A Hybrid Forecasting Model for Short-Term Power Load Based on Sample Entropy, Two-Phase Decomposition and Whale Algorithm Optimized Support Vector Regression

WENWU LI\textsuperscript{1,2}, QIANG SHI\textsuperscript{1,2}, MUHAMMAD SIBTAIN\textsuperscript{1,2}, DAN LI\textsuperscript{1,2}, AND DANIEL ELIOTE MBANZE\textsuperscript{3}

\textsuperscript{1}College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002, China
\textsuperscript{2}Hubei Key Laboratory of Cascaded Hydropower Stations Operation and Control, China Three Gorges University, Yichang 443002, China
\textsuperscript{3}Electrical Engineering Department, Higher Education Polytechnic Institute, Tete 2304, Mozambique

Corresponding author: Qiang Shi (sparkqiang1995@ctgu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51807109, in part by the Open Fund of Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station under Grant 2019KJX08, and in part by the Research Fund for Excellent Dissertation of China Three Gorges University under Grant 2020SSPY055.

\section*{ABSTRACT}
To improve the accuracy and reliability of short-term power load forecasting and reduce the difficulty caused by load volatility and non-linearity, a hybrid forecasting model (CEEMDAN-SE-VMD-PSR-WOA-SVR) is proposed. Firstly, the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) is employed to generate multiple intrinsic modal functions (IMF) by decomposing the historical power load series. Then the sample entropy (SE) of each IMF is calculated to quantitatively evaluate the corresponding complexity. Afterward, variational mode decomposition (VMD) is adopted to achieve secondary decomposition for the component with the maximum sample entropy. Subsequently, the phase space reconstruction (PSR) is applied to reconstruct each IMF into a high-dimensional feature space matrix, which is formed as the input of support vector regression (SVR). Finally, SVR optimized by whale optimization algorithm (WOA) is used for the prediction, where the predicted values of all IMFs are accumulated to obtain the final prediction results. The experimental result demonstrates that the proposed hybrid model can effectively decompose the load series with non-linear characteristic and provide more accurate forecasting results by comparing the other models.

\section*{INDEX TERMS}
Short-term power load forecasting, two-phase decomposition, sample entropy, whale optimization algorithm, support vector regression.

\section*{I. INTRODUCTION}
Power load forecasting plays an important role in power system planning and economic operation. However, the environment and human activities greatly affect the power load. The fluctuation of the short-term power load sequence occurs in real-time, and changes periodically in units of days and weeks. Therefore, the time series of short-term power load has certain regularity as well as large volatility and non-linearity, which brings great difficulties during the accurate prediction. At present, the commonly used power load prediction methods are mainly divided into the following two categories [1]: mathematical statistics model, and artificial intelligence algorithm model.

The mathematical statistics model is relatively simple to be built based on a strong theoretical basis. For instance, Zhao \textit{et al.} [2] proposed a data-driven nonparametric auto-regressive (AR) model and obtained higher short-term load forecasting accuracy by comparing AR models. To obtain a better prediction effect than the AR and ARMA model [3],
Lee et al. [4] proposed an improved autoregressive integrated moving average (ARIMA) model, which improved short-term load prediction performance. Considering that the seasonal information integrated into the load can improve the prediction performance of the above statistical model, Bozkurt et al. [5] proposed a hybrid model based on SARIMA, and the experimental results verify the feasibility of the method. However, the statistical model also has the problem of insufficient ability for fitting the non-linear load sequence.

Over the past decade, artificial intelligence (AI) attracts a lot of attention. Intelligent algorithms with high solving efficiency, such as back propagation neural network (BPNN) and support vector regression (SVR), have been widely applied in load forecasting and achieved fine results [6]–[9]. Neural network models like BP have the good non-linear fitting ability, but they have more parameters, longer training time, and lack of training stability. By contrast, the SVR model possesses fewer parameters and faster running speeds on small samples, but its internal parameters have an impact on the prediction effect. Consequently, the SVR was optimized by a radial basis function network and chaotic artificial bee colony algorithm, which improved the prediction performance of the SVR model [10], [11].

Generally, there are still non-linear load sequences that cannot be weakened in the above models, and thus the more accurate predicted values could not be obtained. In contrast, the hybrid forecasting model usually has the advantages of effectively reducing the non-stationary time series of power load and high solving efficiency. For example, a power load prediction model combining empirical mode decomposition (EMD) with AR achieves a good prediction effect [12]–[14]. Furthermore, some scholars proposed a new short-term load forecasting model based on ensemble empirical mode decomposition and support vector machine (EEMD-SVM) model [15], [16], which effectively improved the prediction performance of the traditional model. Liu et al. [17] proposed a short-term load forecasting model that combines EEMD and SVR optimized by whale algorithm (EEMD-WOA-SVR), which avoids the PSO algorithm easily falling into local optimum and greatly enhances the solving efficiency of the SVR model. Complete EEMD with adaptive noise (CEEMDAN) is an improved method based on EMD and EEMD, which can achieve a better decomposition effect. An optimized dragonfly algorithm was used to optimize the CEEMDAN-SVR model [18]–[20], which effectively promoted the prediction accuracy of the power load. Variational modal decomposition (VMD) is a mature signal decomposition technology, which is also widely used in various fields [21]. Kim et al. [22] proposed the VMD-LSTM model to study the load series and obtained satisfactory forecasting results. Moreover, this kind of hybrid model is also feasible in such research fields as wind speed forecasting, air quality forecast, and runoff prediction [23]–[25]. For instance, Sun and Wang [26] adopted phase space reconstruction (PSR) to optimize the decomposed modal component sequence and established an improved BP model to predict short-term wind speed.

While seeing the advantages of decomposition techniques, we should also realize that a single decomposition may be hard to achieve the potential problem of sufficiently weakening the non-linearity of the sequence.

To obtain a smoother sequence and better prediction effect than a single decomposition, the two-phase decomposition framework CEEMDAN-VMD [24] considering sample entropy (SE) was adopted, and the ELM optimized by the differential evolutionary algorithm was employed to forecast the air quality, and the results showed the efficiency of this method. In addition, the two-phase decomposition (CEEMDAN-VMD) combined with SVM has equal applicability for runoff prediction [25]. In summary, the previous references demonstrate that CEEMDAN and VMD can avoid the drawbacks of other decomposition techniques, and they are successfully applied to the decomposition of the non-linear load time series. SVR is a classical and practical technique in machine learning, which has excellent performance. If the key parameters and kernel functions of the SVR can be optimized sufficiently, the solving efficiency of the SVR model can be enhanced. As a novel population intelligent algorithm, the whale optimization algorithm (WOA) has a stronger global search capability than the traditional intelligent algorithm by introducing a unique spiral update operation, so it is widely used to optimize various prediction models [17], [27], [28].

