# AI-based techniques for multi-step streamflow forecasts: Application for multi-objective reservoir operation optimization and performance assessment

Yuxue Guo, Xinting Yu, Yue-Ping Xu, Hao Chen, Haiting Gu, Jingkai Xie
Institute of Hydrology and Water Resources, Civil Engineering and Architecture, Zhejiang University, Hangzhou, 310058, China

*Correspondence to*: *Yue-Ping Xu (yuepingxu@zju.edu.cn)*

**Abstract.** Streamflow forecasts are traditionally effective in mitigating water scarcity and flood defense. This study developed an Artificial Intelligence (AI)-based management methodology that integrated multi-step streamflow forecasts and multi-objective reservoir operation optimization for water resource allocation. Following the methodology, we aimed to assess forecast quality and forecast-informed reservoir operations performance together due to the influence of inflow forecast uncertainty. Varying combinations of climate and hydrological variables were input into three AI-based models, namely Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Least Squares Support Vector Machine (LSSVM), to forecast short-term streamflow. Based on three deterministic forecasts, the stochastic inflow scenarios were further developed using Bayesian Model Averaging (BMA) for quantifying uncertainty. The forecasting scheme was further coupled with a multi-reservoir optimization model, and the multi-objective programming was solved using the parameterized Multi-Objective Robust Decision Making (MORDM) approach. The AI-based management framework was applied and demonstrated over a multi-reservoir system (25 reservoirs) in the Zhoushan Islands, China. Three main conclusions were drawn from this study: 1) GRU and LSTM performed equally well on streamflow forecasts, and GRU might be the preferred method over LSTM, given that it had simpler structures and less modeling time; 2) Higher forecast performance could lead to improved reservoir operation, while uncertain forecasts were more valuable than deterministic forecasts, regarding two performance metrics, i.e., water supply reliability and operating costs; 3) The relationship between forecast horizon and reservoir operation was complex and depended on the operating configurations (forecast quality and uncertainty) and performance measures. This study reinforces the potential of an AI-based stochastic streamflow forecasting scheme to seek robust strategies under uncertainty.

**Keywords:** AI-based techniques, multi-step streamflow forecasts, reservoir operation, forecast quality, forecast horizon.

## 1 Introduction

Multi-step streamflow forecast is of great importance for reservoir operations to determine optimal water allocations considering the current use and the carry-out storage for mitigating water scarcity risk in the future (Guo et al., 2018; Zhao et al., 2019). Previous studies have identified that real-time reservoir operations are influenced by multiple uncertainties (Xu et

al., 2020), among which inflow forecast uncertainty has been determined as the primary source, resulting in the risk of water shortage when the forecast inflow overestimates the actual inflow. Ensemble forecasting techniques are commonly used to characterize various uncertainties in streamflow forecasts. According to comparative analysis for various probabilistic forecasting techniques (Nott et al., 2012; Fang et al., 2018a; Zhai and Chen, 2018; Zhou et al., 2020b), Bayesian Model Averaging (BMA) (Hoeting et al., 1999) has been found to be an effective and most commonly used method to evaluate uncertainty and thus can be used in streamflow forecast.

Any ensemble forecast approach relies upon model diversity that different models produce, with specific emphasis and different aspects of the features they want to model (Zhou et al., 2020a). In the last few decades, many approaches have been developed to forecast streamflow, including physically-based and data-driven models (Tikhamarine et al., 2020; Zuo et al., 2020). Although physically-based models can help understand underlying physical processes, they usually require a large amount of input information, such as meteorological data, geographic data, soil, and land use characteristics (Guo et al., 2018; Guo et al., 2020a). Different from physically-based models, data-driven models based on statistical modeling have attracted significant interest due to their simplicity and satisfactory forecast results with low information requirements (Al-Sudani et al., 2019; Mehdizadeh et al., 2019; Osman et al., 2020). Artificial intelligence (AI)-based approaches, i.e., machine learning (ML) methods, belong to the latter group. The widely used ML approaches include Artificial neural network (ANN) and Least Squares Support Vector Machine (LSSVM) (Ghumman et al., 2018; Kisi et al., 2019; Meng et al., 2019; Adnan et al., 2020; Ali and Shahbaz, 2020). Such models have been proven to be efficient tools to model qualitative and quantitative hydrological variables and deal with non-linear features in streamflow. In recent years, the booming development of deep learning technology has brought many new approaches, such as recurrent neural networks (RNNs) (Elman, 1990), one of the most popular neural networks in the deep learning field. RNNs can preserve and remember the short-past and long-past information and thus are preferred for a complex and highly non-linear timing problem. Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997) and Gated Recurrent Unit (GRU) (Cho et al., 2014) are two different versions of RNNs. LSTM and GRU networks have been successfully applied in many fields (Greff et al., 2017; Zhang et al., 2018; Jung et al., 2020; Shahid et al., 2020; Ayzel and Heistermann, 2021), and they are demonstrated to generate comparable performances, But GRU has a more straightforward structure and a higher operation speed than LSTM. Recently, many applications that assessed them together are also found in the hydrological field (Gao et al., 2020; Muhammad et al., 2020).

While a considerable research effort has been made to evaluate and improve the quality of streamflow forecasts (Gibbs et al., 2018; Nanda et al., 2019; Sharma et al., 2019; Van Osnabrugge et al., 2019; Feng et al., 2020; Pechlivanidis et al., 2020), how forecasts impact decision-making in the real-time reservoir operations has also gradually gained researchers' attention (Goddard et al., 2010; Shamir, 2017; Anghileri et al., 2019; Alexander et al., 2020; Hadi et al., 2020), e.g., do high-quality forecasts mean improved decision? Traditionally, a skillful forecast is vital for the reliability of the forecasts and is essential to promote the use of forecasts in real-world applications by decision-makers. In fact, forecast value is expected to increase with forecast quality, but it may also vary based on other factors such as reservoir capacity and operating objectives (Anghileri et al., 2016). Some studies even have disproved the intuitive assumption that higher forecast performance always leads to

better operation decisions, for example, in agricultural water management (Chiew et al., 2003) and water resources allocation (Turner et al., 2017). Therefore, when forecasts are used to support reservoir operation, they should be assessed in which conditions they can help make better decisions. Moreover, forecast uncertainty and error generally grow up with the increase of the forecast horizon (Maurer and Lettenmaier, 2004; Denaro et al., 2017; Zhao et al., 2019). A decision-maker may doubt whether longer forecast lead times provide sufficient information for a decision purpose or not. There is often a mismatch between the information needed for reservoir operations and the skillful lead time of the reservoir inflow forecast (Anghileri et al., 2016). It is crucial to demonstrate the applicability and effectiveness of the forecast horizon in a forecast-based reservoir operation system (Xu et al., 2014). Overall, there is a continuous need for in-depth study to conduct posterior evaluations of forecasts with different forecast lead times and obtain the efficient forecast horizon for water allocation.

A decision-maker must allocate limited water to different water use sectors considering the conflicting objectives (e.g., benefits and costs) and multiple uncertainties (e.g., forecast uncertainty) in a forecast-based reservoir operation system. Multi-objective programming (MOP) is a valuable tool for helping decision-makers facilitate decision-making with multiple conflicting objectives (Fang et al., 2018b; Guo et al., 2020c), which can offer feasible methods for generating compromise decision alternatives. Some MOP approaches have been widely developed to tackle the uncertainty associated with the decision making processes, such as multi-objective fuzzy programming (Zimmermann, 1978; Pishvaee and Razmi, 2012; Ren et al., 2017) and multi-objective stochastic programming (Xu et al., 2014; Xu et al., 2020; Zhang et al., 2020). These approaches generally convert the multi-objective functions into a single-objective deterministic problem through a fuzzy programming method or a constraint operator. They can effectively deal with the uncertainties between objectives and/or constraints by integrating the decision-makers aspiration levels. However, they may encounter difficulties due to the need for pre-determined individual preferences or reasonable bounds for all objectives. In comparison, multi-objective robust decision making (MORDM) is an effective way to handle such difficulties (Kasprzyk et al., 2013; Yan et al., 2017). It can generate many alternative solutions (Pareto solutions) that do not require assumptions about decision makers' preferences and enhance the robustness of the optimization process. Besides, MORDM, by parameterizing the decision space, can avoid the curse of dimensionality in some MOP approaches, and simplify computational complexity and reduce the running time (Giuliani et al., 2016; Salazar et al., 2017).

In summary, there are still several challenges in forecast-informed reservoir optimization. To address these challenges, the specific research questions of this study are:

(1) Can GRU achieve the same accuracy in the streamflow forecast compared to LSTM with fewer parameters and more straightforward structures?

(2) In which conditions can an improvement in forecast skill be translated into improved reservoir operation optimization?

(3) How do such short-term inflow forecasts with different forecast horizons be used to optimize the multi-reservoir system to impact operation results?

To answer the questions mentioned above, we build an AI-based management framework, which integrates multi-step streamflow forecasts and multi-reservoir operation optimization. We strive to: (1) simulate inflow using LSTM, GRU, and

LSSVM and verify their effectiveness on short-term deterministic streamflow forecasts; (2) generate stochastic inflow scenarios using BMA for refining uncertainty characterization; (3) develop the parameterized MORDM framework for a multi-reservoir operation system and inform decision making by assessing the value, that is, the operation benefits gain or the induced cost of forecasts with a particular lead time. As a case study, including one recipient reservoir storing water from the continental diversion project and 24 supply reservoirs storing water from local rainfall, 25 reservoirs supplying water for four water plants in the Zhoushan Islands, China, are chosen to assess the performance of the AI-based forecast and the forecast-informed operation.

## 2. Methodology

The experimental approach followed in the study is shown in Figure 1 and described in the following sections.

***Figure 1 is here***

### 2.1 Machine learning (ML) methods

This section gives a brief introduction to long short-term memory (LSTM), gated recurrent unit (GRU), and least square support vector machine (LSSVM). In this study, the mapping function between the forecasted streamflow $Q_t$ and hydrological variables $x_t$ can be represented by $f(\cdot)$. In LSTM and GRU, $Q_t = f(x_t, h_{t-1})$ ($h_{t-1}$ denotes the last hidden cell state and the initial state of $h_t$ is $h_0$= 0), while in GWO-LSSVM, $Q_t = f(x_t)$.

#### 2.1.1. Long short-term memory (LSTM)

LSTM network is one of the recurrent neural networks (RNNs) developed by Hochreiter and Schmidhuber (1997), and the basic structure of an LSTM cell is illustrated in Figure 2(a). It is an improved RNN aiming to solve problems such as gradients in long-term memory and backpropagation. The LSTM cell has three gates maintaining and adjusting its cell state and hidden state, including the forget gate, input gate, and output gate. The forget gate determines what information would be thrown away from the cell state. The input gate decides which information is used to update the cell state. The output gate controls which information stored in the current cell state flows into the new hidden state. In Figure 2(a), the state ($c_t$), and the hidden state ($h_t$) of the LSTM cell are updated as follows (Hochreiter and Schmidhuber, 1997):

Forget gate: $$g_t = \sigma(W_f x_t + U_f h_{t-1} + b_f), \quad (1)$$

Input gate: $$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), \quad (2)$$

Potential cell state: $$\tilde{c}_t = tanh(W_{\tilde{c}} x_t + U_{\tilde{c}} h_{t-1} + b_{\tilde{c}}), \quad (3)$$

Cell state: $$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \quad (4)$$

Output state: $o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$, (5)

Hidden state: $h_t = o_t tanh \odot c_t$, (6)

where $g_t$, $c_t$, $o_t$ and $\tilde{c}_t$ represent the forget gate, input gate, output gate, and potential cell state, respectively. ⊙ denotes the element-wise multiplication of vectors, *tanh*(·) is the hyperbolic tangent; $x_t$ represents the current input vector, $h_{t-1}$ denotes the last hidden cell state and the initial state of $h_t$ is $h_0$= 0. $\sigma(\cdot)$ represents the logistic sigmoid function. $[W_f, W_i, W_o, W_{\tilde{c}}]$, $[U_f, U_i, U_o, U_{\tilde{c}}]$, and $[b_f, b_i, b_o, b_{\tilde{c}}]$ represent the input weight matrix, recurrent weight matrix, and bias vectors for the forget, input-output, and potential cell gates, respectively.

### 2.1.2. Gated recurrent unit (GRU)

GRU networks were proposed as a modification of LSTM networks with a more straightforward structure (Cho et al., 2014). The specific structure of the GRU cell is shown in Figure 2(b). Compared with LSTM, GRU has only two control gates, including a reset gate and an update gate. The update gate is applied to control how much information of the previous step is brought into the current step, while the reset gate is used to control the degree of ignoring the information of the previous state. In this way, GRU is superior to LSTM in terms of computer modelling time and parameter updates. In Figure 2(b), the state ($c_t$) and the hidden state ($h_t$) of the GRU cell are updated as follows (Cho et al., 2014):

Reset state: $r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r)$, (7)

Update state: $z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z)$, (8)

Potential cell state: $\tilde{c}_t = tanh(W_{\tilde{c}} x_t + U_{\tilde{c}}(r_t \odot h_{t-1}) + b_{\tilde{c}})$, (9)

Cell state: $c_t = (1 - z_t) \odot c_{t-1} + z_t \odot \tilde{c}_t$, (10)

Hidden state: $h_t = c_t$, (11)

where $r_t$, $z_t$, and $\tilde{c}_t$ represent the reset, update, and potential cell state, respectively. ⊙ denotes the element-wise multiplication of vectors, *tanh*(·) is the hyperbolic tangent; $x_t$ represents the input vectors, $h_{t-1}$ denotes the last hidden cell state and the initial state of $h_t$ is $h_0$= 0. $\sigma(\cdot)$ represents the logistic sigmoid function. $[W_r, W_z, W_{\tilde{c}}]$, $[U_r, U_z, U_{\tilde{c}}]$, and $[b_r, b_z, b_{\tilde{c}}]$ represent the input weight matrix, recurrent weight matrix, and bias vectors for the reset, update, and potential cell gates, respectively.

***Figure 2 is here***

### 2.1.3. Least squares support vector machine with grey wolf optimizer (GWO-LSSVM)

LSSVM is a modified version of SVM, proposed by Suykens and Vandewalle (1999), to reduce the computational time of SVM. SVM uses the quadratic program to formulate the training process of the modeling procedure, while LSSVM aims to employ the least-squares loss functions. The LSSVM non-linear function is expressed as (Suykens et al., 2002):

$$f(x) = w^T \varphi(x) + b, \tag{12}$$

where $\varphi(\cdot)$ is the mapping function that maps the input $x$ into a d-dimensional feature vector, $w$ is a weight vector, and $b$ represents bias. In LSSVM, a minimum objective function is proposed to estimate $\omega$ and $b$ (Suykens et al., 2002).

$$\min J(w,e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{N} e_i^2, \tag{13}$$

that has the following constraints (Suykens et al., 2002):

$$y_i = w^T \varphi(x_i) + b + e_i, \tag{14}$$

where $e$ is the error variable and $\gamma$ is the regulative constant. The objective function can be obtained to solve the optimization problems in Eq. (15) by introducing the Lagrange multipliers $\alpha$ and transferring the constraint problem into an unconstrained one (Suykens et al., 2002):

$$L(w,b,e,\alpha) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N} \alpha_i \left( w^T \varphi(x_i) + b + e_i - y_i \right), \tag{15}$$

By finding the partial derivative of Eq. (16) with respect to $w$, $b$, $\alpha_i$, and $e_i$, the following equation can be derived:

$$y = \sum_{i=1}^{N} \alpha_i \left( \varphi(x)^T \varphi(x_i) \right) + b = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b, \tag{16}$$

where $K(x, x_i)$ is the kernel function. Many kernel functions such as linear, polynomial, radial basis, and sigmoidal have been proposed for LSSVM (Bemani et al., 2020). We adopt the most widely used kernel function, Radial Basis Function (RBF), in this study. The RBF is expressed as:

$$K(x, x_i) = \exp\left( -\|x - x_i\|^2 / 2\sigma^2 \right), \tag{17}$$

where $\sigma^2$ is the kernel parameter. In this study, the parameter $\gamma$ and $\sigma$ were optimized using grey wolf optimizer (GWO). Please see more details on GWO in Guo et al. (2020d).

## 2.2 Bayesian model averaging (BMA)

Generally, it is difficult to determine which model is the best one, leading to model uncertainty. BMA is proposed to solve the uncertainty of the models through averaged estimations from individual models (Liu and Merwade, 2019; Samadi et al., 2020). The weight for each model is based on the simulated decision probability density function, i.e., the posterior probability of the

model. Suppose $Q$ is the unknown quantity we want to predict, given a subset of model forecasts $f=\{f_1, f_2,\ldots, f_K\}$ ($k$=1,2,…,$K$, $K$ is the number of individual model) and the observed data $D$, the posterior distribution of $Q$ can be calculated as (Hoeting et al., 1999):

$$p(Q|D)=\sum_{i=1}^{K} p(f_k|D)\cdot p_k(Q|f_k,D)=\sum_{i=1}^{K} w_k \cdot p_k(Q|f_k,D), \tag{18}$$

where $p_k(Q|f_k,D)$ is the posterior distribution of $Q$ given the model forecast $f_k$ and the observed data $D$, and $p(f_k|D)$ is the posterior probability. In this case, posterior probabilities are the weighting factor for each model, and $\sum_{k=1}^{K} w_k = 1$. The posterior mean (E) and variance (V) of $Q$ are as follows (Hoeting et al., 1999):

$$E[Q|D]=\sum_{k=1}^{K} w_k \cdot E[p_k(Q|f_k,D)]=\sum_{k=1}^{K} w_k f_k\ , \tag{19}$$

$$V[Q|D]=\sum_{k=1}^{K} w_k \cdot\left[f_k-\sum_{k=1}^{K} w_k f_k\right]^2+\sum_{k=1}^{K} w_k \sigma_k^2\ , \tag{20}$$

where $w_k$ and $\sigma_k^2$ are the posterior mean (weight) and variance of $k^{\text{th}}$ forecast model. In this study, a log-likelihood function is maximized to estimate the parameters (weight $w_k$ and variance $\sigma_k^2$) as shown in Eq. (21).

$$l(\theta)=\log\left(\sum_{k=1}^{K}\left(w_k \cdot g(Q|f_k^t,\sigma_k^2)\right)\right), \tag{21}$$

where $\theta$ is the vector of parameters $\{w_k,\sigma_k^2,k=1,2,...,K\}$. $g(Q|f_k^t,\sigma_k^2)$ is Gaussian distribution function where $w_k$ is the weight and $\sigma_k^2$ is the variance.

The Expectation-Maximization (EM) algorithm (Lee et al., 2020) is used to find out the maximum likelihood with a termination criterion (early stopping or a maximal iteration). As the EM proceeds, the parameters of weight $w_k$ and variance $\sigma_k^2$ are updated as follows.

$$w_k^{(Iter)}=\frac{1}{NT}\left(\sum_{t=1}^{NT} z_k^{t^{(Iter)}}\right), \tag{22}$$

$$\sigma_k^{2^{(Iter)}}=\frac{\sum_{t=1}^{NT} z_k^{t^{Iter}}\cdot\left(Y^t-f_k^t\right)^2}{\sum_{t=1}^{NT} z_k^{t^{Iter}}}, \tag{23}$$

$$z_k^{t^{(Iter)}} = \frac{g\left(Q\middle| f_k^t, \sigma_k^{2^{(Iter-1)}}\right)}{\sum_{k=1}^{K} g\left(Q\middle| f_k^t, \sigma_k^{2^{(Iter-1)}}\right)}, \tag{24}$$

$$l(\theta)^{(Iter)} = \sum_{t=1}^{NT} \log\left(\sum_{k=1}^{K}\left(w_k^{(Iter)} \cdot g\left(Q\middle| f_k^t, \sigma_k^{2^{(Iter)}}\right)\right)\right), \tag{25}$$

where *Iter* is the number of iterations. *NT* is the length of calibration periods. $Y^t$ and $f_k^t$ are the observed and forecast streamflow at $t^{th}$ time step, respectively ($m^3/s$), $z_k^{t^{(Iter)}}$ is the latent variable for the $k^{th}$ model at $t^{th}$ time step in the *Iter* iteration. Then we use the Monte Carlo simulation method to generate BMA ensemble forecasts. Assume *M* is the number of Monte Carlo simulations, and the procedure is described below (Zhou et al., 2020a).

(a) Set the initial cumulative weight $w_0^* = 0$ and calculate the cumulative weight $w_k^* = w_{k-1}^* + w_k$ for *k*=1,2,…, *K*. Create a random variable *u* between 0 and 1. If $w_{k-1}^* \le u \le w_k^*$, the $k^{th}$ forecast model would be used as the target forecast.

(b) Generate a realization of the forecasts $Q_t$ using the Gaussian distribution function $g\left(Q_t\middle| f_k^t, \sigma_k^2\right)$. In such a way, there are a set of alternative forecasts to be chosen from as the final forecast.

