF-BC-MG models were constructed using the transformed data to calibrate the model parameters, respectively. In real-time operation, the inflows of the h_0 , h_k , and s_k for each time period as input data were used by the three models to make probability forecasts, respectively.

RESULTS AND DISCUSSION

Data preprocessing

In this study, the inflow data in a year were divided into 36 time periods, and the inflows for each day in each period with 10-day time horizons were used to calibrate the parameters. In the BPF model, the data of the h_k and s_k should be transformed into the same distribution

space. Then, the data can be used to forecast and simulate. However, in the BPF-BC model, the data of the h_k and s_k may have different transformation parameters. In this study, the parameters in the BPF-BC at the same time period were averaged, and the average parameter was used to transform the data into the same distribution space.

Transformation performances

In the BPF model, the transformation of the forecasted inflow data is relatively easy for the distribution of the forecast residual is closer to the normal distribution. However, the observed inflow data are difficult to transform by the transformation models. To compare the transformation ability of the MG, BC, and BC-MG models on the observed inflows, this study evaluates the efficiency of the models by the AREQ indicator, as shown in Figure 2. The AREQ in Figure 2 indicates that the efficiency of the BC-MG model is significantly better than the MG and BC models from 1 day to 10 days.

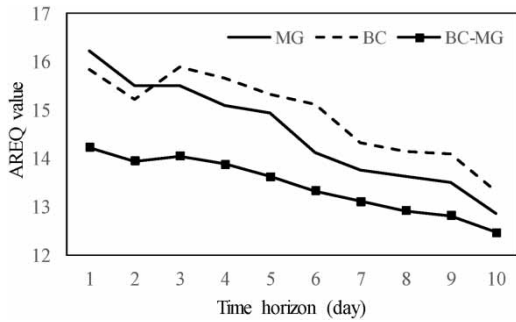


Figure 2 | The variation of AREQ values with time horizon extending.

Due to the large number of time periods and time horizons, this study only lists the 12 time periods in the observed inflow transformation results, as shown in Figure 3, which are transformed by the MG and BC-MG with 1-day time horizon. In Figure 3, the BC-MG model has a better efficiency, and the quantiles of the observed inflows are closer to the 45-degree diagonal. Comparing

the transformed data of the MG model, the results of the extreme value flows deviate from the 45-degree diagonal obviously.

Simulation performances

Based on the BPF-MG, BPF-BC, and BPF-BC-MG models, the daily probabilistic forecasting inflows are simulated during the calibration and validation periods. The values of the 50% quantile of the posterior density function are taken as single probabilistic forecasting inflows, which are in the middle of the posterior probability. Moreover, the confidence intervals of the posterior probability density function as 5–95%, 15–85%, and 25–75% are evaluated as forecasting inflow intervals. The daily single forecasting inflows and 5–95% intervals with 1-day, 3-day, and 7-day forecast horizons from 1st July to 31st August in 2008 are