Based on the above introduction, the single decomposition technique has a limited ability to decompose the non-linearity sequences and the two-phase decomposition technique can further weaken the non-stationary of the time series. At the same time, this combined decomposition technique is rarely used in the field of power load forecasting and current researches lack further analysis of prediction errors. Therefore, this study firstly utilizes the two-phase decomposition framework (CEEMDAN-SE-VMD) to obtain a more stable sequence. Next, PSR is adopted to optimize all sequence components. Then, the SVR model optimized by the WOA algorithm is used for short-term load forecasting. Finally, the predicted values of all components are accumulated to obtain a predicted result. The proposed approach performs load forecasting by taking the real-time power load series of the Jiangyin city, Jiangsu Province, China, for August 2018. To the end, the prediction performance of the proposed model is compared and analyzed with various single prediction models (BP, SVR, and LSTM), hybrid models (EMD-SVR, EEMD-SVR, VMD-BP, and CEEMDAN-SVR), and two-phase decomposition model (CEEMDAN-SE-VMD-SVR).

The main contributions of this paper are presented as follows:

1. An innovative data preprocessing method: Based on the complexity discrimination of sample entropy, the proposed model proposes a two-phase (CEEMDAN-SE-VMD) decomposition technique to preprocess the short-term power load sequence, which fully weakens the non-linearity of
the load sequence, obtains more stable and less complex sequence components, and effectively reduces the difficulty of prediction. Subsequently, PSR is applied for the optimization and reorganization of sequence components, thereby facilitating the training and learning of data by the prediction model.

2. Optimized prediction model: The parameters of the traditional SVR model cannot be selected well, which limits the prediction accuracy of the model. Therefore, the WOA algorithm with strong global optimization ability is proposed to optimize the internal parameters of SVR, which greatly improves the prediction performance of the SVR model.

3. A comprehensive evaluation of the predictive performance of the proposed model: By comparing multiple groups of forecasting models, combined with a variety of mathematical statistics evaluation index (MAE, RMSE, MAPE, and $R^2$), an in-depth analysis of the prediction error of each model, the effectiveness of the proposed model in forecasting short-term power load series is comprehensively considered and verified.

The remainder of this paper is organized as follows: Section 2 presents the base theoretical of CEEMDAN, SE, VMD, PSR, WOA, and SVR. Section 3 introduces the main procedures of the proposed hybrid model in detail. Section 4 provides the experimental data and the specific operation of the model. Section 5 describes a comparison with other prediction models and analyzes prediction errors. The conclusion work is summarized in Section 6.

II. METHODOLOGY

A. COMPLETE ENSEMBLE EMPIRICAL MODE DECOMPOSITION WITH ADAPTIVE NOISE

Empirical mode decomposition (EMD) [13], [29] can be used to analyze non-linear signal sequences. It does not use any defined function as the basis, but adaptively generates the inherent modal function based on the analyzed signal. The purpose of EMD is to decompose a signal $f(t)$ into multiple intrinsic modal functions (IMFs) and a residual (RES). On this basis, EEMD adds a limited amount of white noise to the original signal to divide the frequency range in time-frequency space based on EMD, which makes EEMD effectively reduce the chance of modal aliasing [15], [30].

With the increase of white noise frequency in the EEMD, the efficiency of the algorithm decreases, and the prediction error after reconstruction becomes larger. To overcome this drawback, the CEEMDAN improved the EEMD by adding a finite amount of adaptive white noise [18], [24]. In the CEEMDAN algorithm, the decomposed modal component is represented by $IMF$. The operator $E_{i}(t)$ denotes the $i$-th modal component after the given signal is calculated by the EMD method. $\omega_i$ is the $i$-th added Gaussian white noise satisfying $N(0, 1)$ distribution. If $f(t)$ is the original signal sequence, it is denoted by $f(t) + \varepsilon_i\omega(t)$ after adding the white noise. The signal decomposition process applying CEEMDAN can be briefly described based on references [18], [31], [32], as shown in Equations (1) - (7).

Step 1: Obtain the first modal component decomposed by EMD algorithm:

\[
IMF_1(t) = \frac{1}{N} \sum_{i=1}^{N} IMF_i^1(t) \tag{1}
\]

Step 2: Calculate the first residual margin $R_1(t)$:

\[
R_1(t) = f(t) - IMF_1(t) \tag{2}
\]

Step 3: EMD is used to decompose the signal $R_1(t) + \varepsilon_1E_1(\omega(t))$ repeatedly for $N$ times to obtain the second modal component:

\[
IMF_2(t) = \frac{1}{N} \sum_{i=1}^{N} E_1(R_1(t) + \varepsilon_1E_1(\omega(t))) \tag{3}
\]

Step 4: For $k = 2, 3, \ldots, K$, calculate the $k$-th residual margin:

\[
R_k(t) = R_{k-1}(t) - IMF_k(t) \tag{4}
\]

Step 5: Repeat Step 3 to calculate the $(k + 1)$-th modal component and obtain:

\[
IMF_{k+1}(t) = \frac{1}{N} \sum_{i=1}^{N} E_1(R_1(t) + \varepsilon_kE_k(\omega(t))) \tag{5}
\]

Step 6: Repeat Steps 4 and 5 until the residual margin is not suitable for decomposition. The final residual margin:

\[
R(t) = f(t) - \sum_{i=1}^{K} IMF_i(t) \tag{6}
\]

Step 7: Finally, Equation (7) describes the original signal decomposed by CEEMDAN:

\[
f(t) = R(t) + \sum_{i=1}^{K} IMF_i(t) \tag{7}
\]

B. SAMPLE ENTROPY

Sample entropy (SE) is a method to quantitatively describe the complexity of time series by measuring the probability of generating new patterns from the original sequence [24]. When the SE value is larger, the complexity of the original sample sequence is higher and the prediction difficulty is greater. Therefore, it is necessary to calculate and screen the power load sequence with the highest entropy value. The definition process of SE is given in Step 1 - Step 6:

Step 1: For the one-dimensional sequence sampled at the same time interval, the time series with dimension $N$ is collected: $x(1), x(2), x(3), \ldots, x(N)$;

Step 2: Reconstruct the above sequence to $m$-dimension, namely $X(1), X(2), X(3), \ldots, X(N - m + 1)$, where $X_m(i) = \{x(i), x(i+1), x(i+2), \ldots, x(i+m-1)\}$;

Step 3: When $1 \leq i \leq N - m + 1, i \neq j$, define the distance $D[X_m(i), X_m(j)]$ with the biggest absolute value difference in sequence elements, and its expression
is \( D [X_m(i), X_m(j)] = \max(|x(i+k) - x(j+k)|), \quad k = 0, 1, 2, \ldots, m - 1; \)

Step 4: Calculate the ratio \( r \geq D [X_m(i), X_m(j)] \) of the number of vectors meeting the threshold of \( N - m \) and \( B^m(r) \), and obtain the mean value of the vectors. It is displayed in Equation (8):

\[
B^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} B^m_i \quad (8)
\]

Step 5: When dimension \( m \) increases by 1, repeat Steps 2 - 4 to obtain \( B^{m+1}(r) \), then SE can be defined as Equation (9):

\[
SE(m, r) = \lim_{N \to \infty} - \ln \frac{B^{m+1}(r)}{B^m(r)} \quad (9)
\]

Step 6: Usually, \( N \) is a finite value, and Equation (9) can be estimated as Equation (10):

\[
SE(m, r, N) = \left[ - \ln \frac{B^{m+1}(r)}{B^m(r)} \right] \quad (10)
\]

where \( m \) represents number of dimensions; \( r \) represents conditional threshold.

