

Research Article

Vibration Tendency Prediction Approach for Hydropower Generator Fused with Multiscale Dominant Ingredient Chaotic Analysis, Adaptive Mutation Grey Wolf Optimizer, and KELM

Wenlong Fu ^{1,2}, Kai Wang ^{1,2}, Jiawen Tan ^{1,2} and Kaixuan Shao ^{1,2}

¹College of Electrical Engineering & New Energy, China Three Gorges University, Yichang 443002, China

²Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station, China Three Gorges University, Yichang 443002, China

Correspondence should be addressed to Wenlong Fu; ctgu_fuwenlong@126.com

Received 28 June 2019; Revised 26 October 2019; Accepted 25 November 2019; Published 12 February 2020

Academic Editor: Alex Alexandridis

Copyright © 2020 Wenlong Fu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate vibrational tendency forecasting of hydropower generator unit (HGU) is of great significance to guarantee the safe and economic operation of hydropower station. For this purpose, a novel hybrid approach combined with multiscale dominant ingredient chaotic analysis, kernel extreme learning machine (KELM), and adaptive mutation grey wolf optimizer (AMGWO) is proposed. Among the methods, variational mode decomposition (VMD), phase space reconstruction (PSR), and singular spectrum analysis (SSA) are suitably integrated into the proposed analysis strategy. First of all, VMD is employed to decompose the monitored vibrational signal into several subseries with various frequency scales. Then, SSA is applied to divide each decomposed subseries into dominant and residuary ingredients, after which an additional forecasting component is calculated by integrating the residual of VMD with all the residuary ingredients orderly. Subsequently, the proposed AMGWO is implemented to simultaneously adapt the intrinsic parameters in PSR and KELM for all the forecasting components. Ultimately, the prediction results of the raw vibration signal are obtained by assembling the results of all the predicted prediction components. Furthermore, six relevant contrastive models are adopted to verify the feasibility and availability of the modified strategies employed in the proposed model. The experimental results illustrate that (1) VMD plays a positive role for the prediction accuracy promotion; (2) the proposed dominant ingredient chaotic analysis based on the realization of time-frequency decomposition can further enhance the capability of the forecasting model; and (3) the appropriate parameters for each forecasting component can be optimized by the proposed AMGWO effectively, which can contribute to elevating the forecasting performance distinctly.

1. Introduction

Hydropower generator unit (HGU) is the key equipment of hydropower stations, which plays an important role in emergency reserve as well as regulation of peak load and frequency. Besides, the stability and safety of operation for hydropower stations and power grid depends on the healthy condition of HGU heavily [1–3], which can be revealed by the corresponding vibration signal monitored by specific sensors [4]. Since the faults of HGU are generally revealed by the vibration status of various components in HGU in practice engineering, it is important to achieve the accurate vibrational tendency forecasting for HGU [5]. Accordingly,

the abnormal status of HGU can be disclosed before the accident as well as conducting scientific and reasonable maintenance plans. Nevertheless, the dynamic behaviors of HGU are extremely complicated due to frequently converted working conditions and inherent coupling of hydraulic, mechanical, and electromagnetic factors [6, 7].

As mentioned above, the faults' information and the equipment status can be excavated from the corresponding vibration signal. Hence, the potential fault information can be effectively excavated by accurate prediction of vibration trends, while the forecasting of vibration tendency can be equivalent to the problem of time series prediction that achieves prediction by adequately utilizing historical status.

For this purpose, various state-of-the-art prediction techniques have been developed in practice engineering, which can be classified into statistical models and artificial intelligence (AI) models [8]. Statistical models, such as autoregressive (AR) [9], auto regressive moving average (ARMA) [10], auto regressive integrated moving average (ARIMA) [11], autoregressive fractionally integrated moving average (ARFIMA) [12] and generalized autoregressive conditional heteroscedasticity (GARCH) [13], achieve time series forecasting by adequately extracting the implicit information within the historical datasets. Nevertheless, the prediction capability of such models would be restricted by the non-linearity and nonstationarity of the data. By contrast, AI models possess better adaptability to various data as well as better generalization ability. Among AI models, neural network (NN) [14, 15] can approximate any function in theory, whose network structure will be complex and difficult to determine with the increase of hidden layer number and dataset size. In contrast, support vector regression (SVR) [16] that implements forecasting by appropriate kernel mapping possesses less parameters to be determined, while the computation consumption will increase as the data scale increases. In addition, extreme learning machine (ELM) [17, 18] has been widely developed in various fields due to the advantages of low computational consumption and fewer parameters. Nevertheless, considering that only the empirical risk minimization principle is applied in ELM, regularization coefficient and kernel functions are introduced to ELM by Huang et al. [19], which can contribute to enhancing the generalization performance of ELM ulteriorly as well as weakening the uncertain results caused by randomly generated weights and biases of ELM. Therefore, kernel extreme learning machine (KELM) is investigated to achieve vibration tendency forecasting in this paper.

Due to the fact that the operation of HGU is usually accompanied by background noise and electromagnetic interference, the corresponding vibration signal usually possess strong nonlinearity and nonstationarity, which can greatly affect the predictive performance [6, 20]. To this end, various time-frequency decomposition approaches have been rapidly developed to efficaciously weaken the non-stationary and nonlinear data. For instance, Li et al. [21] performed wavelet transform (WT) on the engine cylinder head vibration signals for knock detection. Fei [22] investigated the prediction of bearing vibration signal combining empirical mode decomposition (EMD) with relevance vector machine optimized by artificial bee colony algorithm. Fu et al. [6] conducted a vibration trend measuring system combined with variational mode decomposition (VMD) and least square support vector machine (LSSVM). Among the decomposition methods mentioned above, VMD has been widely investigated in various fields due to the good adaptability and solid mathematical foundation when comparing with WT and EMD [23, 24]. Hence, VMD is selected as the data preprocessing approach in this paper, which is aimed at preliminarily making a dent in terms of nonstationarity for vibration signal collected in this paper. Additionally, considering that the redundant ingredients may still be included in the decomposed subseries, the

dominant and residuary ingredients of all the subseries are effectively separated by singular spectrum analysis (SSA), with which the forecasting performance could be significantly improved [25, 26]. Furthermore, to adequately excavate the inherent chaotic law for the well-processed subseries as well as deducing the inputs and outputs matrixes for predictor, phase space reconstruction (PSR) is adopted in this paper. Following the previous references [27, 28], prediction accuracy of the models based on PSR can be enhanced to a certain extent by feeding the appropriate parameters for PSR.

Additionally, the parameters of the approaches mentioned above can greatly affect the performance of the combined model. In view of this, various optimization algorithms based on different strategies have been focused for parameters optimization, such as genetic algorithm (GA) [29], region search evolutionary algorithm (RSEA) [30], artificial sheep algorithm (ASA) [31, 32], and grey wolf optimizer (GWO) [33]. To achieve better balance between convergence speed and accuracy, an adaptive mutation grey wolf optimizer (AMGWO) algorithm is proposed in this paper. Based on the above discussion, to achieve accurate vibration tendency forecasting for HGU, multiscale dominant ingredient chaotic analysis based on VMD, SSA, and PSR is organically integrated with KELM and AMGWO in this paper. To begin with, the collected vibration signal is primarily decomposed into several subseries by VMD, after which SSA is investigated to group the dominant and residuary ingredients from all the subseries. As a result, the dominant ingredients are treated as forecasting components, while the VMD residual is accumulated with the residuary ones for further prediction. Afterwards, parameters of PSR and KELM for each component are optimized by the proposed AMGWO algorithm, with which the predicted values corresponding to each component can be generated. Finally, the forecasting results of the raw signal are deduced by accumulating the values of all the predicted components. Furthermore, the effectiveness and superiority of the proposed approach was verified by practical engineering applications as well as detailed contrastive analysis.

The remaining parts of this paper are organized as follows. Section 2 illustrates the background knowledge of VMD, SSA, PSR, and KELM in detail. Section 3 details the contents of multiscale dominant ingredient chaotic analysis, the proposed AMGWO algorithm, optimization strategy, and the specific procedure of the proposed approach. Section 4 demonstrates the availability of the proposed approach by detailing experimental results and analysis. Section 5 summarizes the conclusions obtained in Section 4. In addition, the abbreviations of technical terms appeared in this paper are listed in the Abbreviations section.

2. Methodology

2.1. Variational Mode Decomposition. Variational mode decomposition (VMD), which was proposed by Dragomiretskiy et al. [34, 35], achieves adaptive signal processing by determining center frequency and bandwidth of the corresponding decomposed complement in the phase of

solving a variational problem. Compared with the well-investigated decomposition techniques including wavelet transform (WT) and empirical mode decomposition (EMD), VMD possesses better adaptability for various datasets as well as more complete mathematical theoretical basis. Additionally, following the previous references [6, 36], the superiority of forecasting performance promotion implemented by VMD has been widely investigated and verified. By setting a mode number K in advance, the given original signal f can be decomposed into K band-limited intrinsic mode functions [37], where the description of the corresponding constrained variational problem in VMD is as follows:

$$\begin{aligned} \min_{m_k, \omega_k} & \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * m_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} & \sum_k m_k(t) = f(t), \quad k = 1, 2, \dots, K, \end{aligned} \quad (1)$$

where m_k and ω_k denote the decomposed complements and the corresponding center frequencies, $\delta(t)$ represents the Dirac distribution, and $*$ indicates convolution operator. To solve such constrained problem, quadratic penalty term and Lagrangian multiplication operator λ are introduced, with which the problem can be transformed into an unconstrained one:

$$L(m_k, \omega_k, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * m_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k m_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k m_k(t) \right\rangle, \quad (2)$$

where α represents the balancing parameter constrained by data fidelity. Subsequently, the solution of such constrained variational problem can be treated as searching the saddle point of the augmented Lagrangian L by updating m_k , ω_k , and λ alternately, which is named as alternating direction method of the multipliers (ADMM) [38]. The iterative formulas of which can be deduced as follows:

$$\hat{m}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{m}_i(\omega) + \hat{\lambda}(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2}, \quad (3)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{m}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{m}_k(\omega)|^2 d\omega}, \quad (4)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \gamma \left(\hat{f}(\omega) - \sum_k \hat{m}_k^{n+1}(\omega) \right), \quad (5)$$

where γ denotes the time step of the dual ascents and \hat{m}_k^{n+1} , $\hat{m}_i(\omega)$, $\hat{f}(\omega)$, and $\hat{\lambda}(\omega)$ indicate the Fourier transform corresponding to m_k^{n+1} , $m_i(t)$, $f(t)$, and $\lambda(t)$, respectively. Additionally, the specific main procedures of VMD are exhibited below:

- Step 1: initialize \hat{m}_k^1 , ω_k^1 , β^1 and $n = 1$
- Step 2: execute loop, $n = n + 1$
- Step 3: update \hat{m}_k and ω_k by equations (3) and (4)
- Step 4: update $\hat{\lambda}$ based on equation (5)
- Step 5: if $\sum_k \|\hat{m}_k^{n+1} - \hat{m}_k^n\|_2^2 / \|\hat{m}_k^n\|_2^2 < \varepsilon$, end loop; otherwise, turn to Step 2 for further iteration

It is worth noting that the decomposition effectiveness and efficiency of VMD are affected by the inherent parameters, namely, mode number K and time step of the dual ascents γ , respectively [34]. To achieve better decomposition effectiveness as well as more accurate prediction for vibration tendency, the aforementioned parameters are pre-determined by grid search in this paper.

