

Article

An Ensemble Flow Forecast Method Based on Autoregressive Model and Hydrological Uncertainty Processor

Xin Yang, Jianzhong Zhou *, Wei Fang and Yurong Wang

School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; d201880948@hust.edu.cn (X.Y.); d202081169@hust.edu.cn (W.F.); m201873786@hust.edu.cn (Y.W.)

* Correspondence: jz.zhou@hust.edu.cn

Received: 7 October 2020; Accepted: 4 November 2020; Published: 9 November 2020



Abstract: In the process of hydrological forecasting, there are uncertainties in data input, model parameters, and model structure, which cause a deterministic forecasting to fail to provide useful risk information to decision-makers. Therefore, the study of ensemble forecasting and the analysis of hydrological uncertainty are of great significance to guide the actual operation of reservoirs in the flood season. This study proposed a Bayesian ensemble forecast method, comprising of a Gaussian mixture model (GMM), a hydrological uncertainty processor (HUP), and an Autoregressive (AR) model. First, the GMM is selected as the marginal distribution function to estimate the uncertainty of observed and modelled data. Next, the AR model is used to correct the forecast rainfall data. Then, a modified HUP is used to deal with the uncertainty of hydrological model structure and rainfall input data. In the end, the ensemble flow forecast results are composed of the expected values of the posterior distribution obtained by HUP under different rainfall conditions. Taking the Three Gorges Reservoir (TGR) as a case study, the ensemble flow prediction in the forecast period is calculated by using the above method. Results show that the method proposed in this paper can improve the accuracy of runoff forecasts and reduce the uncertainty of the hydrological forecast.

Keywords: ensemble flow forecast; autoregressive model; hydrological uncertainty processor; Three Gorges Reservoir

1. Introduction

Flooding is the most common natural hazard and third most damaging globally after storms and earthquakes [1]. Flood forecasting can provide key information for disaster warning and flood control, which plays an important role in reducing the damage caused by flooding [2,3]. The common flood forecasting methods generally use a deterministic hydrological model to predict the future flood process [4–7]. However, the traditional deterministic hydrological forecast cannot provide the hydrological forecast value in the forecast period to meet the needs of the corresponding river basin management department [8].

Ensemble flow forecasting can provide more hydrological forecasting information for policymakers in order to better ensure the safety of life and property in downstream areas and improve the social and economic benefits [9,10]. Ensemble weather forecasting (EWF) plays an important role in ensemble flow forecasting [11]. First, EWF contains more information, which can provide more comprehensive meteorological data for a hydrological forecast in a future period [12]. Second, research shows that combining EWF with a hydrological forecast model can improve the reliability and accuracy of forecast results [13]. However, due to the diversity of atmospheric conditions and topography, the simplification of physical and thermodynamic processes, the uncertainties of parameterization of

models, and the limited spatial resolution, the standard EWF offers systematic deviations compared with the observations [14]. Thus, EWF should be corrected first before being applied to the hydrological forecasting process. For example, Hamill et al. [15] used a logistic regression method to process the set rainfall forecast and obtained the conditional distribution function of rainfall under given conditions. Sloughter et al. [16] used a logistic regression function and the Gamma distribution to describe the distribution of rainfall events of zero value and non-zero value, respectively, then built a mixed distribution function to describe the distribution of rainfall events, and improved the Bayes Model Averaging (BMA) method to make it applicable to the post-processing of aggregate rainfall forecasts. Wilks [17] improved the traditional logistic regression approach by introducing quartiles into the function as independent variables, thus, providing a continuous probability distribution function to describe the distribution of rainfall. However, the above methods need to make assumptions about the distribution of rainfall in EWF, and the calculation process is tedious. To overcome the above problems, the Autoregressive (AR) model, which is independent of the rainfall distribution, was used to post-process the precipitation value of EWF.

In recent years, great progress has been made in the research and theory of hydrological ensemble forecasting. Mylne [18] conducted a multi-mode ensemble prediction test, proving that the information given by multi-mode ensemble prediction is more accurate than that provided by a single mode prediction in either the probabilistic sense (such as probability density distribution) or the deterministic sense (such as average ensemble prediction). Brown et al. [19] used the National Center for Environmental Forecasting (NCEP) ensemble prediction model to drive the distributed hydrological model to obtain the hydrological ensemble prediction results, and the results showed that the ensemble prediction could improve the effective flood warning period. Choubin et al. [20] proposed a new flood sensitivity set prediction method based on multiple discriminant analysis (MDA) and classification and regression trees (CART), combined with support vector machine (SVM). However, most of the above methods only consider the uncertainty of hydrological models or the uncertainty of rainfall input data, and fail to reflect the comprehensive impact of uncertainties from different sources in the prediction process and results. Therefore, this paper proposes a method that combines an AR model with a Bayesian forecast system (BFS). The proposed model, comprehensively considering the uncertainties of rainfall input, model structure, and parameters, is designed to yield good ensemble flow forecast results in this paper.

Krzysztofowicz constructed the theoretical system of BFS for the first time in 1999 [21], and pointed out that the source of uncertainty in short-term runoff forecasting is mainly the future time series of rainfall as input to the hydrological model. BFS generates probabilistic forecasts of hydrological variables through deterministic hydrological models. As a result, the basic principle is to divide the sources of hydrological uncertainty into two categories: one is the input uncertainty with rainfall as the core and the other is the structural uncertainty of the hydrological model. Quantitative research, which is then combined into the final uncertainty probability forecast results, verifies that BFS can be combined with any deterministic hydrological model, and is independent of whether the hydrological model structure is linear or not [22,23].

Marshall [24] realized the probabilistic forecast by combining the Markov Monte Carlo (MCMC) with Bayesian method, and compared it with conceptual rainfall-runoff model to illustrate the advantages of probabilistic prediction. Subsequently, Marshall [25] used the Adaptive Metropolis Algorithm to calculate the posterior density of Bayesian probability prediction and verified the feasibility of Bayesian probability prediction as a hydrological prediction model. Henry [26,27] proposed a random ensemble Bayesian probabilistic forecast system (REBFS) based on the BFS. The REBFS model can output multiple set elements, effectively reducing the scale of the set weather prediction system, and making it more feasible to apply the BFS to large-scale regions. However, the process of parameter estimation of the above methods is complicated, which makes their application inconvenient. Feng et al. [28] used the Gaussian mixture model (GMM) to fit the marginal distribution and applied it to the hydrological uncertainty processor (HUP), and concluded that GMM had good applicability in HUP.

However, it did not consider the effect of precipitation uncertainty. In order to solve the above two problems, a method combining precipitation-dependent HUP (PD-HUP) and GMM is proposed in this paper. The GMM function is used as the marginal distribution function in BFS, and the uncertainty of precipitation input data is also considered.