(c) Repeat Steps (a) & (b) for *M* times and obtain a set of streamflow series $Q_{T\times M} = \begin{bmatrix} Q_{1,1} & Q_{2,1} & \cdots & Q_{T,1} \\ Q_{1,2} & Q_{2,2} & \cdots & Q_{T,2} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{1,M} & Q_{1,M} & \cdots & Q_{T,M} \end{bmatrix}$. Furthermore, 90% confidence intervals between the 5% and 95% quantities are employed to represent the uncertainty of BMA ensemble forecasts.

### 2.3 Forecast performance measures

Three performance indicators are applied to assess the deterministic forecast performance of the three data-process models. They are Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970), root mean square errors (RMSE) (Karunanithi et al., 1994), and mean absolute error (MAE) (Legates and McCabe Jr., 1999). They are expressed as below:

$$NSE = 1 - \frac{\sum_{t=1}^{T}\left(Q_{m,t} - Q_{o,t}\right)^2}{\sum_{t=1}^{T}\left(Q_{o,t} - \bar{Q}_o\right)^2}, \tag{26}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Q_{m,t} - Q_{o,t}\right)^2}, \tag{27}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|Q_{m,t} - Q_{o,t}\right|, \tag{28}$$

where $T$ is the number of samples, $Q_{m,t}$ is the forecasted reservoir inflow ($m^3/s$), $Q_{o,t}$ is the observed inflow ($m^3/s$), and $\bar{Q}_o$ is the average of the observed inflow ($m^3/s$). The NSE can be used to evaluate the stability of the forecasted value. In contrast, RMSE and MAE are used to characterize the overall forecast accuracy. NSE value is (−∞, 1], while MAE and RMSE values are (0, +∞). Generally, models with larger values of NSE or smaller values of RMSE and MAE provide better forecasting accuracy.

In addition, two performance indicators are used to evaluate the performance of ensemble forecast models, i.e., the containing ratio (*CR*), and average deviation amplitude (*D*), which were adopted for assessing the goodness of the prediction bounds (Xiong et al., 2009).

$$CR = \frac{1}{T}\sum_{t=1}^{T} N_t \times 100\% \quad N_t = \begin{cases} 1 & \text{if } \hat{Q}_{l,t} \le Q_{o,t} \le \hat{Q}_{u,t} \\ 0 & \text{else} \end{cases}, \tag{29}$$

$$D = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{1}{2}\left(\hat{Q}_{l,t} + \hat{Q}_{u,t}\right) - Q_{o,t}\right|, \tag{30}$$

where $\hat{Q}_{l,t}$ and $\hat{Q}_{u,t}$ represent the lower and upper prediction bounds of streamflow ($m^3/s$), respectively. Clearly, models with higher *CR* values but lower *D* values would produce better performance.

### 2.4 Parameterized multi-objective robust decision making (MORDM)

This study proposes a parameterized multi-objective robust decision-making approach to design operating policies for the multi-objective reservoir operations by combining direct policy search (DPS) and multi-objective robust decision making (MORDM). In the parameterized MORDM, instead of using the volumes of water to be allocated as the decision variables, we prescribe decisions approximated as non-linear functions conditioned on system state variables (e.g., fore-bay water level observed or predicted inflows, and precipitation) (Giuliani et al., 2016; Quinn et al., 2017b; Salazar et al., 2017). The non-linear functions can be realized by the DPS approach. DPS is based on the parameterization of the operating policy $p_\theta$ and the exploration of the parameter space Θ to find a parameterized policy that optimizes the expected function, i.e.,

$$p_\theta^* = \arg\min_{p_\theta}\left(J_1, J_2, \ldots, J_M\right)_{p_\theta} \quad s.t.\ \theta \in \Theta, \tag{31}$$

where $J_1, J_2, \ldots, J_M$ are the objective functions, $M$ is the number of objectives. $p_\theta^*$ is the corresponding optimal policy with parameters $\theta^*$. Different DPS approaches have been proposed, where two nonlinear approximating networks, namely artificial neural networks (ANNs) and radial basis functions (RBFs), have become widely adopted as universal approximators in many applications (Deisenroth et al., 2013). In particular, we parameterize the operating policy as RBFs because they have been

demonstrated to be effective in solving multi-objective water resources management problems (Giuliani et al., 2014; 2015) and the $k^{th}$ decision variable in the vector $u_t$ (with $k = 1, \ldots, K$) is defined as:

$$u_t^k = \sum_{i=1}^{N} \omega_{i,k} \varphi_{i,k}(\Gamma_t), \tag{32}$$

where $N$ is the number of RBFs $\varphi(\cdot)$, $\Gamma_t$ is the policy input vectors at $t^{th}$ time step including exogenous information (e.g., fore-bay water level observed or predicted inflows and precipitation) and $\omega_{i,k}$ is the weight of the $i^{th}$ RBF, $\sum_{i=1}^{N} \omega_{i,k} = 1 \quad \omega_{i,k} > 0$.

The single RBF is defined as follows:

$$\varphi_{i,k}(\Gamma_t) = \exp\left[ -\sum_{j=1}^{M} \frac{\left[ (\Gamma_t)_j - c_{j,i} \right]^2}{b_{j,i}^2} \right], \tag{33}$$

where $(\Gamma_t)_j$ is the $j^{th}$ policy input at $t^{th}$ time step and $M$ denotes the number of policy input vectors $\Gamma_t$, $j$=1,2,…$M$. $c_i$ and $b_i$ are the M-dimensional center and radius vectors of the $i^{th}$ RBF, respectively. The centers of the RBF must lie within the bounded input space (Yang et al., 2017). The parameter vector $\theta$ is defined as $\theta = \left[ c_{i,j,k}, b_{i,j,k}, \omega_{i,j,k} \right]$ with the number of $\theta$ is $n_\theta = N \times K \times (2 \times M + 1)$. In general, when DPS problems involve multiple objectives, they can be coupled with truly multiobjective optimization methods, such as MOEAs, which allow an approximation of the Pareto front in a single run of the algorithm.

In our study, the parameterized MORDM approach will be coupled with a rolling horizon scheme over one year period to solve the multi-objective reservoir operation problem. Given the lead time of 7 days (forecast horizon is equal to operation horizon) as an example, it is operated following two steps: the optimization model is first operated daily over a 7-day horizon using the parameterized MORDM; after implementing current water allocation decisions, the status, inflow, and other information of reservoirs update as time evolves, and then the remainder is subsequently operated. The two steps are repeated until the process (one year period) is completed. In each operating horizon, the main steps of the parameterized MORDM are described below and presented in Figure 3.

(1) Problem formulation, including the performance measures and constraints.

(2) Generate alternative parameterized policies subject to all the constraints, and the objectives are evaluated over stochastic inflows with the following procedures (Giuliani, et al., 2016):

a) The operating policies are parameterized using RBFs;

b) Run a system simulation from t=1,2,…7d upon each individual parameterized policy $p_\theta$ for each inflow series and obtain the system trajectories;

c) Compute the parameterized policies performance in terms of the operating objectives as a function of system trajectories.

(3) Recompute the parameterized policies performance with robust criteria, for instance, the principle of insufficient reason,

minimax, and minimax regret (Guo et al., 2020b). Among them, the principle of insufficient reason transforming the problem under uncertainty into a decision-making problem under risk has been used in the water resources problems (Giuliani and Castelletti, 2016). The principle of insufficient reason suggests that in the absence of knowledge on the probabilities associated with the different states, the decision could be taken by assigning equal probability to all the states (i.e., $P_j = 1/n$ ). The robust parameterized policies performance can be expressed as:

$$\min \left( \frac{1}{n} \sum_{j=1}^{n} Obj(p_\theta, s_j) \right), \tag{34}$$

where $Obj(p_\theta, s_j)$ is the performance function using parameterized policy $p_\theta$ upon $j^{th}$ streamflow series, $s_j$ denotes the scenario of the $j^{th}$ streamflow series, and n is the number of stochastic streamflow series.

(4) Optimizing the parameterized policies using multi-objective evolutionary algorithms (MOEAs) based on the robust performance objectives. Repeat Steps (2), (3), and **(**4) until the times of population iteration are reached and export the optimal Pareto solutions. In this study, the optimization is solved by applying NSGA-II to search the space of decision variables and identify the trajectories.

It should be noted that the parameterized MORDM in this study aims to solve optimization problems under uncertainty, and thereby one streamflow series need to be repeated multiple times.

***Figure 3 is here***

## 3. Case study

### 3.1 Study area and data

The Zhoushan Islands are located in the northeast of Zhejiang Province, China, with a total area of 22,000 $km^2$ and 1,390 islands (Figure 4). The climate is governed by monsoon-influenced subtropical marine weather systems, and the annual mean temperature and precipitation are 17 °C and 1,300 mm, respectively. There are no large rivers in the islands, and the insufficient freshwater resources severely limit the development of industry and population in the Zhoushan Islands. Recently, a continental diversion project transferring water from Ningbo City to Zhoushan Islands is treated as an effective solution to overcome the water scarcity problem partially. The transferred water is stored in Huangjinwan Reservoir and then operated together with the limited freshwater resources in the remaining 24 reservoirs to supply water to four water plants, i.e., Daobei, Hongqiao, Lincheng, and Pingyangpu. Data for this study includes historical inflow and state of reservoirs, water demand of water plants, and climate forcing data over 2002-2008. The climate data, including daily precipitation and evaporation, are observed at one meteorological station and three rainfall stations. The characteristics of the reservoirs are listed in Table 1.

***Figure 4 is here***

***Table 1 is here***

### 3.2 Problem formulation

Figure 5 shows the simplified schematic diagram of the water supply system in Zhoushan Islands, including reservoirs, pumping stations, pipelines, and water plants. The pipeline arrow indicates the direction of the water flow. It covers the processes associated with water abstraction from resources, water distribution through the network involving the use of pumping stations and pipelines, and main activities relevant to water flow. In this study, water resources include local surface water and imported water. The surface water is the water stored in local reservoirs (a number of 24 reservoirs), while the imported water is the water transferred from Ningbo City (stored in Huangjinwang Reservoir). The imported water is transferred from Ningbo City to Zhoushan Islands through Lixidu and Lanshan pumping stations. End-users within the water supply system are generally divided into the household, industry, agriculture, and environmental use. This study mainly considers household and industry use, which water plants can supply. The agriculture and environmental use are satisfied through operating the reservoir storage above a specific value. That is to say, the main goal of the water allocation plan is to ensure sufficient water flows into the four plants in Zhoushan Islands. They are Daobei, Lincheng, Hongqiao, and Pingyangpu plants, respectively. Releases from the reservoirs (Huangjinwan Reservoir and the remaining 24 local reservoirs) must meet the requirements of water plants. As observed in Figure 5, the reservoirs supplying plants can be divided into two categories. Some reservoirs can directly release water into the plants or reservoirs, including Longtan, Ludong, Shatianao, Nanao, Chenao, Cengang, Tuanjie, and Changchunling reservoirs. In contrast, the other reservoirs can only release water into the plants or reservoirs using pumping stations. In such a way, the pumping flow can be obtained by summing reservoir releases through the corresponding pumping station, using the following equation.

$$Q_{t,j}^{p} = \sum_{n=1}^{N_1} Q_{t,n}^{r} , \tag{35}$$

where $Q_{t,j}^{p}$ denotes the $j^{\text{th}}$ pumping flow at $t^{\text{th}}$ time step in ($m^3/s$), $Q_{t,n}^{r}$ denotes the release of the $n^{\text{th}}$ reservoir at $t^{\text{th}}$ time step in ($m^3/s$), and $N_1$ is the number of reservoirs pumped by the $j^{\text{th}}$ pumping station.

It can be noted in Figure 5 that there are no specific hydraulic connections between most of the reservoirs, while Chahe, Hongwei, Chengbei, and Xiamen reservoirs can release water into Hongqiao Reservoir (the largest reservoir in Zhoushan Islands). With a water plant as a center, the whole islands are divided into four districts, i.e., Daobei, Lincheng, Hongqiao, and Dongbu. The dashed line represents the district boundary. Each district includes a water plant, several pumping stations, and reservoirs to supply water for the water plant. The hydraulic connection between such a water plant and corresponding pumping stations and reservoirs can be expressed as:

$$W_t^{s} = \sum_{j=1}^{J} Q_{t,j}^{p} \Delta t + \sum_{n=1}^{N_2} Q_{t,n}^{r} \Delta t , \tag{36}$$

where $W_t^{s}$ is the amount of water supply for a water plant at $t^{\text{th}}$ time step ($m^3$), $J$ is the number of pumping stations flowing into the water plant, and $N_2$ is the number of reservoirs directly releasing into the water plant.

In Figure 5, every two system elements are connected by the pipelines, e.g., reservoir and reservoir, reservoir and pumping station, and pumping station and water plant. In some cases, more than one reservoir or pumping station share one pipeline, leading to competition on channel flow. However, since the multi-objective optimization problem is operated on a daily time step in our study and we assume that reservoir releases or pumping station flows into the water plant without considering the channel flow limitation, and thereby, regardless of the specific hydrologic connections between channels or pipelines.

***Figure 5 is here***

Three objectives are identified to evaluate the performance of the strategies. The conflicting objectives are to minimize the water deficiency ratio of the Daobei Plant, minimize the water deficiency ratio of the remaining three plants (Hongqiao, Lincheng, and Pingyangpu), and maximize the net benefits. The three plants can feed each other and thus are considered together in our study. A decision-maker would consider a different suite of costs depending on whether an existing system is being managed or a completely new system is being designed. As water supply occurs in an existing system, costs considered in this study are the operating costs. These objective functions are given as follows:

$$\text{Min} \quad Obj_1(x) = \left( \sum_{t=1}^{T} W_t^{n,db} - \sum_{t=1}^{T} W_t^{s,db} \right) \Big/ \sum_{t=1}^{T} W_t^{n,db} \times 100\%, \tag{37}$$

$$\text{Min} \quad Obj_2(x) = \left( \sum_{k=1}^{3} \sum_{t=1}^{T} W_{t,k}^{n,th} - \sum_{k=1}^{3} \sum_{t=1}^{T} W_{t,k}^{s,th} \right) \Big/ \sum_{i=1}^{3} \sum_{t=1}^{T} W_{t,k}^{n,th} \times 100\%, \tag{38}$$

$$\text{Min} \quad Obj_3(x) = \left( M_c^{lw} + M_c^{iw} \right) - M_r, \tag{39}$$

where $Obj_1$ and $Obj_2$ are the water deficiency ratio of Daobei Plant and the sum of the remaining three plants, respectively (%); $Obj_3$ is the net operating costs (RMB); $W_t^{s,db}$ and $W_t^{n,db}$ are the amount of water supply and demand for Daobei Plant at $t^{th}$ time step, respectively ($m^3$); $W_{t,k}^{n,th}$ and $W_{t,k}^{n,th}$ are the amount of water supply and demand for the $k^{th}$ plant (one of the remaining three plants) at $t^{th}$ time step, respectively ($m^3$); $M_c^{lw}$ and $M_c^{iw}$ are the operating costs for water supply using local reservoir water and imported water, respectively (RMB); $M_r$ is the revenue (RMB). The costs and revenue can be obtained according to:

1) Operating costs for water supply using local reservoir water ( $M_c^{lw}$ , RMB)

$$M_c^{lw} = M_{c,1}^{lw} + M_{c,2}^{lw} + M_{c,3}^{lw}, \tag{40}$$

$$M_{c,1}^{lw} = c_1^{lw} \times \sum_{t=1}^{T} W_t^{s,lw}, \tag{41}$$

$$M_{c,2}^{lw} = c_2^{lw} \times \sum_{t=1}^{T} W_t^{s,lw}, \tag{42}$$

$$M_{c,3}^{lw} = c_3^{lw} \times \sum_{j=1}^{J} \sum_{t=1}^{T} \frac{Q_{t,j}^{p,lw} \times P_j^{lw}}{Q_{j,\max}^{p,lw}}, \tag{43}$$

where $M_{c,1}^{lw}$, $M_{c,2}^{lw}$, and $M_{c,3}^{lw}$ represent the water resource fees paid to the government, water fees paid to reservoir managers, and the electricity fees in Zhoushan Islands, respectively (RMB); $c_1^{lw}$, $c_2^{lw}$, and $c_3^{lw}$ denote the constant vectors, representing the unit price of water resources, water, and electricity in Zhoushan Islands, respectively (RMB/m$^3$). $\Delta t$ is the time step; *i* is the index of a reservoir; *j* is the index of a pumping station; *I* denotes the number of reservoirs; *J* denotes the number of pumping stations in Zhoushan Islands; $W_t^{s,lw}$ denotes the amount of water supply for plants using local reservoir water at $t^{\text{th}}$ time step (m$^3$); $P_j^{lw}$ denotes the supporting motor power of the $j^{\text{th}}$ pumping station (Kw); $Q_{t,j}^{p,lw}$ denotes the flow through the $j^{\text{th}}$ pumping station at $t^{\text{th}}$ time step in (m$^3$/s); $Q_{j,\max}^{p,lw}$ denotes the upper flow boundary of the $j^{\text{th}}$ pumping station in Zhoushan Islands (m$^3$/s).

2) Operating costs for water supply using imported water ($M_c^{iw}$, RMB)

$$M_c^{iw} = M_{c,1}^{iw} + M_{c,2}^{iw} + M_{c,3}^{iw}, \tag{44}$$

$$M_{c,1}^{iw} = c_1^{iw} \times \sum_{t=1}^{T} W_t^{s,iw}, \tag{45}$$

$$M_{c,2}^{iw} = c_2^{iw} \times \sum_{t=1}^{T} W_t^{s,iw}, \tag{46}$$

$$M_{c,3}^{iw} = c_3^{iw} \times \sum_{j=1}^{J} \sum_{t=1}^{T} \frac{L_j + Q_{t,j}^{p,iw}}{Q_{j,\max}^{p,iw}}, \tag{47}$$

where $M_{c,1}^{iw}$, $M_{c,2}^{iw}$, and $M_{c,3}^{iw}$ represent the water resources fees paid to the government, water fees paid to the river managers, and electricity fees in Ningbo City, respectively (RMB); $c_1^{iw}$, $c_2^{iw}$, and $c_3^{iw}$ denote the constant vectors, representing the unit price of water resources, water, and electricity in Ningbo City, respectively (RMB/m$^3$); $W_t^{s,iw}$ denotes the amount of water supply for plants using imported water at $t^{\text{th}}$ time step (m$^3$); $Q_{t,j}^{p,iw}$ denotes the flow through the $j^{\text{th}}$ pumping station at $t^{\text{th}}$ time step, *J* is the number of pumping stations transferring water from Ningbo City, *J*=2, $Q_{t,1}^{p,iw}$=$Q_{t,2}^{p,iw}$. $L_j$ denotes the length of the continental diversion pipeline using $j^{\text{th}}$ pumping station (m) and $Q_{i,\max}^{p,iw}$ denotes the upper flow boundary of the $i^{\text{th}}$ pumping station for water transfer (m$^3$/s).

3) Revenues ($M_r$, RMB)

$$M_r = b \times \left( \sum_{t=1}^{T} W_t^{s,db} + W_t^{s,th} \right), \tag{48}$$

where *b* denotes the unit price of water supply revenue (RMB/m$^3$).

The optimization model is subject to the following constraints:

(1) Water balance: $$V_{t+1,i} = V_{t,i} + \left(I_{t,i} - Q^r_{t,i}\right)\Delta t\,, \tag{49}$$

(2) Reservoir storage limits: $$V_{t,i,\min} \le V_{t,i} \le V_{t,i,\max}\,, \tag{50}$$

(3) Reservoir release limits: $$Q^r_{t,i} \le Q^r_{t,i,\max}\,, \tag{51}$$

(4) Pumping station limits: $$Q^p_{t,j} \le Q^p_{\max,j}\,, \tag{52}$$

where $I_{t,i}$ is the inflow of the $i^{\text{th}}$ reservoir at $t^{\text{th}}$ time step ($m^3/s$); $V_{t,i}$ is the storage of $i^{\text{th}}$ reservoir at $t^{\text{th}}$ time step ($m^3$); $V_{\min}$ and $V_{\max}$ are the lower and upper storage boundaries, respectively ($m^3$); $Q^r_{t,i,\max}$ is the maximum release of the $i^{\text{th}}$ reservoir at $t^{\text{th}}$ time step ($m^3/s$). In some cases, $Q^p_{t,j}$ obtained by the RBF policies can be greater than $Q^p_{j,\max}$, and we will do the following step to modify $Q^r_{t,n}$.

$$Q^r_{t,n}{}' = \frac{Q^r_{t,n}}{\sum_{n}^{N_j} Q^r_{t,n}} \times Q^p_{j,\max} \tag{53}$$

### 3.3 Model development

In this study, five input combination scenarios are considered to investigate whether the use of data-driven methods with climate forcing is efficient in inflow forecasts or not. These scenarios are described in Table 2. $P_a$ represents antecedent precipitation, $E_a$ represents antecedent evaporation, $Q_a$ represents antecedent streamflow, $P_f$ represents forecast precipitation, and $E_f$ represents forecast evaporation.