2.2. Singular Spectrum Analysis. Singular spectrum analysis (SSA) is a novel time series preprocessing technique, which has been widely investigated to identify and extract periodic, quasiperiodic, and oscillating components from raw data [39]. To separate the characteristic tendency terms and the residuary ingredients from the decomposed subseries quickly and easily, such feature selection method, namely, singular value decomposition (SVD), possessing small computational cost and outstanding effect, is employed to handle the extraction task. Four main procedures, namely, embedding, SVD, grouping, and diagonal averaging, are contained in SSA [26]. Specifically, the corresponding detailed description of these four operations is exhibited as follows [24]:

- (1) *Embedding.* Given a time series with N length $x = \{x_i | i = 1, 2, \dots, N\}$, the series can be embedded into a Hankel matrix [40] in advance:

$$H = (x_{ij})_{i,j=1}^{l,t} = \begin{bmatrix} x_1 & x_2 & \cdots & x_t \\ x_2 & x_3 & \cdots & x_{t+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_l & x_{l+1} & \cdots & x_N \end{bmatrix}, \quad (6)$$

where l denotes the window length and $t = N - l + 1$. In addition, all the elements along the diagonal $i + j = \text{const}$ are equal.

- (2) *SVD.* By implementing SVD on the reconstructed matrix H , the i -th singular value σ_i as well as singular vectors of matrixes HH^T and $H^T H$, namely, U_i and V_i , can be obtained effectively. Thus, the Hankel matrix H can be expressed as follows:

$$\begin{aligned} H &= H_1 + H_2 + \cdots + H_l, \\ H_i &= \sigma_i U_i V_i^T. \end{aligned} \quad (7)$$

- (3) *Grouping.* In this phase, the aforementioned matrices H_i will be partitioned into m group. Let $Z = \{Z_1, Z_2,$

$\dots, Z_r\}$, Z_1, Z_2, \dots, Z_r denote the indices of each group; the Z -th group matrix H_Z is expressed as follows:

$$H_Z = H_{Z_1} + H_{Z_2} + \dots + H_{Z_r}. \quad (8)$$

It is worth mentioning that the set of indices $Z = \{1, \dots, l\}$ are partitioned into two subsets, namely, dominant and residuary ingredients, in this paper, that is, $Z_1 = \{1, 2, \dots, s\}$ and $Z_2 = \{s+1, s+2, \dots, l\}$, with which the matrix H_Z could be represented as

$$H_Z = H_{Z_1} + H_{Z_2}. \quad (9)$$

(4) *Diagonal Averaging.* With the diagonal averaging strategy, the grouped matrix can be converted into a new series with length N . Here, assume matrix X is an $L \times S$ matrix with elements x_{ij} , where $i \leq L \leq L$ and $1 \leq j \leq S$. Let $x_{ij}^* = x_{ij}$ when $L < S$; otherwise, let $x_{ij}^* = x_{ji}$. Then, the m -th restructured point V_m ($m = 1, 2, \dots, N$) is calculated by following formulas:

$$V_m = \begin{cases} \frac{1}{m} \sum_{i=1}^m x_{i,m-i+1}^*, & \text{for } 1 \leq m < L^*, \\ \frac{1}{L^*} \sum_{i=1}^{L^*} x_{i,m-i+1}^*, & \text{for } L^* \leq m \leq S^*, \\ \frac{1}{N-m+1} \sum_{i=m-S^*+1}^{T-S^*} x_{i,m-i+1}^*, & \text{for } S^* < m \leq N, \end{cases} \quad (10)$$

where $L^* = \min(L, S)$ and $S^* = \max(L, S)$. Furthermore, the visual representation of diagonal averaging is depicted in Figure 1.

2.3. Phase Space Reconstruction. Phase space reconstruction (PSR) is the basis of chaotic time series prediction, which can explore the potential law of the series by reconstructing a time series with chaotic characteristics into a low-order nonlinear dynamic system. Delay coordinates method proposed by Packard et al. [41] is one of the most popular approaches that could restore the original dynamic system effectively. PSR considering various time delays is adopted to construct appropriate input matrix for predictors, with which the potential associated information contained in historical data can be adequately exploited compared with the case of fixed time delay with 1. Meanwhile, the essence of such approach is to reconstruct a one-dimensional time series into a d -dimensional phase space vector with time delay τ . Therefore, the PSR expression for the monitored vibration series $x = \{x_i | i = 1, 2, \dots, N\}$ can be indicated as follows:

$$X = [X_1 \ X_2 \ \dots \ X_J]^T = \begin{bmatrix} x_1 & x_{1+\tau} & \dots & x_{1+(d-1)\tau} \\ \vdots & \vdots & \ddots & \vdots \\ x_i & x_{i+\tau} & \dots & x_{i+(d-1)\tau} \\ \vdots & \vdots & \ddots & \vdots \\ x_J & x_{J+\tau} & \dots & x_{J+(d-1)\tau} \end{bmatrix}, \quad (11)$$

where $J = N - \tau \cdot (d-1)$, N represents the total number of wind speed samples, X_i denotes the i -th space vector in the reconstructed phase space, and d and τ denote embedded dimension and time delay, respectively. Furthermore, the i -th output point O_i corresponding to vector X_i can be deduced as follows:

$$O = [O_1 \ O_2 \ \dots \ O_J]^T = [x_{1+(d-1)\tau} \ x_{2+(d-1)\tau} \ \dots \ x_N]^T. \quad (12)$$

The key to accurate prediction applying PSR is to preset appropriate parameters. For this purpose, the embedded dimension and time delay within PSR will be synchronously optimized with the parameters of predictor by the newly modified optimization algorithm, which will be detailed in later sections.

2.4. Kernel Extreme Learning Machine. For the normal extreme learning machine (ELM), the input weights and biases are generated randomly as well as being fixed in subsequent calculations. Besides, the minimal norm least square method is employed in ELM, with which the output weights β can be deduced by solving the set of linear equations $\mathbf{H}\beta = \mathbf{T}$ and described as follows:

$$\beta = \mathbf{H}^\dagger \mathbf{T}, \quad (13)$$

where \mathbf{H}^\dagger denotes the Moore–Penrose generalized inverse of matrix \mathbf{H} . As mentioned above, the input weights and biases are randomly generated in ELM, with which the output values generally performed randomness as well as generating negative impacts on the ultimate results. To enhance the generalization and the universality of ELM [42], a modified version of ELM combining kernel functions is proposed by Huang et al. [19]. By considering both minimizing training errors and output weights norms for ELM, a better generalization performance of the networks could be achieved. Hence, the regularization coefficient C is introduced in the optimization stage, thus deducing the output weights β as follows [19]:

$$\beta = \mathbf{H}^T \left(\mathbf{H}\mathbf{H}^T + \frac{\mathbf{I}}{C} \right)^{-1} \mathbf{T}, \quad (14)$$

where \mathbf{I} represents an identity matrix of dimension N . Additionally, kernel matrix is adopted to handle the cases that the hidden layer feature mapping $h(\bullet)$ would be unknown, where the kernel matrix for kernel extreme learning machine (KELM) is calculated as follows [19]:

$$\Omega = \mathbf{H}\mathbf{H}^T: \Omega_{\text{ELM},ij} = h(x_i) \cdot h(x_j) = K(x_i, x_j), \quad (15)$$

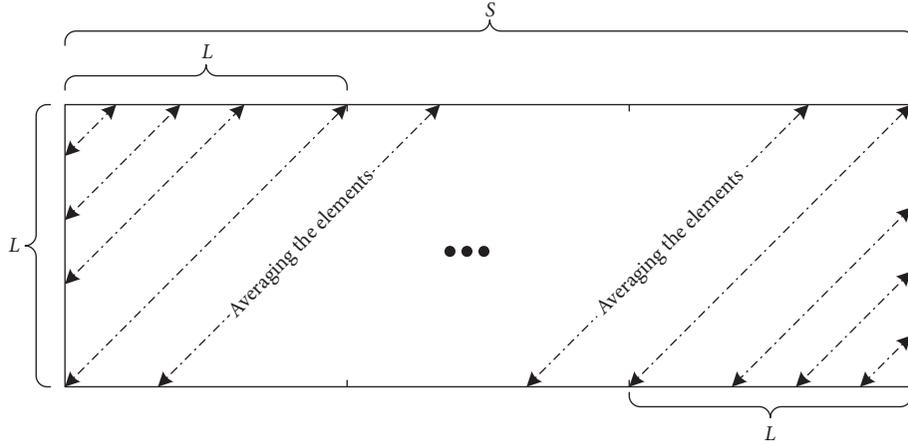


FIGURE 1: Visualization of diagonal averaging using singular spectrum analysis.

where $K(\bullet, \bullet)$ denotes the kernel functions. Organizing equations (13)–(15), the output functions of KELM can be formulated as follows:

$$f(x) = h(x) \cdot \beta = h(x) \cdot \mathbf{H}^T \left(\mathbf{H}\mathbf{H}^T + \frac{\mathbf{I}}{C} \right)^{-1} \mathbf{T}$$

$$= \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix} \left(\Omega_{\text{ELM}} + \frac{\mathbf{I}}{C} \right)^{-1} \mathbf{T}. \quad (16)$$

According to the previous references [43, 44], radial basis function (RBF) is one of the most popular kernel functions and has been recognized as an effective one, which is defined as follows:

$$K(x, y) = \exp\left(\frac{-\|x - y\|^2}{\sigma^2}\right), \quad (17)$$

where σ^2 denotes the kernel parameter. Following the investigation [19], the forecasting capability of KELM will be significantly influenced by the regularization coefficient C and the kernel parameter σ^2 . Hence, to achieve better generalization performance of the networks, the above two parameters need to be determined appropriately. Specifically, the aforementioned two parameters will be simultaneously optimized with the parameters of PSR.

3. The Proposed Approach

3.1. Multiscale Dominant Ingredient Chaotic Analysis. Due to the chaotic nature and intrinsic complexity of the original vibration signal, the prediction performance of single model or models without signal preprocessing would be severely restricted. To this end, VMD is employed to preliminarily decompose the monitored vibration signal into several components with various scale frequencies. According to references [45, 46], the decomposition efficiency and effectiveness of VMD are greatly affected by parameters K and γ mentioned in Section 2.1. In addition, considering that the noises may still be contained in the

decomposed series, SSA is implemented to extract the dominant and residuary ingredients corresponding to each decomposed series severally, thus extracting characteristic trends of the nonstationary subseries. As mentioned in Section 2.3, the indices Z introduced in the grouping phase of SSA is divided into two discrete subsets, i.e., $\{1, 2, \dots, s\}$ and $\{s+1, s+2, \dots, l\}$; it is worth mentioning that the forecasting accuracy can be affected by the parameter s to a certain extent [47]. Additionally, the residual of VMD calculated by $m_r = f - \sum_{k=1}^K m_k$ is assembled with the residuary ingredients separated from all the subseries, which is dedicated to ulteriorly improving the capabilities of the forecasting model. Subsequently, as an effective chaotic sequence analysis tool, PSR is employed to generate the inputs and outputs for the prediction models in advance [47, 48]. It is worth noting that the time delay τ and the embedded dimension d can affect the recovery of the PSR dynamic system to some extent, which would restrict the prediction performance significantly. As a result, the proposed multiscale dominant ingredient chaotic analysis can be implemented by setting the parameters for each module appropriately, which is also an urgent problem to be solved. To this end, a novel parameters optimization based on strategy adaptive mutation grey wolf optimizer (AMGWO) is proposed to achieve a significant improvement of the forecasting model, which will be detailed later.

3.2. Adaptive Mutation Grey Wolf Optimizer. Grey wolf optimizer (GWO) developed by Faris et al. [49] divides the wolf pack into four categories for simulating the leadership hierarchy, i.e., α , β , δ and ω , which are determined by the corresponding positions (fitness value). Then, the mathematical model of the encircling behavior can be formulated as

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)|, \quad (18)$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D},$$

where t indicates the t -th iteration, \vec{X}_p represents the position of the prey, \vec{X} indicates the position vector of a grey

wolf, and \vec{A} and \vec{C} are coefficient vectors calculated as follows:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a}, \quad (19)$$

$$\vec{C} = 2\vec{r}_2, \quad (20)$$

where r_1 and r_2 are random vectors in the scopes of $[0, 1]$. Considering that the parameter a updates with the linear rules in the normal GWO, which would attribute to make the solutions fall into local optimum, a nonlinear quadratic convex function is employed to guide the decreasing process of [50]. The specific updating equation is formulated as

$$a = 2 \times \left(1 - \left(\frac{t}{T} \right)^2 \right), \quad (21)$$

where t and T denote the current and the maximum iterations, respectively. In addition, Figure 2 intuitively demonstrates the variation trends of the factor over the course of the iterations with two functions. As can be observed from equation (21) and Figure 2, the convergence factors can adaptively and nonlinearly change over the course of iterations, thus achieving valid modulation of global and local searching capabilities for the algorithm.