In summary, the main objective of this paper is to propose a Bayesian ensemble forecast method considering uncertainty of hydrological models and precipitation. First, the Autoregressive (AR) model is used to correct the rainfall data of Ensemble Weather Forecasts (EWF). Then, a precipitation-dependent HUP (PD-HUP) based on GMM is proposed to deal with the uncertainty of both the input data and the hydrological model. Finally, the ensemble flow forecast results are generated from the results determined by the PD-HUP method.

2. Methods and Data

2.1. Postprocessing of Ensemble Weather Forecasts

The Autoregressive (AR) model was used to postprocess the precipitation series of the EWF. The AR model only reflects the influence and effect of related factors on the prediction target through its own modelling of historical observations of time series variables. It is not restricted by the assumptions that the model variables are independent of each other. The model constructed can eliminate the difficulties caused by independent variable selection and multicollinearity in general regression prediction methods. The mathematical expression of the AR model is shown below.

$$\begin{cases} e_t^j = \alpha_1^j e_{t-1}^j + \alpha_2^j e_{t-2}^j + \dots + \alpha_p^j e_{t-p}^j + \xi_j \\ w_c^j = w_{fore,t}^j - e_t^j \end{cases}, \quad (1)$$

where $w_{fore,t}^j$ is the uncorrected rainfall forecast value of j th ensemble member, $e_t^j (t = 1, 2, \dots, N)$ is the prediction error sequence of the j th ensemble member at time t and represents the difference between the EWF result and the observed precipitation. N is the length of the precipitation sequence of each ensemble member. $e_{t-1}^j, e_{t-2}^j, \dots, e_{t-p}^j$ is the prediction residual series for the p periods of the j th ensemble member before time t , $\alpha_1^j, \alpha_2^j, \dots, \alpha_p^j$ are the regression coefficients of the j th ensemble member, p is the order, which can be determined by the Akaike information criterion (AIC) criterion, ξ_j represents a white noise sequence whose mean value is 0, and variance is determined from the parameterization and w_c^j being the corrected prediction of the j th ensemble member. The calculation steps of the AR model are as follows.

First, determine the order of p according to the AIC criterion. Letting p take different values, calculate the AIC value of each ensemble member, and select the p value with the minimum AIC value as the regression order of the AR model. AIC is expressed as follows:

$$AIC = 2k + n \ln(L), \quad (2)$$

where k is the number of parameters, referring to the order of AR in the model in this paper, and L is the variance of the residuals.

$$Y = [e_{p+1}, e_{p+2}, \dots, e_N]^T, \quad (3)$$

and

$$\xi = [\xi_{t-1}, \xi_{t-2}, \dots, \xi_{t-n}]^T, \quad (4)$$

$$A = [\alpha_1, \alpha_2, \dots, \alpha_p]^T, \quad (5)$$

$$X = \begin{bmatrix} e_p & e_{p-1} & \cdots & e_1 \\ e_{p+1} & e_p & \cdots & e_2 \\ \vdots & \vdots & \vdots & \vdots \\ e_{N-1} & e_{N-2} & \cdots & e_{N-p} \end{bmatrix} \tag{6}$$

Then the p -order AR model can be expressed as:

$$Y = XA + \xi, \tag{7}$$

The estimation of model parameters can be obtained from the least square principle.

$$A = (X^T X)^{-1} X^T Y, \tag{8}$$

In the end, the AR models with the parameters determined above were used to correct the precipitation forecast errors.

2.2. Establishment of the Ensemble Flow Forecast Model

The Bayesian probability forecast system (BFS) was first proposed by Roman Krzysztofowicz in 1999. As an important part of BFS, the hydrologic uncertainty processor (HUP) presents the hydrological prediction values in a probabilistic form through Bayesian analysis and calculation of the results of the deterministic hydrological prediction model.

2.2.1. The Hydrological Uncertainty Processor Methodology

Let n ($n = 0, 1, 2, \dots, N$) denote forecast time. Suppose that h_n ($n = 0, 1, 2, \dots, N$) represents the observed inflow to the Three Gorges Reservoir (TGR) and s_n ($n = 0, 1, 2, \dots, N$) represents the reservoir inflow simulated by the hydrological model. H_n and S_n are, respectively, the implementation values of the random variables. The prior probability density function (PDF) under different prediction periods can be expressed as the following equation.

$$g_n(h_n|h_0) = \frac{\gamma(h_n)}{(1 - c^{2n})^{1/2} q(Q^{-1}(\Gamma(h_n)))} \times q\left(\frac{Q^{-1}(\Gamma(h_n)) - c^n Q^{-1}(\Gamma(h_0))}{(1 - c^{2n})^{1/2}}\right), \tag{9}$$

where $\gamma(\cdot)$ is the marginal PDF of observed inflow, $q(\cdot)$ is the standard normal density function, $Q^{-1}(\cdot)$ is the standard normal quantile function, $\Gamma(\cdot)$ is the marginal cumulative probability function (CDF) of H_n , and c is the Pearson correlation coefficient obtained from the normal quantile transform.

The PDF under different prediction periods can be expressed as follows.

$$\varphi(h_n|s_n, h_0) = \frac{\gamma(h_n)}{T_n q(Q^{-1}(\Gamma(h_n)))} \times \left(\frac{Q^{-1}(\Gamma(h_n)) - A_n Q^{-1}(\Lambda_n(s_n)) - D_n Q^{-1}(\Gamma(h_0)) - B_n}{T_n} \right) \tag{10}$$

where $\Lambda_n(\cdot)$ is the marginal distribution function of the reservoir inflow simulated by the hydrological model, and A_n, B_n, D_n , and T_n are the parameters calculated in the normal quantile conversion space and the expression are as follows.

$$A_n = \frac{a_n \tau_n^2}{a_n^2 \tau_n^2 + \sigma_n^2}, \tag{11}$$

$$B_n = \frac{-a_n b_n \tau_n^2}{a_n^2 \tau_n^2 + \sigma_n^2} \tag{12}$$

$$D_n = \frac{-a_n d_n \tau_n^2}{a_n^2 \tau_n^2 + \sigma_n^2} \tag{13}$$

$$T_n^2 = \frac{\tau_n^2 \sigma_n^2}{a_n^2 \tau_n^2 + \sigma_n^2} \tag{14}$$

In the normal space, the stochastic dependence between W_n and W_{n-1} is governed by a normal-linear equation.

$$W_n = c_n W_{n-1} + \Xi_n, \tag{15}$$

where c_n is Pearson’s correlation coefficient, Ξ is a stochastically independent and normally distributed variable of W_n with a mean of zero and a variance of $\tau_n^2 = 1 - c_n^2$.

The stochastic dependence between X_n and W_n, W_{n-1}, W_0 is governed by a normal-linear equation.

$$X_n = a_n W_n + d_n W_0 + b_n + \Theta_n, \tag{16}$$

where $a_n, b_n,$ and d_n are the regression coefficients, and Θ_n is a stochastically independent and normally distributed variable of (W_n, W_0) with a mean of zero and a variance of σ_n^2 .

