Hydrologic Uncertainty for Bayesian Probabilistic Forecasting Model Based on BP ANN

Chun-tian Cheng[1] Kwok-wing Chau[2] Xiang-yang Li[1]

[1] *Department of Civil Engineering, Dalian University of Technology, Dalian 116024, China*

[2]*Department of Civil & Structural Engineering, The Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong*

E-mail:ctcheng@dlut.edu.cn

Abstract

The Bayesian forecasting system (BFS) consists of three components which can be deal with independently. Considering the fact that the quantitative rainfall forecasting has not been fully developed in all catchment areas in China, the emphasis is given to the hydrologic uncertainty for Bayesian probabilistic forecasting. The procedure of determining the prior density and likelihood functions associated with hydrologic uncertainty is very complicated and there is a requirement to assume a linear and normal distribution within the framework of BFS. These pose severe limitation to its practical application to real-life situations. In this paper, a new prior density and likelihood function model is developed with BP artificial neural network (ANN) to study the hydrologic uncertainty of short-term reservoir stage forecasts based on the BFS framework. Markov chain Monte Carlo (MCMC) method is employed to solve the posterior distribution and statistics of reservoir stage. A case study is presented to investigate and illustrate these approaches using 3 hours rainfall-runoff data from the ShuangPai Reservoir in China. The results show that Bayesian probabilistic forecasting model based on BP ANN not only increases forecasting precision greatly but also offers more information for flood control, which makes it possible for decision makers consider the uncertainty of hydrologic forecasting during decision-making and estimate risks of different decisions quantitatively.

1. Introduction

Conceptual rainfall-runoff models are an important tool for flood forecasting(e.g., Duan et al., 1992). These models commonly include a large number of parameters, which cannot be directly obtained from measurable quantities of catchment characteristics, and hence model calibration is entailed(Cheng et al.,2002, 2005). However, regardless of the methodology used, most hydrologic models suffer from a generic problem that different models and parameter sets have a similar performance to simulate flood forecasting. The intricacy is originated by several causes, such as input uncertainty, parameters uncertainty and model uncertainty. Recently, considerable attention has been given to assess the uncertainty of rainfall-runoff simulations. In order to deal with the uncertainty problem of hydrologic forecasting, Krzysztofowicz (1999) proposed a Bayesian forecasting system (BFS), which describes the uncertainty of hydrologic forecast quantitatively by using probabilistic distribution function. BFS furnishes more information in making flood operation decisions, so that the decision makers can consider different types of uncertainty and estimate risks and consequences of various alternatives quantitatively. However, the procedure of determining the prior density and likelihood functions is very complicated and there is a requirement to assume a linear and normal distribution within the framework of BFS. These pose severe limitation to its practical application to real-life situations, details on which are described next.

The purpose of this article is to present a new method to estimate prior density and likelihood function model for adopting BFS framework to forecast runoff. Considering the fact that BP ANN technique has the capability to model various characteristics of hydrologic resources system, including randomness, fuzziness, non-linearity, etc., it is appropriate to develop prior density and likelihood functions. In this paper, a new prior density and likelihood function model is developed, on the basis of the BFS framework, with back propagation (BP) artificial neural network (ANN) to study the uncertainty of hydrologic forecasting using 3 hours rainfall-runoff data from the ShuangPai Reservoir in China.

Third International Conference on Natural Computation (ICNC 2007)
0-7695-2875-9/07 $25.00 © 2007 IEEE

