Quantification of the forecast uncertainty using conditional probability and updating models
Huanhuan Ba, Shenglian Guo, Yixuan Zhong, Shaokun He and Xushu Wu

ABSTRACT

Quantifying forecast uncertainty is of great importance for reservoir operation and flood control. However, deterministic hydrological forecasts do not consider forecast uncertainty. This study develops a conditional probability model based on copulas to quantify forecast uncertainty. Three updating models, namely auto-regressive (AR) model, AR exogenous input model, and adaptive neuro fuzzy inference system model, are applied to update raw deterministic inflow forecasts of the Three Gorges Reservoir on the Yangtze River, China with lead times of 1d, 2d, and 3d. Results show that the conditional probability model provides a reasonable and reliable forecast interval. The updating models both enhance the forecast accuracy and improve the reliability of probabilistic forecasts. The conditional probability model based on copula functions is a useful tool to describe and quantify forecast uncertainty, and using an updating model is an effective measure to improve the accuracy and reliability of probabilistic forecast.

Key words | conditional probability, copula function, flood forecasting, Three Gorges Reservoir, uncertainty estimation, updating model

INTRODUCTION

Flooding is one of the most frequent and severe natural disasters in China; thus, flood forecasting plays a critical role in flood control, disaster reduction, and water resources management (Shen et al. 2015; Wu et al. 2017). Due to the extreme complexity of hydrological processes and the limitation of human knowledge, there inevitably remains uncertainty in the hydrological model output, which mainly results from the errors of model input, model structure, and model parameter (Liu & Gupta 2007). However, traditional and deterministic hydrological forecasting does not consider forecast uncertainty. Decision-makers cannot make a quantitative assessment of the possible risk of flood control and reservoir operation in the absence of uncertainty information. Incorrect decisions may result in casualties and economic losses (Palmer 2002). Therefore, it is essential to quantitatively assess the inherent uncertainty of hydrological forecasts.

To overcome the shortage of deterministic forecasts, many researchers have concentrated on the study of probabilistic forecast in recent years (Krzysztofowicz 1999; Duan et al. 2007; Montanari & Grossi 2008; Todini 2008; Pianosi & Raso 2012; Madadgar & Moradkhani 2014; Khajehei & Moradkhani 2017; He et al. 2018; Liu et al. 2018). The current hydrological uncertainty study mainly focuses on two approaches. The first one is the total factor coupling approach, which quantifies the uncertainties of each link in the rainfall–runoff process, such as rainfall input uncertainty, model structure uncertainty, and model parameter uncertainty. One example of this approach is ensemble forecast that generates a set of forecast values through different inputs, different models, and different parameter optimization methods to express forecast uncertainty (Palmer et al. 2005). Many statistical post-processing methods such as Bayesian model averaging, model conditional processor, and ensemble model output statistics are proposed
to accommodate the under-dispersion of ensemble forecasts (Gneiting et al. 2005; Raftery et al. 2005; Zappa et al. 2008; Coccia & Todini 2011). The other one is the total error analysis approach, which directly processes the deterministic forecasts to realize the probabilistic forecasts, such as hydrological uncertainty processor, quantile regression, and Bayesian joint probability model (Krzysztofowicz & Kelly 2000; Weerts et al. 2010; Zhao et al. 2015). The core of the probabilistic forecast literature mentioned above is to establish multivariate joint distribution, where the multivariate Normal distribution is used commonly (Biondi & Todini 2018). Considering that hydrological variables often do not follow the Normal distribution, many data transformation methods such as Box-Cox transformation (Box & Cox 1964) or Normal quantile transform (Kelly & Krzysztofowicz 1997) have been proposed to handle any marginal distributions.

Most previous probabilistic flood forecasting models only consider the uncertainty of the hydrologic model structure and parameters; this study will consider the uncertainty of hydrologic model structure and parameters, as well as model input. So, a conditional probability model based on copula is developed to post-process daily streamflow forecasts and estimate uncertainty in this study. For the purpose of improving the accuracy and reliability of daily streamflow forecasts, three updating models, including auto-regressive (AR), AR exogenous input (ARXM), and adaptive neuro fuzzy inference system (ANFIS), are applied to correct forecasting errors. The effect of three updating models on the accuracy and reliability of probabilistic forecast is also explored and discussed. The remainder of this paper is organized as follows: the conditional probability model, three updating models, and evaluation criteria are explained in the next section. The study area and the hydrological forecasting model are presented in the section ‘Case studies’. In the following section, the results and discussion are given. Finally, the main conclusions are summarized in the last section.

**METHODOLOGY**

**Copula function**

The Copula function, introduced by Sklar (1959), is capable of constructing the multivariate joint distributions through marginal distribution and correlation structure, which is simple and flexible. It can construct multi-dimensional joint distributions through marginal distributions of several variables and their correlations. Suppose that the marginal distributions of n-dimensional random variables \((X_1, X_2, \cdots, X_n)\) are \(F_{X_i}(x_i) = P_{X_i}(X_i \leq x_i), i = 1, 2, \cdots, n\), where \(n\) is the number of random variables and \(x_i\) is the value of random variables \((X_1, X_2, \cdots, X_n)\). Let \(H_{X_1, X_2, \cdots, X_n}(x_1, x_2, \cdots, x_n)\) be the joint distribution function of random variables \((X_1, X_2, \cdots, X_n)\). According to Sklar’s theorem (Sklar 1959), the multivariate cumulative distribution function (CDF) can be written as follows:

\[
H_{X_1, X_2, \cdots, X_n}(x_1, x_2, \cdots, x_n) = C_\theta(F_{X_1}(x_1), F_{X_2}(x_2), \cdots, F_{X_n}(x_n))
\]

(1)

where \(C\) is the copula function and \(\theta\) is the parameter of \(C\). If \(F_{X_i}(x_i)\) are continuous, then \(C\) is unique.

There are different families of copulas that have been employed in a variety of applications (Nelsen 2006). Of all the copula families, Archimedean copulas find a wide range of applications in hydrology, since they can be easily constructed and possess many good properties like simplicity and adaptability. Gumbel-Hougaard, Clayton, and Frank copulas are three commonly used one-parameter Archimedean copulas, which are widely applied in hydrology by many authors (Zhang & Singh 2006; Duan et al. 2016; Zhong et al. 2017; Yin et al. 2018a, 2018b), and chosen in this study to establish the joint distribution between forecasts and observations. The expression of these copulas and the correlation between their parameters \(\theta\) and Kendall correlation coefficients \(r\) are listed in Table 1. The parameters of the two-dimensional Archimedean copulas can be estimated by Kendall correlation coefficients.