2.2.2. Precipitation-Dependent Hydrological Uncertainty Processor Based on Gaussian Mixture Model

In the calculation process of the HUP, the selection of marginal distribution function directly affects the accuracy of calculation results and the complexity of parameter estimation. Feng et al. [28] proposed that the Gaussian mixture model has better performance in fitting the marginal probability distribution of the observed flow and simulated flow, which can improve the performance of the prediction method based on HUP.

The marginal density function and distribution function in the HUP are estimated as follows.

$$\gamma(\cdot) \approx gmm(\cdot|\theta_\Gamma), \tag{17}$$

$$\Gamma(\cdot) \approx GMM(\cdot|\theta_\Gamma) \tag{18}$$

$$\Lambda_n(\cdot) \approx GMM(\cdot|\theta_{\Lambda_n}) \tag{19}$$

where $gmm(\cdot|\theta_\Gamma)$ and $GMM(\cdot)$ approximately represent the PDF and CDF of H_n , where θ_Γ and θ_{Λ_n} are the distribution parameters to be estimated. More details about the GMM can be found in Feng et al. [28]. Then, on the conditional of observed inflow at time t_0 , the prior distribution of H_n can be expressed as:

$$g_n(h_n|h_0) = \frac{gmm(h_n|\theta_\Gamma)}{(1-c^{2n})^{1/2} q(Q^{-1}(GMM(h_n|\theta_\Gamma)))} \times q\left(\frac{Q^{-1}(GMM(h_n|\theta_\Gamma)) - c^n Q^{-1}(GMM(h_0|\theta_\Gamma))}{(1-c^{2n})^{1/2}}\right) \tag{20}$$

Similarly, according to Equation (10), under the condition that the observed inflow at time t_0 is h_0 and the simulated flow is s_n , the posterior probability density function of H_n is the following.

$$\varphi_n(h_n|s_n, h_0) = \frac{gmm(h_n|\theta_\Gamma)}{T_n q(Q^{-1}(GMM(h_n|\theta_\Gamma)))} \times q\left(\frac{Q^{-1}(GMM(h_n|\theta_\Gamma)) - A_n Q^{-1}(GMM(s_n|\theta_{\Lambda_n})) - D_n Q^{-1}(GMM(h_0|\theta_\Gamma)) - B_n}{T_n}\right) \tag{21}$$

In order to accurately describe the uncertainty of the rainfall input data, the parameters of the posterior probability need to be discussed, according to the precipitation during the lead-time period. Krzysztofowicz [23] discussed the HUP model in the case of rainfall occurrence and non-occurrence, and proposed the precipitation-dependent HUP (PD-HUP) model considering the uncertainty of rainfall. However, there are only a few periods during the flood season when no rainfall is generated, so it is difficult to accurately estimate the marginal distribution function of observed and predicted flow

in the case of no rainfall. Therefore, this paper establishes a PD-HUP model based on the precipitation in the lead time period, divided into two types of effective rainfall and negligible rainfall, and then weighs the two to form a PD-HUP based on Gaussian mixture model (PD-HUP-GMM) model that takes rainfall uncertainty into account.

Let Variable V represent different rainfall conditions and W represent precipitation. When $v = 1$, it means effective rainfall ($w \geq 1$ mm) and $v = 0$ ($0 \leq w \leq 1$ mm) means negligible rainfall. Let $\gamma_{ov}(h_0) = p(h_0|V = v)$ represent the marginal PDF of h_0 under the condition $V = v$ where $v = P(V = 1)$ represents the probability of effective rainfall in the lead time period. Then, the marginal PDF of H_n is as follows.

$$\gamma_0(h_0) = \gamma_{00}(h_0)(1 - v) + \gamma_{01}(h_0)v, \tag{22}$$

Under the condition that the observed inflow is h_0 , the probability of effective and negligible rainfall is as follows.

$$P(V = 1|H_0 = h_0) = \frac{\gamma_{01}(h_0)v}{\gamma_0(h_0)}, \tag{23}$$

$$P(V = 0|H_0 = h_0) = \frac{\gamma_{00}}{(h_0)(1 - v)\gamma_0(h_0)} \tag{24}$$

Under the condition that the observed inflow is h_0 , the PDF of the S_n is as follows.

$$k_n(s_n|h_0) = k_{n0}(s_n|h_0)\frac{\gamma_{00}(h_0)(1 - v)}{\gamma_0(h_0)} + k_{n1}(s_n|h_0)\frac{\gamma_{01}(h_0)v}{\gamma_0(h_0)}, \tag{25}$$

Under the condition that the observed inflow is h_0 and simulated inflow is s_n , the probability of effective and negligible rainfall is as follows.

$$P(V = 1|S_n = s_n, H_0 = h_0) = \frac{k_{n1}(s_n|h_0)\gamma_{01}(h_0)v}{k_n(s_n|h_0)\gamma_0(h_0)}, \tag{26a}$$

$$P(V = 0|S_n = s_n, H_0 = h_0) = \frac{k_{n0}(s_n|h_0)\gamma_{00}(h_0)(1 - v)}{k_n(s_n|h_0)\gamma_0(h_0)} \tag{26b}$$

Thus, finally, the posterior probability can be expressed as:

$$\begin{aligned} \varphi(h_n | s_n, h_0) &= \varphi_{n0}(h_n | s_n, h_0) \frac{k_{n0}(s_n | h_0)\gamma_{00}(h_0)(1 - v)}{k_n(s_n | h_0)\gamma_0(h_0)} \\ &+ \varphi_{n1}(h_n | s_n, h_0) \frac{k_{n1}(s_n | h_0)\gamma_{01}(h_0)v}{k_n(s_n | h_0)\gamma_0(h_0)} \end{aligned} \tag{27}$$

2.2.3. Generation of Ensemble Flow Forecasts

In order to convert the probability flow forecast result obtained by the PD-HUP-GMM method into the ensemble flow forecast result, the expected value of the probability flow forecast result under each rainfall condition is taken as the member of the flow forecast set to constitute the flow forecast ensemble.

First, the value of flood expectation under a different rainfall forecast is as follows.

$$E(H_{n,t}) = \int_{-\infty}^{+\infty} h_{n,t} \varphi(h_{n,t}|s_{n,t}, h_0) dh_{n,t}, \tag{28}$$

where $H_{n,t}$ represents the simulated value of the n th flood ensemble forecast member at time t , and $E(H_{n,t})$ is the expected value of $H_{n,t}$. Then, the ensemble flow forecast result can be expressed as follows.

$$\begin{bmatrix} E(H_{1,1}) & E(H_{1,2}) & \cdots & E(H_{1,T}) \\ E(H_{2,1}) & E(H_{2,2}) & \cdots & E(H_{2,T}) \\ \vdots & \vdots & \ddots & \vdots \\ E(H_{N,1}) & E(H_{N,2}) & \cdots & E(H_{N,T}) \end{bmatrix}, \tag{29}$$

where T represents the number of periods of simulated runoff sequence, and each row in the matrix represents a member of a flood ensemble forecast. The flow chart describing the Bayesian ensemble forecast method is shown in Figure 1.