***Table 2 is here***

Several strategies have been proposed in the literature to tackle a multi-step-ahead forecast task (Kline, 2004), such as the recursive, direct, combination of direct and recursive strategies. In this study, we chose one of the most carried out strategies, i.e., the direct strategy (Ben Taieb et al., 2012), to forecast multi-step streamflow over the short-term horizon (1-7 days). In this case, the streamflow is forecasted using the following equations, given S3 as an example.

$$\begin{aligned}
&\text{1d: } Q^f_{t+1} = f\left(Q_t, Q_{t-1}, ..., Q_{t-k}, E_t, E_{t-1}, ..., E_{t-k}, P_{t-1}, ..., P_{t-k}\right)\\
&\text{2d: } Q^f_{t+2} = f\left(Q_t, Q_{t-1}, ..., Q_{t-k}, E_t, E_{t-1}, ..., E_{t-k}, P_{t-1}, ..., P_{t-k}\right)\\
&...\\
&\text{7d:: } Q^f_{t+7} = f\left(Q_t, Q_{t-1}, ..., Q_{t-k}, E_t, E_{t-1}, ..., E_{t-k}, P_{t-1}, ..., P_{t-k}\right)
\end{aligned}, \tag{54}$$

where $f(\cdot)$ is the mapping function between inputs and outputs, which can be modelled by LSTM, GRU, and GWO-LSSVM in our case. The hydrological variables normalized to the same scale of [0, 1] are used as the inputs in the three ML methods. The normalization equation is given as follows:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{55}$$

where $x$ and $x'$ are the original and normalized values, respectively. $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the original series, respectively.

An issue with the ML methods is that they can easily overfit training data. To avoid this, the entire data is divided into three subsets in RNNs: (i) a training set, which is used to compute the gradient and update the weights and biases of the network, (ii) a validation set over which the errors are monitored during the training process and is used to decide when to stop training, (iii) a test set, which is used to assess the expected performance in the future. In addition, dropout is a regularization method where input and recurrent connections to LSTM and GRU units are probabilistically excluded from activation and weight updates while training a network. The strategies mentioned above have the effect of reducing overfitting and improving model performance in RNNs. Both LSTM and GRU are trained based on truncated Back Propagation Through Time (BPTT) (Cheng et al., 2020), which uses a back propagation network to update the parameters in iterations. The NSE function is used as the loss function to calibrate the LSTM and GRU models. As for LSSVM, we avoid overfitting by minimizing the NSE during the calibration and validation periods, while the test period is also used to assess the performance. In this study, Jan 2002 to Dec 2006 is used as the training period, while the validation and tests extend from Jan 2007 – Dec 2007 and Jan 2008 - Dec 2008, respectively.

The multi-reservoir operation optimization using inflow forecasts is performed over one year (April 1st, 2007- March 31st, 2008) with 25 reservoirs. The period is selected to ensure that it does not cover the calibration datasets. For the short-term forecasting and reservoir operation purpose, a forecast horizon of 1-7 days ahead is chosen. In this study, we use the parametrized MORDM approach to design operating policies for the multi-objective reservoir operations under uncertainty. The optimized operations are regulated based on both deterministic and uncertain forecast inflow. To keep fair, we perform a simulation to generate deterministic and observed ensemble forecasts that each deterministic forecast and observed data are repeated 900 times, respectively. Using the uncertain streamflow forecasts (BMA, deterministic or observed ensemble forecasts) as policy inputs in the parametrized MORDM method, we can generate alternative RBF policies subjecting to all the constraints, and the objectives are evaluated over stochastic inflows. Under the parameterized MORDM, the decision variables in the optimization problem are not the volumes of water to be transferred from Ningbo City and the remaining 24 reservoirs each day. Instead, the decision variables are the parameters of the RBF policies. The best operation is obtained by conditioning the operating policies upon the following two input variables, e.g., the initial fore-bay water level and current inflow of reservoir. The optimization is solved at each time step (a particular forecast horizon, e.g., 1-7 days) by applying NSGA-II to search the space of decision variables and identify the islands' water allocation trajectories.

### 3.4. Results and discussion

#### 3.4.1 Multi-step deterministic forecasts based on ML methods

We consider five different input scenarios described in Section 3.3. Table 3 demonstrates the forecast analysis carried out with the different configurations (input combination and forecast model), tabulating the NSE ranges for lead times from 1 day-ahead to 7 day-ahead over all reservoirs during the calibration, validation, and test periods. It can be seen that S1 using only the flow variables and S2 using only the antecedent climate variables are inferior to the other scenarios. The performance is generally improved when the flow variables are used in combination with the antecedent precipitation and evaporation under S3. However, in this case, the antecedent variables succeed to forecast only at 1-day ahead. The forecast performance decreases significantly as the forecast horizon increases from 1-day to7-day ahead. Herein, we suppose that the following precipitation and evaporation have been forecasted. It is clear that S4 and S5, with the forecast climate variables, make significant increments in streamflow forecasting. The NSE can remain relatively stable at different horizons. There are no apparent differences between the three forecast models during the calibration period. However, the two RNNs perform better than GWO-LSSVM during the validation period, while GWO-LSSVM outperforms during the test periods. Besides, given that GRU has more superficial structures and fewer parameters and requires less time for model training, it may be the preferred method for short-term streamflow forecast compared with LSTM. The same results have been obtained in Gao et al. (2020) when they used LSTM and GRU to model short-term rainfall-runoff relationships.

***Table 3 is here***

We aim to compare how the forecasted climate variables impact the streamflow forecast and reservoir operation performance. For the sake of brevity, S3 and S5 are compared in detail in the following section. Recall that S3 uses flow variables, antecedent precipitation, and evaporation as inputs, while S5 uses flow variables as well as the antecedent and forecast climate forcing. After assessing model validity, the next step is to compare the performance across the 24 reservoirs. The coefficient of variation (COV), defined as the ratio of the standard deviation of the inflow time series, is used to capture the varying characteristics of the incoming flow into the reservoir. Figure 6 reveals a strong negative relationship between COV and forecast performance under S3 at all lead times. The forecast performance decreases as the COV increases for all forecast models. This indicates that the more variation the flow has, the harder it is for data-driven methods to learn the flow pattern when there doesn't exist enough input information. However, the negative signal under S5 (Figure 7) with forecasted climate variables (precipitation and evaporation in this study) is not as strong as that under S3, indicating that the forecast climate variables can help AI-based models mapping functions between inputs and outputs. The improvements are more significant for the two RNN models, i.e., LSTM and GRU, than LSSVM. This result demonstrates that the efficiency of deep-learning RNN methods is better and more accurate than LSSVM.

***Figure 6 is here***

***Figure 7 is here***

### 3.4.2 Multi-step stochastic forecasts based on BMA method

Based on the forecast results of three data-driven models in the calibration period, the BMA method determines weights for LSTM, GRU, and GWO-LSSVM models. The weights reflecting the performance of the ensemble models during the calibration period are shown only for lead times of 1 and 7 days for the sake of brevity under S3 and S5 in Figure 8. The model weights reflect the comparative importance of all the competitive modelling predictions on one level. Figure 8 indicates that it is difficult to conclude which individual model provides the best prediction. For example, GRU outperforms the remaining two models for Hongqiao Reservoir, while LSTM performs best for Cenggang Reservoir in Figure 8(a1). Similar results can be obtained from Figure 8(b1). Comparatively, Figure 8(a2) shows that LSTM and GWO-LSSVM influence the BMA model more than GRU. This higher weight is assigned because the forecasts are more similar to observations than those less similar to observations using the BMA posterior processor. However, observed from Figure 8(b2), the prediction accuracy of GWO-LSSVM is seriously affected and much less than that of GRU. It is consistent with the results obtained in Figure 7, indicating that RNNs outperform GWO-LSSVM when there exists more input information under S5. Overall, model uncertainty always exists whether forecast climate variables are involved or not, and it is necessary to analyze and evaluate the model uncertainty involved using BMA.

***Figure 8 is here***

To access model validity, the evaluation of the modelled streamflow is performed over calibration, validation, and test periods using NSE, RMSE, and MAE metrics. Table 4 shows the performance metric ranges for all 24 reservoirs of BMA methods under S3 and S5. Apparently, both the replicative (forecast performance in calibration sets) and predictive (forecast performance in validation and test sets) validity under S5 for forecast horizons are significantly better than those under S3. For example, Figure 9 demonstrates the improvement rates in terms of NSE, RMSE, and MAE of the BMA model compared with the three individual models. BMA produces the maximum NSE, minimum RMSE, and minimum MAE during the calibration period for both two scenarios, indicating that BMA has the best goodness-of-fit. This is because the weights are derived according to the individual forecast model in this period. With respect to validation and test periods, the BMA method shows better forecasts than the three comparative models except for the GRU modelling validation datasets under S5. Thus, it is shown that the BMA model well matches the actual streamflow.

***Table 4 is here***

***Figure 9 is here***

The model validity is then assessed using (i) hydrographs and (ii) scatter plots of observed and modelled streamflow, as shown in Figure 10 and Figure 11. Herein, we only show three reservoirs, i.e., Hongqiao (the largest reservoir), Goushan (the medium reservoir), and Nanao (the smallest reservoir), for the sake of brevity. From Figure 10, it is clearly shown that the modelled streamflow deviates gradually from the 1:1 line and the forecast skill decreases with the increase of lead time under S3 as expected, which is consistent with the statistical results shown in Table 4. In contrast, the scatters of the observed and modelled streamflow implemented with forecasted climate variables fit well across the 1:1 line at different lead times under S5, observed

from Figure 11. The performance for Hongqiao Reservoir is affected explicitly by an extreme peak event that hit the reservoir during the calibration period shown in Figure 10, which does not occur over the training set of data. This causes heavy underestimations in the streamflow forecast. A more extended calibration period is required to improve the performance over such extreme peak flow events. However, the BMA method performs well on this extreme peak flow in Hongqiao Reservoir at all lead times when the forecast climate forcing is applied as inputs. This is because the reservoirs in Zhoushan Islands have relatively small drainage areas, and thus the flow concentrates in a very short time after an extreme rain event.

***Figure 10 is here***

***Figure 11 is here***

We use the Monte Carlo simulation method to generate BMA ensemble forecasts. The number of simulations is set as 1000 in this study. To demonstrate the optimization results of multi-reservoir operations based on the data-driven forecast models under uncertainty, 90% confidence intervals associated with the deterministic predictions at BMA are further calculated. The confidence interval provides more alternatives that are possibly useful to a tradeoff between multiple objectives, such as flood control, hydropower generation, and improved navigation (Zhang et al., 2015). The interval performance metrics of *Cr* and *D* described in Section 2.3 are adopted to assess the performance of uncertain forecasts. Table 5 displays the averaged metrics for all the 24 reservoirs under S3 and S5. Both indicators under S5 are superior to those under S3. The 90% streamflow interval between the fifth and ninety-fifth percentiles of some representative reservoirs, e.g., Hongqiao, Goushan, and Nanao reservoirs, are presented in Figure 12 and Figure 13. The results are consistent with those in Figure 10 and Figure 11. It is observed from Figure 12 that the streamflow interval fails to capture the extreme peak flow for Hongqiao Reservoir under S3. The BMA performs gradually worse with increasing lead times for the three reservoirs. However, in Figure 13, the red dots represent the observed streamflow, most of which are covered by the 90% interval at both 1-day ahead and 7-days ahead. Therefore, the forecast climate variables will be conducive to reduce the predictive uncertainty of real-time streamflow forecasting.

***Table 5 is here***

***Figure 12 is here***

***Figure 13 is here***

### 3.4.3 Multi-objective reservoir operation performance evaluation

The optimized operations are regulated based on both deterministic and uncertain forecast inflow. To demonstrate the relationship between the conflicting objectives, a set of Pareto solutions over a 7-day horizon at different periods under S5 is given as an example, as shown in Figure 14. The optimization using the Pareto concept allows the operator to choose an appropriate solution depending on the prevailing circumstances and analyzing the tradeoff between the conflicting objectives. In each of the plots, the water deficiency ratio of Daobei Plant and the sum of the remaining plants are plotted on the *x* and *y* axes, respectively. The color of the markers indicates the net operating costs with colors ranging from red, representing low value, to blue, representing high value. Thus, an ideal solution should be located at the left corner (low value of the water

deficiency ratio of Daobei Plant and the sum of the remaining three plants) of the plot and represented by a red (low net operating costs) marker. The black arrows have been added in the figure to guide the reader in understanding the directions of optimization. Generally, the water deficiency ratio of Daobei Plant has an inverse relationship with that of the sum of the remaining plants (inverse relationship, i.e., the former decrease with the increase of the latter). In contrast, the water deficiency ratio of the remaining three plants has a positive relationship with the net costs (positive relationship, i.e., the former increase with the increase of the latter).

It is interesting to compare the performances associated with deterministic and uncertain forecasts. Uncertain conditions (Figure 14(b)) show a much broader scale on the three objectives than deterministic conditions (Figure 14(a)). For instance, uncertain forecasts produce the water deficiency ratio of Daobei Plant, ranging from -40% to 80%, during 2007 August 12th - 2007 August 18st, while deterministic forecasts have a smaller range with a value from 30% to 100%. The water supply deficits under deterministic forecasts are due to the high demand happening in August, which can be mitigated when informing the reservoir operations with uncertain forecasts. In this way, we expect that if the ensemble streamflow forecasts are used in a stochastic optimization scheme, the reservoir operation could be further enhanced because the optimization considers possible uncertainty provided by uncertain forecasts and thus takes advantage of correcting the influences of uncertainty.

***Figure 14 is here***

**(1) Performance evaluation with different forecast skills**

In general, forecasts are always useful for reservoir operations. The annual revenues, costs, and water supply reliability, are chosen as metrics to compare the performance of the operating policies derived from different configurations. Reliability is a measure of how well the water demand for users is met in a water transfer system. In this case, reliability is expressed as a percentage. The system performances are averaged over a set of solutions. The annual values during the period from 2007 April 1$^{st}$ to 2008 March 31$^{st}$ at various configurations are provided in Table 6 with two decision horizons of 1 and 7 days. The multi-reservoir operation based on observation is designed as a benchmark. It can be seen from Table 6 that the performance indicators from the 1-day forecast horizon are better than those from 7-day using deterministic inflows (in the case of observed and forecasted inflows). Two scenarios (S3 and S5) with the 1-day forecast horizon show similar operating performance, which is consistent with the performance of the inflow forecast listed in Table 3. Recall again that S3 uses flow variables, antecedent precipitation, and evaporation as forecast inputs, while S5 uses flow variables as well as the antecedent and forecast climate forcing. In contrast to S3, the operating results of S5 with a 7-day forecast horizon are closest to that of the observation. This is due to the improved inflow forecast performance under S5. However, it is depicted in Table 6 that the indicator of water supply reliability and net costs under S5 are inferior to those under S3. As for the stochastic forecasts, S5 outperforms S3 with lower net costs and approximate water supply reliability. In this case, the improved performance may not lead to improved decisions in deterministic forecasts.

***Table 6 is here***

The results obtained in Table 6 show that system performance derived from the observed inflows is inferior to that from other configurations. This finding cannot confirm the effectiveness of inflow forecasts. The reason for that is the forecast inflows may overestimate the actual inflows. For example, the mean value (0.14 $m^3/s$) of the observed inflow of Hongqiao Reservoir is lower than that of the forecasted inflow (0.17 $m^3/s$). In this case, the good performance presented in Table 6 is 'fake'. That is to say, although decision-makers can follow the strategies determined by the forecasted inflows, the system performance should be assessed using the actual inflows (i.e., observed inflows). We further re-evaluate the operating strategies optimized from different configurations mentioned above using the observed inflows. The performance metrics are listed in Table 7. It is expected that the results can reveal the maximum efficiency and reliability that could be achieved based on accurate information. In general, the indicator values under deterministic forecasts in Table 7 are reduced compared with those in Table 6. The reason is that reservoir operating decisions in Table 6 are optimized based on a higher inflow series.

***Table 7 is here***

In terms of both deterministic and uncertain forecasts, net operating costs of S5 are improved significantly compared with that of S3, while water supply reliability is increased slightly. This result may suggest that improved forecasts are more skillful in making decisions when using forecast climate variables as inputs. We highlight that this result we obtained is specific for Zhoushan Islands. Indeed, many studies show that higher forecast performance did not lead to better operation decisions (Chiew et al., 2003; Goddard et al., 2010; Turner et al., 2017). However, some researchers draw the same conclusions as us. For instance, Anghileri et al. (2016) declared that inflow forecasts with accurate weather components would produce much smaller water supply deficits. Moreover, Anghileri et al. (2019) found that preprocessed forecasts (higher performance) were more valuable than the raw forecasts (less performance) regarding to two operation performance metrics, i.e., mean annual revenues and spilled water volume.

There is also an interesting finding that the operating performance upon deterministic forecasts deteriorates, while the performance upon uncertain forecasts can keep relatively stable. This implies that the use of uncertain forecasts in reservoir operation can be more efficient and reliable than that of deterministic forecasts. The reason is that in a stochastic optimization scheme, the value can be further enhanced because the optimization can account for the total uncertainty provided by the ensemble forecasts. Similar results were obtained by Roulston and Smith (2003), who reported that the hydroelectric power production derived from the ensemble forecasts was increased compared with the deterministic forecasts. Boucher et al. (2012) also found that stochastic forecasts outperformed deterministic ones with the lower turbinate flow, higher generation production, and less spillage during a flood period. Overall, in most cases, a noticeable improvement can be achieved through the use of the stochastic decision-making assistance tool.

We then assess the performance metrics of water supply reliability over different seasons. It is noted in Figure 15 that the deterministic forecasts are less skillful than the uncertain forecast when used in spring (JFM), summer (AMJ), autumn (JAS), and winter (OND) with the two forecast horizons. Although the operating performance using the deterministic forecast is lower due to its deterministic character, the main characteristics of the relationship between forecast quality and value remain

unchanged. That is to say, the benefits of considering the forecasts are more significant when the forecast quality is higher. It indicates that the optimization is capable of exploiting efficient information to improve reservoir operations. In our multi-objective optimization modelling, we would like to make the best use of water resources and maximize water supply. However, the operating performance in autumn shows a lower value with respect to that in other seasons. This is because the water demand in autumn is usually much higher. The shortage does not imply the non-effectiveness of our proposed forecast-based management framework but is due to the limitation of available water and pies capacity.

***Figure 15 is here***

**(2) Performance evaluation with different forecast horizons**

The impact of different forecast horizons on the operation performance is further evaluated under different configurations, as shown in Figure 16. It is noted that the operating policy optimized from uncertain forecast inflows upon S5 outperforms that from S3. In terms of deterministic conditions, S5 improves the operation on the metrics of water supply reliability of Daobei Plant, water supply reliability of the other plants, and net costs with a variation of 2.11~13.58%, 2.74~7.38%, and -19.94~-10.30%, respectively, compared with S3. As for uncertain conditions, S5 improves by 0.24~1.90%, 0.06~1.32%, and -59.45~-176.19%, respectively. Although the increments in water supply reliability are not insignificant, S5 can secure water demand with much less operating costs than S3, which decision-makers value most. Furthermore, uncertain forecasts produce an improved ratio of 31.52~65.01%, 19.98~46.60%, -116.45~-56.95% than deterministic forecasts regarding to the three metrics, respectively. Our results again highlight that uncertain forecasts are more valuable than deterministic forecasts when designing forecast-informed reservoir operations.

***Figure 16 is here***

With an increase in forecast horizon from 1 to 7 days, the performance in water supply reliability and net operating costs upon deterministic conditions are generally reduced. This suggests that considering a longer forecast horizon (up to 7 days) does not necessarily improve reservoir operation without future forecast climate variables as inputs (low forecast quality). The reduced performance in water supply reliability might be due to the fact that the optimization explores strategies to secure the whole water demand in a longer horizon, which results in reliability sacrifice on some particular days. This result is similar to the finding proposed in Xu et al. (2014), who argued that the use of longer-horizon (an efficient forecast horizon longer than one day) inflows could not improve hydropower performance when they set the forecast horizon as one to five days. Nevertheless, the increasing forecast horizon may not generate improved or decreased water supply reliability in uncertain conditions. Approximate water supply volume can lead to similar revenues or fees paid to the government and managers (water fees and water resources fees). Accordingly, the growing trend in net costs is caused by the increased operating costs, mainly dominated by electricity prices, when the multi-reservoir is operated to supply the demand in a longer horizon. In this case, the operation performance varies at different conditions. This demonstrates that the relationship between forecast horizon and reservoir operation is rather complex and depends not only on the configurations (i.e., inflow forecast quality and uncertainty) used to determine operating rules but also on the performance metrics used to assess operation.