The other stage, namely, hunting, is usually led by α wolf, while β and δ wolves participate in hunting occasionally [51]. Hence, α , β , and δ wolves are assumed to possess more knowledge about the potential location of prey, where the corresponding updating equations of α , β , and δ wolves are defined as follows:

$$\begin{aligned} \vec{D}_\alpha &= \left| \vec{C}_1 \times \vec{X}_\alpha - \vec{X} \right|; \\ \vec{D}_\beta &= \left| \vec{C}_2 \times \vec{X}_\beta - \vec{X} \right|; \\ \vec{D}_\delta &= \left| \vec{C}_3 \times \vec{X}_\delta - \vec{X} \right|, \end{aligned} \quad (22)$$

$$\begin{aligned} \vec{X}_1 &= \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha; \\ \vec{X}_2 &= \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta; \\ \vec{X}_3 &= \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta, \end{aligned} \quad (23)$$

where \vec{X}_1 , \vec{X}_2 , and \vec{X}_3 indicate the positional information owned by α , β , and δ wolves so far. Considering that the individuals' position in normal GWO is merely calculated by averaging the positions of α , β , and δ wolves, the difference in the importance of three wolves cannot be revealed by such simple averaging strategy [26]. To this end, a weighted averaging strategy is proposed for the iteration of individuals, where α , β , and δ wolves are separately assigned a weight value that is deduced by inverting the corresponding fitness values of the wolves in sequence. The detailed calculations are as follows:

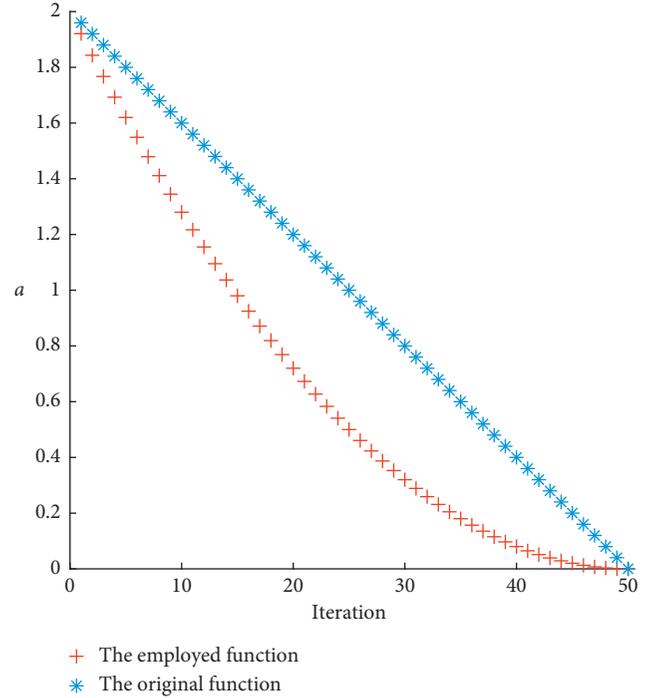


FIGURE 2: Convergence curves of \vec{a} over the course of iterations obtained by the employed function and the original one.

$$\begin{aligned} w_\alpha &= \frac{1}{\text{fitness}_\alpha(t)}, \\ w_\beta &= \frac{1}{\text{fitness}_\beta(t)}, \\ w_\delta &= \frac{1}{\text{fitness}_\delta(t)}, \end{aligned} \quad (24)$$

$$\vec{X}(t+1) = \frac{w_\alpha \cdot \vec{X}_1(t) + w_\beta \cdot \vec{X}_2(t) + w_\delta \cdot \vec{X}_3(t)}{w_\alpha + w_\beta + w_\delta}, \quad (25)$$

where w and fitness denote the weight and fitness value of the corresponding individual. To enrich the population diversity in the late iterations of the algorithm, mutation operation is employed to avert falling into local optimum effectively [52]. It is worth noting that the mutation strategy is merely implemented for updating α wolf, which can be described as

$$X_1(t) = \begin{cases} X_1(t-1) \times [1 + M \times (0.5 - \text{rand}())], & t = nT_p, \\ X_\alpha(t) - A_1 \cdot D_\alpha(t), & t \neq nT_p, \end{cases} \quad (26)$$

where $n = 1, 2, \dots$, $\text{rand}()$ represents the random number conforming to uniform distribution $U(0, 1)$ and M and T_p denote the coefficient and period of mutation, respectively. With the mutation operation introduced above, the positions of α wolf will be periodically mutated over the course of iterations, which makes the algorithm possess better capability of jumping out of local optimum. To verify the validity

and performance of the proposed AMGWO based on adaptive strategy and mutation operators, a set of benchmark functions are investigated to demonstrate the performance of various algorithms, which are detailed in Table 1. Besides, seven popular swarm intelligence optimization algorithms including particle swarm optimization (PSO), sine cosine algorithm (SCA), ant lion optimization (ALO), whale optimization algorithm (WOA), bat algorithm (BA), moth-flame optimization (MFO), and GWO are adopted to compare with the proposed AMGWO. For all the experimental algorithms, the number of searching agents is given as 50, while the maximum iterations are set as 200. Moreover, the ultimate values obtained by various algorithms are averaged by operating 30 times with various random seeds, where the convergence curve corresponding to each algorithm is depicted in Figure 3.

It can be observed from Figure 3 that the proposed AMGWO possesses faster convergence speed and much more appropriate solutions compared with the remaining algorithms. Specifically, following the comparison between GWO and AMGWO, it can be found that the proposed modified version of GWO based on adaptive strategy and mutation operator possesses better global search capability, which is attributed to the polytropic particles in the later stage of iterations caused by mutation operation. Furthermore, to describe the proposed AMGWO more clearly and intuitively, the pseudocode of the proposed AMGWO algorithm is exhibited in Algorithm 1.

3.3. Optimization Strategy. To effectively enhance the accuracy and effectiveness of the proposed hybrid structure, the proposed AMGWO is implemented to optimize the parameters of PSR and KELM for each subseries. To begin with, the key parameters K and γ in VMD are predetermined by grid search (GS), of which least-squares error index (LSEI) [6] is employed to be the objective function in GS. Then, the parameters of SSA are given according to the conclusions obtained in [47]. Since the parameters in KELM are not affected by the number of input layers, the parameters of PSR and KELM for each subseries are considered to be optimized synchronously. In addition, due to the accumulation of residuary ingredients and VMD residual, the total number of all the forecasting components is $K + 1$, while the coding strategy of the agents is exhibited in Figure 4. Furthermore, metric root-mean-square error (RMSE) formulated in the first row of Table 2 is employed as the objective function.

3.4. Specific Procedures. The major procedures of the proposed vibration tendency prediction approach combined with VMD, SSA, PSR, KELM, and AMGWO-based parameters optimization strategies are described as follows:

- Step 1: collect the monitoring vibration sequence
- Step 2: predetermine the mode number K and the updating step τ of VMD by minimizing the index LSEI with GS

Step3: decompose the raw vibration signal into K subseries and calculate the residual m_r of VMD

Step 4: implement SSA for the i -th sub-series, thus extracting the dominant ingredients as well as accumulating all the residuary ingredients with m_r

Step 5: obtain the optimal parameter sets $(\tau_i, d_i, C_i, \sigma_i^2)$ for the i -th forecasting component by AMGWO, where $i = 1, 2, \dots, K + 1$

Step 6: predict all the components with the well-trained model

Step 7: accumulate the predicted values of all the components to obtain ultimate forecasting values of the vibration signal

The overall process of the proposed hybrid vibration tendency prediction approach is exhibited in Figure 5.

4. Experimental Design

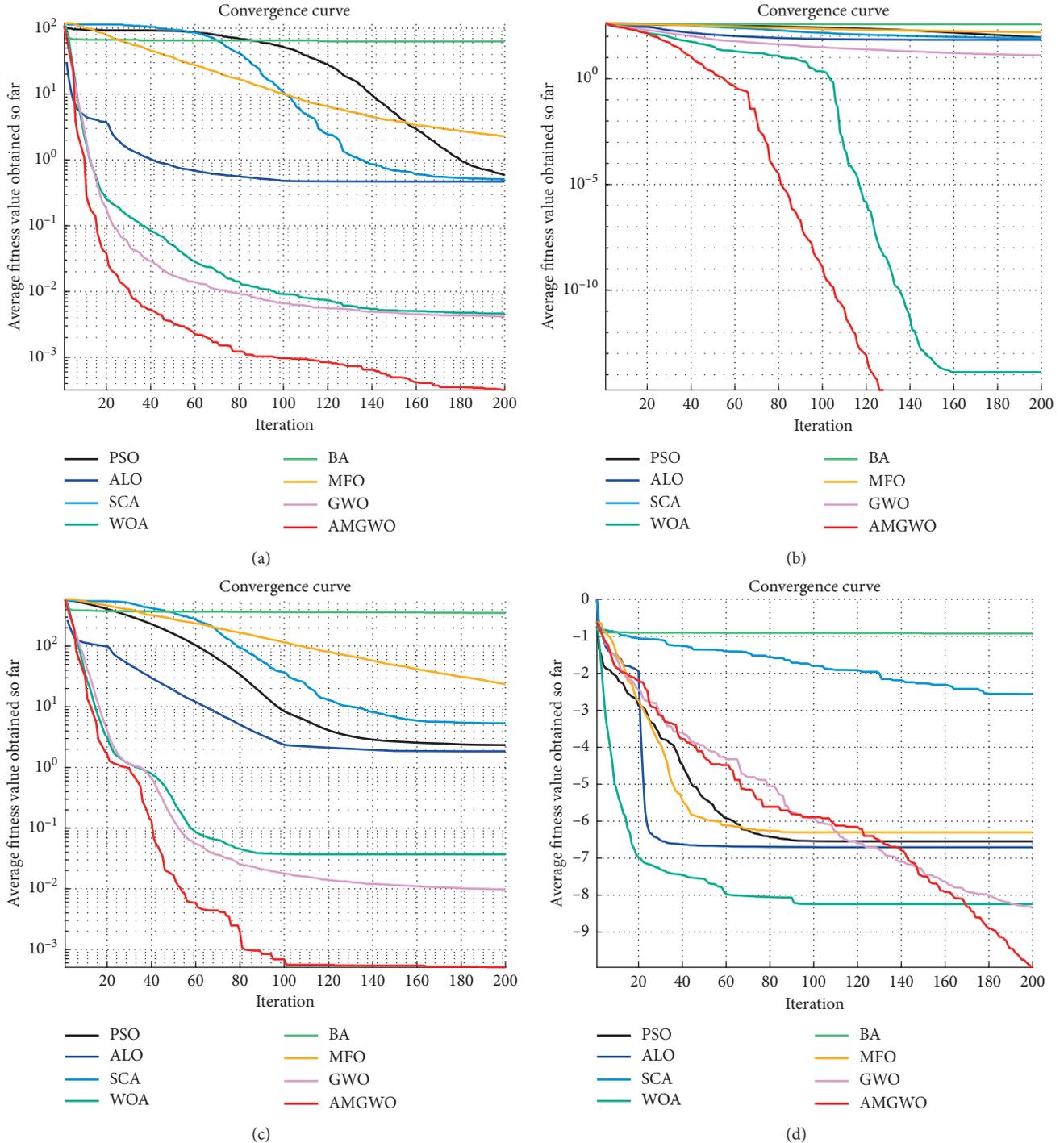
4.1. Data Description. In this section, a series of the vibration signal that are collected from the Ertan Hydropower Station in China is employed to verify the validity of the proposed hybrid forecasting approach. The general structure of a mixed current hydropower unit is depicted in Figure 6 [53]. Owing to the complex structures and frequently converted operation conditions within HGU, it is hard to guarantee that various monitored signals possess uniform time intervals. In this view, the monitored vibration signal is selected with the average time intervals for experimental analysis, thus conforming to the practical engineering [6]. For this purpose, there are total 216 samples collected from the swing data of upper guide in Y direction in this study, which come from 24 February 2011 to 4 March 2011 with the mean time interval of an hour. In addition, the collected vibration signal is exhibited in Figure 7, for which the maximum and minimum of the raw data are highlighted with red point. Meanwhile, the detailed statistical information of the monitored signal containing skewness (Skew.), kurtosis (Kurt.), standard deviation (Std.), minimum (Min.) value, maximum (Max.) value, and mean value is indicated in the bottom right corner of Figure 7. It can be observed that the raw vibrational data are accompanied by strong nonlinearity and nonstationarity, with which the forecasting accuracy may be significantly limited. It is worth noting that PSR will be implemented to generate the phase space matrices corresponding to the forecasting components before the predictor executed, among which the last 70 ones are taken as the testing set while the remaining is applied as training set.