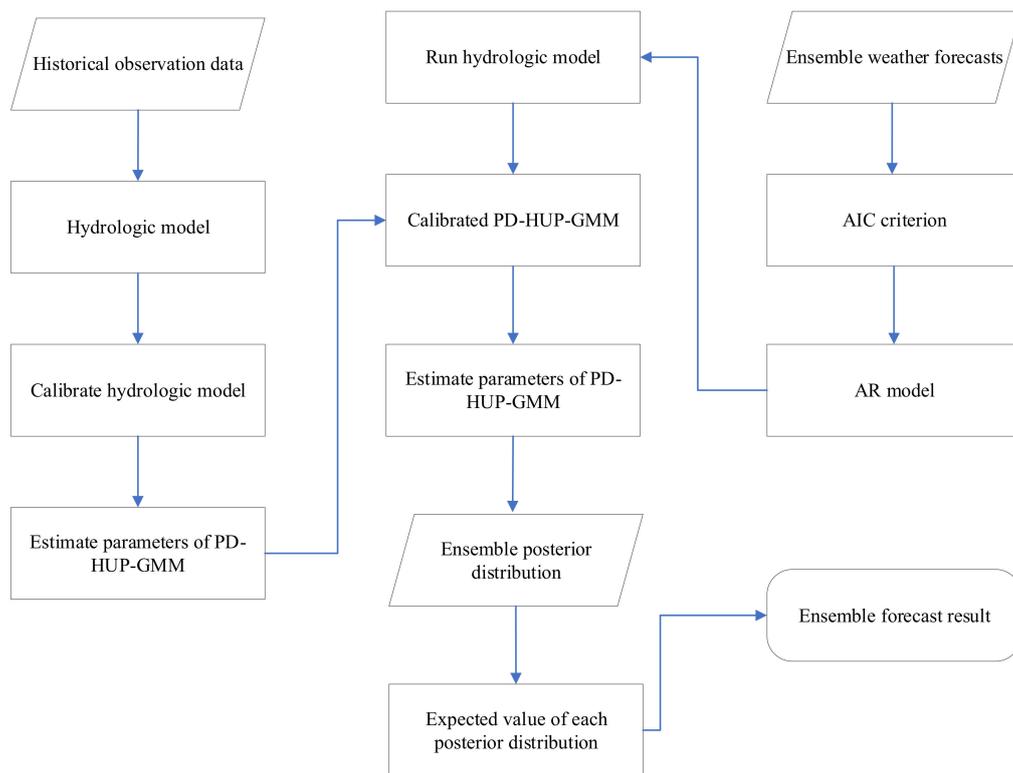


Figure 1. Flow of methods. The PD-HUP-GMM represents the precipitation-dependent hydrological uncertainty processor (PD-HUP) based on a Gaussian mixture model (GMM).

2.3. Data

The Three Gorges Reservoir (TGR) located on the Yangtze River in China was selected as a case study and located at Yichang on the Yangtze river in China, which is the world’s largest hydroelectric project. Its geographical location is shown in Figure 2. The main functions of the Three Gorges Reservoir are flood control, power generation, and navigation, etc. The basin area of the Three Gorges Reservoir is 1×10^6 km², the surface area of the reservoir is about 1080 km², and the average width is about 1100 m [29].

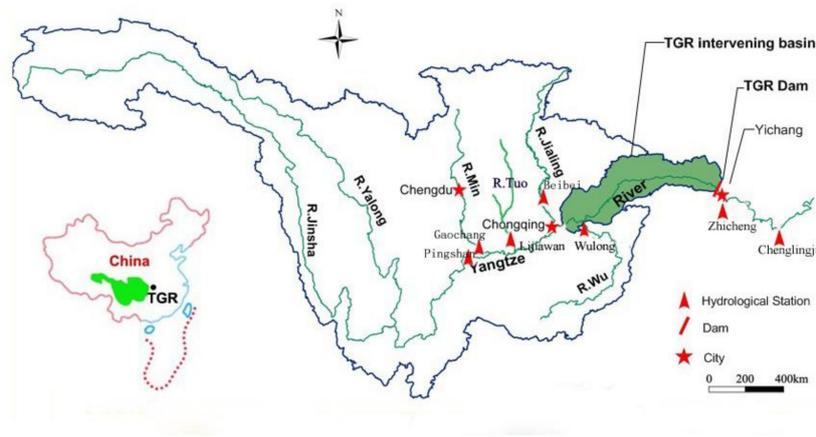


Figure 2. Schematic of the Three Gorges Reservoir (TGR) in China.

The historical daily inflow data of TGR and precipitation from 2006 to 2017 during the flood season were used in this study. This study uses the EWF data from the second edition of the Global Ensemble Forecast System (GEFS) developed by the National Center for Environmental Forecasting (NCEP). This is one of the most widely used numerical forecasting systems in the world. The data set contains 11 set forecast members, and the spatial resolution in latitude and longitude is $1^\circ \times 1^\circ$. Data from January to September in 2017 was used in this study. The Xin'an Jiang model is selected as the deterministic prediction model of the PD-HUP-GMM method with the precipitation and evaporation in the study area used as the input. We decided to restrict the forecast period to one day ahead.

3. Results and Discussion

3.1. Correction of Ensemble Weather Forecasts

First, the value of order p was determined according to the AIC criterion. In this paper, p takes different values (1, 2, ..., 10). When we calculate the AIC value of each ensemble member, the result is shown in Figure 3. The p value ($p = 8$) with the minimum AIC value was selected as the regression order of the AR model.

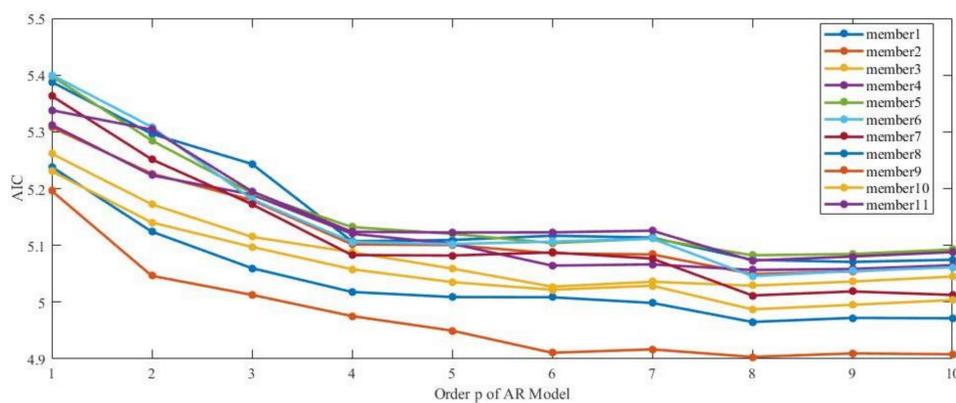


Figure 3. The AIC value of each ensemble member.