To obtain the optimal copula function, the root mean square error (RMSE) is used to measure the goodness of fit between the empirical and the theoretical joint distribution (Zhang & Singh 2007), and defined as follows:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_{et} - P_i)^2}
\]

(2)

where \(P_{et}\) and \(P_i\) are the empirical and theoretical joint CDF values, respectively. The smaller the value of \(RMSE\) is, the better the copula function.
Table 1 | Three commonly used two-dimensional Archimedean copula functions

<table>
<thead>
<tr>
<th>Copula</th>
<th>(C(u, v))</th>
<th>(\theta)</th>
<th>(\tau)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gumbel-Hougaard</td>
<td>(\exp{-[(\ln u)^\theta + (\ln v)^\theta]^{1/\theta}})</td>
<td>([1, \infty))</td>
<td>(1 - \theta^{-1})</td>
</tr>
<tr>
<td>Clayton</td>
<td>((u^{-\theta} + v^{-\theta} - 1)^{-1/\theta})</td>
<td>((0, \infty))</td>
<td>(\theta/((\theta + 2)\theta))</td>
</tr>
<tr>
<td>Frank</td>
<td>(-\frac{1}{\theta} \ln\left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right])</td>
<td>(R\setminus{0})</td>
<td>(1 + \frac{4}{5} \frac{\theta^2}{\theta} \int_0^t \exp(t) - 1 dt - 1)</td>
</tr>
</tbody>
</table>

### Conditional probability

Let predictand \(H\) be the observed discharge whose realization \(h\) is being forecast. Let estimator \(S\) be the output discharge generated by a corresponding deterministic forecast model whose realization \(s\) constitutes a point estimate of \(H\). Let random variable \(H_0\) represent the observed discharge at the time \(k = 0\) when the forecast is prepared; then \(H_k\) \((k = 1, 2, \ldots, N)\) is the observed discharge and \(S_k\) \((k = 1, 2, \ldots, N)\) is the corresponding deterministic forecast discharge at lead time \(k\). What the rational decision-maker then needs is not a single number \(s_k\), but the distribution function of predictand \(H_k\) conditional on \(S_k = s_k\) (Liu et al. 2018).

Let \(u = F_H(h)\) and \(v = F_S(s)\) be the marginal CDFs, respectively. The corresponding probability density functions (PDFs) are \(f_H(h)\) and \(f_S(s)\). According to Sklar’s theorem, the joint CDF of the \(H\) and \(S\) can be expressed as follows:

\[
F(h, s) = C_h(F_H(h), F_S(s)) = C_\theta(u, v) \tag{3}
\]

For a given \(S = s\), the conditional CDF of observation \(h\) can be derived as follows:

\[
F(h|s) = P(H \leq h|S = s) = P(U \leq u|V = v) = \frac{\partial C_\theta(u, v)}{\partial v} \tag{4}
\]

The corresponding conditional PDF of \(h\) can be expressed as follows:

\[
f(h|s) = dF(h|s)/dh = \frac{\partial^2 C_\theta(u, v)}{\partial u \partial v} \cdot \frac{du}{dh} = c_\theta(u, v) \cdot f_H(h) \tag{5}
\]

where \(c_\theta(u, v) = (\partial^2 C_\theta(u, v)/\partial u \partial v)\) is the PDF of the two-dimensional copula function.

The solution of conditional probability model based on copula (Equation (5)) is given by the following steps:

1. In the model calibration period, the parameters of marginal distribution and copula function are estimated by the data series of observations \((H)\) and forecast values \((S)\), respectively.
2. In the real-time flood forecasting, the observations are unknown for the lead times of 1d, 2d, and 3d. We discretize and determine the distribution range \([h_{min}, h_{max}]\) for the observation \(H\), according to its real flood probability distribution characteristics. Then the discrete \(h\) can be expressed as follows:

\[
h_i = h_{min} + (i - 1) \cdot \Delta h \quad i = 1, 2, \ldots, n \tag{6}
\]

\[
\Delta h = (h_{max} - h_{min})/(n - 1) \tag{7}
\]

where \(h_{min}\) and \(h_{max}\) are the minimum and maximum observation, \(n\) is the number of discretization, and \(\Delta h\) is the discretization spacing.

3. Calculate the density function of copula \(c_\theta(u, v)\) and the density function value \(f_H(h)\), according to the forecast values \((s_k)\) and the discrete value \(h\).
4. Calculate the conditional density function value \(f(h|s)\) by Equation (5).

The median value of the random variable \(H\) is used as a point estimate result. The median value \(h_m\) is the 50% quantile of the random variable \(H\), having the following form:

\[
F_{h_m}(h_m) = 0.5 \tag{8}
\]

For a given confidence level \(\alpha = 0.1\), let \(h_l\) and \(h_u\) be the 5% and 95% quantiles of the random variable \(h\), respectively. Then the 90% confidence interval \([h_l, h_u]\) of \(h\) is used to reflect the forecast uncertainty.
Updating models

AR updating model

The most widely used and the simplest updating model is the AR model, whose procedure consists of three steps. First, the forecast errors series and the corresponding distance average series can be calculated by

\[ e_t = h_t - s_t \]  \hspace{1cm} (9)
\[ e_t = \bar{e} - e_t \]  \hspace{1cm} (10)

where \( e_t \) is the forecast error at the current time \( t \), \( h_t \) and \( s_t \) denote the observed and forecasted flow discharge at the current time \( t \), respectively, \( \bar{e} \) is the mean value of forecast error series, and \( e_t \) is the distance average series of \( e_t \), where the mean value of \( e_t \) equals zero.

Then, the estimates of \( e_t \) for one-step ahead can be expressed as follows:

\[ \hat{e}_{t+1} = a_1 \cdot e_t + a_2 \cdot e_{t-1} + \cdots + a_p \cdot e_{t-p+1} \]  \hspace{1cm} (11)

where \( a_1, a_2, \cdots, a_p \) are the coefficients of the AR model, which can be estimated by the ordinary least square method; \( \hat{e}_{t+1} \) is the estimate of \( e_{t+1} \), and \( p \) is the order of the AR model, which can be determined by autocorrelation and partial autocorrelation and Bayesian information criterion (BIC).

Finally, the updated forecasts can be obtained by adding up the forecast error estimates and the original forecast value, having the following form:

\[ \hat{s}_{t+1} = s_{t+1} + \hat{e}_{t+1} + \bar{e} \]  \hspace{1cm} (12)

where \( \hat{s}_{t+1} \) is the updated forecast and \( s_{t+1} \) is the output of the deterministic hydrological model.

Equation (11) only describes the one-step-ahead updating procedure, while the lead-2 and lead-3 updating procedures can be expressed as follows:

\[ \hat{e}_{t+2} = a_1 \cdot \hat{e}_{t+1} + a_2 \cdot e_t + \cdots + a_p \cdot e_{t-p+2} \]  \hspace{1cm} (13)
\[ \hat{e}_{t+3} = a_1 \cdot \hat{e}_{t+2} + a_2 \cdot \hat{e}_{t+1} + a_3 \cdot e_t + \cdots + a_p \cdot e_{t-p+3} \]  \hspace{1cm} (14)

where \( \hat{e}_{t+1} \) and \( \hat{e}_{t+2} \) are calculated by Equations (11) and (13), respectively.