### C. VARIATIONAL MODE DECOMPOSITION

Variational mode decomposition (VMD) is a self-adaptive, completely non-recursive modal variation and signal processing technique [21], [23]. It is considered that the single decomposition (CEEMDAN) usually cannot weaken the non-linear and irregular short-term load series sufficiently. To fully decompose the load sequence components, this study intends to use VMD for two-phase decomposition [24], [25]. VMD overcomes the problem of end effect and modal component aliasing in the EMD method and has a more solid mathematical theoretical basis, which can reduce the non-linearity of the time series. The VMD decomposition obtains multiple different frequency scales with relatively stable subsequences.

Suppose the original signal \( f \) is decomposed into \( k \) components to ensure that the decomposition sequence is a modal component with limited bandwidth and central frequency. The specific variational constraint expression is as Equation (11):

\[
\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[ \left( \delta(t) + j/\pi t \right) * u_k(t) \right] e^{-j\omega t} \right\|_2^2 \right\}
\]

s.t. \( \sum_{k=1}^{K} u_k = f \) \quad (11)

where \( u_k \) and \( \omega_k \) respectively corresponding to the \( k \)-th modal component and the center frequency; \( K \) is a positive integer and represents the number of modals that need to be decomposed; \( \delta(t) \) is a Dirac function; \( * \) is a convolution operator. Furthermore, the constraint condition is that the sum of all the modal components obtained by decomposition is consistent with the original sequence.

To solve the Equation (11) with variational constraints, a Lagrange multiplication operator is introduced, and quadratic penalty factor \( \alpha \) is introduced to reduce the interference of Gaussian noise. The augmented Lagrange expression is expressed as Equation (12):

\[
L (\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[ \left( \delta(t) + j/\pi t \right) * u_k(t) \right] e^{-j\omega t} \right\|_2^2
\]

\[
+ \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\{ \lambda(t), f(t) - \sum_k u_k(t) \right\} \quad (12)
\]

The alternating direction multiplier (ADMM) iterative algorithm is used in combination with Parseval-Plancherel and Fourier isometric transform to optimize each modal component and center frequency [33]. The expressions of \( u_k, \omega_k \) and \( \lambda \) can be described as Equations (13) - (15):

\[
\check{u}_k^{n+1}(\omega) \leftarrow \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \lambda(\omega)/2}{1 + 2\alpha (\omega - \omega_k)^2} \quad (13)
\]

\[
\omega_k^{n+1} \leftarrow \int_0^\infty \omega \left\| \hat{u}_k^{n+1}(\omega) \right\|_2^2 d\omega \quad (14)
\]

\[
\lambda^{n+1}(\omega) \leftarrow \check{\lambda}^{n}(\omega) + \gamma \left( \hat{f}(\omega) - \sum_k \check{u}_k^{n+1}(\omega) \right) \quad (15)
\]

where \( \gamma \) is the noise tolerance, which satisfies the fidelity requirement of signal decomposition, and \( \check{u}_k^{n+1}(t), \hat{u}_i(\omega), \check{\lambda}^{n}(\omega) \) and \( \hat{\lambda}(\omega) \) correspond to the Fourier transform forms of \( u_k^{n+1}(t), u_i(t), f(t) \) and \( \lambda(t) \), respectively.

The decomposition steps of VMD can be briefly expressed as Step 1 – Step 4:

Step 1: Initialize \( \hat{u}_1, \omega_1 \) and \( \check{\lambda}_1 \), and set the maximum number of iterations \( N, n \leftarrow 0; \)

Step 2: Apply Equations (13) and (14) to update \( \hat{u}_k \) and \( \omega_k \);

Step 3: Update \( \check{\lambda} \) using Equation (15);

Step 4: The accuracy convergence criterion \( \varepsilon \) needs to be greater than zero. If \( \sum_k \left\| \check{u}_k^{n+1} - \check{u}_k^n \right\|_2^2 / \left\| \check{u}_k^n \right\|_2^2 < \varepsilon \) and \( n < N \) are not satisfied, the Step 2 is returned. Otherwise, the iteration is completed, and the final \( \hat{u}_k \) and \( \omega_k \) are output.

### D. PHASE SPACE RECONSTRUCTION

Phase space reconstruction (PSR) technique is proposed by Packard et al. [34] in 1980, which is an effective and common method to analyze chaotic time series. In this paper, PSR technology is applied to "process and optimize" the two-phase decomposition sequence to form the input of the SVR.

The coordinate delay reconstruction method is adopted to construct the phase space vector of the modal component sequence under delay time \( \tau \) and embedded dimension \( d \). Therefore, the PSR [21], [34] expression of \( x = \{x_i | i = 1, 2, \ldots, N \} \) at different prediction layers for the
modal component sequence $h$ decomposed by the two-phase decomposition can be expressed as Equation (16):

$$
X = \begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_L \\
\end{bmatrix} = \begin{bmatrix}
x_1 & x_1 + \tau & \cdots & x_1 + (d-1)\tau \\
\vdots & \vdots & \ddots & \vdots \\
x_l & x_l + \tau & \cdots & x_l + (d-1)\tau \\
\vdots & \vdots & \ddots & \vdots \\
x_L & x_L + \tau & \cdots & x_L + (d-1)\tau \\
\end{bmatrix}
$$  \hspace{1cm} (16)

where $X_i$ represents the $i$-th space vector in the phase space; $L = N - (d - 1) \cdot \tau - h$; $N$ is the total number of load samples. The corresponding output matrix $O$ could be shown as Equation (17):

$$
O = \begin{bmatrix} O_1 & O_2 & \cdots & O_L \end{bmatrix}^T = \begin{bmatrix} x_1 + h + (d-1)\tau & x_2 + h + (d-1)\tau & \cdots & x_N \end{bmatrix}^T
$$  \hspace{1cm} (17)

