



1 Quantify and reduce flood forecast uncertainty by the CHUP- 2 BMA method

3 Zhen Cui¹, Shenglian Guo^{1*}, Hua Chen¹, Dedi Liu¹, Yanlai Zhou¹, Chong-Yu Xu²

4 ¹ State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, China

5 ² Department of Geoscience, University of Oslo, Oslo, Norway

6 *Correspondence to:* Shenglian Guo (slguo@whu.edu.cn)

7 **Abstract.** The Bayesian model averaging (BMA), hydrological uncertainty processor (HUP), and HUP-BMA methods have
8 been widely used to quantify flood forecast uncertainty. This study, for the first time, introduced a copula-based HUP in the
9 framework of BMA and proposed the CHUP-BMA method to bypass the need for normal quantile transformation of the
10 HUP-BMA method. The proposed ensemble forecast scheme consists of 8 members (two forecast precipitation inputs, two
11 advanced long short-term memory (LSTM) models, and two objective functions used to calibrate parameters) and is applied
12 to the interval basin between Xiangjiaba and Three Gorges Reservoir (TGR) dam-site. The ensemble forecast performance of
13 the HUP-BMA and CHUP-BMA methods is explored in the 6-168h forecast horizons. The TGR inflow forecasting results
14 show that the two methods can improve the forecast accuracy over the selected member with the best forecast accuracy, and
15 the CHUP-BMA performs much better than the HUP-BMA. Compared with the HUP-BMA method, the forecast interval
16 width with the 90% confidence level and continuous ranked probability score metrics of the CHUP-BMA method are highest
17 reduced by 28.42% and 17.86%, respectively. The probability forecast of the CHUP-BMA method has better reliability and
18 sharpness and is more suitable for flood ensemble forecasts, providing reliable risk information for flood control decision-
19 making.

20 1 Introduction

21 Accurate and reliable flood forecasting is one of the necessary measures to reduce flood disasters and improve water
22 resource utilization (Zhou et al., 2019; Vegad and Mishra, 2022). With the development of hydrological theory and flood
23 forecasting techniques, the flood forecasting accuracy and lead time have been significantly improved in recent years (Xu et
24 al., 2022; Cui et al., 2023). However, neither physically-based and conceptual hydrological models nor data-driven models
25 can guarantee to obtain perfect forecasting in real conditions. Because of the influence of the changing environment and the
26 limitations of human perception of complex hydrological processes, the meteorological forcing and other inputs,
27 hydrological model structure, and parameters, etc., contain significant uncertainties (Cloke et al., 2009), which leads to the
28 simulation and forecast results of the model inevitably containing integrated uncertainties from multiple sources (Liu et al.,
29 2022). Traditional flood forecasting schemes are mostly deterministic forecast results without considering forecast



30 uncertainty (Zhong et al., 2018a; Gelfan et al., 2018), which makes decision-makers unable to grasp useful risk information
31 beyond the forecast value. Excessive superstition on a single forecast value will likely lead to poor decision-making
32 (Krzysztofowicz et al., 1999). Therefore, it is essential to quantify and reduce flood forecast uncertainty in practical
33 applications.

34 Probabilistic flood forecasting is one of the effective methods to quantify integrated forecast uncertainty (Matthews et
35 al., 2022). It not only provides a deterministic forecast value, but also provides forecast uncertainty (or risk) information by
36 means of quantile, confidence interval, or density function (Biondi and Todini, 2018; Ferretti et al., 2020; Zhou et al., 2022),
37 which is more scientifically reasonable and practically useful compared with deterministic forecasts and helps decision-
38 makers consider forecast risk quantitatively (Todini, 2008). Various probabilistic forecasting methods based on statistical
39 post-processing of numerical forecast data have been developed in recent years. Among these methods, probabilistic
40 ensemble forecasting is considered to overcome the limitations of a single model or a simple average with fixed model
41 weights (Han and Coulibaly, 2017) and contains richer forecast information because it can consider the ensemble forecast
42 results of multiple models to quantify and reduce integrated uncertainty that contains uncertainties in the inputs, model
43 structure, and parameters (Li et al., 2017; Saleh et al., 2016). Bayesian model averaging (BMA), proposed by Raftery et al.
44 (2005), uses the Bayesian theory and a total probability formulation to transform ensemble forecasts into probabilistic
45 forecasts and is one of the most representative and reliable methods that has been widely used to supplement uncertainty
46 information beyond point estimates (Shu et al., 2022).

47 The BMA method is initially successfully applied to the ensemble forecast of meteorological elements such as
48 temperature and precipitation (Raftery et al. 2005; Sloughter et al., 2010). After confirming that the BMA method can
49 effectively quantify forecast uncertainty and obtain highly accurate deterministic forecasts, it is widely used in hydrological
50 forecasting to quantify forecast uncertainty from different sources, such as model inputs, structure, and parameters. The
51 standard BMA method assumes that each member's posterior probability distribution approximately obeys a normal
52 distribution (Huang et al., 2019; Guo et al., 2021). However, some variables, such as wind speed, rainfall, runoff, etc.,
53 usually obey skewed distributions and require methods such as Box-Cox to convert non-Gaussian variables to standard
54 normal variables that affect the accuracy of probability distribution estimation (Duan et al., 2007; Liu et al., 2018). Many
55 authors have investigated the applicability of BMA in flood ensemble forecasting and tried to overcome its limitations
56 (Madadgar and Moradkhani, 2014; Darbandsari and Coulibaly, 2020). Sloughter et al. (2010) proposed an improved BMA
57 method by assuming that the posterior probability distribution of each member could obey a specific non-normal distribution
58 (e.g., Gamma distribution) and using the member forecast values to estimate the mean and variance of the distribution.
59 Madadgar and Moradkhani (2014) introduced the Copula function to solve the posterior probability distribution of members
60 in the BMA method and proposed the Copula-based BMA method, which avoids the assumption of the posterior probability
61 distribution and further reduces the application limitation of the BMA method. Meanwhile, the BMA method usually
62 ensembles the forecast results of multiple models to be as close to the actual values as possible. However, too many
63 ensemble members may generate redundant information. Darbandsari and Coulibaly (2020) introduced the Shannon entropy



64 theory to select the forecast members that satisfy the above conditions before applying BMA. Their results showed that the
65 BMA method incorporating entropy could improve the probabilistic forecasting performance for high flows over the
66 standard BMA method. In addition, some studies have developed various methods based on the BMA principle, such as the
67 multi-model ensemble forecasting method based on Vine Copula (Zhang et al., 2022) and the combination of BMA and data
68 assimilation techniques (Parrish et al., 2012). However, most studies ignore an essential issue: the BMA does not consider
69 the constraint of initial conditions (i.e., observed flow at the start of the forecast). When the member forecasts are the same,
70 the ensemble forecast will produce the same conditional probability distribution and lack rationality.

71 The hydrological uncertainty processor (HUP) can obtain the posterior distribution function of the actual value under
72 the condition of the forecast value and the observed flow at the start of the forecast based on Bayesian principles and the
73 assumption of perfect rainfall forecasting (Krzysztofowicz and Kelly, 2000). Darbandsari and Coulibaly (2021) utilized the
74 HUP method to derive the posterior distribution of each member and used the BMA method to weight the posterior
75 distribution of all members to obtain the final posterior distribution, which is called the HUP-BMA method. Their results
76 showed that the HUP-BMA method outperforms the HUP method and improves the BMA method in short-term probabilistic
77 forecasting. In addition, the derivability of the posterior distribution for the ensemble members is theoretically enhanced, the
78 heteroskedasticity of the ensemble members is considered, and the interpretability and logical rationality of the BMA
79 method are improved

80 Although it has been demonstrated that considering initial conditions in the BMA method can improve ensemble
81 forecast performance, there are still issues to be explored. The HUP-BMA method requires a normal quantile conversion
82 method to convert the flow data series to Gaussian space to solve the posterior distribution. The process is not only tedious
83 and complicated, but also prone to bias in the inverse conversion. To this end, Liu et al. (2018) proposed a Copula-based
84 HUP (CHUP) and found that it could bypass the normal transformation process and improve the probabilistic forecasting. It
85 is anticipated that coupling CHUP to the BMA may improve the HUP-BMA accuracy and applicability, which motivates the
86 current study.

87 Therefore, the main aims and research steps of this study are as follows: (1) to propose the novel CHUP-BMA method
88 by coupling CHUP into BMA to realize the consideration of the initial condition of the forecast while bypassing the need for
89 data normal quantile transformation; (2) to construct an ensemble forecast containing 8 members combining two types of
90 forecast precipitation, two long short-term memory (LSTM) models, i.e., the recursive encoder-decoder structure-based
91 LSTM-RED model and the feature-temporal dual attention-based DA-LSTM-RED model, and two objective functions of
92 model calibration; and (3) to analyze and discuss the ensemble forecast performance of the proposed method in terms of the
93 deterministic and probabilistic forecast as compared with the HUP-BMA benchmark method. The interval basin between
94 Xiangjiaba Dam and the Three Gorges Dam is selected for the proposed study.

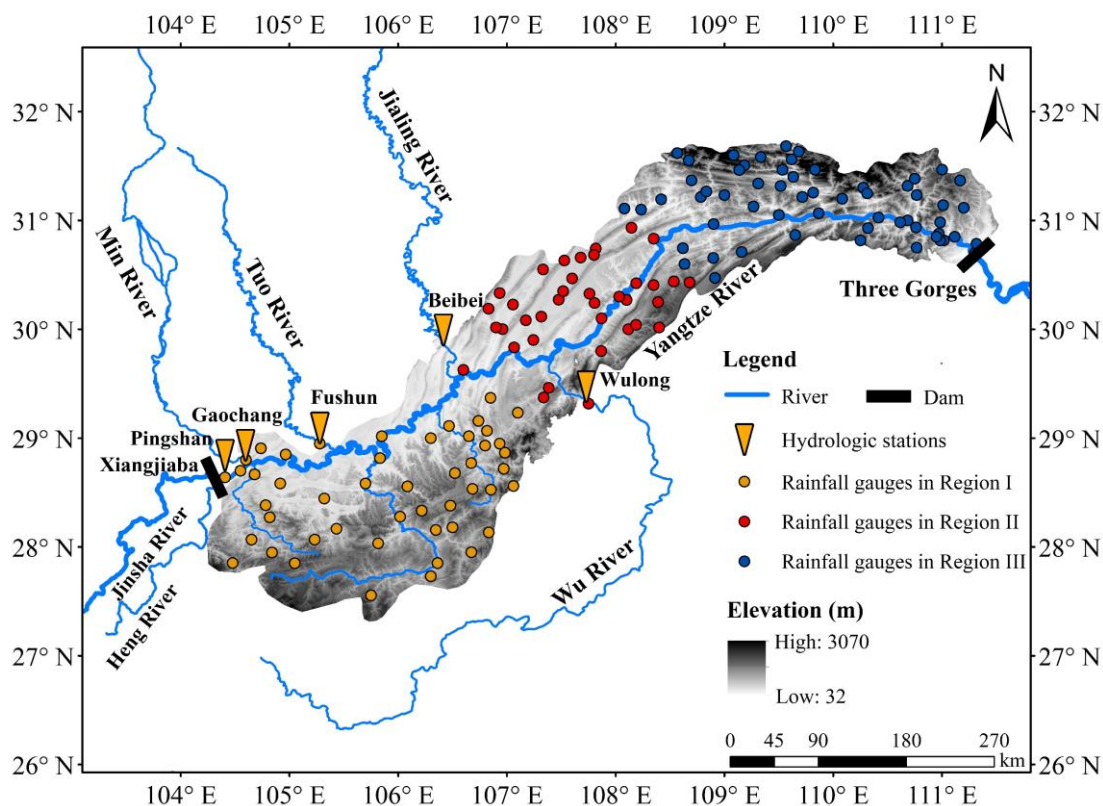
95 The rest of the paper is organized as follows. Section 2 introduces the case study and materials. The methods are
96 presented in Section 3. Section 4 evaluates the deterministic and ensemble forecast results. Conclusions and prospects are
97 given in Section 5.