## 4 Limitations and future work

Our work suffers from some limitations which could be overcome in future studies. One of the limitations is that only one single indicator was used to calibrate the forecast models, while multiple indicators were used in assessing the performance of the models. It should be a more fair practice by using multi-criteria to do both calibration and assessment and can be interesting for future work. Another limitation is that we used the average observed price to calculate the revenues and operating costs. In an operational and deregulated market setting, the prices may fluctuate significantly (Anghileri et al., 2019). For instance, forecasting electricity prices is likely to improve short-term operation efficiency significantly. The combined effects of price and streamflow forecasts on water resource allocation are worth investigating in future studies. Our study also suffers from the drawback that instead of using the short-term weather forecasts from the Global Forecast System (GFS) or European Centre for Medium‐Range Weather Forecasts (ECMWF) model (Choong and El-Shafie, 2015; Schwanenberg et al., 2015; Peng et al., 2018; Ahmad and Hossain, 2019; Liu et al., 2019), we used the observed weather conditions as alternatives, which may result in an overestimation in forecast quality. However, forecast uncertainty and error that generally grow up with lead time. The usefulness of the forecast information can be reduced with the increase of the forecast horizon and thus the operating performance. This may influence the finding we highlight above that the relationship between forecast horizon and reservoir operation is not constant and specific. It would be interesting to analyze the reservoir operations performance when accounting for an ensemble numerical weather prediction.

## 5. Conclusions

In this study, we proposed an AI-based management methodology to assess forecast quality and forecast-informed reservoir operation performance together. The approach was tested on a water resources allocation system in Zhoushan Islands, China. Specifically, the findings are summarized below.

A data-driven reservoir inflow forecasting system using ML methods (LSTM, GRU, and GWO-LSSVM) was first developed with a comprehensive calibration-validation-testing framework. The validity of the deterministic forecast was demonstrated by applying it over 25 reservoirs with varying climate and hydrological characteristics. Results showed that the more variation the streamflow has (a high COV value), the harder it was for the ML methods to learn the flow pattern when there didn't exist enough input information. The forecast skill deteriorated with increasing lead times under such scenarios. However, short-term forecast climate forcing was efficient and scalable in forecasting the multi-reservoir inflow over the forecast horizon (1-7 days). LSTM and GRU models generated comparable performance under different configurations. Given that GRU has simpler structures and fewer parameters and required less time for modeling, it might be the preferred method for streamflow forecasts than LSTM.

Then we used BMA to generate stochastic inflow scenarios for quantifying uncertainty based on LSTM, GRU, and GWO-LSSVM deterministic forecasts. The results demonstrated that it was difficult to conclude which individual model provided the best prediction, but the BMA did display better forecast skills in comparison to the individual ones. Including one scenario

with antecedent conditions and one scenario with both antecedent and forecast information, two input combination scenarios were compared on the uncertain forecast performance in detail. The comparison indicated that forecast climate variables would help reduce the predictive uncertainty of short-term streamflow forecasting.

The forecasting scheme was further coupled with a multi-objective reservoir operation model to optimize water resources allocation. Using a MORDM approach, we identified strategies that tradeoff between water supply reliability and operating costs in Zhoushan Islands. A rolling horizon scheme was employed to obtain optimal operating policy over the horizon of 1-7 days. The long-term assessment over a year based on deterministic and stochastic forecasts showed quite different performances in terms of water supply reliability and net operating costs. Our averaged annual results showed that uncertain forecasts were more valuable than deterministic forecasts. The operating benefits of considering the forecasts were more significant when the forecast quality was higher. Similar results could be obtained at a seasonal scale. While showing the unquestionable benefit of implementing forecast-based reservoir operations, our results also demonstrated that the relationship between forecast horizon and reservoir operation was complex and depended on the operating configurations (forecast quality and uncertainty) and performance measures for the Zhoushan Islands system.

Overall, the developed AI-based management framework has demonstrated a clear advantage in quantifying the uncertainties of inflow forecasts to improve the overall system performance of water allocation systems. Such a framework can be further applied to other study sites with similar problems. However, the results we obtained in this study are only specific for the Zhoushan Islands and should be exported with care to other study sites.

*Data availability. The data used to support the findings of this study are available from the corresponding author upon request.*

*Author contribution. Yuxue Guo and Yue-Ping Xu designed all the experiments. Hao Chen and Haiting Gu collected and preprocessed the data. Yuxue Guo and Xinting Yu conducted all the experiments and analyzed the results. Yuxue Guo wrote the first draft of the manuscript with contributions from Jingkai Xie. Yue-Ping Xu supervised the study and edited the manuscript.*

*Conflicts of interest. The authors declare that they have no conflict of interest.*

*Acknowledgments. This study is financially supported by the Key Project of Zhejiang Natural Science Foundation (LZ20E090001), Zhejiang Key Research and Development Program (2021C03017), and the Fundamental Research Funds for the Zhejiang Provincial Universities （2021XZZX015). Finally, editors and two reviewers are greatly acknowledged for their constructive comments to improve the quality of this manuscript.*

**References**


Adnan, R. M., Liang, Z., Heddam, S., et al. Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. Journal of Hydrology, 586, 124371, https://doi.org/10.1016/j.jhydrol.2019.124371, 2020.

Ahmad, S. K., Hossain, F. A generic data-driven technique for forecasting of reservoir inflow: Application for hydropower maximization. Environmental Modelling & Software, 119, 147-165, https://doi.org/10.1016/j.envsoft.2019.06.008, 2019.

Alexander, S., Yang, G., Addisu, G., et al. Forecast-informed reservoir operations to guide hydropower and agriculture allocations in the Blue Nile basin, Ethiopia. International Journal of Water Resources Development, 1-26, https://doi.org/10.1080/07900627.2020.1745159, 2020.

Al-Sudani, Z. A., Salih, S. Q., Yaseen, Z. M. Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation. Journal of Hydrology, 573, 1-12, https://doi.org/10.1016/j.jhydrol.2019.03.004, 2019.

Ali, S., Shahbaz, M. Streamflow forecasting by modeling the rainfall–streamflow relationship using artificial neural networks. Modeling Earth Systems and Environment, 6(3), 1645-1656, https://doi.org/10.1007/s40808-020-00780-3, 2020.

Anghileri, D., Monhart, S., Zhou, C., et al. The Value of Subseasonal Hydrometeorological Forecasts to Hydropower Operations: How Much Does Preprocessing Matter? Water Resources Research, 55(12), 10159-10178, https://doi.org/10.1029/2019WR025280, 2019.

Anghileri, D., Voisin, N., Castelletti, A., et al. Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments. Water Resources Research, 52(6), 4209-4225, https://doi.org/10.1002/2015WR017864, 2016.

Ayzel, G., Heistermann, M. The effect of calibration data length on the performance of a conceptual hydrological model versus LSTM and GRU: A case study for six basins from the CAMELS dataset. Computers & Geosciences, 149, 104708, https://doi.org/10.1016/j.cageo.2021.104708, 2021.

Bemani, A., Xiong, Q., Baghban, A., et al. Modeling of cetane number of biodiesel from fatty acid methyl ester (FAME) information using GA-, PSO-, and HGAPSO- LSSVM models. Renewable Energy, 150, 924-934, https://doi.org/10.1016/j.renene.2019.12.086, 2020.

Ben Taieb, S., Bontempi, G., Atiya, A. F., et al. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. Expert Systems with Applications, 39(8), 7067-7083, https://doi.org/10.1016/j.eswa.2012.01.039, 2012.

Boucher, M. A., Tremblay, D., Delorme, L., et al. Hydro-economic assessment of hydrological forecasting systems. Journal of Hydrology, 416, 133-144, https://doi.org/10.1016/j.jhydrol.2011.11.042, 2012.

Cheng, M., Fang, F., Kinouchi, T., et al., 2020. Long lead-time daily and monthly streamflow forecasting using machine learning methods. Journal of Hydrology, 590, 125376.

Chiew, F., Zhou, S., McMahon, T. Use of seasonal streamflow forecasts in water resources management. Journal of Hydrology, 270(1-2), 135-144, https://doi.org/10.1016/S0022-1694(02)00292-5, 2003.

Cho, K., Merrienboer, B. v., Gulcehre, C., et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. Association for Computational Linguistics, 1724-1734, 2014.

Choong, S.-M., El-Shafie, A. State-of-the-Art for Modelling Reservoir Inflows and Management Optimization. Water Resources Management, 29(4), 1267-1282, https://doi.org/10.1007/s11269-014-0872-z, 2015.

Deisenroth, M., Neumann, G., Peters, J. A Survey on Policy Search for Robotics. Foundations and Trends in Robotics, 2(1-2), 1-142, 10.1561/2300000021, 2013.

Denaro, S., Anghileri, D., Giuliani, M., et al. Informing the operations of water reservoirs over multiple temporal scales by direct use of hydro-meteorological data. Advances in water resources, 103, 51-63, https://doi.org/10.1016/j.advwatres.2017.02.012, 2017.

Elman, J. L. Finding Structure in Time. Cognitive Science, 14(2), 179-211, https://doi.org/10.1207/s15516709cog1402_1, 1990.

Fang, G., Guo, Y., Huang, X., et al. Combining Grey Relational Analysis and a Bayesian Model Averaging Method to Derive Monthly Optimal Operating Rules for a Hydropower Reservoir. Water, 10(8), 1099, https://doi.org/10.3390/w10081099, 2018a.

Fang, G., Guo, Y., Wen, X., et al. Multi-Objective Differential Evolution-Chaos Shuffled Frog Leaping Algorithm for Water Resources System Optimization. Water Resources Management, 32(12), 3835-3852, https://doi.org/10.1007/s11269-018-2021-6, 2018b.

Feng, D., Fang, K., Shen, C. Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. Water Resources Research, 56(9), e2019WR026793, https://doi.org/10.1029/2019WR026793, 2020.

Gao, S., Huang, Y., Zhang, S., et al. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. Journal of Hydrology, 589, 125188, https://doi.org/10.1016/j.jhydrol.2020.125188, 2020.

Ghumman, A. R., Ahmad, S., Hashmi, H. N. Performance assessment of artificial neural networks and support vector regression models for stream flow predictions. Environmental monitoring assessment, 190(12), 704, https://doi.org/10.1007/s10661-018-7012-9, 2018.

Gibbs, M. S., McInerney, D., Humphrey, G., et al. State updating and calibration period selection to improve dynamic monthly streamflow forecasts for an environmental flow management application. Hydrology and Earth System Sciences, 22(1), 871-887, https://doi.org/10.5194/hess-22-871-2018, 2018.

Giuliani, M., Castelletti, A. Is robustness really robust? How different definitions of robustness impact decision-making under climate change. Climatic Change, 135(3-4), 409-424, https://doi.org/10.1007/s10584-015-1586-9, 2016.

Giuliani, M., Castelletti, A., Pianosi, F., et al. Curses, tradeoffs, and scalable management: Advancing evolutionary multiobjective direct policy search to improve water reservoir operations. Journal of Water Resources Planning and Management, 142(2), 04015050, https://doi.org/10.5334/jors.293, 2016.

Giuliani, M., Herman, J., Castelletti, A., Reed, P. Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management. Water Resources Research. 50, 3355–3377, http://.doi.org/10.1002/2013WR014700, 2014.

Giuliani, M., Pianosi, F., Castelletti, A. Making the most of data: an information selection and assessment framework to improve water systemsoperations. Water Resources Research, 51(11), 9073–9093, http://doi.org/10.1002/2015WR017044, 2015.

Goddard, L., Aitchellouche, Y., Baethgen, W., et al. Providing Seasonal-to-Interannual Climate Information for Risk Management and Decision-making. Procedia Environmental Sciences, 1, 81-101, https://doi.org/10.1016/j.proenv.2010.09.007, 2010.

Greff, K., Srivastava, R. K., Koutník, J., et al. LSTM: A Search Space Odyssey. IEEE Transactions on Neural Networks and Learning Systems, 28(10), 2222-2232, https://doi.org/10.1109/TNNLS.2016.2582924, 2017.

Guo, Y., Fang, G., Wen, X., et al. Hydrological responses and adaptive potential of cascaded reservoirs under climate change in Yuan River Basin. Hydrology Research, 50(1), 358–378, https://doi.org/10.2166/nh.2018.165, 2018.

Guo, Y., Fang, G., Xu, Y.-P., et al. Identifying how future climate and land use/cover changes impact streamflow in Xinanjiang Basin, East China. Science of the Total Environment, 710, 136275, https://doi.org/10.1016/j.scitotenv.2019.136275, 2020a.

Guo, Y., Fang, G., Xu, Y.-P., et al. Responses of hydropower generation and sustainability to changes in reservoir policy, climate and land use under uncertainty: A case study of Xinanjiang Reservoir in China. Journal of Cleaner Production, 124609, https://doi.org/10.1016/j.jclepro.2020.124609, 2020b.

Guo, Y., Tian, X., Fang, G., et al. Many-objective optimization with improved shuffled frog leaping algorithm for inter-basin water transfers. Advances in Water Resources, 138, 103531, https://doi.org/10.1016/j.advwatres.2020.103531, 2020c.

Guo, Y., Xu, Y.-P., Sun, M., et al. Multi-step-ahead forecast of reservoir water availability with improved quantum-based GWO coupled with the AI-based LSSVM model. Journal of Hydrology, 125769, https://doi.org/10.1016/j.jhydrol.2020.125769, 2020d.

Hadjimichael, A., Gold, D., Hadka, D., et al. Rhodium: Python Library for Many-Objective Robust Decision Making and Exploratory Modeling. Journal of Open Research Software, 8(1), 12, 10.5334/jors.293, 2020.

Hadi, S. J., Tombul, M., Salih, S. Q., et al. The capacity of the hybridizing wavelet transformation approach with data-driven models for modeling monthly-scale streamflow. IEEE Access, 8, 101993-102006, https://doi.org/10.1109/ACCESS.2020.2998437, 2020.

Hochreiter, S., Schmidhuber, J. Long Short-Term Memory. Neural Computation, 9(8), 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Hoeting, J. A., Madigan, D., Raftery, A. E., et al. Bayesian Model Averaging: A Tutorial. Statistical Science, 14(4), 382-417, https://www.jstor.org/stable/2676803, 1999.

Jung, Y., Jung, J., Kim, B., et al. Long short-term memory recurrent neural network for modeling temporal patterns in long-term power forecasting for solar PV facilities: Case study of South Korea. Journal of Cleaner Production, 250, 119476, https://doi.org/10.1016/j.jclepro.2019.119476, 2020.

Karunanithi, N., Grenney, W. J., Whitley, D., et al. Neural networks for river flow prediction. Journal of computing in civil engineering, 8(2), 201-220, https://doi.org/10.1061/(ASCE)0887-3801(1994)8:2(201), 1994.

Kasprzyk, J. R., Nataraj, S., Reed, P. M., et al. Many objective robust decision making for complex environmental systems undergoing change. Environmental Modelling & Software, 42, 55-71, https://doi.org/10.1016/j.envsoft.2012.12.007, 2013.

Kisi, O., Choubin, B., Deo, R. C., et al. Incorporating synoptic-scale climate signals for streamflow modelling over the Mediterranean region using machine learning models. Hydrological Sciences Journal, 64(10), 1240-1252, https://doi.org/10.1080/02626667.2019.1632460, 2019.

Kline, D. 2004. Methods for Multi-Step Time Series Forecasting with Neural Networks. Neural Networks in Business Forecasting. 226-250.

Lee, S., Yen, H., Yeo, I.-Y., et al. Use of multiple modules and Bayesian Model Averaging to assess structural uncertainty of catchment-scale wetland modeling in a Coastal Plain landscape. Journal of Hydrology, 582, 124544, https://doi.org/10.1016/j.jhydrol.2020.124544, 2020.

Legates, D. R., McCabe Jr., G. J. Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. Water Resources Research, 35(1), 233-241, https://doi.org/10.1029/1998WR900018, 1999.

Liu, L., Xu, Y., Pan, S., et al. Potential application of hydrological ensemble prediction in forecasting floods and its components over the Yarlung Zangbo River basin, China. Hydrology and Earth System Sciences, 23, 3335-3352, https://doi.org/10.5194/hess-23-3335-2019, 2019.

Liu, Z., Merwade, V. Separation and prioritization of uncertainty sources in a raster based flood inundation model using hierarchical Bayesian model averaging. Journal of Hydrology, 578, 124100, https://doi.org/10.1016/j.jhydrol.2019.124100, 2019.

Maurer, E. P., Lettenmaier, D. P. Potential effects of long-lead hydrologic predictability on Missouri River main-stem reservoirs. Journal of Climate, 17(1), 174-186, https://doi.org/10.1175/1520-0442(2004)017<0174:PEOLHP>2.0.CO;2, 2004.

Mehdizadeh, S., Fathian, F., Safari, M. J. S., et al. Comparative assessment of time series and artificial intelligence models to estimate monthly streamflow: A local and external data analysis approach. Journal of Hydrology, 579, 124225, https://doi.org/10.1016/j.jhydrol.2019.124225, 2019.

Meng, E., Huang, S., Huang, Q., et al. A robust method for non-stationary streamflow prediction based on improved EMD-SVM model. Journal of Hydrology, 568, 462-478, https://doi.org/10.1016/j.jhydrol.2018.11.015, 2019.

Nanda, T., Sahoo, B., Chatterjee, C. Enhancing real-time streamflow forecasts with wavelet-neural network based error-updating schemes and ECMWF meteorological predictions in Variable Infiltration Capacity model. Journal of Hydrology, 575, 890-910, https://doi.org/10.1016/j.jhydrol.2019.05.051, 2019.

Muhammad A.U., Li X., Feng J. Using LSTM GRU and Hybrid Models for Streamflow Forecasting. Machine Learning and Intelligent Communications 2019. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer, 294, 510-524, https://doi.org/10.1007/978-3-030-32388-2_44, 2019.

Nash, J. E., Sutcliffe, J. V. River flow forecasting through conceptual models part I—A discussion of principles. Journal of Hydrology, 10(3), 282-290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Nott, D. J., Marshall, L., Brown, J. Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: What's the connection? Water Resources Research, 48(12), https://doi.org/10.1029/2011WR011128, 2012.

Osman, A., Afan, H. A., Allawi, M. F., et al. Adaptive Fast Orthogonal Search (FOS) algorithm for forecasting streamflow. Journal of Hydrology, 124896, https://doi.org/10.1016/j.jhydrol.2020.124896, 2020.

Pechlivanidis, I., Crochemore, L., Rosberg, J., et al. What are the key drivers controlling the quality of seasonal streamflow forecasts? Water Resources Research, 56(6), e2019WR026987, https://doi.org/10.1029/2019WR026987, 2020.

Peng, A., Zhang, X., Peng, Y., et al. The application of ensemble precipitation forecasts to reservoir operation. Water Supply, 19(2), 588-595, https://doi.org/10.2166/ws.2018.105, 2018.

Pishvaee, M. S., Razmi, J. Environmental supply chain network design using multi-objective fuzzy mathematical programming. Applied Mathematical Modelling, 36(8), 3433-3446, https://doi.org/10.1016/j.apm.2011.10.007, 2012.

Quinn, J. D., Reed, P. M., Giuliani, M., et al. Rival framings: A framework for discovering how problem formulation uncertainties shape risk management trade-offs in water resources systems. Water Resources Research, 53(8), 7208-7233, https://doi.org/10.1002/2017WR020524, 2017a.

Quinn, J. D., Reed, P. M., Keller, K. Direct policy search for robust multi-objective management of deeply uncertain socio-ecological tipping points. Environmental Modelling & Software, 92, 125-141, https://doi.org/10.1016/j.envsoft.2017.02.017, 2017b.

Ren, C., Guo, P., Tan, Q., et al. A multi-objective fuzzy programming model for optimal use of irrigation water and land resources under uncertainty in Gansu Province, China. Journal of Cleaner Production, 164, 85-94, https://doi.org/10.1016/j.jclepro.2017.06.185, 2017.

Roulston, M. S., Smith, L. A. Combining dynamical and statistical ensembles. Tellus A: Dynamic Meteorology and Oceanography, 55(1), 16-30, https://doi.org/10.3402/tellusa.v55i1.12082, 2003.

Salazar, J. Z., Reed, P. M., Quinn, J. D., et al. Balancing exploration, uncertainty and computational demands in many objective reservoir optimization. Advances in water resources, 109, 196-210, https://doi.org/10.1016/j.advwatres.2017.09.014, 2017.

Samadi, S., Pourreza-Bilondi, M., Wilson, C., et al. Bayesian model averaging with fixed and flexible priors: Theory, concepts, and calibration experiments for rainfall-runoff modeling. Journal of Advances in Modeling Earth Systems, 12(7), e2019MS001924, https://doi.org/10.1029/2019MS001924, 2020.

Schwanenberg, D., Fan, F. M., Naumann, S., et al. Short-Term Reservoir Optimization for Flood Mitigation under Meteorological and Hydrological Forecast Uncertainty. Water Resources Management, 29(5), 1635-1651, https://doi.org/10.1007/s11269-014-0899-1, 2015.