E. SUPPORT VECTOR REGRESSION

Support vector machine (SVM) is a kind of machine learning algorithm, which has been widely applied in regression prediction. Compared with the neural network model, SVR has a stable and efficient solving capability. SVR [12], [18], [25] mainly realizes the problem of the linear inseparability of sample data through penalty coefficient and kernel function technology. It is assumed that there are $n$ non-linear load data sets $(x_i, y_i)$, $x_i$ is the load sample input, $y_i$ is the output, $\phi(x_i)$ represents the feature vector after $x_i$ is mapped to the high-dimensional feature space. The corresponding optimal formula is as Equation (18):

$$
f(x_i) = \omega^T \phi(x_i) + b
$$  \hspace{1cm} (18)

where $\omega$ represents the normal vector and $b$ is the displacement term. The essence of the SVR model training process is to find the optimal $\omega$ and $b$ to make $f(x_i)$ approach $y_i$, and its convex optimization function is shown as Equation (19):

$$
\min_{\omega, b} \frac{1}{2} ||\omega||^2 + C \sum_{i=1}^{m} (\xi_i + \hat{\xi}_i)
$$  \hspace{1cm} (19)

The constraint conditions are illustrated by Equation (20):

$$
\begin{aligned}
&f(x_i) - y_i \leq \varepsilon + \xi_i \\
y_i - f(x_i) \leq \varepsilon + \hat{\xi}_i \\
\xi_i, \hat{\xi}_i \geq 0, i = 1, 2, \ldots, m \\
C > 0
\end{aligned}
$$  \hspace{1cm} (20)

where $C$ represents the penalty coefficient, which is used to balance structural risk $\frac{1}{2} ||\omega||^2$ and empirical risk $\sum_{i=1}^{m} (\xi_i + \hat{\xi}_i)$; $\xi_i$ and $\hat{\xi}_i$ are relaxation variables on both sides of the hyperplane, which determine the error size. Relaxation factor is the deviation value between $f(x_i)$ and $y_i$. To solve this convex optimization function, Lagrange multipliers $a_i$ and $\hat{a}_i$ are introduced into the original function. The parameters $\omega$ and the regression model function $f(x)$ in the SVR model can be obtained by Equations (21)-(22):

$$
\begin{aligned}
&\omega = \sum_{i=1}^{m} (\hat{a}_i - a_i) \phi(x_i) \hfill (21) \\
&f(x) = \sum_{i=1}^{m} \hat{a}_i \phi(x_i) K(x, x_i) + b \hfill (22)
\end{aligned}
$$

In the above Equation, $g$ is a kernel parameter; the kernel function $K(x, x_i)$ is used to replace the linear term in the linear function, and the traditional linear model can deal with the non-linear regression problem.

III. PROPOSED HYBRID MODELLING

A. MODEL INTRODUCTION AND OPTIMIZATION

To weaken the non-linearity of power load time series, reduce the difficulty of short-term power load prediction and improve the prediction accuracy, a hybrid prediction model based on CEEMDAN-SE-VMD-PSR-SVR was proposed to predict short-term power load. Firstly, the original load data is preprocessed by the composite modal decomposition framework (CEEMDAN-SE-VMD), which fully weakens the non-linearity of the load time series and reduces the difficulty of forecasting. And then the PSR technique assembles the series of load components to form the input of the SVR. Finally, this study adopts the optimized SVR model to make a prediction.

Based on fully decomposing the non-linear sequences, this paper proposes a whale optimization algorithm (WOA) optimized CEEMDAN-SE-VMD-PSR-SVR hybrid model to achieve SVR parameter optimization and obtain better prediction performance. WOA is a swarm intelligence optimization algorithm inspired by humpback whale hunting and preying behavior. This algorithm has the advantages of simple parameter adjustment, fast convergence speed, and strong global optimization ability than general intelligent algorithms [17], [27], [28]. Furthermore, the WOA flow is briefly described in Figure 1.

1) Wrap around [27]. The humpback whales constantly adjust their position according to the position of prey combined with spiral movement to obtain the best hunting strategy. This search strategy can be described as Equation (23):

$$
\begin{aligned}
&D = |CX^*(t) - X(t)| \\
&X(t + 1) = X^*(t) - AD
\end{aligned}
$$  \hspace{1cm} (23)

where $D$ is the distance between the current humpback whale and its prey; $X^*(t)$ is the current position vector to obtain the optimal solution; $X(t)$ is the current position vector; $X(t + 1)$ is the target vector after the next iteration; $A$ and $C$ are coefficient vectors, which can be calculated by Equation (24):

$$
\begin{aligned}
&A = 2ar_1 - a \\
&C = 2r_2, \hfill a = 2 - \frac{2t}{T}
\end{aligned}
$$  \hspace{1cm} (24)
where $A \in [-1, 1]$; $r_1$ and $r_2$ are random vectors in $(0, 1)$, the constant $a$ linearly decreases during the iteration process; $t$ represents the current number of iterations; $T$ is the maximum number of iterations.

(2) Hunting behavior. The humpback whale uses a spiral equation to update the position of the humpback whale after searching and surrounding its prey:

$$
X(t + 1) = \begin{cases} 
X^*(t) - AD & P < 0.5 \\
D^* e^{bl \cdot \cos(2\pi l)} + X^*(t) & P \geq 0.5
\end{cases}
$$

(25)

where $D^*$ is the distance between the current best position of the $i$-th whale and its prey; $b$ is the shape constant of the logarithmic spiral; $l$ is the random number between $[-1, 1]$; $P$ is a random number between $[0, 1]$. The spiral position is updated by Equation (26):

$$
X(t + 1) = \begin{cases} 
X^*(t) - AD & P < 0.5 \\
D^* e^{bl \cdot \cos(2\pi l)} + X^*(t) & P \geq 0.5
\end{cases}
$$

(26)

(3) Search for prey. In actual hunting, humpback whales combine spiral search and random search, that is, random search within the range of $|A| > 1$, which is convenient to find other suitable foods and enhance WOA’s global search and predation ability:

$$
X(t + 1) = X_{\text{rand}} - AD \\
D = \|C \times X_{\text{rand}} - X(t)\|
$$

(27)

where $X_{\text{rand}}$ is a randomly selected position in the whale group. In the subsequent iterations, the current optimal solution and the global optimal solution are selected [17], [27].

B. SPECIFIC PROCEDURES

Since considering the sample entropy of the load component sequences, the non-linearity original sequence is fully weakened by the two-phase decomposition, and the optimized SVR model is used for prediction. Therefore, the prediction accuracy of the traditional model could be improved. The specific flow of the proposed hybrid model is shown in Figure 2.

The main steps of the proposed hybrid model can be explained as given below:

Step1. CEEMDAN. Decompose the original power load sequence by employing CEEMDAN to get the modal components;

Step2. Sample entropy. Calculate the sample entropy value of each component obtained by CEEMDAN;

Step3. Two-phase decomposition. Adopt the composite framework (CEEMDAN-SE-VMD) fully weakens the non-linearity of the load time series;

Step4. Optimization and reorganization. PSR is applied to optimize all the modal components decomposed by two-phase decomposition, and then input the SVR model for training;

Step5. Parameters optimization. Use the WOA algorithm to search the optimal parameters $C$ and $g$ of SVR model for each component;

Step6. Training and prediction. The SVR model optimized by WOA is used to train and predict;

Step7. Getting results. Carry out the summation of the predicted values of Step (6) to get the final result of the power load prediction.

