

Received 20 August 2022, accepted 11 September 2022, date of publication 14 September 2022,  
date of current version 30 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3206486

## RESEARCH ARTICLE

# Short-Term Power Load Prediction Based on VMD-SG-LSTM

QIU SUN<sup>1</sup> AND HUAFENG CAI<sup>1,2</sup>

<sup>1</sup>School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, China

<sup>2</sup>Xiangyang Industrial Institute, Hubei University of Technology, Xiangyang 210023, China

Corresponding author: Huafeng Cai (whgkzj@hbut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61473116.

**ABSTRACT** Power load prediction plays an important role in the safety and stability of national power system. However, due to the nonlinear and multi-frequency characteristics of the power system itself, power load prediction is difficult. To address this problem, we propose a short-term power load prediction model based on variational mode decomposition (VMD). First, original data are decomposed into intrinsic mode function (IMF) of different frequencies using the VMD algorithm, and the decomposed sub-functions are reconstructed. After smoothing the reconstructed data by Savitzky-Golay (S-G) filtering algorithm, the change trend of raw data (CTRD) is obtained. Then, IMF, CTRD and raw data are used as inputs to predict short-term power load by long short-term memory network (LSTM). Finally, the proposed prediction model is compared with the other two groups of prediction models. The results show that the proposed VMD-SG-LSTM prediction model has high fitting ability and high prediction accuracy, and is an effective method for short-term power load prediction.

**INDEX TERMS** Power load prediction, variational mode decomposition, intrinsic mode function, S-G filtering, long short-term memory network.

## I. INTRODUCTION

In recent years, with the rapid development of science and technology in the world and the increasing population, people's dependence on electricity in their daily lives has gradually increased, which has led to a dramatic increase in the global demand for electricity. In this environment, some serious power problems can arise. For example, it is a very important task for a country to maintain the balance between supply and demand in its power system. If the power generation capacity of the national power system is lower than the supply demand, there will be large-scale blackouts, which will affect people's normal life and social production and cause immeasurable losses. On the contrary, if the power generation capacity of the national power system is higher than the supply demand, this will lead to the waste of resources caused by leaving the power plants idle for a certain period of time [1]. Therefore, how to accurately predict the future

electricity demand so as to maintain the supply and demand balance of the power system has become a top priority.

In the field of power load prediction, many prediction methods have been proposed, including physical model method, statistical method, and artificial intelligence method [2], [3], [4], [5], [6], [7]. The physical model can be used to accurately predict the power load, but the difficulty of the physical model design is that it is difficult to accurately describe the necessary information of the system components in the face of a more complex model. Also, its transferability is poor. When the designed model is applied to other systems for prediction, its accuracy will be greatly reduced. Statistical methods rely more on the periodicity and outliers of data [8]. In the prediction process, statistical methods tend to identify the intrinsic patterns from the historical data of power load and compare them with other parameters, and then summarize the data for prediction. However, the relationship between power load and other parameters is generally complex and nonlinear, and for these reasons, it is difficult to obtain accurate predicted values with statistical

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

methods [9]. Therefore, more and more scholars are using artificial neural networks to predict energy consumption. The reason is that artificial neural networks have the function of autonomous learning. It can keep adjusting its model parameters by learning historical data so as to gradually approach the real value [10]. The neural networks can deal with complex nonlinear problems and has certain portability. The trained model can be transplanted to other energy consumption models and still ensure high prediction accuracy. Among many researchers, Bian *et al.* [11] proposed a short-term power load forecasting model based on K-means and FCM-BP. Its advantage is that the improved mean clustering method can filter out the local similar data in the data, so as to improve the accuracy of power load forecasting by taking the important features as the input of BP network. However, BP network as a feedforward neural network has no obvious advantages in time series data forecasting. Liao *et al.* [12] proposed a multi-wavelet convolution neural network for power load forecasting. Its advantage is that different types of wavelet are used to reconstruct the original load data to obtain different model inputs, and the ability of convolution neural network feature extraction is used to improve the accuracy of power load forecasting. However, compared with LSTM network, convolutional neural network has lower advantages in learning time series data. These methods allow better adaption to nonlinear spikes, more accurate modeling of data features, and better prediction accuracy. In recent years, they have become a major research direction in energy prediction.

With the rapid development of neural networks and the continuous improvement of prediction accuracy, more and more scholars have started to add wavelet transform, empirical mode decomposition, ensemble empirical mode decomposition, variational mode decomposition and other [13], [14], [15], [16], [17], [18], [19], [20] time series decomposition methods into the prediction model to improve the prediction effect. Bahrami *et al.* [21] proposed a method based on wavelet transform and grey model combination for short-term power load forecasting, which has the advantage of using wavelet transform to eliminate the high frequency components in power load data and improve the accuracy of forecasting. However, due to the lack of adaptability of wavelet transform itself, it has less advantages than other decomposition algorithms. On this basis, Liang *et al.* [22] proposed a short-term power load forecasting method based on empirical mode decomposition and improved regression neural network. The advantage of this method is that the original data is decomposed into IMF with different frequencies by empirical mode decomposition to weaken the volatility of data and improve the prediction accuracy. However, the empirical mode decomposition itself has a modal aliasing phenomenon, which has less advantages than VMD. From the above study, it can be seen that the inclusion of time series decomposition method in the prediction model can better improve the prediction accuracy of the model.

Due to the randomness and volatility of power load data, this paper applies data enhancement techniques to the field of

power load prediction and proposes a short-term power load prediction model based on VMD-SG-LSTM. By using IMF, CTRD and raw data as inputs, a new power load dataset is constructed and the capability of LSTM to process long-term sequences is used to predict short-term power load, thus improving the fitting ability and prediction accuracy of the model. Its main contributions are as follows:

- (1) An improved parameter selection method for Savitzky-Golay filter is proposed. By setting the value of MAE, the appropriate window width and polynomial fitting order can be automatically allocated, which can ensure the filtering effect and greatly save the selection time of parameters.
- (2) A short-term power load prediction model based on VMD-SG-LSTM is proposed. By taking IMF, CTRD and raw data as input, and using the ability of LSTM to process long time series to predict short-term power load, high prediction accuracy can be achieved.
- (3) Compared with the other two groups of prediction models, the VMD-SG-LSTM model not only has high prediction accuracy, but also has good fitting effect on the predicted peak and trough parts. Compared with the other three latest prediction models, this model has the best prediction performance.