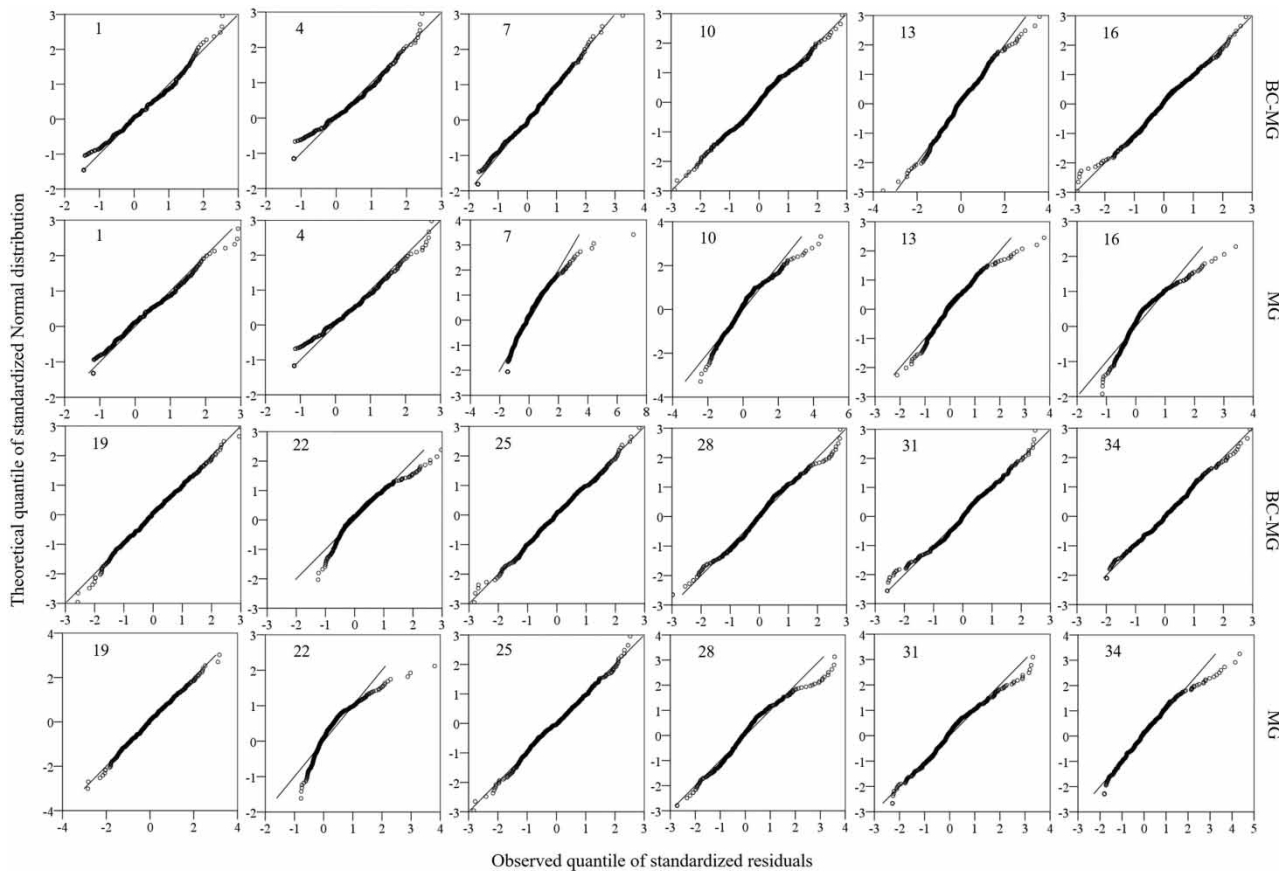


Figure 3 | The QQ (quantile-quantile) plot of MG and BC-MG for observed inflows with 1-day time horizon.