Then, the mean absolute error (MAE) and root-mean square error (RMSE) were used to evaluate the performance of the AR model. The MAE and RMSE are defined as follows.

$$MAE = \frac{1}{N} \sum_{t=1}^N \frac{|F_t - O_t|}{O_t}, \tag{30}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (F_t - O_t)^2} \quad (31)$$

where F_t represents the forecast precipitation, and O_t represents the observed precipitation. The MAE and RMSE value of each GEFS member before and after correction are shown in Table 1.

According to Table 1, after AR model correction, the MAE of precipitation data of GEFS members decreased by an average of 18.10 m³/s, while RMSE decreased by an average of 16.05 m³/s. It shows that the AR model can effectively reduce the precipitation deviation of GEFS. However, there are still some bias and uncertainties in the corrected GEFS precipitation data. Therefore, the PD-HUP model was adopted in this paper to further deal with the uncertainty of hydrological forecast caused by rainfall forecast bias.

Table 1. Comparison of MAE and RMSE before and after correction.

GEFS Member	MAE		RMSE	
	Before Correction	After Correction	Before Correction	After Correction
1	30.46	11.25	32.42	15.98
2	29.95	11.49	31.89	16.36
3	29.45	12.09	31.54	16.30
4	28.87	11.49	31.03	15.82
5	28.13	11.75	30.29	15.14
6	28.54	10.44	30.86	14.17
7	29.47	10.73	31.83	14.94
8	29.74	11.61	31.72	15.36
9	28.92	10.55	30.71	14.18
10	28.76	10.83	30.97	14.91
11	29.98	10.91	32.19	15.77

3.2. Estimated Parameters of Ensemble Flow Forecast Model

3.2.1. Generation of Ensemble Flow Forecasts

The observed flow data in the flood season from 2006 to 2016 were used to fit the marginal distribution. In order to obtain the marginal distribution function of the simulated flow, the Xin'an Jiang model was used to calculate the simulated flow value in the same period as the observed flow at first, and the lead time n was set to one day. The number of days with effective rainfall is 1178, and the number of days with negligible rainfall is 164.

Following the ideas of Feng [28], the GMM with three Gaussian components was used to fit the observed inflow and simulated inflow data of TGR. The Maximum Likelihood Estimate (MLE) was used to estimate the parameters of the marginal distributions under heavy effective and negligible rainfall conditions, respectively. The estimated parameter values are given in Tables 2 and 3. The fitting of marginal CDF is shown in Figure 4. As can be seen from Figure 4, GMM fits the observed and simulated flow marginal distribution under effective and negligible rainfall conditions.

Table 2. Estimated parameters of GMM distributions under effective rainfall.

Dataset	Components	Weight	Mean	Variance
Observed value	1	0.49	1.59×10^4	1.16×10^7
	2	0.17	3.61×10^4	9.36×10^7
	3	0.34	2.49×10^4	2.20×10^7
Simulated value	1	0.51	1.75×10^4	1.26×10^7
	2	0.12	3.96×10^4	9.19×10^7
	3	0.36	2.77×10^4	2.18×10^7

Table 3. Estimated parameters of GMM distributions under negligible rainfall.

Dataset	Components	Weight	Mean	Variance
Observed value	1	0.46	1.63×10^4	1.12×10^7
	2	0.16	2.59×10^4	6.03×10^7
	3	0.38	1.06×10^4	3.29×10^7
Simulated value	1	0.57	1.25×10^4	4.55×10^7
	2	0.12	2.82×10^4	4.79×10^7
	3	0.31	1.92×10^4	1.20×10^7

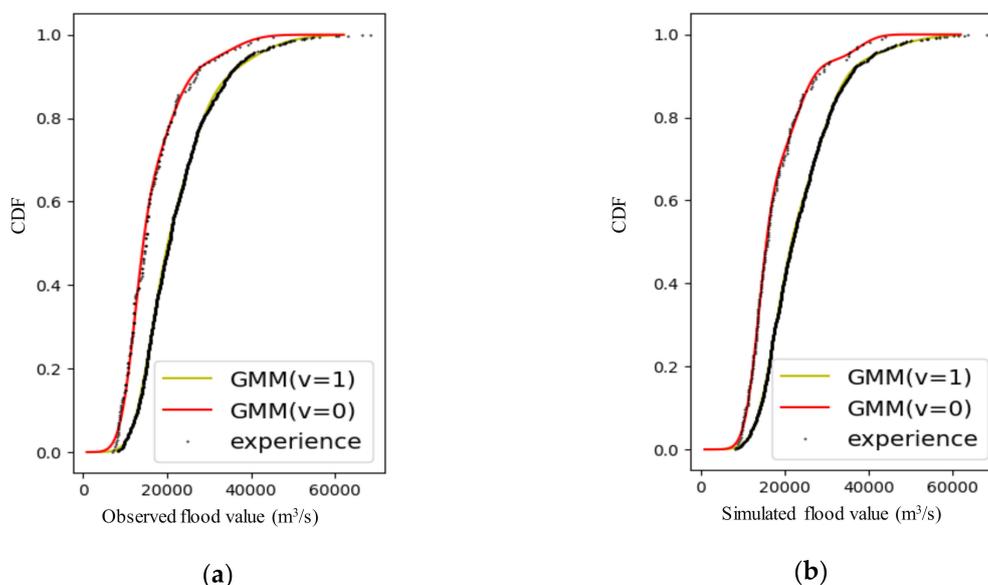


Figure 4. Fitting of the marginal CDF of observed value and simulated flood value. (a) Marginal CDF of observed value. (b) Marginal CDF of simulated value. GMM ($v = 1$) and GMM ($v = 0$), respectively, represent the fitting of the distribution function in the case of effective rainfall and negligible rainfall.

3.2.2. Parameters of PD-HUP

The calculation method of prior density, likelihood function, and posterior is shown in Equations (15)–(18). The parameters of prior density and likelihood function in the transformed space is shown in Table 4 and the parameters of parameters of post distribution in the transformed space is shown in Table 5.

Table 4. Parameters of prior density and likelihood function in the transformed space.

V	c	a_n	d_n	b_n	σ_n	t_n
1	0.931	1.058	−0.150	−0.080	0.420	0.365
0	0.974	0.868	0.715	0.043	0.384	0.225

Table 5. Parameters of parameters of post distribution in the transformed space.

V	A_n	B_n	D_n	T_n
1	0.434	0.065	0.538	0.269
0	0.236	−0.169	0.765	0.200

3.3. Ensemble Forecast Analysis

In order to verify the effectiveness of the method proposed in this paper, precipitation and inflow data of TGR in the flood season of 2017 were taken as model input data, and the above method was

used to calculate the ensemble inflow forecast results of TGR. Three scenarios as shown in Table 6 were set up for comparative analysis. The Nash-Sutcliffe efficiency (NSE) and RMSE were selected to evaluate the accuracy of forecast results under different scenarios and the variance was used to evaluate the uncertainty of the Bayesian ensemble forecast result, as discussed in the following subsections.