The main contents of this paper are organized as follows. Section II mainly introduces the algorithms used in this paper. Section III presents the proposed VMD-SG-LSTM prediction model and the performance evaluation method of the model. Section IV presents the experimental and comparative analysis of the proposed method. Finally, the conclusion is given in Section V.

## II. PRINCIPLE OF MODEL CONSTRUCTION

### A. VARIATIONAL MODE DECOMPOSITION

VMD is a variational structured signal processing method which integrates the Hilbert transform method, the alternating direction multiplier method, and the Wiener filter method [23]. More specifically, the VMD decomposition algorithm can decompose an actual signal or original data into K bandwidth-based eigenmode functions. The advantage of VMD is that it can decompose unstable and nonlinear time-series signals and filter out some noise and interference from the original signal. Theoretically, the modal aliasing phenomenon can be effectively suppressed by selecting an appropriate value of K [24]. The main flow of VMD decomposition algorithm is as follows:

Assuming that the original signal is composed of K finite bandwidth intrinsic mode components  $u_k(t)$ , the central frequency of each IMF is  $\omega_k$ , and the model  $u_k(t)$  of the analyzed signal is calculated by Hilbert transform, the unilateral spectrum can be expressed as:

$$(\delta(t) + \frac{j}{\pi t}) * u_k(t) \quad (1)$$

In the analysis signal of the model is multiplied by the operator  $e^{-j\omega_k t}$ , and the model  $u_k(t)$  can be modulated to the

corresponding baseband:

$$\left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \quad (2)$$

The square norm  $L^2$  of the demodulation gradient is calculated and the bandwidth of each modal signal is estimated:

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|^2 \right\} \\ s.t. \sum_k u_k(t) = s \end{cases} \quad (3)$$

In the above equation,  $\{u_k\} = \{u_1, u_2, \dots, u_k\}$  represents the decomposed IMF, and  $\{\omega_k\} = \{\omega_1, \omega_2, \dots, \omega_k\}$  represents the central frequency of each component.

The constrained optimization problem is transformed into an unconstrained optimization problem by introducing the Lagrange function  $\lambda_t$  and the second-order penalty factor  $\alpha$ :

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) \\ = \alpha \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|^2 \\ + \left\| s(t) - \sum_k u_k(t) \right\|^2 + \left\langle \lambda(t), s(t) - \sum_k u_k(t) \right\rangle \end{aligned} \quad (4)$$

The alternating direction multiplier method is used to update each component and its central frequency continuously, and finally the optimal solution of the unconstrained model is obtained. According to the frequency domain space, all components can be obtained from the following formula:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{s}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (5)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| u_k^{n+1}(\omega) \right|^2 d\omega}{\int_0^\infty \left| u_k^{n+1}(\omega) \right|^2 d\omega} \quad (6)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \lambda \left( \hat{s}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right) \quad (7)$$

The above steps are repeated until the iteration stopping condition is satisfied. Finally, we can decompose the original input signal into K IMFs by VMD algorithm. However, the selection of parameter K needs to be determined in advance, otherwise it will affect the performance of VMD decomposition. Therefore, this paper adopts a method to select K values based on the improved signal energy (ISE). The principle is that the parameter K can be determined when the ratio of residual energy to original energy is small enough and there is no obvious downward trend [25]. The formula

is defined as follows:

$$\text{Erse} = \frac{\sum_{n=1}^N \left| f[n] - \sum_{k=1}^K u_k[n] \right|^2}{\sum_{n=1}^N f[n]^2} \times 100\% \quad (8)$$

In the formula, Erse represents the ratio of residual energy to original energy. When the ratio is less than 1%, we can consider the ratio small enough. When the decreasing trend tends to be level off from steep, the corresponding K value can be selected as the number of IMFs.

### B. SAVITZKY-GOLAY FILTER

The S-G filter is characterized by the ability to filter out noise and interference while ensuring that the shape and width of the original signal does not change, thus the change trend of the original signal can be more effectively preserved and analyzed [26]. The principle of the S-G filter is to convolve a certain length of filter with the data to be processed using a weighted average algorithm of moving windows, while fitting a weighted polynomial to the data to be processed that minimizes the root mean square error of the fitted target, thereby discarding some edge points far from the majority of points [27]. The basic formula of the S-G filter is:

$$Y_j^* = \frac{\sum_{i=-m}^m C_i \times Y_{j+i}}{N} \quad (9)$$

where  $Y^*$  is the fitting value,  $Y$  is the original value of the signal, and  $C$  is the coefficient of the S-G polynomial fitting, indicating the coefficient of the  $i$ th filtering from the beginning of the filter.  $m$  is the width of the half filtering window, and  $N$  is the length of the filter, which is equal to the width of the sliding array  $2m + 1$ . Smoothing filtering with the S-G algorithm improves the smoothness of the original data and reduces noise interference. Another core of the S-G algorithm is to set two parameters in the algorithm, namely the window size and the polynomial fitting order. For a given signal, the correct choice of two parameters will directly lead to different filtering effects. When a low-order large window is selected, the intensity of the absorption peak diminishes and the absorption line becomes wider, leading to distortion of the signal and difficult in retaining the required information. When the parameters choose high-order small window, although the original information of the signal is better retained, the filtering effect on noise is also reduced. So how to choose the S-G filter window size and polynomial fitting order is the key to affect the filtering effect. In this paper, an improved S-G filter parameter selection method is proposed according to the actual control requirements. The method flow chart is shown in figure 1.