C. MODEL PERFORMANCE EVALUATION INDICATORS

In this paper, the prediction performance of the proposed model on the test set is assessed and analyzed by the statistical indexes [35], [36] such as the mean square error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and the determination coefficient ($R^2$). And the indexes are shown in Equations (28) - (31).

$$
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i| \\
\text{RMSE} = \left( \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - \bar{y}_i)^2 \right)^{\frac{1}{2}} \\
\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{\hat{y}_i - \bar{y}_i}{\hat{y}_i} \right| \times 100 \\
R^2 = \frac{\sum_{i=1}^{N} (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2}
$$

(28) - (31)

where $y_i$, $\bar{y}_i$, and $\hat{y}_i$ denotes the actual, average, and predicted values of power load sequences.

IV. CASE STUDY

A. DATA COLLECTION

The reliability of data and prediction model is the significant factors affecting the accuracy of short-term load forecasting. The experimental data in this paper are from the power load.
data measured from August 1 to August 10, 2018, in Jiangyin city, Jiangsu province, China. The sampling frequency is 15 minutes, and the total number of samples is 960. Reasonable training set and test set also have some influence on the prediction effect of the model. The data of the first 8 days (768 samples) is selected for training, while the remaining 192 samples (last two days) are considered for testing to verify the accuracy and reliability of the proposed prediction model. Experimental platform configuration of this study: Windows 8.1 professional, Intel i5 CPU, 4GB. Simulation software: Matlab R2017a and JetBrains PyCharm 2018.3.3. The historical data of the original load sequence is shown in Figure 3.

**B. ENGINEERING APPLICATION**

In the first place, the original load sequence is decomposed by CEEMDAN to obtain multiple modal components. Then, the SE value of each sequence component is calculated, and the modal component with the maximum SE value is decomposed by VMD. As a modal decomposition technology with high decomposition efficiency, VMD can fully reduce the series non-stationary property with high complexity and reduce the difficulty of short-term power load prediction. For this reason, several key parameters of CEEMDAN were...
| Component | IMF1  | IMF2  | IMF3  | IMF4  | IMF5  | IMF6  | IMF7  | IMF8  | IMF9  | RES   |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Sample Entropy | 1.5717 | 0.53  | 0.8248 | 0.5451 | 0.5028 | 0.2193 | 0.1062 | 0.0556 | 0.0194 | 0.0034 |

**FIGURE 4.** Decomposition of power load time series using CEEMDAN.

**FIGURE 5.** Decomposition of power load time series using VMD.

set as [18]: noise standard deviation, \( N_{std} = 0.05 \); number of white noise groups, \( NR = 500 \); maximum iterations, \( MaxIter = 5000 \). Figure 4 shows the 9 modal components \( IMF \) and 1 residual component \( RES \) at different frequency scales obtained by CEEMDAN.

To accurately calculate the modal component of the highest SE value obtained by CEEMDAN, the core parameter \( m \) in the SE algorithm is set as 2, and the conditional threshold \( r \) choose 0.2 times of the standard deviation (\( \sigma \)) of the modal sequences [26]. Table 1 illustrates the SE value of each modal component decomposed by CEEMDAN.

It can be seen from Table 1 that the SE value of IMF1 is the largest, indicating that IMF1 is the sequence with the highest complexity after CEEMDAN decomposition. Therefore, the two-phase decomposition of IMF1 is conducted by VMD. To ensure that IMF1 is effectively decomposed and avoid the influence of too many modal components on parameter optimization of the WOA algorithm, the parameter K of VMD is set to 4, and the remaining parameters are the default values [21], [33]. As illustrated in Figure 5, the modal components VMF1-VMF4 were obtained through VMD.

To explore the embedded features of the power load time series, the phase space reconstruction (PSR) technique, an effective and commonly used chaotic time series analysis tool, is used to process the decomposed modal component sequences of CEEMDAN and VMD [21], [34], [35]. As a result of PSR, the sample can better reflect the change of the actual power load and enter the SVR forecasting model. However, the selection of the delay time \( \tau \) cannot be too small and the embedded dimension of \( d \) should be appropriate. Otherwise, the correlation of the coordinates will be too strong, making it difficult to display the internal information of the power load. Hence, the two key parameters of PSR [36]: delay time \( \tau \) chooses 1, embedded dimension \( d \) chooses 5, specific reconstruction process as shown in Equations (16) - (17). After PSR optimization, the modal component data still need to be normalized before input into the SVR model to improve the training efficiency.

Considering that the WOA algorithm has the characteristics with simple structure, few parameters, strong search ability, and high convergence accuracy, this paper adopts WOA to optimize the SVR model. To improve the searching ability of the whale algorithm and obtain the global optimal solution, relevant parameters should be set reasonably. The size of the whale group, the maximum number of iterations, and the search range of SVR parameters are specified in Table 2. With the help of the WOA algorithm, the hyperparameters of SVR including penalty coefficient \( C \) and kernel parameter \( g \) are optimized and selected reasonably [17], [27], [36]. Finally, the predicted values of all modal components are summed up to complete the final short-term load forecasting. The optimization parameters of the involved models are shown in Table 2, and \( \sigma \) represents the standard deviation value of the sequence obtained by CEEMDAN. The grid search algorithm is used to select the internal parameters of SVR in the comparison model [37], and the search range is \([2^{-10}, 2^{10}]\). Other important parameters are set by trial and error and predetermined parameter values, as detailed in Table 2.

The proposed model is applied to the actual load prediction by adopting the process mentioned in Figure 2 and combining it with the preferred parameters in Table 2. The short-term power load forecasting effect diagram is shown in Figure 6.

When carrying out research on the prediction of non-linear and fluctuant power load time series, Figure 6 shows
that the proposed two-phase decomposition technique combined WOA-SVR (CEEMDAN-SE-VMD-PSR-WOA-SVR) model can well fit and predict the future sequences. It can obtain accurate results in short-term urban load forecasting and provide reasonable arrangements for power grid operation in summer.