98 2 Case study and materials

99 2.1 Study basin

100 Three Gorges Reservoir (TGR) is the largest hydraulic project in the world and plays a vital role in flood control, power
101 generation, and other water resource management issues (Zhong et al., 2020). The TGR controls a watershed area of about 1
102 million km². The total reservoir capacity is about 39.3 billion m³, with a flood control capacity of about 22.15 billion m³.
103



104
105 **Fig. 1** Schematic diagram of the interval-basin between Xiangjiaba and TGR dam-site which is divided into three sub-regions.
106

107 The TGR inflow is directly influenced by the runoff yield of the cascade reservoir interval-basin between Xiangjiaba
108 and TGR (Fig.1), with a basin area of about 127,400 km² (Zhou et al., 2019). The inflow of the TGR consists of the outflow
109 discharge from the Xiangjiaba Reservoir, the inflow of several tributaries such as Min, Tuo, Jialing, and Wu Rivers, and the
110 rainfall of the interval-basin. The flow sources are complex and have different effects on the TGR inflow. Moreover, TGR is
111 a river-type reservoir with a length of about 600 km at the normal storage level (175m) and an average width of only 1.1 km,
112 resulting in uncertainty in rainfall intensity and storm-centre positioning (Zhong et al., 2020). Therefore, there is significant
113 uncertainty in the flood forecast of TGR. It has been a major challenge to quantify and reduce forecast uncertainty.



114 Table 1 shows the flow propagation time from the hydrological control stations of the mainstream and tributaries to the
115 TGR dam. The outflow discharge of Xiangjiaba Reservoir, located on the Jinsha River, is observed at the Pingshan
116 hydrological station and represents the mainstream flow. The discharge values from large tributaries (Min, Jialing, Tuo, and
117 Wu Rivers) are observed at the Gaochang, Fushun, Beibei, and Wulong hydrological stations, respectively.

118 Considering the uneven distribution of rainfall intensity because of the narrow and long basin, the interval basin
119 between Xiangjiaba and TGR dam-site is divided into three sub-basins: Pingshan-Cuntan, Cuntan-Wanxian, and Wanxian-
120 TGR dam-site. Their watershed areas are 76,900, 22,900, and 27,600 km² respectively. Meanwhile, there are 45, 38, and 60
121 gauged rainfall stations in these three sub-regions, respectively.

122
123

Table 1 List of flow propagation time for hydrological control stations to TGR dam-site

Rivers	Hydrological control stations	To TGR dam flow propagation time (h)
Jinsha	Pingshan	48-66
Min	Gaochang	48-66
Jialing	Beibei	24-42
Tuo	Fushun	42-60
Wu	Wulong	15-30

124

125 2.2 Study materials

126 This study collects 6h observed flow discharges at TGR dam-site and five hydrological stations (Table 1), and 6h
127 observed rainfall in the interval-basin during the 2010-2021 flood season (May-September). The Tyson polygon method is
128 used to calculate areal average rainfall using rainfall station data for each sub-basin area. Meanwhile, this study collects the
129 forecasted precipitation data issued by the European Centre for Medium-Range Weather Forecasts (ECMWF) and the
130 Hydrology Bureau of the Yangtze River Water Resources Commission (HBYRWRC) for the 2017-2021 flood season in the
131 three sub-basins. Their forecast time starts at 8:00, with the 6-168h forecast horizons and the 6h forecast interval. The spatial
132 resolution of each grid for the ECMWF forecasted precipitation is 0.125°×0.125°. The HBYRWRC forecasted precipitation
133 is the areal average forecasted precipitation data.

134 The training period is from 2010 to 2016, and the validation period is from 2017 to 2021. Since the precipitation
135 forecast starts at 8:00 a.m., the forecasted flow for the 6-168h forecast horizons is also calculated from the daily 8:00 a.m. in
136 the validation period.



137 3 Methods

138 3.1 Proposed CHUP-BMA method

139 3.1.1 Bayesian model averaging (BMA)

140 Bayesian model averaging (BMA) method's principle is as follows.

$$p(Q_o | Q_{f,1}, Q_{f,2}, \dots, Q_{f,k}) = \sum_{i=1}^k w_i \cdot p(Q_o | Q_{f,i}) \quad (1)$$

141 where, where, $p(\cdot)$ denotes the probability density function. Q_o denotes the observed flow corresponding to the forecast
 142 moment (target value). k is the number of ensemble members. Q_f denotes the forecasted flow of ensemble members. w_i
 143 denotes the weight of the i -th model. $p(Q_o | Q_{f,i})$ denotes the conditional probability density of Q_o conditional on $Q_{f,i}$, which is
 144 assumed to approximately obey a normal distribution with the expectation of $\mu_i = a_i + b_i \cdot Q_{f,i}$ and variance of σ_i . a_i and b_i are the
 145 bias correction coefficients obtained by linear fitting of $Q_{f,i}$ to Q_o .

146 Therefore, Eq. (1) can be rewritten as follows.

$$p(Q_o | Q_{f,1}, Q_{f,2}, \dots, Q_{f,k}) = \sum_{i=1}^k w_i \cdot N(Q_o | \mu_i, \sigma_i) \quad (2)$$

147 From Eq. (2), it can be seen that the BMA method does not consider the influence of the initial state (the actual
 148 observed flow at the start of the forecast) on the posterior distribution. When the member forecasts at different times are the
 149 same, the posterior probability distribution generated by the BMA is also the same, which lacks logical rationality.

150 3.1.2 Hydrological uncertainty processor (HUP)

151 Based on the assumption that the precipitation uncertainty is zero, under the condition that the i -th ensemble member
 152 forecasts ($Q_{f,i}$) and the observed flow at the start of the forecast (Q_b), the posterior distribution of Q_o derived by the HUP
 153 method is as follows.

$$p(Q_o | Q_{f,i}, Q_b) = \frac{p(Q_{f,i} | Q_o, Q_b) \cdot p(Q_o | Q_b)}{\int_{-\infty}^{+\infty} p(Q_{f,i} | Q_o, Q_b) \cdot p(Q_o | Q_b) dQ_o} \quad (3)$$

154 where, $p(Q_o | Q_b)$ is the prior density function, $p(Q_{f,i} | Q_o, Q_b)$ is the likelihood density function. $p(Q_o | Q_{f,i}, Q_b)$ is the posterior
 155 density function.

156 The HUP method is a meta-Gaussian model assuming that the runoff series obeys a normal distribution, the core of
 157 which is the normal quantile transformation (Liu et al., 2016).

$$\hat{Q}_o = N^{-1}(P(Q_o)), \hat{Q}_{f,i} = N^{-1}(P(Q_{f,i})) \quad (4)$$

158 where, $P(\cdot)$ denotes the probability distribution function. $N^{-1}(\cdot)$ denotes the inverse function of the standard normal
 159 distribution. \hat{Q}_o and $\hat{Q}_{f,i}$ are the observed and forecasted flow transformed to the normal space, respectively.



160 The HUP method assumes that the observed flow obeys a first-order Markov process (Krzysztofowicz and Kelly, 2000),
 161 i.e., the flows between adjacent forecast horizons obey the linear constraint after the normal transformation.

$$\hat{Q}_{o,t} = c_t \times \hat{Q}_{o,t-1} + \varepsilon_t \quad (5)$$

162 where, $\hat{Q}_{o,t}$ is the observed flow corresponding to the t -th forecast horizon. c is the regression coefficient. ε is the residual,
 163 obeying $N(0, 1-c^2)$.

164 The prior density function expressions are as follows.

$$p(\hat{Q}_{o,t} | \hat{Q}_b) = \frac{1}{(1 - C_t^2)^{0.5}} n \left\{ \frac{\hat{Q}_{o,t} - C_t \times \hat{Q}_b}{(1 - C_t^2)^{0.5}} \right\}, C_t = \prod_{i=1}^t c_i \quad (6)$$

165 where, $n(\cdot)$ denotes standard normal density function; \hat{Q}_b is the observed flow at the start of the forecast transformed to the
 166 normal space.

167 \hat{Q}_b , \hat{Q}_o , and $\hat{Q}_{f,i}$ are assumed to obey a linear relationship. The expression of the likelihood function in normal space is
 168 as follows.

$$p(\hat{Q}_{f,i,t} | \hat{Q}_{o,t}, \hat{Q}_b) = \frac{1}{\sigma_t} n \left\{ \frac{\hat{Q}_{f,i,t} - (a_t \times \hat{Q}_{o,t} + d_t \times \hat{Q}_b + b_t)}{\sigma_t} \right\} \quad (7)$$

169 where, θ_t is an independent variable obeying $N(0, \sigma_t^2)$. a_t , d_t , and b_t are regression coefficients.

170 The posterior density function under normal space can be derived by substituting Eqs. (6) and (7) into Eq. (3).

$$p(\hat{Q}_{o,t} | \hat{Q}_{f,i,t}, \hat{Q}_b) = \frac{1}{Y_t} n \left\{ \frac{\hat{Q}_{o,t} - (A_t \times \hat{Q}_{f,i,t} + D_t \times \hat{Q}_b + B_t)}{Y_t} \right\}, \quad (8)$$

$$A_t = \frac{a_t y_t^2}{a_t^2 y_t^2 + \sigma_t^2}, B_t = \frac{-a_t b_t y_t^2}{a_t^2 y_t^2 + \sigma_t^2}, D_t = \frac{C_t \sigma_t^2 - a_t d_t y_t^2}{a_t^2 y_t^2 + \sigma_t^2}, Y_t = \left(\frac{y_t^2 \sigma_t^2}{a_t^2 y_t^2 + \sigma_t^2} \right)^{0.5}, y_t^2 = 1 - C_t^2$$

171 The posterior distribution function under the normal space can be converted to the original space by Jacobian
 172 transformation (Liu et al., 2016). The posterior density function of $Q_{o,t}$ under $Q_{f,i,t}$ and Q_b conditions is as follows.

$$p(Q_{o,t} | Q_{f,i,t}, Q_b) = \frac{J(Q_{o,t})}{Y_t} n \left\{ \frac{N^{-1}(P(Q_{o,t})) - A_t N^{-1}(P(Q_{f,i,t})) - D_t N^{-1}(P(Q_b)) - B_t}{Y_t} \right\}, \quad (9)$$

$$J(Q_{o,t}) = \frac{p(Q_{o,t})}{n(N^{-1}(P(Q_{o,t})))}$$

173 where, $J(\cdot)$ is the Jacobian transformation function.

174 3.1.3 HUP-BMA method

175 Darbandsari and Coulibaly et al. (2021) applied the hydrological uncertainty processor (HUP) to the ensemble forecast
 176 members, substituted the posterior density function obtained by the HUP method (Eq. (9)) into the BMA framework (Eq.



177 (2)), and then obtained the posterior distribution function of the target flow based on the initial state and the forecasted flow
 178 of the ensemble member. Therefore, the expression of the HUP-BMA method is as follows.

$$p(Q_o|Q_{f,1}, Q_{f,2}, \dots, Q_{f,k}, Q_b) = \sum_{i=1}^k w_i \cdot \frac{J(Q_{o,t})}{Y_t} n \left\{ \frac{N^{-1}(P(Q_{o,t})) - A_t N^{-1}(P(Q_{f,i,t})) - D_t N^{-1}(P(Q_b)) - B_t}{Y_t} \right\} \quad (10)$$

179 3.1.4 Copula-based HUP-BMA (CHUP-BMA) method

180 (1) Copula-based HUP

181 According to Sklar's theorem (Sklar, 1959), the joint distribution of m variables is as follows.

$$P(x_1, x_2, \dots, x_m) = C_m(P(x_1), P(x_2), \dots, P(x_m)) \quad (11)$$

182 where, $C_m(\cdot)$ denotes the m -dimensional copula distribution.

183 The copula-based HUP method (CHUP) was proposed by Liu et al. (2018), which can avoid the normal quantile
 184 transformation process of the flow series in the standard HUP method. With the help of the copula function, the prior density
 185 function in Eq. (3) can be derived as follows.

$$p(Q_o|Q_b) = \frac{\partial^2 C_2(P(Q_o), P(Q_b))}{\partial P(Q_o) \partial P(Q_b)} \cdot \frac{dP(Q_o)}{dQ_o} = c_2(P(Q_o), P(Q_b)) \cdot p(Q_o) \quad (12)$$

186 where, $c_m(\cdot)$ denotes the m -dimensional copula density function.