Table 6. Brief descriptions of the four scenarios.

Scenarios	Description
Observed data + Xin'an Jiang model	Use observed data only to run Xin'an Jiang model and get deterministic forecast
GEFS data + Xin'an Jiang model	Use GEFS to run the Xin'an Jiang model and get ensemble forecast
Corrected GEFS data + Xin'an Jiang model	Corrected GEFS data to run the Xin'an Jiang model and get ensemble forecast
Bayesian ensemble forecast	Use HUP to postprocess the result of corrected GEFS data with the Xin'an Jiang model and get the ensemble result

3.3.1. Comparison of Forecast Accuracy

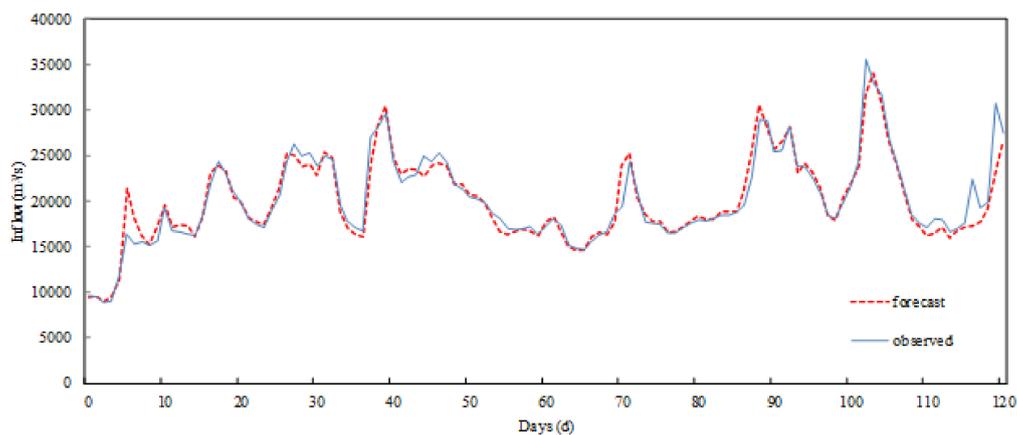
For the deterministic forecast, the NSE and RMSE were calculated by observed inflow and simulated result of the Xin'an Jiang model. For the ensemble forecast evaluate, the mean value of ensemble members and observed were used to calculate NSE and RMSE. The forecast results of different scenarios are shown in Figure 5. The formula for calculating NSE is as follows and the calculated results of indicators are shown in Table 7.

$$NSE = 1 - \frac{\sum_{i=1}^N (Q_{sim} - Q_{obs})^2}{\sum_{i=1}^N (Q_{sim} - \bar{Q}_{obs})^2}, \quad (32)$$

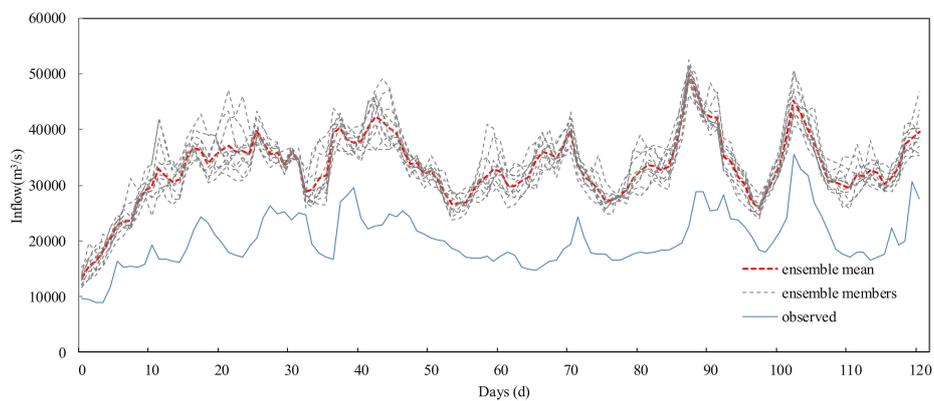
It can be seen from the simulation results that, when the observed historical rainfall data is used as the input of the hydrological model, the NSE of the simulation results can reach 0.92, as shown in Table 7, indicating that the Xin'an Jiang model can be used as a deterministic model for hydrological prediction in the research area. When GEFS data is directly used as input data of the Xin'an Jiang model, the simulation results of the hydrological model are significantly larger than the actual runoff process $NSE = -7.14$. When the GEFS data is corrected by using the AR model and the corrected GEFS data is used as the input of the Xin'anjiang model, the simulation accuracy of the Xin'an Jiang model has been significantly improved. However, by comparing the ensemble average with the observed inflow process, the NSE improves to only 0.56, which is difficult to meet the demand for prediction accuracy in actual production. This is mainly because the uncertainty of rainfall input data affects the accuracy of hydrological model simulation results. Furthermore, by using the PD-HUP-GMM method proposed in this paper, it can be found that the simulation results of the hydrological model have been effectively improved. The NSE of the ensemble average value is up to 0.91 and the RMSE decreases by 52.6% when compared with the scenario of directly adopting the corrected GEFS data, greatly improve the accuracy of the runoff simulation, and enough to meet demand for hydrological forecast accuracy in practical production. It shows that the method proposed in this paper greatly improves the accuracy of inflow forecast of TGR and is sufficient to meet the demand of hydrological prediction accuracy in actual production.

Table 7. NSE and RMSE of the four scenarios.

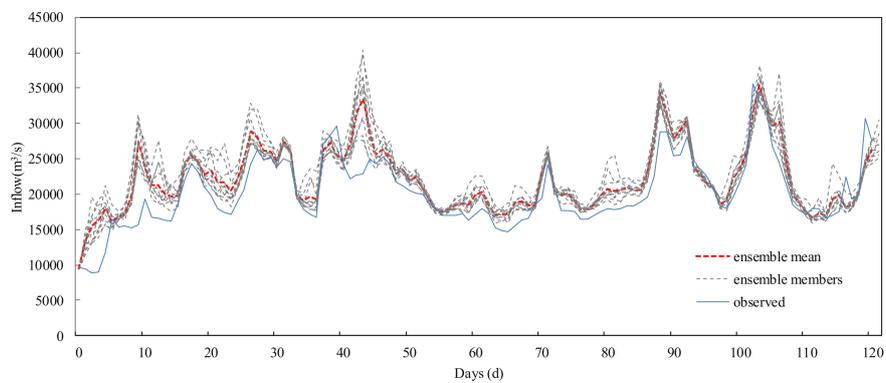
Scenarios	NSE	RMSE
Observed data + Xin'an Jiang model	0.92	1364.02
GEFS data + Xin'an Jiang model	-7.14	13613.12
Corrected GEFS data + Xin'an Jiang model	0.56	3167.20
Bayesian ensemble forecast	0.91	1501.23



(a)



(b)



(c)

Figure 5. Cont.