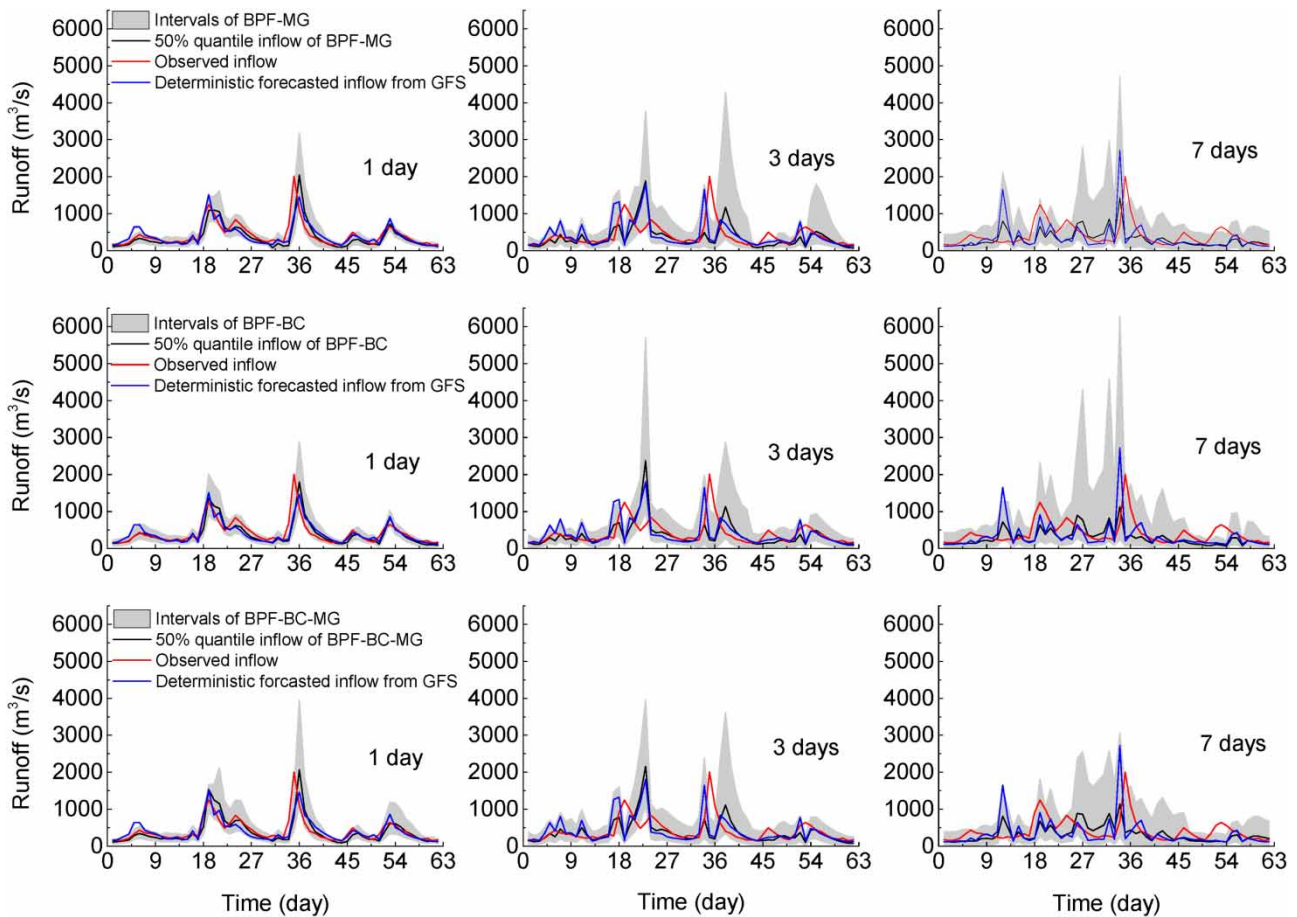


Figure 4 | The daily forecasting inflows and intervals with different forecast horizons.

forecasted, respectively, using the three models, as shown in Figure 4.

In Figure 4, the deviations between observed and 50% quantile inflows are not obviously between the three BPF models. However, the intervals with different models and forecast horizons have significant deviation. When using the forecasting precipitation with 1-day forecast horizon, the accuracy of the forecasting inflows is high for the three models. Thus, the fewer deviations between the inflows with 1-day forecast horizon and the intervals are narrow. Comparing the performances with the forecast horizon extending, the intervals diffuse and the accuracy of the forecasting inflows diminishes. The results indicate that the uncertainty of the forecasting inflows increases with the forecast horizon extending. Comparing the performances of the three BPF models, the intervals of the BPF-BC-MG are narrower than the others, and the BPF-BC model is

the worst performance model. In the BPF-BC model, the BC model is used to transform the inflows by using the uniform parameter. When the observed and forecasted inflows do not obey the same marginal distribution, the transformation efficiency of the two sets of serial data will be affected by the parameter, and this is the main reason that the intervals of the BPF-BC are more diffuse than the others.

Performance evaluation for single values

The values of the RMSE with different forecast horizons during calibration and verification were evaluated, respectively, as shown in Figure 5. The results show that the RMSE values increase with the forecast horizon extending both in calibration and verification. It means that the