In figure 1, the mean absolute error (MAE) is used as the parameter selection criteria for the S-G filter. First, the desired filtering effect  $X$  is input to the system, and then the system automatically assigns the window size  $i$  and the

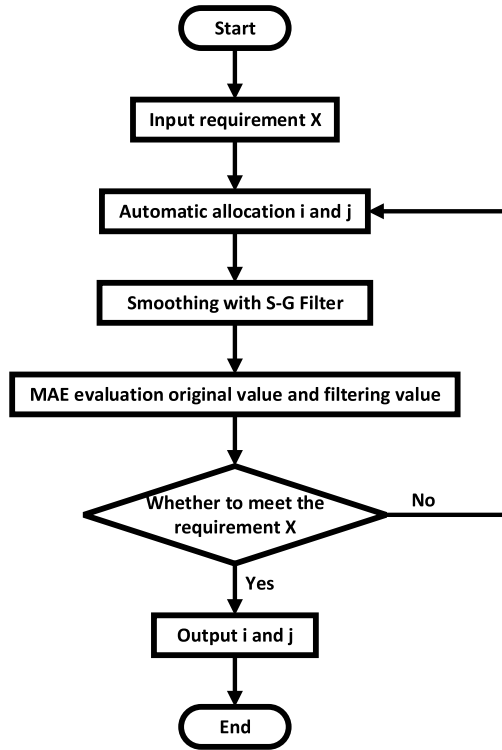


FIGURE 1. Flow chart of improved S-G filter parameter selection method.

polynomial fitting order  $j$  in the order from smallest to largest. The original signal, the allocated window size and the polynomial fitting order are then brought into the S-G filter to obtain the filtering value. Finally, the original signal and the filtering value are brought into the MAE algorithm, and the final result is calculated and compared with the desired filtering effect  $X$ . When the final result is the same as  $X$ , the distribution value is output. When the final result is different from  $X$ , the window size and polynomial fitting order are redistributed until the filtering effect is satisfied. This improved S-G filter parameter selection method can effectively save time in parameter selection and also express CTRD more accurately.

C. LSTM NETWORK

LSTM neural network is a special kind of recurrent neural network (RNN). But unlike RNN, LSTM can apply previous information to the current task, i.e., RNN has a certain memory capability [28]. within contrast to other neural networks, long short-term memory neural networks introduce input gate, forget gate and output gate in each unit to control the input value, memory value and output value respectively [29]. This makes the long short-term memory neural network have a strong memory function when dealing with time series-like models, and at the same time, it can solve some gradient vanishing problems according to the characteristics of the model. However, it also has the obvious advantage that the gradient disappears when the sequence it processes exceeds its own upper limit, and in terms of computation, as the network span becomes longer, the computation of the model

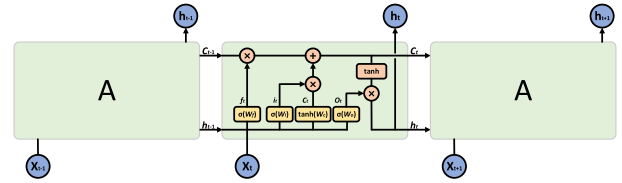


FIGURE 2. LSTM neural network structure diagram.

keeps increasing. The LSTM model structure is shown in figure 2.

As can be seen from the graph, LSTM network contains three gate functions, from left to right, the forgetting gate, the input gate and the output gate, and  $f_t$ ,  $i_t$  and  $o_t$  represent the state values of the forgetting gate, the input gate and the output gate respectively. The sigmoid layer in the forgetting gate  $C_{t-1}$  determines the information to be forgotten in the past historical data, while the current layer  $x_t$  and the output  $h_{t-1}$  of the previous layer are used as input to the forgetting gate. The output of the forgetting door is:

$$f_t = \sigma \left( W_f [x_t, h_{t-1}]^T + b_f \right) \tag{10}$$

In the input gate, the function is to update the data state in three steps. First, the input gate updates the information to be remembered by the result of sigmoid layer, and then a new candidate value is generated by the tanh layer. Finally, a new data state is obtained by adding the part to be forgotten and the part to be remembered. The formula is as follows:

$$i_t = \sigma \left( W_i [x_t, h_{t-1}]^T + b_i \right) \tag{11}$$

$$\tilde{C}_t = \tanh \left( W_c [x_t, h_{t-1}]^T + b_c \right) \tag{12}$$

$$C_t = C_{t-1} \times f_t + i_t \times \tilde{C}_t \tag{13}$$

In the output gate, the sigmoid layer is used to determine which part of the output data state  $C_t$  is, and then the tanh layer is used to scale the updated data state value between  $[-1, 1]$  and multiply it with the output of the sigmoid layer to obtain the final output  $h_t$ , the formula is as follows:

$$o_t = \left( W_o [x_t, h_{t-1}]^T + b_o \right) \tag{14}$$

$$h_t = o_t \times \tanh (C_t) \tag{15}$$

III. POWER LOAD PREDICTION MODEL

A. STRUCTURE OF PREDICTION MODEL

The VMD-SG-LSTM power load prediction model is proposed by combining decomposition algorithm, filtering algorithm, data enhancement and neural network techniques. The model mainly consists of decomposition module, enhancement module and prediction module. Its structure is shown in figure 3. It can accurately predict the future short-term power load, and the model structure is described below.

In the decomposition module, the VMD algorithm is used to decompose the original data to obtain IMFs of different frequencies. Using IMF as input can improve the accuracy

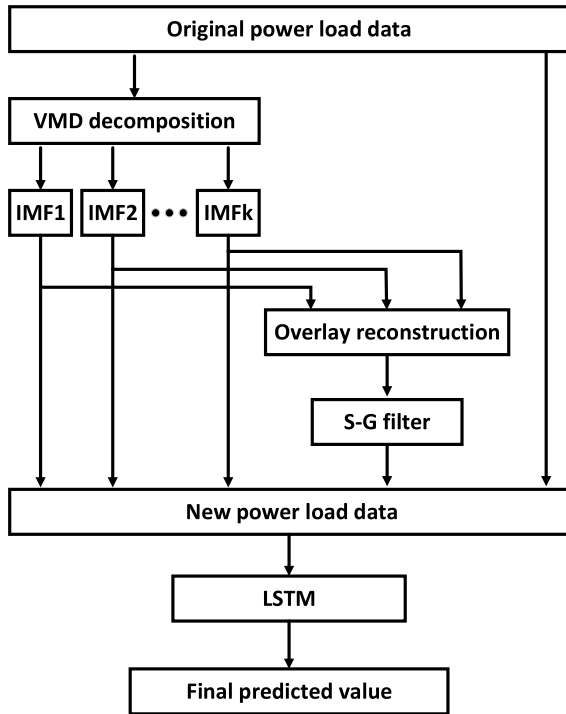


FIGURE 3. Structure diagram of VMD-SG-LSTM power load prediction model.

of prediction. At the same time, as a decomposition technology with noise reduction function, it can remove part of the interference existing in the power load, thereby improving the reliability of the data. In the process of decomposing original data using VMD algorithm, parameter K is determined by formula (8), that is, parameter K can be determined when the ratio of residual energy to original energy is small enough and there is no obvious downward trend.