ARXM model

In the ARXM updating procedure, the recent values of the observed discharge together with the non-updated outputs of the hydrological model are used as inputs of the ARXM model to produce the updated discharge forecast. The ARXM updating procedure for one-step lead-time can be expressed as follows:

\[ \hat{s}_{t+1} = \sum_{k=1}^{p} a_k h_{t-k+1} + \sum_{k=0}^{q} b_k s_{t-k+1} \]  \hspace{1cm} (15)

where \( \hat{s}_{t+1} \) denotes the one-step-ahead updated flow discharge, \( h_{t-k+1}(k = 1, 2, \ldots, p) \) denotes the antecedent observed flow discharge, \( s_{t-k+1}(k = 0, 1, \ldots, q) \) denotes the current and recent output of the hydrological model, and \( p \) and \( q \) are the orders of auto-regressive and the exogenous input parts of the ARXM, respectively. \( a_k, b_k \) are the coefficients of the ARXM model, which are estimated by the ordinary least square method.

For a lead-time \( m > 1(m = 2, 3) \), the updated flow discharge forecasts can be given by

\[ \hat{s}_{t+m|t} = \sum_{k=1}^{m-1} a_k \hat{s}_{t+m-k|t} + \sum_{k=0}^{p} a_k h_{t+m-k} + \sum_{k=0}^{m-1} b_k s_{t+m-k} \]  \hspace{1cm} (16)

The ARXM updating model differs from the AR updating model. In the ARXM updating procedure, the model acquires the updated forecasts directly, whereas during the AR updating process, the forecast errors at the current time are firstly estimated, and then added to the original model output to obtain the updated forecasts.

Adaptive neural fuzzy inference system

According to the types of fuzzy reasoning and fuzzy if-then rules, the ANFIS can be categorized into three types (Jang 1993). In this study, Takagi and Sugen’s fuzzy if-then rules are used, according to which the output of the whole
fuzzy inference system is a weighted average of each rule’s output, the weight values of each rule output are determined by the membership function (a mathematical tool used to represent the ‘degree of truth’ of the element belonging to a fuzzy set).

As shown in Figure 1, the simplest example is given to explain the structure of ANFIS. We assume that the ANFIS has two inputs \( x_1 \) and \( x_2 \) and the fuzzy system has only two if-then rules of Takagi and Sugeno’s type. Then, its typical fuzzy inference rule can be described as follows:

Rule 1: If \( x_1 \) is \( A_1 \) and \( x_2 \) is \( B_1 \), Then \( f_1 = p_1 x_1 + q_1 x_2 + r_1 \) \hspace{1cm} (17)

Rule 2: If \( x_1 \) is \( A_2 \) and \( x_2 \) is \( B_2 \), Then \( f_2 = p_2 x_1 + q_2 x_2 + r_2 \) \hspace{1cm} (18)

where \( A_1 \) and \( A_2 \) are the linguistic label (small or large) of the inputs \( x_1 \) and \( B_1 \) and \( B_2 \) are the linguistic label (small or large) of the inputs \( x_2 \).

As shown in Figure 1, the ANFIS network structure consists of five layers, the functions of each layer node are explained as follows:

Layer 1: Each node in this layer is an adaptive node, which is denoted by a square node. The node function describes the membership degree to which the given input \( x_1 \) belongs to the corresponding fuzzy set \( A_i \), having the following form:

\[ O_1^i = \mu_{A_i}(x_1) \] \hspace{1cm} (19)

where \( \mu_{A_i}(x) \) is the membership function of the fuzzy set \( A_i \), the Gaussian function is chosen in this study.

Layer 2: Each node in this layer is a fixed node, denoted by a circle node. The output of the node is the multiple of all the incoming signals.

\[ w_i = \mu_{A_i}(x_1) \times \mu_{B_i}(x_2) \] \hspace{1cm} (20)

Layer 3: The output of the \( i \)th node is the ratio of the \( i \)th rule confidence \( w_i \) to the sum of all the fuzzy if-then rules confidence, having the following form:

\[ \bar{w}_i = \frac{w_i}{\sum_i w_i}, \quad i = 1, 2 \] \hspace{1cm} (21)

Layer 4: Each node in this layer is also an adaptive node. The output of each node in this layer can be expressed as follows:

\[ O_4^i = \bar{w}_i f_i = \bar{w}_i (p_1 x_1 + q_2 x_2 + r_i) \] \hspace{1cm} (22)

Layer 5: The node in this layer is a fixed node. It calculates the sum of all incoming signals as the total final output:

\[ O_5 = \bar{f} = \sum_i \bar{w}_i f_i = \frac{\sum_i \bar{w}_i f_i}{\sum_i \bar{w}_i} \] \hspace{1cm} (23)

The input and output of the ANFIS updating model in each lead time are the same as the ARXM updating model. For the lead times of 1d, 2d, and 3d, the relationship between the input and output of the ANFIS model can be expressed as follows, respectively:

\[ \hat{s}_{t+1|t} = F_{ANFIS}(h_t, h_{t-1}, \ldots, h_{t-p+1}, s_{t+1|t}, s_t, \ldots, s_{t-q+1}) \] \hspace{1cm} (24)

\[ \hat{s}_{t+2|t} = F_{ANFIS}(\hat{s}_{t+1|t}, h_t, h_{t-1}, \ldots, h_{t-p+2}, s_{t+2|t}, s_{t+1|t}, s_t, \ldots, s_{t-q+2}) \] \hspace{1cm} (25)

\[ \hat{s}_{t+3|t} = F_{ANFIS}(\hat{s}_{t+2|t}, \hat{s}_{t+1|t}, h_t, h_{t-1}, \ldots, h_{t-p+3}, s_{t+3|t}, s_{t+2|t}, s_{t+1|t}, s_t, \ldots, s_{t-q+3}) \] \hspace{1cm} (26)

where \( F_{ANFIS} \) denotes a non-linear function.

**Evaluation criteria**

**Deterministic forecast**

1. Nash-Sutcliffe efficiency (NSE)

\[ NSE = \left( 1 - \frac{\sum_{t=1}^{n} (h_t - s_t)^2}{\sum_{t=1}^{n} (h_t - \bar{h})^2} \right) \] \hspace{1cm} (27)
where \( h_t \) and \( s_t \) denote the observed and forecasted flow discharges at time \( t \), respectively; and \( \bar{h} \) is the average value of observed flows. The closer the value of NSE is to 1, the better the model performs. NSE indicators are sensitive to the evaluation of high flow, while being insensitive to the additive and proportional differences between forecasts and observations.

2. Relative error (RE)

\[
RE = \frac{\sum_{t=1}^{n}s_t - \sum_{t=1}^{n}h_t}{\sum_{t=1}^{n}h_t} \times 100\% \quad (28)
\]

where \( RE \) represents the water balance error, which is an evaluation of the total water volume. The value closer to zero indicates a better forecast. The forecasting errors of flood volume are also very important for large reservoirs.