V. DISCUSSION

To further illustrate the performance of the proposed hybrid model, it is compared with other predictive models (BP, SVR, LSTM, EMD-SVR, EEMD-SVR, VMD-BP, CEEMDAN-SVR, and CEEMDAN-SE-VMD-SVR). Finally, the evaluation indexes (MAE, RMSE, MAPE and $R^2$) are used to test and verify the superiority and accuracy of the proposed model. Table 3 shows the number of modal components obtained by decomposing the original load sequence for each model. Table 4 shows the prediction effect of each modal component. Table 5 shows the evaluation indexes of the specific prediction effect of each model. Figure 7 shows the prediction effect of the proposed model and other comparable

| Models         | Modules | Parameter              | Determination approach | Range/Determined value |
|----------------|---------|------------------------|------------------------|------------------------|
| BP             | BP      | Network layers         | Preset                 | 3                      |
|                |         | Batch_size             | Preset                 | 64                     |
|                |         | Epochs                 | Trial and error        | 100                    |
| SVR            | SVR     | Penalty coefficient C  | Grid search            | $2^{-10}, 2^{10}$     |
|                |         | Kernel parameter g     | Grid search            | $2^{-10}, 2^{10}$     |
| LSTM           | LSTM    | Network layers         | Preset                 | 3                      |
|                |         | Batch_size             | Preset                 | 64                     |
|                |         | Epochs                 | Trial and error        | 100                    |
| EMD-SVR        | EEMD    | Penalty coefficient C  | Grid search            | $2^{-10}, 2^{10}$     |
|                |         | Kernel parameter g     | Grid search            | $2^{-10}, 2^{10}$     |
| VMD-BP         | VMD     | Number of modes K      | Trial and error        | 5                      |
|                |         | Network layers         | Preset                 | 3                      |
|                |         | Batch_size             | Trial and error        | 32                     |
|                |         | Epochs                 | Trial and error        | 100                    |
| CEEMDAN-SVR    | CEEMDAN | Penalty coefficient C  | Grid search            | $2^{-10}, 2^{10}$     |
|                |         | Kernel parameter g     | Grid search            | $2^{-10}, 2^{10}$     |
| CEEMDAN-SE-VMD-SVR | CEEMDAN | Penalty coefficient C  | Grid search            | $2^{-10}, 2^{10}$     |
|                |         | Kernel parameter g     | Grid search            | $2^{-10}, 2^{10}$     |
|                |         | NR                     | Preset                 | 0.05                   |
|                |         | MaxIter                | Preset                 | 500                    |
|                |         | SE                     | Number of dimensions m | Preset                 | 2                      |
|                |         |                        | Conditional threshold r | Preset                 | $0.2\sigma$            |
|                |         | VMD                    | Number of modes K      | Trial and error        | 4                      |
|                |         | SVR                    | Penalty coefficient C  | Grid search            | $2^{-10}, 2^{10}$     |
|                |         |                        | Kernel parameter g     | Grid search            | $2^{-10}, 2^{10}$     |
|                |         | CEEMDAN                | Preset                 | 0.05                   |
|                |         | NR                     | Preset                 | 500                    |
|                |         | MaxIter                | Preset                 | 5000                   |
|                |         | SE                     | Number of dimensions m | Preset                 | 2                      |
|                |         |                        | Conditional threshold r | Preset                 | $0.2\sigma$            |
|                |         | VMD                    | Number of modes K      | Trial and error        | 4                      |
|                |         | PSR                    | Delay time r           | Trial and error        | 1                      |
|                |         |                        | Embedded dimension d   | Trial and error        | 5                      |
|                |         | WOA                    | Searching agents X     | Preset                 | 30                     |
|                |         |                        | Maximum iterations T   | Preset                 | 50                     |
|                |         | SVR                    | Penalty coefficient C  | WOA                    | $2^{-10}, 2^{10}$     |
|                |         |                        | Kernel parameter g     | WOA                    | $2^{-10}, 2^{10}$     |

| Model | EMD-SVR | EEMD-SVR | VMD-BP | CEEMDAN-SVR | CEEMDAN-SE-VMD-SVR | The proposed |
|-------|---------|----------|--------|-------------|---------------------|--------------|
| Number | 8       | 10       | 5      | 10          | 10                  | 9+4          |
models. Figure 8 compares the prediction error and the fitting effect of each model. It should be noted that the BP, SVR, and LSTM models have not undergone modal decomposition, and the number of
remaining model decompositions is shown in Table 3. Firstly, The proposed model adopts CEEMDAN to obtain 10 components. The VMD then decomposes IMF1 into 4 components. Finally, 13 modal components are obtained from the proposed model.

Table 4 gives the prediction error and fitting effect of the predicted value of each component compared to the true value. It shows that the proposed model can achieve better a prediction effect on each component. Accumulating the predicted values of all components can realize the final load forecast.

From Table 5 and Figure 7, the proposed model has obtained the optimal prediction effect on the test set, with the smallest error index (MAE, RMSE, and MAPE) and the highest fitting index ($R^2$), which is respectively 19.53 MW, 23.81 MW, 0.51% and 99.32% on the test set.

For example, compared with a single prediction model, the prediction error of the BP model is the largest, and the proposed hybrid model is the smallest. The prediction effect of SVR and LSTM is better than that of BP. However, the prediction effect of SVR and the LSTM model is not ideal. The neural network model has a good fitting performance, but in the case of non-linear load prediction, the prediction performance is not stable and needs a long time of training due to the requirement to adjust many parameters. The mode decomposition technology can decompose non-linearity power load sequences. Therefore, the hybrid model is much better than that of BP, SVR, and LSTM. Taking MAPE as an example, compared with the SVR model, the EMD-SVR model decreased by 0.04% in the test set. The VMD-BP model benefits from the high decomposition efficiency of VMD, and the MAPE index is 0.43% lower than that of the BP model. However, the VMD-BP model has been trained for a long time in the experiment. The predictive performance of the CEEMDAN-SVR model is better than that of EMD-SVR, EEMD-SVR, and VMD-BP models, and the fitting effect on the test set reaches 93.09%.
Since the above single decomposition technique could not deal with the non-linear data sequence well, and there is room for improvement. The proposed model takes the sample entropy of the power load sequence into consideration and combines two-phase decomposition techniques, CEEMDAN and VMD, to sufficiently weaken the non-linearity of the original sequence and further reduce the difficulty of prediction. PSR is used to optimize the sequence components and input of the SVR prediction model, and the WOA is used to optimize the SVR model, which improved the solving efficiency of the SVR. The above optimization operation enhances the prediction performance of the traditional CEEMDAN-SVR and CEEMDAN-SE-VMD-SVR model. For example, MAPE decreases by 0.67% and 0.5%, while $R^2$ increases by 6.23% and 4.37%, respectively. Therefore, we can conclude that the proposed model uses the two-phase decomposition technique and WOA optimized SVR hybrid model to get a smoother load sequence, which reduces the difficulty of prediction and is of great help to improve the prediction accuracy and stability of the traditional model.

To show the prediction effect of each model more clearly and intuitively, Figure 7 shows the prediction effect of each model, and Figure 8 compares the prediction error and fitting capability of each model.

It can be observed from Figure 7 and Figure 8 that the proposed model effectively predicted the real load sequence, with the minimum prediction error and the best fitting effect. To further study and analyze the distribution characteristics and prediction effect of the prediction errors of each model, the Gaussian mixture distribution test and statistical analysis of the prediction errors of each model can be carried out.