In the enhancement module, the decomposed sub-functions are superimposed and reconstructed to obtain the denoised power load data. Then, the S-G filtering algorithm is used to smooth the denoised power load data to obtain CTRD, which is used to improve the fitting ability and prediction accuracy of the model. Finally, the IMF and CTRD are added to the raw data set to obtain a new data set, thus achieving the effect of data enhancement. In the design of the enhancement module, the difficulty lies in how to calculate the CTRD quickly and effectively, and the key lies in selecting the appropriate window size and polynomial fitting order. Here we propose an improved S-G filter parameter selection method, and the specific process is shown in figure 1.

In the prediction module, the newly generated power load data set is taken as input and brought into the LSTM neural network for prediction. The ability of feature extraction of LSTM time series is used to further improve the prediction accuracy, so as to obtain the final power load prediction value. In the design of the prediction module, the parameters of the prediction model need to be set to achieve better prediction results, and the specific parameter adjustment process will be introduced in Section IV.

### B. EVALUATION INDEX OF PREDICTION MODEL

In this experiment, it is also necessary to evaluate the prediction ability of VMD-SG-LSTM model, so that the pros and cons of each model can be seen by comparison. In this paper, two evaluation methods are selected, namely Mean Absolute Scaled Error (MASE) and Mean Absolute percentage error (MAPE) [30]. Among them, MASE is an evaluation method based on MAE. Its characteristic is that the greater the error is, the greater the value is, where  $y_i$  is the real value of the current moment,  $y_{i-1}$  is the real value of the previous moment, and  $\hat{y}_i$  is the predicted value of the current moment. MASE formula is as follows:

$$MASE = \frac{\frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|}{\frac{1}{N-1} \sum_{i=2}^N |y_i - y_{i-1}|} \tag{16}$$

MAPE considers not only the error between the predicted value and the actual value, but also the ratio between the error and the actual value. It is generally believed that MAPE measures the accuracy of the prediction. The smaller the MAPE value, the higher the prediction accuracy. The formula is as follows:

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right| \tag{17}$$

## IV. EXPERIMENTS AND ANALYSIS

### A. DATA PROCESSING

In order to verify the feasibility and effectiveness of the proposed power load prediction model, this experiment uses the power load situation of the Belgian power grid company Elia for forecasting and analysis [31]. It includes the measured net generation from local power stations that inject power to the Elia grid, the netto inflows from the distribution to the Elia grid and the netto import at the borders. The dataset mainly recorded the power load every 15 minutes from January 1, 2020 to December 31, 2021, with a total length of 70176 and three missing values, all processed using the mean value method. The power load situation of grid company Elia is shown in figure 4.

Figure 4 can be analyzed in terms of power load trend and power load size. From the power load trend, Elia’s power load shows a cyclical characteristic. From the power load size, the minimum power load of Elia is less than 0.4MW and the maximum power load is nearly 1.4MW. It can be seen that its variation range and magnitude are particularly obvious. At the same time, many abnormal values can be found in the diagram. Through the analysis we judge that Elia’s power load is nonlinear and unstable, which will have a certain impact on power load prediction, but at the same time this difference can also detect the performance of the prediction model more comprehensively.

### B. MODEL PARAMETERS SETTING

In the parameter setting of the VMD-SG-LSTM model, the K value in the VMD algorithm is determined by the

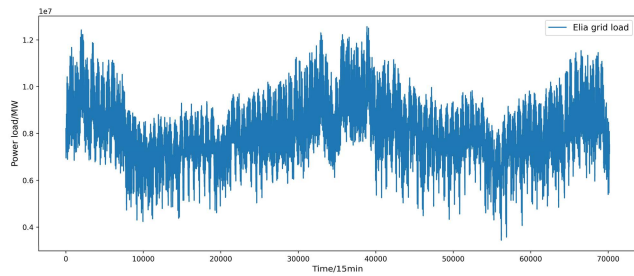


FIGURE 4. Power load diagram of power grid company Elia.

improved ISE rule. The original time series is decomposed into multiple IMF components by the VMD algorithm, and the decomposed IMF components are reconstructed to obtain the denoised power load data. Then, the reconstructed data are smoothed by the improved S-G filtering algorithm to obtain the CTRD, and then the IMF, CTRD and the raw data are used as inputs. Finally, the LSTM network is used to predict the short-term power load to obtain the final power load prediction value.

In the selection of K value, first set other parameters of VMD, set penalty factor  $\alpha$  as 2000, set other parameters as default values, and then calculate the ratio of remaining energy to original energy with K value from 1 to 9. The ratio of residual energy to original energy is shown in figure 5. ISE stipulates that when the ratio is less than 1%, Erse can be considered to be small enough at this time, and when the declining trend tends to be level off gradually from steep, the corresponding K value can be selected as the number of IMFs at this time. As can be seen from the figure, when K is 2, the Erse at this time is less than 1%, but there is no obvious downward trend. When K is 4, the Erse value flat at this time, and 4 is finally selected as the number of IMFs through comparison. After the K value is selected, the original power load data are decomposed by the VMD algorithm, and the decomposed components are shown in figure 6.

After decomposing the original power load data into four IMF signals and one residual signal by VMD algorithm, the denoised power load data can be obtained by reconstructing the four IMF signals. The denoised power load data are shown in figure 7.

It can be seen that by noise reduction of the raw power load data, some fluctuations and outliers in the raw data can be removed, so that the curve after noise reduction becomes smoother and also retains the important information of the raw data to the maximum extent. However, there are still some jagged fluctuations that are not easy to remove after noise reduction, which makes the prediction model unable to accurately represent the CTRD. At this time, we use the improved S-G filtering algorithm to smooth the original data after noise reduction. The sawtooth fluctuation after noise reduction is shown in figure 8.

To smooth the noise-reduced original data, it is necessary to select a suitable MAE as the evaluation index. In order to

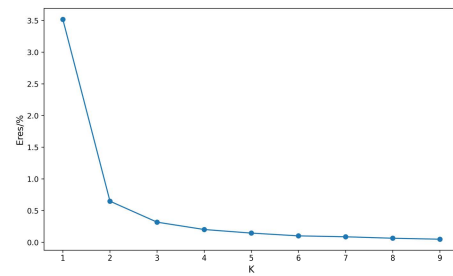


FIGURE 5. Ratio of residual energy to original energy.