3. RMSE

\[
RMSE = \sqrt{\frac{\sum_{t=1}^{n}(h_t - s_t)^2}{n}} \quad (29)
\]

where \( RMSE \) represents the standard deviation of the differences between the observed values and simulated values, which is sensitive to outliers. The value closer to zero indicates a better forecast. In the evaluation of medium and high flows, the RMSE indicator can better reflect the pros and cons of the model.

4. Mean absolute error (MAE)

\[
MAE = \frac{1}{n} \sum_{t=1}^{n}|h_t - s_t| \quad (30)
\]

where \( MAE \) is the average absolute error between the observed and forecasted value, it can reflect the forecast error of each point more intuitively. Small \( MAE \) values are preferred. In the evaluation of low and medium flows, the \( MAE \) indicator can better reflect the pros and cons of the model. The units of \( RMSE \) and \( MAE \) are \( m^3/s \).

**Probabilistic forecast**

In this section, continuous ranked probability score (CRPS), quantile-quantile (Q-Q) plot, and other indicators are used to evaluate the quality of probabilistic forecast.

1. CRPS

CRPS is the most widely used criterion to give a comprehensive evaluation of the quality of the probabilistic forecast, which can address the reliability and sharpness of predictive distribution simultaneously (Gneiting & Raftery 2007). The CRPS is defined as follows:

\[
CRPS = \frac{1}{n} \sum_{t=1}^{n} \int_{-\infty}^{\infty} (F_t(r) - H_t(r - h_t))^2 \, dr \quad (31)
\]

where \( F_t(r) \) is the cumulative probability function of the forecast at time \( t \); \( h_t \) is the observed flow at time \( t \); and \( H_t(r - h_t) \) denotes the Heaviside step function, which is calculated by

\[
H_t(r - h_t) = \begin{cases} 
1, & r \geq h_t \\
0, & r < h_t 
\end{cases} \quad (32)
\]

For the deterministic forecasts, the CRPS reduces to generalize the absolute error, which is equivalent to the MAE (Zhao et al. 2015; Liu et al. 2018). Thus, it can also be used to directly compare the accuracy of the deterministic forecast and probabilistic forecast. Smaller CRPS values indicate better performance.

2. Q-Q plot

The reliability of the probabilistic forecast can be evaluated by the Q-Q plot (Laio & Tamea 2007). The closer the Q-Q plot to the 1:1 line indicates a more reliable probabilistic forecast. In order to evaluate the distance between the Q-Q plot and 1:1 line, the \( \alpha \) index indicator presented by Renard et al. (2010) is used and calculated as follows:

\[
\alpha_{index} = 1 - \frac{2}{n} \sum_{t=1}^{n} \left| P_t(h_t) - U_t(h_t) \right| \quad (33)
\]

where \( P_t(h_t) \) is the value that the predictive CDF derives at the observation and \( U_t(h_t) \) is the non-exceedance uniform probability of the observation. The value of the \( \alpha \) index ranges from 0 to 1, large \( \alpha \) index values mean it has greater reliability.

3. Assess the 90% confidence interval

The 90% confidence interval is used to reflect the uncertainties of forecast results. To assess the quality of the 90% confidence interval, four indices were evaluated: containing ratio (CR), average band-width (B), average relative
band-width ($RB$), and percentage of observations bracketed by the unit confidence interval ($PUCI$). These indices are used and calculated as follows:

$$CR = \frac{n_c}{n}$$  \hspace{1cm} (34)

where $n_c$ is the number of measured data contained in the prediction interval. It describes the ratio of the number of observed points within the confidence interval to the total number of the observed points. The larger the $CR$ value indicates the higher percentage of the observed flows within the confidence interval

$$B = \frac{1}{n} \sum_{t=1}^{n} (h_{tu} - h_{tl})$$  \hspace{1cm} (35)

where $h_{tu}$ denotes the upper bound of the 90% confidence interval corresponding to the 95% quantile and $h_{tl}$ denotes the lower bound of the 90% confidence interval corresponding to the 5% quantile

$$RB = \frac{1}{n} \sum_{t=1}^{n} \left( \frac{h_{tu} - h_{tl}}{h_t} \right)$$  \hspace{1cm} (36)

where $h_t$ is the observed flow at time $t$. The average band-width and the average relative band-width are also commonly used to assess the quality of the confidence interval. For a given confidence level $\alpha$, it is best that the values of $B$ and $RB$ should be as small as possible.

According to the above, we can know that the smaller $RB$ value indicates that the average relative band-width of the confidence interval is narrower; the larger the $CR$ value means higher percentage of the observations bracketed by the confidence interval. However, Xiong et al. (2009) indicated that as the $CR$ increased, the $RB$ value of the confidence interval also increased. The two indices are often contradictory when assessing the quality of the 90% confidence interval. Therefore, a comprehensive index of the percentage of observations bracketed by the unit confidence interval ($PUCI$) proposed by Li et al. (2011) was used to evaluate the quality of the uncertainty interval estimate. It can be expressed as follows:

$$PUCI = \frac{CR}{RB}$$  \hspace{1cm} (37)

The larger the $PUCI$ value, the better the performance of the 90% confidence interval is.

### CASE STUDIES

#### Study area and data

The Yangtze River is the largest river in China and the third longest river in the world. The Three Gorges Reservoir (TGR), located in the upper reaches of the Yangtze River, was selected as a case study. It is the largest and the most important water conservancy project in China. The river channel above the dam site of the TGR is 4,500 km long with a contributing drainage area of $1 \times 10^6$ km$^2$. The TGR plays an important role in flood control in the middle and lower sections of the Yangtze River, with the combined benefits of power generation, navigation improvement, etc.

The TGR intervening basin, shown in Figure 2, has a catchment area of 55,907 km$^2$, accounting for 5.6% of the TGR basin area. The mean annual rainfall and evaporation in the basin are 1,100 mm and 900 mm, respectively. The mean annual inflow is 14,300 m$^3$/s. The observed and forecasted daily inflows of Cuntan and Wulong hydrological stations, and the daily rainfall data in the intervening basin from 2012 to 2015 are collected for the inflow forecasting of the TGR. The data series from 2012 to 2014 are used for calibration, and the data of 2015 are used for validation. Numerical precipitation forecasting is necessary to prolong the forecast horizon. The numeric weather prediction daily precipitation predictions from the European Center for Medium-range Weather Forecast (ECMWF) in lead times

![Figure 2](https://iwaponline.com/hr/article-pdf/50/6/1751/636106/nh0501751.pdf)
of 1d, 2d, and 3d are used as a hydrological model input in this study. The statistical characteristics of calibration and validation data are also shown in Figure 3. The maximum and minimum flow during the validation period are included in the calibration period, and the samples in the validation period have similar statistical characteristics during the calibration period, indicating that the calibration and validation period are reasonably divided and can be used for model calibration.