Shahid, F., Zameer, A., Muneeb, M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. Chaos, Solitons & Fractals, 140, 110212, https://doi.org/10.1016/j.chaos.2020.110212, 2020.

Shamir, E. The value and skill of seasonal forecasts for water resources management in the Upper Santa Cruz River basin, southern Arizona. Journal of Arid Environments, 137, 35-45, https://doi.org/10.1016/j.jaridenv.2016.10.011, 2017.

Sharma, S., Siddique, R., Reed, S., et al. Hydrological Model Diversity Enhances Streamflow Forecast Skill at Short-to Medium-Range Timescales. Water Resources Research, 55(2), 1510-1530, https://doi.org/10.1029/2018WR023197, 2019.

Suykens, J. A. K., Gestel, V., T., D. B., et al. Least Squares Support Vector Machines. World Scientific Publishing.2002. Singapore.

Tikhamarine, Y., Souag-Gamane, D., Najah Ahmed, A., et al. Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. Journal of Hydrology, 582, 124435, https://doi.org/10.1016/j.jhydrol.2019.124435, 2020.

Turner, S., Bennett, J., Robertson, D., et al. Complex relationship between seasonal streamflow forecast skill and value in reservoir operations. Hydrology and Earth System Sciences, 21, 4841-4859, https://doi.org/10.5194/hess-21-4841-2017, 2017.

Van Osnabrugge, B., Uijlenhoet, R., Weerts, A. Contribution of potential evaporation forecasts to 10-day streamflow forecast skill for the Rhine River. Hydrology and Earth System Sciences, 23(3), 1453-1467, https://doi.org/10.5194/hess-23-1453-2019, 2019.

Xiong, L., Wan, M., Wei, X., et al. Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation. Hydrological Sciences Journal, 54(5), 852-871, https://doi.org/10.1623/hysj.54.5.852, 2009.

Xu, B., Zhong, P.-a., Lu, Q., et al. Multiobjective stochastic programming with recourses for real-time flood water conservation of a multireservoir system under uncertain forecasts. Journal of Hydrology, 590, 125513, https://doi.org/10.1016/j.jhydrol.2020.125513, 2020.

Xu, W., Zhang, C., Peng, Y., et al. A two stage B ayesian stochastic optimization model for cascaded hydropower systems considering varying uncertainty of flow forecasts. Water Resources Research, 50(12), 9267-9286, https://doi.org/10.1002/2013WR015181, 2014.

Yan, D., Ludwig, F., Huang, H. Q., et al. Many-objective robust decision making for water allocation under climate change. Science of the Total Environment, 607, 294-303, https://doi.org/10.1016/j.scitotenv.2017.06.265, 2017.

Yang, G., Guo, S., Liu, P., et al. Multiobjective reservoir operating rules based on cascade reservoir input variable selection method. Water Resources Research, 53(4), 3446-3463, https://doi.org/10.1002/2016WR020301, 2017.

Zeff, H. B., Kasprzyk, J. R., Herman, J. D., et al. Navigating financial and supply reliability tradeoffs in regional drought management portfolios. Water Resources Research, 50(6), 4906-4923, https://doi.org/10.1002/2013WR015126, 2014.

Zhai, B., Chen, J. Development of a stacked ensemble model for forecasting and analyzing daily average PM2. 5 concentrations in Beijing, China. Science of the Total Environment, 635, 644-658, https://doi.org/10.1016/j.scitotenv.2018.04.040, 2018.

Zhang, D., Lindholm, G., Ratnaweera, H. Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring. Journal of Hydrology, 556, 409-418, https://doi.org/10.1016/j.jhydrol.2017.11.018, 2018.

Zhang, F., Guo, S., Liu, X., et al. Towards sustainable water management in an arid agricultural region: A multi-level multi-objective stochastic approach. Agricultural Systems, 182, 102848, https://doi.org/10.1016/j.agsy.2020.102848, 2020.

Zhang, J., Liu, P., Wang, H., et al. A Bayesian model averaging method for the derivation of reservoir operating rules. Journal of Hydrology, 528, 276-285, https://doi.org/10.1016/j.jhydrol.2015.06.041, 2015.

Zhao, Q., Cai, X., Li, Y. Determining Inflow Forecast Horizon for Reservoir Operation. Water Resources Research, 55(5), 4066-4081, https://doi.org/10.1029/2019WR025226, 2019.

Zhou, Y., Chang, F.-J., Chen, H., et al. Exploring Copula-based Bayesian Model Averaging with multiple ANNs for PM2.5 ensemble forecasts. Journal of Cleaner Production, 263, 121528, https://doi.org/10.1016/j.jclepro.2020.121528, 2020a.

Zhou, Y., Chang, L.-C., Chang, F.-J. Explore a Multivariate Bayesian Uncertainty Processor driven by artificial neural networks for probabilistic PM2.5 forecasting. Science of the Total Environment, 711, 134792, https://doi.org/10.1016/j.scitotenv.2019.134792, 2020b.

Zimmermann, H.-J. Fuzzy programming and linear programming with several objective functions. Fuzzy sets systems, 1(1), 45-55, https://doi.org/10.1016/0165-0114(78)90031-3, 1978.

Zuo, G., Luo, J., Wang, N., et al. Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting. Journal of Hydrology, 124776, https://doi.org/10.1016/j.jhydrol.2020.124776, 2020.

**List of Figure Captions**

Figure 1: Framework of the AI-based management methodology.

Figure 2: Structure of an (a) LSTM and (b) GRU cell.

Figure 3: Schematization of the parameterized MORDM methods.

Figure 4: Location of the Zhoushan Islands.

Figure 5: Schematic diagram of the Zhoushan Islands.

Figure 6: NSE values at lead times of 1 to 7 days plotted against the coefficient of variation (COV) for all the 24 reservoirs during the period of (a) calibration, (b) validation, and (c) test under S3.

Figure 7: NSE values at lead times of 1 to 7 days plotted against the coefficient of variation (COV) for all the 24 reservoirs during the period of (a) calibration, (b) validation, and (c) test S5.

Figure 8: Weights of three individual forecast models for the BMA model for all reservoirs under (a) S3 and (b) S5.

Figure 9: Improvement rates in terms of averaged (a) NSE, (b) RMSE, and (c) MAE of the BMA model for forecasts as compared with the three individual models.

Figure 10: Forecast results of (a) Hongqiao, (b) Goushan, and (c) Nanao reservoirs under S3.

Figure 11: Forecast results of (a) Hongqiao, (b) Goushan, and (c) Nanao reservoirs under S5.

Figure 12: 90% streamflow interval of the BMA method under S3.

Figure 13: 90% streamflow interval of the BMA method under S5.

Figure 14: A set of Pareto solutions at different periods over a 7-day horizon under (a) deterministic and (b) uncertain forecasts.

Figure 15: Seasonal system performance of water supply reliability.

Figure 16: Annual system performance with different forecast horizons.

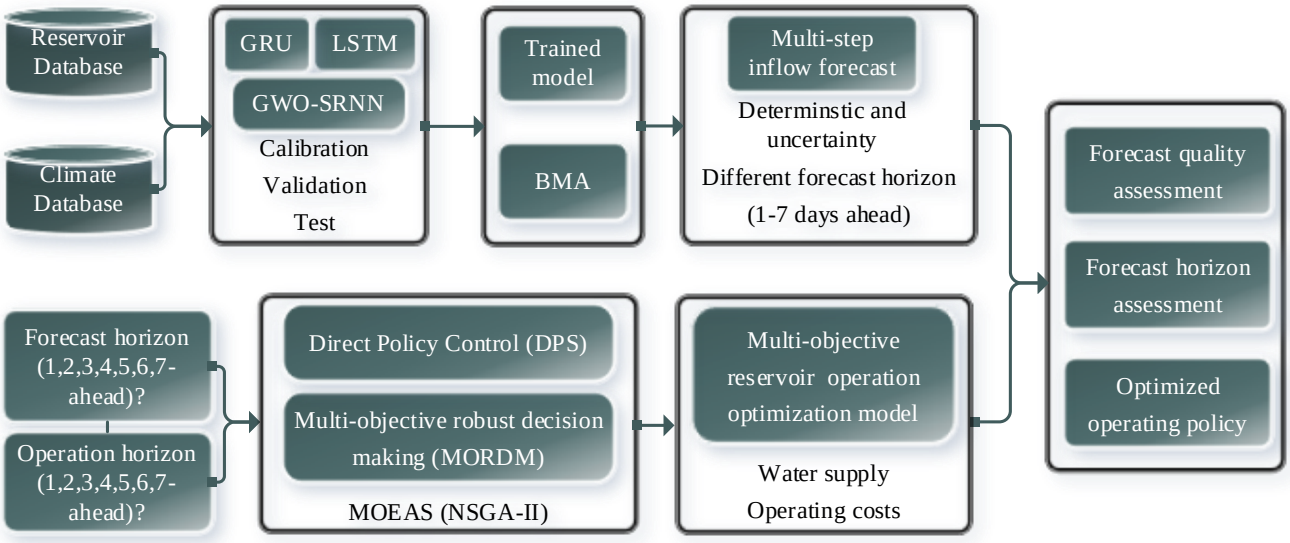
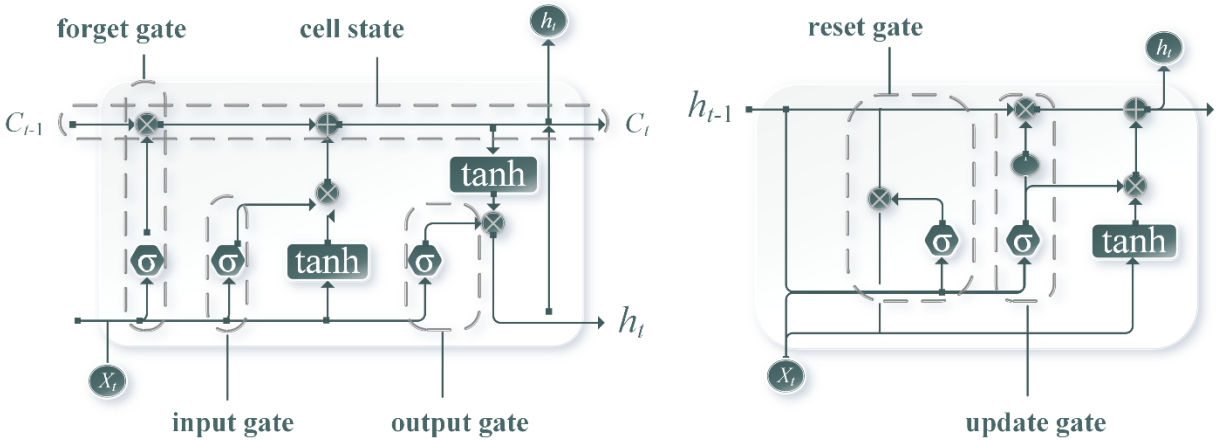



**Figure 1: Framework of the AI-based management methodology.**




(a) LSTM (b) GRU

**Figure 2: Structure of an (a) LSTM and (b) GRU cell.**

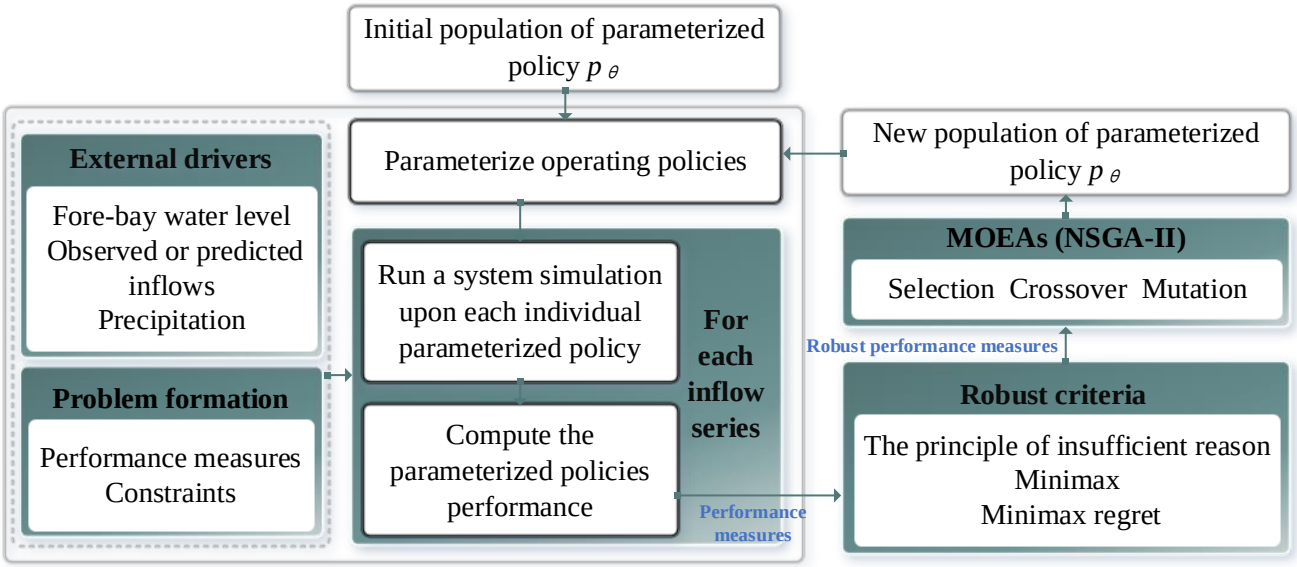
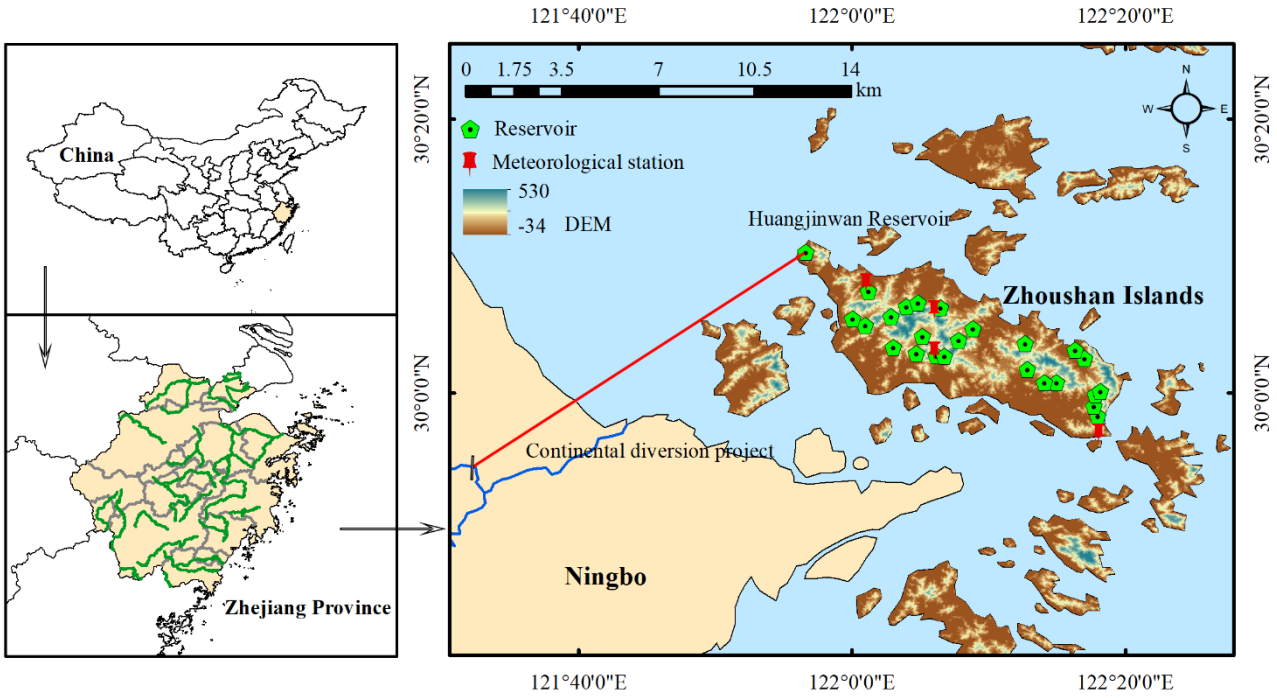



**Figure 3: Schematization of the parameterized MORDM methods.**




875

**Figure 4: Location of the Zhoushan Islands.**

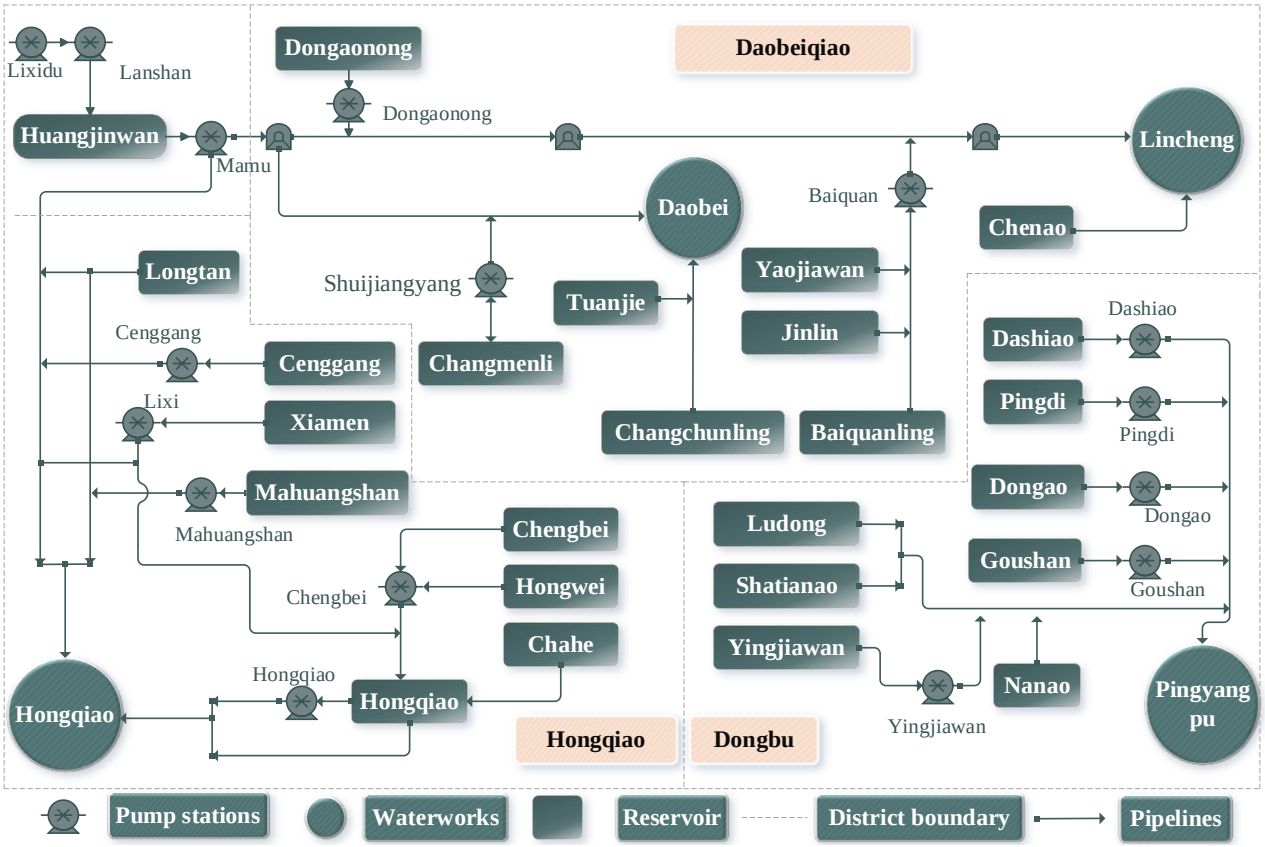
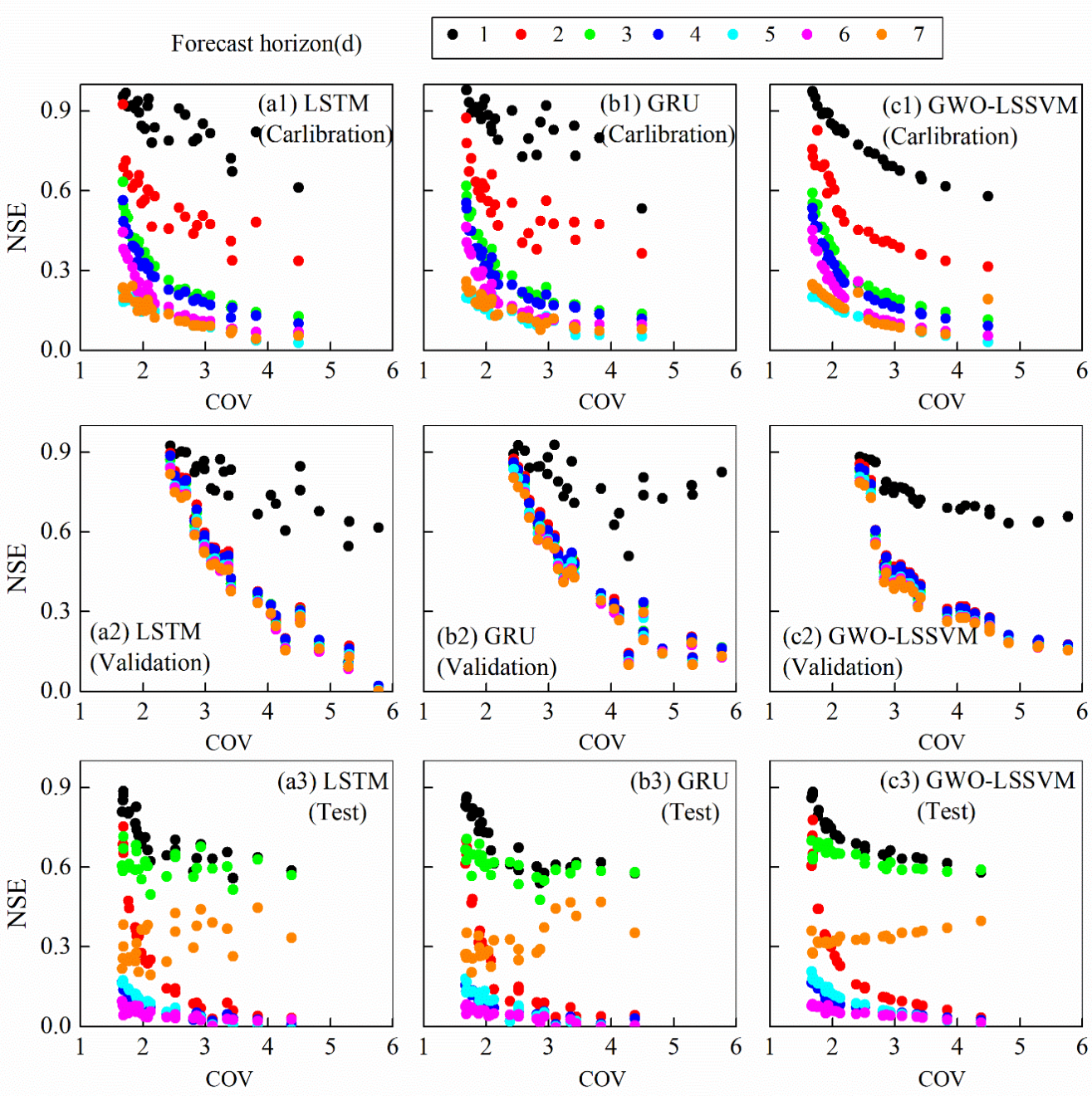