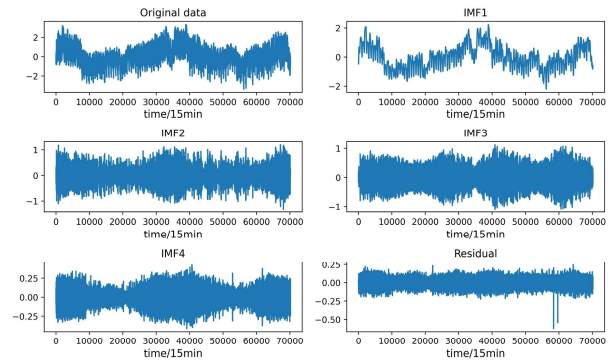


FIGURE 6. Each component after VMD decomposition.

reflect the CTRD more clearly and retain the important information of the raw data to the maximum extent, we compare different MAE parameters with the noise-reduced original data, as shown in figure 9.

It can be seen that when MAE is 0.02, the important information of the noise reduction data can be well retained, but the performance in removing the jagged fluctuation is not good, and there are still obvious fluctuations. When MAE is 0.04, it can be found that the jagged fluctuations are completely eliminated, but some important information of the noise reduction data is also lost. Therefore, in order to better reflect the CTRD and maximize the retention of important information from the raw data, the evaluation index of MAE is finally set to 0.03. At this point, using the improved S-G parameter selection method, the window size and the polynomial fitting order can be quickly calculated to be 25 and 2, respectively. Then, each parameter is brought into the S-G filter for smoothing, and finally the change trend of the raw data is obtained. It can be seen that the jagged fluctuations after noise reduction have been significantly improved. At the same time, the important information of the raw data is retained to the maximum extent, and the change trend of the original data can be reflected more. The decomposed IMF components and CTRD are added to the raw power load data to obtain a new power load data set. The new power load data set is shown in Table 1.

When constructing a new power load data set, it is also necessary to bring the data set into the LSTM model for short-term power load prediction. In the design of LSTM

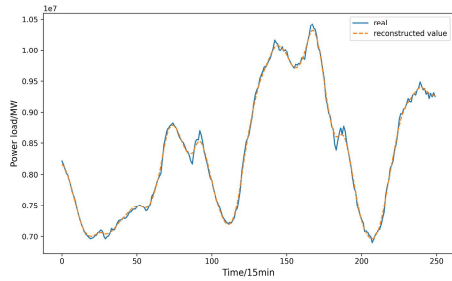


FIGURE 7. Power load data after noise reduction.

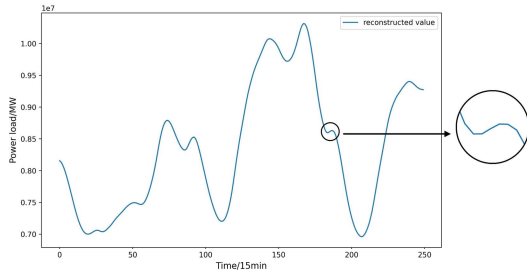


FIGURE 8. Sawtooth fluctuation after noise reduction.

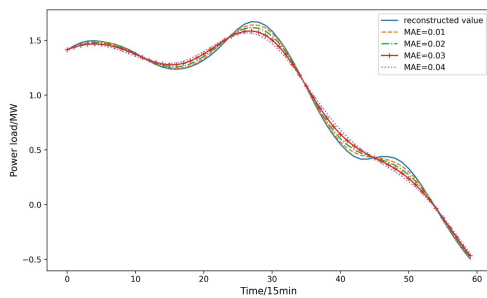


FIGURE 9. Change of MAE in noise reduction data.

prediction model, the number of hidden layers is set to two layers, namely the LSTM layer and the full connection layer. It is generally believed that increasing the number of hidden neurons can reduce the training error and test error of the neural network to improve the prediction accuracy of the model. In this paper, MAE is used as the standard of model performance evaluation to select the number of hidden layer neurons. The number of neurons in each layer is selected as shown in Table 2 and 3.

From Table 2 and Table 3, it can be seen that the MAE value is the smallest when the number of neurons in LSTM layer is 11 and the number of neurons in the fully connected layer is 10. Therefore, the number of neurons in the LSTM layer is set to 11, and the number of neurons in the full connection layer is set to 10. At this time, the LSTM prediction model is designed and completed, and the new power load data set is brought into the LSTM model for training and testing. If the new power load data set is directly used as the input of the model, it will lead to the internal overflow of the computer and cannot be trained. Therefore,

it is necessary to generate samples for the data set. When the length of each training sample is set to 672 and the step length of sliding window is set to 1, a set of data sets with the size of [69503, 672, 6] can be obtained. According to the ratio of 90% and 10%, the new power load data set is divided into training set and test set. At the same time, the Min-Max function is used to normalize the data to accelerate the speed of gradient descent to obtain the optimal solution [32]. The normalization formula is as follows:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (18)$$

At the same time, in order to improve the training effect of the model, this experiment the mini-batch training method and the Adam optimizer to train the model. The number of samples Batch size for each input model is set to 128, and the number of iterations is set to 1000. The error curve of model training is shown in figure 10. It can be seen that the model is basically convergent and stable after 1000 iterations, and there is no overfitting problem.

### C. COMPARISON OF PREDICTION MODELS

In this experiment, in order to better verify the feasibility and effectiveness of the proposed prediction model, the improved prediction model and the latest prediction model are compared with the VMD-SG-LSTM model to predict the power load in the next hour, and MASE and MAPE are used to evaluate the prediction performance of each model. Firstly, the LSTM model and VMD-LSTM model before improvement are compared with VMD-SG-LSTM model, and the prediction results are shown in figure 11.

It can be seen from figure 11 that the prediction performance of LSTM is poor, and there is obvious oscillation in the process of power load prediction, which will increase the instability of prediction and lead to large deviation from the actual value. Compared with the LSTM model, the prediction accuracy of VMD-LSTM model is higher, which is because the VMD algorithm can extract the characteristic information of the original signal, so as to further improve the prediction accuracy. The VMD-SG-LSTM model proposed in this paper adds the S-G algorithm on the basis of the original prediction model, which can be closer to the real value curve in the peak part. At the same time, it also makes the prediction curve smoother, and has a certain fitting effect in predicting some values with large time span.