**Hydrological forecasting model**

The inflow of TGR consists of three components, the main streamflow measured by the Cuntan gauge station, the tributary inflow from the Wu River measured by the Wulong gauge station, and the rainfall runoff of the TGR intervening basin, as shown in Figure 2. The flowchart of deterministic and probabilistic flow forecasting for the TGR is shown in Figure 4, in which the deterministic hydrological forecasting of TGR is implemented by a mixed model. The rainfall–runoff relationships of the intervening basin are modeled by the Xinanjiang model (Zhao 1980). The Muskingum method is applied for flood routing (Cunge 1969). Then, the inflow forecast of the TGR can be calculated by

\[
Q_{\text{cal}}(t) = Q_{\text{interval}}(t) + C_1 \cdot Q_{\text{ct}}(t - 1) + C_2 \cdot Q_{\text{wl}}(t - 1)
\]  

(38)

where \(Q_{\text{cal}}(t)\), \(Q_{\text{ct}}(t)\), and \(Q_{\text{wl}}(t)\) denote the inflow of TGR, the flow discharges of Cuntan station and Wulong station at time \(t\), respectively.

The Xinanjiang model proposed by Zhao (1980) is suitable for humid and semi-humid regions and has been widely used for flood simulation and forecasting in China. The model is mainly characterized by the concept of runoff formation as natural storage in humid regions, which implies that runoff is produced after the soil water content reaches field capacity. The storage capacity curve is used to solve the problem of the uneven space distribution of the runoff formation throughout the basin. There are 15 parameters, and each parameter has a defined physical meaning. A more detailed description of the model is available in Zhao (1980).

In the model simulation calibration and verification periods, the observed flood series of the Cuntan and Wulong hydrological stations as well as the observed rainfall of the interval basin area are used as model inputs. The output of the model is the inflow of the TGR. The parameters of the Xinanjiang model and Muskingum method are optimized by the genetic algorithm (Goldberg 1989). The simulation results for the NSE and RE are 99.05% and −0.57% in the calibration period, 98.30% and −1.57% in the validation period, respectively. The simulation results indicate that the established model can be used as a forecasting scheme in practice.

**RESULTS AND DISCUSSION**

**The order of the three updating models**

The order of the three updating models can be determined by using the autocorrelation function (ACF), partial ACF (PACF) and the BIC. In the ARXM and ANFIS updating procedure, the recent values of the observed discharge together with the non-updated outputs of the hydrological model are used as inputs, whereas during the AR updating process, the distance average series of the forecast error...
series is used as its input, the order of models is determined separately. The BIC results of the AR model under different model orders are plotted in Figure 5 (red line). The ACF and PACF of the error time series are shown in Figure 6. In Figure 5, the BIC has the minimum value at the order $p = 5$. In Figure 6, the ACF of the error time series tends to be zero with the increase in the lags. However, the partial autocorrelation coefficient is truncated, which occurs roughly at the order $p = 5$. Therefore, the order of the AR model is determined as 5.

For the ARXM and ANFIS updating models, the order $p$ and $q$ need to be determined. The ACF and PACF of the observed TGR inflow series are calculated and shown in Figure 7. In Figure 7, the PACF is truncated at the order $p = 3$. The BIC results of different orders $q$ at order $p = 3$ are shown in Figure 5 (blue line), on which the BIC has the minimum value at the order $q = 3$. Thus, the order $p = 3$ and $q = 3$ are used to update the forecasts by the ARXM and ANFIS models.

The marginal distributions

First, the marginal distributions of the observed flow, raw forecasts, and updated forecasts need to be determined. Seven commonly used distributions, including Gamma, Log-normal, Generalized Extreme Value (GEV), Pearson type III (P-III), Gumbel, Weibull, and Log-logistic, are chosen as candidate distributions to fit the marginal distributions of all variables. The parameters of each distribution are
estimated by the maximum likelihood estimation method. The modified Shapiro-Wilk test developed by Ashkar & Aucoin (2012) is applied to evaluate the goodness of fit of the probability distributions. The test results of each candidate distribution are listed in Table 2. The best marginal distribution is the one that has the maximum test values.

Table 2 | Test results for seven selected marginal distributions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Gamma</th>
<th>Lognormal</th>
<th>GEV</th>
<th>P-III</th>
<th>Gumbel</th>
<th>Weibull</th>
<th>Log-logistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>0.9150</td>
<td>0.9278</td>
<td>0.9272</td>
<td>0.9412</td>
<td>0.9007</td>
<td>0.9049</td>
<td>0.9164</td>
</tr>
<tr>
<td>$S_1$</td>
<td>0.9071</td>
<td>0.9246</td>
<td>0.9340</td>
<td>0.9451</td>
<td>0.8918</td>
<td>0.8961</td>
<td>0.9142</td>
</tr>
<tr>
<td>$S_2$</td>
<td>0.9029</td>
<td>0.9183</td>
<td>0.9282</td>
<td>0.9555</td>
<td>0.8882</td>
<td>0.8928</td>
<td>0.9079</td>
</tr>
<tr>
<td>$S_3$</td>
<td>0.9060</td>
<td>0.9183</td>
<td>0.9260</td>
<td>0.9520</td>
<td>0.8927</td>
<td>0.8968</td>
<td>0.9070</td>
</tr>
<tr>
<td>$S_{AR1}$</td>
<td>0.9125</td>
<td>0.9278</td>
<td>0.9304</td>
<td>0.9442</td>
<td>0.8979</td>
<td>0.9016</td>
<td>0.9169</td>
</tr>
<tr>
<td>$S_{AR2}$</td>
<td>0.9082</td>
<td>0.9217</td>
<td>0.9247</td>
<td>0.9564</td>
<td>0.8959</td>
<td>0.8972</td>
<td>0.9111</td>
</tr>
<tr>
<td>$S_{AR3}$</td>
<td>0.9106</td>
<td>0.9210</td>
<td>0.9222</td>
<td>0.9536</td>
<td>0.9016</td>
<td>0.8998</td>
<td>0.9098</td>
</tr>
<tr>
<td>$S_{ARXM1}$</td>
<td>0.9129</td>
<td>0.9274</td>
<td>0.9295</td>
<td>0.9428</td>
<td>0.8982</td>
<td>0.9023</td>
<td>0.9164</td>
</tr>
<tr>
<td>$S_{ARXM2}$</td>
<td>0.9101</td>
<td>0.9223</td>
<td>0.9229</td>
<td>0.9548</td>
<td>0.8962</td>
<td>0.9005</td>
<td>0.9111</td>
</tr>
<tr>
<td>$S_{ARXM3}$</td>
<td>0.9162</td>
<td>0.9249</td>
<td>0.9224</td>
<td>0.9545</td>
<td>0.9039</td>
<td>0.9076</td>
<td>0.9127</td>
</tr>
<tr>
<td>$S_{ANFIS1}$</td>
<td>0.9151</td>
<td>0.9273</td>
<td>0.9259</td>
<td>0.9595</td>
<td>0.9011</td>
<td>0.9053</td>
<td>0.9158</td>
</tr>
<tr>
<td>$S_{ANFIS2}$</td>
<td>0.9166</td>
<td>0.9278</td>
<td>0.9257</td>
<td>0.9583</td>
<td>0.9028</td>
<td>0.9071</td>
<td>0.9159</td>
</tr>
<tr>
<td>$S_{ANFIS3}$</td>
<td>0.9178</td>
<td>0.9265</td>
<td>0.9237</td>
<td>0.9556</td>
<td>0.9042</td>
<td>0.9100</td>
<td>0.9142</td>
</tr>
</tbody>
</table>

Figure 6 | The ACF and PACF of errors time series.