187 The likelihood density function in Eq. (3) can be derived as follows.

$$p(Q_{f,i}|Q_o, Q_b) = \frac{\frac{\partial^3 C_3(P(Q_o), P(Q_{f,i}), P(Q_b))}{\partial P(Q_o) \cdot \partial P(Q_{f,i}) \cdot \partial P(Q_b)} \cdot \frac{dP(Q_{f,i})}{dQ_{f,i}}}{\frac{\partial^2 C_2(P(Q_o), P(Q_b))}{\partial P(Q_o) \cdot \partial P(Q_b)}} = \frac{c_3(P(Q_o), P(Q_{f,i}), P(Q_b))}{c_2(P(Q_o), P(Q_b))} \cdot p(Q_{f,i}) \quad (13)$$

188 The posterior density function in Eq. (3) can be derived as follows.

$$p(Q_o|Q_{f,i}, Q_b) = \frac{c_3(P(Q_o), P(Q_{f,i}), P(Q_b))}{\int_0^1 c_3(P(Q_o), P(Q_{f,i}), P(Q_b)) dP(Q_o)} \cdot p(Q_o) \quad (14)$$

189 (2) Copula-based HUP-BMA method

190 Applying CHUP to the i -th ensemble member, the posterior probability distribution function $p(Q_o|Q_{f,i}, Q_b)$ of Q_o based
 191 on $Q_{f,i}$ and Q_b can be obtained. Coupling $p(Q_o|Q_{f,i}, Q_b)$ into the BMA framework, the copula-based HUP-BMA (CHUP-
 192 BMA) method can be constructed, and Eq. (2) can become as follows.

$$p(Q_o|Q_{f,1}, Q_{f,2}, \dots, Q_{f,k}, Q_b) = \sum_{i=1}^k w_i \cdot \frac{c_3(P(Q_o), P(Q_{f,i}), P(Q_b))}{\int_0^1 c_3(P(Q_o), P(Q_{f,i}), P(Q_b)) dP(Q_o)} \cdot p(Q_o) \quad (15)$$

193 The forecast uncertainty is quantified by the forecast interval with a 90% confidence level. Before constructing the
 194 copula, selecting the marginal distribution and the copula type is usually necessary. This study intends to select the
 195 appropriate marginal distribution and copula function from five common distribution functions, such as Pearson type III (P-



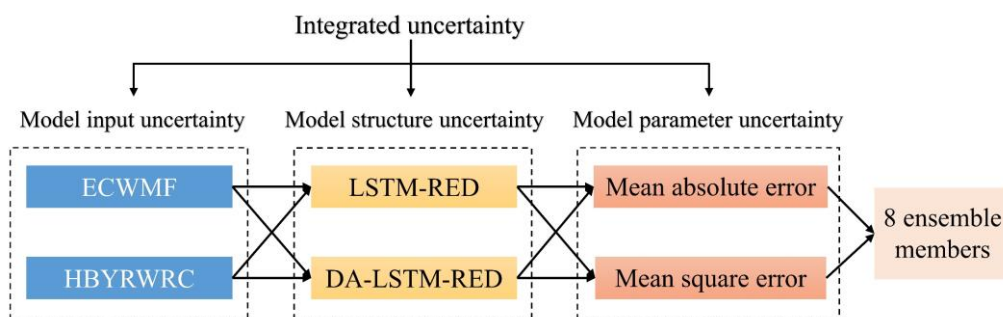
196 III), Gamma, Normal, Lognormal, and Weibull, and five common copula functions, such as Gumbel-Hougaard, Frank,
197 Clayton, Student-t (Student) and Gaussian copula, according to the root mean square error (RMSE) minimization criterion,
198 respectively. The definition and mathematical expressions of copula functions can be referred to Liu et al. (2018) and Chen
199 and Guo (2019).

200 Darbandsari and Coulibaly (2021) demonstrated that the HUP-BMA method could improve the probabilistic forecasting
201 performance of the HUP and BMA methods in the short forecast horizons. Therefore, this paper focuses on analyzing and
202 comparing the performance of the HUP-BMA and CHUP-BMA methods. The HUP-BMA and CHUP-BMA methods only
203 calibrate the ensemble members' weights through the Expectation-Maximization (EM) algorithm (Darbandsari and
204 Coulibaly, 2021). Meanwhile, since the forecast accuracy of ensemble members may change with time due to seasonality
205 and other factors (Zhong et al., 2020), the sliding window approach is used to update the weighting parameters. It has been
206 studied that the BMA method with sliding windows can obtain better probabilistic forecast performance (Parrish et al., 2012;
207 Darbandsari and Coulibaly, 2019).

208 3.2 Ensemble forecasting scheme

209 An ensemble forecast scheme containing multi-source uncertainties in the model input, the model structure, and the
210 parameter is constructed using a multi-member approach consisting of two forecasted precipitation, two models, and two
211 objective functions used to calibrate parameters, as shown in Fig. 2.

212



213

214

215

Fig. 2 The TGR's flood ensemble forecast scheme

216 3.2.1 Model input uncertainty

217 This study adopts forecasted precipitation released by ECMWF and HBYRWRC to consider the uncertainty of model
218 inputs. The TGR is a river-type reservoir, so building a river confluence model for flood forecasting is necessary. The
219 observed and forecasted precipitations are converted to effective precipitation, i.e., the precipitation after the loss of plant
220 retention, infiltration, evaporation, etc., to consider the runoff yield in the three sub-basin areas. The rainfall-runoff



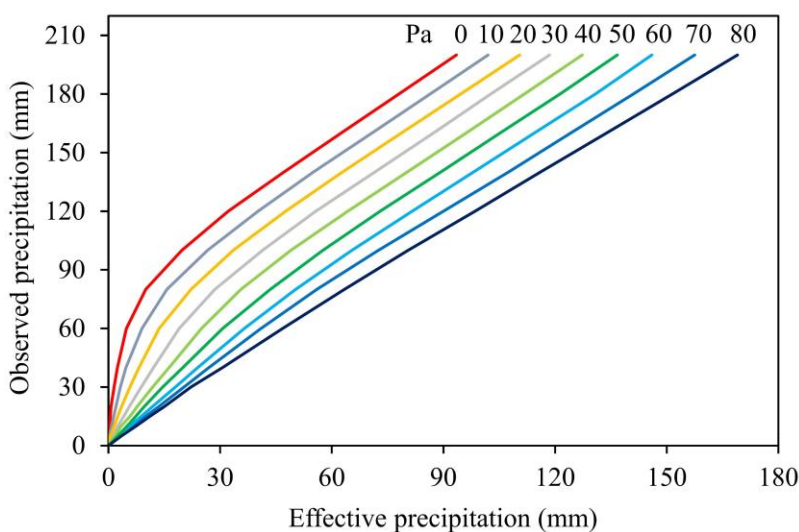
221 relationship graph method (Fedora and Beschta, 1989) commonly used in the Yangtze River basin (Fig. 3) calculates the
 222 effective precipitation in the three sub-basin areas. The core of this method is to calculate the antecedent precipitation index
 223 representing soil moisture (Zhong et al., 2018b), shown in the following equation.

$$P_{a,t+1} = k(P_{a,t} + P_t) \quad (16)$$

$$P_{a,t+1} \leq I_m \quad (17)$$

224 where, P_a denotes the antecedent precipitation index. P_t is the daily precipitation. I_m is the water storage capacity of the basin.
 225 k denotes evaporation reduction index. The k and I_m for the three sub-basin areas are shown in Table 2.

226



227

228

Fig. 3 Rainfall-runoff relationship between Xiangjiaba and Three Gorges Dam-site uncontrolled interval basin

229

230

Table 2 The k and I_m values for the three sub-basin areas

Sub-basin	k	I_m
Pingshan-Cuntan	0.90	50
Cuntan-Wanxian	0.95	80
Wanxian-TGR dam-site	0.95	80

231

232 After obtaining the daily antecedent precipitation index at 8:00, the antecedent precipitation index for the 6-h time scale
 233 is calculated as follows.

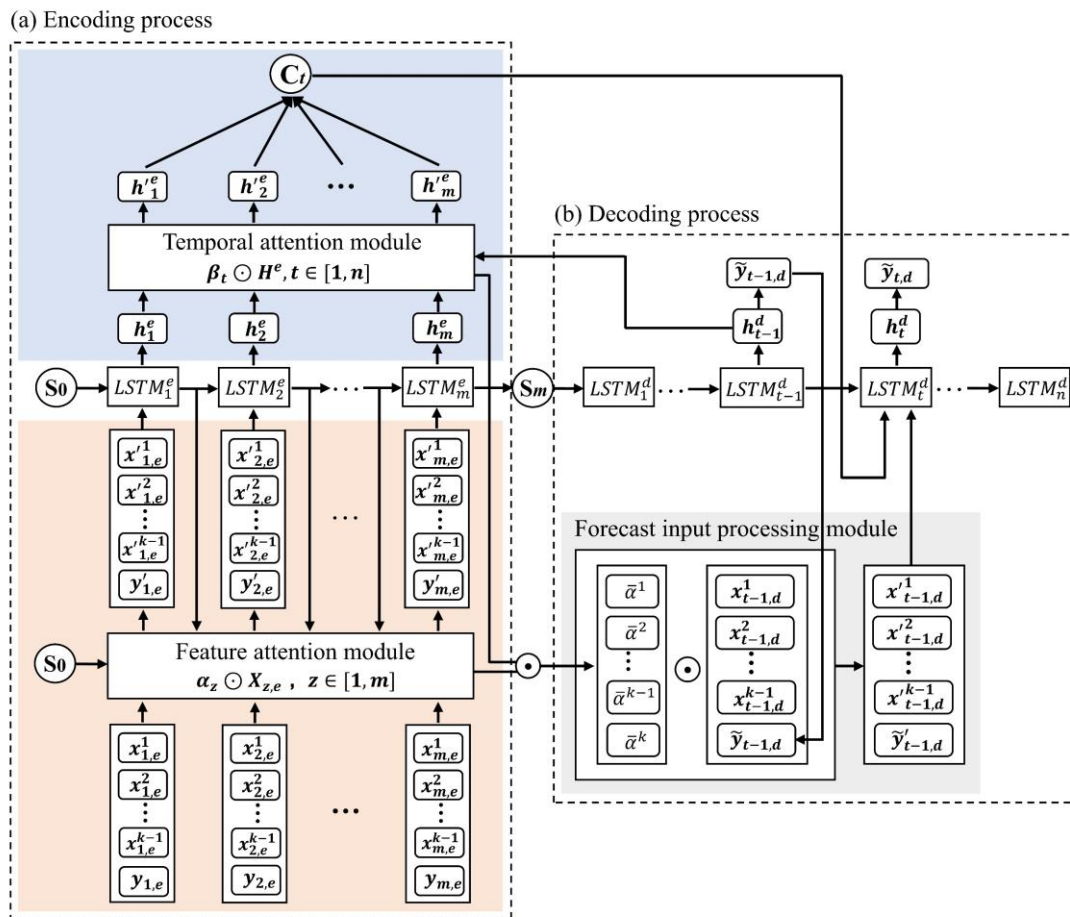
$$P_{a,t,m} = (P_{a,t} + \sum P_{t,n}) \times k^{\frac{h}{24}} \quad (18)$$

234 where, $P_{a,t,m}$ denotes the antecedent precipitation index at $m:00$ on the t -th day. $\sum P_{t,n}$ denotes the cumulative observed
 235 precipitation from 8:00 to $m:00$ on the t -th day. h denotes the time gap from 8:00 to $m:00$ on the t -th day.

236 **3.2.2 Model structure uncertainty**

237 The TGR inflow forecasting is influenced by the upstream mainstream and tributary reservoir scheduling decisions, the
 238 rainfall intensity and distribution in the interval basin, and the changes in the subsurface characteristics, which is challenging
 239 to establish complex and physical-based hydrological models (Yang et al., 2019; Cho et al., 2022; Hauswirth et al., 2023).
 240 The simulation or forecast accuracy in this interval-basin needs to be improved to meet the needs of the work. Therefore, two
 241 advanced data-driven models for obtaining multi-step-ahead flood processes forecasting, namely the long short-term
 242 memory (LSTM-RED) model based on an encoder-decoder structure and the coupled dual attention LSTM-RED (DA-
 243 LSTM-RED) model, are used for confluence calculations as a way to consider the uncertainty in the model structure.

244



245

246 **Fig. 4** Schematic diagram of the DA-LSTM-RED model. e and d are the encoding and decoding processes, respectively. k is the number of
 247 input types. $X_{z,e}$ is the input variables of encoding process, $X_{z,e} = \{x_{z,e}^1, x_{z,e}^2, \dots, x_{z,e}^{k-1}, y_{z,e}\}$. α_z denotes the weights of the input variables,
 248 $\alpha_z = \{\alpha_z^1, \alpha_z^2, \dots, \alpha_z^k\}$. m is the input time-steps in the encoding process. S is the hidden layer output. n is the maximum forecast horizon. H^e



249 is the hidden layer state, $H^e = \{h_1^e, h_2^e, \dots, h_m^e\}$. β_t denotes the weights of the hidden layer states of the encoding process, $\beta_t = \{\beta_t^1, \beta_t^2, \dots, \beta_t^m\}$.
250 C denotes the key information highlighted by the temporal attention. $\bar{\alpha}$ denotes the forecast input weights.

