IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 21, 2020, accepted September 30, 2020, date of publication October 6, 2020, date of current version October 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3028959

A Sequential Structure for Water Inflow Forecasting in Coal Mines Integrating Feature Selection and Multi-Objective Optimization

SHI CHEN[1,2,3] AND SHUNING DONG[4]

[1]College of Water Resources and Architectural Engineering, Northwest A&F University, Xianyang 712100, China
[2]Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas, Ministry of Education, Northwest A&F University, Xianyang 712100, China
[3]State Key Laboratory of Water Resource Protection and Utilization in Coal Mining, Beijing 100011, China
[4]Xi'an Research Institute of China Coal Technology and Engineering Group Corporation, Xi'an 710054, China

Corresponding authors: Shi Chen (chensun2006@nwafu.edu.cn) and Shuning Dong (dongshuning@cctegxian.com)

This work was supported in part by the Open Fund of State Key Laboratory of Water Resource Protection and Utilization in Coal Mining under Grant GJNY-18-73.15, in part by the National Natural Science Foundation of China (NSFC) under Grant 41602254 and Grant 41807221, in part by the Science and Technology Innovation Project of Northwest A&F University, double-class discipline group dry area hydrology and water resources regulation research funding project, under Grant Z102022011, and in part by the Special Funding Project for Basic Scientific Research Business Fees of Central Universities under Grant 2452016179.

ABSTRACT To construct an accurate and stable approach for water inflow forecasting, a series of advanced and effective techniques, such as variational mode decomposition (VMD), outlier robust extreme learning machine (ORELM) and multi-objective grey wolf optimizer (MOGWO), are appropriately integrated into this study. Considering that the influence of the mode number on the VMD decomposition effectiveness, such an argument is determined by observing the converged centre frequency distribution among the components. Then the characteristic items of water inflow series are extracted by VMD, thus obtaining a series of sub-components. Afterwards, ORELM is applied to predict each component, where the parameters of ORELM are optimized by MOGWO with multi-objective functions including forecasting accuracy and stability. Correspondingly, the aggregation of all components' prediction values is considered as the final results. The experimental results obtained by performing eight various models on real-time data demonstrate that the supplementary modules achieve positive effects on the improvement of prediction accuracy, where the proposed model implements an average performance promotion of 48.43% compared with the contrastive models.

INDEX TERMS Water inflow prediction, variational mode decomposition, centre frequency distribution, outlier robust extreme learning machine, multi-objective grey wolf optimizer.

I. INTRODUCTION

The precise water inflow prediction for deep mine has received widespread focus in the past few decades, which plays a vital role in designing sound mine drainage system and practical arrangements for water prevention [1].

A brief classification for the existing water inflow prediction approaches can be claimed as analytical methods, numerical simulations and data-based uncertainty models. A series of empirical formulas based on Dupuit formula and Theis formula have been developed in past decades, which describe the relationship between the water inflow and the different factors, such as groundwater influence radius, surrounding groundwater level and permeability coefficient of surrounding rock [2]. However, the anisotropy of rock mass and changes in the seepage field are generally neglected among the above process, since the hydrogeological conditions are idealized. For the numerical simulations, which focus on estimation modelling by finite element methods applying the permeability coefficient, recharge intensity and mechanical parameters, have been widely investigated among the past decades. Oda [3] proposed an equivalent model by coupling the seepage field, stress field of the rock mass and combining the stress tensor method with the permeability tensor.

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

Besides, Molinero *et al.* [4] realized the numerical solution of the nonlinear water flow equation with step iterative process. It can be confident that the numerical simulations consider the complex conditions of studying hell comprehensively, such as the coexistence of laminar turbulence and so on. However, the above methods require high computational resources, due to the redundant parameters, thus resulting in time-consuming. Different from the above two types of models, the data-based uncertainty models applying machine learning, deep learning networks and the relevant stochastic process methods can achieve water inflow prediction without comprehensive hydrogeological conditions of target areas. Considering such a prediction task as the data-based time series forecasting, the implementation of the data-based uncertainty models do not impose too strict requirements on the quantity and quality of data, which can be directly performed on the water inflow series without structural characteristics of aquifer media and aquifer flow in the target area. For instance, Wang *et al.* [5] developed a GM(1, 1) metabolic model for the estimation of the spring discharge, where the results indicate that the predicted spring flow values can well-converged to the actual values. Kong-A-Siou *et al.* [6] applied a neural network (NN) with a recurrent multilayer perceptron to predict the water table levels, where the assessments of the proposed model demonstrate that the accurate estimations for maximal drawdown can be effectively generated by the neural model. Deng *et al.* [7] developed a feed forward back propagation neural network (FBNN) model to predict the final velocity of single bubbles in still water, where the coefficient of determination value of the final velocity prediction model was 0.83. Additionally, Li *et al.* [8] and Goyal *et al.* [9] developed the gaussian process regression model (GPR) and support vector regression (SVR) for water inflow prediction, respectively. Chen *et al.* [10] applied the GPR model to handle the mapping relationship between outflow of the upstream reservoir and inflow of the downstream reservoir, where the the simulation results illustrate that the smaller mean absolute deviation can be obtained by the developed GPR model. Liu *et al.* [11] proposed three novel hybrid wind speed forecasting models based on multi-decomposing strategy and extreme learning machine (ELM), where ELM is performed as the predictor for subseries prediction. The above simulation results illustrate that the nonlinear estimation methods, such as NN, GPR, SVR, etc., can achieve satisfied estimation results for the nonlinear inflow series. However, the forecasting performance of NN and SVR are restricted by the inherent parameters and the dataset scale significantly [12]. For this purpose, extreme learning machine (ELM) is developed based on the regular single-hidden layer feed-forward network (SLFN) [13], which takes into account both training speed and generalization performance. Considering that the forecasting stability of normal ELM is affected by the outliers existed in datasets, Zhang and Luo [14] introduced the l_1-norm into ELM for enhancing the prediction robustness, which is defined as outlier-robust extreme learning machine (ORELM). To our knowledge, such an advanced estimation model has not been investigated in the field of water inflow forecasting. Hence, the efficiency and effectiveness of ORELM applied for water inflow forecasting will be testified in this study.

To obtain better performance of the aforementioned nonlinear methods, scholars pay attention to the field of parameter optimization, which contributes to enhancing the forecasting robustness and accuracy. Among the optimization strategies, intelligent optimization algorithms including genetic algorithm (GA), particle swarm optimization (PSO), grey wolf optimizer (GWO) and so on, have been widely investigated in past decades [15]–[17]. However, the objective function used in the previous studies is the forecasting error indicator, guaranteeing the forecasting accuracy virtually, while the prediction stability is neglected. Therefore, the multi-objective optimization methods developed based on intelligent optimization algorithms are employed to balance the forecasting accuracy and stability simultaneously [18]–[20]. Given this, multi-objective grey wolf optimizer (MOGWO) equipped with unique leadership mechanism is selected to achieve parameter optimization for ORELM in this study, which selects the indicators in the aspects of accuracy and stability, namely root-mean-square error (RMSE) and standard deviation, as the objective functions in the optimization process.

It is worth noting that in the field of time series forecasting, the data preprocessing strategies play a vital role in further improving the forecasting performance. Among the strategies, feature selection based on the filter method has received widespread focus due to the characteristic of easy realization. The central idea of above feature selection is to decompose the one-dimensional time series into multiple components, thus extracting the tendency item with better smoothness. Adarsh and Janga Reddy [21] adopted the multivariate empirical mode decomposition (MEMD) to perform the multiscale characterization of hydroclimatic time series, where the forecasting accuracy is significantly improved. Nazir *et al.* [22] compared the improvements achieved by the combined models adopting empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD) and variational mode decomposition (VMD), where the evaluations for the results illustrate that the VMD-based model possesses higher prediction previous. For the well-investigated decomposition approaches including EMD, EEMD and VMD, the modal-aliasing that is existed in EMD and is modified by EEMD to some extent may pose a challenge of accurate prediction for the decomposed components. By contrast, VMD implements the separation of components effectively by determining the correlation frequency band with the variational model, overcoming the modal-aliasing problem. Nevertheless, the mode number of VMD need to be predetermined based on the frequency scales of signals. In contrast, such frequency conditions of the actual water inflow data are typically unknown. Hence, by the observation of the centre frequency distribution

with different mode numbers, the optimal mode number of VMD can be obtained [23]. For this purpose, VMD coupling the determination approach for the above mode number is employed to obtain the characteristic components of the water inflow series in this study.

In summary, a novel data-based hybrid model for water inflow prediction is developed by integrating VMD, ORELM and MOGWO-based multi-objective parameter optimization, where the main contributions and innovations of our study are summarized as below:

1) The mode number of VMD is successfully determined for the water inflow series by observing the converged centre frequency distributions with mode numbers in certain boundaries. Subsequently, VMD with the appropriate argument is performed on the inflow series, extracting the characteristic components for decreasing the prediction difficulty.
2) Four benchmark functions for evaluating multi-objective optimization algorithms are employed to compare the optimization capabilities of multi-objective ant lion optimizer (MOALO), multi-objective multi-universe optimization (MOMVO) and MOGWO. The visualizations of the optimized Pareto optimal solutions and actual ones are depicted for the intuitive observation.
3) To balance the accuracy and stability of the combined water inflow forecasting model simultaneously, the parameters of ORELM for each decomposed component are optimized by MOGWO with two objective functions, i.e., RMSE and standard deviation of prediction errors.
4) Comprehensive evaluations for all the experimental models in the aspects of forecasting accuracy, correlation to the actual values and differences among the experimental models are carried out for evaluating the forecasting performance of the proposed model reasonably and scientifically.

