IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 2, 2020, accepted August 20, 2020, date of publication August 26, 2020, date of current version September 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3019574

Forecasting Monthly Runoff Time Series by Single-Layer Feedforward Artificial Neural Network and Grey Wolf Optimizer

XIONG CHENG[1], ZHONG-KAI FENG[2], AND WEN-JING NIU[3]

[1]College of Hydraulic and Environmental Engineering, China Three Gorges University, Yichang 443002, China
[2]School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
[3]Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan 430010, China

Corresponding author: Zhong-Kai Feng (myfellow@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0405900, in part by the National Natural Science Foundation of China under Grant 51609124 and Grant 51709119, and in part by the Natural Science Foundation of Hubei Province under Grant 2020CFB340 and Grant 2018CFB573.

ABSTRACT Generally, accurate hydrological forecasting information plays an increasingly important role in promoting the comprehensive benefit of hydropower reservoirs. With satisfying generalization ability and search rate, the extreme learning machine (ELM), a famous single-layer feedforward neural network, has been widely used to address regression and classification problem. However, the standard ELM method often falls into second-best solutions with a high probability due to the random assignments of network parameters. In order to overcome this problem, this article aims at developing a hybrid model for monthly runoff time series forecasting. In the hybrid method, an effective swarm intelligence method, grey wolf optimizer (GWO), is adopted to optimize the input-hidden weights and hidden biases of the ELM method; and then the Moore-Penrose generalized inverse method is adopted to determine the hidden-output weights. The world's largest hydropower reservoir, Three Gorges, is chosen to compare the performances of various forecasting methods. Based on the simulation results, the presented method outperforms several traditional forecasting methods (like artificial neural network and support vector machine) in several quantitative indexes. Thus, a novel alternative is presented to predict the nonlinear hydrological time series in China.

INDEX TERMS Hydrologic forecasting, grey wolf optimizer, extreme learning machine, artificial neural network.

I. INTRODUCTION

Accurate monthly runoff forecasting plays an important role in promoting the scientific management of water resource, like flood control [1], power generation [2]–[4], environmental protection [5]–[7], operational rule [8]–[10], peak shaving [11] and multi-energy complement scheduling [12]–[14]. In the past decades, a variety of streamflow forecasting methods have been successfully proposed by scholars all over the world [15]–[18]. From the viewpoint of model mechanism, the existing runoff forecasting methods can be roughly divided into two different kinds of groups: theory-based method and data-based method [19]–[21]. Generally, the theory-based method uses some well-designed equations to reflect the complex runoff formation process in nature, but may fail to accurately mimic the nonlinear streamflow in some cases, especially as the observed runoff data is not available [22]–[24]. Besides, without knowing the exact priori input-output formula of the target problems, the data-based method can provide satisfying runoff forecasting results by iteratively learning implicit knowledge from a large number of data samples, promoting its widespread applications in various engineering problems [25]–[28]. Based on these considerations, this article focuses on the development of the data-based methods for monthly runoff series forecasting.

As one of the most classical data-based methods, artificial neural network based on gradient training technique has been widely used in hydrologic forecasting due to its unique advantages of strong flexibility and mapping ability [29]–[31]. However, the applications of the traditional gradient training methods are often limited by some defects, like local optima

The associate editor coordinating the review of this manuscript and approving it for publication was Nadeem Iqbal.

and slow convergence [32]–[34]. To alleviate this problem, the novel extreme learning machine (ELM) is proposed in recent years [35]–[37]. In ELM, the weight vectors linking input layer and hidden layer and bias of hidden layer are randomly placed in the preset state space, and then the Moore-Penrose generalized inverse method is employed to analytically solve the weight vectors linking hidden layer and output layer. Without the time-consuming iterative search, the ELM method can immediately finish its parameter learning process and produce satisfying performance. In other words, the ELM method has the merits of faster learning speed, stronger generalization capability and less execution parameters compared with the standard gradient-based training methods. As a result, the applications of the ELM-based models are becoming an increasingly popular research topic in a variety of engineering problems [37]–[39].

Nevertheless, it was found that the standard ELM method easily falls into local minimum, and there are certain rooms to improve the ELM performance in practice [40]. Based on the existing literatures, the swarm intelligence method proves to be an effective tool to optimize the model parameters of the artificial neural network [41], [42]. Inspired by the hunting behavior and coordination mechanism of the grey wolves, a novel swarm intelligence method called grey wolf optimizer (GWO) is developed in recent years [43]–[45]. With satisfying search ability, GWO has been successfully used in many research fields, like reservoir operation, imagine recognition and multi-energy complementation. However, up to this day, there are few publications about using the GWO method to optimize the ELM parameters in monthly runoff forecasting. For the sake of refilling this gap, this study proposes a practical hybrid ELM-GWO model based on the ELM and GWO method for monthly runoff forecasting. Specially, in the presented method, the GWO method is adopted to search for satisfying parameters of the ELM method, while the Moore-Penrose generalized inverse method is used to solve the hidden-output weights in an analytical way. The results from the world's largest hydropower project demonstrate that the GWO method can effectively improve the predetermined parameters of the ELM method, while the ELM-GWO method can produce better forecasting results than several traditional methods in terms of all the statistical indexes. Thus, this article provides an effective alternative for monthly runoff prediction under the changing environment.

The rest of this article is organized as below: the details of the hybrid method are given in Section II; in Section III, the proposed method is used to forecast the monthly runoff of a real-world reservoir and the conclusions are given in the end.

II. METHODS

A. EXTREME LEARNING MACHINE (ELM)

As a typical single-layer feedforward neural network, the extreme learning machine (ELM) is a novel learning tool developed to solve the defects of gradient-based method, like slow convergence and parameter tuning [40]. In theory, it has been successfully proved that when the input-hidden weights and hidden biases are randomly assigned while the mapping functions of all the hidden neurons are infinitely continuously differentiable, the ELM model will become a typical linear system where the hidden-output weight vectors are directly deduced by using the generalized inverse method. Then, the ELM method is able to possess good generalization ability while avoiding stopping condition and training epochs.

Without loss of generality, it is assumed that the standard ELM model in Fig. 1 contains three layers: one n-node input layer, one L-node hidden layer, and one m-node output layer. Then, the ELM method can approximate N data samples with zero error, which can be mathematically described as below

$$\boldsymbol{t}_i = \sum_{l=1}^{L} \boldsymbol{\beta}_l \cdot g\left(\boldsymbol{w}_l \cdot \boldsymbol{q}_i + b_l\right), \quad i = 1, 2, \cdots, N \tag{1}$$

where $\boldsymbol{q}_i$ and $\boldsymbol{t}_i$ are the input vector and output vector in the ith training sample. $\boldsymbol{\beta}_l$ is the weight vector linking the lth hidden nodes and the output layer. $\boldsymbol{\omega}_l$ is the weight vector linking the lth hidden node and the input layer. b_l and $g(\cdot)$ denote the threshold and nonlinear mapping function of the lth hidden nodes.

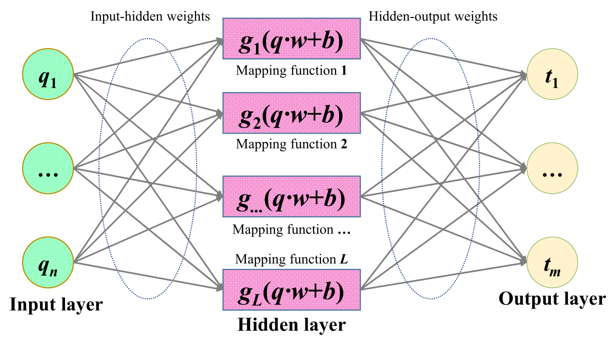

FIGURE 1. Sketch map of the ELM method.