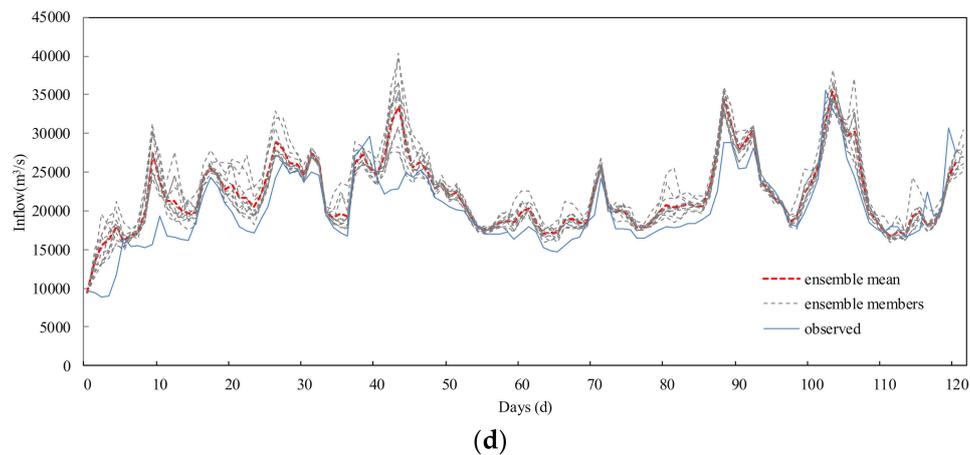


Figure 5. Simulated result of the four scenarios: (a) Observed data + Xin’an Jiang model, (b) GEFS data + Xin’an Jiang model, (c) Corrected GEFS data + Xin’an Jiang model, and (d) Bayesian ensemble forecast.

3.3.2. Uncertainty Evaluation of Ensemble Prediction Results

In order to analyze the uncertainty of hydrological model simulation results under different scenarios, the variances of ensemble members in different time periods and different scenarios are calculated. The formula for variance used is shown below.

$$Var_t = \frac{\sum_{i=1}^n (Q_{i,t} - \bar{Q}_t)^2}{n - 1}, \tag{33}$$

where n is the number of ensemble members, $Q_{i,t}$ represents the prediction result of the i th ensemble member at time t , and \bar{Q}_t is the mean value of the prediction result of all ensemble members at time t . The variance of ensemble members in each scenario changes over time as shown in Figure 6.

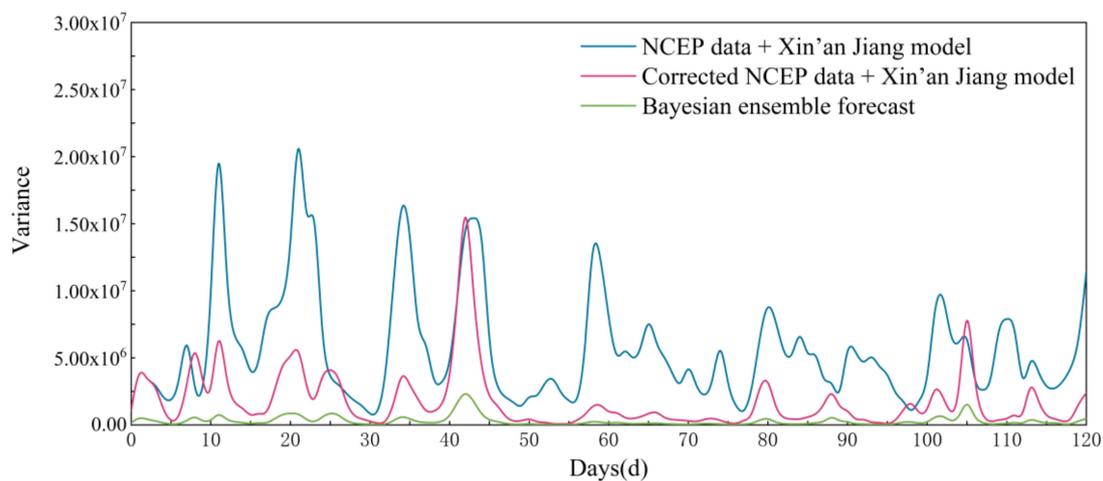


Figure 6. The variance of ensemble flow forecast members at different moments of the 2017 flood season under different scenarios.

As can be seen from Figure 6, when different rainfall input data are used, the uncertainty of hydrological model simulation results is significantly different. When GEFS rainfall data is used to directly drive the hydrological model, the variance of the ensemble member of the hydrological model simulation results can reach 2.49×10^7 , indicating that the simulation results are highly uncertain. However, by using the adjusted GEFS rainfall data to drive the hydrological model, the variance of

the ensemble members of the inflow prediction in different periods will significantly decrease, and the uncertainty of the prediction results will significantly decrease. When the Bayesian ensemble forecast method proposed by this paper is adopted, the uncertainty of hydrological prediction results is further reduced.

4. Conclusions

In this paper, the precipitation data of the Global Ensemble Forecast System (GEFS) was postprocessed by an Autoregressive (AR) model, and a precipitation-dependent hydrological uncertainty processor based on GMM (PD-HUP-GMM) was proposed to generate the ensemble flow forecast. Then, the Nash-Sutcliffe efficiency (NSE), root-mean square error (RMSE), and variances of ensemble members were used to evaluate the accuracy and uncertainty of the ensemble flow forecast. The TGR was selected as a case study. The main conclusions of this study are summarized as follows.

(1) The results of the AR model show that this model can effectively correct GEFS precipitation data, and it is simple and feasible. The use of AR models to correct the GEFS rainfall data also helps to improve the simulation accuracy of the deterministic hydrological model.

(2) The proposed method was compared with three different forecast methods. Results of the case study show that the PD-HUP-GMM method combined with the corrected GEFS data can significantly improve the NSE and RMSE of flow forecasting. From this point of view, the proposed method performs well.

(3) The proposed method can effectively deal with the uncertainty of precipitation input data and the hydrological model and, thus, improve the accuracy of the ensemble flow forecast average value to a useful level.

Future work will involve extending the forecast period to a longer horizon to further evaluate the potential of HUP in operational flow forecasting and improve its practical significance.

Author Contributions: Conceptualization, X.Y. and J.Z. Methodology, X.Y. Data curation, J.Z. Writing—original draft preparation, X.Y. Writing—review and editing, X.Y, W.F., Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation Key Project of China (No. 52039004) and National Natural Science Foundation of China (No. U1865202).