Based on the Gaussian Distribution [23, 38], the distribution of prediction errors of each model is tested and analyzed. It is assumed that the prediction error sequence $X$ is the probability distribution subject to the position parameter $\mu$ and scale parameter $\sigma$, where $\mu$ represents the expected mean.
of the error value, $\sigma$ is the standard deviation, and the normal distribution [39] satisfying the $N(\mu, \sigma^2)$ prediction error is as shown in Equation (32):

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X-\mu)^2}{2\sigma^2}} \quad (32)$$

As demonstrated in Figure 9, it shows the prediction error value of each model at each test point and calculates $\mu$ and $\sigma$ of the error value of each model by using Equation (32). Then describe the Gaussian frequency histogram and fitting distribution curve of the prediction error of each model.

As can be seen from the prediction error accumulation diagram in Figure 9, the error value of the proposed model is smaller than that of other models at specific prediction points, and the standard deviation is the smallest in the whole test set. The mathematical statistics analysis of the prediction error value describes the Gaussian distribution histogram and the fitting curve of the prediction error [23], [39], [40]. The result shows that the error of the proposed model is mainly concentrated in the near zero, and the probability of points with low prediction error values is higher. In contrast, for other models, the error value offset zero points on both sides, and the occurrence frequency of the points with larger prediction error value is higher. It shows that the error distribution of the proposed model is more reasonable and has higher observation accuracy. On the test set, the intelligent algorithm model (BP, SVR, and LSTM) model has a wide error distribution range and low prediction accuracy. The single decomposition model (EMD-SVR, EEMD-SVR, VMD-BP, and CEMDAN-SVR) and two-phase decomposition model (CEMDAN-SE-VMD-SVR) reduce the non-linearity of the power load sequence due to mode decomposition, thus ensuring the prediction effect. The proposed model obtained a more stable load sequence by utilizing two-phase decomposition and then adopted the WOA optimized SVR model for prediction. Hence, the prediction effect is better than all the comparison models, and the solution effect is efficient and stable. The actual prediction error distribution of the proposed model is more reasonable and concentrated, with less volatility and higher observation accuracy and prediction stability.

VI. CONCLUSION

Short-term load forecasting plays an important role in power system dispatching and market transaction. To reduce the non-linearity of load series and improve the forecasting accuracy, this paper proposes a hybrid model for short-term load forecasting, which combines two-phase decomposition and optimized support vector regression (CEMDAN-SE-VMD-PSR-WOA-SVR). The proposed model firstly uses CEEMDAN to decompose the original sequence and calculate the sample entropy of all components. Subsequently, the sequence with the largest entropy value is used for two-phase decomposition using VMD technology, which fully weakens the non-linearity of the sequence. Then PSR optimizes all the modal components and inputs them into the SVR model. Next, the SVR model optimized by the WOA algorithm is used to train the data and predict the future sequence. Finally, the prediction accuracy of the proposed model is compared with BP, SVR, LSTM, EMD-SVR, EEMD-SVR, VMD-BP, CEEMDAN-SVR, and CEMDAN-SE-VMD-SVR. And the distribution of the prediction error of each model is analyzed by mathematical statistics. According to the detailed prediction results and the corresponding comprehensive analysis, the conclusions can be drawn:

1. The proposed hybrid model achieves the highest prediction performance and the smallest prediction error among all models. It is verified that the proposed hybrid model can effectively improve the forecasting accuracy of short-term power load.

2. The performance of the hybrid models (EMD-SVR, EEMD-SVR, VMD-BP, CEEMDAN-SVR, and CEMDAN-SE-VMD-SVR) preprocessed by modal decomposition is better than that of the BP, SVR, and LSTM models, indicating the effectiveness of modal decomposition technology in reducing non-linearity short-term load sequences.

3. As two kinds of signal processing technologies with high decomposition efficiency, CEEMDAN and VMD have good complementary advantages. After the sample entropy screening, the proposed hybrid model is decomposed into a two-phase mode to obtain a more stable load sequence component, which reduces the difficulty of prediction. Compared with the traditional single-phase decomposition technology, it is improved to a certain extent and has better practicability.

4. PSR can effectively optimize the sequence components and form the input of SVR. Moreover, the WOA algorithm is very effective for optimizing and selecting the key parameters (penalty coefficient $C$ and kernel parameter $g$) inside the SVR model. It has a distinct effect on improving the prediction accuracy of the hybrid model and the fitting effect reaches 99.32%.

5. The hybrid model proposed in this paper combines historical data preprocessing, prediction model optimization, and prediction error analysis to carry out research on short-term load forecasting, which is more conducive to achieving higher forecasting accuracy.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

REFERENCES

[1] E. A. Foinberg and D. Genethliou, “Load forecasting,” in Applied Mathematics for Restructured Electric Power Systems. Cham, Switzerland: Springer, 2005, pp. 269–285.

[2] Y. Zhao, X. Zhang, and K. Xie, “Application of nonparametric autoregression to short-term load forecasting,” Gaodianya Jishu/High Voltage Eng., vol. 37, no. 2, pp. 429–435, 2011.

[3] S.-J. Huang and K.-R. Shih, “Short-term load forecasting via ARMA model identification including non-Gaussian process considerations,” IEEE Trans. Power Syst., vol. 18, no. 2, pp. 673–679, May 2003.
[4] C.-M. Lee and C.-N. Ko, “Short-term load forecasting using lifting scheme and ARIMA models,” Expert Syst. Appl., vol. 38, no. 5, pp. 5902–5911, May 2011.

[5] O. O. Bozkurt, G. Biricik, and Z. C. Tayi, “Artificial neural network and ARIMA based models for power load forecasting in Turkish electricity market,” PLoS ONE, vol. 12, no. 4, Apr. 2017. Art. no. e0175915.

[6] P. C. Jackson, Introduction to Artificial Intelligence. New York, NY, USA: Dover, 2019.

[7] G.-C. Liao and T.-P. Tsoa, “Application of fuzzy neural networks and artificial intelligence for load forecasting,” Electr. Power Syst. Res., vol. 70, no. 3, pp. 237–244, Aug. 2004.

[8] E. Alpaydin, Introduction to Machine Learning. Cambridge, MA, USA: MIT Press, 2020.

[9] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, May 2015.

[10] C. Cecati, J. Kolbusz, P. Rozycki, P. Siano, and B. M. Wilamowski, “A novel RBF training algorithm for short-term electric load forecasting and comparative studies,” IEEE Trans. Ind. Electron., vol. 62, no. 10, pp. 6519–6529, Oct. 2015.

[11] W.-C. Hong, “Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm,” Energy, vol. 36, no. 9, pp. 5568–5578, Sep. 2011.

[12] G.-F. Fan, L.-L. Peng, W.-C. Hong, and F. Sun, “Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression,” Neurocomputing, vol. 173, pp. 958–970, Jan. 2016.