Furthermore, the theoretical foundation of VMD, ORELM and MOGWO are introduced in Section II. In Section III, the support strategies employed in the proposed model are explained in sequence, where the evaluation for MOGWO is performed. Section IV describes the study dataset, experimental design and contrastive assessments, respectively. The conclusions are drawn out in Section V.

II. MATERIALS AND METHODS

A. VARIATIONAL MODE DECOMPOSITION

Considering that variational mode decomposition (VMD) is a self-adaptive, completely non-recursive modal variation and signal processing technique, it possesses solid mathematical theory and superior decomposition performance compared with empirical mode decomposition (EMD) [24]. Each decomposed band-limited mode is compressed around the centre pulsation determined by the decomposition. For the given signal f, the primary procedures of VMD can be summarized as follows:

1) CONSTRUCTING VARIATIONAL PROBLEM

The unilateral frequency spectrum corresponding to mode u_k, can be obtained by applying the Hilbert transform to deduce a correlation analysis signal [25]. Then the obtained spectrum is transferred to the baseband by regulating the mixed exponents to the estimated central frequencies, after which the bandwidth can be estimated based on the H^1 Gauss smoothness of the demodulated signal [26]. Correspondingly, the constrained variational problem can be defined as below:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}$$
$$s.t. \sum_{k=1}^{K} u_k = f \tag{1}$$

where $\delta(t)$ denotes the Dirac distribution, u_k and ω_k indicate the sets of modes and centre pulsations, respectively [24].

2) SOLVING THE VARIATIONAL PROBLEM

Then the quadratic penalty term and Lagrange multiplier are applied to convert the above constrained problem into a dual unconstrained problem [24], where the redefined unconstrained problem is expressed as below:

$$\begin{aligned} &L\left(\{u_k\},\{\omega_k\},\lambda\right) \\ &= \alpha \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ &\quad + \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle \end{aligned} \tag{2}$$

Subsequently, alternate direction method of multipliers (ADMM) is employed to solve the above unconstrained problem. By seeking out the saddle points of the Lagrangian expression based on the iteration of u_k^{n+1}, ω_k^{n+1} and λ^{n+1}. Hence, the solutions of u_k and ω_k can be obtained as follows:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \alpha)} \tag{3}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k(\omega) \right|^2 d\omega} \tag{4}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right) \tag{5}$$

where $\hat{u}_k^{n+1}$, $\hat{u}_i(\omega)$, $\hat{f}(\omega)$ and $\hat{\lambda}(\omega)$ are the Fourier transform results of u_k^{n+1}, $u_i(t)$, $f(t)$ and $\lambda(t)$ one by one.

B. OUTLIER-ROBUST EXTREME LEARNING MACHINE

On the basis of single-hidden-layer feedforward network (SLFN), a unique network structure applying random inherent network parameters is proposed by Huang *et al.* [13], which is defined as extreme learning machine (ELM). To further enhance the generalization performance and robustness

to outlier points a novel version of ELM, namely outlier robust ELM (ORELM), is proposed by Zhang and Luo [14]. To obtain a convex relaxation objective function, l_1 and l_2-norm are employed for the training error and output weight, respectively, which is expressed as below:

$$\min_{\beta} \|\mathbf{e}\|_1 + \frac{1}{C}\|\beta\|_2^2 \ \ s.t. \mathbf{y} - \mathbf{H}\beta = \mathbf{e} \tag{6}$$

where C is the regularization parameter and $\mathbf{H}$ denotes the hidden layer output matrix. For the constrained convex optimization problem constructed in (6), the augmented Lagrangian function can be deduced to obtain the solutions, which is defined as follows:

$$L_\mu(\mathbf{e}, \beta, \lambda) = \|\mathbf{e}\|_1 + \frac{1}{C}\|\beta\|_2^2 + \lambda^T(\mathbf{y} - \mathbf{H}\beta - \mathbf{e}) + \frac{\mu}{2}\|\mathbf{y} - \mathbf{H}\beta - \mathbf{e}\|_2^2 \tag{7}$$

where μ represents a penalty parameter defined as $2N/\|y\|_1$, λ indicates the Lagrange multiplier vector. Then the optimal solutions can be deduced by minimizing those above augmented Lagrangian function, where the specific process can be found in [27], [28].

C. MULTI-OBJECTIVE GREY WOLF OPTIMIZER

Grey wolf optimizer (GWO) proposed by Mirjalili *et al.* [29] is a novel population-based meta-heuristics, which mimics the behaviours of grey wolves in terms of leadership and hunting mechanisms. Among the algorithm, four groups of grey wolves, namely alpha, beta, delta, and omega wolves, are separated according to the pack acknowledge of various wolves. To achieve multi-objective optimization based on GWO, the archive component and a leader selection strategy are introduced into the normal GWO, thus performing multi-objective grey wolf optimizer (MOGWO) effectively. Therefore, the principal mechanisms of MOGWO can be summarized as follows [20]:

1) ENCIRCLING BEHAVIOR

Besides the social leadership mentioned above, the encircling behaviour among the hunting process is simulated with the formulations blow:

$$\vec{D} = \left|\vec{C} \cdot \overrightarrow{X_p}(t) - \vec{X}\right| \tag{8}$$

$$\vec{X}(t+1) = \overrightarrow{X_p}(t) - \vec{A} \cdot \vec{D} \tag{9}$$

$$\vec{C} = 2 \cdot \vec{r}_1$$

$$\vec{A} = 2\vec{a} \cdot \vec{r}_2 - \vec{a}$$

where $\vec{C}$ and $\vec{A}$ are coefficient vectors that generate random values in the intervals of $[0, 2]$ and $[-1, 1]$, a is the linear decreasing factor in the scoped of $[0, 2]$, r_1 and r_2 are random values in $[0, 1]$, $\vec{X}_p$ and $\vec{X}$ denote the position vectors of the prey and searching wolf, respectively.

2) HUNTING BEHAVIOR

For the obtained three best solutions, i.e., α, β and δ wolves, the positions of such wolves are updated with the formulas to simulate the hunting behaviour.

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1\left(\left|\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}\right|\right) \tag{10}$$

$$\vec{X}_2 = \vec{X}_\beta - \vec{A}_2\left(\left|\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}\right|\right) \tag{11}$$

$$\vec{X}_3 = \vec{X}_\delta - \vec{A}_3\left(\left|\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}\right|\right) \tag{12}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{13}$$

3) EXPLOITATION AND EXPLORATION

To achieve a better balance between the searching and convergence phases, an adaptive strategy controlling the transition from exploitation to exploration in the course of the optimization process is introduced in MOGWO, which is guaranteed by the status of $\vec{A}$. Specifically, the search agents will tend against the grey for better simulation of the natural behaviours of wolves. In contrast, the search agents will converge toward the grey. It is worth noting that $\vec{C}$ is generated randomly among the phases between the beginning and end of the iteration to obtain a better global searching capability [29].

4) ARCHIVE

The archive is introduced to the normal GWO frame for the storage and retrieval of the non-dominated Pareto optimal solutions. Considering the Dominant relationship between the new member and the archive residences, the status of the archive can be updated. Additionally, the grid mechanism is performed to re-arrange the segmentation of objective space, thus enhancing the variety of the approximate Pareto optimal frontier [20].

5) LEADER SELECTION MECHANISM

Based on a roulette-wheel approach, the leader selection mechanism is introduced to handle the difficulty of comparing the solutions obtained so far. Besides, the probability for each hypercube defined in the roulette-wheel approach is expressed as follows [20]:

$$P_i = \frac{c}{N_i} \tag{14}$$

III. STRUCTURE OF THE PROPOSED MODEL

A. SELECT APPROPRIATE MODE NUMBER FOR VMD

Similar to the relevant decomposition approaches, such as empirical mode decomposition (EMD) and complete ensemble EMD with adaptive noise (CEEMDAN), the unidimensional time series can be decomposed into a set of components. However, the number of decomposed components are adaptively controlled in EMD and CEEMDAN, while the mode number should be pre-given for VMD [24]. For the time series including wind speed, water inflow and so on, whose frequency scales are difficult to determine in advance due to the characteristics of randomness, intermittence and volatility, the mode number of VMD will be difficult to determine. Following the previous literature [23], [30]–[32], the mode number of VMD is predetermined

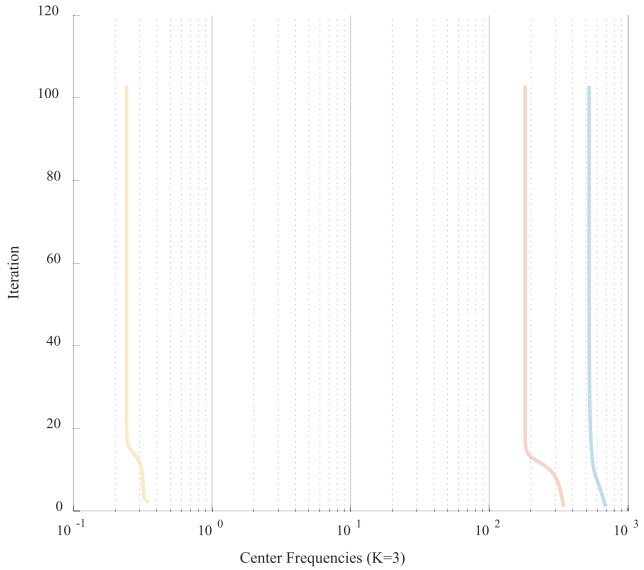

FIGURE 1. Center frequency distribution with $K = 3$.