251

252 (1) Long short-term memory model based on encoder-decoder structure

253 This study nests an LSTM neural network into a recursive encoder-decoder (RED) structure that can be obtained for
254 forecasting flood processes to build an LSTM-RED model. Among them, the RED structure is similar to that of Kao et al.
255 (2020). The description of the LSTM neural network can be found in Cui et al. (2022). The encoding process of the RED
256 structure is used to extract the critical information (C_t) of the input (Xiang et al., 2020). In the decoding process, forecast
257 information of the same category as the encoding process is input to neural network of the latter moment, besides the C_t and
258 the output of the hidden layer at the previous moment.

259 (2) LSTM-RED neural network coupled dual attention mechanism

260 The LSTM-RED model based on the dual attention mechanism (DA-LSTM-RED) is established by adding the feature-
261 temporal dual attention mechanism to the LSTM-RED model, which can enable the model to highlight effective information
262 in different types and moments of the input. The DA mechanism (Fig. 4) consists of the feature attention module, the
263 temporal attention module, and the forecast input processing module.

264 The feature attention module can adaptively highlight the critical input types by assigning feature weights to the input
265 of the encoding process (Qin et al., 2017). The temporal attention module can highlight the information (hidden layer states)
266 extracted at critical time step by assigning temporal weights to the information extracted at all time step in the encoding
267 process (Ding et al., 2020). Meanwhile, the feature weights are averaged based on temporal weights and applied them to the
268 forecast information inputted in the decoding process, thus highlighting the key forecast input variables. The principle of the
269 DA-LSTM-RED model can be found in Cui et al. (2023).

270 (3) Model input and hyperparameter selection

271 In this study, the input types for the encoding process include effective precipitation in the three sub-basins, flow
272 discharge in the mainstream and tributaries (i.e., five hydrological stations in Table 1), and previously observed inflow to the
273 TGR for a total of nine types data. In order to make the model learn comprehensive information, input variables with the last
274 11-time steps (66h) are inputted to the encoding process according to the flow propagation times from the hydrological
275 stations to the TGR dam site in Table 1.

276 The forecasted effective precipitation, the forecasted flow of the mainstream and tributaries, and the forecasted flow for
277 the previous forecast horizon are used as inputs of the decoding process. Among them, the forecasted effective precipitation
278 is calculated by the observed precipitation during the training period and by the forecast precipitation during the validation
279 period. The forecasted flow of the mainstream and tributaries is replaced by the observed flow during the training and
280 validation periods. The TGR's observed inflow for the 6-168h forecast horizons is the target output, needed for practical
281 forecasting.



282 The input and output data are handled by the normalization method. Moreover, the trial-and-error method is used for
283 debugging the network hyperparameters. The model is trained by the Adam method.

284 3.2.3 Model parameter uncertainty

285 Different parameter-optimization objective functions may focus on different forecast results (Zhong et al., 2020). For
286 example, the mean absolute error function considers flow errors of different magnitudes equally. The mean square error
287 function usually magnifies the errors in the high flow, which may make the model parameters with different objective
288 functions produce forecast results with different focus points (Duan et al., 2007). Therefore, it is necessary to consider the
289 uncertainty of the model parameters. Neural network models usually train model parameters (such as model internal weights
290 and bias values, etc.) based on loss functions, so this paper uses two common loss functions, namely the mean absolute error
291 and the mean square error, to train the model as a way to consider the uncertainty of model parameters.

292 3.3 Evaluation metrics

293 3.3.1 Deterministic forecast evaluation metrics

294 The accuracy of deterministic forecast is evaluated by three metrics: the Nash-Sutcliffe efficiency (Nash and Sutcliffe,
295 1970) (NSE), the mean absolute error (MAE) and the relative error of total runoff (RE).

$$NSE = 1 - \frac{\sum_{i=1}^N (Q_{o,i} - Q_{f,i})^2}{\sum_{i=1}^N (Q_{o,i} - \bar{Q}_o)^2} \quad (19)$$

$$RE = \frac{\sum_{i=1}^N Q_{f,i} - \sum_{i=1}^N Q_{o,i}}{\sum_{i=1}^N Q_{o,i}} \times 100\% \quad (20)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Q_{o,i} - Q_{f,i}| \quad (21)$$

296 where, N is the sample number. \bar{Q}_o and \bar{Q}_f are the average of the observed and forecasted flow, respectively.

297 3.3.2 Probabilistic forecast evaluation metrics

298 (1) Forecast interval evaluation metrics

299 The forecast interval is evaluated by three metrics: the average coverage cate (CR), average interval width (IW), and the
300 percentage of observations bracketed by the unit confidence Interval (PUCI) (Li et al., 2011).

$$CR = \frac{n_c}{N} \quad (22)$$

$$IW = \frac{1}{N} \sum_{i=1}^N (Q_{u,i} - Q_{l,i}) \quad (23)$$



$$PUCI = \frac{CR}{\frac{1}{N} \sum_{i=1}^N \left(\frac{Q_{u,i} - Q_{l,i}}{Q_{o,i}} \right)} \quad (24)$$

301 where, n_c denotes the number of Q_o located in the forecast interval. Q_u and Q_l are the upper and lower boundaries of the
 302 forecast interval with a 90% confidence level, respectively.

303 (2) Probabilistic forecast evaluation metrics

304 The probabilistic forecast is evaluated by three metrics: the α _index (Renard et al., 2010), the ignorance score (IGS)
 305 (Gneiting et al., 2005), and continuous ranked probability score (CRPS) (Raftery et al., 2005).

$$\alpha_index = 1 - \frac{2}{N} \sum_{i=1}^N |q_{e,i} - q_{th,i}| \quad (25)$$

$$IGS = -\frac{1}{N} \sum_{i=1}^N \ln(p(Q_{o,i})) \quad (26)$$

$$CRPS = \frac{1}{N} \sum_{i=1}^N \int_0^{+\infty} (P_i(r) - I(r - Q_{o,i}))^2 dr, \quad (27)$$

$$I(r - Q_{o,i}) = \begin{cases} 1 & r \geq Q_{o,i} \\ 0 & r < Q_{o,i} \end{cases}$$

306 where, $q_{e,i}$ and $q_{th,i}$ denote observed and theoretical p-values of $Q_{o,i}$, respectively. p-value denotes the posterior probability
 307 distribution value of the $Q_{o,i}$ (Renard et al., 2010). $I(\cdot)$ denotes the indicative function. r denotes the flow variable.

308 The α _index metric can indicate the probabilistic forecast reliability and quantitatively evaluate the difference between
 309 the quantile-quantile (Q-Q) graph curve and the 1:1 line (Thyer et al., 2009; Laio and Tamea, 2007). The closer the α _index
 310 value is to 1, the more reliable the probabilistic forecast is. The IGS metric indicates the sharpness of the probabilistic
 311 forecast (Gneiting et al., 2005). The lower the IGS value, the better the sharpness and the lower the forecast uncertainty. The
 312 CRPS metric can reflect the reliability and sharpness of the probabilistic forecast and indicate the fit performance between
 313 the posterior probabilistic distribution and the actual probabilistic distribution of Q_o (Raftery et al., 2005).

314 4 Results evaluation

315 4.1 Deterministic forecast results of ensemble member

316 Since the study focuses on the differences in ensemble forecast performance between the HUP-BMA and CHUP-BMA
 317 methods, the overall forecast accuracy of members is analysed (Fig. 5), and the differences in forecast accuracy between
 318 members are not explicitly analysed. As shown in Fig. 5, using the observed values as input during the training period, high
 319 forecast accuracy can be acquired in different forecast horizons, with the NSE values exceeding 0.95 and the MAE values
 320 below 1400 m³/s, and the absolute value of RE within 4%.



321

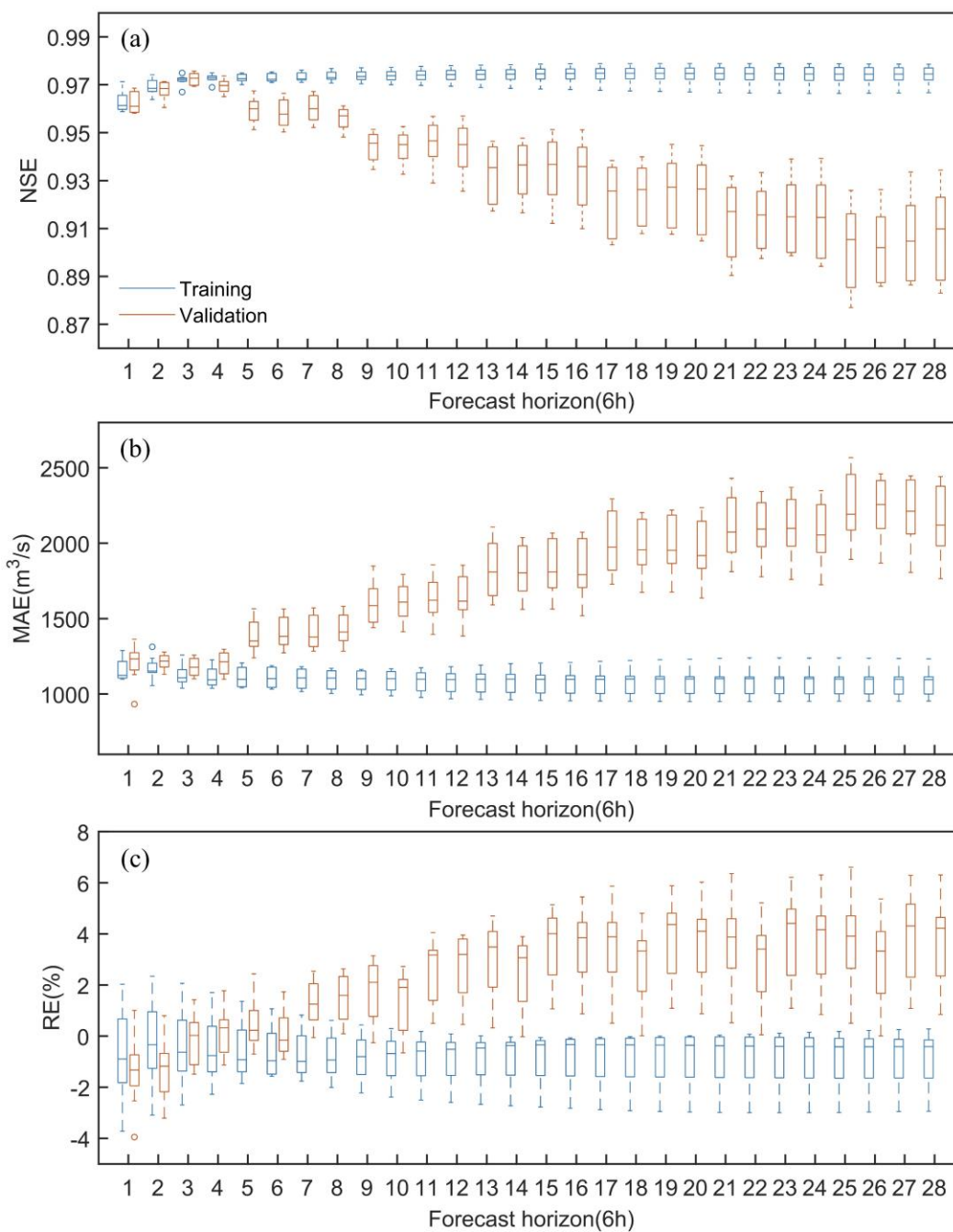


Fig. 5 Statistical chart of evaluation metrics of 8 ensemble members

322
323
324



325 After combining the forecasted precipitation during the validation period, the NSE values show a decreasing trend, and
 326 the MAE and RE values show an increasing trend with the increase of the forecast horizon. Taking the NSE metrics of the 1-
 327 7d forecast horizons as an example (Table 3), the average value of the NSE metric decreases from 0.97 to 0.89, which
 328 indicates that the forecast accuracy gradually decreases. Meanwhile, the range of evaluation metrics gradually increases with
 329 the increase of the forecast horizon. It can be seen from Table 3 that the difference between the maximum and minimum
 330 values of NSE indicators for the 1d forecast horizon is only 0.01. In contrast, the difference for the 7d forecast horizon is as
 331 high as 0.05, which indicates that the difference in forecast accuracy of members is also more significant, and the forecast
 332 uncertainty gradually increases. Overall, the NSE values of the forecast members in the 6-168h forecast horizons are higher
 333 than 0.88, and the absolute values of the RE metrics are within 7%. Hence, the forecast accuracy of members is relatively
 334 high, and the forecast error is low, which can be used for flood ensemble forecasting.