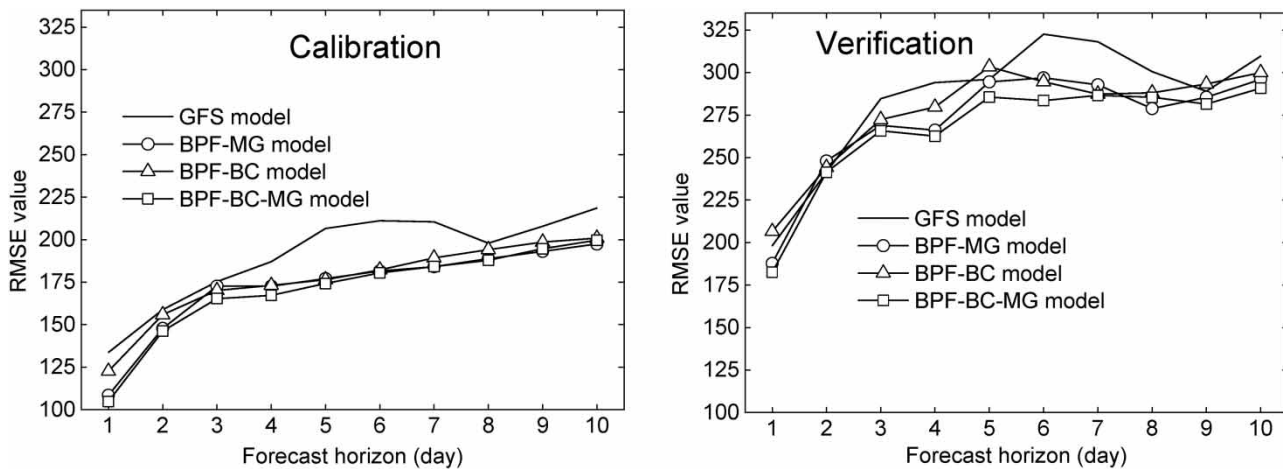


Figure 5 | The RMSE values of the probabilistic and deterministic forecasts with different forecast horizons.

uncertainties of the forecasted inflows increase with the forecast horizon extending.

The RMSE values of the three BPF models are lower than that of the deterministic forecasts from GFS in the entire forecast horizon. The deterministic forecasts from GFS are represented as the GFS model in Figure 5. The results in Figure 5 demonstrate that the three BPF models have correction capability by combining the prior and likelihood density, especially during the forecast horizons from 4 days to 7 days. In general, the 50% quantile values of the BPF-BC-MG have higher accuracy than the others. During the forecast horizons from 1 day to 3 days, the deviations of the values between the GFS model and three BPF models are less than those of the other forecast horizons. Thus, the linear relationship between the forecasted and observed data is relatively consistent from 1 day to 3 days, and the likelihood density has little effect on the posterior density in the BPF. Thus, the 50% quantile values during 1 day to 3 days are close to the observed inflows.

Performance evaluation for intervals

The indicator values of the ARIL represent the variation of the uncertainties of the inflows with forecast horizon extending, as shown in Figure 6. Comparing the variations of the 5–95%, 15–85%, and 25–75%, the results demonstrate that the posterior density functions are more

diffuse with the forecast horizon extending. However, relying only on the ARIL indicators is not enough to show the forecast efficiency between the three BPF models.

The indicator of the PCI represents the percentage of the observed inflows occurred in the forecasting intervals, as shown in Figure 7. The larger values of the PCI represent the better performances of the forecasting intervals. The PCI values in Figure 7 are gradually reduced with the interval narrowing. However, the PCI values of the BPF-BC-MG model maintain high level, and the results indicate that this model has better efficiency than the others.

The indicators of the ARIL and PCI reflect the different characteristics of the intervals. The lower value of the ARIL may include fewer observed inflows and have lower values of the PCI. Thus, the PUCI are used to evaluate the performances by combining the values of the PCI and ARIL. The higher the value of the PUCI, the confidence intervals have better performance.

According to Equation (21), the PUCI values of the three models during calibration and verification are evaluated, as shown in Figure 8. With the forecast horizon extending, the PUCI values of the forecast intervals in the three models gradually decreased. Comparing the performances of the PUCI, the BPF-BC-MG model has greater PUCI values than those of the BPF-MG and BPF-BC models both in calibration and verification.