By this time, the above equation can be rewritten as below:

$$\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{T} \tag{2}$$

where

$$\boldsymbol{H} = \begin{bmatrix} g\left(\boldsymbol{w}_1 \cdot \boldsymbol{q}_1 + b_l\right) & \cdots & g\left(\boldsymbol{w}_L \cdot \boldsymbol{q}_1 + b_L\right) \\ \vdots & \cdots & \vdots \\ g\left(\boldsymbol{w}_1 \cdot \boldsymbol{q}_N + b_l\right) & \cdots & g\left(\boldsymbol{w}_L \cdot \boldsymbol{q}_N + b_L\right) \end{bmatrix}_{N \times L} \tag{3}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{1,1} & \cdots & \beta_{1,m} \\ \vdots & \cdots & \vdots \\ \beta_{L,1} & \cdots & \beta_{L,m} \end{bmatrix}_{L \times m} \tag{4}$$

$$\boldsymbol{T} = \begin{bmatrix} t_{1,1} & \cdots & t_{1,m} \\ \vdots & \cdots & \vdots \\ t_{N,1} & \cdots & t_{N,m} \end{bmatrix}_{N \times m} \tag{5}$$

where $\boldsymbol{H}$ denotes the output matrix of the hidden layer. $\boldsymbol{\beta}$ denotes the weight matrix linking the hidden layer and output layer in ELM. $\boldsymbol{T}$ is the matrix of all samples' output.

When $\boldsymbol{H}$ and $\boldsymbol{T}$ are respectively seen as the independent variables and dependent variables, $\boldsymbol{\beta}$ is equivalent to the coefficients to be optimized and then the problem in Eq. (2) will be converted to a typical linear system. Thus, the weight matrix linking the hidden layer and the output layer can be directly obtained by seeking out the least-square solution of the above linear system, which can be expresses as below:

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{H}^{\dagger}\boldsymbol{T} \tag{6}$$

where $\boldsymbol{H}^{\dagger}$ is the Moore–Penrose generalized inverse matrix with respect to $\boldsymbol{H}$.

B. GREY WOLF OPTIMIZER (GWO)

Grey wolf optimizer (GWO) based on the leadership and hunting mechanism of grey wolves is a novel meta-heuristic method for solving global optimization problems [46]–[48]. As illustrated in Fig. 2, a number of grey wolves living together obey the strict social hierarchy. To effectively hunt the preys, the wolves in the pack are usually divided into four different classes. The first class only contains a leader called alpha wolf (a) which is in charge of making decisions of the pack. The second class contains the coleader called beta wolf that helps alpha wolf make decision and deliver the information to other wolves. The third class contains the henchman called delta wolf (δ) who is under the guidance of both alpha and beta wolves. The remaining wolves at the lowest level in the population are called omegas wolves (ω) that take charge of some important tasks (like detection, protection and feeding).

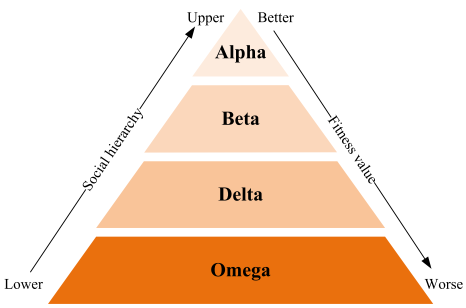

FIGURE 2. The leadership structure of grey wolf.

Based on the above analysis about social behaviors of grey wolves, three elite solutions having the best, second-best and third-best performances are considered as the alpha, beta and delta wolves, while the remaining solutions are considered as omega wolves. All the wolves iteratively improve their performances by obeying the orders of three leader wolves.

Firstly, the wolves' encircling behavior around the prey can be expressed as below:

$$\boldsymbol{X}^{k+1} = \boldsymbol{X}_p^k - \boldsymbol{\lambda} \cdot \boldsymbol{D} \tag{7}$$

$$\boldsymbol{D} = \left|\boldsymbol{C} \circ \boldsymbol{X}_p^k - \boldsymbol{X}^k\right| \tag{8}$$

$$\boldsymbol{\lambda} = 2\boldsymbol{a} \circ \boldsymbol{r}_1 - \boldsymbol{a} \tag{9}$$

$$\boldsymbol{C} = 2\boldsymbol{r}_2 \tag{10}$$

where $\circ$ is the entry wise product of two vectors. $\boldsymbol{D}$ is the absolute value between two vectors. $\boldsymbol{X}_p^k$ is the position vector of the prey at the kth cycle. $\boldsymbol{X}^k$ is the position vector of the wolf at the kth cycle. $\boldsymbol{\lambda}$ and $\boldsymbol{C}$ are two coefficient vectors. r_1 and r_2 are two random vectors whose elements are randomly distributed in the range of 0 and 1. $\boldsymbol{a}$ is the coefficient vector whose jth element (a_j) is linearly reduced by the following equation

$$a_j = a_0 \cdot (1 - \frac{k}{\bar{k}}) \tag{11}$$

where a_0 is the initial value. $\bar{k}$ is the maximum iteration.

Secondly, the hunting behavior of grey wolves is executed by gradually approximating the prey positions under the leadership of three elite solutions. Then, the position of each wolf is dynamically updated by the following equation:

$$\boldsymbol{X}_1' = \boldsymbol{X}_\alpha - \boldsymbol{A}_1 \circ \boldsymbol{D}_\alpha \tag{12}$$

$$\boldsymbol{X}_2' = \boldsymbol{X}_\beta - \boldsymbol{A}_2 \circ \boldsymbol{D}_\beta \tag{13}$$

$$\boldsymbol{X}_3' = \boldsymbol{X}_\delta - \boldsymbol{A}_3 \circ \boldsymbol{D}_\delta \tag{14}$$

$$\boldsymbol{X}^{k+1} = \frac{(\boldsymbol{X}_1' + \boldsymbol{X}_2' + \boldsymbol{X}_3')}{3} \tag{15}$$

where $\boldsymbol{X}_\alpha$, $\boldsymbol{X}_\beta$ and $\boldsymbol{X}_\delta$ denotes the position vectors of the alpha, beta and delta wolves.

From the above analysis, it can be obviously found that the wolf pack will lead off an attack and finish the hunting task when the prey comes to rest. In the search process, all wolves will enhance the global exploration in the entire state space at the early evolutionary stage, while the local exploitation will be improved with the increasing number of iterations. In this way, the swarm can gradually converge to promising areas of the complex global optimization problem [49]–[51].

C. THE HYBRID ELM-GWO METHOD FOR MONTHLY RUNOFF SERIES FORECASTING

Different from the gradient-based method, the ELM model can maintain the generalization ability while sharply reduce the execution efficiency by avoiding the iterative search and parameter setting, promoting its wide applications in many engineering problems. However, because the preset network parameters often remain unchanged during the training stage, some superfluous or inapposite hidden biases and input-hidden weights may produce negative influences on the forecasting accuracy of the ELM model in practice. In order to improve the ELM performance, this article presents a hybrid forecasting method called extreme learning machine based on grey wolf optimizer (ELM-GWO for short). In the ELM-GWO method, the GWO approached is employed to determine the optimal parameters at the learning stage, while the deduced ELM method is employed for operational forecasting. In this way, the dynamic combinations of parameter optimization and analytic optimization can help produce an

ELM model with more compact network structure than its original version.