335

336 **Table 3** Mean, minimum, and maximum values of NSE metrics for 8 ensemble members in the validation period

Forecast horizon (h)	Mean	Max	Min	Forecast horizon (h)	Mean	Max	Min
6	0.96	0.97	0.96	90	0.93	0.95	0.91
12	0.97	0.97	0.96	96	0.93	0.95	0.91
18	0.97	0.98	0.97	102	0.92	0.94	0.90
24	0.97	0.97	0.97	108	0.92	0.94	0.91
30	0.96	0.97	0.95	114	0.93	0.95	0.91
36	0.96	0.97	0.95	120	0.92	0.94	0.90
42	0.96	0.97	0.95	126	0.91	0.93	0.89
48	0.96	0.96	0.95	132	0.91	0.93	0.90
54	0.94	0.95	0.93	138	0.92	0.94	0.90
60	0.94	0.95	0.93	144	0.91	0.94	0.89
66	0.95	0.96	0.93	150	0.90	0.93	0.88
72	0.94	0.96	0.93	156	0.90	0.93	0.89
78	0.93	0.95	0.92	162	0.91	0.93	0.89
84	0.93	0.95	0.92	168	0.91	0.93	0.88

337

338 4.2 Ensemble forecast results

339 4.2.1 Marginal distribution and copula function selection

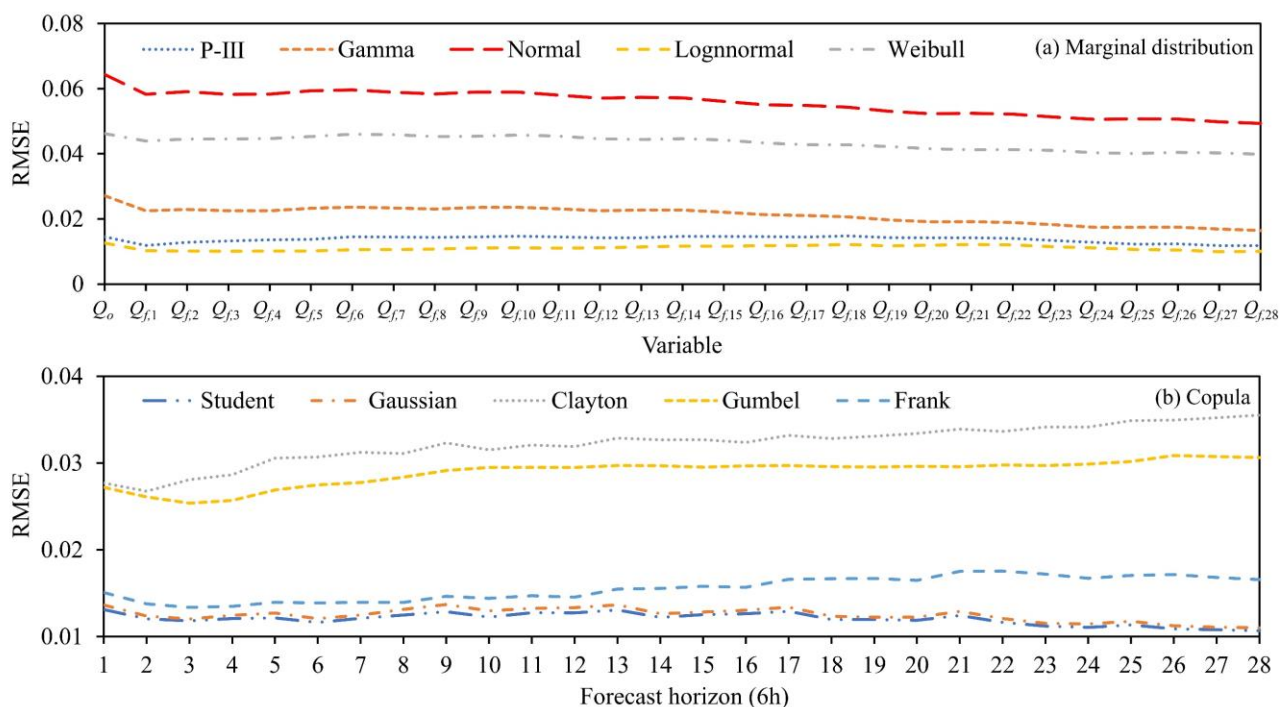
340 It is necessary first to fit the marginal distributions of the observed flow and the forecasted flow of the 6~168h forecast
 341 horizons. The Q_o and Q_b obey the same distribution. The RMSE criterion is used to select the marginal distribution type. In



342 each forecast horizon, the RMSE values of the 8 members are averaged to obtain the marginal distribution suitable for the
 343 forecasted flow intuitively. Meanwhile, according to Eq. (14), the three-dimensional joint distribution of Q_o , Q_b , and Q_f
 344 needs to be constructed. The RMSE criterion is used to select the copula function. Similarly, the RMSE values for the 8
 345 members of each forecast horizon were averaged.

346 Figs. 6 (a) and (b) show the RMSE values generated by fitting the marginal distribution and copula function,
 347 respectively. It can be seen from Fig. 6(a) that the Lognormal distribution has the lowest RMSE value among the five
 348 alternative marginal distributions and is chosen as the sequence marginal distribution type. As shown in Fig. 6(b), the student
 349 copula has the lowest RMSE value in the 6-168h forecast horizons and is chosen to construct the three-dimensional joint
 350 distribution function of Q_o , Q_b , and Q_f .

351



352

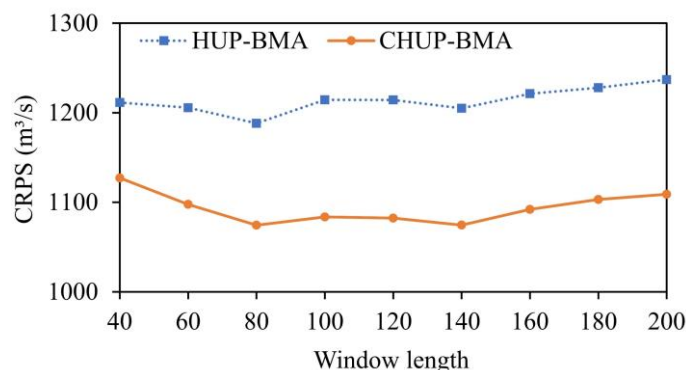
353 **Fig. 6** The RMSE values of Q_o , Q_b and Q_f sequence marginal distributions and copula functions. 1, 2, ..., 28 denote 6h, 12h, ..., 168h
 354 forecast horizons, respectively.

355 4.2.2 Sliding window length selection

356 Since there is no specific method or rule to calculate the sliding window length, this study adopts the CRPS metric as
 357 the objective function and the trial-and-error method to select the sliding window length. The range of window lengths is [40,
 358 200]. To facilitate the selection of the sliding window lengths, Fig. 7 shows the average CRPS values of the HUP-BMA and
 359 CHUP-BMA methods for all forecast horizons with different window lengths. It can be seen from Fig. 7 that the HUP-BMA



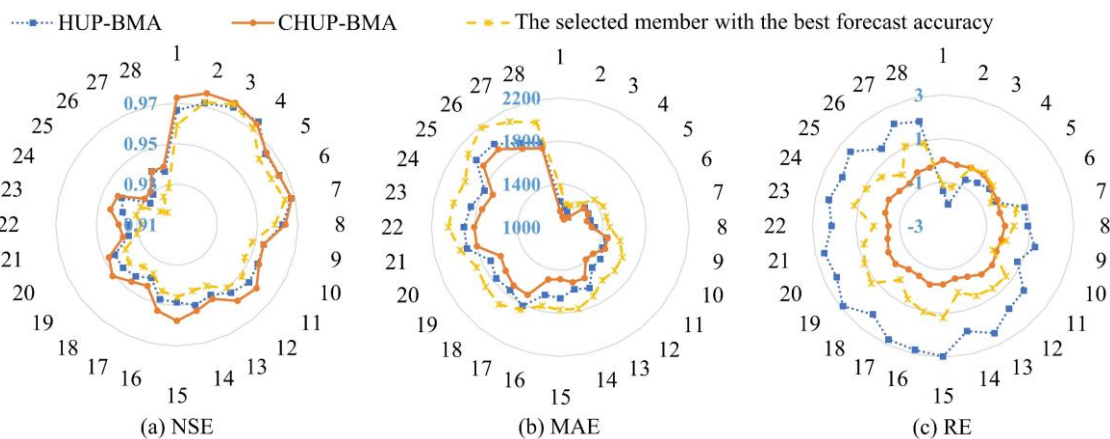
360 and CHUP-BMA methods all have the lowest CRPS values at the sliding window length of 80. Therefore, 80 is the optimal
 361 window length for the ensemble forecasting study.
 362



363
 364 **Fig. 7** The average CRPS values of the CHUP-BMA and HUP-BMA methods with different window lengths
 365

366 **4.2.3 Deterministic forecast results of ensemble forecast**

367 The HUP-BMA and CHUP-BMA methods use expectation values of ensemble forecasts as deterministic forecast
 368 results. In order to analyze the deterministic forecast performance of ensemble forecasts, one member with the best forecast
 369 accuracy is selected for comparative analysis based on the criteria of the relatively low RE and MAE values and relatively
 370 high NSE values.
 371



372
 373 **Fig. 8** Deterministic forecast evaluation metrics for the HUP-BMA, the CHUP-BMA, and the selected member with the best forecast
 374 accuracy
 375



376 Fig. 8 (a), (b), and (c) show the NSE, MAE, and RE metrics of three deterministic forecast results, respectively. It can
377 be seen that the NSE metrics show a decreasing trend, and the MAE metrics show an increasing trend as the forecast horizon
378 increases, indicating a gradual decrease in forecast accuracy.

379 As shown in Fig. 8(a), the NSE metrics of three forecast results are at least 0.92 during the 6-168h forecast horizons.
380 The difference between the two is small, not more than 0.02. Among them, the CHUP-BMA method has the best NSE
381 metrics. However, the advantage value gradually decreases as the forecast horizon increases. The NSE metrics of the HUP-
382 BMA method are better than those of the selected forecast member in most forecast horizons. From Fig. 8(b), the maximum
383 and mean values of MAE are 1923 and 1513 m³/s for the CHUP-BMA method, 1999 and 1582 m³/s for the HUP-BMA
384 method, and 2179 and 1719 m³/s for the selected forecast member, respectively. The CHUP-BMA method has the best MAE
385 metric, with the maximum and average reduction of 10.69% and 4.36% relative to the HUP-BMA method, respectively.
386 Meanwhile, the MAE values of two ensemble forecasting methods are lower than those of the selected forecast members. As
387 shown in Fig. 8(c), the maximum and mean of the RE metric are 0.02% and -0.27% for the CHUP-BMA method, 2.97% and
388 1.36% for the HUP-BMA method, and 1.20% and 0.34% for the selected forecast member, respectively. The CHUP-BMA
389 method can reduce the RE metrics of the selected forecast member in most forecast horizons, while the HUP-BMA method
390 has no advantage in the RE metric.

391 Overall, ensemble forecast methods can somewhat improve the selected best member forecast accuracy. The CHUP-
392 BMA method's expectation forecast has the best accuracy, which indicates that the copula-based CHUP-BMA method can
393 improve the performance of the HUP-BMA method in correcting errors.

394 To further analyze the accuracy of ensemble forecast methods, seven floods with peaks exceeding 50,000 m³/s during
395 the 24 and 168h forecast horizons in the validation period (2017-2021) are selected for analyzing. The average relative error
396 metric of peak (PRE) (Cui et al., 2022) is added to analyze the forecasting performance for flood peaks.

397 Table 4 demonstrates the forecast evaluation metrics for the seven flood events. With the increase in the forecast
398 horizon, the NSE metric shows a decreasing trend, and the RE and MAE metrics show an increasing trend, indicating a
399 gradual decrease in forecasting performance. It can be seen from Table 4 that (1) in the 24h forecast horizon, the forecast
400 accuracy of the two methods is similar for most flood events and quality metrics, (2) in the 168h forecast horizon, the
401 forecast accuracy of the CHUP-BMA method is better than HUP-BMA method in most flood events and quality metrics. The
402 average values of the NSE, RE, MAE, and PRE are 0.88, -0.63%, 2980m³/s, and -4.55% for CHUP-BMA, and 0.84, -2.38%,
403 3188m³/s, and -6.46% for HUP-BMA, respectively, indicating an overall improvement of CHUP-BMA over HUP-BMA in
404 forecasting accuracy.

405 To further demonstrate the accuracy of flood process forecasting and applicability of the two methods, four relatively
406 large flood events are selected for comparative analysis for 168h forecast horizon (Fig. 9).

407 In the 20180703-flood event (Fig. 9a), the two methods have similar forecast performance, underestimating the peak
408 and rising water processes and overestimating the receding water process. The CHUP-BMA method has relatively low PRE
409 values and total runoff error. The HUP-BMA method accurately forecasts the peak present time.