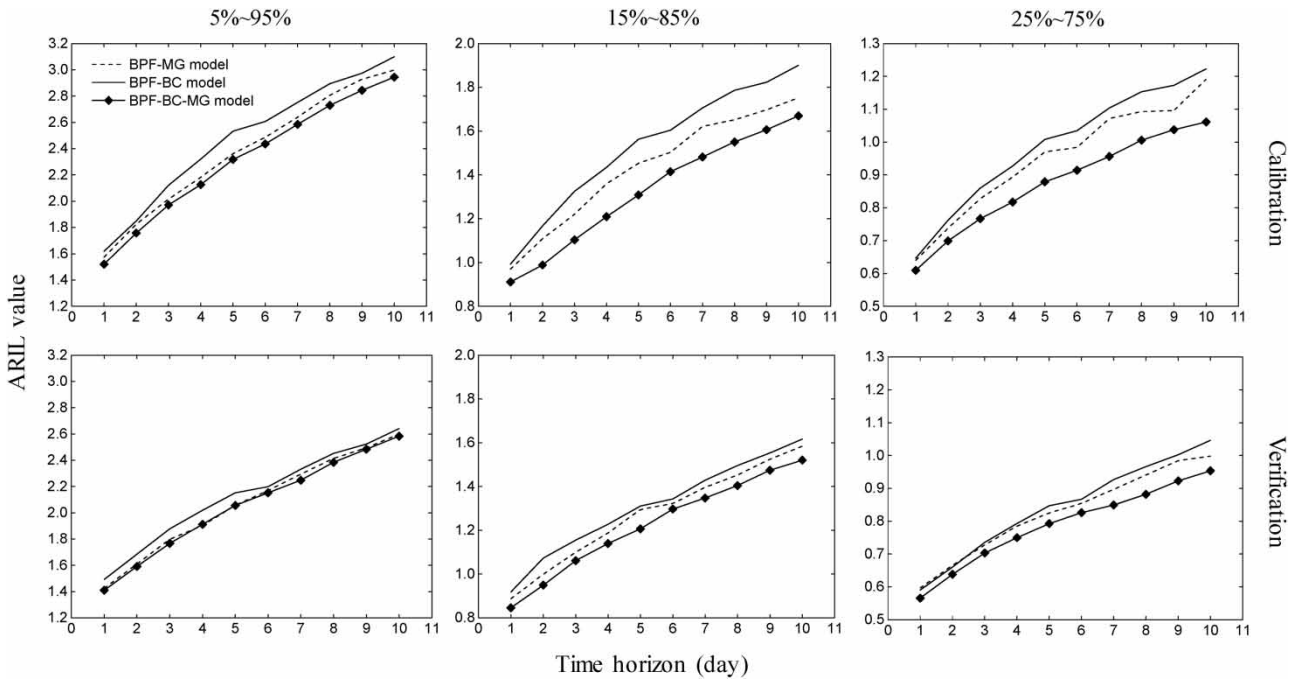


Figure 6 | The variations of ARIL values with different time horizons and confidence intervals.

With the forecast interval narrows, the PUCI value gap of the three models becomes obvious, and the efficiency of the BPF-BC-MG model is fully demonstrated.

Meanwhile, the PUCI values of the 25–75% are the highest, and it indicates that the interval of the 25–75% has a smaller interval width and has higher inclusion efficiency