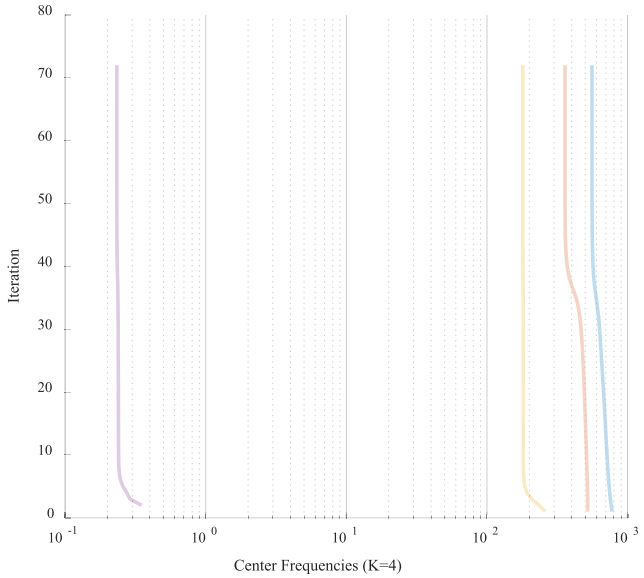

FIGURE 2. Center frequency distribution with $K = 4$.

by the subjective experience of scholars and optimized by swarm intelligence algorithms, which are not adaptive to different data as well as being time-consuming when adopting the iterative based algorithm. By observing the centre frequency distribution of each component with various mode number K, it can be seen that the centre frequencies of adjacent components will be convergent as the value of K increases, thus resulting in mode mixing problem [23]. Hence, to select an appropriate mode number of VMD, the visualizations of the centre frequency distribution for each decomposed component with $K = 3$, $K = 4$ and $K = 5$ are expressed in Fig. 1-Fig. 3. It can be seen from Fig. 2 that the iterative curves of the components start to converge on each other when $K \geqslant 4$, which illustrates that excessive

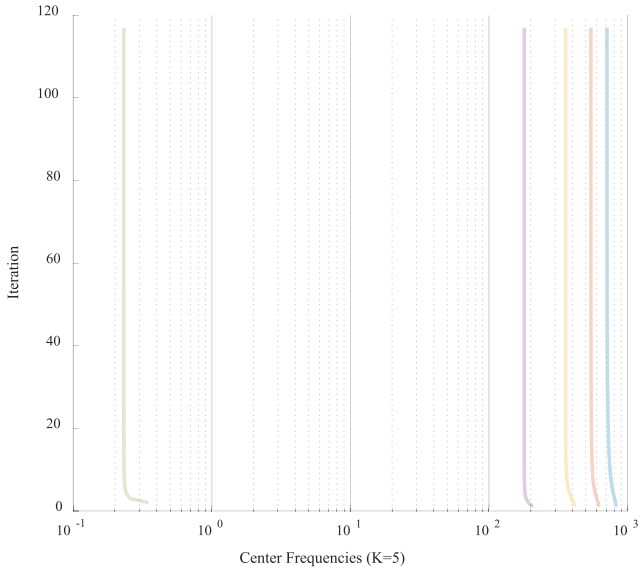

FIGURE 3. Center frequency distribution with $K = 5$.

decomposition has occurred. Therefore, the mode number K is set to 3 for the collected water inflow series in this study.

B. PARAMETER OPTIMIZATION FOR ORELM BASED ON MOGWO

1) EVALUATION FOR MOGWO

In this section, four typical benchmark functions illustrated in Table 1 are employed to visually testify the superiority of MOGWO [33] . Meanwhile, two newly developed multi-objective optimizers including multi-objective ant lion optimizer (MOALO) [34] and multi-objective multi-universe optimization (MOMVO) [35], are carried out for contrastive experiments. Additionally, the maximum iteration number, the number of searching agents and the archive capacity of above three algorithms are set to 100, 100 and 100 in sequence. The Pareto optimal solutions of the functions mentioned above deduced by all the optimizers are demonstrated in Fig. 4-Fig. 7. It can be claimed that the obtained Pareto optimal fronts of all the algorithms are distributed to the true Pareto optimal front. At the same time, there exist more Pareto optimal solutions obtained by MOGWO. Hence, the stability and superiority of MOGWO performing on these problems (ZDT1 with linear PF, ZDT1, ZDT 2 and ZDT3) can be verified from the quantitative results.

2) PARAMETER OPTIMIZATION

As mentioned in reference [36], [37], the forecasting performance of the newly constructed ORELM model is affected by the inherent parameters, such as the regularization coefficient and the number of hidden. To implement stable prediction results, parameter optimization for predictors has been received widespread attention, where various swarm intelligence optimization algorithms are employed to optimize the parameters with single-objective function [38], [39].

TABLE 1. Benchmark functions.

ZDT1 with linear PF	ZDT1
$Minimize: f_1(x) = x_1$ $Minimize: f_2(x) = g(x) \times h(f_1(x), g(x))$ $Where: G(x) = 1 + \frac{9}{N-1}\sum_{i=2}^{N} x_i$ $h(f_1(x), g(x)) = 1 - \frac{f_1(x)}{g(x)}$ $0 \leqslant x_i \leqslant 1, 1 \leqslant i \leqslant 30$	$Minimize: f_1(x) = x_1$ $Minimize: f_2(x) = g(x) \times h(f_1(x), g(x))$ $Where: G(x) = 1 + \frac{9}{N-1}\sum_{i=2}^{N} x_i$ $h(f_1(x), g(x)) = 1 - \sqrt{\frac{f_1(x)}{g(x)}}$ $0 \leqslant x_i \leqslant 1, 1 \leqslant i \leqslant 30$
ZDT2	**ZDT3**
$Minimize: f_1(x) = x_1$ $Minimize: f_2(x) = g(x) \times h(f_1(x), g(x))$ $Where: G(x) = 1 + \frac{9}{N-1}\sum_{i=2}^{N} x_i$ $h(f_1(x), g(x)) = 1 - \left(\frac{f_1(x)}{g(x)}\right)^2$ $0 \leqslant x_i \leqslant 1, 1 \leqslant i \leqslant 30$	$Minimize: f_1(x) = x_1$ $Minimize: f_2(x) = g(x) \times h(f_1(x), g(x))$ $Where: G(x) = 1 + \frac{9}{29}\sum_{i=2}^{N} x_i$ $h(f_1(x), g(x)) = 1 - \sqrt{\frac{f_1(x)}{g(x)}}$ $- \left(\frac{f_1(x)}{g(x)}\right) \sin(10\pi f_1(x))$ $0 \leqslant x_i \leqslant 1, 1 \leqslant i \leqslant 30$

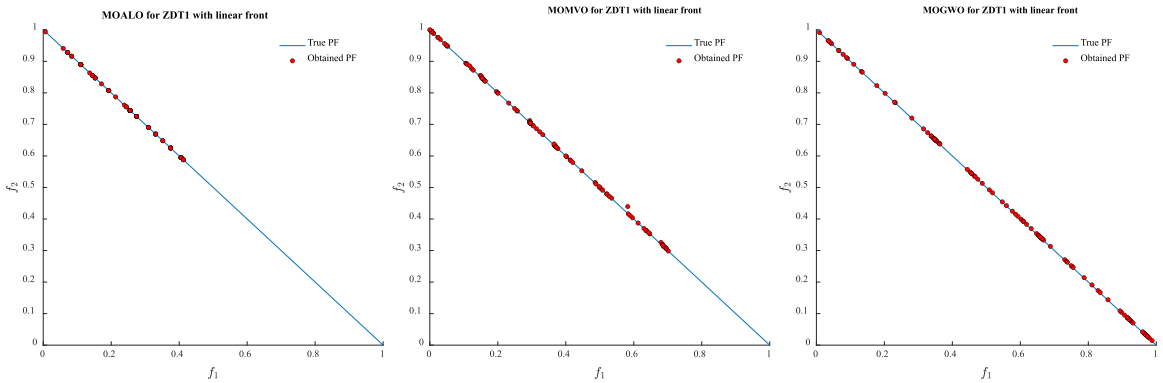

FIGURE 4. Pareto optimal solutions for ZDT1 with linear front.

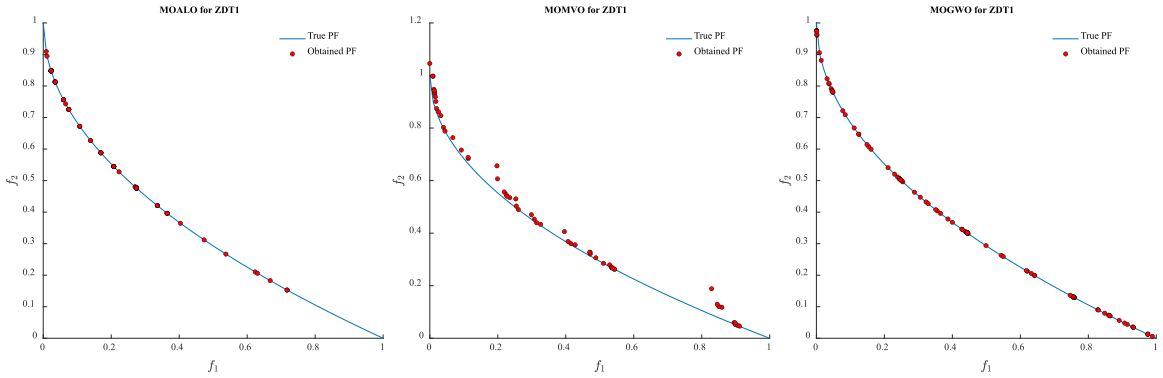

FIGURE 5. Pareto optimal solutions for ZDT1.