Generally, a common way is to forecast 1-step-ahead monthly runoff with the aid of n input variables. In other words, only 1 output node and n input nodes should be considered. As the number of hidden nodes is set as L, the structure of the ELM model will become n-L-1. It should be pointed out that all the input data should be normalized into the range of 0 and 1 before starting the simulations, while the outputs of the ELM model should be renormalized to the original range of the monthly runoff time series. Then, the execution procedures of the ELM model are given as below:

Step 1: Set the calculating parameters of the ELM-GWO approach, including the number of iterations K and wolves I in GWO, and the nonlinear mapping functions of the hidden nodes in ELM. Here, the classical sigmoid function is chosen as the transfer function, which is given as below:

$$g[x] = \frac{1}{1+e^{-x}} \tag{16}$$

Step 2: All the samples in the training and testing datasets are normalized to the range of [0,1] by:

$$X_i' = \frac{X_i - X_i^{\min}}{X_i^{\max} - X_i^{\min}} \tag{17}$$

where $X_i^{\max}$ and $X_i^{\min}$ are the maximum and minimum of the ith attribute data. X_i' and X_i are the normalized and original values of the ith attribute data.

Step 3: Set the counter $k = 1$. Then, I wolves are randomly assigned in the feasible state space and each wolf defined in Eq. (18) represents one possible network structure that is composed of the input-hidden weights and hidden biases.

$$\boldsymbol{\theta}_i(k) = \left[\boldsymbol{w}_{1,(i,k)}^T, \cdots, \boldsymbol{w}_{l,(i,k)}^T, \cdots, \boldsymbol{w}_{L,(i,k)}^T, b_{1,(i,k)}, \cdots, b_{l,(i,k)}, \cdots, b_{L,(i,k)}\right] \tag{18}$$

where $\boldsymbol{w}_{l,(i,k)}^T$ is the weight vector linking the lth hidden node and the input layer. b_l is the bias of the lth hidden nodes.

Step 4: Set $k = k + 1$, and then calculate the fitness value of all the wolves in the pack by the following equation:

$$F[\boldsymbol{\theta_i}(k)] = \sqrt{\frac{1}{N}\sum_{s=1}^{L}\left(t_s - \sum_{l=1}^{L}\widehat{\boldsymbol{\beta}}_{l,(i,k)} g\left(\boldsymbol{w}_{l,(i,k)}\cdot \boldsymbol{q}_s + b_{l,(i,k)}\right)\right)^2} \tag{19}$$

$$\widehat{\boldsymbol{\beta}}_{(i,k)} = \boldsymbol{H}_{(i,k)}^{\dagger}\boldsymbol{T} \tag{20}$$

$$\boldsymbol{H}_{(i,k)} = \begin{bmatrix} g(\boldsymbol{w}_{1,(i,k)}\cdot\boldsymbol{q}_1 + b_{l,(i,k)}) & \cdots & g(\boldsymbol{w}_{L,(i,k)}\cdot\boldsymbol{q}_1 + b_{L,(i,k)}) \\ \vdots & \cdots & \vdots \\ g(\boldsymbol{w}_{1,(i,k)}\cdot\boldsymbol{q}_N + b_{l,(i,k)}) & \cdots & g(\boldsymbol{w}_{L,(i,k)}\cdot\boldsymbol{q}_N + b_{L,(i,k)}) \end{bmatrix}_{N\times L} \tag{21}$$

where t_s is the sth output value in the training dataset. N is the size of training dataset. $\boldsymbol{H}_{(i,k)}$ is the output matrix of the hidden-layer associated with the ith wolf, and $\boldsymbol{H}_{(i,k)}^{\dagger}$ is the Moore-Penrose generalized inverse of $\boldsymbol{H}_{(i,k)}$.

Step 5: Update the positions of three leader wolves to calculate the position of all the particles by Eqs. (12)~(15).

Step 6: If $k < K$, go to Step 4 for the next cycle; otherwise, stop the calculation. The alpha wolf is seen as the final preset parameters of the ELM model, and then the Moore-Penrose generalized inverse method is used to determine the hidden-output weights.

III. CASE STUDIES

A. STDUY AREA AND DATASETS

Here, the monthly runoff time series of Three Gorges project located on the mainstream of Yangtze River is used to test the feasibility of the ELM-GWO method. As the world's largest hydropower project, Three Gorges has the installed capacity of 22.5GW and storage volume of 393×10^8 m^3. The catchment area of Three Gorges is about 10000 km^2 and the multi-year average runoff is about 15000 m^3/s. Three Gorges plays an important role in promoting the healthy economic development of China by providing multiple benefits, like flood control, power generation, ecological protection. As a multi-purpose reservoir, in the flood season, the flood control is the main function of Three Gorges; in the dry season, the power generation and shipping become the main function.

The studied monthly streamflow data collected from Three Gorges is from January 1890 to December 2015. According to the previous research results [26], [38], there is no unifying principle to split the training and testing data. In order to ensure the forecasting ability, the representative dataset is often chosen to train the model so that the testing samples can be well considered in the learning process. Then, the proportions of model training in the entire monthly runoff data are about 70%, while the left for testing. From the data in Fig. 3, it can be found that the monthly runoff varies in a relatively large range, demonstrating the complexity of operational predication; while the training data can basically cover the variation amplitude of the testing data, which can effectively guarantee the feasibility of the obtained data.

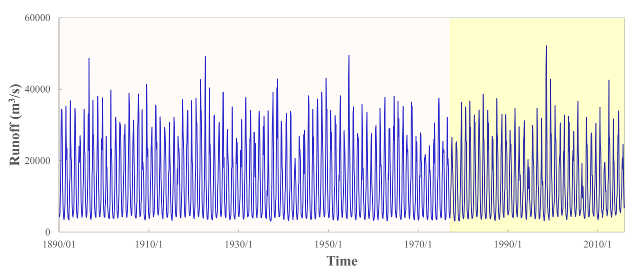

FIGURE 3. Monthly streamflow data from the Three Gorge.

B. EVALUATION INDEXES

In this section, four statistical indexes are introduced to fully compare the performances of various forecasting models, including Coefficient of correlation (R) and Nash-Sutcliffe efficiency (CE), Root mean squared error ($RMSE$), and Mean absolute percentage error (*MAPE*). The *R* index is able to effectively reflect the incidence relation between the original and forecasted data; the *CE* index can evaluate the model's performance keeping away from the mean value; the *RMSE* index can reflect the performance of the forecasting model in large data values; the *MAPE* index is used to evaluate the relative error of the forecasting model. Generally, the larger the values of *R* and *CE* while the smaller the values of the values of *RMSE* and *MAPE*, the better the forecasting model. Then, the detailed computational formulas of four evaluation indexes are given as below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2} \tag{22}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\tilde{y}_i - y_i}{y_i}\right| \times 100\% \tag{23}$$

$$CE = 1 - \frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{\sum_{i=1}^{n}\left(y_i - y_{avg}\right)^2} \tag{24}$$

$$R = \frac{\sum_{i=1}^{n}\left[\left(y_i - y_{avg}\right)\left(\tilde{y}_i - \tilde{y}_{avg}\right)\right]}{\sqrt{\sum_{i=1}^{n}\left(y_i - y_{avg}\right)^2\left(\tilde{y}_i - \tilde{y}_{avg}\right)^2}} \tag{25}$$

where y_{avg} and $\tilde{y}_{avg}$ are the average of all the observed and predicted value. y_i and $\tilde{y}_i$ are the *i*th observed and predicted data.