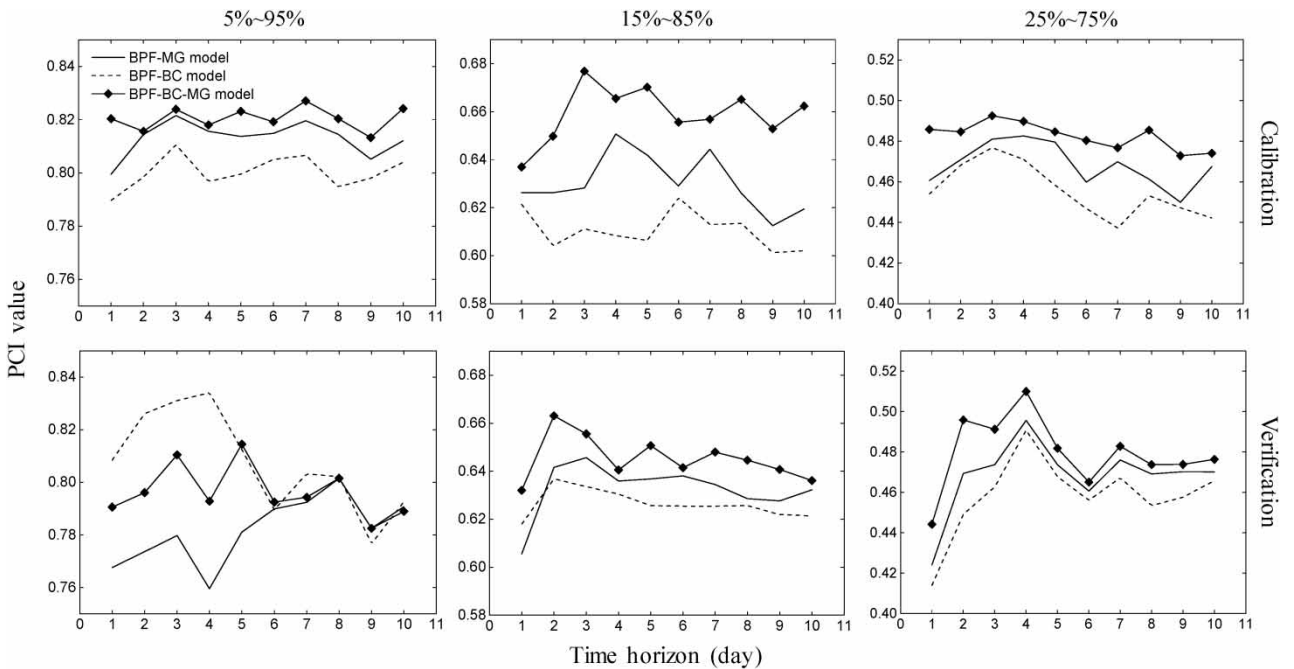


Figure 7 | The variations of PCI values with different time horizons and confidence intervals.

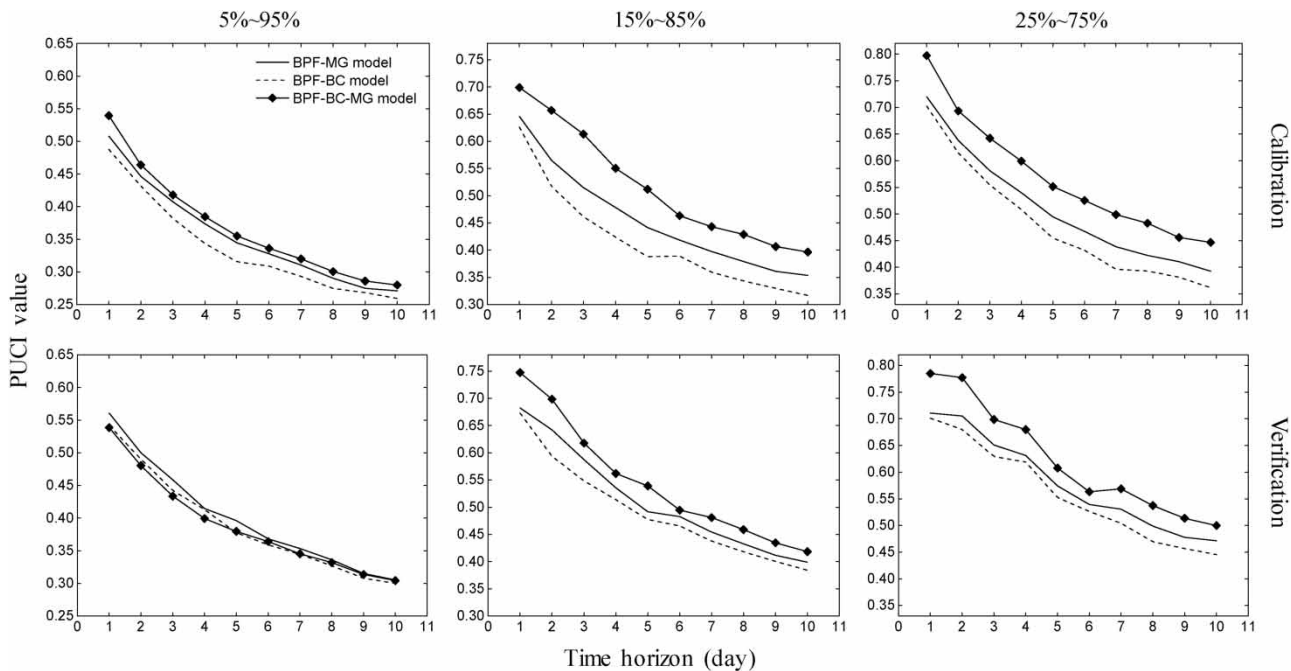


Figure 8 | The variations of PUCI values with different time horizons and confidence intervals.

for the inflow data. The results illustrate that the forecasting intervals of the BPF-BC-MG model have better efficiency than that of the BPF-MG and the BPF-BC models.