4.2. Experimental Description

4.2.1. Contrastive Models and Evaluation Metrics. To achieve comprehensive verification of the availability and the superiority for the proposed hybrid forecasting approach, two benchmark methods and several hybrid models fused with different methods are adopted to achieve contrastive experiments. Among the contrast models, SVR and

TABLE 1: Benchmark functions employed in this section.

Function	Dim	Range	f_{\min}
Unimodal benchmark problems			
$F_7 = \sum_{i=1}^n ix_i^4 + \text{rand}[0, 1)$	30	$[-1.28, 1.28]$	0
Multimodal benchmark problems			
$F_9 = \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i) + 10]$	30	$[-5.12, 5.12]$	0
$F_{11} = 1/4000 \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos(x_i/\sqrt{i}) + 1$	30	$[-600, 600]$	0
Fixed-dimension multimodal benchmark problems			
$F_{21} = -\sum_{i=1}^5 [(X - a_i)(X - a_i)^T + c_i]^{-1}$	4	$[0, 10]$	-10.1532

FIGURE 3: Convergence curves of PSO, ALO, SCA, WOA, BA, MFO, GWO, and the proposed AMGWO on different functions: (a) F_7 , (b) F_9 , (c) F_{11} , and (d) F_{21} .

(1) Parameters:	
$iter_max$: Maximum iterations	\bar{X}_β : The second-best search agent
$search_agents$: The number of search agents	\bar{X}_δ : The third-best search agent
t : Current iteration	T_p : Period of variation
\bar{X}_α : The best search agents	M : Mutation extent coefficient
(2) Initialize the population $\bar{X}_i (i = 1, 2, \dots, search_agents)$	
(3) Initialize a , \bar{A} and \bar{C}	
(4) Calculate the fitness of each search member	
(5) $t = 0$	
(6) While ($t < iter_max$)	
(7) For each search agent:	
(8) Update the position of the current search agent on the basis of equations (24) and (25)	
(9) End for:	
(10) Update a , \bar{A} and \bar{C} by equations (21), (19) and (20), respectively.	
(11) Calculate the fitness of all grey wolves	
(12) Update the positions of α , β and δ wolves by equations (22) and (23), while the position of α wolf will be mutated every T iterations:	
(13) If $\text{mod}(t + 1 + T_p, T_p) = 0$ /* mod represents the remainder function */	
(14) $\bar{X}_1 = \bar{X}_1 \times (1 + M \times (0.5 - \text{rand}()))$	
(15) Else:	
(16) $\bar{X}_1 = \bar{X}_\alpha - \bar{A}_1 \cdot \bar{D}_\alpha$	
(17) End if	
(18) End else	
(19) $t = t + 1$	
(20) End while	
(21) Return \bar{X}_α	

ALGORITHM 1: The pseudocode of the proposed AMGWO algorithm.

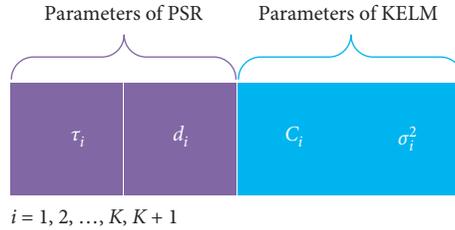


FIGURE 4: Coding strategy of the population in the proposed model based on AMGWO.

TABLE 2: Six metrics employed for quantitative evaluation of the proposed model.

Metrics	Definition	Equation
RMSE	Root-mean-square error	$RMSE = \sqrt{1/N \sum_{i=1}^N (Y_i - \hat{Y}_i)^2}$
MAE	Mean absolute error	$MAE = 1/N \sum_{i=1}^N Y_i - \hat{Y}_i $
MAPE	Absolute percentage error	$MAPE = 1/N \sum_{i=1}^N 100 \times (Y_i - \hat{Y}_i)/Y_i $
P_{RMSE}	Promoting percentages of root-mean-square error	$P_{RMSE} = ((RMSE_a - RMSE_b)/RMSE_a) \times 100$
P_{MAE}	Promoting percentages of mean absolute error	$P_{MAE} = ((MAE_a - MAE_b)/MAE_a) \times 100$
P_{MAPE}	Promoting percentages of mean absolute percentage error	$P_{MAPE} = ((MAPE_a - MAPE_b)/MAPE_a) \times 100$

KELM perform the prediction on the raw vibration signal without preprocessing, while GS is carried out for optimizing the parameters in these two models. Besides, the models merely combining time-frequency signal decomposition, namely, EMD-KELM and VMD-KELM, are adopted for comparing the performance between EMD and VMD. Moreover, the combined models, namely, EMD-SSA-PSR-KELM and VMD-SSA-PSR-KELM, achieve prediction

by implementing the proposed dominant ingredient chaotic analysis on the basis of decomposition technology. Furthermore, the parameters of KELM for the models mentioned above are all optimized by GS.

To achieve the quantitative assessment of various forecasting models, three common metrics including RMSE, mean absolute error (MAE), and mean absolute percentage error (MAPE) are employed, which can

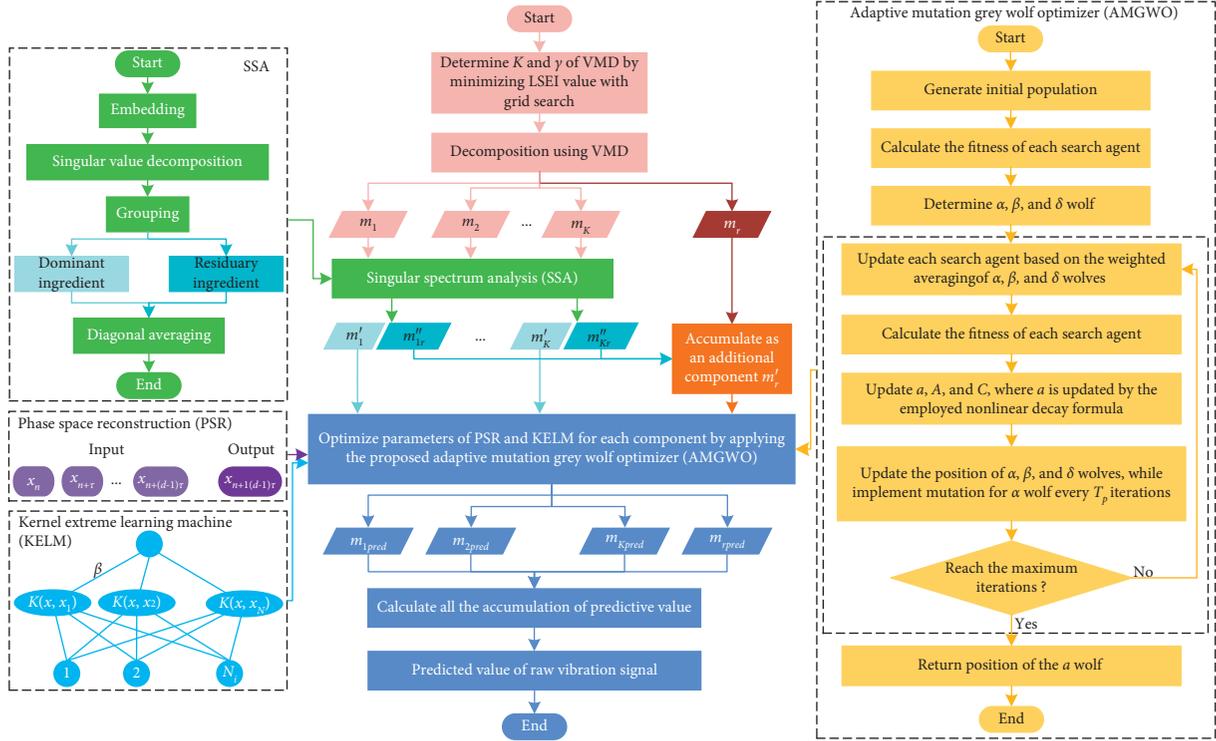


FIGURE 5: The procedures of the proposed approach for vibration tendency prediction.

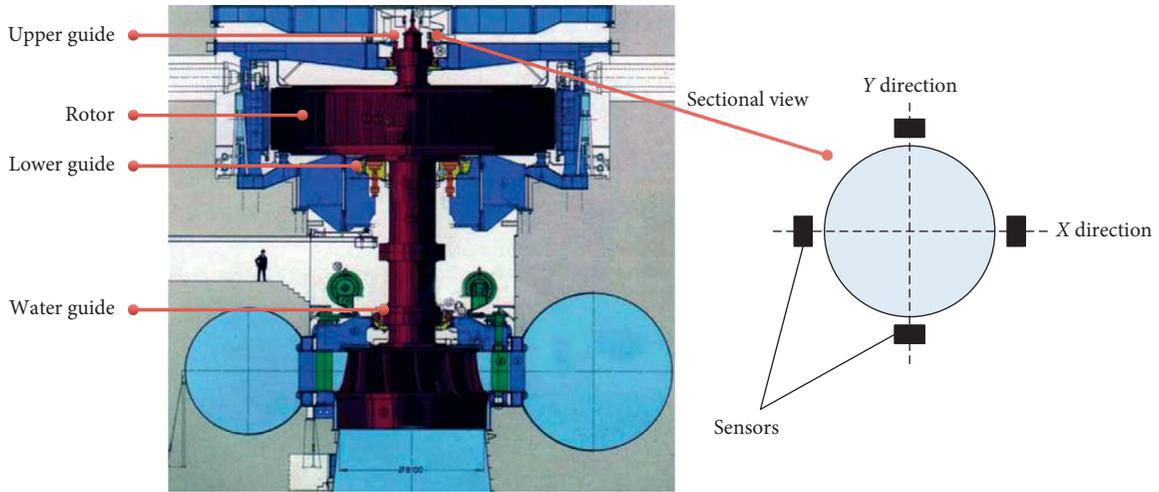


FIGURE 6: The structure of HGU and the corresponding monitoring location of sensors.

effectively represent the deviation between the predicted and collected values [54, 55]. Additionally, to further quantitatively compare the performance between the contrastive models and the proposed one, the descent ratios of the metrics mentioned above are adopted, which are expressed as P_{RMSE} , P_{MAE} , and P_{MAPE} . The specific definitions of these six metrics are minutely described in Table 2. Here, N denotes the total number of testing set. \hat{Y} and Y illustrate the predicted and monitored data, respectively. Furthermore, subscript a indicates the contrastive model, while subscript b expresses the proposed approach in this study.