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Chen, L.; Singh, V.P.; Guo, S.; Zhou, J.; Ye, L. Copula entropy coupled with artificial neural network for rainfall-runoff simulation. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 1755–1767. [[CrossRef](#)]
2. Yaseen, Z.M.; Naganna, S.R.; Sa'adi, Z.; Samui, P.; Ghorbani, M.A.; Salih, S.Q.; Shahid, S. Hourly River Flow Forecasting: Application of Emotional Neural Network Versus Multiple Machine Learning Paradigms. *Water Resour. Manag.* **2020**, *34*, 1075–1091. [[CrossRef](#)]
3. Ramaswamy, V.; Saleh, F. Ensemble Based Forecasting and Optimization Framework to Optimize Releases from Water Supply Reservoirs for Flood Control. *Water Resour. Manag.* **2020**, *34*, 989–1004. [[CrossRef](#)]
4. Bashir, A.; Shehzad, M.A.; Hussain, I.; Rehmani, M.I.A.; Bhatti, S.H. Reservoir Inflow Prediction by Ensembling Wavelet and Bootstrap Techniques to Multiple Linear Regression Model. *Water Resour. Manag.* **2019**, *33*, 5121–5136. [[CrossRef](#)]
5. Fu, J.-C.; Huang, H.-Y.; Jang, J.-H.; Huang, P.-H. River Stage Forecasting Using Multiple Additive Regression Trees. *Water Resour. Manag.* **2019**, *33*, 4491–4507. [[CrossRef](#)]
6. Bai, Y.; Bezak, N.; Sapač, K.; Klun, M.; Zhang, J. Short-Term Streamflow Forecasting Using the Feature-Enhanced Regression Model. *Water Resour. Manag.* **2019**, *33*, 4783–4797. [[CrossRef](#)]
7. Zhou, Q.; Chen, L.; Singh, V.P.; Zhou, J.; Chen, X.; Xiong, L. Rainfall-runoff simulation in karst dominated areas based on a coupled conceptual hydrological model. *J. Hydrol.* **2019**, *573*, 524–533. [[CrossRef](#)]

8. Li, W.; Zhou, J.; Sun, H.; Feng, K.; Zhang, H.; Tayyab, M. Impact of Distribution Type in Bayes Probability Flood Forecasting. *Water Resour. Manag.* **2017**, *31*, 961–977. [[CrossRef](#)]
9. Cloke, H.L.; Pappenberger, F. Ensemble flood forecasting: A review. *J. Hydrol.* **2009**, *375*, 613–626. [[CrossRef](#)]
10. Wu, W.; Emerton, R.; Duan, Q.; Wood, A.W.; Wetterhall, F.; Robertson, D.E. Ensemble flood forecasting: Current status and future opportunities. *Wiley Interdiscip. Rev. Water* **2020**, *7*, e1432. [[CrossRef](#)]
11. Li, X.-Q.; Chen, J.; Xu, C.-Y.; Li, L.; Chen, H. Performance of Post-Processed Methods in Hydrological Predictions Evaluated by Deterministic and Probabilistic Criteria. *Water Resour. Manag.* **2019**, *33*, 3289–3302. [[CrossRef](#)]
12. Zhang, J.; Chen, J.; Li, X.; Chen, H.; Xie, P.; Li, W. Combining Postprocessed Ensemble Weather Forecasts and Multiple Hydrological Models for Ensemble Streamflow Predictions. *J. Hydrol. Eng.* **2020**, *25*, 1–17. [[CrossRef](#)]
13. Boucher, M.-A.; Anctil, F.; Perreault, L.; Tremblay, D. A comparison between ensemble and deterministic hydrological forecasts in an operational context. *Adv. Geosci.* **2011**, *29*, 85–94. [[CrossRef](#)]
14. Han, S.; Coulibaly, P. Probabilistic Flood Forecasting Using Hydrologic Uncertainty Processor with Ensemble Weather Forecasts. *J. Hydrometeorol.* **2019**, *20*, 1379–1398. [[CrossRef](#)]
15. Hamill, T.M.; Whitaker, J.S.; Wei, X. Ensemble Reforecasting: Improving Medium-Range Forecast Skill Using Retrospective Forecasts. *Mon. Weather. Rev.* **2004**, *132*, 1434–1447. [[CrossRef](#)]
16. Sloughter, J.M.L.; Raftery, A.E.; Gneiting, T.; Fraley, C. Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging. *Mon. Weather. Rev.* **2007**, *135*, 3209–3220. [[CrossRef](#)]
17. Wilks, D.S. Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteorol. Appl.* **2009**, *16*, 361–368. [[CrossRef](#)]
18. Mylne, K.R. Decision making from probability forecasts using calculations of forecast value. *Meteorol. Appl.* **2001**, *9*, 307–315. [[CrossRef](#)]
19. Brown, J.D.; Seo, D.-J. A Nonparametric Postprocessor for Bias Correction of Hydrometeorological and Hydrologic Ensemble Forecasts. *J. Hydrometeorol.* **2010**, *11*, 642–665. [[CrossRef](#)]
20. Choubin, B.; Moradi, E.; Golshan, M.; Adamowski, J.; Sajedi-Hosseini, F.; Mosavi, A. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Sci. Total. Environ.* **2019**, *651*, 2087–2096. [[CrossRef](#)]
21. Krzysztofowicz, R. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.* **1999**, *35*, 2739–2750. [[CrossRef](#)]
22. Krzysztofowicz, R.; Kelly, K.S. Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.* **2000**, *36*, 3265–3277. [[CrossRef](#)]
23. Krzysztofowicz, R.; Herr, H.D. Hydrologic uncertainty processor for probabilistic river stage forecasting: Precipitation-dependent model. *J. Hydrol.* **2001**, *249*, 46–68. [[CrossRef](#)]
24. Marshall, L.; Nott, D.; Sharma, A. A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. *Water Resour. Res.* **2004**, *40*, 183–188. [[CrossRef](#)]
25. Marshall, L.; Nott, D.; Sharma, A. Hydrological model selection: A Bayesian alternative. *Water Resour. Res.* **2005**, *41*. [[CrossRef](#)]
26. Herr, H.D.; Krzysztofowicz, R. Ensemble Bayesian forecasting system Part I: Theory and algorithms. *J. Hydrol.* **2015**, *524*, 789–802. [[CrossRef](#)]
27. Herr, H.D.; Krzysztofowicz, R. Ensemble Bayesian forecasting system Part II: Experiments and properties. *J. Hydrol.* **2019**, *575*, 1328–1344. [[CrossRef](#)]
28. Feng, K.; Zhou, J.; Liu, Y.; Lu, C.; He, Z. Hydrological Uncertainty Processor (HUP) with Estimation of the Marginal Distribution by a Gaussian Mixture Model. *Water Resour. Manag.* **2019**, *33*, 2975–2990. [[CrossRef](#)]
29. Zhou, Y.; Guo, S. Risk analysis for flood control operation of seasonal flood-limited water level incorporating inflow forecasting error. *Hydrol. Sci. J.* **2014**, *59*, 1006–1019. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).