However, the forecasting stability of the approaches is neglected in the above optimization process, which is comparably essential for the construction of the forecasting model. For this purpose, the newly developed MOGWO is employed to achieve multi-objective optimization for ORELM in terms of forecasting accuracy and stability, where the indicators, namely root-mean-square error (RMSE) and standard deviation (STD) of the forecasting error, are employed as the objective functions. The detailed definitions of such two indexes are illustrated in formulas (15) and (16), respectively.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2} \tag{15}$$

$$STD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(e_i - \bar{e}_i)^2}$$
$$e_i = y_i - \hat{y}_i \tag{16}$$

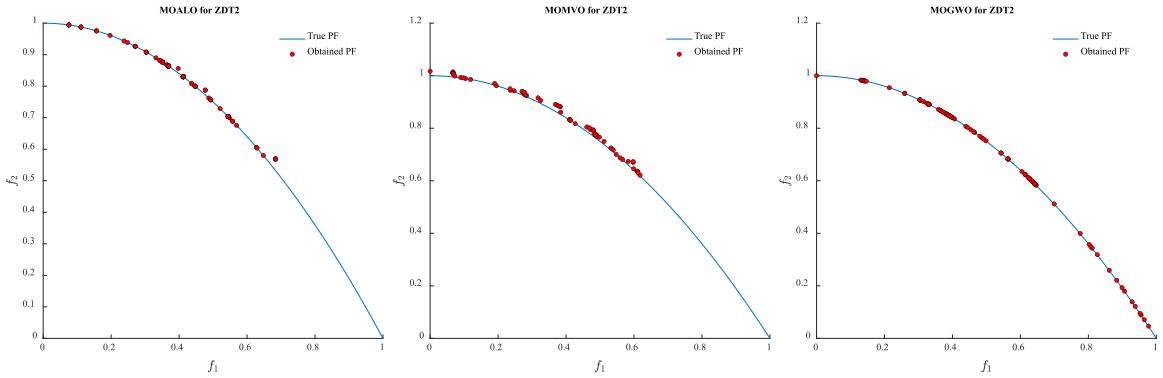

FIGURE 6. Pareto optimal solutions for ZDT2.

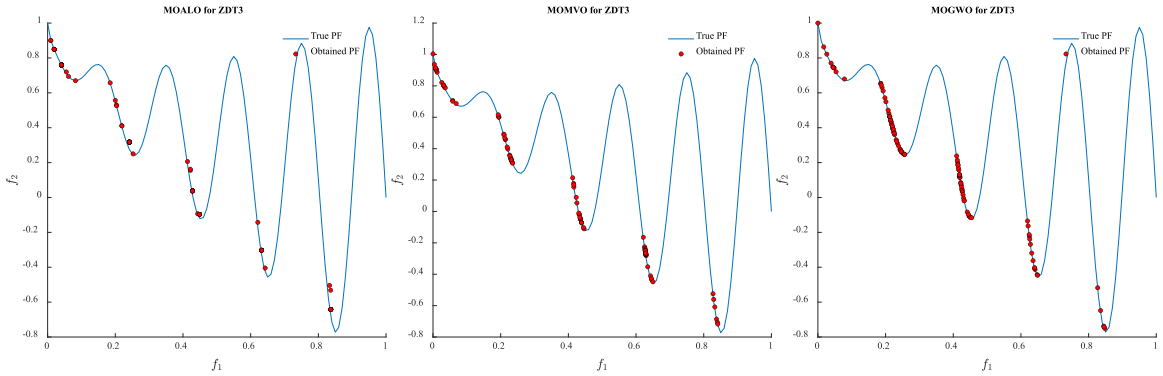

FIGURE 7. Pareto optimal solutions for ZDT3.

where y_i and $\hat{y}_i$ denote the actual value and the predicted value, $\bar{e}_i$ is the mean value of e_i, respectively. Specifically, the pseudocode of aforementioned objective optimization for ORELM is depicted in Algorithm 1.

C. STUDY AREA AND DATA

The second coal mine of Yangquan Coal Mine is a large modern coal mine located in Yangquan, Shanxi, northern China. Its reconnaissance and exploration began in the 1950s and officially went into production on May 1, 1951. The main minable coal seams include No. 3 coal seam in the Shanxi group and No. 8, No. 12 and No. 15 coal seam in the Taiyuan group of the Carboniferous. The water source mainly comes from the confined aquifer of the fractured sandstone between coal seams, where the collected water inflow data is exhibited in Fig. 9. Besides, the water inflow sequence data is divided into three parts with the splitting ratio of 3:1:1, namely the training, validation and testing sets, where the training and validation sets are employed for the parameter optimization, playing the roles of model construction and model selection, respectively.

D. SPECIFIC FRAMEWORK SEQUENCES OF THE PROPOSED APPROACH

In summary, the integrated framework of the proposed forecasting approach is depicted in Fig. 10, while the specific procedures are summarized below:

Step 1: Split the collected water inflow series into the training, validation and testing parts.
Step 2: Select an optimal mode number of VMD based on the observation of converged centre frequency distributions.
Step 3: Obtain the optimal parameters of ORELM adopting MOGWO-based multi-objective optimization on the training and validation sets.
Step 4: Integrate the predicted components into the final prediction series of the actual testing sets.

IV. EXPERIMENTAL DESIGN

A. EXPERIMENTAL DESCRIPTION

1) CONTRASTIVE MODELS

To achieve sufficient verification for the proposed hybrid forecasting method, several benchmark prediction models and the decomposition-based combined models are developed for contrastive assessments. Support vector regression (SVR), back propagation neural network (BPNN), long short term memory (LSTM) network and ORELM implementing water inflow prediction with the original data are adopted as benchmark models, which can further reveal the improvements obtained by the proposed model. Besides, two adaptive decomposition techniques, namely EMD and CEEMDAN, are combined with ORELM verify the superiority of VMD, where the parameters of ORELM are determined by

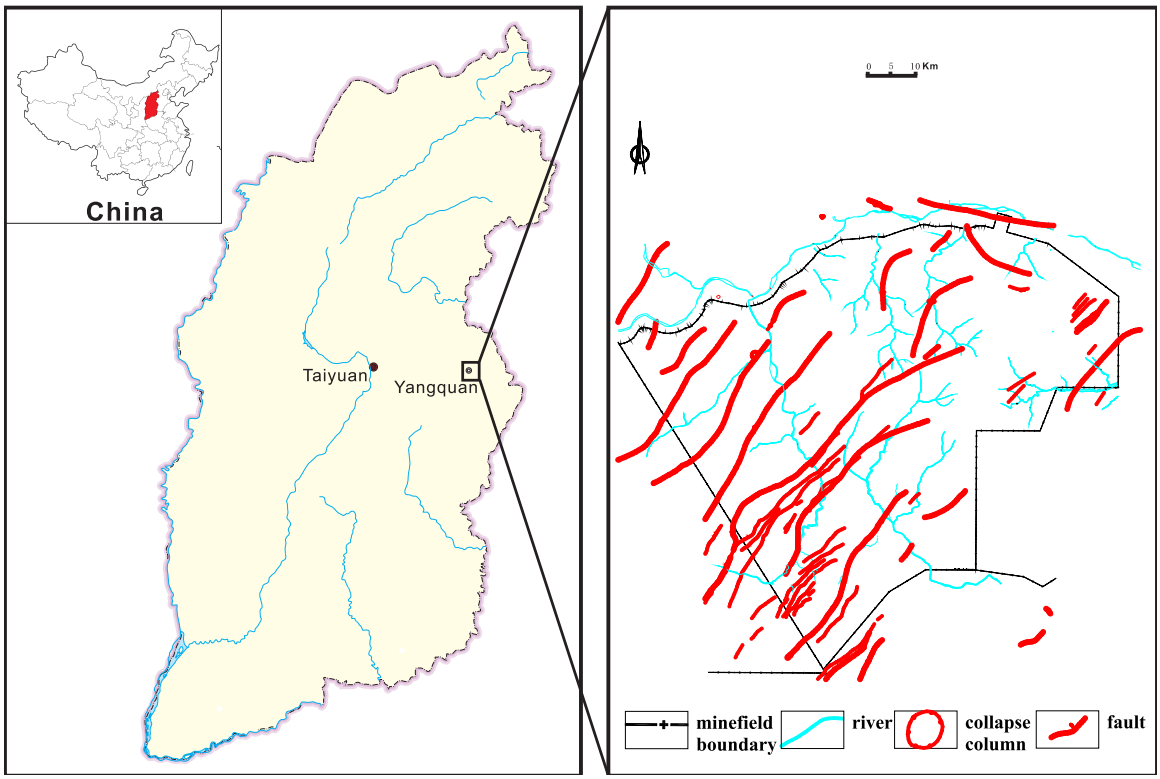

FIGURE 8. Location of the study area.