4.2.2. Parameters Setting for All Experimental Models. For the contrastive models that achieve forecasting applying SVR and KELM, the regularization coefficient C and the kernel parameter σ^2 in such predictors are all optimized by GS. Thus, these two parameters are searched in $[2^{-8}, 2^8]$ and $[2^{-5}, 2^{-5}]$, respectively, while the searching step is 0.5 applied on the exponent. In addition, for the contrast models based on SSA and PSR, the window length l of Hankel and the grouping parameters s in SSA are given as 100 and 21, respectively [47], while the time delay time τ and embedding dimension d in PSR are set as 1 and 10 orderly. For all the experimental models fused with VMD, the parameters K and

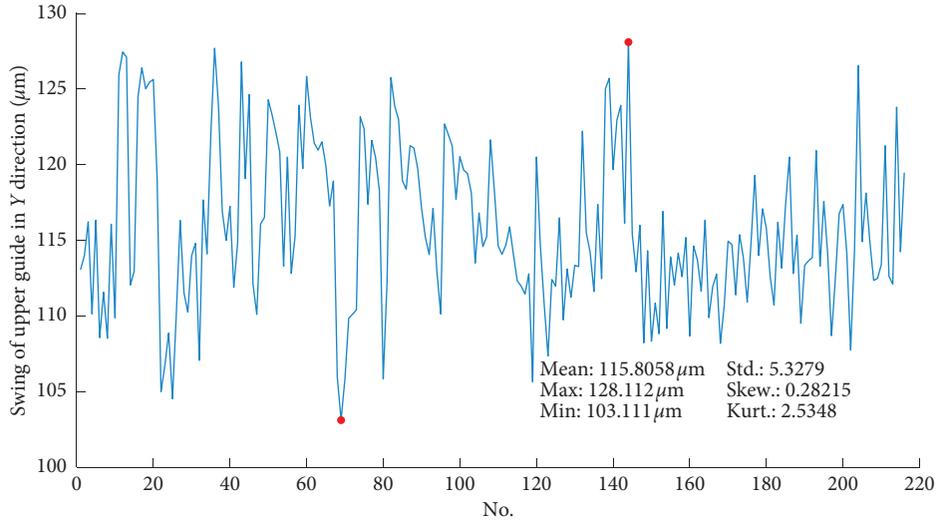
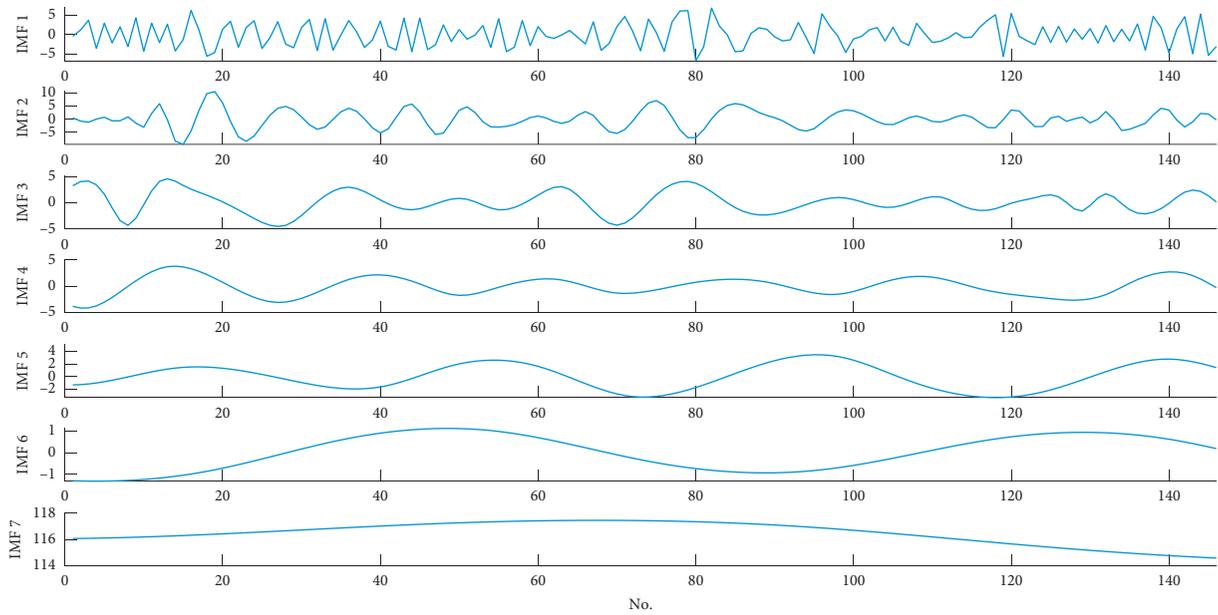


FIGURE 7: The swing data of upper guide in Y direction.

TABLE 3: Optimal parameters corresponding to each forecasting component for the proposed model.

Parameters	Forecasting components ($K = 10, \gamma = 1$)										
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_r
τ	1	1	1	1	1	1	1	1	1	2	3
d	3	5	4	10	6	10	22	4	2	2	17
C	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	11.67
σ^2	1.543	12.510	7.087	3.453	8.908	2.614	40.430	2.401	163.3	169.2	442



(a)

FIGURE 8: Continued.

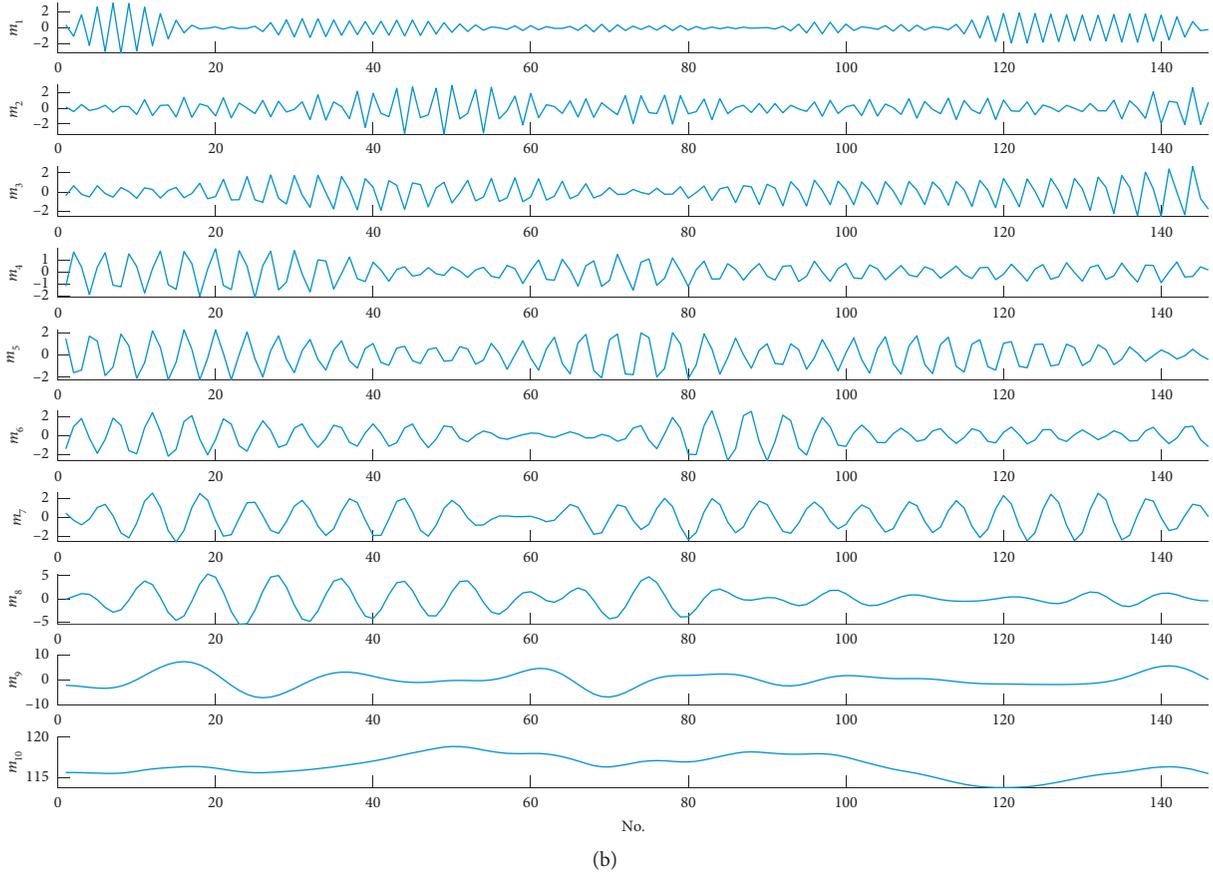


FIGURE 8: Subseries of the vibration signal decomposed by (a) empirical mode decomposition (EMD) and (b) variational mode decomposition (VMD).

TABLE 4: Evaluation metrics of various forecasting models on the test data.

Models	Metrics			
	RMSE (μm)	MAE (μm)	MAPE (%)	Time (s)
SVR (GS)	4.2948	3.4677	3.0594	10.51
KELM (GS)	3.8864	3.0813	2.6974	9.2
EMD-KELM (GS)	2.6751	2.1658	1.8931	80.47
VMD-KELM (GS)	0.7082	0.6046	0.5300	114.56
EMD-SSA-PSR-KELM (GS)	1.531	1.1842	1.0358	89.93
VMD-SSA-PSR-KELM (GS)	0.4936	0.3893	0.3431	124.96
The proposed model	0.2454	0.2034	0.1791	274.78

γ are predetermined by GS [18], where K is searched in the range of [2, 10] with increasing step 1, and γ is searched in the scope of [0, 1] with increasing step 0.1. Furthermore, the parameters of PSR and KELM for the proposed model are optimized by the proposed AMGWO, whose number of search agents, maximum iterations, period of variation, and mutation extent coefficient are set as 30, 50, 5, and 1 severally. Besides, the boundary of the parameters τ , d , C , and σ^2 are set in the interval [1, 2, 5, 25], [0.001, 1000] and [0.001, 1000] orderly, while the parameters of SSA in the proposed model are set as same as the contrastive models introduced above. The parameter sets (τ , d , C , σ^2) for all the forecasting components optimized by AMGWO are depicted in Table 3.

Meanwhile, the subseries decomposed by EMD and VMD (with optimal parameters) are illustrated in Figure 8 in detail, from which it can be seen that the strong nonstationary raw signal is decomposed into several subseries with major characteristic tendencies.

4.3. Contrastive Analysis. In this section, the experimental results of various forecasting models will be discussed in detail. The metrics RMSE, MAE, and MAPE obtained by all the comparison models and the proposed model are illustrated in Table 4. In addition, the descent ratio of metrics obtained by the proposed model when

TABLE 5: Promotion percentage of the proposed model compared with the contrastive models.