C. INPUT VARIALBE SELECTION

The set of input variables often have a great influence on the model performance. Up to this day, there is no widely accepted methods for input variable determination. In order to produce satisfying forecasting model, 10 different input variable combinations in Table 1 are designed, where the $x(t\text{-}d)$ denotes the $t\text{-}d$th value in the target time series, while S_i denotes the set of input variable considered in the *i*th case. Besides, considering the hidden layer has a direct influence on the model performance, three different model structure with various number of hidden nodes ($2d$-1, $2d$ and $2d + 1$) are designed. Obviously, 30 models will be developed for the ANN-based forecasting methods, demonstrating the amount of workload in selecting the suitable forecasting model.

D. FORECASTING MODEL DEVELOPEMT

1) ELM MODEL

For the ELM model, the input-hidden weights and hidden biases are randomly determined while the Moore-Penrose generalized inverse method is used to find the hidden-output weights. The famous sigmoid function is set as the transfer function of the hidden nodes, while the best model is chosen from all the developed ELM models.

TABLE 1. Different models for the three gorge project.

Model name	Input variables	Output	Input number
M1	S_1={$x(t$-1), $x(t$-2)}	$x(t)$	2
M2	$S_2 = S_1 \cup x(t\text{-}3)$	$x(t)$	3
M3	$S_3 = S_2 \cup x(t\text{-}4)$	$x(t)$	4
M4	$S_4 = S_3 \cup x(t\text{-}5)$	$x(t)$	5
M5	$S_5 = S_4 \cup x(t\text{-}6)$	$x(t)$	6
M6	$S_6 = S_5 \cup x(t\text{-}7)$	$x(t)$	7
M7	$S_7 = S_6 \cup x(t\text{-}8)$	$x(t)$	8
M8	$S_8 = S_7 \cup x(t\text{-}9)$	$x(t)$	9
M9	$S_9 = S_8 \cup x(t\text{-}10)$	$x(t)$	10
M10	$S_{10} = S_9 \cup x(t\text{-}11)$	$x(t)$	11

The statistical indexes of the forecasting results obtained by the ELM method are listed in Table 2. It can be found that for a model with the same input variables and output variable, the results will change with the varying number of hidden nodes. For instance, the model M2 with $2d + 1$ hidden nodes achieves the best performances at both training and testing phases; for the mode M7 (the name of the 7th model in Table 1), the best values of four indexes at the training phase is achieved by the model with $2d$ hidden nodes, while the testing phase is achieved by the model with $2d + 1$ hidden nodes. Besides, the performance of model M1 is inferior to other models, which means that the forecasting results may be unsatisfying due to the unsuitable selections of input variables. To guarantee the ELM performance, the statistical indexes of all the models are fully compared. Based on the comprehensive analysis, the model M8 with the 9-19-1 structure is chosen to forecast the monthly runoff of Three Gorge Project due to its satisfying performances at both training and testing phases.

2) ELM-GWO MODEL

For the ELM-GWO model, the grey wolf optimizer and Moore-Penrose generalized inverse method are used to find the suitable model parameters. In the training process, the sigmoid function is chosen as the transfer function of hidden layer, while the number of wolves and iterations are set as 50 and 1000, respectively.

The statistical indexes of the forecasting results obtained by the ELM-GWO method are listed in Table 3. It is also showed that the model structure has a great influence on the performance of the ELM-GWO method. For instance, the model M8 with $2d + 1$ hidden nodes can produce the best performance at the training phase while the best values of *MAPE* and *CE* at the testing phase; the model M9 with $2d$ hidden nodes also achieve the best performance at the training phase while the best *R* value at the testing phase. Besides, like the ELM model, the ELM-GWO model tends to yield unsatisfying forecasting results, demonstrating the importance of input variables and model parameters; the ELM-GWO method under the same input variables often is better than the standard ELM method, demonstrating the effectiveness of the GWO method in optimizing model parameters. Then, to ensure the ELM-GWO performance, the statistical indexes

TABLE 2. Results of various ELM models for the three gorge project.

No.	Structure	Model	Training				Testing			
			RMSE	*MAPE*	*R*	*CE*	*RMSE*	*MAPE*	*R*	*CE*
1	2-3-1	M1	7742.0259	64.2650	0.7812	0.4386	7099.8627	61.8648	0.7660	0.4398
2	2-4-1		6532.1009	40.8274	0.7768	0.6004	6191.3238	38.9686	0.7579	0.5740
3	2-5-1		6474.4799	39.9000	0.7811	0.6074	6122.1270	38.1225	0.7642	0.5835
4	3-5-1	M2	5923.4207	36.0589	0.8251	0.6714	5730.8022	35.9796	0.7980	0.6350
5	3-6-1		6026.5539	36.1139	0.8160	0.6598	5813.6755	35.7132	0.7907	0.6244
6	3-7-1		5619.5868	30.1111	0.8428	0.7042	5389.6786	30.2391	0.8231	0.6772
7	4-7-1	M3	6017.0272	38.2357	0.8278	0.6609	5604.6067	38.5444	0.8170	0.6509
8	4-8-1		5458.3671	28.1482	0.8536	0.7209	5205.6135	28.6205	0.8364	0.6989
9	4-9-1		5541.7865	30.1934	0.8488	0.7124	5343.8418	30.8288	0.8268	0.6827
10	5-9-1	M4	5810.4113	37.7552	0.8409	0.6838	5617.7499	38.2230	0.8116	0.6493
11	5-10-1		5180.0363	27.5522	0.8707	0.7487	4944.8358	27.5779	0.8544	0.7283
12	5-11-1		5106.2783	26.0365	0.8749	0.7558	4871.4665	26.6528	0.8593	0.7363
13	6-11-1	M5	5436.4350	28.0625	0.8560	0.7232	5264.6900	30.0070	0.8333	0.6920
14	6-12-1		5071.5692	26.0906	0.8770	0.7591	4942.0930	26.8657	0.8543	0.7286
15	6-13-1		5007.6970	25.2155	0.8805	0.7651	4859.5344	25.3074	0.8595	0.7376
16	7-13-1	M6	5468.6856	32.5433	0.8558	0.7199	5236.7257	34.4540	0.8370	0.6953
17	7-14-1		5154.5654	27.5199	0.8723	0.7511	4961.7259	28.3769	0.8536	0.7264
18	7-15-1		5002.4813	25.8900	0.8808	0.7656	4793.5920	25.2661	0.8639	0.7447
19	8-15-1	M7	5314.4124	30.9617	0.8643	0.7355	5147.1378	32.0266	0.8432	0.7056
20	8-16-1		5035.6956	26.0458	0.8791	0.7625	4858.1017	26.1027	0.8600	0.7377
21	8-17-1		5046.2822	26.2607	0.8784	0.7615	4785.5583	25.5375	0.8648	0.7455
22	9-17-1	M8	5337.8533	30.9642	0.8621	0.7331	5012.3992	31.9468	0.8532	0.7208
23	9-18-1		5137.5735	27.4728	0.8734	0.7528	4872.3754	28.0109	0.8601	0.7362
24	**9-19-1**		**4984.3275**	**26.1877**	**0.8819**	**0.7673**	**4785.4143**	**25.7889**	**0.8645**	**0.7455**
25	10-19-1	M9	5202.4272	28.3787	0.8698	0.7465	4902.4738	28.9888	0.8593	0.7329
26	10-20-1		5016.3642	26.6114	0.8803	0.7643	4820.1503	26.3813	0.8629	0.7418
27	10-21-1		4983.4339	26.2571	0.8823	0.7674	4810.1408	25.7830	0.8633	0.7429
28	11-21-1	M10	5148.1188	27.0045	0.8730	0.7518	4983.0512	28.6666	0.8529	0.7241
29	11-22-1		4998.1675	26.1875	0.8815	0.7660	4876.6580	26.2432	0.8594	0.7357
30	11-23-1		4939.1440	25.7312	0.8846	0.7715	4789.3600	25.2862	0.8650	0.7451

Note: Bold denote the selected ELM model.

of all the models at both training and testing phase are compared. Based on the comprehensive analysis, due to its satisfying performance, the model M6 with the structure of 7-15-1 is chosen to forecast the monthly runoff of the Three Gorge Project.