CONCLUSIONS

This paper investigates the efficiency and stability of the BPF model to describe the uncertainties of forecasting inflows by using different methodologies to transform non-normal distribution data into normal distribution space. Three models, i.e., BPF-MG, BPF-BC, and BPF-BC-MG models, have been developed to describe the uncertainties of the observed and forecasted inflows, respectively. The models are tested on the Huanren hydropower reservoir in China. The findings obtained are summarized below.

1. The MG model is the most popular transform model in the BPF model. In real-time operation, the marginal distribution functions for each time period should be identified. However, it is difficult to identify the functions, and this is the main barrier that applies the BPF model to real-time operation.
2. The BC model can transform non-normal distribution data into normal distribution space without marginal distribution function. However, in the BPF model, the optimal transform parameters of the observed and forecasted inflows may be different. Thus, the average value of the parameters is used to transform the two sets of inflows into the same normal distribution space. The efficiency and stability of the BPF-BC will be affected by the transformed data.
3. In the BPF-BC-MG model, the observed and forecasted inflows are transformed into different normal distribution spaces by using the BC model. Then, the transformed inflows are transformed again into standard normal distribution space by using the MG model. In real-time operation, the density distribution functions of the prior and likelihood are determined quickly, and the probabilistic forecasts with high accuracy are obtained.
4. In this study, the medium-term precipitations of the GFS are used to forecast the inflows. The uncertainties of the inflows increase with the forecast horizon extending. The results demonstrate that the BPF model is a useful methodology to describe the uncertainties of the observed and forecasted inflows.

ACKNOWLEDGEMENTS

This research is supported by the National Natural Science Foundation of China (Grant Nos. 51609025, 91547111, and 51709108), the Open Fund Approval (SKHL1713, 2017), Chongqing Technology Innovation and Application Demonstration Project (cstc2018jscx-msybX0274 and cstc2016shmszx30002), and the Hun River Cascade Hydropower Reservoirs Development Company, Ltd.