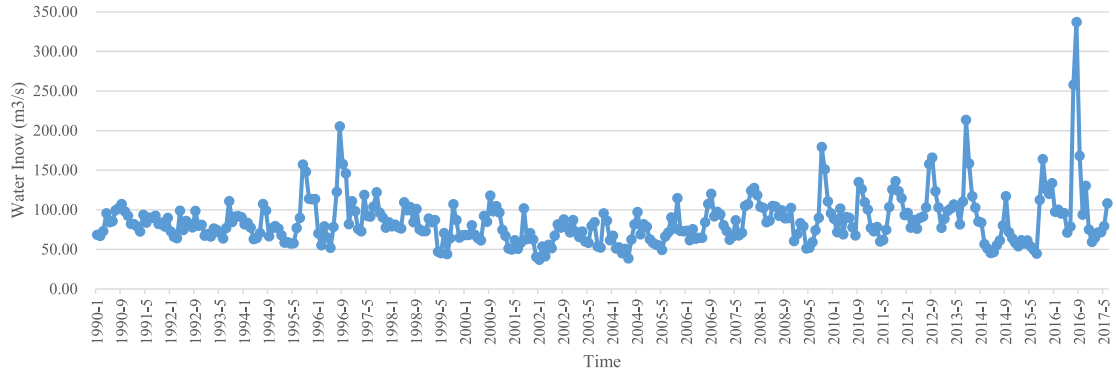

FIGURE 9. Raw water inflow data.

grid searching. Likewise, VMD is composited with the grid searching-based ORELM, thus evaluating the effectiveness of VMD and the MOGWO-based parameter optimization.

Additionally, the parameter setting for the above contrast models and the proposed one will be recommended in this part. Among the contrastive models, the hyper-parameters including the regularization coefficient C, kernel parameter δ and the number of hidden layer neurons within SVR and ORELM are determined by grid searching in the scope of $[2^{-10}, 2^{10}]$, $[2^{-8}, 2^{8}]$ and $[10, 100]$. For BPNN and LSTM network, the number of hidden layer for them is set as 1, while the number of the hidden layer neurons are determined to 32 according to the trial-and-error procedure [40], [41]. Moreover, the key parameters of CEEMDAN including the maximum number of sifting iterations, standard deviation of the added white noise, and the number of realizations are preset as 5000, 500 and 0.2 following the previous investigations [42], [43]. For the proposed model, the arguments of MOGWO, such as searching agents, maximum iteration and archive size are set as 100, 50 and 100, while the boundaries of the parameters within ORELM are $[0.01, 1000]$ and $[10, 100]$, respectively.

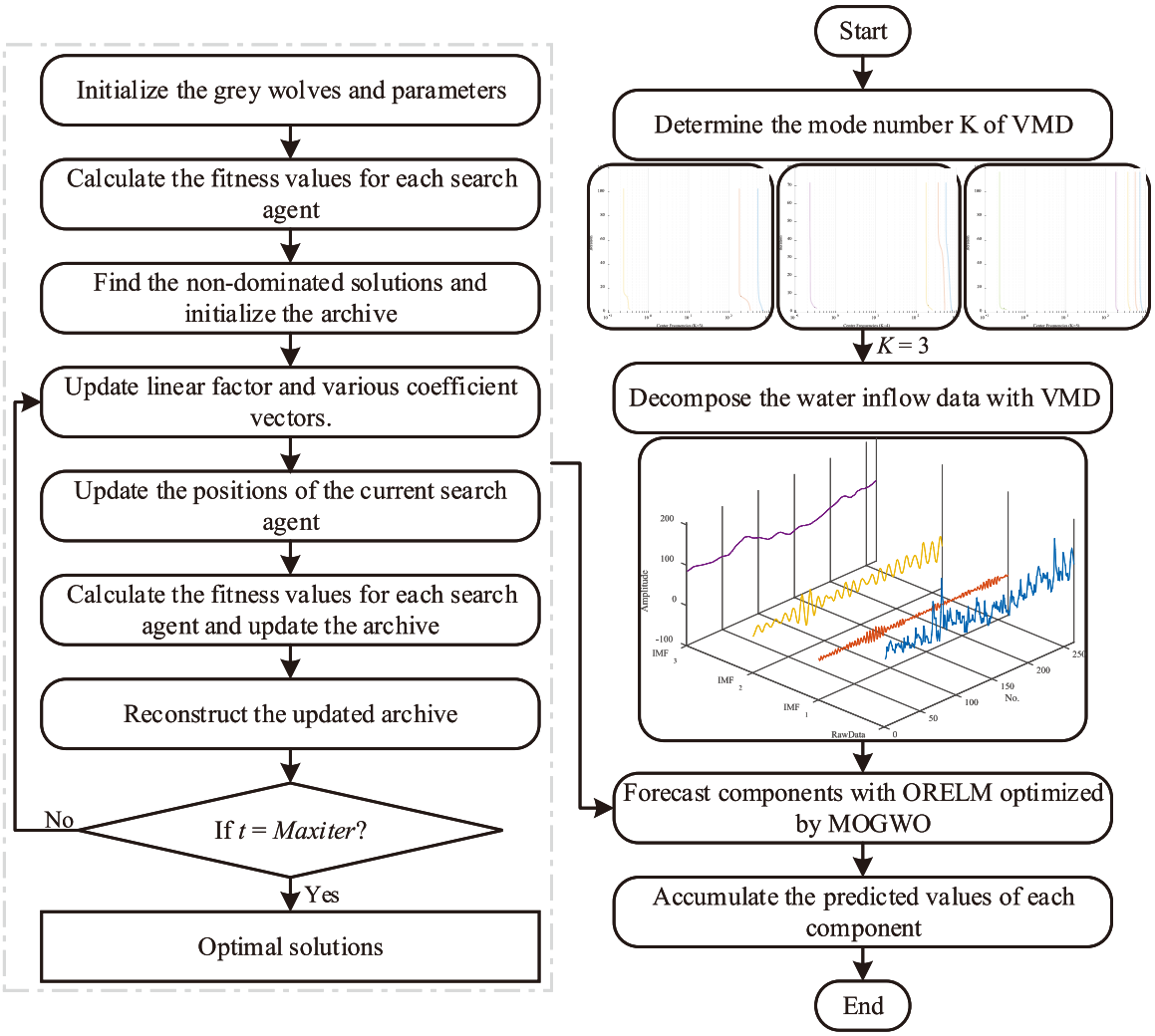

FIGURE 10. Flow chart of the proposed model.

Specifically, the decomposed components obtained by EMD, CEEMDAN and VMD are illustrated in Fig. 11-Fig. 13, respectively. It can be observed from Fig. 11 that there exist modal-aliasing problem among components, which has been weakened in the diagrams of CEEMDAN. Furthermore, as depicted in Fig. 12, VMD possesses a smaller number of components, which contributes to training the combined model with less time-consuming.

2) EVALUATION METRICS

To achieve adequate assessment for the experimental models, mean absolute error (MAE), RMSE, mean absolute percentage error (MAPE), correlation coefficient (R^2) and Nash-Sutcliffe efficiency coefficient (CE) are employed to reveal the forecasting accuracy and the differences between the proposed approach and the contrastive ones [44]. The detailed description of above five indicators are presented in Table 2, where y_i, $\hat{y}_i$ and $\bar{y}_i$ denote the actual value, the predicted value and the mean value of y_i. Furthermore, the simple decreasing

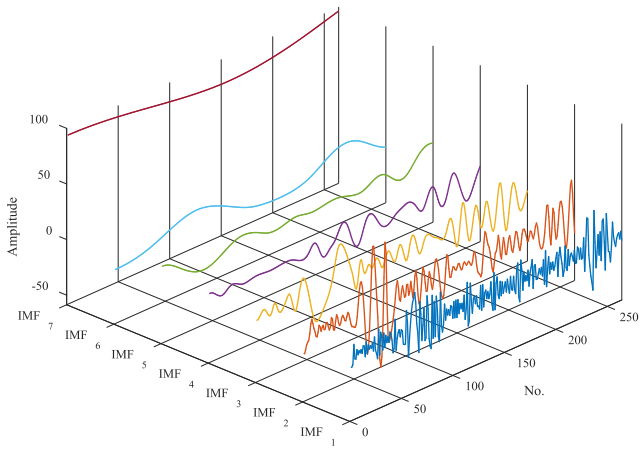

FIGURE 11. Decomposed components applying EMD.

of metrics RMSE, MAE and MAPE is adopted to observe the differences between two models, where the specific formula is expressed as (17). Additionally, Diebold-Mariano (DM)

Algorithm 1 Optimized ORELM Based on MOGWO

Fitness Function:

$$\min \begin{cases} f_1 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \\ f_2 = std\left(y_i - \hat{y}_i\right) \end{cases}$$

Parameters: *maxIter*: maximum iteration number
N: number of gray wolves
F_i: fitness value of i-th gray wolves
X_i: position of i-th gray wolves
t: current iteration
d: dimension of the problem
a: linear factor decreasing from 2 to 0 with the iterations
A and C: coefficient vectors

Initialize the vector $\{X_i | i = 1, 2, \ldots, N\}$, a, A and C.
for $i : 1 \leq i \leq N$ **do**
  Calculate the fitness values F_i for each search agent.
end for
Seek out the non-dominated solutions and initialize the archive with them
$X_\alpha = SelectLeader(archive)$
Remove the alpha wolf from the archive temporarily to avoid selecting the same leader
$X_\beta = SelectLeader(archive)$
Remove the beta wolf from the archive temporarily to avoid selecting the same leader
$X_\delta = SelectLeader(archive)$
Re-add the alpha and beta wolves to the archive
$t = 1$
while $t < maxIter$ **do**
  for $i : 1 \leq i \leq N$ **do**
    Update the position of current search agent by formulas (10), (11) and (12).
  end for
  Update the arguments of a, A, and C
  Calculate the objective values of all search agents
  Seek out the non-dominated solutions
  Update the archive based on the non-dominant solution obtained so far
end while
return Archive
Obtain the optimal parameters for ORELM.

test [45] and grey relational analysis (GRA) [46] are carried out for further assessment.

$$P_{Metric} = \frac{Metric_a - Metric_b}{Metric_a} \times 100\% \tag{17}$$

where *Metric* represents RMSE, MAE or MAPE, $Metric_a$ and $Metric_b$ denote the metric values of model a and b, respectively.