Figure 7 | The ACF and PACF of the observed TGR inflow series.
As shown in Table 2, the P-III distribution performs the best fitting with the maximum test values for all variables. The upper tail of the observed flow series is relatively large, which is consistent with the characteristics of the P-III distribution and has been widely used by many authors (Zhong et al. 2017; Yin et al. 2018a, 2018b). The estimated parameters of P-III distributions for all variables are listed in Table 3.

### Table 3 | Estimated parameters of P-III distributions for all variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>α</th>
<th>β</th>
<th>α₀</th>
<th>Variables</th>
<th>α</th>
<th>β</th>
<th>α₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>1.09</td>
<td>0.000096</td>
<td>2,919.72</td>
<td>S_ARXM1</td>
<td>1.03</td>
<td>0.000093</td>
<td>3,059.47</td>
</tr>
<tr>
<td>S₁</td>
<td>0.94</td>
<td>0.000087</td>
<td>3,340.38</td>
<td>S_ARXM2</td>
<td>1.07</td>
<td>0.000096</td>
<td>2,885.52</td>
</tr>
<tr>
<td>S₂</td>
<td>0.97</td>
<td>0.000091</td>
<td>3,150.95</td>
<td>S_ARXM3</td>
<td>1.25</td>
<td>0.000108</td>
<td>2,679.98</td>
</tr>
<tr>
<td>S₃</td>
<td>1.09</td>
<td>0.000104</td>
<td>2,915.85</td>
<td>S_ANFIS1</td>
<td>1.11</td>
<td>0.000098</td>
<td>2,857.52</td>
</tr>
<tr>
<td>S_AR1</td>
<td>1.01</td>
<td>0.000091</td>
<td>3,151.49</td>
<td>S_ANFIS2</td>
<td>1.17</td>
<td>0.000102</td>
<td>2,702.51</td>
</tr>
<tr>
<td>S_AR2</td>
<td>1.05</td>
<td>0.000096</td>
<td>3,247.06</td>
<td>S_ANFIS3</td>
<td>1.25</td>
<td>0.000107</td>
<td>2,495.05</td>
</tr>
<tr>
<td>S_AR3</td>
<td>1.21</td>
<td>0.000111</td>
<td>3,351.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As shown in Table 2, the P-III distribution performs the best fitting with the maximum test values for all variables. The upper tail of the observed flow series is relatively large, which is consistent with the characteristics of the P-III distribution and has been widely used by many authors (Zhong et al. 2017; Yin et al. 2018a, 2018b). The estimated parameters of P-III distributions for all variables are listed in Table 3.

### Construction of the joint distribution based on copula

Three commonly used Archimedean copula functions, including Gumbel, Clayton, and Frank, are selected to construct the joint distribution of the forecasts and observations. The Kendall correlation coefficient $\tau$ between observed and forecasted series is calculated, and then the parameters of three copula functions are computed, according to the relationship between $\tau$ and $\theta$. The $RMSE$ values for evaluating the goodness of fit between theoretical and empirical joint distributions are calculated for each copula function. The results listed in Table 4 show that the Gumbel copula has the best fit with the smallest $RMSE$ values for lead times of 1d, 2d, and 3d. Thus, the Gumbel copula is applied for constructing the joint distribution between observations and forecasts. It can also be seen in Table 4 that the Kendall correlation coefficient decreases as the lead time

### Table 4 | Estimated parameters of copula function and corresponding $RMSE$ values

<table>
<thead>
<tr>
<th>Variables</th>
<th>Copula</th>
<th>$\tau$</th>
<th>$\theta$</th>
<th>$RMSE$</th>
</tr>
</thead>
<tbody>
<tr>
<td>H,S₁</td>
<td>Gumbel-Hougaard</td>
<td>0.940</td>
<td>16.67</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td>Clayton</td>
<td>31.34</td>
<td>0.0120</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>64.99</td>
<td>0.0119</td>
<td></td>
</tr>
<tr>
<td>H,S₂</td>
<td>Gumbel-Hougaard</td>
<td>0.904</td>
<td>10.37</td>
<td>0.0029</td>
</tr>
<tr>
<td></td>
<td>Clayton</td>
<td>18.75</td>
<td>0.0065</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>39.78</td>
<td>0.0076</td>
<td></td>
</tr>
<tr>
<td>H,S₃</td>
<td>Gumbel-Hougaard</td>
<td>0.877</td>
<td>8.10</td>
<td>0.0055</td>
</tr>
<tr>
<td></td>
<td>Clayton</td>
<td>14.20</td>
<td>0.0115</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>30.66</td>
<td>0.0126</td>
<td></td>
</tr>
<tr>
<td>H,S_AR₁</td>
<td>Gumbel-Hougaard</td>
<td>0.954</td>
<td>21.90</td>
<td>0.0030</td>
</tr>
<tr>
<td></td>
<td>Clayton</td>
<td>41.80</td>
<td>0.0098</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>85.92</td>
<td>0.0089</td>
<td></td>
</tr>
<tr>
<td>H,S_AR₂</td>
<td>Gumbel-Hougaard</td>
<td>0.908</td>
<td>10.91</td>
<td>0.0035</td>
</tr>
<tr>
<td></td>
<td>Clayton</td>
<td>19.83</td>
<td>0.0122</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>41.94</td>
<td>0.0115</td>
<td></td>
</tr>
<tr>
<td>H,S_AR₃</td>
<td>Gumbel-Hougaard</td>
<td>0.883</td>
<td>8.52</td>
<td>0.0069</td>
</tr>
<tr>
<td></td>
<td>Clayton</td>
<td>15.05</td>
<td>0.0111</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>32.36</td>
<td>0.0135</td>
<td></td>
</tr>
<tr>
<td>H,S_ARXM₁</td>
<td>Gumbel-Hougaard</td>
<td>0.955</td>
<td>22.36</td>
<td>0.0042</td>
</tr>
<tr>
<td></td>
<td>Clayton</td>
<td>42.71</td>
<td>0.0103</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>87.75</td>
<td>0.0091</td>
<td></td>
</tr>
<tr>
<td>H,S_ARXM₂</td>
<td>Gumbel-Hougaard</td>
<td>0.905</td>
<td>10.52</td>
<td>0.0039</td>
</tr>
<tr>
<td></td>
<td>Clayton</td>
<td>19.04</td>
<td>0.0105</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>40.36</td>
<td>0.0075</td>
<td></td>
</tr>
<tr>
<td>H,S_ARXM₃</td>
<td>Gumbel-Hougaard</td>
<td>0.879</td>
<td>8.28</td>
<td>0.0035</td>
</tr>
<tr>
<td></td>
<td>Clayton</td>
<td>14.56</td>
<td>0.0061</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>31.38</td>
<td>0.0061</td>
<td></td>
</tr>
<tr>
<td>H,S_ANFIS₁</td>
<td>Gumbel-Hougaard</td>
<td>0.957</td>
<td>23.33</td>
<td>0.0081</td>
</tr>
<tr>
<td></td>
<td>Clayton</td>
<td>44.65</td>
<td>0.0164</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>91.62</td>
<td>0.0124</td>
<td></td>
</tr>
<tr>
<td>H,S_ANFIS₂</td>
<td>Gumbel-Hougaard</td>
<td>0.919</td>
<td>12.38</td>
<td>0.0026</td>
</tr>
<tr>
<td></td>
<td>Clayton</td>
<td>22.76</td>
<td>0.0079</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>47.82</td>
<td>0.0097</td>
<td></td>
</tr>
<tr>
<td>H,S_ANFIS₃</td>
<td>Gumbel-Hougaard</td>
<td>0.887</td>
<td>8.87</td>
<td>0.0035</td>
</tr>
<tr>
<td></td>
<td>Clayton</td>
<td>15.75</td>
<td>0.0060</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>33.77</td>
<td>0.0085</td>
<td></td>
</tr>
</tbody>
</table>
Figure 8 | Comparison of probability median forecast and raw forecast for lead times of 1d, 2d, and 3d.
increases as expectedly, since forecast quality decreases as the lead time increases.