410

411

Table 4 Evaluation metrics for forecast flood events for 24 and 168h forecast horizons

Flood event	Method	Forecast horizon (h)	Evaluation metric			
			NSE	RE (%)	MAE (m ³ /s)	PRE (%)
20180703 (2018/7/1-7/7)	HUP-BMA	24	0.93	1.95	1697	-3.29
		168	0.80	1.69	2709	-8.60
	CHUP-BMA	24	0.94	3.63	1667	1.64
		168	0.78	1.30	2988	-6.26
20180714 (2018/7/11-7/17)	HUP-BMA	24	0.85	-1.38	2768	-8.04
		168	0.97	0.11	1101	0.88
	CHUP-BMA	24	0.84	-1.97	2874	-7.70
		168	0.95	-2.37	1587	-1.23
20200717 (2020/7/14-7/20)	HUP-BMA	24	0.91	-7.02	3094	-10.02
		168	0.64	-11.67	5965	-19.00
	CHUP-BMA	24	0.91	-4.75	3211	-8.80
		168	0.75	-7.45	5255	-13.58
20200727 (2020/7/25-7/31)	HUP-BMA	24	0.97	-0.22	1371	0.02
		168	0.84	-4.73	3044	-13.47
	CHUP-BMA	24	0.94	4.40	1819	3.62
		168	0.88	0.04	3155	-7.79
20200815 (2020/8/12-8/17)	HUP-BMA	24	0.93	-1.31	2714	-8.21
		168	0.94	-1.96	2259	-9.25
	CHUP-BMA	24	0.96	2.06	2062	-3.53
		168	0.95	3.05	2167	-3.82
20200820 (2020/8/18-8/24)	HUP-BMA	24	0.95	-0.79	2772	0.22
		168	0.92	5.74	3509	11.72
	CHUP-BMA	24	0.96	2.58	2125	2.60
		168	0.96	4.08	2816	6.06
20210907 (2021/9/4-9/10)	HUP-BMA	24	0.94	-3.26	2231	-7.43
		168	0.87	-4.66	3042	-13.15
	CHUP-BMA	24	0.97	-0.64	1722	-4.07
		168	0.94	-0.99	2016	-6.82

412

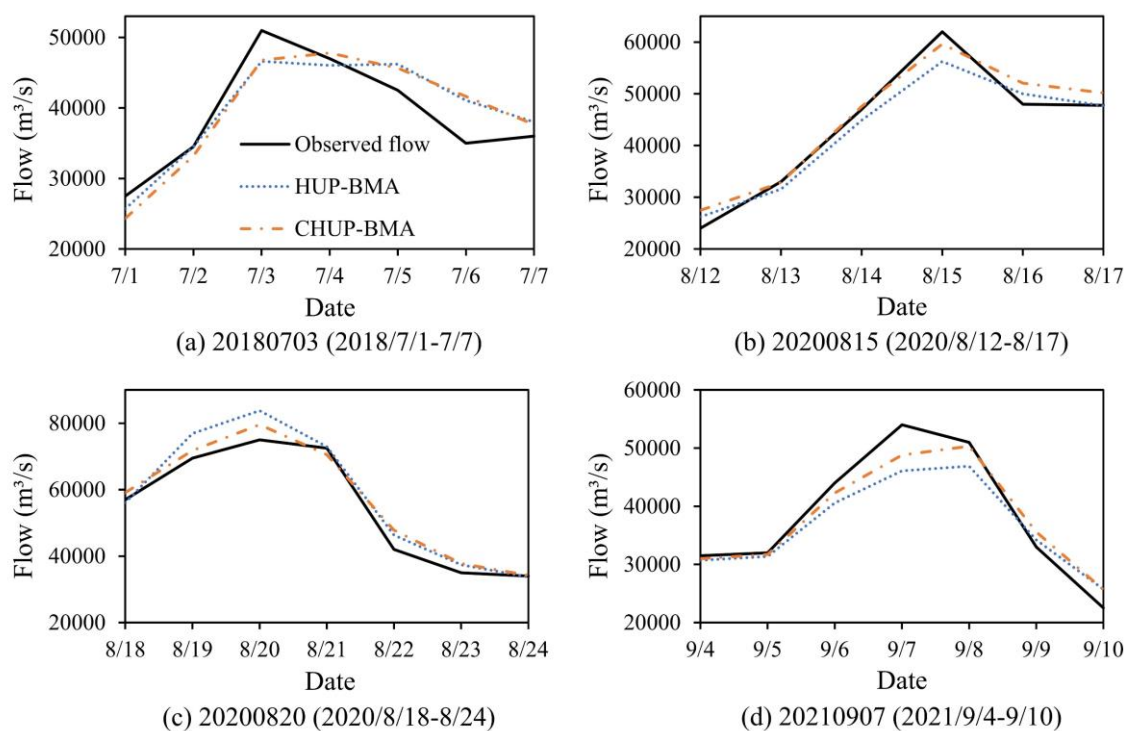


413 In the 20200815-flood event (Fig. 9b), two methods underestimate the flood peak and overestimate the receding water
414 process. The HUP-BMA method has a larger flood peak error, and the CHUP-BMA method has a better fitting performance.

415 In the 20200820-flood event (Fig. 9c), two methods overestimate the observed flood process, with the CHUP-BMA
416 method having the lower peak and total runoff error than the HUP-BMA method.

417 In the 20210907-flood event (Fig. 9d), the CHUP-BMA and HUP-BMA methods underestimate the flood peak and
418 delay the forecast peak occurring time. The former has smaller peak and water volume error.

419



420

421

422

Fig. 9 Forecasted flood events during 168h forecast horizon for the HUP-BMA and the CHUP-BMA methods

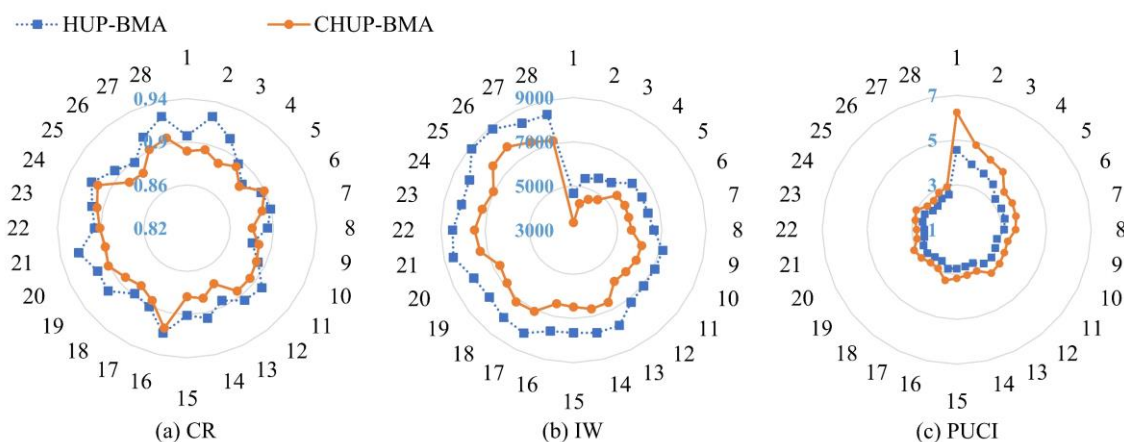
423 4.2.4 Probabilistic forecast results of ensemble forecast

424 (1) Evaluation of forecast interval

425 Figs. 10 (a), (b), and (c) show the CR, IW, and PUCI metrics for the forecast interval with a 90% confidence level,
426 respectively. Fig. 10(a) shows that during the 6-168h forecasting period, the maximum, minimum, and mean of the CR
427 metric for the forecast interval of the CHUP-BMA method are 0.92, 0.88, and 0.89, respectively, and 0.93, 0.88, and 0.91 for
428 the HUP-BMA method, respectively. The CR values of the two methods' forecast intervals are close to or exceed the 90%
429 confidence level, indicating that the forecast intervals are reliable.



430 It is obvious from Fig. 10(b) that the forecast interval width tends to increase with the increase of the forecast horizon,
431 indicating that the forecast uncertainty gradually increases. The maximum, minimum, and mean of the IW metrics for the
432 forecast interval of the CHUP-BMA method are 7820, 3337, and 6257 m³/s, respectively, and 8888, 4662, and 7345 m³/s for
433 the HUP-BMA method, respectively. The forecast intervals of the CHUP-BMA method are significantly narrower than those
434 of the HUP-BMA method, with the maximum and average reduction of 28.42% and 15.32%, respectively, which indicates
435 that the CHUP-BMA method can effectively reduce the interval width and forecast uncertainty.
436



437
438 **Fig. 10** Evaluation metrics of forecast intervals with the 90% confidence level of the HUP-BMA and CHUP-BMA methods
439

440 From Fig. 10(c), the maximum, minimum, and mean of the PUCI metric for the forecast interval of the CHUP-BMA
441 method are 6.24, 2.65, and 3.48, respectively, and 4.55, 2.35, and 2.95 for the HUP-BMA method, respectively. The CHUP-
442 BMA method has the higher PUCI values, indicating that the forecast interval of the CHUP-BMA method reflects the
443 forecast uncertainty relatively well.

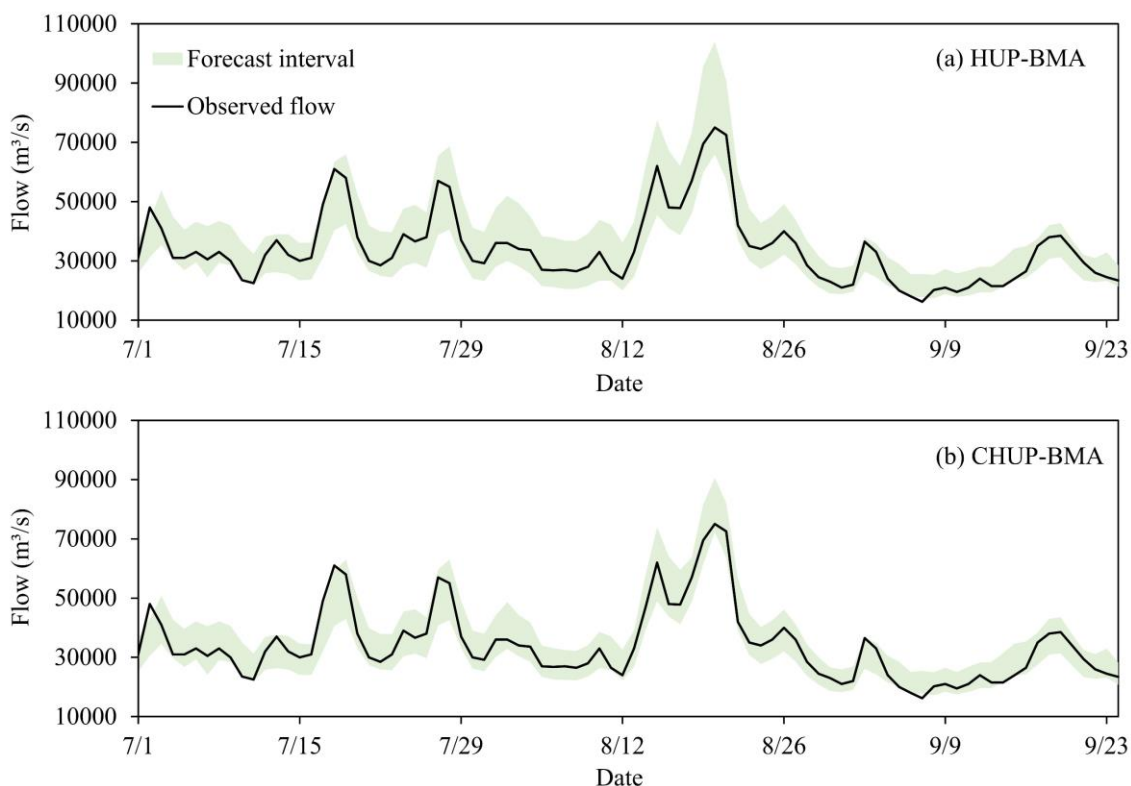
444 In summary, the CHUP-BMA outperforms the HUP-BMA method under the premise that the CR values are close to or
445 exceed the 90% confidence level. The CHUP-BMA method has narrower forecast intervals and better performance in
446 quantifying forecast uncertainty. Although the HUP-BMA method has a higher CR value, its IW value is larger, and the
447 PUCI value is smaller for the long forecast horizon, indicating that the forecast interval is too conservative to estimate the
448 uncertainty range reasonably.

449 In order to visually analyze the ability of the CHUP-BMA method to quantify forecast uncertainty, the forecast intervals
450 with a 90% confidence level of the HUP-BMA and CHUP-BMA methods for 168h forecast horizon in the 2020 flood season
451 are compared. It can be seen from Fig. 11 that the forecast intervals of the two ensemble forecasts can cover most of the
452 observed flows and always cover the annual maximum flood peak, indicating that the forecast intervals are reliable.
453 Meanwhile, the forecast intervals of the CHUP-BMA method are remarkably narrower than those of the HUP-BMA method,



454 indicating that the forecast uncertainty of the former is relatively low, which can provide more reasonable risk information
455 for TGR flood control decisions.