Secondly, in the comparison of the latest prediction models, this paper uses the VMD-Bi-LSTM prediction model proposed by Tang *et al.* [33], the VMD-CISSA-LSSVM prediction model proposed by Wang *et al.* [34] and the VMD-GWO-SVR prediction model proposed by Zhou *et al.* [35] to compare with the VMD-SG-LSTM. Each parameter is set according to the value in the article to predict the power load in the next hour. The prediction results are shown in figure 12.

It can be seen from figure 12 that the four models have high prediction accuracy, but careful observation shows that

TABLE 1. New power load data.

Number	IMF 1	IMF 2	IMF 3	IMF 4	CTRD/MW	Elia Grid Load/MW
1	-0.4649	0.2353	0.2123	0.1091	8267955.45	8209190
2	-0.4655	0.2322	0.2066	0.1021	8174069.48	8155130
3	-0.4665	0.2266	0.1960	0.0892	8082959.70	8084900
4	-0.4672	0.2189	0.1811	0.0719	7994626.11	8016320
5	-0.4677	0.2094	0.1626	0.0515	7909068.70	7985210
6	-0.4682	0.1978	0.1406	0.0289	7826287.48	7886800
7	-0.4687	0.1844	0.1157	0.0056	7746282.44	7799050
...	...	...	...	...	...	...
70170	-0.6012	0.1085	-0.0639	0.0342	7251993.52	7388656
70171	-0.5996	0.1039	-0.0826	0.0874	7280233.29	7430679
70172	-0.5989	0.0991	-0.0999	0.1341	7316584.53	7411482
70173	-0.5988	0.0947	-0.1153	0.1726	7361047.23	7528237
70174	-0.6000	0.0898	-0.1289	0.2003	7413621.39	7461745
70175	-0.6018	0.0853	-0.1396	0.2174	7474307.01	7417540
70176	-0.6032	0.0824	-0.1458	0.2252	7543104.09	7314050

TABLE 2. Selection of neuron number in LSTM layer.

Number of neurons in LSTM layer	Epochs	Batch size	MAE
1	1000	128	45643.45
3	1000	128	44221.16
5	1000	128	41870.88
7	1000	128	40054.71
9	1000	128	38314.82
11	1000	128	36812.70
13	1000	128	38443.50
15	1000	128	40337.95

TABLE 3. Selection of neuron number in fully connected layer.

Number of neurons in fully connected layer	Epochs	Batch size	MAE
0	1000	128	36917.31
2	1000	128	34762.92
4	1000	128	31439.48
6	1000	128	30385.62
8	1000	128	28449.70
10	1000	128	27791.05
12	1000	128	28663.15
14	1000	128	30613.49

the VMD-SG-LSTM model is closer to the real curve. This is because the VMD-SG-LSTM model has a certain fitting effect in predicting some values with large time span, which will make the predicted curve more gentle and closer to the actual value. The prediction performance of the four models is shown in Table 4.

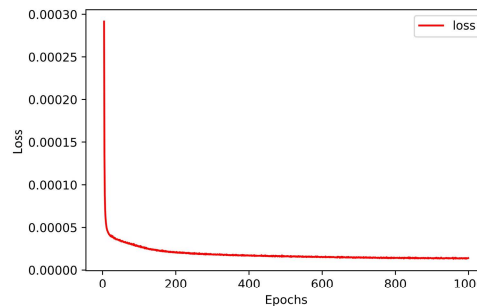


FIGURE 10. Error curve of model training.

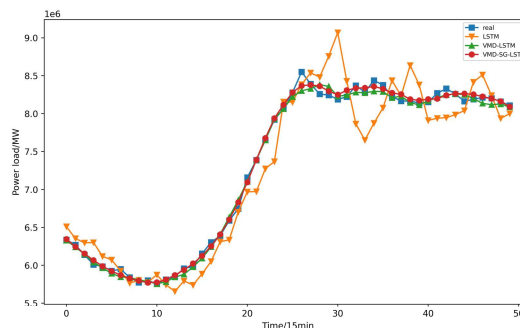


FIGURE 11. Predict the power load for the next hour.

Table 4 records the performance evaluation of the four prediction models in the next one hour. In the prediction and evaluation of the next one hour, the prediction performance of VMD-SG-LSTM model is reduced by 0.057, 0.071 and 0.076 in MASE compared with VMD-Bi-LSTM, VMD-CISSA-LSSVM and VMD-GWO-SVR, and 0.067, 0.084 and 0.087 in MAPE. In summary, the prediction performance of VMD-SG-LSTM model is better than that of the other three models.



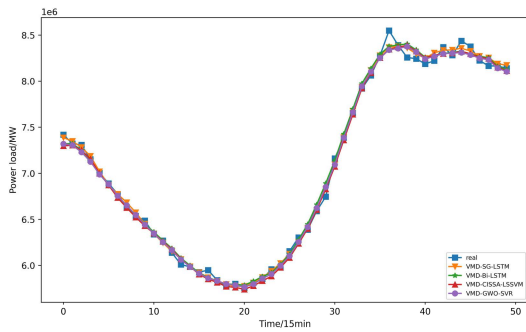


FIGURE 12. Comparison of the latest prediction models.

TABLE 4. Evaluation indexes of power load predicted by each model.

Prediction model	Evaluation indicator	Error
VMD-SG-LSTM	MASE	0.387
	MAPE/%	0.463
VMD-Bi-LSTM	MASE	0.444
	MAPE/%	0.530
VMD-CISSA-LSSVM	MASE	0.458
	MAPE/%	0.547
VMD-GWO-SVR	MASE	0.463
	MAPE/%	0.550

V. CONCLUSION

Aiming at the characteristics of randomness and volatility of power load data, this paper applies data enhancement techniques to the field of power load prediction model based on VMD-SG-LSTM. Specifically, the original data are decomposed into IMFs with different frequencies by VMD data preprocessing technique to improve the predictability of the data. The decomposed sub-functions are reconstructed, and the reconstructed data are smoothed by the S-G filtering algorithm to obtain CTRD. Then the IMF, CTRD and the original data are used as inputs to predict the short-term power load using the LSTM network to obtain the final predicted value. In addition, the prediction performance of the proposed model is compared with that of LSTM, VMD-LSTM and EMD-LSTM models. The main conclusions are as follows:

- (1) LSTM prediction model has the ability of time series feature extraction, which can predict power load, but when predicting some values with large time span, the prediction performance gradually decreases;
- (2) In the VMD-LSTM prediction model, since the VMD algorithm can extract the feature information of the original signal, the prediction performance is better than that of the single LSTM prediction model;
- (3) Compared with the VMD-LSTM model, the VMD-SG-LSTM model proposed in this paper adds the S-G algorithm on the basis of the original prediction model. It can make the peak part closer to the real value curve, and also make the prediction curve more flat. It has

a certain fitting effect in predicting some values with large time span;

- (4) Compared with the other three latest prediction models, the proposed VMD-SG-LSTM model has higher prediction accuracy and stronger prediction performance.