Extant models vs proposed	Promoting percentages		
	P_{RMSE} (%)	P_{MAE} (%)	P_{MAPE} (%)
SVR	94.29	94.13	94.15
KELM	93.69	93.40	93.36
EMD-KELM	90.83	90.61	90.54
VMD-KELM	65.35	66.36	66.21
EMD-SSA-PSR-KELM	83.97	82.82	82.71
VMD-SSA-PSR-KELM	50.28	47.75	47.80

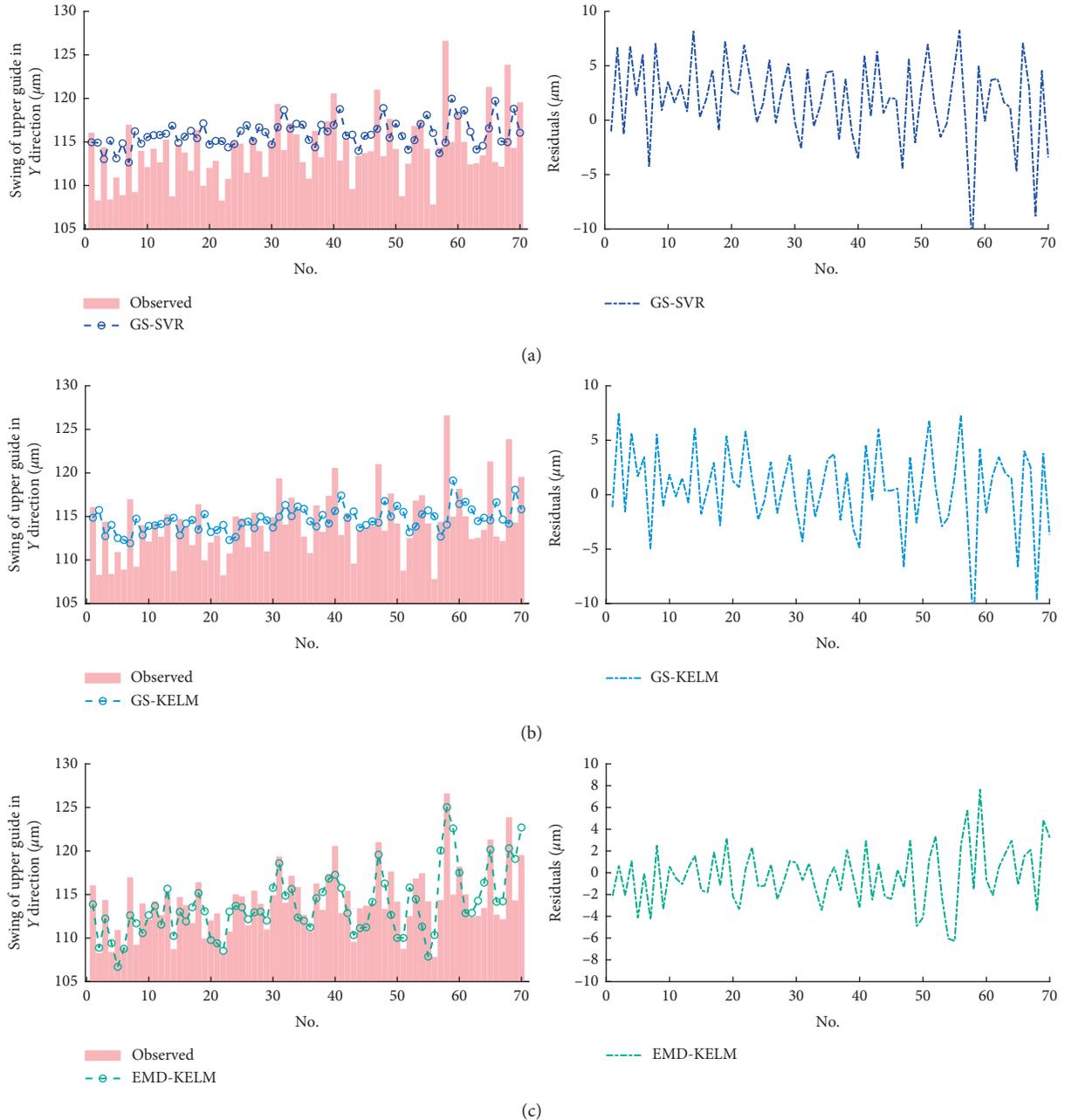


FIGURE 9: Continued.

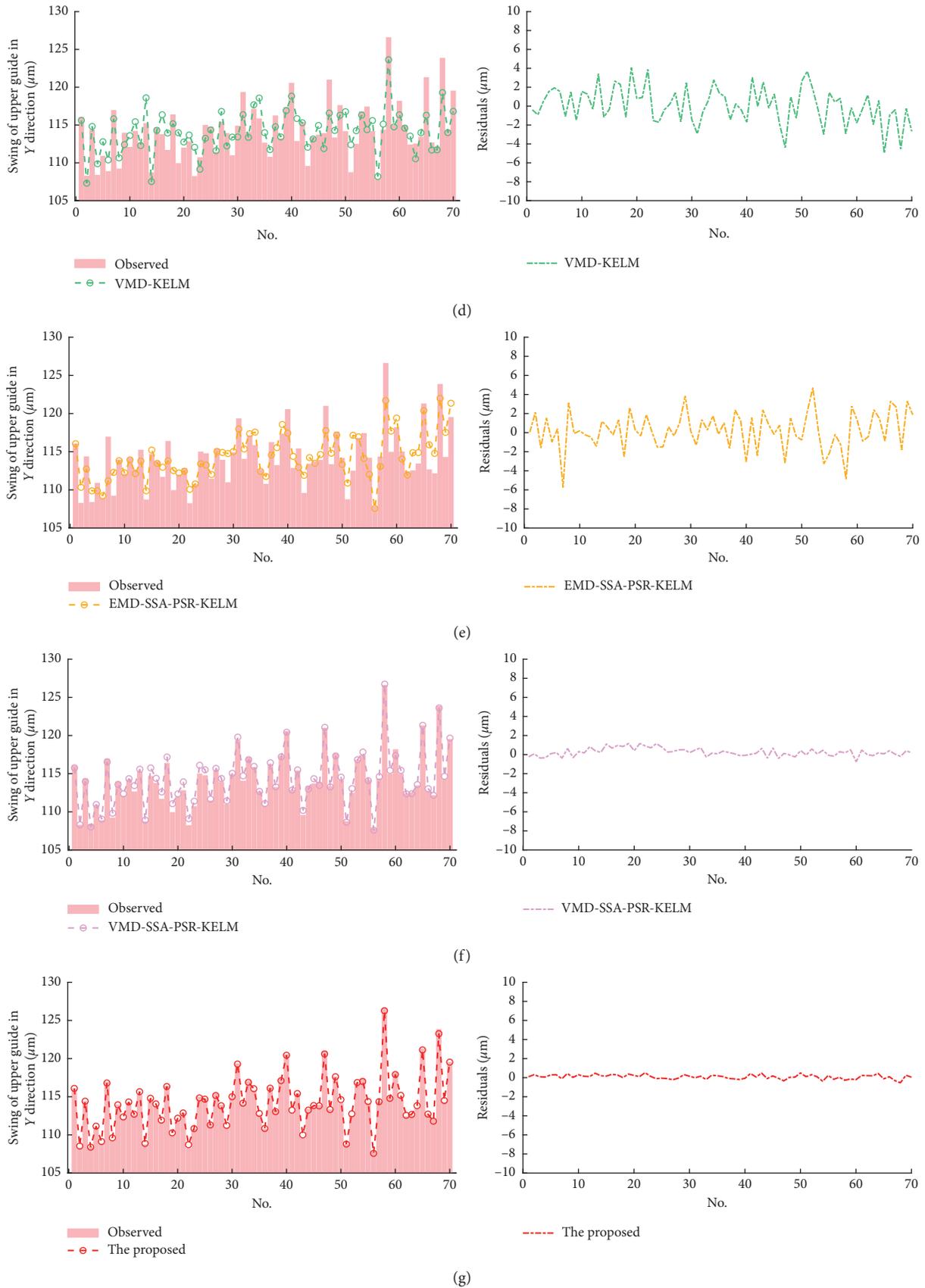


FIGURE 9: The forecasting results of various experimental models: (a) SVR, (b) KELM, (c) EMD-KELM, (d) VMD-KELM, (e) EMD-SSA-PSR-KELM, (f) VMD-SSA-PSR-KELM, and (g) the proposed model.

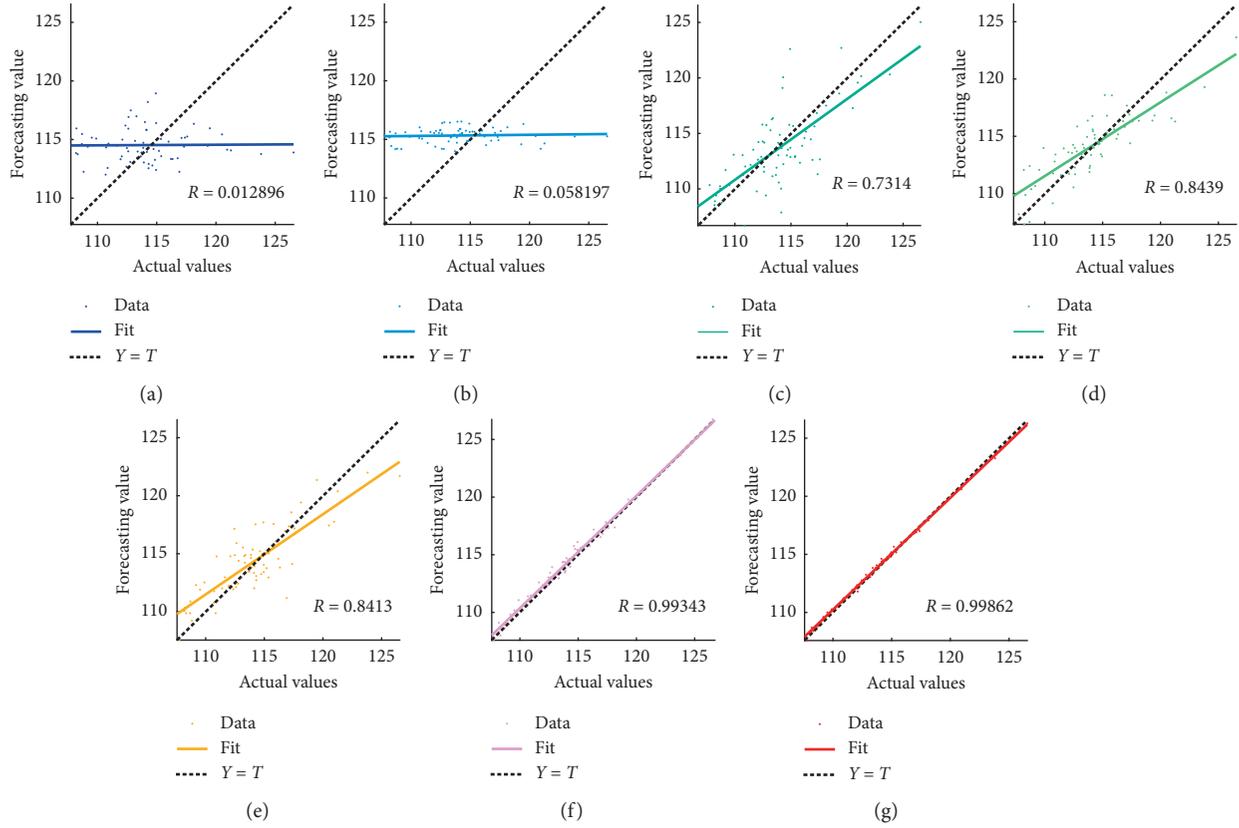
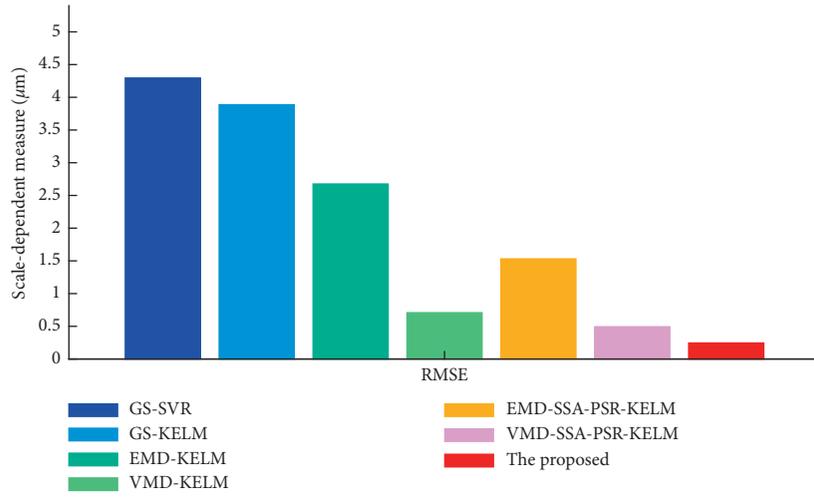