2. Hydrologic uncertainty for Bayesian Probabilistic Forecasting

The Bayesian forecasting system is a general theoretical framework for probabilistic forecasting via deterministic hydrologic model **(**Krzysztofowicz , 1999)**.** The BFS decomposes the total uncertainty into input uncertainty and hydrologic uncertainty which are quantified independently and then integrated into a predictive distribution. Uncertainty associated with input and hydrologic can be independently dealt with according to the further studies from Krzysztofowicz and his co-investigators (Krzysztofowicz and Herr, 2001;Krzysztofowicz,2002;Krzysztofowicz and Maranzano,2004).Here, we only discuss the hydrologic uncertainty because the quantitative rainfall forecasting isn't available at the current stage in China. It is considered that $H_0 = H_{t_0}, H_{t_0-1}, ..., H_{t_0-p+1}$ denotes the early stages of observed flow series at the prediction time; $\{H_n, n = 1,2,...,N\}$ represents the actual flow series to be predicted; $\{S_n, n = 1,2,...,N\}$ denotes the output flow series generated by a corresponding deterministic hydrologic forecast model; n is the prediction period. Moreover, the observed values of H_0 and H_n and the estimated value of S_n are donated by the small letter case h_0, h_n and s_n, respectively. For any lead time $n(n = 1, 2, ..., N)$ and any observed flow stage $H_0 = h_0$,the expected density of model output is given by the total probability

$$\kappa(s_n \mid h_0) = \int_{-\infty}^{\infty} f(s_n \mid h_n, h_0) g(h_n \mid h_0) dh_n \quad (1)$$

and the posterior density function of actual discharge H_n, conditional on model discharge stage $S_n = s_n$ is as follows:

$$\phi(h_n \mid s_n, h_0) = \frac{f(s_n \mid h_n, h_0) g(h_n \mid h_0)}{\kappa(s_n \mid h_0)} \quad (2)$$

where $g(h_n \mid h_0)$ is the prior density function representing the uncertainty of the observed flow series; $f(s_n \mid h_n, h_0)$ is the likelihood function of h_n representing the prediction capability of the forecast model.

Equation 2 demonstrates that the posterior density depends on the prior density function and likelihood function. Currently, the research on prior density and likelihood functions are still in an exploratory stage. Krzysztofowicz and his co-investigators proposed a linear-normal model in addressing this problem (Krzysztofowicz and Herr, 2001; Krzysztofowicz, 2002; Krzysztofowicz and Maranzano, 2004). In their model, the actual flow H_n and the predicted flow S_n generated by the deterministic model are firstly undergoing a series of transformation. The transformed $\{h_n \mid h_0\}$ and $\{h_n \mid s_n, h_0\}$ are assumed to be linear and normally distributed. Linear regression method is then employed to determine the posterior density of H_n in the transformed space, from which the posterior density function of H_n in the original space can be found. However, the procedure in this method is very complicated and there is a requirement to assume a linear and normal distribution. These pose severe limitation to its practical application to real-life situations. Since BP ANN technique has the capability to model various characteristics of hydrologic resources system, including randomness, fuzziness, non-linearity, etc., it is appropriate to develop prior density and likelihood functions

3. BP ANN with Prior Density and Likelihood Functions

Generally, error of hydrologic simulation can be estimated in normal distribution. Contraditionally, prior density and likelihood functions can be determined by statistics methods based on historical observations. Here, we presented a BP ANN to estimate them.

It is generally recognized that the flow series in reservoirs can be simulated by the pth order Markov process. Sample series

$\{(h_n, h_0)_i :, n = 1,2,..., N; i = 1,2,...m\}$ and

$\{(s_n, h_n, h_0)_i :, n = 1,2,..., N; i = 1,2,...m\}$ can be acquired from historical observation data. In the above formulation, n is the prediction period; m is the length of time series; h_0 is the flow series in the early stage, $h_0 = h_{t_0}, h_{t_0-1}, h_{t_0-2}, ..., h_{t_0-p+1}$; p is the model order. Based on these two sample series, a three layer BP ANN model with the prior density and likelihood functions is developed.

The prior density distribution model defined by a BP ANN can be represented by the following equation:

$$H_n = g(H_n \mid H_0) = g(H_0) + \varepsilon_n \quad (3)$$

where n is the prediction period; t_0 is the time at the prediction; g is the non-linear representation of the prior density by BP ANN model; H_0 is the early stages of observed flow series at time t_0,

$H_0 = H_{t_0-p+1},...,H_{t_0-1},...,H_{t_0}$; p is the model order; ε_n is the model residue error with a random nature. Assuming that the residue error follows a normal distribution, then $\varepsilon_n \sim N(0, RMSE^2)$ where RMSE is the root mean square error of the model. It can be deduced from eq (3) that H_n will follow a normal distribution $H_n \sim N(g(h_{t_0}, h_{t_0-1},...,h_{t_0-p+1}), RMSE^2)$ under the conditions of $H_{t_0} = h_{t_0}$, $H_{t_0-1} = h_{t_0-1}$, …, $H_{t_0-p+1} = h_{t_0-p+1}$.