**Figure 5: Schematic diagram of the Zhoushan Islands.**




880

**Figure 6: NSE values at lead times of 1 to 7 days plotted against the coefficient of variation (COV) for all the 24 reservoirs during the period of (a) calibration, (b) validation, and (c) test under S3.**

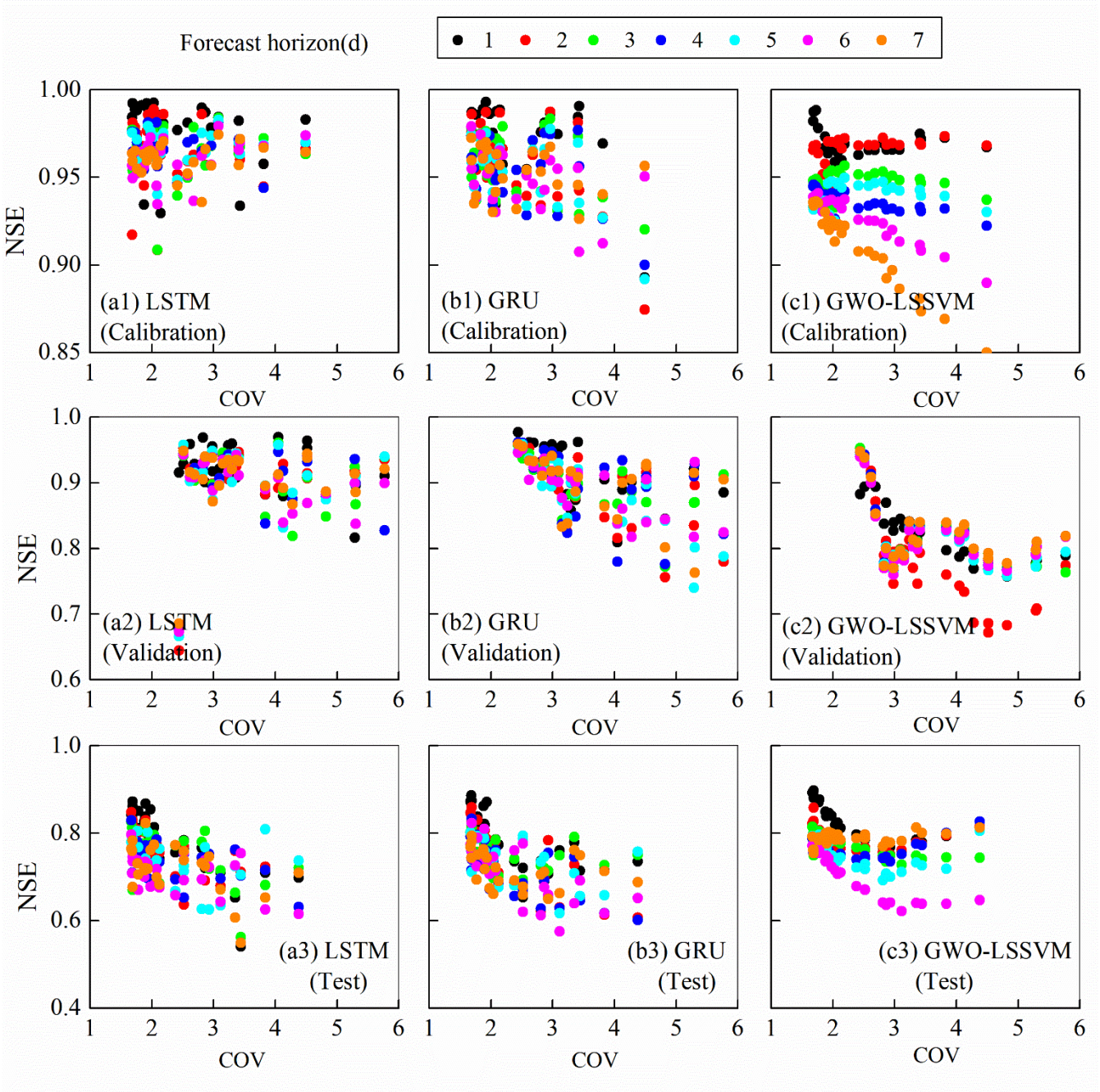



**Figure 7: NSE values at lead times of 1 to 7 days plotted against the coefficient of variation (COV) for all the 24 reservoirs during the period of (a) calibration, (b) validation, and (c) test under S5.**

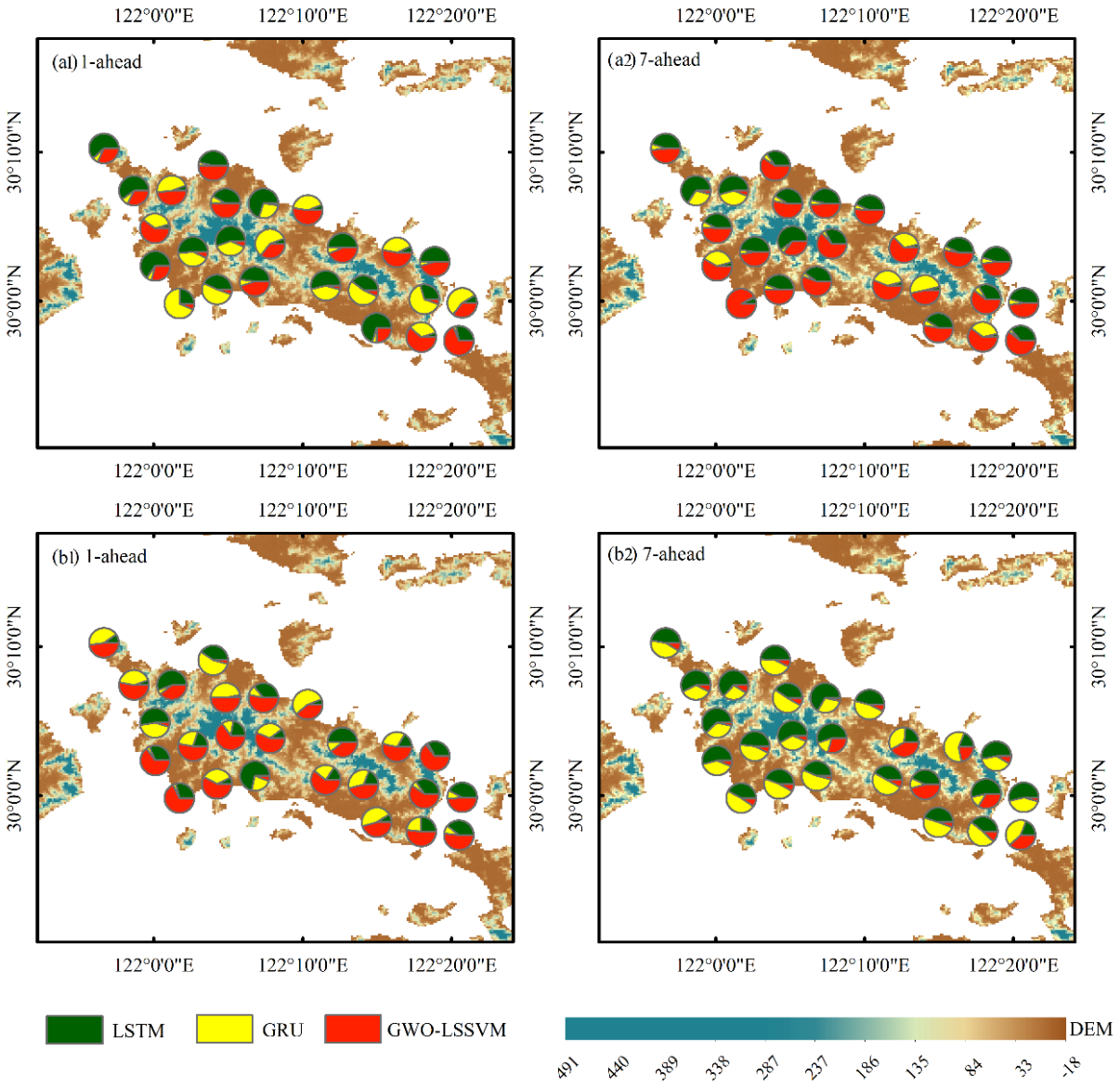



**Figure 8: Weights of three individual forecast models for the BMA model for all reservoirs under (a) S3 and (b) S5.**

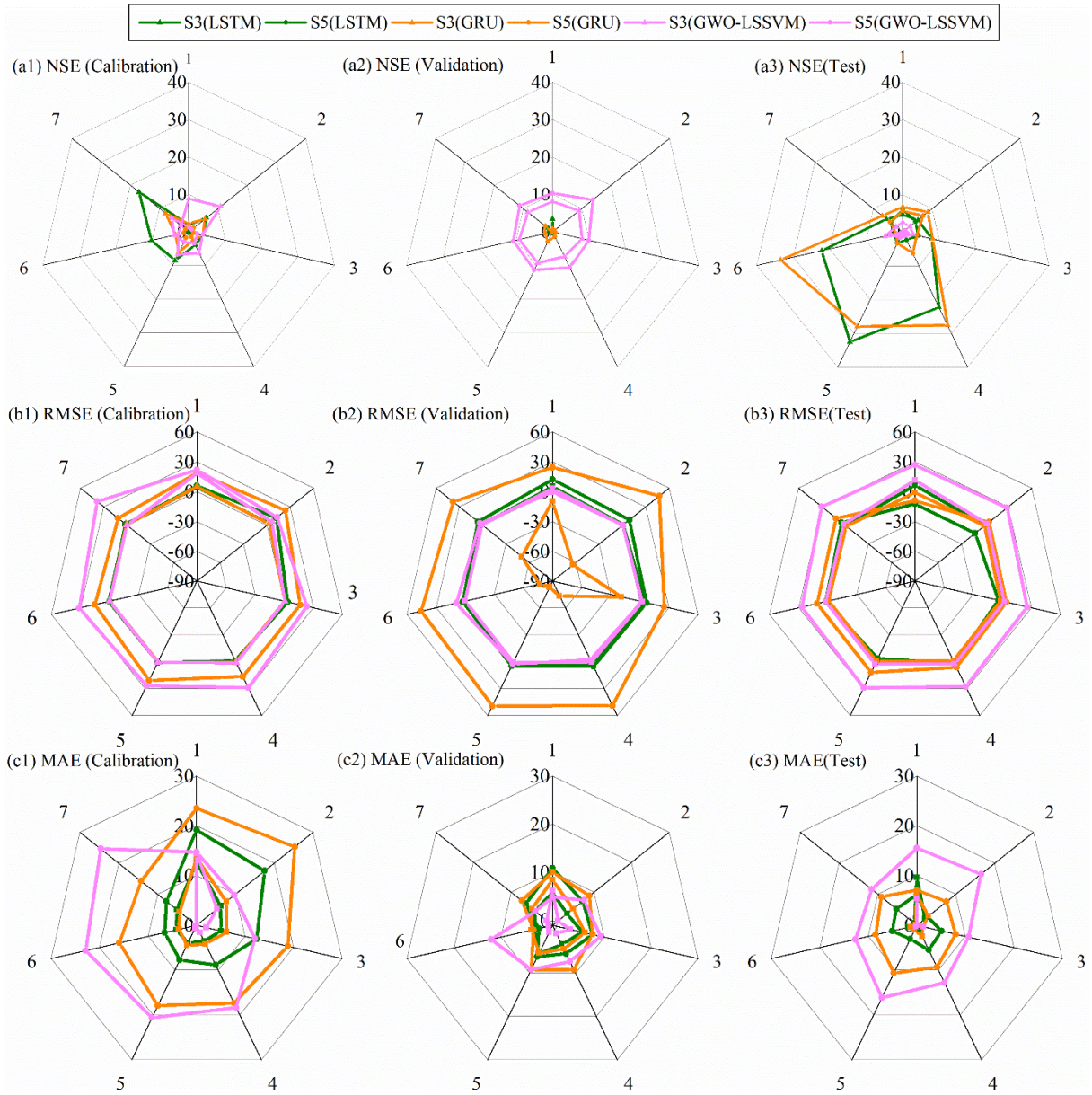



**Figure 9: Improvement rates in terms of averaged (a) NSE, (b) RMSE, and (c) MAE of the BMA model for forecasts as compared with the three individual models.**