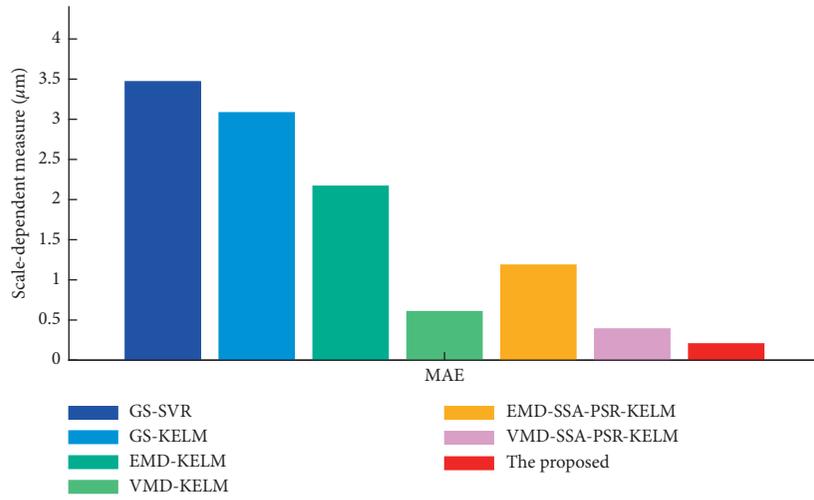
FIGURE 10: Scatter plots illustrating the fitting degree of each model in the order of Table 4.

compared to the relevant models is presented in Table 5 integrally. Analysis of the experimental results described in these two tables can lead to the following conclusions:

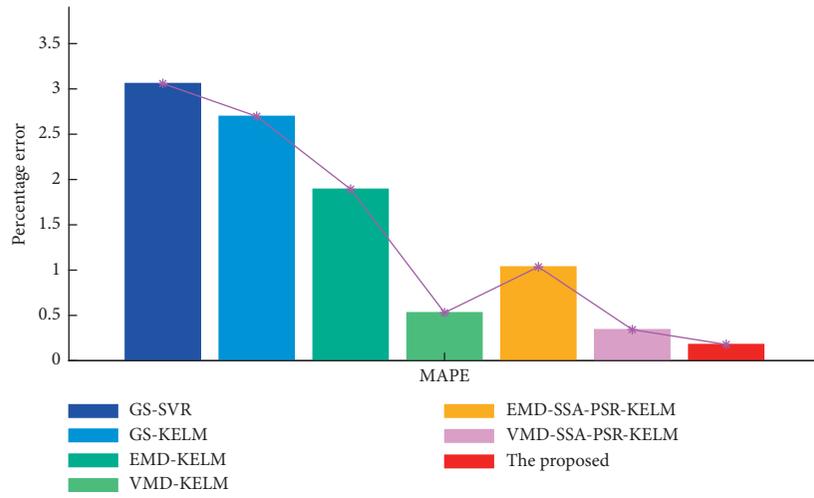
- (1) Compared with SVR, the metrics RMSE, MAE, and MAPE obtained by KELM are generally lower, from which the verdict that a better forecasting performance could be obtained by KELM can be proved preliminarily. In addition, the reducing ratio in terms of RMSE, MAE, and MAPE for KELM is 9.51%, 11.14%, and 11.83%. It can be observed that the performance gap between these two models are not significant, which can be attributed to the strong nonstationary traits within raw vibration signal that severely restrict the capabilities of the models. Hence, implementing the time-frequency signal preprocessing approaches would be the key point to achieve prediction performance promotions.
- (2) By comparing the results obtained by KLEM, EMD-KELM, and VMD-KELM, it can be demonstrated that the prediction accuracy can be significantly enhanced by the time-frequency signal processing approaches applied. Among the results, compared with single model KELM, the metrics RMSE, MAE, and MAPE of EMD-KELM are $2.6751 \mu\text{m}$, $2.1658 \mu\text{m}$, and 1.8931%, which have been decreased by 31.17%, 29.71%, and 29.82%. Meanwhile, the descent ratio achieved by VMD-KELM in terms of RMSE, MAE, and MAPE are 81.78%, 80.38%, and 80.35%, which are much larger than the decreasing ratio obtained by EMD-KELM. Hence, forecasting error can be significantly reduced by signal decomposition method-based preprocessing strategy, of which the VMD-based model possesses better prediction capability. Hence, the availability and the superiority of VMD can be proved convincingly.
- (3) Based on the decomposition methods implemented, the proposed dominant ingredient chaotic analysis fused with SSA and PSR can ulteriorly improve the prediction accuracy. Following the comparison between EMD-KELM and EMD-SSA-PSR-KELM, it can be observed that the evolution metrics obtained by EMD-SSA-PSR-KELM are $1.5310 \mu\text{m}$, $1.1842 \mu\text{m}$, and 1.0358%, which achieves the decreasing percentage of 42.77%, 45.32%, and 45.29% by comparing with EMD-KELM, respectively. Meanwhile, compared with VMD-KELM, VMD-SSA-PSR-KELM possesses much lower forecasting indexes of $0.4936 \mu\text{m}$, $0.3893 \mu\text{m}$, and 0.3431%, where the average decrease of these three metrics is 33.72%.
- (4) Further comparing VMD-SSA-PSR-KELM with the proposed approach, it can be seen that both models contain the same structures, while the parameters of PSR and KELM for each forecasting component in the proposed one are optimized by the proposed AMGWO algorithm. The metrics RMSE, MAE, and MAPE obtained by the proposed model are $0.2454 \mu\text{m}$, $0.2034 \mu\text{m}$, and 0.1791%, respectively,



(a)



(b)



(c)

FIGURE 11: Comparison of all the experimental models in terms of (a) RMSE, (b) MAE, and (c) MAPE.

which are much lower than the metrics obtained by VMD-SSA-PSR-KELM. Besides, the reduction ratio of all the metrics is 50.28%, 47.75%, and 47.80%,

respectively. Hence, the necessity and the superiority of the proposed AMGWO can be demonstrated positively. Furthermore, it can be observed from the

computational cost of each model that the computation complexity of each single model is almost the same, while the cost of KELM is slightly smaller. Meanwhile, the time consumption of the combined models applying decomposition techniques and GS increases correspondingly as the number of the decomposed subseries increases. It is worth noting that the proposed model possesses greatest time consumption, while the improvement implemented by the proposed model is significant compared with the remaining models. Meanwhile, the averaged metrics' decreasing ratios obtained by the proposed model are 79.74%, 79.18%, and 79.35% in contrast to all the contrastive ones.

Additionally, the comparisons of the predicted and monitored values, as well as prediction errors of each experimental model, are depicted in Figure 9 one by one, thus achieving intuitive observation of the prediction results. The predicted curve of the proposed approach can much better approximate to the actual values as well as possessing error curves that are distributed around zero with smaller fluctuations. Similarly, it can be observed from the comparisons between Figures 9(d) and 9(f) and Figures 9(c) and 9(e) that the error curves of the models based on dominant ingredient chaotic analysis are much more approximate to zero as well as achieving lower undulation, with which the availability and the superiority of the proposed dominant ingredient chaotic analysis fused with SSA and PSR can be demonstrated ulteriorly.

Furthermore, the scatter plots illustrating the fitting degree of all the experimental models are demonstrated in Figure 10 in the order of SVR (GS), KELM (GS), EMD-KELM (GS), VMD-KELM (GS), EMD-SSA-PSR-KELM (GS), VMD-SSA-PSR-KELM (GS), and the proposed model. It can be observed that distribution of the proposed approach on the regression line is the most uniform, and the corresponding R values are 0.99862, which is the largest among all models. Meanwhile, the combined models have generally achieved significant improvement compared with single models, of which the proposed structure, namely, VMD-SSA-PSR-KELM, is superior to the remaining combined models and achieves suboptimum performance among all the models.

Furthermore, the histograms of the evolution metrics calculated by all the experimental models are denoted in Figure 11 with which the conclusions summarized above can be intuitively observed. Among the histograms, the metric MAPE of various models are demonstrated in Figure 11(c) with the forms of histogram and dotted lines, with which the MAPE fluctuation of each models can be intuitively observed. The proposed approach possesses the minimum values of all the assessment metrics, while the model that contains the same frameworks as the proposed one achieves suboptimum performance. Furthermore, VMD-based models generally own smaller indicator values, such as VMD-KELM, VMD-SSA-PSR-KELM, and the proposed model.

5. Conclusions

To achieve accuracy forecasting for the vibration tendency, a novel hybrid approach combined with VMD, SSA, PSR, KELM, and AMGWO-based parameters optimization strategy is proposed in this paper. Concretely, VMD that can decompose the vibration signal into several subseries was employed to preliminarily weaken the nonstationary and nonlinear raw signal, while the optimal parameters of VMD are determined by minimizing LSEI by GS. Afterwards, SSA was implemented to separate the dominant and residuary ingredients from each subseries, after which all the residuary ingredients are accumulated with residual of VMD for further forecasting. To maximize the capability of the proposed prediction structure, i.e., VMD-SSA-PSR-KELM, adaptive updating strategies and mutation operator are introduced to the normal GWO for enhancing the corresponding parameters optimization performance. Therefore, the parameters of PSR and KELM for each forecasting component can be optimized effectively. Ultimately, the forecasting values corresponding to the vibration tendency are deduced by accumulating the values of all the predicted components. In the experimental phase, two single models and four combined models were adopted to compare with the proposed one. The corresponding intensive analysis at various levels demonstrated that (1) the forecasting models based on VMD can achieve much better performance than the models based on EMD and the models without time-frequency decomposition in this study; (2) the forecasting accuracy can be significantly enhanced by the proposed dominant ingredient chaotic analysis; and (3) the capability of the proposed hybrid forecasting framework can be maximized by implementing the proposed AMGWO algorithm to determine the appropriate parameters for each component. Compared with the relevant comparison models, the average decreasing ratio achieved by the proposed approach in terms of RMSE, MAE, and MAPE is 79.73%, 79.18%, and 79.13%, respectively. Therefore, the proposed hybrid approach can be considered as a credible tool for vibration tendency forecasting.

Abbreviations

ADMM:	Alternating direction method of the multipliers
AMGWO:	Adaptive mutation grey wolf optimizer
AR:	Autoregressive
ARIMA:	Auto regressive integrated moving average
ARMA:	Auto regressive moving average
ASA:	Artificial sheep algorithm
ELM:	Extreme learning machine
EMD:	Empirical mode decomposition
GA:	Genetic algorithm
GARCH:	Generalized autoregressive conditional heteroscedasticity
GS:	Grid search
GWO:	Grey wolf optimizer
HGU:	Hydropower generator unit
KELM:	Kernel extreme learning machine
Kurt.:	Kurtosis

LSEI:	Least-squares error index
LSSVM:	Least square support vector machine
MAE:	Mean absolute error
MAPE:	Mean absolute percentage error
Max.:	Maximum
Min.:	Minimum
NN:	Neural network
PSO:	Particle swarm optimization
PSR:	Phase space reconstruction
RMSE:	Root-mean-square error
RSEA:	Region search evolutionary algorithm
SCA:	Sine cosine algorithm
Skew:	Skewness
SSA:	Singular spectrum analysis
Std.:	Standard deviation
SVD:	Singular value decomposition
SVR:	Support vector regression
VMD:	Variational mode decomposition
WT:	Wavelet transform.

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (NSFC) (51741907), the Open Fund of Hubei Provincial Key Laboratory for the Operation and Control of a Cascaded Hydropower Station (2017KJX06), and the Hubei Provincial Major Project for Technical Innovation (2017AAA132). This work was also partly funded by the Research Fund for Excellent Dissertation of China Three Gorges University (2019SSPY053).