REFERENCES

- Ahmadisharaf, E., Kalyanapu, A. J., Lillywhite, J. R. & Tonn, G. L. 2018 A probabilistic framework to evaluate the uncertainty of design hydrograph: case study of Swannanoa river watershed. *Hydrological Sciences Journal* **63** (12), 1776–1790.
- Barbetta, S., Coccia, G., Moramarco, T. & Todini, E. 2016 Case study: a real-time flood forecasting system with predictive uncertainty estimation for the Godavari River, India. *Water* **8** (10), 465.
- Barbetta, S., Coccia, G., Moramarco, T. & Todini, E. 2018 Real-time flood forecasting downstream river confluences using a Bayesian approach. *Journal of Hydrology* **565**, 516–523.
- Biondi, D. & Luca, D. 2013 Performance assessment of a Bayesian forecasting system (BFS) for real-time flood forecasting. *Journal of Hydrology* **479**, 51–63.
- Biondi, D., Versace, P. & Sirangelo, B. 2010 Uncertainty assessment through a precipitation dependent hydrologic uncertainty processor: an application to a small catchment in southern Italy. *Journal of Hydrology* **386** (1–4), 38–54.
- Chen, F., Jiao, M. & Chen, J. 2013 The meta-Gaussian Bayesian processor of forecasts and associated preliminary experiments. *Acta Meteorologica Sinica* **27** (2), 199–210.
- Chen, Y., Yuan, W., Xia, J., Fisher, J. B., Dong, W., Zhang, X., Liang, S., Ye, A., Cai, W. & Feng, J. 2015 Using Bayesian model averaging to estimate terrestrial evapotranspiration in China. *Journal of Hydrology* **528**, 537–549.
- Han, S. & Coulibaly, P. 2017 Bayesian flood forecasting methods: a review. *Journal of Hydrology* **551**, 340–351.
- Hao, Z., Hao, F., Singh, V. P., Sun, A. Y. & Xia, Y. 2016 Probabilistic prediction of hydrologic drought using a conditional probability approach based on the meta-Gaussian model. *Journal of Hydrology* **542**, 772–780.
- Krzysztofowicz, R. 1999 Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research* **35** (9), 2739–2750.
- Krzysztofowicz, R. 2014 Probabilistic flood forecast: exact and approximate predictive distributions. *Journal of Hydrology* **517**, 643–651.
- Krzysztofowicz, R. & Kelly, K. S. 2000 Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resources Research* **36** (11), 3265–3277.
- Krzysztofowicz, R. & Maranzano, C. J. 2004 Hydrologic uncertainty processor for probabilistic stage transition forecasting. *Journal of Hydrology* **293** (1–4), 57–73.
- Li, L., Xu, C. Y., Xia, J., Engeland, K. & Reggiani, P. 2011 Uncertainty estimates by Bayesian method with likelihood of AR (1) plus normal model and AR (1) plus multi-normal model in different time-scales hydrological models. *Journal of Hydrology* **406** (1–2), 54–65.
- Liang, Z., Li, Y., Hu, Y., Li, B. & Wang, J. 2017 A data-driven SVR model for long-term runoff prediction and uncertainty analysis based on the Bayesian framework. *Theoretical and Applied Climatology* **133** (1–2), 137–149.
- Liu, D., Li, X., Guo, S., Rosbjerg, D. & Chen, H. 2015 Using a Bayesian probabilistic forecasting model to analyze the uncertainty in real-time dynamic control of the flood limiting water level for reservoir operation. *Journal of Hydrologic Engineering* **20** (2), 04014036.
- Liu, Z., Guo, S., Zhang, H., Liu, D. & Yang, G. 2016 Comparative study of three updating procedures for real-time flood forecasting. *Water Resources Management* **30** (7), 2111–2126.
- Liu, Z., Guo, S., Xiong, L. & Xu, C. Y. 2017 Hydrological uncertainty processor based on a copula function. *Hydrological Sciences Journal* **63** (1), 74–86.
- Ma, Z., Li, Z., Zhang, M. & Fan, Z. 2013 Bayesian statistic forecasting model for middle-term and long-term runoff of a hydropower station. *Journal of Hydrologic Engineering* **18** (11), 1458–1463.
- Ma, Y., Hong, Y., Chen, Y., Yang, Y., Tang, G., Yao, Y., Long, D., Li, C., Han, Z. & Liu, R. 2018 Performance of optimally merged multisatellite precipitation products using the dynamic Bayesian model averaging scheme over the Tibetan plateau. *Journal of Geophysical Research: Atmospheres* **123** (2), 814–834.
- Madadgar, S. & Moradkhani, H. 2014 Improved Bayesian multimodeling: integration of copulas and Bayesian model averaging. *Water Resources Research* **50** (12), 9586–9603.
- Qu, B., Zhang, X., Pappenberger, F., Zhang, T. & Fang, Y. 2017 Multi-model grand ensemble hydrologic forecasting in the Fu river basin using Bayesian model averaging. *Water* **9** (2), 74.
- Reggiani, P. & Weerts, A. H. 2008 A Bayesian approach to decision-making under uncertainty: an application to real-time forecasting in the river Rhine. *Journal of Hydrology* **356** (1–2), 56–69.
- Tanessong, R. S., Vondou, D. A., Igri, P. M. & Kamga, F. M. 2017 Bayesian processor of output for probabilistic quantitative precipitation forecast over central and West Africa. *Atmospheric and Climate Sciences* **7** (3), 263–286.
- Weerts, A. H., Winsemius, H. C. & Verkade, J. S. 2011 Estimation of predictive hydrological uncertainty using quantile regression: examples from the national flood forecasting system (England and Wales). *Hydrology and Earth System Sciences* **15** (1), 255–265.

- Xie, W., Zhang, P., Chen, R. & Zhou, Z. 2019 A nonparametric Bayesian framework for short-term wind power probabilistic forecast. *IEEE Transactions on Power Systems* **34** (1), 371–379.
- Xu, W., Peng, Y. & Wang, B. 2013 Evaluation of optimization operation models for cascaded hydropower reservoirs to utilize medium range forecasting inflow. *Science China Technological Sciences* **56** (10), 2540–2552.
- Zhang, X., Liang, F., Yu, B. & Zong, Z. 2011 Explicitly integrating parameter, input, and structure uncertainties into Bayesian neural networks for probabilistic hydrologic forecasting. *Journal of Hydrology* **409** (3–4), 696–709.
- Zhong, Y., Guo, S., Ba, H., Xiong, F., Chang, F.-J. & Lin, K. 2018 Evaluation of the BMA probabilistic inflow forecasts using TIGGE numeric precipitation predictions based on artificial neural network. *Hydrology Research* **49** (5), 1417–1433.

First received 26 February 2019; accepted in revised form 17 June 2019. Available online 10 July 2019