IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 March 2024, accepted 30 March 2024, date of publication 2 April 2024, date of current version 16 April 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3384246*

RESEARCH ARTICLE

# Short Term Power Load Combination Forecasting Method Based on Feature Extraction

**QINGBIN CHEN[1], GENGHUANG YANG[1,2], AND LIQING GENG[1,2]**
[1]School of Automation and Electrical Engineering, Tianjin University of Technology and Education, Tianjin 300222, China
[2]Tianjin Key Laboratory of Information Sensing and Intelligent Control, Tianjin University of Technology and Education, Tianjin 300222, China

Corresponding author: Genghuang Yang (ygh@tute.edu.cn)

This work was supported in part by the Science and Technology Development Fund of Tianjin Higher Education under Grant 2022ZD037, and in part by the Science and Technology Project of Tianjin under Grant 23YDTPJC00320.

**ABSTRACT** Accurate and fast forecasting of short-term load is conducive to the safe and stable operation of the power system, and a short-term power load combination forecasting method based on feature extraction is proposed. Firefly Sparrow Algorithm (FSA) is applied to find the optimal combination of influencing parameters in Variational Mode Decomposition (VMD) to obtain the signal components with the best effect. Since the signal components contain different influence characteristics and timing information, the Maximum Information Coefficient (MIC) is used to screen the features of each signal component, establish the feature matrix, and use the over-zero rate as an index to determine the high and low frequency signal demarcation points. Based on the different characteristics of high and low frequency signals, the Informer model is used to forecast the high frequency signal components, and the LSTM is used to forecast the low frequency signal components. All the forecasting results are reconstructed to obtain the final forecasting value. Taking the Spanish power load data as an example, considering the actual seasonal factors, and experimentally comparing with other forecasting models, the results show that after the feature screening, the errors are significantly reduced, and the decidability coefficient is significantly improved, which verifies the accuracy and universality of the model proposed in this paper.

**INDEX TERMS** Power load forecasting, feature selection, hybrid model, variational mode decomposition, informer.

## I. INTRODUCTION

Power demand is affected by various factors, such as weather, electricity prices and accidents, resulting in random volatility and the corresponding fluctuations in power load. Therefore, power load forecasting has become increasingly important for the power supply plans and the balance between power demand in the power grid [1]. According to the length of forecast time, power load forecasting is categorized into long-term load forecasting (LTLF), medium-term load forecasting (MTLF) and short-term load forecasting (STLF) [2]. Accurate STLF not only makes the system run normally with low energy waste, but also improves power supply reliability [3].

The forecasting methods is challenging in improving the accuracy of power load forecasting. Enhancing model accuracy is particularly crucial accordingly. Due to the time series and nonlinear characteristics of load sequences, numerous scholars have proposed a variety of models by simple or complex methods such as machine learning [4], exponential smoothing method [5], autoregressive integrated moving average model [6], multiple linear regression method [7], Kalman filter algorithm [8], grey prediction theory [9] and support vector machine [10]. Nevertheless, their overall forecasting accuracy has room for improvement, especially by the, challenge of applying large-scale data with universality. With the continuous development and application of deep learning [11] in various fields, recurrent neural network [12] is applied to improve the accuracy of MTLF and LTLF. The long short-term memory neural network [13] is applied to solve the problem of weight disappearance in the back

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

propagation of recurrent neural network so as to improve the forecasting accuracy. The Informer model [14] has effectively reduced the spatial-temporal complexity, memory usage and decoder decoding time of the Transformer model, based on the model proposed by Google. Additionally, it has improved the forecasting accuracy of time series and has a stronger ability to capture the long-range correlation coupling between input and output data. Meanwhile, a variety of combination forecasting methods are widely used in STLF. The method primarily based on Empirical Mode Decomposition (EMD) [15], can decompose the original load sequence. However, the decomposed components are susceptible to model mixing, which can impact the accuracy, Ensemble Empirical Mode Decomposition (EEMD) [16] adds Gaussian white noise prior to decomposition, which can effectively alleviate the mode mixing phenomenon of empirical mode decomposition mode decomposition. However, the method's robustness to measurement noise needs to be strengthened. Variational Mode Decomposition (VMD) [17] decomposes the data into mode functions with different characteristics, effectively avoiding mode mixing. However, the effectiveness of load sequence decomposition is affected by the number of mode components and the penalty parameter. In the field of power load forecasting, there is a lack of scientific evaluation standards for determining these parameters, which are often determined empirically and subjectively, leading to an impact on effect of the load sequence decomposition and the total accuracy. In this regard, parameter optimization through optimization algorithms is of greater significance [18], [19], [20].

This paper proposes a STLF method that utilizes feature selection and hybrid model. In order to extract signal components with feature information from the original load sequence, which has more noise and achieve optimal signal processing. First of all, Firefly Sparrow Algorithm (FSA) is applied to find the optimal combination of influencing parameters in VMD to obtain the signal components with the best effect. Secondly, the adaptive algorithm is employed to determine the optimal zero crossing rate [21], and each signal component is divided into two parts of high frequency and low frequency. Thirdly, the Maximum Information Coefficient (MIC) [22] is used to filter the features of each signal component. Then, the high frequency signal components and their features are used to forecast with the Informer model, while the low frequency signal components and their features are used to forecast with the LSTM model [23]. Finally, the forecasting result of all signal components are reconstructed to obtain the final forecasting value. The accuracy of the method is verified by comparison experiments.