References

- [1] Z.-K. Feng, W.-J. Niu, S. Wang et al., "Developing a successive linear programming model for head-sensitive hydropower system operation considering power shortage aspect," *Energy*, vol. 155, pp. 252–261, 2018.
- [2] Z.-K. Feng, W.-J. Niu, and C.-T. Cheng, "China's large-scale hydropower system: operation characteristics, modeling challenge and dimensionality reduction possibilities," *Renewable Energy*, vol. 136, pp. 805–818, 2019.
- [3] J. Cheng, C. Zhu, W. Fu, C. Wang, and J. Sun, "An imitation medical diagnosis method of hydro-turbine generating unit based on Bayesian network," *Transactions of the Institute of Measurement and Control*, vol. 41, no. 12, pp. 3406–3420, 2019.
- [4] Z. Li, Y. Tao, A. Abu-Siada et al., "A new vibration testing platform for electronic current transformers," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 3, pp. 704–712, 2018.
- [5] X. An and J. Yang, "Denoising of hydropower unit vibration signal based on variational mode decomposition and approximate entropy," *Transactions of the Institute of Measurement and Control*, vol. 38, no. 3, pp. 282–292, 2016.
- [6] W. Fu, K. Wang, C. Li, X. Li, Y. Li, and H. Zhong, "Vibration trend measurement for a hydropower generator based on optimal variational mode decomposition and an LSSVM improved with chaotic sine cosine algorithm optimization," *Measurement Science and Technology*, vol. 30, no. 1, Article ID 015012, 2019.
- [7] X. Yuan, Z. Chen, Y. Yuan, and Y. Huang, "Design of fuzzy sliding mode controller for hydraulic turbine regulating system via input state feedback linearization method," *Energy*, vol. 93, pp. 173–187, 2015.
- [8] K.-B. Zhou, J.-Y. Zhang, Y. Shan, M.-F. Ge, Z.-Y. Ge, and G.-N. Cao, "A hybrid multi-objective optimization model for vibration tendency prediction of hydropower generators," *Sensors*, vol. 19, no. 9, p. 2055, 2019.
- [9] A. Masood, J. Bahrawi, and A. Elfeki, "Modeling annual rainfall time series in Saudi Arabia using first-order autoregressive AR(1) model," *Arabian Journal of Geosciences*, vol. 12, no. 6, p. 191, 2019.
- [10] M. Baptista, S. Sankararaman, I. P. de Medeiros, C. Nascimento, H. Prendinger, and E. M. P. Henriques, "Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling," *Computers & Industrial Engineering*, vol. 115, pp. 41–53, 2018.
- [11] D. S. D. O. Santos Júnior, J. F. L. de Oliveira, and P. S. G. de Mattos Neto, "An intelligent hybridization of ARIMA with machine learning models for time series forecasting," *Knowledge-Based Systems*, vol. 175, pp. 72–86, 2019.
- [12] X. Yuan, Q. Tan, X. Lei, Y. Yuan, and X. Wu, "Wind power prediction using hybrid autoregressive fractionally integrated moving average and least square support vector machine," *Energy*, vol. 129, pp. 122–137, 2017.
- [13] H. T. Pham and B.-S. Yang, "Estimation and forecasting of machine health condition using ARMA/GARCH model," *Mechanical Systems and Signal Processing*, vol. 24, no. 2, pp. 546–558, 2010.
- [14] Z. Zhang, H. Qin, Y. Liu et al., "Long short-term memory network based on neighborhood gates for processing complex causality in wind speed prediction," *Energy Conversion and Management*, vol. 192, pp. 37–51, 2019.
- [15] Z. Zhao, J. Yang, W. Yang, J. Hu, and M. Chen, "A coordinated optimization framework for flexible operation of pumped storage hydropower system: nonlinear modeling, strategy optimization and decision making," *Energy Conversion and Management*, vol. 194, pp. 75–93, 2019.
- [16] A. Rai and S. H. Upadhyay, "An integrated approach to bearing prognostics based on EEMD-multi feature extraction, Gaussian mixture models and Jensen-Rényi divergence," *Applied Soft Computing*, vol. 71, pp. 36–50, 2018.
- [17] 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," *Applied Energy*, vol. 237, pp. 103–116, 2019.
- [18] C. Zhang, J. Zhou, C. Li, W. Fu, and T. Peng, "A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting," *Energy Conversion and Management*, vol. 143, pp. 360–376, 2017.
- [19] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513–529, 2012.

- [20] C. Li, Z. Xiao, X. Xia, W. Zou, and C. Zhang, "A hybrid model based on synchronous optimisation for multi-step short-term wind speed forecasting," *Applied Energy*, vol. 215, no. 2017, pp. 131–144, 2018.
- [21] N. Li, J. Yang, R. Zhou, and Q. Wang, "Knock detection in spark ignition engines using a nonlinear wavelet transform of the engine cylinder head vibration signal," *Measurement Science and Technology*, vol. 25, no. 11, Article ID 115002, 2014.
- [22] S.-W. Fei, "Kurtosis forecasting of bearing vibration signal based on the hybrid model of empirical mode decomposition and RVM with artificial bee colony algorithm," *Expert Systems with Applications*, vol. 42, no. 11, pp. 5011–5018, 2015.
- [23] W. Fu, J. Tan, X. Zhang, T. Chen, and K. Wang, "Blind parameter identification of MAR model and mutation hybrid GWO-SCA optimized SVM for fault diagnosis of rotating machinery," *Complexity*, vol. 2019, Article ID 3264969, 17 pages, 2019.
- [24] W. Fu, K. Wang, J. Zhou, Y. Xu, J. Tan, and T. Chen, "A hybrid approach for multi-step wind speed forecasting based on multi-scale dominant ingredient chaotic analysis, KELM and synchronous optimization strategy," *Sustainability*, vol. 11, no. 6, p. 1804, 2019.
- [25] Y. Gao, C. Qu, and K. Zhang, "A hybrid method based on singular spectrum analysis, firefly algorithm, and BP neural network for short-term wind speed forecasting," *Energies*, vol. 9, no. 10, 2016.
- [26] W. Fu, K. Wang, J. Tan, and K. Zhang, "A composite framework coupling multiple feature selection, compound prediction models and novel hybrid swarm optimizer-based synchronization optimization strategy for multi-step ahead short-term wind speed forecasting," *Energy Conversion and Management*, vol. 205, Article ID 112461, 2020.
- [27] D. Wang, H. Luo, O. Grunder, and Y. Lin, "Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction," *Renewable Energy*, vol. 113, pp. 1345–1358, 2017.
- [28] W. Fu, J. Zhou, Y. Zhang, W. Zhu, X. Xue, and Y. Xu, "A state tendency measurement for a hydro-turbine generating unit based on aggregated EEMD and SVR," *Measurement Science and Technology*, vol. 26, no. 12, Article ID 125008, 2015.
- [29] C. Zhang, T. Peng, C. Li, W. Fu, X. Xia, and X. Xue, "Multi-objective optimization of a fractional-order PID controller for pumped turbine governing system using an improved NSGA-III algorithm under multiworking conditions," *Complexity*, vol. 2019, Article ID 5826873, 18 pages, 2019.
- [30] Y. Liu, H. Qin, Z. Zhang et al., "A region search evolutionary algorithm for many-objective optimization," *Information Sciences*, vol. 488, pp. 19–40, 2019.
- [31] Z. Wang, C. Li, X. Lai, N. Zhang, Y. Xu, and J. Hou, "An integrated start-up method for pumped storage units based on a novel artificial sheep algorithm," *Energies*, vol. 11, no. 1, p. 151, 2018.
- [32] X. Lai, C. Li, J. Zhou, and N. Zhang, "Multi-objective optimization of the closure law of guide vanes for pumped storage units," *Renewable Energy*, vol. 139, pp. 302–312, 2019.
- [33] C. Li, W. Wang, and D. Chen, "Multi-objective complementary scheduling of hydro-thermal-RE power system via a multi-objective hybrid grey wolf optimizer," *Energy*, vol. 171, pp. 241–255, 2019.
- [34] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
- [35] R. Wang, C. Li, W. Fu, and G. Tang, "Deep learning method based on gated recurrent unit and variational mode decomposition for short-term wind power interval prediction," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2019.
- [36] W. Fu, K. Wang, C. Zhang, and J. Tan, "A hybrid approach for measuring the vibrational trend of hydroelectric unit with enhanced multi-scale chaotic series analysis and optimized least squares support vector machine," *Transactions of the Institute of Measurement and Control*, vol. 41, no. 15, pp. 4436–4449, 2019.
- [37] 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," *Applied Energy*, vol. 230, pp. 1108–1125, 2018.
- [38] R. H. Chan, M. Tao, and X. Yuan, "Constrained total variation deblurring models and fast algorithms based on alternating direction method of multipliers," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 680–697, 2013.
- [39] N. Chen, Z. Qian, and X. Meng, "Multistep wind speed forecasting based on wavelet and Gaussian processes," *Mathematical Problems in Engineering*, vol. 2013, Article ID 461983, 8 pages, 2013.
- [40] R. Golafshan and K. Yuce Sanliturk, "SVD and Hankel matrix based de-noising approach for ball bearing fault detection and its assessment using artificial faults," *Mechanical Systems and Signal Processing*, vol. 70–71, pp. 36–50, 2016.
- [41] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Physical Review Letters*, vol. 45, no. 9, pp. 712–716, 1980.
- [42] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, pp. 985–990, IEEE, Budapest, Hungary, July 2014.
- [43] L. Duan, S. Dong, S. Cui, and W. Ma, "Extreme learning machine with Gaussian kernel based relevance feedback scheme for image retrieval," in *Proceedings of ELM-2015*, vol. 1, pp. 397–408, Springer, Cham, Switzerland, 2016.
- [44] Q. Li, H. Chen, H. Huang et al., "An enhanced grey wolf optimization based feature selection wrapped kernel extreme learning machine for medical diagnosis," *Computational and Mathematical Methods in Medicine*, vol. 2017, Article ID 9512741, 15 pages, 2017.
- [45] W. Fu, K. Shao, J. Tan, and K. Wang, "Fault diagnosis for rolling bearings based on composite multiscale fine-sorted dispersion entropy and SVM with hybrid mutation SCA-HHO algorithm optimization," *IEEE Access*, vol. 8, no. 1, pp. 13086–13104, 2020.
- [46] 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," *Applied Sciences*, vol. 9, no. 8, p. 1676, 2019.
- [47] T. Niu, J. Wang, K. Zhang, and P. Du, "Multi-step-ahead wind speed forecasting based on optimal feature selection and a modified bat algorithm with the cognition strategy," *Renewable Energy*, vol. 118, pp. 213–229, 2018.
- [48] W. Sun and Y. Wang, "Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network," *Energy Conversion and Management*, vol. 157, pp. 1–12, 2018.

- [49] H. Faris, I. Aljarah, M. A. Al-Betar, and S. Mirjalili, "Grey wolf optimizer: a review of recent variants and applications," *Neural Computing and Applications*, vol. 30, no. 2, pp. 413–435, 2018.
- [50] N. Mittal, U. Singh, and B. S. Sohi, "Modified grey wolf optimizer for global engineering optimization," *Applied Computational Intelligence and Soft Computing*, vol. 2016, Article ID 7950348, 16 pages, 2016.
- [51] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.
- [52] Y. V. Pehlivanoglu, "A new particle swarm optimization method enhanced with a periodic mutation strategy and neural networks," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 3, pp. 436–452, 2013.
- [53] Y. Wu, S. Li, S. Liu, H. Dou, and Z. Qian, "Vibration of hydraulic machinery," in *Mechanisms and Machine Science*, vol. 11, Dordrecht, Netherlands, Springer, 2013.
- [54] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.
- [55] S. Zhu, X. Yuan, Z. Xu, X. Luo, and H. Zhang, "Gaussian mixture model coupled recurrent neural networks for wind speed interval forecast," *Energy Conversion and Management*, vol. 198, Article ID 111772, 2019.