B. EXPERIMENTAL RESULTS AND ANALYSIS

1) EXPERIMENT 1

In experiment 1, SVR, BPNN, LSTM network and ORELM are performed on the collected dataset without decomposition

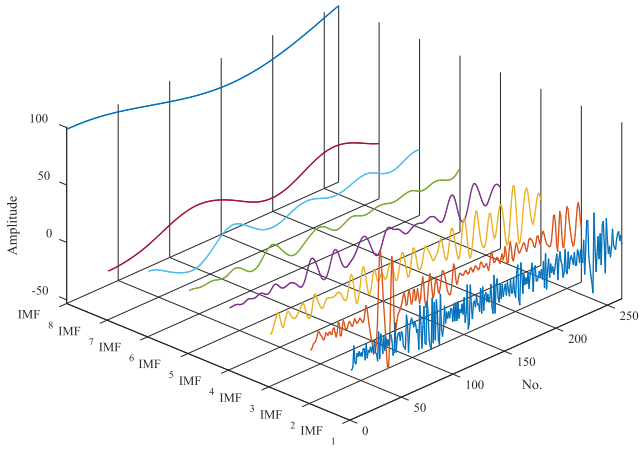

FIGURE 12. Decomposed components applying CEEMDAN.

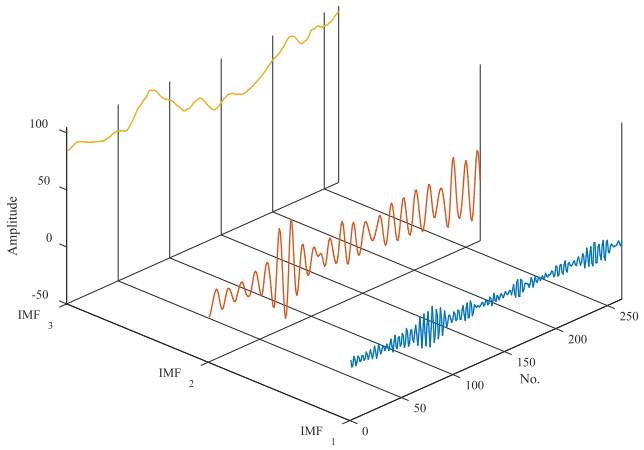

FIGURE 13. Decomposed components applying VMD.

methods, which is designed to reveal the performance gap among ORELM and the remaining approaches. The evaluation results of all the single models are presented in Table 3. It can be found that ORELM achieves minimum values of RMSE, MAE and MAPE, i.e., 36.6539 $\left(m^3/s\right)$, 21.6489 $\left(m^3/s\right)$ and 0.2294 (%). Hence, it can be certain that ORELM possesses strong generalization performance and the capability of processing outlier points. However, there exist small differences between ORELM and the remaining models, which indicates that it is difficult to achieve water inflow forecasting without data preprocessing.

2) EXPERIMENT 2

In this part, EMD, CEEMDAN and VMD are combined with ORELM to construct EMD-ORELM, CEEMDAN-ORELM and VMD-ORELM, respectively. Following the comparisons of evaluation results obtained by ORELM, EMD-ORELM, CEEMDAN-ORELM, VMD-ORELM and the proposed approach, the effectiveness and necessity of adopting decomposition techniques could be intuitively demonstrated. The specific evaluation indicators of ORELM and above combined models are presented in Table 4. It can be seen that the combined mode applying EMD, CEEMDAN and VMD

TABLE 2. Evaluation metrics.

Indicators	Definition	Equation
RMSE	Root-Mean-Square Error	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$
MAE	Mean Absolute Error Error	$MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert y_i - \hat{y}_i\rvert$
RMSE	Absolute Percentage Error	$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left\lvert\frac{100\times(y_i-\hat{y}_i)}{y_i}\right\rvert$
R^2	Correlation Coefficient	$R^2 = \frac{\left(N\sum_{i=1}^{N}\hat{y}_i y_i - \sum_{i=1}^{N}\hat{y}_i\sum_{i=1}^{N}y_i\right)^2}{\left(N\sum_{i=1}^{N}\hat{y}_i^2 - \left(\sum_{i=1}^{N}\hat{y}_i\right)^2\right)\left(N\sum_{i=1}^{N}y_i^2 - \left(\sum_{i=1}^{N}y_i\right)^2\right)}$
CE	Nash-Sutcliffe Efficiency Coefficient	$CE = 1 - \frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{N}(y_i-\bar{y}_i)^2}$

TABLE 3. Metrics RMSE, MAE and MAPE of single models.

Models	RMSE	MAE	MAPE
SVR	38.5872	22.5896	0.2444
BPNN	38.4179	22.6208	0.2436
LSTM	37.4548	22.2223	0.2356
ORELM	36.6539	21.6489	0.2294

TABLE 4. Metrics RMSE, MAE and MAPE of combined models.

Models	RMSE	MAE	MAPE
ORELM	36.6539	21.6489	0.2294
EMD-ORELM	26.5276	19.4596	0.1963
CEEMDAN-ORELM	24.5819	17.6416	0.1646
VMD-ORELM	16.843	10.9454	0.1096
The Proposed	13.4975	10.0226	0.1009

obtain the significant improvements of forecasting accuracy compared with ORELM. For instance, the metrics RMSE, MAE and MAPE of EMD-ORELM are 26.5276 (m^3/s), 19.4596 (m^3/s) and 0.1963 (%), which achieve an average metric decreasing rate of 17.39% compared with ORELM. Likewise, the metric decreasing rates of RMSE, MAE and MAPE obtained by CEEMDAN-ORELM and VMD-ORELM are 32.93%, 18.51%, 28.26% and 54.05%, 49.44%, 52.21%, respectively. It can be found that the performance promotion of VMD-based model is more significant compared with EMD-ORELM and CEEMDAN-ORELM, thus demonstrating the superiority of VMD. Further comparing the results of VMD-ORELM and the proposed model, the metric RMSE of above two models are 16.843 (m^3/s) and 13.4975 (m^3/s), respectively. Hence, the experimental results illustrate that the MOGWO-based parameter optimization can contribute to further improving the forecasting performance. In addition, the histograms of RMSE, MAE and MAPE are presented in Fig. 14 and Fig. 15, respectively. It can be seen that the proposed model obtains the minimum values among all the experimental models, thus verifying the effectiveness of VMD and MOGWO-based multi-objective parameter optimization.

3) DIFFERENCE ANALYSES

To further reveal the performance promotion achieved by the proposed model, the comprehensive difference analyses

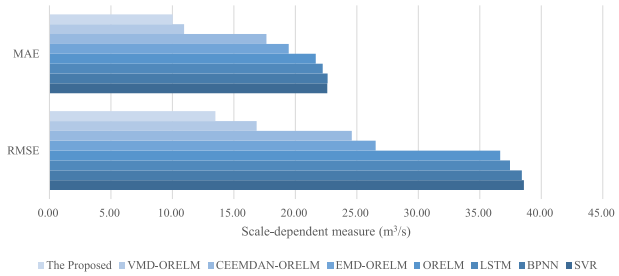

FIGURE 14. Histogram of metrics MAE and RMSE.

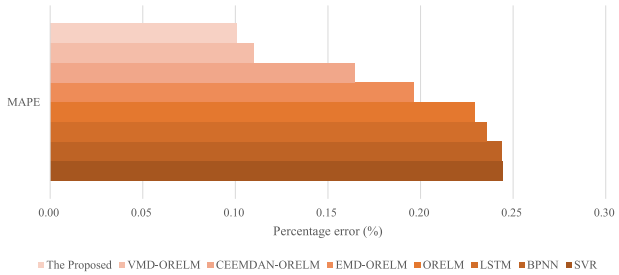

FIGURE 15. Histogram of metrics MAPE.

TABLE 5. Metrics R^2, CE and GRA of all the models.

Models	R^2	CE	GRA
SVR	0.6832	0.3885	0.7713
BPNN	0.6449	0.3939	0.8317
LSTM	0.6635	0.4239	0.8298
ORELM	0.6802	0.4482	0.8343
EMD-ORELM	0.8479	0.711	0.8047
CEEMDAN-ORELM	0.8674	0.7518	0.8038
VMD-ORELM	0.9563	0.8835	0.8356
The Proposed	0.9685	0.9252	0.8902

will be carried out in this section. To begin with, the metrics R^2, CE and GRA of all the forecasting approaches are presented in Table 5, which can intuitively account the correlation between the actual values and the prediction results of each model. As expressed in Table 5, the proposed model are maximum values among the experimental models, which indicates that the forecasting results of the proposed model can converge the actual values better. Hence, aforeinferred conclusions can be obtained here further. Furthermore, the histograms of indexes R^2 and CE are depicted

TABLE 6. Evaluation of the performance promotion for the proposed model.