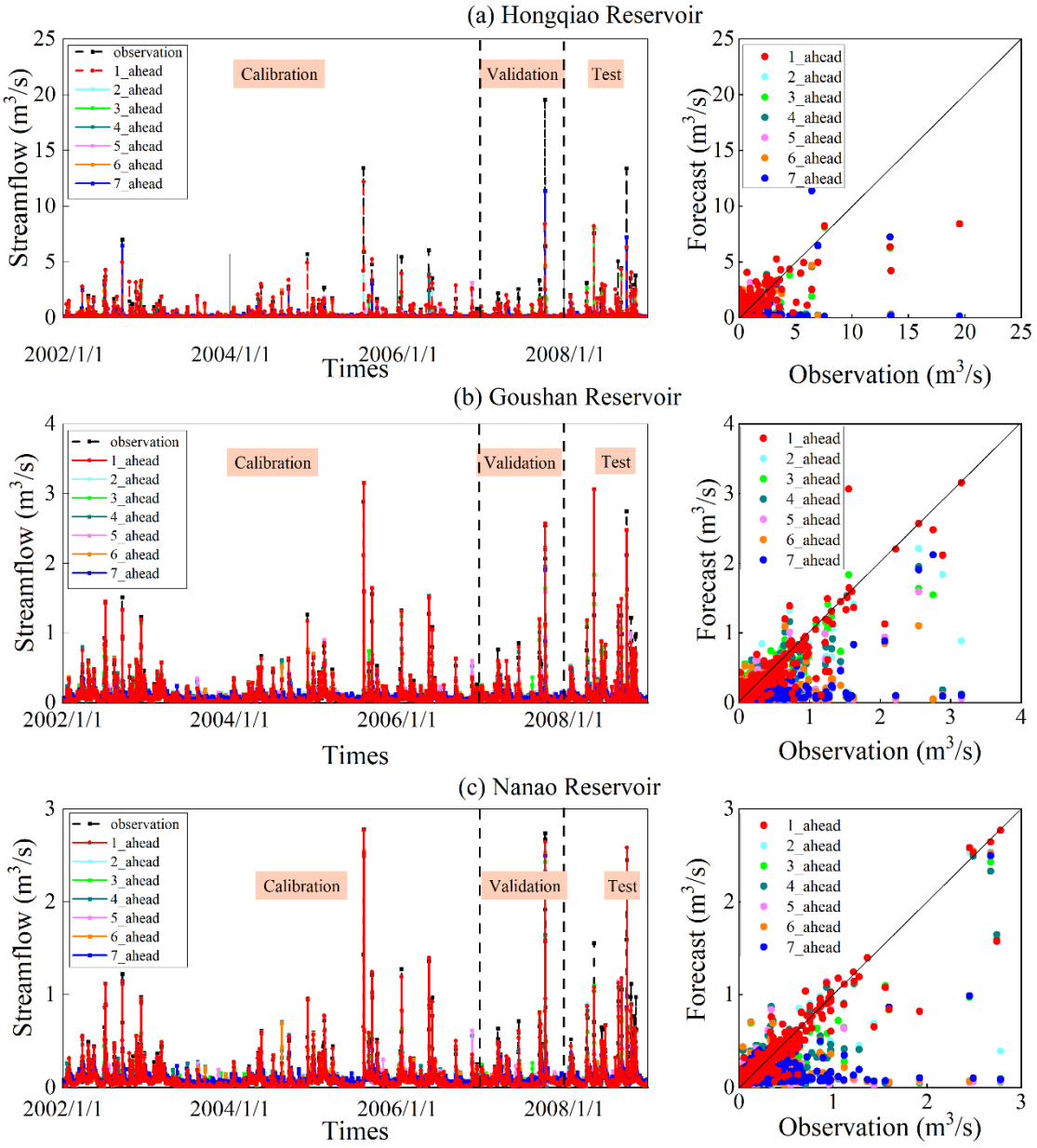



**Figure 10: Forecast results of (a) Hongqiao, (b) Goushan, and (c) Nanao reservoirs under S3.**

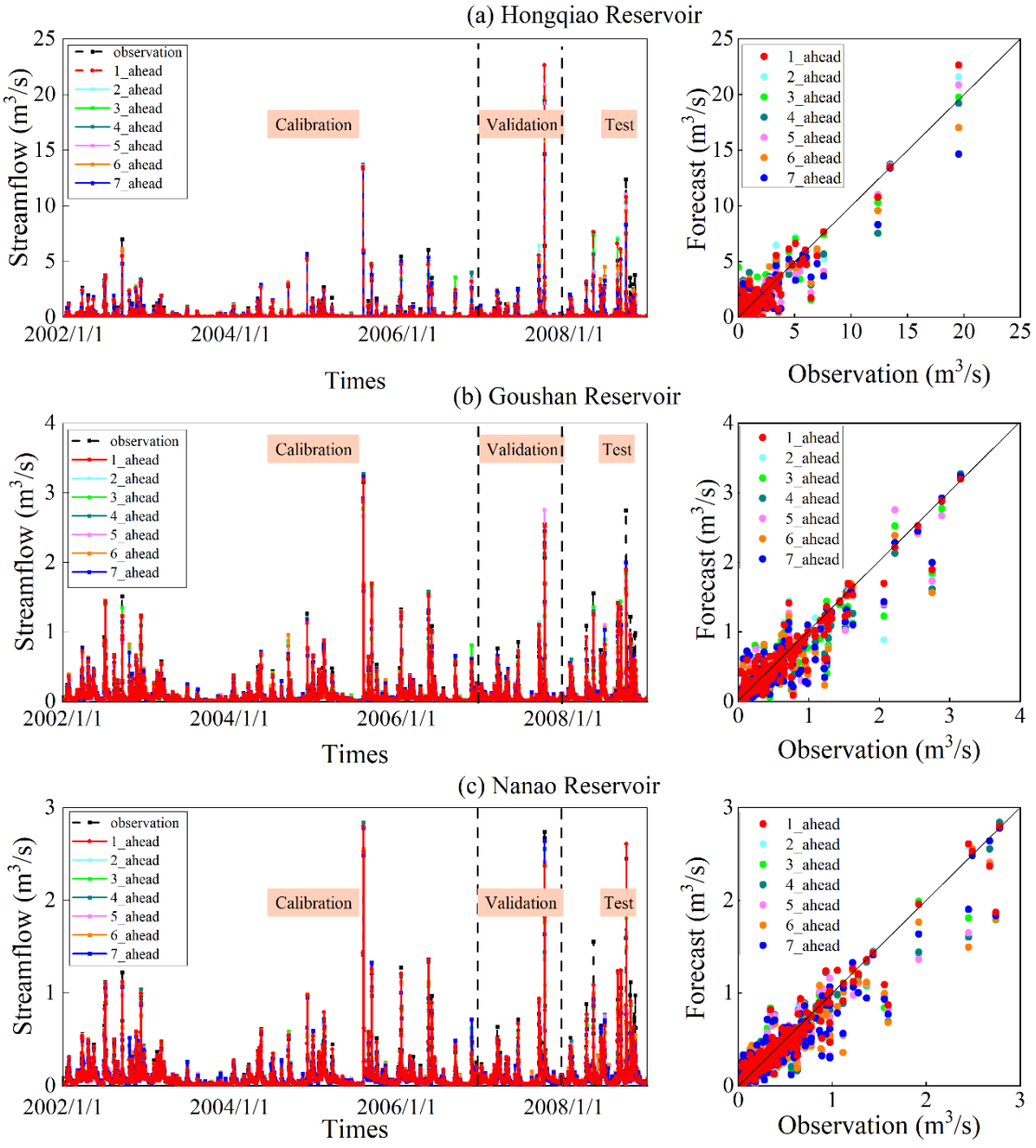



**Figure 11: Forecast results of (a) Hongqiao, (b) Goushan, and (c) Nanao reservoirs under S5.**

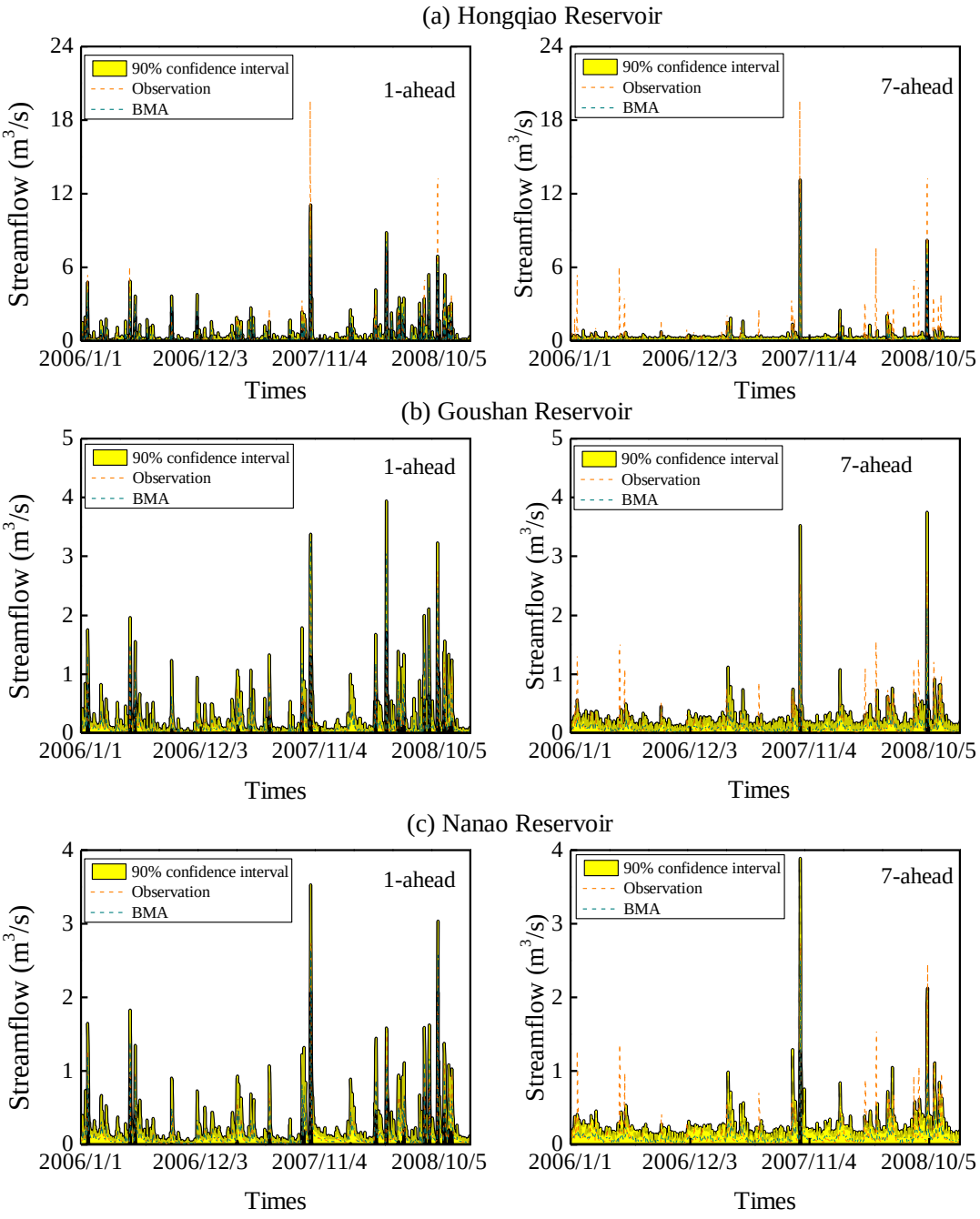



**Figure 12: 90% streamflow interval of the BMA method under S3.**

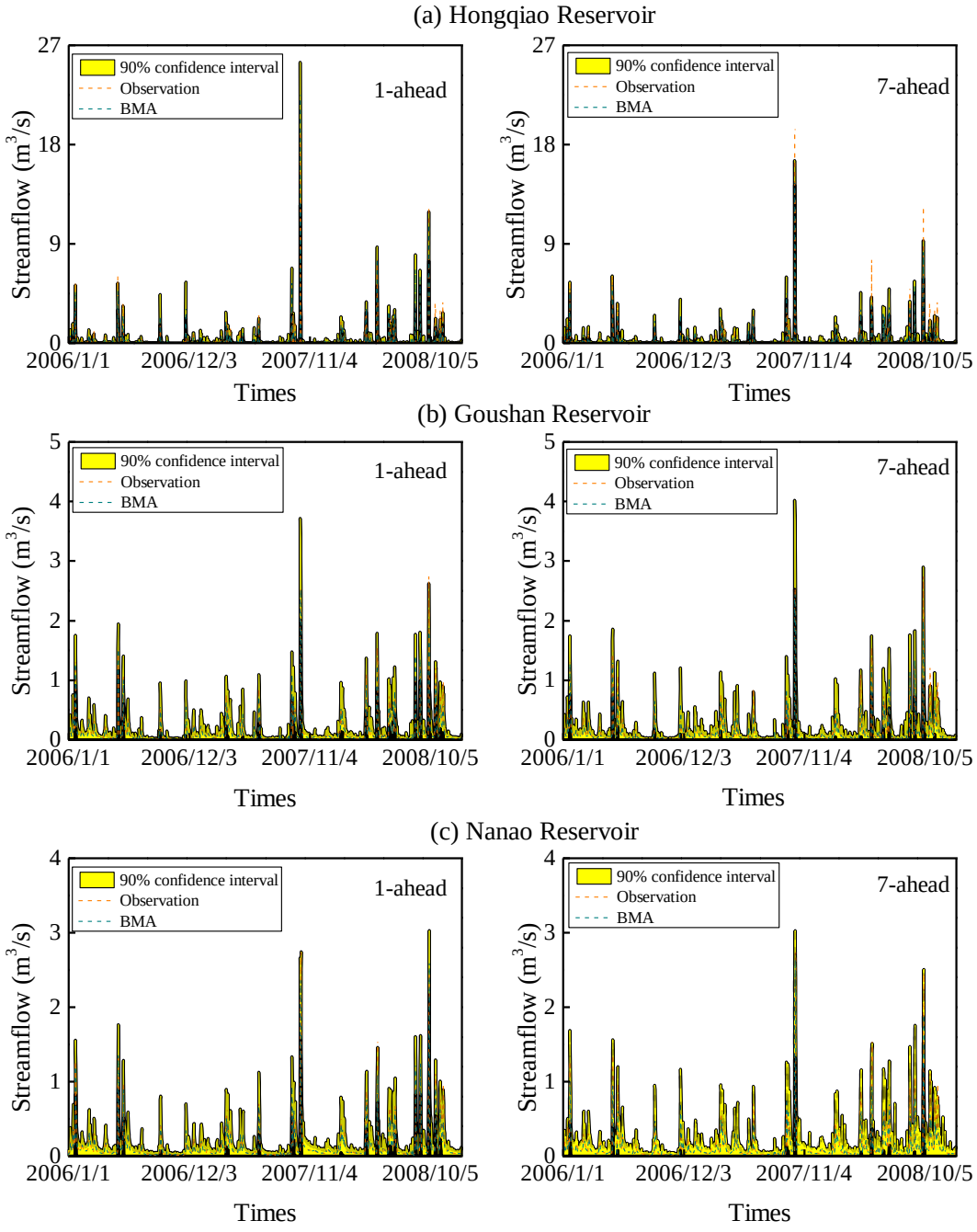
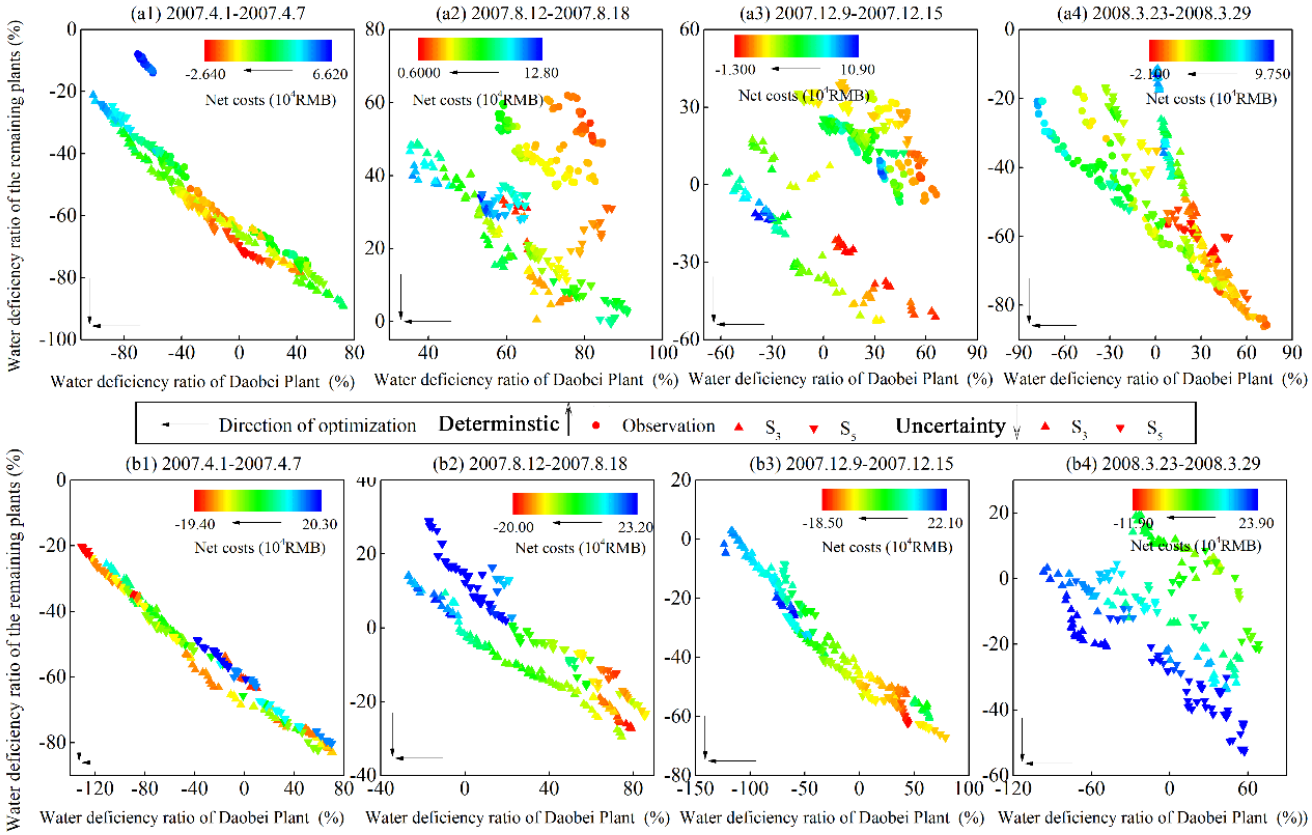



**Figure 13: 90% streamflow interval of the BMA method under S5.**




900 **Figure 14: A set of Pareto solutions at different periods over a 7-day horizon under (a) deterministic and (b) uncertain forecasts.**

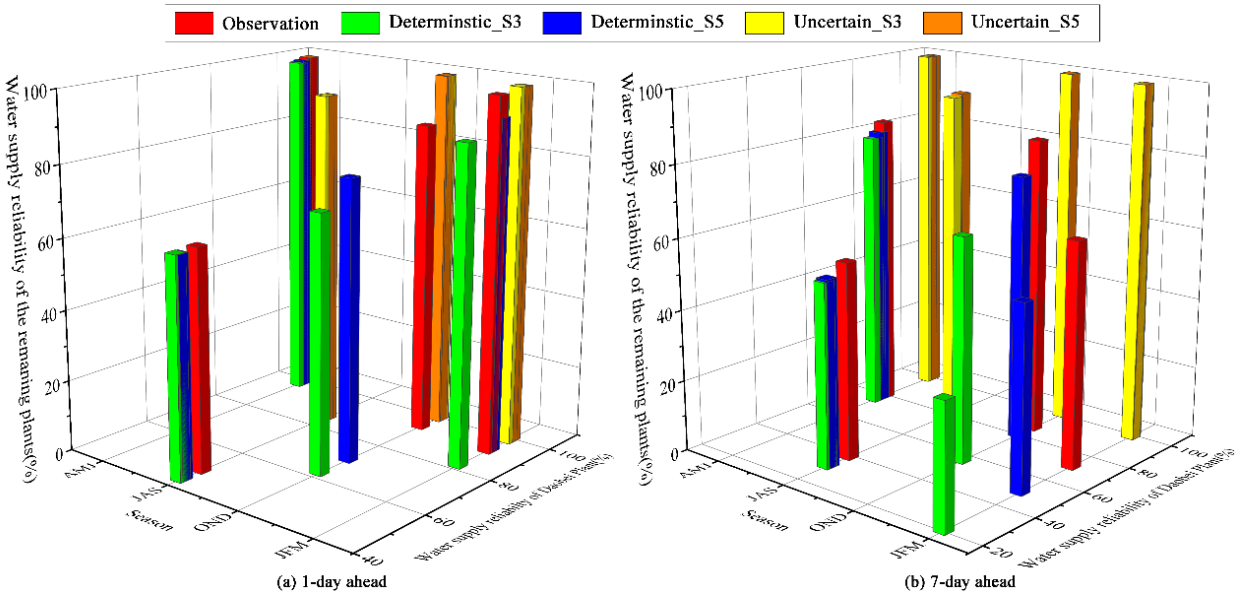



**Figure 15: Seasonal system performance of water supply reliability.**

905

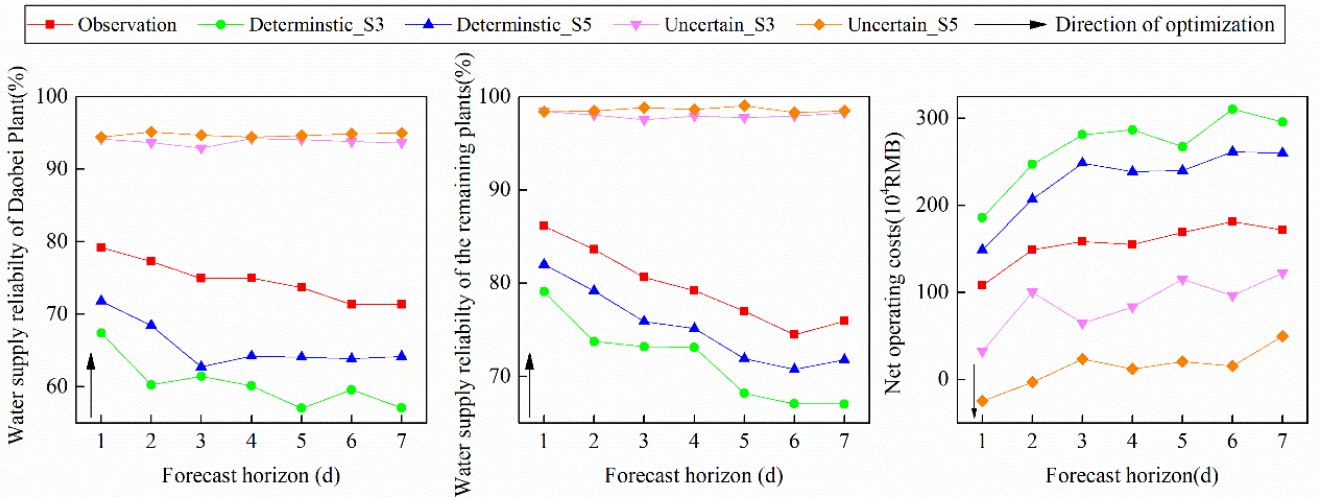



**Figure 16: Annual system performance with different forecast horizons.**

**List of Table Captions**

Table 1: Reservoir characteristics in the Zhoushan Islands.

Table 2: Five input combination scenarios.

Table 3: NSE ranges ([min, max]) for all reservoirs with the different configurations during the calibration, validation, and test periods.

Table 4: Performance metric ranges ([min, max]) for all 24 reservoirs of BMA methods under S3 and S5.

Table 5: Ranges of interval performance metrics ([min, max]) for all the 24 reservoirs under S3 and S5.

Table 6: Annual system performance using forecast inflow information.

Table 7: Annual system performance using observed inflow information.