The likelihood function model defined by a BP ANN can be represented by the following equation:

$$S_n = f(S_n \mid H_n, H_0) = f(H_n, H_0) + \varepsilon_n \tag{4}$$

where the various parameters have the same meaning as in the above section, f is the non-linear representation of the likelihood function by BP ANN model. S_n will follow a normal distribution $S_n \sim N(f(h_n, h_0), RMSE^2)$ under the conditions of $H_n = h_n$, $H_0 = h_0$.

Through BP ANN, we develop the prior density function $g(h_n \mid h_0)$ and the likelihood function $f(s_n \mid h_n, h_0)$ of the observed flow series in Shuangpai Reservoir. Since the universal constant $\kappa(s_n \mid h_0)$ entailed for the posterior density function $\phi(h_n \mid s_n, h_0)$ as shown in eq (1) is unknown, the actual value of the flow posterior density function $\phi(h_n \mid s_n, h_0)$ cannot be determined. Under such conditions, the Markov Chain Monte Carlo (MCMC) method can be employed to sample the posterior flow series h_n at random, based on the universal probability distribution $f(s_n \mid h_n, h_0) g(h_n \mid h_0)$. In this way, the extreme distribution of h_n can also be found.

The basic principle of the MCMC method is to generate initially a Markov chain such that its extreme distribution will converge to the flow posterior density function $\phi(h_n \mid s_n, h_0)$. Monte Carlo method is then used to sample this Markov chain. The probability distribution of the resulting flow sample series $h_n^0, h_n^1, h_n^2, ...$ will converge to $\phi(h_n \mid s_n, h_0)$.

The MCMC method adopted here is the commonly used Metropolis-Hasting algorithm, which is employed to sample and search the flow in Shuangpai Reservoir. The basic procedures of the Metropolis-Hasting algorithm are briefly introduced as follows:

1)Initialize to set I = 0 and $h_n^i = s_n$;

2)Generate a new h_n^* based on the transformation probability $G(h_n^* \mid h_n^i)$;

3)Compute the acceptance probability $A(h_n^i \mid h_n^*)$ of h_n^*

$$A(h_n^i,\ h_n^*) = \min\{1,\ \frac{G(h_n^i \mid h_n^*) f(s_n \mid h_n^*, h_0) g(h_n^* \mid h_0)}{G(h_n^* \mid h_n^i) f(s_n \mid h_n^i, h_0) g(h_n^i \mid h_0)}\} \tag{5}$$

4)Generate an evenly distributed random number u ~ U[0,1]

5)If $u < A(h_n^i \mid h_n^*)$, then $h_n^{i+1} = h_n^*$. Otherwise, $h_n^{i+1} = h_n^i$.

6)Set i = i+1 and repeat steps 2) to 6).

In order to obtain better sampling effect and faster convergence speed in using the Metropolis-Hasting algorithm, an appropriate transformation probability $G(h_n^* \mid h_n^i)$ of h_n^* is required. If the variation of h_n generated by $G(h_n^* \mid h_n^i)$ is too small, the convergence speed in searching the posterior density function will become very slow. On the contrary, if the variation of h_n generated by $G(h_n^* \mid h_n^i)$ is too large, it might lead to an inconsistency with the posterior distribution, which in turn might also result in a slow convergence speed. After having considered the characteristics of flow distribution in Shuangpai Reservoir, it is assumed that the distribution limit of h_n is $[\max\{0, s_n - 1000\}, s_n + 1000]$. In this paper, the transformation probability $G(h_n^* \mid h_n^i)$ is defined by a randomly generated and evenly distributed number $h_n^* \sim U[\max\{0, s_n - 1000\}, s_n + 1000]$.