456



457

458 **Fig. 11** Forecast intervals with the 90% confidence level for the HUP-BMA and CHUP-BMA methods from 2020/7/1 8:00 to 9/24 8:00

459

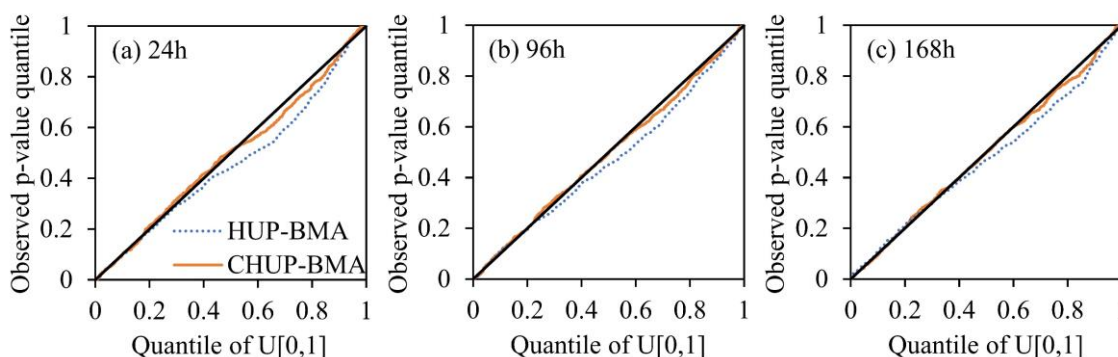
460 (2) Evaluation of overall probabilistic forecast

461 Fig. 12 shows the Q-Q plots of the HUP-BMA and CHUP-BMA methods for the 24, 96, and 168h forecast horizons.
462 The Q-Q curves of the CHUP-BMA method are closer to the 1:1 line, indicating that the probabilistic forecasts are the more
463 reliable. Meanwhile, the Q-Q curves of both methods are skewed to the lower right of the 1:1 line, indicating that the
464 forecasts are slightly overestimated, especially for HUP-BMA method.

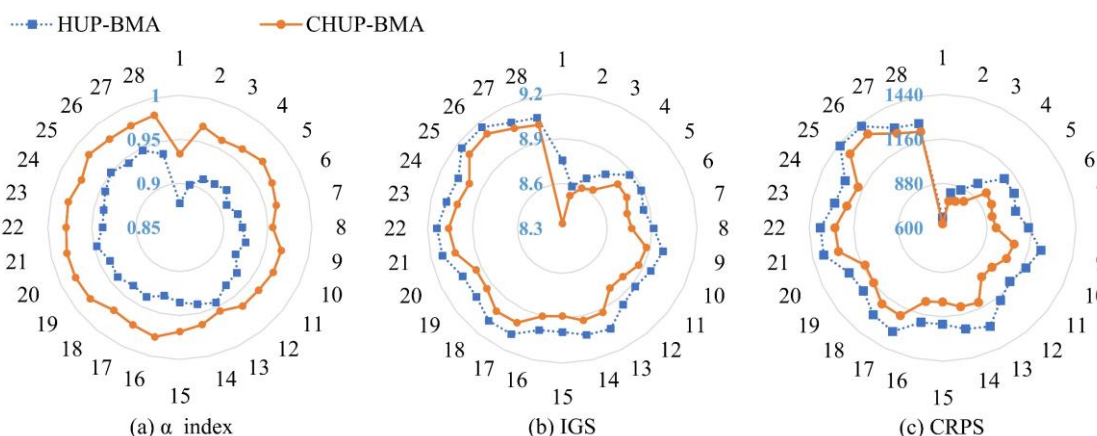
465 Meanwhile, Fig. 13 (a), (b), and (c) show the evaluation metrics of reliability (α _index), concentration (IGS), and
466 overall performance (CRPS) for the two ensemble probabilistic forecasts, respectively. It can be seen from Fig. 13(a) that the
467 α _index metrics of the CHUP-BMA method-based probabilistic forecasts are significantly higher than those of the HUP-
468 BMA method in the 6-168h forecast horizons. Among them, the maximum, minimum, and mean of the α _index metric for
469 CHUP-BMA method-based probabilistic forecasts are 0.98, 0.93, and 0.97, respectively, and 0.95, 0.88, and 0.93 for the



470 HUP-BMA method, respectively. The α _index metric of the CHUP-BMA method-based probabilistic forecast is closer to the
 471 perfect value of 1, indicating that its probability forecast is the more reliable.
 472



473
 474 **Fig. 12** Q-Q plots of the HUP-BMA and CHUP-BMA methods for the ensemble forecasts of the 24, 96, and 168h forecast horizons. U[0,1]
 475 denotes the uniform distribution on the interval [0,1].
 476



477
 478 **Fig. 13** Evaluation metrics of reliability, sharpness, and overall performance of two ensemble forecasts
 479

480 It can be seen from Fig. 13(b) that the IGS values of the two methods gradually increase with the increase of the
 481 forecast horizon, indicating that the forecast uncertainty gradually increases. The maximum, minimum, and mean of the IGS
 482 metric for the CHUP-BMA method are 9.10, 8.33, and 8.87, respectively, and 9.16, 8.59, and 8.98 for the HUP-BMA
 483 method, respectively. It can be seen that the IGS metrics of the CHUP-BMA method are consistently lower than those of the
 484 HUP-BMA method, which indicates that the CHUP-BMA method has a superior sharpness, assigns a high probability
 485 density around the actual values, and has a low forecast uncertainty.

486 As shown in Fig. 13(c), the CRPS values of the two methods are lower than the MAE values of the selected member
 487 (Fig. 8(b)), indicating that the probabilistic forecasts are effective and can fit the probabilistic distribution of the target values



488 well. Meanwhile, during the 6-168h forecast horizons, the maximum, minimum, and mean of the CRPS metric for the
489 CHUP-BMA method are 1356, 625, and 1074 m³/s, respectively, and 1425, 662, and 1188 m³/s for the HUP-BMA method,
490 respectively. It can be seen that the CRPS values of the CHUP-BMA method are lower than those of the HUP-BMA method,
491 with a maximum and average reduction of 17.86% and 9.71%, respectively. It can be seen that the CHUP-BMA method can
492 better fit the posterior distribution of the actual values and effectively improve the probabilistic forecast performance of the
493 HUP-BMA method.

494 In summary, the CHUP-BMA method considers the influence of the initial state on the ensemble forecast, bypasses the
495 normal quantile transformation of the HUP-BMA method, derives the posterior distribution of the target flow without
496 restrictions, and improves the probabilistic forecast performance of the HUP-BMA method. Therefore, the ensemble
497 forecasting by CHUP-BMA method can provide more reasonable and reliable risk information for the TGR.

498 **5 Conclusion and prospects**

499 This study proposed the CHUP-BMA method by coupling the copula-based HUP with the BMA method and
500 established an ensemble forecast scheme that consists of two forecasted precipitation, two hydrological models, and two
501 objective functions of parameter calibration. The ensemble forecasting performance of the HUP-BMA and CHUP-BMA
502 methods is discussed from the perspective of deterministic and probabilistic forecasts. The flood ensemble forecasting
503 experiment with 6-168h forecast horizons is conducted in the Xiangjiaba-TGR dam-site interval basin. The main conclusions
504 are summarized as follows.

505 (1) The two ensemble forecasting methods can improve the members' forecast accuracy. The proposed CHUP-BMA
506 method performs better than the HUP-BMA method, and the MAE metric value is reduced by a maximum of 10.69%.

507 (2) The coverage rate of the forecast interval of the CHUP-BMA method is close to or exceeds the specified 90%
508 confidence level, and the forecast interval is significantly narrower than that of the HUP-BMA method, with a maximum
509 reduction of 28.42%, which can effectively reduce the forecast uncertainty.

510 (3) The probabilistic forecast of the CHUP-BMA method has better reliability and sharpness, and its CRPS values are
511 reduced by a maximum of 17.86% relative to the HUP-BMA method, which indicates that the CHUP-BMA method can
512 better fit the posterior distribution of the actual values.

513 (4) The CHUP-BMA method can derive the posterior distribution of the target flow without restriction under the
514 condition of considering the initial constraint, which makes the BMA method more towards perfection. Therefore, it is more
515 suitable for the flood forecasting in the 6-168h forecast horizons and provides reliable risk information for reservoir
516 scheduling decision-making.

517 The present study focuses on flood ensemble forecasting for the TGR's 6-168h forecast horizons. Future studies can
518 explore the ensemble forecasting performance of the proposed CHUP-BMA method for longer forecast horizons and further
519 validate the effectiveness of the proposed method in global basins. Meanwhile, the effective way or method of guiding



520 reservoir scheduling based on ensemble forecasts can be further explored so that ensemble forecasts can be widely used in
521 decision-making.

522 **Code availability**

523 The code used to support the findings of this study are available from the corresponding author upon request.

524 **Data availability**

525 The data generated and/or analyzed during the current study are not publicly available for legal/ethical reasons but are
526 available from the corresponding author on reasonable request.

527 **Author contributions**

528 Zhen Cui and Shenglian Guo conceived and designed the experiments; Zhen Cui performed the experiments and wrote the
529 manuscript draft; Zhen Cui, Shenglian Guo, Chong-Yu Xu, Hua Chen, Dedi Liu, and Yanlai Zhou reviewed and edited the
530 manuscript.

531 **Competing interests**

532 The authors declare that they have no conflict of interest.

533 **Acknowledgments**

534 This study was financially supported by the National Key Research and Development Program of China (2022YFC3202801,
535 2021YFC3200305), and China Three Gorges Cooperation (0799254).

536 **References**

- 537 Biondi, D., and Todini, E. Comparing hydrological postprocessors including ensemble predictions into full predictive
538 probability distribution of streamflow. *Water Resour. Res.*, 54(12), 9860-9882, <https://doi.org/10.1029/2017WR022432>,
539 2018.
- 540 Chen, L., and Guo, S. *Copulas and its application in hydrology and water resources*. Springer Water, Springer Singapore,
541 <https://doi.org/10.1007/978-981-13-0574-0>, 2019.



- 542 Cho, K., and Kim, Y. Improving streamflow prediction in the WRF-Hydro model with LSTM networks. *J. Hydrol.*, 605,
543 127297. <https://doi.org/10.1016/j.jhydrol.2021.127297>, 2022.
- 544 Cloke, H. L., and Pappenberger, F. Ensemble flood forecasting: A review. *J. Hydrol.*, 375(3-4), 613-626.
545 <https://doi.org/10.1016/j.jhydrol.2009.06.005>, 2009.
- 546 Cui, Z., Guo, S., Zhou, Y., Wang, J. Exploration of dual-attention mechanism-based deep learning for multi-step-ahead flood
547 probabilistic forecasting. *J. Hydrol.*, 622, 129688. <https://doi.org/10.1016/j.jhydrol.2023.129688>, 2023.
- 548 Cui, Z., Zhou, Y., Guo, S., Wang, J., and Xu, C. Y. Effective improvement of multi-step-ahead flood forecasting accuracy
549 through encoder-decoder with an exogenous input structure. *J. Hydrol.*, 609, 127764.
550 <https://doi.org/10.1016/j.jhydrol.2022.127764>, 2022.
- 551 Darbandsari, P., and Coulibaly, P. Introducing entropy-based Bayesian model averaging for streamflow forecast. *J. Hydrol.*,
552 591, 125577. <https://doi.org/10.1016/j.jhydrol.2020.125577>, 2020.
- 553 Darbandsari, P., and Coulibaly, P. HUP-BMA: An Integration of Hydrologic Uncertainty Processor and Bayesian Model
554 Averaging for Streamflow Forecasting. *Water Resour. Res.*, 57(10), e2020WR029433.
555 <https://doi.org/10.1029/2020WR029433>, 2021.
- 556 Ding, Y., Zhu, Y., Feng, J., Zhang, P., and Cheng, Z. Interpretable spatial-temporal attention LSTM model for flood
557 forecasting. *Neurocomputing*, 403, 348-359. <https://doi.org/10.1016/j.neucom.2020.04.110>, 2020.
- 558 Duan, Q., Ajami, N. K., Gao, X., and Sorooshian, S. Multi-model ensemble hydrologic prediction using Bayesian model
559 averaging. *Adv. Water Resour.*, 30(5), 1371-1386. <https://doi.org/10.1016/j.advwatres.2006.11.014>, 2007.
- 560 Fedora, M. A., and Beschta, R. L. Storm runoff simulation using an antecedent precipitation index (API) model. *J. Hydrol.*,
561 112(1-2), 121-133. [https://doi.org/10.1016/0022-1694\(89\)90184-4](https://doi.org/10.1016/0022-1694(89)90184-4), 1989.
- 562 Ferretti, R., Lombardi, A., Tomassetti, B., Sangelantoni, L., Colaiuda, V., Mazzarella, V., Maiello, I., Verdecchia, M., and
563 Redaelli, G.: A meteorological-hydrological regional ensemble forecast for an early-warning system over small Apennine
564 catchments in Central Italy, *Hydrol. Earth Syst. Sci.*, 24, 3135-3156, <https://doi.org/10.5194/hess-24-3135-2020>, 2020.
- 565 Gelfan, A., Moreydo, V., Motovilov, Y., and Solomatine, D. P. Long-term ensemble forecast of snowmelt inflow into the
566 Cheboksary Reservoir under two different weather scenarios. *Hydrol. Earth Syst. Sci.*, 22(4), 2073-2089.
567 <https://doi.org/10.5194/hess-22-2073-2018>, 2018.
- 568 Gneiting, T., Raftery, A. E., Westveld, A. H., and Goldman, T. Calibrated probabilistic forecasting using ensemble model
569 output statistics and minimum CRPS estimation. *Mon. Weather Rev.*, 133(5), 1098-1118.
570 <https://doi.org/10.1175/MWR2904.1>, 2005.
- 571 Guo, Y., Yu, X., Xu, Y. P., Chen, H., Gu, H., and Xie, J. AI-based techniques for multi-step streamflow forecasts:
572 application for multi-objective reservoir operation optimization and performance assessment. *Hydrol. Earth Syst. Sci.*,
573 25(11), 5951-5979. <https://doi.org/10.5194/hess-25-5951-2021>, 2021.
- 574 Han, S., and Coulibaly, P. Bayesian flood forecasting methods: A review. *J. Hydrol.*, 551, 340-351.
575 <https://doi.org/10.1016/j.jhydrol.2017.06.004>, 2017.