3) ANN MODEL

For the ANN model, the trial-and-error method is used to choose the best one from 30 forecasting models, while the classical back-propagation method is chosen to optimize the network parameters. Besides, the famous sigmoid function is chosen as the transfer function of hidden nodes. Based on four statistical indexes, 30 different ANN models are used to select the suitable model for operational forecasting in Three Gorge Project.

4) SVM MODEL

For the SVM model, the kernel function is set as the famous radial basis function, while the grid search strategy is used to find feasible parameters. After comparative analysis, the SVM model obtained in the 8^{th} run is selected as the best one for monthly runoff forecasting of Three Gorge Project.

E. RESULTS COMPARISON

Table 4 lists the statistical results of four developed forecasting modes during the training and testing stages. It can be found that the traditional ANN model provides the worst results among all the considered models, which proves the modeling difficulty of nonlinear monthly streamflow time series. Besides, three other methods produce different results while the results of the standard ELM and SVM model are obviously inferior to the proposed ELM-GWO model at both training and testing stage, demonstrating the feasibility of swarm intelligence in optimizing the model parameters. For instance, compared with the ANN, ELM and SVM models at the training phase, the ELM-GWO method can make about 25.6%, 20.1% and 19.8% improvements in the *RMSE* value, and about 37.9%, 33.7% and 37.7% improvements in the *MAPE* value; at the testing phase, the improvements in the *R* value are close to about 7.4%, 5.9% and 6.4% while the improvements in the *CE* value are about 15.0%,

TABLE 3. Results of various ELM-GWO models for the three gorge project.

No.	Structure	Model	Training				Testing			
			RMSE	*MAPE*	*R*	*CE*	*RMSE*	*MAPE*	*R*	*CE*
1	2-3-1	M1	6525.7713	38.3958	0.7812	0.6011	6131.4397	36.6127	0.7651	0.5822
2	2-4-1		6452.3759	41.9975	0.7811	0.6101	6143.9463	40.3853	0.7639	0.5805
3	2-5-1		6450.7057	42.0594	0.7812	0.6103	6129.4885	40.5049	0.7652	0.5825
4	3-5-1	M2	5315.1367	27.4637	0.8576	0.7354	5385.6269	29.8771	0.8285	0.6777
5	3-6-1		5148.2898	25.5510	0.8670	0.7518	5093.0513	28.0260	0.8474	0.7118
6	3-7-1		5093.3992	25.9293	0.8701	0.7570	5204.5377	28.8328	0.8402	0.6990
7	4-7-1	M3	4663.7456	24.5832	0.8923	0.7963	4510.6809	26.6789	0.8834	0.7739
8	4-8-1		4554.0802	22.4873	0.8976	0.8058	4392.5150	25.0398	0.8896	0.7856
9	4-9-1		4511.5822	21.7802	0.8996	0.8094	4472.9404	24.7045	0.8856	0.7777
10	5-9-1	M4	4365.2000	22.5944	0.9064	0.8215	4326.9608	24.4761	0.8953	0.7919
11	5-10-1		4181.9580	19.4763	0.9144	0.8362	4047.4582	20.9361	0.9085	0.8180
12	5-11-1		4113.6991	19.4312	0.9173	0.8415	4149.9531	23.2052	0.9051	0.8086
13	6-11-1	M5	4145.9612	19.7035	0.9160	0.8390	4156.0809	22.0796	0.9050	0.8081
14	6-12-1		4061.0897	18.5193	0.9195	0.8455	4067.7997	20.8259	0.9087	0.8161
15	6-13-1		4019.6021	17.4471	0.9212	0.8487	4077.4066	20.7363	0.9083	0.8153
16	7-13-1	M6	4161.2520	19.2345	0.9153	0.8378	4153.5186	21.9620	0.9051	0.8083
17	7-14-1		4020.4882	17.4240	0.9212	0.8486	4068.2482	20.1581	0.9089	0.8161
18	**7-15-1**		**3980.8747**	**17.3576**	**0.9228**	**0.8516**	**3938.7247**	**21.3592**	**0.9157**	**0.8276**
19	8-15-1	M7	4186.0318	20.0413	0.9143	0.8359	4119.3425	22.6764	0.9070	0.8114
20	8-16-1		4075.9760	18.5290	0.9189	0.8444	4065.1854	20.4455	0.9095	0.8164
21	8-17-1		4002.5047	17.0931	0.9219	0.8500	4115.2876	20.9720	0.9067	0.8118
22	9-17-1	M8	4248.6113	20.3044	0.9116	0.8309	4137.8368	23.2303	0.9065	0.8097
23	9-18-1		4344.9631	20.6032	0.9076	0.8232	4069.7717	23.5754	0.9095	0.8159
24	9-19-1		4228.5578	19.6285	0.9126	0.8325	4074.4737	22.2998	0.9101	0.8155
25	10-19-1	M9	4163.6486	20.3390	0.9153	0.8376	4118.1690	23.2269	0.9055	0.8115
26	10-20-1		4051.6939	18.3642	0.9199	0.8462	4126.7553	21.7332	0.9057	0.8108
27	10-21-1		4413.4717	19.2477	0.9049	0.8176	4122.2413	21.1917	0.9048	0.8112
28	11-21-1	M10	4078.4736	18.8730	0.9188	0.8442	4079.6262	21.1575	0.9059	0.8151
29	11-22-1		3889.9697	17.4514	0.9264	0.8583	4096.7249	21.1372	0.9062	0.8135
30	11-23-1		4024.6053	18.3488	0.9211	0.8483	4131.4912	21.3073	0.9044	0.8103

Note: Bold denote the selected ELM-GWO model.

TABLE 4. Results of various forecasting models for the three gorge project.

No.	Model	Training				Testing			
		RMSE	*MAPE*	*R*	*CE*	*RMSE*	*MAPE*	*R*	*CE*
1	ANN	5349.3581	27.9727	0.8698	0.7320	5020.5173	27.5668	0.8527	0.7199
2	ELM	4984.3275	26.1877	0.8819	0.7673	4785.4143	25.7889	0.8645	0.7455
3	SVM	4961.6625	27.8436	0.8810	0.7694	4856.8091	30.0002	0.8609	0.7379
4	ELM-GWO	3980.8747	17.3576	0.9228	0.8516	3938.7247	21.3592	0.9157	0.8276

11.0% and 12.2%. Thus, the proposed method proves to be an effective method for the monthly hydrological forecasting.