**Comparison of deterministic and probabilistic forecasting results**

**Deterministic forecasting results**

According to the distribution range of the observation, let \( h_{\text{min}} = 3,000 \, \text{m}^3/\text{s} \) and \( h_{\text{max}} = 90,000 \, \text{m}^3/\text{s} \) (the cumulative probability of which is above 0.999999). It is theoretically better to select a small discrete spacing. However, smaller discrete spacing introduces more computational complexity into the calculation. Therefore, the discrete spacing \( \Delta h = 2 \, \text{m}^3/\text{s} \) is selected in this study. Then the discrete \( H \) of each time step can be calculated by Equation (6), where the number of discretization is \( n = 43,501 \). Based on the principle of mathematical statistics, the probability median forecast is calculated by Equation (8) and used as the point estimate results of probabilistic forecast. Four evaluation indexes, namely \( NSE, RE, RMSE, \) and \( MAE \), are used to evaluate the accuracy of the median forecast. The comparison with the raw forecast is presented in Figure 8, which shows that in terms of all evaluation indicators, the accuracy of probability median forecast is significantly improved compared with the raw forecast. Compared with the raw forecast results, the \( RE \) of the probability median forecast are reduced by 0.12%, 3.38%, and 5.01%, and the \( NSE \) are improved by 0.1%, 0.6%, and 1.6% for lead times of 1d, 2d, and 3d in the validation period, respectively.

The deterministic evaluation indexes of raw forecast and three updating models for lead times of 1d, 2d, and 3d in calibration and verification periods are listed in Table 5. The comparison between the forecast results of the three updating models and the original model is also shown in Figure 9. It is obvious that the performances of three updating models are superior to that of the raw forecasts in terms of almost all evaluation indicators. The ANFIS updating model performs the best for almost all lead times, while the ARXM updating model performs slightly better than the AR model for 1d, 2d, and 3d lead times. The ANFIS model performs the best because it not only can effectively identify the non-linear relationship in hydrological process but also effectively classify the inputs into high and low levels, which can consider the relationship between errors and flood magnitude. In addition, according to the results of \( NSE \) and \( RE \) indexes, the improvement degree becomes more obvious with the increase in lead time. However, for the \( RMSE \) and \( MAE \) indexes, the improvement degree decreases with lead time increases. For the ANFIS updating model, the \( MAE \) indicator in 1d, 2d and 3d lead times has improved compared with the original forecast, while for the AR and ARXM models, the improvement in the \( MAE \) index is no longer achieved for long lead times (3d) compared to the original forecast. This indicates that the ability of the updating model for correcting medium and low flow is gradually lost as the lead time increases.

### Table 5 | Deterministic evaluation indexes of raw forecast and three updating models for lead times of 1d, 2d, and 3d in calibration and verification periods

<table>
<thead>
<tr>
<th>Lead time</th>
<th>Model</th>
<th>NSE (%)</th>
<th>RE (%)</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1d</td>
<td>RAW</td>
<td>98.61</td>
<td>98.35</td>
<td>–0.41</td>
<td>–0.79</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>98.99</td>
<td>98.92</td>
<td>0.06</td>
<td>–0.05</td>
</tr>
<tr>
<td></td>
<td>ARXM</td>
<td>99.00</td>
<td>99.03</td>
<td>–0.57</td>
<td>–0.71</td>
</tr>
<tr>
<td></td>
<td>ANFIS</td>
<td>99.10</td>
<td>99.06</td>
<td>0.00</td>
<td>0.23</td>
</tr>
<tr>
<td>2d</td>
<td>RAW</td>
<td>95.59</td>
<td>95.57</td>
<td>–2.72</td>
<td>–4.23</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>96.16</td>
<td>96.33</td>
<td>0.07</td>
<td>–0.73</td>
</tr>
<tr>
<td></td>
<td>ARXM</td>
<td>96.19</td>
<td>96.60</td>
<td>–0.85</td>
<td>–2.36</td>
</tr>
<tr>
<td></td>
<td>ANFIS</td>
<td>96.43</td>
<td>96.79</td>
<td>–0.03</td>
<td>–0.96</td>
</tr>
<tr>
<td>3d</td>
<td>RAW</td>
<td>90.43</td>
<td>91.94</td>
<td>–5.81</td>
<td>–5.97</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>91.55</td>
<td>93.30</td>
<td>0.11</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>ARXM</td>
<td>92.09</td>
<td>93.75</td>
<td>0.29</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>ANFIS</td>
<td>92.60</td>
<td>93.60</td>
<td>–0.02</td>
<td>–0.51</td>
</tr>
</tbody>
</table>
Deterministic evaluation indexes of three updating models for lead times of 1d, 2d, and 3d. Note: AR1, ARXM1, ANFIS1 models with 1d lead time, AR2, ARXM2, ANFIS2 models with 2d lead time, AR3, ARXM3, ANFIS3 models with 3d lead time.