**Table 1: Reservoir characteristics in the Zhoushan Islands.**

| District | Reservoir name | Reservoir storage ($10^4m^3$) | Dead storage ($10^4m^3$) | Normal storage ($10^4m^3$) | Drainage area ($km^2$) | Mean flow ($m^3/s$) | Standard deviation of flow ($m^3/s$) | COV |
|---|---|---|---|---|---|---|---|---|
| Hongqiao | Hongqiao | 1307 | 12 | 1015 | 13.4 | 0.15 | 0.77 | 5.08 |
| | Chahe | 254 | 35.08 | 185 | 8 | 0.11 | 0.49 | 4.34 |
| | Hongwei | 85 | 36 | 76.1 | 1.94 | 0.10 | 0.21 | 2.06 |
| | Chengbei | 123 | 45 | 111.1 | 4.98 | 0.10 | 0.33 | 3.34 |
| | Mahuangshan | 354 | 17.15 | 286.4 | 4.87 | 0.10 | 0.32 | 3.28 |
| | Xiamen | 281 | 42 | 240 | 4 | 0.10 | 0.29 | 2.92 |
| | Cenggang | 733 | 14.2 | 627 | 6.6 | 0.10 | 0.41 | 3.91 |
| | Longtan | 160 | 9 | 133.6 | 2.27 | 0.10 | 0.22 | 2.18 |
| Daobei | Dongaonong | 185 | 3.4 | 159.84 | 2.6 | 0.10 | 0.23 | 2.28 |
| | Changmenli | 205 | 49.49 | 179.5 | 2.3 | 0.10 | 0.22 | 2.17 |
| | Tuanjie | 122 | 30.4 | 106.6 | 2.05 | 0.10 | 0.21 | 2.09 |
| | Changchunling | 410 | 34.3 | 368.3 | 5.41 | 0.10 | 0.35 | 3.53 |
| | Yaojiawan | 124 | 31.09 | 105 | 1.46 | 0.10 | 0.22 | 2.23 |
| | Jinlin | 154 | 40.48 | 125.9 | 2.42 | 0.10 | 0.20 | 1.96 |
| | BaiquanLing | 204 | 12.56 | 177.4 | 3 | 0.10 | 0.24 | 2.44 |
| | Chenao | 236 | 59 | 195.2 | 4.13 | 0.10 | 0.29 | 2.99 |
| Dongbu | Dashiao | 293 | 49.1 | 254 | 2.8 | 0.10 | 0.23 | 2.37 |
| | Pingdi | 317 | 1 | 317.2 | 0.6 | 0.10 | 0.19 | 1.87 |
| | Dongao | 457.8 | 47.5 | 384 | 6.4 | 0.10 | 0.40 | 3.87 |
| | Goushan | 194 | 4.59 | 170 | 2.73 | 0.10 | 0.23 | 2.34 |
| | Nanao | 73 | 15.8 | 66.7 | 1.21 | 0.10 | 0.20 | 1.94 |
| | Ludong | 142 | 54 | 118.5 | 3.7 | 0.10 | 0.27 | 2.75 |
| | Yingjiawan | 124 | 31.09 | 105 | 1.46 | 0.10 | 0.31 | 3.17 |
| | Shatianao | 127 | 20.8 | 116.5 | 4.54 | 0.10 | 0.19 | 1.89 |

**Table 2: Five input combination scenarios.**

| ID | Scenario | Input combination |
|---|---|---|
| 1 | S1 | $Q_a$ |
| 2 | S2 | $P_{a,} E_a$ |
| 3 | S3 | $Q_{a,} P_{a,} E_a$ |
| 4 | S4 | $P_a$, $P_f$, $E_{a,} E_f$ |
| 5 | S5 | $Q_a$, $P_a$, $P_f$, $E_{a,} E_f$ |

**Table 3: NSE ranges ([min, max]) for all reservoirs with the different configurations during the calibration, validation, and test periods.**

| Period | Forecast horizon (d) | LSTM | | | | | GRU | | | | | GWO-LSSVM | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | S1 | S2 | S3 | S4 | S5 | S1 | S2 | S3 | S4 | S5 | S1 | S2 | S3 | S4 | S5 |
| Calibr- | 1 | [0.11, | [0.57, | [0.61, | [0.8, | [0.93, | [0.18, | [0.57, | [0.53, | [0.66, | [0.89, | [0.17, | [0.54, | [0.58, | [0.87, | [0.96, |
| ation | | 0.87] | 0.94] | 0.97] | 0.96] | 0.99] | 0.87] | 0.87] | 0.98] | 0.96] | 0.99] | 0.91] | 0.86] | 0.97] | 0.97] | 0.99] |
| | 2 | [0.05, | [0.27, | [0.34, | [0.82, | [0.91, | [0.07, | [0.29, | [0.36, | [0.78, | [0.87, | [0.08, | [0.27, | [0.31, | [0.87, | [0.94, |
| | | 0.58] | 0.72] | 0.93] | 0.95] | 0.99] | 0.58] | 0.66] | 0.87] | 0.96] | 0.99] | 0.58] | 0.72] | 0.83] | 0.97] | 0.97] |
| | 3 | [0.03, | [0.11, | [0.13, | [0.75, | [0.91, | [0.05, | [0.10, | [0.14, | [0.79, | [0.92, | [0.05, | [0.13, | [0.11, | [0.86, | [0.93, |
| | | 0.48] | 0.55] | 0.63] | 0.94] | 0.98] | 0.51] | 0.52] | 0.62] | 0.95] | 0.98] | 0.51] | 0.55] | 0.59] | 0.94] | 0.96] |
| | 4 | [0.03, | [0.08, | [0.10, | [0.84, | [0.94, | [0.04, | [0.08, | [0.12, | [0.80, | [0.90, | [0.05, | [0.1, | [0.09, | [0.87, | [0.92, |
| | | 0.44] | 0.49] | 0.56] | 0.95] | 0.98] | 0.45] | 0.45] | 0.56] | 0.95] | 0.98] | 0.45] | 0.8] | 0.54] | 0.92] | 0.95] |
| | 5 | [0.01, | [0.02, | [0.03, | [0.74, | [0.94, | [0.02, | [0.02, | [0.05, | [0.86, | [0.89, | [0.03, | [0.05, | [0.03, | [0.87, | [0.93, |
| | | 0.17] | 0.16] | 0.22] | 0.95] | 0.98] | 0.17] | 0.17] | 0.22] | 0.95] | 0.98] | 0.16] | 0.46] | 0.23] | 0.93] | 0.95] |
| | 6 | [0.01, | [0.06, | [0.07, | [0.83, | [0.93, | [0.02, | [0.05, | [0.09, | [0.8, | [0.91, | [0.03, | [0.07, | [0.05, | [0.87, | [0.89, |
| | | 0.39] | 0.39] | 0.44] | 0.95] | 0.98] | 0.4] | 0.38] | 0.46] | 0.95] | 0.98] | 0.41] | 0.87] | 0.45] | 0.90] | 0.94] |
| | 7 | [0.01, | [0.04, | [0.04, | [0.84, | [0.94, | [0.02, | [0.04, | [0.07, | [0.86, | [0.93, | [0.02, | [0.06, | [0.06, | [0.84, | [0.85, |
| | | 0.18] | 0.19] | 0.24] | 0.96] | 0.97] | 0.19] | 0.19] | 0.26] | 0.95] | 0.97] | 0.19] | 0.81] | 0.25] | 0.88] | 0.94] |
| Valid- | 1 | [0.09, | [0.45, | [0.50, | [0.79, | [0.82, | [0.11, | [0.47, | [0.51, | [0.34, | [0.81, | [0.04, | [0.5, | [0.58, | [0.70, | [0.76, |
| ation | | 0.90] | 0.93] | 0.92] | 0.96] | 0.97] | 0.87] | 0.87] | 0.98] | 0.96] | 0.99] | 0.79] | 0.95] | 0.88] | 0.93] | 0.90] |
| | 2 | [0.08, | [0.01, | [0.01, | [0.42, | [0.64, | [0.09, | [0.09, | [0.07, | [0.54, | [0.76, | [0.00, | [0.01, | [0.03, | [0.70, | [0.67, |
| | | 0.85] | 0.87] | 0.90] | 0.95] | 0.95] | 0.58] | 0.66] | 0.87] | 0.96] | 0.99] | 0.74] | 0.83] | 0.86] | 0.93] | 0.95] |
| | 3 | [0.08, | [0.02, | [-0.01, | [0.79, | [0.68, | [0.09, | [0.09, | [0.08, | [0.52, | [0.77, | [0.00, | [0.02, | [0.03, | [0.74, | [0.76, |
| | | 0.83] | 0.83] | 0.87] | 0.96] | 0.96] | 0.51] | 0.52] | 0.62] | 0.95] | 0.98] | 0.74] | 0.8] | 0.83] | 0.94] | 0.95] |
| | 4 | [0.08, | [0.01, | [0.02, | [0.80, | [0.68, | [0.09, | [0.08, | [0.07, | [0.52, | [0.78, | [0.00, | [0.01, | [0.04, | [0.73, | [0.77, |
| | | 0.83] | 0.84] | 0.89] | 0.96] | 0.95] | 0.45] | 0.45] | 0.56] | 0.95] | 0.98] | 0.74] | 0.81] | 0.84] | 0.94] | 0.95] |
| | 5 | [0.08, | [0.01, | [0.01, | [0.75, | [0.67, | [0.08, | [0.07, | [0.06, | [0.52, | [0.74, | [0.00, | [-0.01, | [0.02, | [0.70, | [0.76, |
| | | 0.81] | 0.82] | 0.85] | 0.96] | 0.96] | 0.17] | 0.17] | 0.22] | 0.95] | 0.98] | 0.72] | 0.78] | 0.81] | 0.94] | 0.94] |
| | 6 | [0.08, | [0.00, | [0.00, | [0.80, | [0.67, | [0.09, | [0.07, | [0.05, | [0.51, | [0.80, | [0.01, | [0.00, | [0.02, | [0.73, | [0.76, |
| | | 0.80] | 0.80] | 0.84] | 0.95] | 0.94] | 0.4] | 0.38] | 0.46] | 0.95] | 0.98] | 0.71] | 0.77] | 0.79] | 0.94] | 0.94] |
| | 7 | [0.07, | [0.01, | [0.00, | [0.76, | [0.69, | [0.08, | [0.07, | [0.06, | [0.53, | [0.76, | [0.00, | [0.00, | [0.02, | [0.77, | [0.77, |
| | | 0.78] | 0.79] | 0.82] | 0.96] | 0.95] | 0.19] | 0.19] | 0.26] | 0.95] | 0.97] | 0.70] | 0.76] | 0.79] | 0.95] | 0.95] |
| Test | 1 | [-0.04, | [0.50, | [0.56, | [0.58, | [0.54, | [-0.09, | [0.48, | [0.54, | [0.54, | [0.65, | [0.04, | [0.53, | [0.58, | [0.69, | [0.76, |
| | | 0.69] | 0.73] | 0.89] | 0.77] | 0.87] | 0.71] | 0.74] | 0.87] | 0.76] | 0.89] | 0.72] | 0.71] | 0.88] | 0.79] | 0.90] |
| | 2 | [-0.13, | [0.04, | [0.03, | [0.41, | [0.63, | [-0.16, | [0.04, | [0.03, | [0.58, | [0.61, | [-0.01, | [0.03, | [0.03, | [0.69, | [0.75, |
| | | 0.69] | 0.62] | 0.75] | 0.78] | 0.85] | 0.66] | 0.59] | 0.7] | 0.77] | 0.86] | 0.71] | 0.63] | 0.78] | 0.79] | 0.86] |
| | 3 | [0.01, | [0.48, | [0.50, | [0.58, | [0.56, | [0.03, | [0.51, | [0.48, | [0.61, | [0.69, | [0.03, | [0.50, | [0.58, | [0.67, | [0.73, |
| | | 0.20] | 0.65] | 0.72] | 0.73] | 0.83] | 0.2] | 0.65] | 0.71] | 0.78] | 0.82] | 0.2] | 0.63] | 0.70] | 0.74] | 0.81] |
| | 4 | [-0.01, | [0.02, | [0.01, | [0.41, | [0.63, | [-0.02, | [0.02, | [0.01, | [0.57, | [0.60, | [-0.01, | [0.01, | [0.02, | [0.68, | [0.74, |
| | | 0.15] | 0.13] | 0.17] | 0.77] | 0.83] | 0.13] | 0.13] | 0.15] | 0.77] | 0.83] | 0.13] | 0.12] | 0.18] | 0.81] | 0.83] |
| | 5 | [-0.01, | [0.01, | [-0.01, | [0.63, | [0.63, | [-0.02, | [-0.01, | [0.00, | [0.53, | [0.62, | [0.00, | [0.01, | [0.01, | [0.65, | [0.69, |
| | | 0.14] | 0.14] | 0.17] | 0.78] | 0.81] | 0.14] | 0.15] | 0.18] | 0.75] | 0.80] | 0.12] | 0.15] | 0.21] | 0.77] | 0.80] |
| | 6 | [-0.01, | [0.01, | [0.00, | [0.62, | [0.62, | [-0.01, | [0.01, | [0.00, | [0.56, | [0.58, | [-0.01, | [0.02, | [0.01, | [0.59, | [0.62, |
| | | 0.05] | 0.08] | 0.10] | 0.76] | 0.80] | 0.05] | 0.08] | 0.08] | 0.82] | 0.82] | 0.04] | 0.07] | 0.08] | 0.75] | 0.77] |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 7 | [-0.02, 0.04] | [0.23, 0.40] | [0.19, 0.45] | [0.57, 0.74] | [0.55, 0.82] | [-0.02, 0.03] | [0.22, 0.47] | [0.20, 0.47] | [0.59, 0.81] | [0.65, 0.79] | [-0.01, 0.03] | [0.24, 0.34] | [0.27, 0.40] | [0.69, 0.82] | [0.75, 0.81] |

925

**Table 4: Performance metric ranges ([min, max]) for all 24 reservoirs of BMA methods under S3 and S5.**

| Period | Forecast horizon (d) | NSE | | RMSE($m^3/s$) | | MAE($m^3/s$) | |
|---|---|---|---|---|---|---|---|
| | | S3 | S5 | S3 | S5 | S3 | S5 |
| Calibration | 1 | [0.60, 0.98] | [0.97, 0.99] | [0.02, 0.36] | [0.01, 0.09] | [0.01, 0.09] | [0.01, 0.03] |
| | 2 | [0.36, 0.92] | [0.96, 0.99] | [0.04, 0.46] | [0.02, 0.10] | [0.02, 0.13] | [0.01, 0.04] |
| | 3 | [0.14, 0.63] | [0.95, 0.98] | [0.09, 0.53] | [0.02, 0.11] | [0.03, 0.15] | [0.01, 0.04] |
| | 4 | [0.12, 0.57] | [0.94, 0.98] | [0.10, 0.54] | [0.02, 0.12] | [0.04, 0.17] | [0.01, 0.04] |
| | 5 | [0.04, 0.23] | [0.95, 0.98] | [0.14, 0.56] | [0.02, 0.12] | [0.04, 0.17] | [0.01, 0.04] |
| | 6 | [0.07, 0.46] | [0.94, 0.98] | [0.11, 0.55] | [0.02, 0.10] | [0.04, 0.17] | [0.01, 0.04] |
| | 7 | [0.06, 0.25] | [0.94, 0.97] | [0.13, 0.52] | [0.03, 0.11] | [0.05, 0.15] | [0.01, 0.04] |
| Validation | 1 | [0.60, 0.92] | [0.84, 0.96] | [0.08, 0.66] | [0.06, 0.39] | [0.02, 0.13] | [0.02, 0.10] |
| | 2 | [0.07, 0.93] | [0.80, 0.96] | [0.07, 1.09] | [0.06, 0.33] | [0.03, 0.19] | [0.02, 0.09] |
| | 3 | [0.06, 0.90] | [0.82, 0.95] | [0.09, 1.09] | [0.06, 0.30] | [0.04, 0.21] | [0.02, 0.09] |
| | 4 | [0.08, 0.91] | [0.85, 0.96] | [0.08, 1.07] | [0.06, 0.34] | [0.04, 0.19] | [0.02, 0.09] |
| | 5 | [0.09, 0.85] | [0.85, 0.96] | [0.11, 1.08] | [0.05, 0.29] | [0.05, 0.22] | [0.02, 0.09] |
| | 6 | [0.06, 0.83] | [0.86, 0.95] | [0.11, 1.08] | [0.06, 0.34] | [0.05, 0.21] | [0.03, 0.09] |
| | 7 | [0.04, 0.82] | [0.87, 0.96] | [0.12, 1.10] | [0.06, 0.35] | [0.05, 0.22] | [0.02, 0.10] |
| Test | 1 | [0.60, 0.89] | [0.76, 0.89] | [0.08, 0.68] | [0.08, 0.47] | [0.03, 0.20] | [0.03, 0.16] |
| | 2 | [0.05, 0.76] | [0.68, 0.87] | [0.12, 1.05] | [0.09, 0.50] | [0.05, 0.27] | [0.04, 0.16] |
| | 3 | [0.59, 0.73] | [0.68, 0.83] | [0.13, 0.69] | [0.10, 0.53] | [0.05, 0.23] | [0.04, 0.16] |
| | 4 | [0.03, 0.18] | [0.69, 0.83] | [0.22, 1.06] | [0.10, 0.54] | [0.08, 0.29] | [0.04, 0.16] |
| | 5 | [0.01, 0.21] | [0.68, 0.81] | [0.21, 1.08] | [0.11, 0.51] | [0.08, 0.30] | [0.04, 0.16] |
| | 6 | [0.02, 0.09] | [0.64, 0.81] | [0.23, 1.07] | [0.11, 0.63] | [0.09, 0.30] | [0.04, 0.18] |
| | 7 | [0.25, 0.43] | [0.67, 0.80] | [0.19, 0.84] | [0.12, 0.55] | [0.09, 0.28] | [0.05, 0.16] |

**Table 5: Ranges of interval performance metrics ([min, max]) for all the 24 reservoirs under S3 and S5.**

| Period | Forecast horizon | *Cr*(%) | | *D*($m^3/s$) | |
|---|---|---|---|---|---|
| | (d) | S3 | S5 | S3 | S5 |
| Calibration | 1 | [94.36, 99.86] | [96.29, 99.86] | [0.01, 0.06] | [0.01, 0.03] |
| | 2 | [93.67, 99.17] | [94.91, 99.66] | [0.01, 0.05] | [0.01, 0.03] |
| | 3 | [94.36, 98.21] | [95.05, 99.59] | [0.01, 0.04] | [0.01, 0.03] |
| | 4 | [92.98, 96.97] | [95.67, 99.72] | [0.02, 0.04] | [0.01, 0.04] |
| | 5 | [93.26, 96.22] | [94.57, 99.79] | [0.01, 0.04] | [0.01, 0.04] |
| | 6 | [93.26, 96.70] | [95.74, 99.66] | [0.02, 0.05] | [0.01, 0.04] |
| | 7 | [92.64, 96.15] | [95.53, 99.72] | [0.02, 0.05] | [0.01, 0.03] |
| Validation | 1 | [92.88, 99.73] | [96.44, 100.00] | [0.01, 0.05] | [0.01, 0.03] |
| | 2 | [94.25, 99.45] | [93.97, 100.00] | [0.01, 0.05] | [0.01, 0.03] |
| | 3 | [92.33, 97.26] | [94.52, 100.00] | [0.01, 0.02] | [0.01, 0.04] |
| | 4 | [92.60, 96.16] | [93.42, 99.73] | [0.01, 0.06] | [0.01, 0.04] |
| | 5 | [91.78, 94.79] | [95.07, 100.00] | [0.01, 0.04] | [0.01, 0.04] |
| | 6 | [91.78, 94.79] | [95.07, 100.00] | [0.01, 0.04] | [0.01, 0.04] |
| | 7 | [90.68, 93.42] | [93.70, 99.73] | [0.00, 0.03] | [0.01, 0.03] |
| Test | 1 | [90.83, 99.32] | [93.84, 99.73] | [0.03, 0.22] | [0.03, 0.15] |
| | 2 | [92.75, 97.95] | [94.25, 99.73] | [0.04, 0.26] | [0.03, 0.16] |
| | 3 | [92.48, 97.40] | [94.39, 99.73] | [0.05, 0.26] | [0.03, 0.15] |
| | 4 | [91.66, 95.35] | [94.39, 99.59] | [0.07, 0.28] | [0.04, 0.16] |
| | 5 | [90.70, 94.12] | [94.66, 99.45] | [0.07, 0.29] | [0.04, 0.15] |
| | 6 | [90.83, 93.98] | [93.43, 99.45] | [0.08, 0.29] | [0.05, 0.18] |
| | 7 | [89.88, 92.48] | [93.57, 99.45] | [0.08, 0.28] | [0.04, 0.16] |

930 **Table 6: Annual system performance using forecast inflow information.**

| Inflow Configuration | | Forecast horizon (d) | Revenues ($10^4$RMB) | Costs ($10^4$RMB) | Net costs ($10^4$RMB) | Reliability (%) | |
|---|---|---|---|---|---|---|---|
| | | | | | | Daobei | Others |
| Observation | | 1 | 3228.38 | 3336.52 | 108.15 | 79.22 | 86.10 |
| | | 7 | 2651.01 | 2822.95 | 171.94 | 71.31 | 75.94 |
| Deterministic | S3 | 1 | 3541.27 | 3633.50 | 92.23 | 79.68 | 96.90 |
| | S5 | 1 | 3596.23 | 3690.32 | 94.09 | 79.59 | 96.31 |
| | S3 | 7 | 3262.51 | 3401.53 | 139.02 | 80.93 | 91.84 |
| | S5 | 7 | 2945.42 | 3118.95 | 173.53 | 76.10 | 85.27 |
| Uncertain | S3 | 1 | 3931.58 | 3837.88 | -93.70 | 93.49 | 99.80 |
| | S5 | 1 | 3988.70 | 3791.54 | -197.16 | 92.63 | 99.58 |
| | S3 | 7 | 3946.61 | 3902.59 | -44.02 | 94.03 | 100.00 |
| | S5 | 7 | 3911.55 | 3846.44 | -65.11 | 92.83 | 99.38 |

**Table 7: Annual system performance using observed inflow information.**

| Inflow Configurations | | Forecast horizon (d) | Revenues ($10^4$RMB) | Costs ($10^4$RMB) | Net costs ($10^4$RMB) | Reliability (%) | |
|---|---|---|---|---|---|---|---|
| | | | | | | Daobei | Others |
| Observation | | 1 | 3228.38 | 3336.52 | 108.15 | 79.22 | 86.10 |
| | | 7 | 2651.01 | 2822.95 | 171.94 | 71.31 | 75.94 |
| Deterministic | S3 | 1 | 2597.17 | 2783.31 | 186.14 | 67.42 | 79.10 |
| | S5 | 1 | 2735.64 | 2884.66 | 149.02 | 71.79 | 82.00 |
| | S3 | 7 | 2159.05 | 2454.79 | 295.75 | 57.08 | 67.04 |
| | S5 | 7 | 2371.45 | 2631.57 | 260.11 | 64.15 | 71.80 |
| Uncertain | S3 | 1 | 3788.08 | 3820.27 | 32.18 | 94.18 | 98.45 |
| | S5 | 1 | 3805.98 | 3781.46 | -24.52 | 94.42 | 98.38 |
| | S3 | 7 | 3762.07 | 3884.75 | 122.68 | 93.64 | 98.28 |
| | S5 | 7 | 3785.55 | 3835.29 | 49.75 | 94.99 | 98.46 |