However, the proposed method has certain limitations due to the need to process the current and past power load data in real time in the test set before making predictions. In future work, it is planned to combine some improved decomposition methods with the latest prediction models to further improve the accuracy of short-term power load prediction. In addition, the influence of different input parameters on the prediction model will be carefully studied to further improve the prediction performance of the model.

REFERENCES

- [1] M. E. Günay, "Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: Case of Turkey," *Energy Policy*, vol. 90, pp. 92–101, Mar. 2016, doi: 10.1016/j.enpol.2015.12.019.
- [2] H. Ji, J. Yang, H. Wang, K. Tian, M. O. Okoye, and J. Feng, "Electricity consumption prediction of solid electric thermal storage with a cyber-physical approach," *Energies*, vol. 12, no. 24, pp. 1–18, 2019. [Online]. Available: https://econpapers.repec.org/RePEc:gam:jeners:v:12:y:2019:i:24:p:4744-d:297254
- [3] Y. Iino, M. Murai, D. Murayama, and I. Motoyama, "Physical and JIT model-based hybrid modeling approach for building thermal load prediction," *Electr. Eng. Jpn.*, vol. 185, no. 2, pp. 30–39, 2013, doi: 10.1002/eej.22293.
- [4] A. Yamaguchi, P. W. Sarli, and T. Ishihara, "Extreme load estimation of the wind turbine tower during power production," *Wind Eng.*, vol. 45, no. 1, pp. 93–106, 2021, doi: 10.1177/0309524X19872766.
- [5] M. G. D. Giorgi, P. M. Congedo, and M. Malvoni, "Photovoltaic power forecasting using statistical methods: Impact of weather data," *IET Sci. Meas. Technol.*, vol. 8, no. 3, pp. 90–97, May 2014, doi: 10.1049/iet-smt.2013.0135.
- [6] Y. J. Ma and M. Y. Zhai, "Day-ahead prediction of microgrid electricity demand using a hybrid artificial intelligence model," *Processes*, vol. 7, no. 6, p. 320, 2019, doi: 10.3390/pr7060320.
- [7] J. Shi, Y. L. Shi, J. Tan, L. Zhu, and H. Li, "Research on electricity consumption forecast based on mutual information and random forests algorithm," in *Proc. Int. Conf. Energy Eng. Environ. Protection (EEEP)*, vol. 121. Nanjing, China, 2018, Art. no. 052089, doi: 10.1088/1755-1315/121/5/052089.
- [8] H.-X. Zhao and F. Magouls, "A review on the prediction of building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 16, no. 6, pp. 3586–3592, 2012, doi: 10.1016/j.rser.2012.02.049.
- [9] D. F. Pereira, F. D. C. Lopes, and E. H. Watanabe, "Nonlinear model predictive control for the energy management of fuel cell hybrid electric vehicles in real time," *IEEE Trans. Ind. Electron.*, vol. 68, no. 4, pp. 3213–3223, Apr. 2021, doi: 10.1109/TIE.2020.2979528.
- [10] F. J. Ardakani and M. M. Ardehali, "Long-term electrical energy consumption forecasting for developing and developed economies based on different optimized models and historical data types," *Energy*, vol. 65, pp. 452–461, Feb. 2014, doi: 10.1016/j.energy.2013.12.031.
- [11] H. Bian, Y. Zhong, J. Sun, and F. Shi, "Study on power consumption load forecast based on K-means clustering and FCM-BP model," *Energy Rep.*, vol. 6, pp. 693–700, Dec. 2020, doi: 10.1016/j.egy.2020.11.148.
- [12] Z. Liao, H. Pan, X. Fan, Y. Zhang, and L. Kuang, "Multiple wavelet convolutional neural network for short-term load forecasting," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 9730–9739, Sep. 2021, doi: 10.1109/JIOT.2020.3026733.
- [13] Y. Liu, "Wind power short-term prediction based on LSTM and discrete wavelet transform," *Appl. Sci.*, vol. 9, no. 6, p. 1108, 2019, doi: 10.3390/app9061108.
- [14] Y. Liu, J. Shi, Y. Yang, and W.-J. Lee, "Short-term wind-power prediction based on wavelet transform-support vector machine and statistic-characteristics analysis," *IEEE Trans. Ind. Appl.*, vol. 48, no. 4, pp. 1136–1141, Jul./Aug. 2012, doi: 10.1109/TIA.2012.2199449.