[13] G.-F. Fan, S. Qing, H. Wang, W.-C. Hong, and H.-J. Li, “Support vector regression model based on empirical mode decomposition and auto regression for electric load forecasting,” Energies, vol. 6, no. 4, pp. 1887–1901, Apr. 2013.

[14] Z. He, C. Li, Y. Shen, and A. He, “A hybrid model equipped with the minimum cycle decomposition concept for short-term forecasting of electrical load time series,” Neural Process. Lett., vol. 46, no. 3, pp. 1099–1081, Dec. 2017.

[15] Z. Liu, W. Sun, and J. Zeng, “A new short-term load forecasting method of power system based on EEMD and SS-PSO,” Neural Comput. Appl., vol. 24, nos. 3–4, pp. 973–983, Mar. 2014.

[16] D. Liu, K. Sun, H. Huang, and P. Tang, “Monthly load forecasting based on economic data by decomposition integration theory,” Sustainability, vol. 10, no. 9, p. 3282, Sep. 2018.

[17] T. Liu, Y. Jin, and Y. Gao, “A new hybrid approach for short-term electric load forecasting applying support vector machine with ensemble empirical mode decomposition and whale optimization,” Energies, vol. 12, no. 8, p. 1520, Apr. 2019.

[18] Z. Zhang and W.-C. Hong, “Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm,” Nonlinear Dyn., vol. 98, no. 2, pp. 1107–1136, Oct. 2019.

[19] W. Lu, Y. Rui, Z. Yi, B. Ran, and Y. Gu, “A hybrid model for lane-level traffic flow forecasting based on complete ensemble empirical mode decomposition and extreme gradient boosting,” IEEE Access, vol. 8, pp. 42042–42054, 2020.

[20] Y. Ren, P. N. Suganthan, and N. Srikanth, “A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods,” IEEE Trans. Sustain. Energy, vol. 6, no. 1, pp. 236–244, Jan. 2015.

[21] 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, pp. 13086–13104, 2020.

[22] S. Kim, G. Lee, G.-Y. Kwon, D.-I. Kim, and Y.-J. Shin, “Deep learning based on multi-decomposition for short-term load forecasting,” Energies, vol. 11, no. 12, p. 3433, Dec. 2018.

[23] 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 Convers. Manage., vol. 205, Feb. 2020. Art. no. 112461.

[24] D. Wang, S. Wei, H. Luo, C. Yue, and O. Grunder, “A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine,” Sci. Total Environ., vol. 580, pp. 719–733, Feb. 2017.

[25] M. Sibtain, X. Li, G. Nabi, M. I. Azam, and H. Bashir, “Development of a three-stage hybrid model by utilizing a two-stage signal decomposition methodology and machine learning approach to predict monthly runoff at swat river basin, Pakistan,” Discrete Dyn. Nature Soc., vol. 2020, pp. 1–19, May 2020.

[26] 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 Convers. Manage., vol. 157, pp. 1–12, Feb. 2018.

[27] S. Mirjalili and A. Lewis, “The whale optimization algorithm,” Adv. Eng. Softw., vol. 95, pp. 51–67, May 2016.

[28] I. Aljarah, H. Faris, and S. Mirjalili, “Optimizing connection weights in neural networks using the whale optimization algorithm,” Soft Comput., vol. 22, no. 1, pp. 1–15, Jan. 2018.

[29] P. Flandrin, G. Rilling, and P. Goncavles, “Empirical mode decomposition as a filter bank,” IEEE Signal Process. Lett., vol. 11, no. 2, pp. 112–114, Feb. 2004.

[30] Z. Wu and N. E. Huang, “Ensemble empirical mode decomposition: A noise-assisted data analysis method,” Adv. Adapt. Data Anal., vol. 1, no. 1, pp. 1–41, Jan. 2009.

[31] M. E. Torres, M. A. Coloninas, G. Schlotthauer, and P. Flandrin, “A complete ensemble empirical mode decomposition with adaptive noise,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2011, pp. 4144–4147.

[32] S. Tian, X. Bian, Z. Tang, K. Yang, and L. Li, “Fault diagnosis of gas pressure regulators based on CEEMDAN and feature clustering,” IEEE Access, vol. 7, pp. 132492–132502, 2019.

[33] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” IEEE Trans. Signal Process., vol. 62, no. 3, pp. 531–544, Feb. 2014.

[34] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, “Geometry from a time series,” Phys. Rev. Lett., vol. 45, no. 9, p. 712, Sep. 1980.

[35] W. Fu, K. Wang, C. Li, and J. Tan, “Multi-step short-term wind speed forecasting approach based on multi-scale dominant ingredient chaotic analysis, improved hybrid GWO-SCA optimization and ELM,” Energy Convers. Manage., vol. 187, pp. 356–377, May 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,” Trans. Inst. Meas. Control, vol. 41, no. 15, pp. 4436–4449, Nov. 2019.

[37] W. Fu, K. Wang, J. Tan, and K. Shao, “Vibration tendency prediction approach for hydropower generator fused with multiscale dominant ingredient chaotic analysis, adaptive mutation grey wolf optimizer, and KELM,” Complexity, vol. 2020, pp. 1–20, Feb. 2020.

[38] M. Ohmi and H. Mori, “A Gaussian processes technique for short-term load forecasting with consideration of uncertainty,” IEEE Trans. Power Energy, vol. 126, no. 2, pp. 202–208, 2006.

[39] S. Zhu, X. Yuan, Z. Xu, X. Luo, and H. Zhang, “Gaussian mixture model coupled recurrent neural networks for wind speed interval forecast,” Energy Convers. Manage., vol. 198, Oct. 2019, Art. no. 111772.

[40] G. Xi and Z. Liao, “Short-term load forecasting based on phase space reconstruction and Gaussian process regression,” Power Syst. Protection Control, vol. 45, no. 5, pp. 73–79, 2017.
QIANG SHI is currently pursuing the master’s degree in electrical engineering with China Three Gorges University (CTGU). His research interests include machine learning, power load forecasting, and optimal reservoir operation.

MUHAMMAD SIBTAI N is currently pursuing the Ph.D. degree in electrical engineering with China Three Gorges University (CTGU). His research interests include machine learning and optimal reservoir operation.

DAN LI received the Ph.D. degree in electrical engineering from Chongqing University (CQU), Chongqing, China, in 2017. She is currently working with the Electrical Engineering Department, China Three Gorges University (CTGU). Her research interests include artificial intelligence and new energy power prediction.

DANIEL ELIOTE MBANZE received the M.S. degree in electrical engineering from China Three Gorges University (CTGU), Yichang, China, in 2019. He is currently with the Electrical Engineering Department, Higher Education Polytechnic Institute, Mozambique. His research interests include reinforcement learning and optimal reservoir operation.