Models	P_{RMSE}	P_{MAE}	P_{MAPE}	DM-Test
SVR	65.0209	55.632	58.7228	2.2466**
BPNN	64.8668	55.6931	58.584	2.3551**
LSTM	63.9634	54.8986	57.1829	2.3067**
ORELM	63.1759	53.704	56.022	2.2468**
EMD-ORELM	49.1193	48.4957	48.6086	3.3818***
CEEMDAN-ORELM	45.0919	43.1879	38.6993	3.2603***
VMD-ORELM	19.8633	8.4317	7.9691	1.2372

** and *** denote the significance levels of 5% and 1%, respectively.

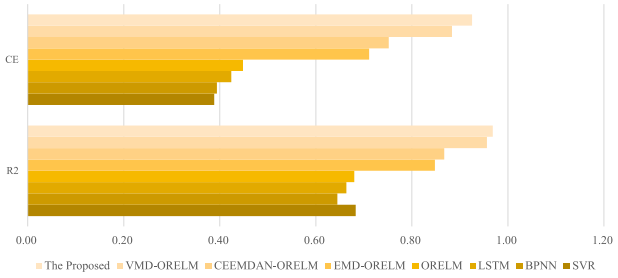

FIGURE 16. Histogram of metrics R^2 and CE.

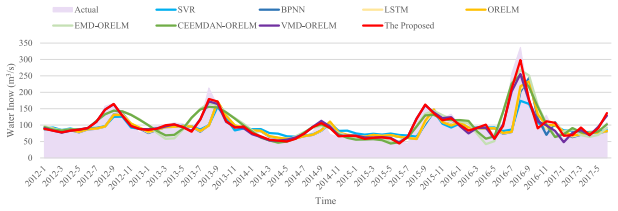

FIGURE 17. Fitting curves and the corresponding error areas diagrams of all the experimental models.

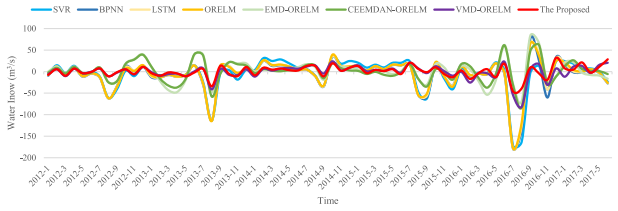

FIGURE 18. Diagrams of error curves for all the experimental models.

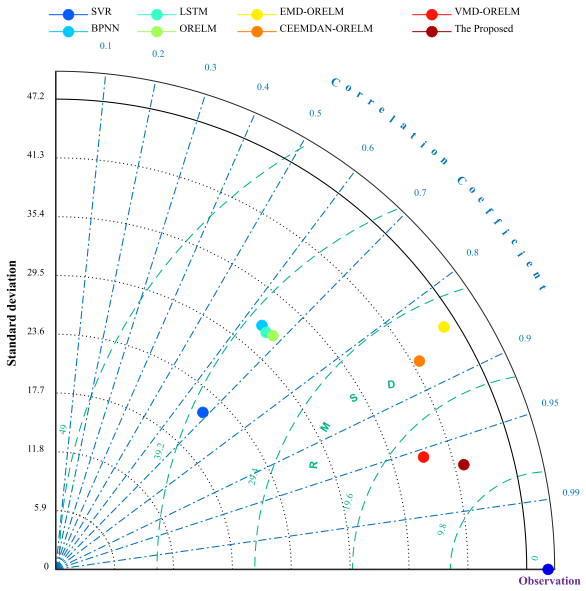

FIGURE 19. Taylor diagram for all the experimental models.

in Fig. 16, from which the conclusion obtained above can be intuitively observed here.

Additionally, the DM-test values and the ratios of performance improvement by comparing the proposed model with the contrastive ones are exhibited in Table 6. Compared with all the single models, the MAPE values of the proposed model are decreased by 58.72%, 58.58%, 57.18% and 56.02%, respectively. Besides, the averaged reduction of all the indexes achieved by the proposed model are 38.02%, 33.37% and 31.76%, respectively. In addition, it can be seen from the column illustrating the DM-test values that the null hypothesis can be accepted at the 5% significance level, indicating that the proposed model achieves a significant diversity with the significance level of 95%. Additionally, it can be observed from Fig. 17 and Fig. 18 that the VMD-based hybrid models can obtain the fitting curves that are approximate to the actual values, while the proposed model is superior to VMD-ORELM significantly. Meanwhile, the error curves of all the experimental models illustrate that the proposed approach achieves smoother fluctuations as well as distributing the forecasting error points approximately to zero points better. Hence, following the intuitive visualization results depicted in Fig. 17 and Fig. 18, it can be concluded that both forecasting accuracy and stability can be obtained by the proposed composite approach based on multi-objective parameter optimization. As demonstrated in Fig. 19, Taylor diagram [47] visualizing three statistics information simultaneously depicts that the proposed model can better approximate the observation point.

V. CONCLUSION

A novel forecasting structure applying VMD, ORELM and MOGWO-based multi-objective parameter optimization is developed in this study to achieve accurate water inflow prediction. The application of VMD contributes to improving

the forecasting performance significantly, where the mode number of VMD is determined by the observation of centre frequency distribution with different mode numbers. Besides, MOGWO is performed to achieve multiple parameter optimization for ORELM, considering the prediction accuracy and stability simultaneously. Furthermore, the following conclusions can be drawn out according to the above evaluation analyses:

1) The appropriate mode number of VMD can be effectively determined by observing the converged center frequency distribution under various mode numbers. Compared with EMD and CEEMDAN, VMD possesses better decomposition effectiveness, thus making the combined models applying VMD possess exceptional prediction results. Therefore, the application of VMD-based feature selection technique plays a vital role in significantly enhancing the forecasting accuracy.
2) The visible results of performing MOGWO on four commonly employed benchmark functions demonstrated that MOGWO introducing the unique leadership mechanism could obtain more Pareto optimal solutions, indicating that MOGWO possesses better optimization ability compared with the relevant competitors.
3) Accurate and stable forecasting results can be obtained by the proposed model based on ORELM optimized by MOGWO considering two objective functions. Comprehensive evaluation measurements for all the experimental models illustrated the superiority of the application of VMD, ORELM and multi-objective parameter optimization. Hence, the proposed hybrid model can be an effective technique for water inflow forecasting, to improve fore- casting performance with higher accuracy and stability.

REFERENCES

[1] Z. Li, T. Gao, C. Guo, and H.-A. Li, "A gated recurrent unit network model for predicting open channel flow in coal mines based on attention mechanisms," *IEEE Access*, vol. 8, pp. 119819–119828, 2020.

[2] Press, Pergamon, and Incorporated, "Recommendations for the treatment of water inflows and outflows in operated underground structures," *Tunnelling Underground Space Technol.*, vol. 4, no. 3, pp. 343–407, 1989.

[3] M. Oda, "An equivalent continuum model for coupled stress and fluid flow analysis in jointed rock masses," *Water Resour. Res.*, vol. 22, no. 13, pp. 1845–1856, Dec. 1986.

[4] J. Molinero, J. Samper, and R. Juanes, "Numerical modeling of the transient hydrogeological response produced by tunnel construction in fractured bedrocks," *Eng. Geol.*, vol. 64, no. 4, pp. 369–386, Jun. 2002.

[5] Y. Wang, Y. Hao, and X. Wang, "Simulation of karst hydrological processes using GM(1,1) metabolic model," in *Proc. IEEE Int. Conf. Grey Syst. Intell. Services (GSIS)*, Nov. 2009, pp. 588–593.

[6] L. Kong-A-Siou, A. Johannet, V. B. Estupina, and S. Pistre, "Neural networks for karst groundwater management: Case of the Lez spring (Southern France)," *Environ. Earth Sci.*, vol. 74, no. 12, pp. 7617–7632, Dec. 2015.

[7] B. Deng, R. J. Chin, Y. Tang, C. Jiang, and S. H. Lai, "New approach to predict the motion characteristics of single bubbles in still water," *Appl. Sci.*, vol. 9, no. 19, p. 3981, Sep. 2019.

[8] S.-C. Li, P. He, L.-P. Li, S.-S. Shi, Q.-Q. Zhang, J. Zhang, and J. Hu, "Gaussian process model of water inflow prediction in tunnel construction and its engineering applications," *Tunnelling Underground Space Technol.*, vol. 69, pp. 155–161, Oct. 2017.

[9] M. K. Goyal, A. Sharma, and K. L. Katsifarakis, "Prediction of flow rate of karstic springs using support vector machines," *Hydrol. Sci. J.*, vol. 62, no. 13, pp. 2175–2186, Oct. 2017.

[10] X. Chen, J. Zhou, B. Jia, X. Yang, and C. Zhou, "Characterizing the hydraulic connection of cascade reservoirs for short-term generation dispatching via Gaussian process regression," *IEEE Access*, vol. 8, pp. 145489–145502, 2020.

[11] H. Liu, X. Mi, and Y. Li, "An experimental investigation of three new hybrid wind speed forecasting models using multi-decomposing strategy and ELM algorithm," *Renew. Energy*, vol. 123, pp. 694–705, Aug. 2018.

[12] G. Huang, G.-B. Huang, S. Song, and K. You, "Trends in extreme learning machines: A review," *Neural Netw.*, vol. 61, pp. 32–48, Jan. 2015.

[13] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 2, Jul. 2004, pp. 985–990. [Online]. Available: http://ieeexplore.ieee.org/document/1380068/

[14] K. Zhang and M. Luo, "Outlier-robust extreme learning machine for regression problems," *Neurocomputing*, vol. 151, pp. 1519–1527, Mar. 2015, doi: 10.1016/j.neucom.2014.09.022.

[15] S. Wang, N. Zhang, L. Wu, and Y. Wang, "Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method," *Renew. Energy*, vol. 94, pp. 629–636, Aug. 2016.