Then, Fig. 4 draws the observed and predicted streamflow data produced by various methods at the testing stage. It can be seen that four forecasting models can produce valid results in capturing the changing tendency of monthly streamflow time series in the Three Gorge, demonstrating the feasibility of the artificial intelligence algorithm. Besides, the developed model is obviously better than three other control models since its tested line is close to the ideal line, demonstrating the superiority of the developed hybrid forecasting model. Thus, the combinations of two effective tools, i.e. ELM and GWO, are capable of producing satisfying forecasting results as used to model the complex runoff process of Three Gorge.

Next, the peak value of the observed and predicted runoff series produced by various methods at the testing stage are drawn in Fig. 5. It can be observed that the ELM-GWO method can generate better results than the control methods. For instance, the proportions of four forecasting models in the mean peak flows are about 33.1%, 29.0%, 28.7% and 14.2%; besides, the ANN, ELM, SVM and hybrid methods make about 46.0%, 41.1%, 41.1% and 24.5% improvements in the maximum peak flow. Thus, it can be concluded that

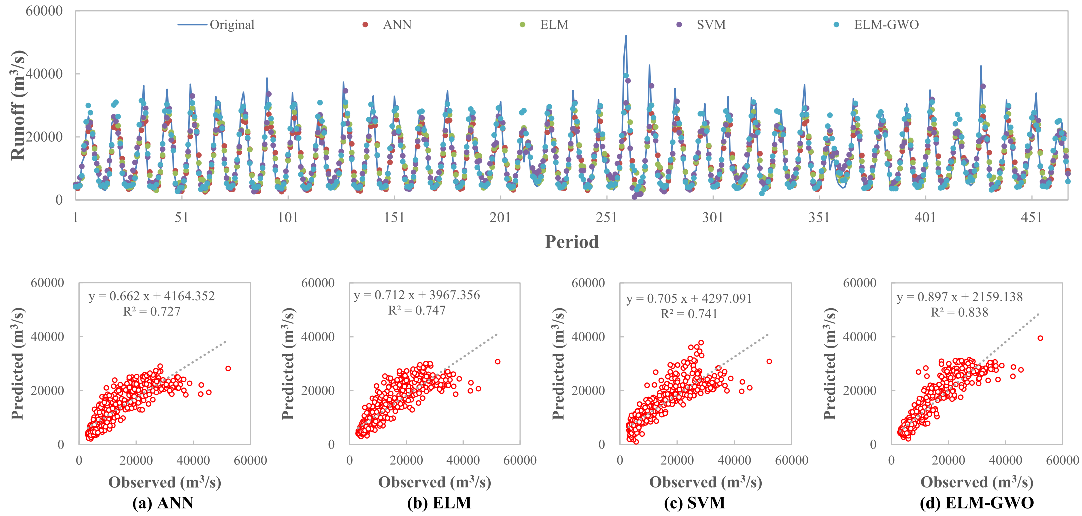

FIGURE 4. Detailed results of various forecasting models for the Three Gorges Project at the testing phase.

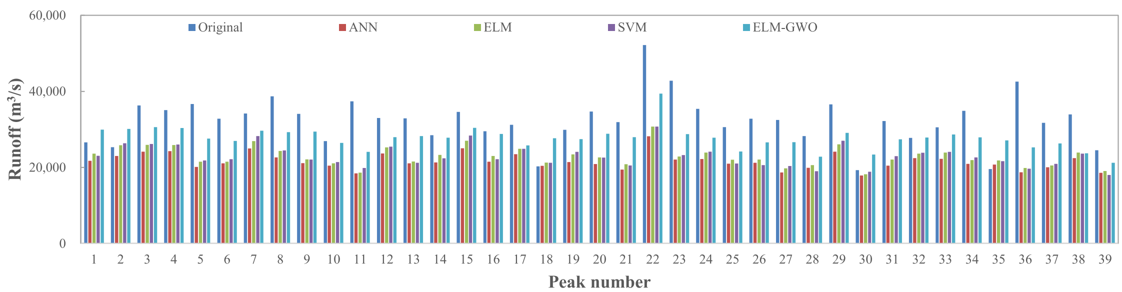

FIGURE 5. Peak flows of various forecasting models for the Three Gorges Project at the testing phase.

the hybrid method can effectively track the variation trend of extreme streamflow data of the Three Gorge Project.

F. DISCUSSIONS

From the above experiments, it can be clearly seen that the hybrid model is better than several control methods. The superior performance of the developed ELM-GWO model is contributed by the dynamic integration of swarm intelligence and artificial neural network. Firstly, the GWO method is employed to determine the suitable input-hidden weights and hidden biases in the ELM network, which can help the model avoid falling into local optima; secondary, the famous Moore–Penrose generalized inverse method is used to seek out the least-squared solution for the linear system, which can determine satisfying hidden-output weights at an express speed. In this way, the computation parameters of the special artificial neural network can be improved by iteratively searching in the problem space. As a result, the carefully-designed ELM model will have a better network structure and then produce higher forecasting precisions in comparison with the randomly determined model, providing a novel alternative tool for the monthly streamflow prediction under the rapidly changing environment. On the other hand, the developed method can provide high-quality prediction results than three traditional forecasting methods, several promising research directions can be considered in the future: one is to test the feasibility of the recent modified GWO versions where the algorithm's swarm diversity and global search ability are enhanced by some hybrid strategies, like parallel computing, local search and chaotic mutations; besides, the other is to design more effective initial modules that can enhance the network compactness of the extreme leaning machine; while the last one is to explore the feasibility of ELM-GWO method in various time-scaled hydrologic forecasting for Three Gorges and other reservoirs around the world.

IV. CONCLUSION

In recent years, growing attention is paid to develop effective methods for accurate hydrologic forecasting. In order to meet the practical necessity, this article successfully develops a hybrid ELM-GWO method to forecast monthly streamflow time series, where the famous grey wolf optimizer (GWO) is employed to determine the optimal parameters combinations (like input-hidden weights and hidden biases) of the classical extreme learning machine (ELM) network. In order to accurately forecast the runoff collected from the world's largest hydropower reservoir, Three Gorges, the proposed method is executed dozens of times to select the best model structure. The simulations show that the presented ELM-GWO method can produce superior results compared with several traditional forecasting methods. For instance, the values of correlation coefficient during the testing period are improved by 7.4%, 5.9% and 6.4% in comparison with the standard artificial neural network, extreme learning machine as well as support vector machine, respectively. To sum up, by using GWO to explore the ELM model parameters, an effective tool is provided to make significant improvements in the forecasting accuracy of long and mid- term runoff.

REFERENCES

[1] A. Mosavi, P. Ozturk, and K.-W. Chau, "Flood prediction using machine learning models: Literature review," *Water*, vol. 10, no. 11, p. 1536, Oct. 2018.

[2] Z.-K. Feng, W.-J. Niu, J.-Z. Zhou, and C.-T. Cheng, "Linking Nelder–Mead simplex direct search method into two-stage progressive optimality algorithm for optimal operation of cascade hydropower reservoirs," *J. Water Res. Planning Manage.*, vol. 146, no. 5, 2020, Art. no. 04020019, doi: 10.1061/(ASCE)WR.1943-5452.0001194.

[3] B. Xu, P. A. Zhong, Y. F. Zhao, Y. Z. Zhu, and G. Q. Zhang, "Comparison between dynamic programming and genetic algorithm for hydro unit economic load dispatch," *Water Sci. Eng.*, vol. 7, no. 4, pp. 420–432, 2014.

[4] Y. Zhang, Z. Jiang, C. Ji, and P. Sun, "Contrastive analysis of three parallel modes in multi-dimensional dynamic programming and its application in cascade reservoirs operation," *J. Hydrol.*, vol. 529, pp. 22–34, Oct. 2015.