4. Case study

Shuangpai Reservoir in Hunan Province is used as case study here. The catchment basin in Shuangpai Reservoir is located at the wet area with plentiful amount of rainfall. The prediction model adopted here is the XAJ model with 17 parameters(Cheng et al. , 2002, 2005). 10 flooding events with 3 hours time step from 1984 to 1990 in Shuangpai Reservoir are used to train the prior density and likelihood functions of the BP ANN model. 10 flooding events from 1990 to 2002 are used to verify the model as well as to analyze the posterior distribution.

Third International Conference on Natural Computation (ICNC 2007)
0-7695-2875-9/07 $25.00 © 2007 IEEE

In the modeling process, the data sets of discharge were scaled to the range between 0 and 1 as follow:

$$q_i' = \frac{q_i - q_{min}}{q_{max} - q_{min}} \quad (6)$$

where q_i' is the scaled value, q_i is the original discharge value and q_{min}, q_{max} are respectively the minimum and maximum of discharge series.

The correlation coefficient (CORR) and the root mean square error (RMSE) are used to evaluate the training performances.

$$\text{CORR} = \frac{\frac{1}{m}\sum_{i=1}^{m}(h(i)-\bar{h})(h_f(i)-\overline{h_f})}{\sqrt{\frac{1}{m}\sum_{i=1}^{m}(h(i)-\bar{h})^2} * \sqrt{\frac{1}{m}\sum_{i=1}^{m}(h_f(i)-\overline{h_f})^2}} \quad (7)$$

and

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(h(i)-h_f(i))^2} \quad (8)$$

Where $h(i)$ is observed value; $h(f)$ is simulated value; $\bar{h}$ is the average value of observed series; $\bar{h}_f$ is the average value of simulated series using ANN.

The results during the training of the of BP ANN show that the most effective modeling outcome is attained with the model order p = 3 and the number of hidden layer r = 5 (represented by ANN(3,5) later). In such case, the correlation coefficient (CORR) has the largest value whilst the root mean square error (RMSE) has the smallest value. Hence, this set of parameters is adopted in the prior density and likelihood function BP ANN model for flow prediction.

The application of the abovementioned prior density and likelihood function BP ANN model to Shuangpai Reservoir on 16 April 2000 is used to illustrate a typical example. In this example, the prediction period (n) is 1 and the MCMC method (Metropolis-Hasting) is adopted for 10,000 iterations. Figures 1 and 2 show the observed flow, forecasted flow by XAJ model and posterior mean flow with prediction periods (n = 1, 2) during the entire process of the flooding event no. 20000416, respectively. It can be observed that the use of the BFS based on BP ANN will enhance the accuracy of flood forecasting to different degree at different prediction period. In this regard, the accuracy of Bayesian probabilistic forecasting generally decreases when the prediction period becomes longer. This can be explained by the increasing uncertainty of the prior density when the prediction period is longer.

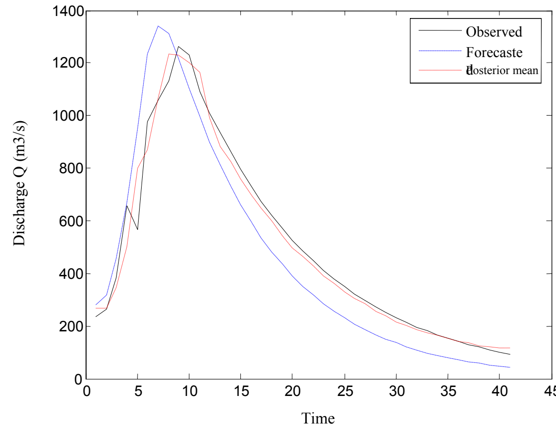

Fig.1 The observed flow, forecasted flow (by XAJ model) and posterior mean flow (n = 1)

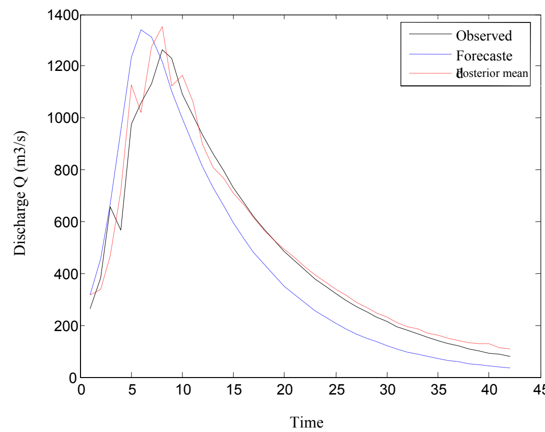

Fig.2 The observed flow, forecasted flow (by XAJ model) and posterior mean flow (n = 2)