[16] H. Duan, G. R. Lei, and K. Shao, "Forecasting crude oil consumption in China using a grey prediction model with an optimal fractional-order accumulating operator," *Complexity*, vol. 2018, pp. 1–12, Aug. 2018.

[17] M. Wang, H. Chen, H. Li, Z. Cai, X. Zhao, C. Tong, J. Li, and X. Xu, "Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction," *Eng. Appl. Artif. Intell.*, vol. 63, pp. 54–68, Aug. 2017.

[18] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[19] C. A. C. Coello, G. T. Pulido, and M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 256–279, Jun. 2004.

[20] S. Mirjalili, S. Saremi, S. M. Mirjalili, and L. D. S. Coelho, "Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization," *Expert Syst. Appl.*, vol. 47, pp. 106–119, Apr. 2016, doi: 10.1016/j.eswa.2015.10.039.

[21] S. Adarsh and M. J. Reddy, "Multiscale characterization and prediction of reservoir inflows using MEMD-SLR coupled approach," *J. Hydrol. Eng.*, vol. 24, no. 1, Jan. 2019, Art. no. 04018059.

[22] H. M. Nazir, I. Hussain, M. Faisal, E. E. Elashkar, and A. M. Shoukry, "Improving the prediction accuracy of river inflow using two data preprocessing techniques coupled with data-driven model," *PeerJ*, vol. 7, p. e8043, Dec. 2019.

[23] J. Tan, W. Fu, K. Wang, X. Xue, W. Hu, and Y. Shan, "Fault diagnosis for rolling bearing based on semi-supervised clustering and support vector data description with adaptive parameter optimization and improved decision strategy," *Appl. Sci.*, vol. 9, no. 8, p. 1676, Apr. 2019.

[24] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014. [Online]. Available: http://link.springer.com/10.1007/978-3-319-14612-6_15 and http://ieeexplore.ieee.org/document/6655981/

[25] H. Luo, D. Wang, C. Yue, Y. Liu, and H. Guo, "Research and application of a novel hybrid decomposition-ensemble learning paradigm with error correction for daily PM 10 forecasting," *Atmos. Res.*, vol. 201, pp. 34–45, Mar. 2018.

[26] Y. Zhang, K. Liu, L. Qin, and X. An, "Deterministic and probabilistic interval prediction for short-term wind power generation based on variational mode decomposition and machine learning methods," *Energy Convers. Manage.*, vol. 112, pp. 208–219, Mar. 2016, doi: 10.1016/j.enconman.2016.01.023.

[27] Z. Qian, Y. Pei, H. Zareipour, and N. Chen, "A review and discussion of decomposition-based hybrid models for wind energy forecasting applications," *Appl. Energy*, vol. 235, pp. 939–953, Feb. 2019, doi: 10.1016/j.apenergy.2018.10.080.

[28] B. He, L.-Z. Liao, D. Han, and H. Yang, "A new inexact alternating directions method for monotone variational inequalities," *Math. Program.*, vol. 92, no. 1, pp. 103–118, Mar. 2002. [Online]. Available: http://link.springer.com/10.1007/s101070100280

[29] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0965997813001853, doi: 10.1016/j.advengsoft.2013.12.007.
[30] Y. Zhang, Y. Zhao, C. Kong, and B. Chen, "A new prediction method based on VMD-PRBF-ARMA-E model considering wind speed characteristic," *Energy Convers. Manage.*, vol. 203, Jan. 2020, Art. no. 112254, doi: 10.1016/j.enconman.2019.112254.
[31] X. Wen, Q. Feng, R. C. Deo, M. Wu, Z. Yin, L. Yang, and V. P. Singh, "Two-phase extreme learning machines integrated with the complete ensemble empirical mode decomposition with adaptive noise algorithm for multi-scale runoff prediction problems," *J. Hydrol.*, vol. 570, pp. 167–184, Mar. 2019, doi: 10.1016/j.jhydrol.2018.12.060.
[32] F. He, J. Zhou, Z.-K. Feng, G. Liu, and Y. Yang, "A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm," *Appl. Energy*, vol. 237, pp. 103–116, Mar. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0306261919300571, doi: 10.1016/j.apenergy.2019.01.055.
[33] J. Wang, S. Wang, and W. Yang, "A novel non-linear combination system for short-term wind speed forecast," *Renew. Energy*, vol. 143, pp. 1172–1192, Dec. 2019.
[34] S. Mirjalili, P. Jangir, and S. Saremi, "Multi-objective ant lion optimizer: A multi-objective optimization algorithm for solving engineering problems," *Appl. Intell.*, vol. 46, no. 1, pp. 79–95, 2017.
[35] S. Mirjalili, P. Jangir, S. Z. Mirjalili, S. Saremi, and I. N. Trivedi, "Optimization of problems with multiple objectives using the multi-verse optimization algorithm," *Knowl.-Based Syst.*, vol. 134, pp. 50–71, Oct. 2017.
[36] H. Liu, Z. Duan, Y. Li, and H. Lu, "A novel ensemble model of different mother wavelets for wind speed multi-step forecasting," *Appl. Energy*, vol. 228, pp. 1783–1800, Oct. 2018.
[37] H. Liu and C. Chen, "Multi-objective data-ensemble wind speed forecasting model with stacked sparse autoencoder and adaptive decomposition-based error correction," *Appl. Energy*, vol. 254, Nov. 2019, Art. no. 113686.
[38] H. Liu, H. Wu, and Y. Li, "Smart wind speed forecasting using EWT decomposition, GWO evolutionary optimization, RELM learning and IEWT reconstruction," *Energy Convers. Manage.*, vol. 161, pp. 266–283, Apr. 2018.
[39] P. Lu, L. Ye, W. Zhong, Y. Qu, B. Zhai, Y. Tang, and Y. Zhao, "A novel spatio-temporal wind power forecasting framework based on multi-output support vector machine and optimization strategy," *J. Cleaner Prod.*, vol. 254, May 2020, Art. no. 119993.
[40] P. J. Gemperline, J. R. Long, and V. G. Gregoriou, "Nonlinear multivariate calibration using principal components regression and artificial neural networks," *Anal. Chem.*, vol. 63, no. 20, pp. 2313–2323, Oct. 1991.
[41] Y.-W. Huang and D.-H. Lai, "Hidden node optimization for extreme learning machine," *AASRI Procedia*, vol. 3, pp. 375–380, Jan. 2012.
[42] Z. Yang and J. Wang, "A combination forecasting approach applied in multistep wind speed forecasting based on a data processing strategy and an optimized artificial intelligence algorithm," *Appl. Energy*, vol. 230, pp. 1108–1125, Nov. 2018.
[43] T. Niu, J. Wang, H. Lu, and P. Du, "Uncertainty modeling for chaotic time series based on optimal multi-input multi-output architecture: Application to offshore wind speed," *Energy Convers. Manage.*, vol. 156, pp. 597–617, Jan. 2018.
[44] S. Chen, S. Dong, Z. Cao, and J. Guo, "A compound approach for monthly runoff forecasting based on multiscale analysis and deep network with sequential structure," *Water*, vol. 12, no. 8, p. 2274, Aug. 2020.
[45] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Statist.*, vol. 20, no. 1, pp. 134–144, Jan. 2002.
[46] B. Wilamowski, "Neural network architectures and learning algorithms," *IEEE Ind. Electron. Mag.*, vol. 3, no. 4, pp. 56–63, Dec. 2009.
[47] K. E. Taylor, "Summarizing multiple aspects of model performance in a single diagram," *J. Geophys. Res., Atmos.*, vol. 106, no. D7, pp. 7183–7192, Apr. 2001.

SHI CHEN was born in Yichang, Hubei, China, in 1983. He received the B.S. and M.S. degrees in hydrogeology engineering from the China University of Geosciences, Wuhan, China, in 2009, and the Ph.D. degree in mineral prospecting and exploration from the China Coal Research Institute, Xi'an, Shaanxi, China, in 2015. From 2015 to 2016, he was an Assistant Researcher of Hydrogeology at Xi'an Research Institute of China Coal Technology and Engineering Group Corporation. Since 2016, he has been a Lecturer with the Department of Hydrology and Water Resources Engineering, College of Water Resources and Architectural Engineering, Northwest A&F University. His main research interests include analytical solution of well flow and deep learning algorithms to predict mine water inflow.

SHUNING DONG was born in 1961. He is currently a Researcher and a Ph.D. Tutor. He is also the Chairman and the General Manager at Xi'an Research Institute of China Coal Technology and Engineering Group Corporation, where he is also a Chief Scientist, enjoying special government allowances of the state council and also an Outstanding Contribution Expert of Shaanxi. He is a part-time Professor with Jilin University, Northwest University, and the Xi'an University of Science and Technology. Moreover, he has published 12 monographs (including two foreign languages) and more than 30 articles and three invention patents. He has been engaged in the work of scientific research and business and management for hydrogeology and engineering geology in the last 30 years. He has hosted or participated in nearly 30 research programs, including the United Nations Development Programme (UNDP) funded program, the National Basic Research Program of China (973 Program), the National 11th Five-Year Plan Science and Technology Support Program, National Industrial Test, and the National Key Scientific and Technological Innovation Program, 16 of these programs have obtained the National and Provincial Awards and the National Science and Technology Progress Awards. He is the Vice Director of the Hydrogeology Group of Committee of Experts of State Administration of Coal Mine Safety, a member of the Expert Group of International Scientific and Technological Cooperation in the Ministry of Science and Technology, the Expert Group of Review of National Natural Science Foundation of China, IAH, and IMWA, and the Director of the Professional Committee of Prevention and Control of Coal Mine Water in Coal Industry Committee of Technology.