- 576 Hauswirth, S. M., Bierkens, M. F., Beijk, V., and Wanders, N. The suitability of a seasonal ensemble hybrid framework
577 including data-driven approaches for hydrological forecasting. *Hydrol. Earth Syst. Sci.*, 27(2), 501-517.
578 <https://doi.org/10.5194/hess-27-501-2023>, 2023.
- 579 Huang, H., Liang, Z., Li, B., Wang, D., Hu, Y., and Li, Y. Combination of multiple data-driven models for long-term
580 monthly runoff predictions based on Bayesian model averaging. *Water Resour. Manag.*, 33, 3321-3338.
581 <https://doi.org/10.1007/s11269-019-02305-9>, 2019.
- 582 Kao, I. F., Zhou, Y., Chang, L. C., and Chang, F. J. Exploring a Long Short-Term Memory based Encoder-Decoder
583 framework for multi-step-ahead flood forecasting. *J. Hydrol.*, 583, 124631. <https://doi.org/10.1016/j.jhydrol.2020.124631>,
584 2020.
- 585 Krzysztofowicz, R. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.*,
586 35(9), 2739-2750. <https://doi.org/10.1029/1999WR900099>, 1999.
- 587 Krzysztofowicz, R., and Kelly, K. S. Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resour.*
588 *Res.*, 36(11), 3265-3277. <https://doi.org/10.1029/2000WR900108>, 2000.
- 589 Kussul, N., Shelestov, A., and Skakun, S. Grid system for flood extent extraction from satellite images. *Earth Sci. Inform.*, 1,
590 105-117. <https://doi.org/10.1007/s12145-008-0014-3>, 2008.
- 591 Laio, F., and Tamea, S. Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrol. Earth Syst.*
592 *Sci.*, 11(4), 1267-1277. <https://doi.org/10.5194/hess-11-1267-2007>, 2007.
- 593 Li, L., Xu, C. Y., Xia, J., Engeland, K., and Reggiani, P. Uncertainty estimates by Bayesian method with likelihood of AR (1)
594 plus Normal model and AR (1) plus Multi-Normal model in different time-scales hydrological models. *J. Hydrol.*, 406(1-2),
595 54-65. <https://doi.org/10.1016/j.jhydrol.2011.05.052>, 2011.
- 596 Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., and Di, Z. A review on statistical postprocessing methods for
597 hydrometeorological ensemble forecasting. *Wiley Interdiscip. Rev.-Water*, 4(6), e1246. <https://doi.org/10.1002/wat2.1246>,
598 2017.
- 599 Liu, J., Yuan, X., Zeng, J., Jiao, Y., Li, Y., Zhong, L., and Yao, L. Ensemble streamflow forecasting over a cascade reservoir
600 catchment with integrated hydrometeorological modeling and machine learning. *Hydrol. Earth Syst. Sci.*, 26(2), 265-278.
601 <https://doi.org/10.5194/hess-26-265-2022>, 2022.
- 602 Liu, Z., Guo, S., Xiong, L., and Xu, C. Y. Hydrological uncertainty processor based on a copula function. *Hydrol. Sci. J.-J.*
603 *Sci. Hydrol.*, 63(1), 74-86. <https://doi.org/10.1080/02626667.2017.1410278>, 2018.
- 604 Liu, Z., Guo, S., Zhang, H., Liu, D., and Yang, G. Comparative study of three updating procedures for real-time flood
605 forecasting. *Water Resour. Manag.*, 30, 2111-2126. <https://doi.org/10.1007/s11269-016-1275-0>, 2016.
- 606 Madadgar, S., and Moradkhani, H. Improved Bayesian multi-modelling: Integration of copulas and Bayesian model
607 averaging. *Water Resour. Res.*, 50(12), 9586-9603. <https://doi.org/10.1002/2014WR015965>, 2014.



- 608 Matthews, G., Barnard, C., Cloke, H., Dance, S. L., Jurlina, T., Mazzetti, C., and Prudhomme, C. Evaluating the impact of
609 post-processing medium-range ensemble streamflow forecasts from the European Flood Awareness System. *Hydrol. Earth*
610 *Syst. Sci.*, 26(11), 2939-2968. <https://doi.org/10.5194/hess-26-2939-2022>, 2022.
- 611 Nash, J. E., and Sutcliffe, J. V. River flow forecasting through conceptual models: part I—A discussion of principles. *J.*
612 *Hydrol.*, 10(3), 282-290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.
- 613 Parrish, M. A., Moradkhani, H., and DeChant, C. M. Toward reduction of model uncertainty: Integration of Bayesian model
614 averaging and data assimilation. *Water Resour. Res.*, 48(3). <https://doi.org/10.1029/2011WR011116>, 2012.
- 615 Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., and Cottrell, G. A dual-stage attention-based recurrent neural network for
616 time series prediction. *arXiv preprint arXiv:1704.02971*. <https://doi.org/10.48550/arXiv.1704.02971>, 2017.
- 617 Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. Using Bayesian model averaging to calibrate forecast
618 ensembles. *Mon. Weather Rev.*, 133(5), 1155-1174. <https://doi.org/10.1175/MWR2906.1>, 2005.
- 619 Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W. Understanding predictive uncertainty in hydrologic
620 modeling: The challenge of identifying input and structural errors. *Water Resour. Res.*, 46(5).
621 <https://doi.org/10.1029/2009WR008328>, 2010.
- 622 Shu, Z., Zhang, J., Wang, L., Jin, J., Cui, N., Wang, G., ..., and Liu, C. Evaluation of the impact of multi-source uncertainties
623 on meteorological and hydrological ensemble forecasting. *Engineering*. <https://doi.org/10.1016/j.eng.2022.06.007>, 2022.
- 624 Saleh, F., Ramaswamy, V., Georgas, N., Blumberg, A. F., and Pullen, J. A retrospective streamflow ensemble forecast for an
625 extreme hydrologic event: a case study of Hurricane Irene and on the Hudson River basin. *Hydrol. Earth Syst. Sci.*, 20(7),
626 2649-2667. <https://doi.org/10.5194/hess-20-2649-2016>, 2016.
- 627 Sklar, M. Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8, 229-231. 1959.
- 628 Sloughter, J. M., Gneiting, T., and Raftery, A. E. Probabilistic wind speed forecasting using ensembles and Bayesian model
629 averaging. *J. Am. Stat. Assoc.*, 105(489), 25-35. <https://doi.org/10.1198/jasa.2009.ap08615>, 2010.
- 630 Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W., and Srikanthan, S. Critical evaluation of parameter
631 consistency and predictive uncertainty in hydrological modelling: A case study using Bayesian total error analysis. *Water*
632 *Resour. Res.*, 45(12). <https://doi.org/10.1029/2008WR006825>, 2009.
- 633 Todini, E. A model conditional processor to assess predictive uncertainty in flood forecasting. *Int. J. River Basin Ma.*, 6(2),
634 123-137. <https://doi.org/10.1080/15715124.2008.9635342>, 2008.
- 635 Vegad, U., and Mishra, V. Ensemble streamflow prediction considering the influence of reservoirs in Narmada River Basin,
636 India. *Hydrol. Earth Syst. Sci.*, 26(24), 6361-6378. <https://doi.org/10.5194/hess-26-6361-2022>, 2022.
- 637 Wei, L., Jiang, S., Dong, J., Ren, L., Liu, Y., Zhang, L., ..., and Duan, Z. Fusion of gauge-based, reanalysis, and satellite
638 precipitation products using Bayesian model averaging approach: Determination of the influence of different input sources. *J.*
639 *Hydrol.*, 618, 129234. <https://doi.org/10.1016/j.jhydrol.2023.129234>, 2023.
- 640 Xiang, Z., Yan, J., and Demir, I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resour.*
641 *Res.*, 56(1), e2019WR025326. <https://doi.org/10.1029/2019WR025326>, 2020.



- 642 Xu, C., Zhong, P. A., Zhu, F., Yang, L., Wang, S., and Wang, Y. Real-time error correction for flood forecasting based on
643 machine learning ensemble method and its uncertainty assessment. *Stoch. Environ. Res. Risk Assess.*, 1-21.
644 <https://doi.org/10.1007/s00477-022-02336-6>, 2022.
- 645 Yang, T., Sun, F., Gentine, P., Liu, W., Wang, H., Yin, J., ..., and Liu, C. Evaluation and machine learning improvement of
646 global hydrological model-based flood simulations. *Environmental Research Letters*, 14(11), 114027.
647 <https://doi.org/10.1088/1748-9326/ab4d5e>, 2019.
- 648 Zhang, B., Wang, S., Qing, Y., Zhu, J., Wang, D., and Liu, J. A vine copula-based polynomial chaos framework for
649 improving multi-model hydroclimatic projections at a multi-decadal convection-permitting scale. *Water Resour. Res.*, 58(6),
650 e2022WR031954. <https://doi.org/10.1029/2022WR031954>, 2022.
- 651 Zhou, Y., Cui, Z., Lin, K., Sheng, S., Chen, H., Guo, S., and Xu, C. Y. Short-term flood probability density forecasting using
652 a conceptual hydrological model with machine learning techniques. *J. Hydrol.*, 604, 127255.
653 <https://doi.org/10.1016/j.jhydrol.2021.127255>, 2022.
- 654 Zhou, Y., Guo, S., and Chang, F. J. Explore an evolutionary recurrent ANFIS for modelling multi-step-ahead flood forecasts.
655 *J. Hydrol.*, 570, 343-355. <https://doi.org/10.1016/j.jhydrol.2018.12.040>, 2019.
- 656 Zhong, Y., Guo, S., Ba, H., Xiong, F., Chang, F. J., and Lin, K. Evaluation of the BMA probabilistic inflow forecasts using
657 TIGGE numeric precipitation predictions based on artificial neural network. *Hydrol. Res.*, 49(5), 1417-1433.
658 <https://doi.org/10.2166/nh.2018.177>, 2018a.
- 659 Zhong, Y., Guo, S., Liu, Z., Wang, Y., and Yin, J. Quantifying differences between reservoir inflows and dam site floods
660 using frequency and risk analysis methods. *Stoch. Environ. Res. Risk Assess.*, 32, 419-433. [https://doi.org/10.1007/s00477-](https://doi.org/10.1007/s00477-017-1401-4)
661 017-1401-4, 2018b.
- 662 Zhong, Y., Guo, S., Xiong, F., Liu, D., Ba, H., and Wu, X. Probabilistic forecasting based on ensemble forecasts and EMOS
663 method for TGR inflow. *Front. Earth Sci.*, 14, 188-200. <https://doi.org/10.1007/s11707-019-0773-9>, 2020.
- 664