Table 1 lists the results of certainty coefficient, RMSE and quantity balance coefficient by both the deterministic XAJ model and the Bayesian probabilistic forecasting for various flooding events with a prediction period (n) of 1. In the Bayesian probabilistic forecasting, the parameters are determined such that the flood forecasting value is taken as the mean value of the posterior probability density. It can be noted from Table 1 that, with the use of Bayesian probabilistic forecasting, there are obvious improvement to the accuracy in flood forecasting for each flooding event.

5. Conclusions

BFS is independent of a specific deterministic hydrologic forecasting model and can be integrated with any deterministic models without attaching any additional assumptions. In this way, this method furnishes a framework for various types of

0-7695-2875-9/07 $25.00 © 2007 IEEE

probabilistic

Table 1 Statistics of outputs by XAJ Model and probabilistic forecasting (n=1)

Flooding event number	Deterministic forecasting			Probabilistic forecasting		
	Certainty coefficient (%)	RMSE	Quantity balance coefficient	Certainty coefficient (%)	RMSE	Quantity balance coefficient
19900530	91.7	204.5	1.046	94.8	162.4	1.096
19940421	93.9	357.6	0.978	95.4	212.3	1.018
19950614	85.2	179.3	0.922	89.3	145.4	1.006
19990526	96.4	258.0	1.001	97.2	229.2	0.958
19990618	83.9	157.6	0.961	91.3	124.5	1.025
19990831	90.1	163.5	1.051	93.7	129.0	1.031
20000416	91.2	103.6	0.883	97.5	55.0	0.985
20010402	87.3	170.7	0.990	91.9	135.9	1.007
20010606	84.5	528.1	1.154	96.4	252.9	1.049
20020513	82.1	303.2	1.132	96.6	132.9	1.064

hydrologic forecasting system. The prior density and likelihood function BP ANN model presented in this paper is capable of simulating the attributes of hydrologic and water resources systems such as randomness, non-linearity, and so forth. Through the application of this model and the use of MCMC method, the posterior density function of the actual flow attains significant enhancement in results over its counterparts by the XAJ forecasting model. The Bayesian probability flood forecasting system is able to fine-tune the deterministic hydrologic models in real-time, by describing quantitatively the uncertainty of hydrologic forecasting with a probability distribution. In addition, it furnishes the posterior density function of the actual flow H_n for different prediction periods. It provides more information in making flood operation decisions, so that the decision makers can consider different types of uncertainty and estimate risks and consequences of various alternatives quantitatively. In this way, the optimal integration of forecast and decision can be realized.

6. Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. 50479055) and the Internal Competitive Research Grant of Hong Kong Polytechnic University (A-PE26).

7. References

[1]Bates B.C., Campbell E. P. A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resources Research*, 2001, 37(4): 937-94

[2]Cheng C.T., Ou.C.P., Chau K.W.. Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration. *Journal of Hydrology*, 2002, 268, 72-86

[3]Cheng C.T., , Wu X.Y, Chau K.W. Multiple criteria rainfall–runoff model calibration using a parallel genetic algorithm in a cluster of computers. *Hydrological Sciences Journal*, 2005, 50(6), 1069-1088

[4]Duan, Q., Sorooshian, S., Gupta, V.. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resource Research*. 1992, 28(4), 1015-1031

[5]Krzysztofowicz R. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research*, 1999, 35(9): 2739-2750

[6]Krzysztofowicz R., Herr H.D. Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model. *Journal of Hydrology*, 2001 249(1-4): 46-68

[7]Krzysztofowicz R. Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology*, 2002, 268(1-4): 16-40

[8]Krzysztofowicz R. and Maranzano C.J.. Bayesian system for probabilistic stage transition forecasting. *Journal of Hydrology, 2004 , 299(1-2), 15-44*

0-7695-2875-9/07 $25.00 © 2007 IEEE