[5] Z.-K. Feng, S. Liu, W.-J. Niu, S.-S. Li, H.-J. Wu, and J.-Y. Wang, "Ecological operation of cascade hydropower reservoirs by elite-guide gravitational search algorithm with Lévy flight local search and mutation," *J. Hydrol.*, vol. 581, Feb. 2020, Art. no. 124425, doi: 10.1016/j.jhydrol.2019.124425.

[6] Z. Jiang, P. Liu, C. Ji, H. Zhang, and Y. Chen, "Ecological flow considered multi-objective storage energy operation chart optimization of large-scale mixed reservoirs," *J. Hydrol.*, vol. 577, Oct. 2019, Art. no. 123949, doi: 10.1016/j.jhydrol.2019.123949.

[7] X. Lei, J. Zhang, H. Wang, M. Wang, S.-T. Khu, Z. Li, and Q. Tan, "Deriving mixed reservoir operating rules for flood control based on weighted non-dominated sorting genetic algorithm II," *J. Hydrol.*, vol. 564, pp. 967–983, Sep. 2018.

[8] Z.-K. Feng, W.-J. Niu, R. Zhang, S. Wang, and C.-T. Cheng, "Operation rule derivation of hydropower reservoir by k-means clustering method and extreme learning machine based on particle swarm optimization," *J. Hydrol.*, vol. 576, pp. 229–238, Sep. 2019, doi: 10.1016/j.jhydrol.2019.06.045.

[9] Q.-F. Tan, X. Wang, H. Wang, C. Wang, X.-H. Lei, Y.-S. Xiong, and W. Zhang, "Derivation of optimal joint operating rules for multi-purpose multi-reservoir water-supply system," *J. Hydrol.*, vol. 551, pp. 253–264, Aug. 2017.

[10] F. Zhu, P.-A. Zhong, B. Xu, Y.-N. Wu, and Y. Zhang, "A multi-criteria decision-making model dealing with correlation among criteria for reservoir flood control operation," *J. Hydroinformat.*, vol. 18, no. 3, pp. 531–543, May 2016.

[11] Z.-K. Feng, W.-J. Niu, X. Cheng, J.-Y. Wang, S. Wang, and Z.-G. Song, "An effective three-stage hybrid optimization method for source-network-load power generation of cascade hydropower reservoirs serving multiple interconnected power grids," *J. Cleaner Prod.*, vol. 246, Feb. 2020, Art. no. 119035, doi: 10.1016/j.jclepro.2019.119035.

[12] C. Ma, H. Wang, and J. Lian, "Short-term electricity dispatch optimization of Ertan hydropower plant based on data by field tests," *J. Renew. Sustain. Energy*, vol. 3, no. 6, Nov. 2011, Art. no. 063109.

[13] B. Ming, P. Liu, S. Guo, L. Cheng, Y. Zhou, S. Gao, and H. Li, "Robust hydroelectric unit commitment considering integration of large-scale photovoltaic power: A case study in China," *Appl. Energy*, vol. 228, pp. 1341–1352, Oct. 2018.

[14] H. Zhang, J. Chang, C. Gao, H. Wu, Y. Wang, K. Lei, R. Long, and L. Zhang, "Cascade hydropower plants operation considering comprehensive ecological water demands," *Energy Convers. Manage.*, vol. 180, pp. 119–133, Jan. 2019.

[15] W.-C. Wang, K.-W. Chau, D.-M. Xu, L. Qiu, and C.-C. Liu, "The annual maximum flood peak discharge forecasting using Hermite projection pursuit regression with SSO and LS method," *Water Resour. Manage.*, vol. 31, no. 1, pp. 461–477, Jan. 2017.

[16] W.-C. Wang, K.-W. Chau, D.-M. Xu, and X.-Y. Chen, "Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition," *Water Resour. Manage.*, vol. 29, no. 8, pp. 2655–2675, Jun. 2015.

[17] C. L. Wu and K. W. Chau, "A flood forecasting neural network model with genetic algorithm," *Int. J. Environ. Pollut.*, vol. 28, nos. 3–4, pp. 261–273, 2006.

[18] T. Zhao, Q. J. Wang, A. Schepen, and M. Griffiths, "Ensemble forecasting of monthly and seasonal reference crop evapotranspiration based on global climate model outputs," *Agricult. Forest Meteorol.*, vol. 264, pp. 114–124, Jan. 2019.

[19] Z.-K. Feng, W.-J. Niu, Z.-Y. Tang, Z.-Q. Jiang, Y. Xu, Y. Liu, and H.-R. Zhang, "Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization," *J. Hydrol.*, vol. 583, Apr. 2020, Art. no. 124627, doi: 10.1016/j.jhydrol.2020.124627.

[20] Q.-F. Tan, X.-H. Lei, X. Wang, H. Wang, X. Wen, Y. Ji, and A.-Q. Kang, "An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach," *J. Hydrol.*, vol. 567, pp. 767–780, Dec. 2018.

[21] W. Qi, C. Zhang, G. Fu, H. Zhou, and J. Liu, "Quantifying uncertainties in extreme flood predictions under climate change for a medium-sized basin in northeastern China," *J. Hydrometeorol.*, vol. 17, no. 12, pp. 3099–3112, Dec. 2016.

[22] C. L. Wu, K. W. Chau, and C. Fan, "Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques," *J. Hydrol.*, vol. 389, nos. 1–2, pp. 146–167, Jul. 2010.

[23] C. L. Wu, K. W. Chau, and J. S. Huang, "Modelling coupled water and heat transport in a soil–mulch–plant–atmosphere continuum (SMPAC) system," *Appl. Math. Model.*, vol. 31, no. 2, pp. 152–169, Feb. 2007.

[24] C. L. Wu, K. W. Chau, and Y. S. Li, "Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques," *Water Resour. Res.*, vol. 45, no. 8, pp. 1–24, Aug. 2009.

[25] W. Niu, Z. Feng, Y. Chen, H. Zhang, and C. Cheng, "Annual streamflow time series prediction using extreme learning machine based on gravitational search algorithm and variational mode decomposition," *J. Hydrol. Eng.*, vol. 25, no. 5, 2020, Art. no. 04020008, doi: 10.1061/(ASCE)HE.1943-5584.0001902.

[26] W.-C. Wang, K.-W. Chau, C.-T. Cheng, and L. Qiu, "A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series," *J. Hydrol.*, vol. 374, nos. 3–4, pp. 294–306, Aug. 2009.

[27] W.-C. Wang, K.-W. Chau, L. Qiu, and Y.-B. Chen, "Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition," *Environ. Res.*, vol. 139, pp. 46–54, May 2015.

[28] W.-C. Wang, D.-M. Xu, K.-W. Chau, and S. Chen, "Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD," *J. Hydroinformat.*, vol. 15, no. 4, pp. 1377–1390, Oct. 2013.

[29] S. Wen, M. Dong, Y. Yang, P. Zhou, T. Huang, and Y. Chen, "End-to-end detection-segmentation system for face labeling," *IEEE Trans. Emerg. Topics Comput. Intell.*, early access, Nov. 6, 2019, doi: 10.1109/TETCI.2019.2947319.

[30] S. Wen, W. Liu, Y. Yang, T. Huang, and Z. Zeng, "Generating realistic videos from keyframes with concatenated GANs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 8, pp. 2337–2348, Aug. 2019, doi: 10.1109/TCSVT.2018.2867934.