Figure 9 | Deterministic evaluation indexes of three updating models for lead times of 1d, 2d, and 3d. Note: AR1, ARXM1, ANFIS1 models with 1d lead time, AR2, ARXM2, ANFIS2 models with 2d lead time, AR3, ARXM3, ANFIS3 models with 3d lead time.
increases. Comparing the results of indicators in Figures 8 and 9, it can be found that the traditional updating models, including AR, ARXM, and ANFIS, are better than the copula-based probability model.

**Probabilistic forecasting results**

The conditional probability density and distribution curve of the observation $H$, given $S_k = s_k$, can be obtained based on the conditional probability for given forecasts. Take 1 July 2015 ($S_1 = 38,462 \text{m}^3/\text{s}$, $S_2 = 33,469 \text{m}^3/\text{s}$, and $S_3 = 26,386 \text{m}^3/\text{s}$) as an example, the PDF and CDF of observation $H$ are presented in Figure 10. As can be seen in Figure 10, the conditional density function becomes fatter with the increase in lead time, indicating that the flood forecasting uncertainty increases as the lead time increases.

In order to evaluate the reliability of the probabilistic forecast, the predictive Q-Q plots for three updating models with different lead times are shown in Figure 11. As the Kolmogorov–Smirnov test is highly sensitive when applied to the large data sets (Serinaldi 2009), none of the Q-Q plots lie within the significance bands. However, overall the Q-Q plots for the updating models are closer to the 1:1 line than the original model for all lead times. It is noted that we cannot obtain enough information from the visual inspection of the Q-Q plot to evaluate their reliability. Therefore, the CRPS and $\alpha$ index are used to quantitatively measure the reliability of the probabilistic forecasts. The values of CRPS and $\alpha$ index for the probability raw forecast and updating models for lead times of 1d, 2d, and 3d are listed in Table 6. The values of $\alpha$ index for three updating models are closer to 1 than those of the original model, which implies that the probabilistic forecasts are more reliable after forecast correction. It is also shown that the CRPS values increase as the lead time increases, which indicates that the probabilistic forecast is less sharp and reliable at longer lead times. Compared with the corresponding raw deterministic forecast, the CRPS values of the probabilistic forecast in the validation period are reduced by 25.76%, 26.86%, and 26.37% for lead times of 1d, 2d, and 3d, respectively. It indicates that the conditional probability model based on copula is a useful tool for quantifying and assessing forecast uncertainty. Table 6 also shows that as the accuracy of deterministic forecast increases after error correction, the corresponding probabilistic forecast becomes more reliable. This is because the conditional probability model achieves probabilistic forecast on the basis of given deterministic forecast and the reliability of probabilistic forecast depends largely on the accuracy of deterministic forecast. In terms of the comprehensive index CRPS, the ANFIS model performs the best in almost all lead times.

The evaluation index results of 90% confidence interval at different lead times are illustrated in Figure 12. The average and relative band-widths gradually increase as lead times increase, indicating that the inflow forecast uncertainty increases. The $B$ and $RB$ values of three updating models are smaller than that of the original forecast. The ANFIS updating model has the lowest $B$ and $RB$ values among the three updating models. In terms of CR values, the coverage rate of different updating models in all lead times is close to or exceeds the given confidence level 90%, indicating that the 90% forecast interval is reasonable.
and reliable. The \textit{PUCI} indicator is used to compare the performance of different updating models. As shown in Figure 12, the \textit{PUCI} values of probabilistic forecast decrease as lead times increase, which means that the forecast uncertainty of reservoir inflow increases with lead times. On the other hand, the \textit{PUCI} value of updated forecast increases significantly compared with that of the raw deterministic forecast, which implies that the probabilistic forecasts are more reliable after error correction. In terms of \textit{PUCI} value, the ANFIS model performs the best with the largest \textit{PUCI} value among the three updating models. There is no significant difference between the ARXM and AR models.

The 90\% confidence intervals of the raw forecast model and ANFIS updating model in the flood season in the
validation period are plotted in Figure 13. It is observed that the 90% confidence interval covers most of the observed flows, indicating that the attained probabilistic forecast interval is reliable. The band-width of 90% confidence intervals attained by the updating models is narrower than the original. Furthermore, the 90% confidence intervals become wider at a longer lead time, indicating that the forecast uncertainty becomes larger as lead time increases.

**CONCLUSIONS**

Quantifying the inherent uncertainty in flood forecasting is of great importance for flood control decision-making and water resources management. Properly communicating uncertainties in flood forecasting provides the basis for reservoir operation, and thus prevents casualties and economic losses. However, traditional flood forecasting does not consider predictive uncertainty. This study develops a conditional probability model based on copula to quantify and estimate predictive uncertainty. Three updating models, namely the AR model, ARXM model, and ANFIS model, are applied to correct raw forecasts. The TGR in China is selected as a case study. The main conclusions of this study can be summarized as follows:

1. Three updating models can significantly improve the raw forecast results in terms of the deterministic evaluation criteria. The degree of improvement in forecast accuracy by the ANFIS updating model outperforms the ARXM and AR models.

2. The conditional probability model based on copula is a useful tool for post-processing deterministic forecast to quantify forecast uncertainty, providing both a deterministic forecast value along with a reasonable and reliable uncertainty interval. The updated probability forecasting results show higher accuracy and reliability than that of raw forecasts in terms of CRPS, α_index, and PUCI, indicating that the updating models are a useful tool to improve the reliability of probabilistic forecast. The ANFIS model is superior to the ARXM and AR models in terms of comprehensive indicator CRPS and PUCI in almost all lead times.

3. The conditional probability forecast model based on copula can quantitatively assess forecast uncertainty, and the three updating models can improve the accuracy and reliability of the probability forecast. In terms of deterministic flood forecasting, the conditional probabilistic forecasting model based on copula can improve the forecast accuracy; however, the degree of improvement by the three updating models is more significant.

In summary, the developed conditional probability model combined with the updating models quantify the hydrological forecasting uncertainty accurately and reliably. Like most existing studies, this paper focuses on
evaluating the pros and cons of probabilistic forecasting schemes; however, how to effectively take advantage of the risk information provided by probabilistic forecasts in the reservoir operation decision-making process needs further study.

**ACKNOWLEDGEMENTS**

This study is financially supported by the National Key R&D Project (2016YFC0402206) and National Natural Science Foundation (51539009, 51879192) of China. We are very thankful to the reviewers for their valuable suggestions. The opinions expressed in this manuscript are those of the authors and do not necessarily reflect the views of the funding agencies.
grateful to the editor and anonymous reviewers for their valuable comments and constructive suggestions, and Dr Bryan O’Reilly for proofreading that helped us to greatly improve the manuscript.

REFERENCES


First received 15 June 2019; accepted in revised form 3 September 2019. Available online 27 September 2019.