- [15] H. T. Zheng, J. B. Yuan, and L. Chen, "Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation," *Energies*, vol. 10, no. 8, p. 1168, 2017, doi: [10.3390/en10081168](https://doi.org/10.3390/en10081168).
- [16] J. C. Li, S. W. Zhu, Q. Q. Wu, and P. F. Zhang, "A hybrid forecasting model based on EMD-GASVM-RBFNN for power grid investment demand," *Math. Probl. Eng.*, vol. 2018, Sep. 2018, Art. no. 7416037, doi: [10.1155/2018/7416037](https://doi.org/10.1155/2018/7416037).
- [17] W. D. Li, X. Yang, H. Li, and L. L. Su, "Hybrid forecasting approach based on GRNN neural network and SVR machine for electricity demand forecasting," *Energies*, vol. 10, no. 1, p. 44, 2017, doi: [10.3390/en10010044](https://doi.org/10.3390/en10010044).
- [18] M. F. Azam and M. S. Younis, "Multi-horizon electricity load and price forecasting using an interpretable multi-head self-attention and EEMD-based framework," *IEEE Access*, vol. 9, pp. 85918–85932, 2021, doi: [10.1109/ACCESS.2021.3086039](https://doi.org/10.1109/ACCESS.2021.3086039).
- [19] Y. M. Shen, Y. X. Ma, S. M. Deng, C. J. Huang, and P. H. Kuo, "An ensemble model based on deep learning and data preprocessing for short-term electrical load forecasting," *Sustainability*, vol. 13, no. 4, p. 1694, 2021, doi: [10.3390/su13041694](https://doi.org/10.3390/su13041694).
- [20] G. J. Wang, X. H. Wang, Z. X. Wang, C. R. Ma, and Z. X. Song, "A VMD-CISSA-LSSVM based electricity load forecasting model," *Mathematics*, vol. 10, no. 1, p. 28, 2022, doi: [10.3390/math10010028](https://doi.org/10.3390/math10010028).
- [21] S. Bahrami, R.-A. Hooshmand, and M. Parastegari, "Short term electric load forecasting by wavelet transform and grey model improved by PSO (particle swarm optimization) algorithm," *Energy*, vol. 72, pp. 434–442, Aug. 2014, doi: [10.1016/j.energy.2014.05.065](https://doi.org/10.1016/j.energy.2014.05.065).
- [22] Y. Liang, D. Niu, and W.-C. Hong, "Short term load forecasting based on feature extraction and improved general regression neural network model," *Energy*, vol. 166, pp. 653–663, Jan. 2019, doi: [10.1016/j.energy.2018.10.119](https://doi.org/10.1016/j.energy.2018.10.119).
- [23] J. Duan, P. Wang, W. Ma, S. Fang, and Z. Hou, "A novel hybrid model based on nonlinear weighted combination for short-term wind power forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 134, Jan. 2022, Art. no. 107452, doi: [10.1016/j.ijepes.2021.107452](https://doi.org/10.1016/j.ijepes.2021.107452).
- [24] Y. Ruan, G. Wang, H. Meng, and F. Qian, "A hybrid model for power consumption forecasting using VMD-based the long short-term memory neural network," *Frontiers Energy Res.*, vol. 9, p. 917, 2022, doi: [10.3389/fenrg.2021.772508](https://doi.org/10.3389/fenrg.2021.772508).
- [25] S. Shen and J. He, "SGCS: A signal reconstruction method based on Savitzky-Golaysgz filtering and compressed sensing for wavelength modulation spectroscopy," *Opt. Exp.*, vol. 29, no. 22, pp. 35848–35863, 2021, doi: [10.1364/OE.437649](https://doi.org/10.1364/OE.437649).
- [26] A. Huang, R. Shen, W. Di, and H. Han, "A methodology to reconstruct LAI time series data based on generative adversarial network and improved Savitzky-Golay filter," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, Dec. 2021, Art. no. 102633, doi: [10.1016/j.jag.2021.102633](https://doi.org/10.1016/j.jag.2021.102633).
- [27] L. Tang, Z. Zhao, P. Tang, and H. Yang, "SURE-based optimum-length S-G filter to reconstruct NDVI time series iteratively with outliers removal," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 18, no. 2, Mar. 2020, Art. no. 2050001, doi: [10.1142/S0219691320500010](https://doi.org/10.1142/S0219691320500010).
- [28] G. Mao, M. Wang, J. Liu, Z. Wang, K. Wang, Y. Meng, R. Zhong, H. Wang, and Y. Li, "Comprehensive comparison of artificial neural networks and long short-term memory networks for rainfall-runoff simulation," *Phys. Chem. Earth, Parts A/B/C*, vol. 123, Oct. 2021, Art. no. 103026, doi: [10.1016/j.pce.2021.103026](https://doi.org/10.1016/j.pce.2021.103026).
- [29] X. H. Yuan, C. Chen, M. Jiang, and Y. B. Yuan, "Prediction interval of wind power using parameter optimized Beta distribution based LSTM model," *Appl. Soft Comput.*, vol. 82, Sep. 2019, Art. no. 105550, doi: [10.1016/j.asoc.2019.105550](https://doi.org/10.1016/j.asoc.2019.105550).
- [30] R. Li, P. Jiang, H. Yang, and C. Li, "A novel hybrid forecasting scheme for electricity demand time series," *Sustain. CITIES Soc.*, vol. 55, Apr. 2020, Art. no. 102036, doi: [10.1016/j.scs.2020.102036](https://doi.org/10.1016/j.scs.2020.102036).
- [31] *ELIA: Elia Grid Load Data [EB/OL]*. [Online]. Available: <https://www.elia.be/en/grid-data/data-download-page>
- [32] P. Zhang, C. Li, C. Peng, and J. Tian, "Ultra-short-term prediction of wind power based on error following forget gate-based long short-term memory," *Energies*, vol. 13, no. 20, p. 5400, Oct. 2020, doi: [10.3390/en13205400](https://doi.org/10.3390/en13205400).
- [33] J. Tang, J. Zhao, H. Zou, G. Ma, J. Wu, X. Jiang, and H. Zhang, "Bus load forecasting method of power system based on VMD and bi-LSTM," *Sustainability*, vol. 13, no. 19, p. 10526, Sep. 2021, doi: [10.3390/su131910526](https://doi.org/10.3390/su131910526).
- [34] G. Wang, X. Wang, Z. Wang, C. Ma, and Z. Song, "A VMD-CISSA-LSSVM based electricity load forecasting model," *Mathematics*, vol. 10, no. 1, p. 28, 2022, doi: [10.3390/math10010028](https://doi.org/10.3390/math10010028).
- [35] M. Zhou, T. Hu, K. Bian, W. Lai, F. Hu, O. Hamrani, and Z. Zhu, "Short-term electric load forecasting based on variational mode decomposition and grey wolf optimization," *Energies*, vol. 14, no. 16, p. 4890, Aug. 2021, doi: [10.3390/en14164890](https://doi.org/10.3390/en14164890).



**QIU SUN** was born in Jiangsu, China, in 1997. He received the degree from the School of Electrical and Electronic Engineering, Hubei University of Technology, where he is currently pursuing the master's degree. His research interests include control theory and control engineering.



**HUAFENG CAI** was born in Hubei, China, in 1978. He received the master's degree in power electronics and power transmission from the Hubei University of Technology, in 2005. He is currently with the Hubei University of Technology as an Associate Professor of automation research direction for power electronics technology, fault prediction, and optimization.

...