[31] S. Wen, R. Hu, Y. Yang, T. Huang, Z. Zeng, and Y.-D. Song, "Memristor-based echo state network with online least mean square," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 9, pp. 1787–1796, Sep. 2019, doi: 10.1109/TSMC.2018.2825021.

[32] X. Xie, S. Wen, Z. Zeng, and T. Huang, "Memristor-based circuit implementation of pulse-coupled neural network with dynamical threshold generators," *Neurocomputing*, vol. 284, pp. 10–16, Apr. 2018, doi: 10.1016/j.neucom.2018.01.024.

[33] S. Wen, M. Z. Q. Chen, Z. Zeng, X. Yu, and T. Huang, "Fuzzy control for uncertain vehicle active suspension systems via dynamic sliding-mode approach," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 1, pp. 24–32, Jan. 2017, doi: 10.1109/TSMC.2016.2564930.

[34] S. Wen, T. Huang, X. Yu, M. Z. Q. Chen, and Z. Zeng, "Aperiodic sampled-data sliding-mode control of fuzzy systems with communication delays via the event-triggered method," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 5, pp. 1048–1057, Oct. 2016, doi: 10.1109/TFUZZ.2015.2501412.

[35] R. Taormina and K.-W. Chau, "Data-driven input variable selection for rainfall–runoff modeling using binary-coded particle swarm optimization and extreme learning machines," *J. Hydrol.*, vol. 529, pp. 1617–1632, Oct. 2015.

[36] R. Taormina, K.-W. Chau, and B. Sivakumar, "Neural network river forecasting through baseflow separation and binary-coded swarm optimization," *J. Hydrol.*, vol. 529, pp. 1788–1797, Oct. 2015.

[37] Z. M. Yaseen, S. O. Sulaiman, R. C. Deo, and K.-W. Chau, "An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction," *J. Hydrol.*, vol. 569, pp. 387–408, Feb. 2019.

[38] B. Li and C. Cheng, "Monthly discharge forecasting using wavelet neural networks with extreme learning machine," *Sci. China Technol. Sci.*, vol. 57, no. 12, pp. 2441–2452, Dec. 2014.

[39] R. Barzegar, E. Fijani, A. Asghari Moghaddam, and E. Tziritis, "Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models," *Sci. Total Environ.*, vols. 599–600, pp. 20–31, Dec. 2017.

[40] W.-J. Niu, Z.-K. Feng, C.-T. Cheng, and J.-Z. Zhou, "Forecasting daily runoff by extreme learning machine based on quantum-behaved particle swarm optimization," *J. Hydrol. Eng.*, vol. 23, no. 3, 2018, Art. no. 04018002, doi: 10.1061/(ASCE)HE.1943-5584.0001625.

[41] G.-B. Huang, D. H. Wang, and Y. Lan, "Extreme learning machines: A survey," *Int. J. Mach. Learn. Cybern.*, vol. 2, no. 2, pp. 107–122, Jun. 2011.

[42] F. Zheng, Z. Qi, W. Bi, T. Zhang, T. Yu, and Y. Shao, "Improved understanding on the searching behavior of NSGA-II operators using run-time measure metrics with application to water distribution system design problems," *Water Resour. Manage.*, vol. 31, no. 4, pp. 1121–1138, Mar. 2017.

[43] L. Li, L. Sun, W. Kang, J. Guo, C. Han, and S. Li, "Fuzzy multilevel image thresholding based on modified discrete grey wolf optimizer and local information aggregation," *IEEE Access*, vol. 4, pp. 6438–6450, 2016, doi: 10.1109/ACCESS.2016.2613940.

[44] S. K. Goudos, T. V. Yioultsis, A. D. Boursianis, K. E. Psannis, and K. Siakavara, "Application of new hybrid Jaya grey wolf optimizer to antenna design for 5G communications systems," *IEEE Access*, vol. 7, pp. 71061–71071, 2019, doi: 10.1109/ACCESS.2019.2919116.

[45] M. H. Qais, H. M. Hasanien, and S. Alghuwainem, "A grey wolf optimizer for optimum parameters of multiple PI controllers of a grid-connected PMSG driven by variable speed wind turbine," *IEEE Access*, vol. 6, pp. 44120–44128, 2018, doi: 10.1109/ACCESS.2018.2864303.

[46] A. A. Alomoush, A. A. Alsewari, H. S. Alamri, K. Aloufi, and K. Z. Zamli, "Hybrid harmony search algorithm with grey wolf optimizer and modified opposition-based learning," *IEEE Access*, vol. 7, pp. 68764–68785, 2019, doi: 10.1109/ACCESS.2019.2917803.

[47] P. Hu, S. Chen, H. Huang, G. Zhang, and L. Liu, "Improved alpha-guided grey wolf optimizer," *IEEE Access*, vol. 7, pp. 5421–5437, 2019, doi: 10.1109/ACCESS.2018.2889816.

[48] W. Long, T. Wu, S. Cai, X. Liang, J. Jiao, and M. Xu, "A novel grey wolf optimizer algorithm with refraction learning," *IEEE Access*, vol. 7, pp. 57805–57819, 2019, doi: 10.1109/ACCESS.2019.2910813.

[49] K. Albina and S. G. Lee, "Hybrid stochastic exploration using grey wolf optimizer and coordinated multi-robot exploration algorithms," *IEEE Access*, vol. 7, pp. 14246–14255, 2019, doi: 10.1109/ACCESS.2019.2894524.

[50] L. Zhao and X. Wang, "A deep feature optimization fusion method for extracting bearing degradation features," *IEEE Access*, vol. 6, pp. 19640–19653, 2018, doi: 10.1109/ACCESS.2018.2824352.

[51] Z.-K. Feng, S. Liu, W.-J. Niu, Y. Liu, B. Luo, S.-M. Miao, and S. Wang, "Optimal operation of hydropower system by improved grey wolf optimizer based on elite mutation and quasi-oppositional learning," *IEEE Access*, vol. 7, pp. 155513–155529, 2019, doi: 10.1109/ACCESS.2019.2949582.

XIONG CHENG was born in Yingcheng, Hubei, China, in 1984. He received the B.S. degree from the College of Hydraulic and Environmental Engineering, China Three Gorges University, Yichang, Hubei, in 2008, and the Ph.D. degree in hydraulic and hydropower engineering from the Dalian University of Technology, Dalian, Liaoning, China, in 2015.

Since 2015, he has been a Lecturer with the College of Hydraulic and Environmental Engineering, China Three Gorges University. He has developed approximately ten inventions and authored 30 articles published in peer-reviewed journals. His current research interests include optimal reservoir operation, electricity markets, and decision support system development.

ZHONG-KAI FENG was born in Heze, Shandong, China, in 1988. He received the B.S. and Ph.D. degrees in hydraulic and hydropower engineering from the Dalian University of Technology, Dalian, Liaoning, China, in 2011 and 2016, respectively.

Since 2019, he has been an Associate Professor with the School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan, Hubei, China. He is the author of about 30 inventions and 50 articles in peer-reviewed journals. His current research interests include hydropower and renewable energy operation optimization, machine learning, and decision support system development.

WEN-JING NIU was born in Chifeng, Inner Mongolia, China, in 1989. She received the B.S. and Ph.D. degrees in hydraulic and hydropower engineering from the Dalian University of Technology, Dalian, Liaoning, China, in 2012 and 2017, respectively.

Since 2017, she has been an Engineer with the Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan, China. She is the author of about 40 inventions and 30 articles published in peer-reviewed journals. Her current research interests include runoff forecasting, watershed management, and cascade reservoirs operation optimization.