The rest of this article is structured as follows. In Section II, we introduce the basic process of feature extraction. In Section III, we describe the different characteristics of the forecasting algorithms. In Section IV, we show the construction process of the overall model and the evaluation metrics of the forecasting effect. In Section V, we demonstrate the usage process of the overall model, conduct result analysis, and perform comparative experimental verification. Section VI concludes the paper.

## II. FEATURE EXTRACTION METHOD BASED ON FSA-VMD-MIC

The power system load data exhibits large fluctuations and randomness, which are determined by the complexity of the influencing factors. The data may not necessarily reflect the contribution of each influencing factor to the load trend. During the summer months, the use of refrigeration equipment tends to increase as temperatures rise. This increase in usage can lead to a higher demand for power. Although weather variables such as wind speed and humidity do not directly affect power load, accurately identifying the significant factors that do can significantly increase the precision of load forecasting.

In this case, we use VMD to decompose the original power load data, and decompose the original signal into a number of Intrinsic Modal Function (IMF), each of which has different frequency and amplitude characteristics. From the decomposition steps of VMD, it can be seen that before decomposing the signal, it is necessary to set the appropriate number of modes K and the penalty parameter $\alpha$. Different combinations of K and $\alpha$ will have different effects on the decomposition results, and their optimal combinations need to be combined with a specific scenario to determine. In this paper, we propose FSA to optimize the optimal combination of VMD parameters. FSA combines the advantages of Firefly Algorithm (FA) [24] and Sparrow Search Algorithm (SSA) [25], compared with SSA, FSA can effectively improve the global search ability and avoid falling into local optimal solutions. The specific steps taken by FSA to optimize the VMD parameters are shown in Fig. 1.

The specific process for FA and SSA to form an FSA and optimize VMD is as follows,

1) We randomly generate several combinations of [K, $\alpha$] parameters as the initial positions for both the discoverer and follower.

2) After performing the VMD decomposition, the fitness value is calculated and sorted.

3) The discoverer and follower's locations are updated based on the alerts, while the sparrow's location is updated based on the fitness value.

4) FA disturbs the sparrow's location. If the disturbed location is better than the original location, it is adopted. Otherwise, the original sparrow location is used.

5) For each discoverer and follower location, VMD decomposition is performed and the envelope entropy of each signal component is calculated.

6) Randomly select the alerts and update their location, calculating envelope entropy.

7) Determine whether the stop condition is met. If it is satisfied, complete the process and obtain the minimum fitness value. Output the best parameter combinations [K, $\alpha$]. Otherwise, return to step 3.

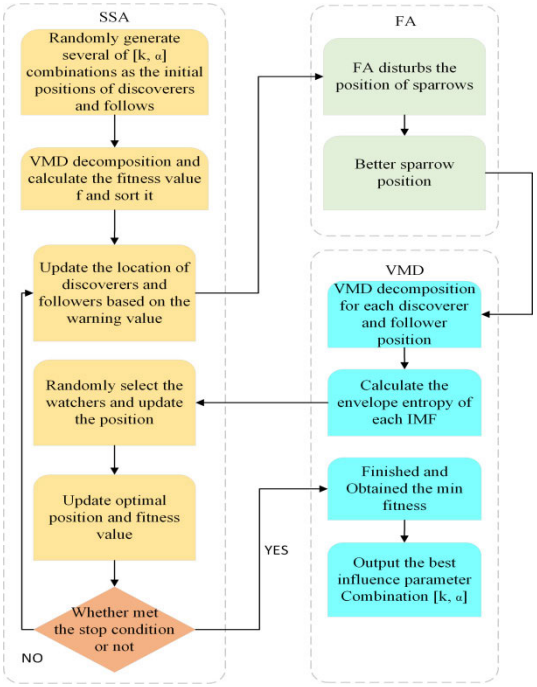



**FIGURE 1.** Flow chart of FSA optimized VMD parameters.

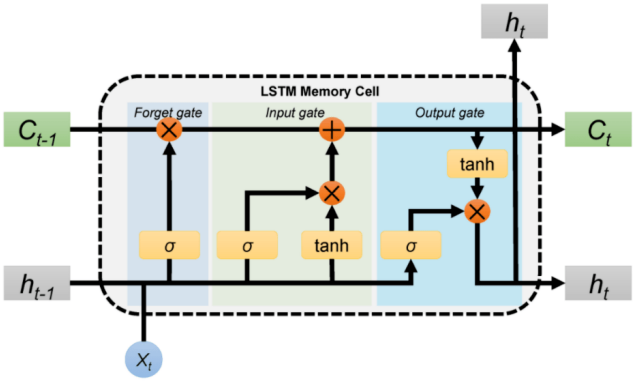



**FIGURE 2.** LSTM basic unit.

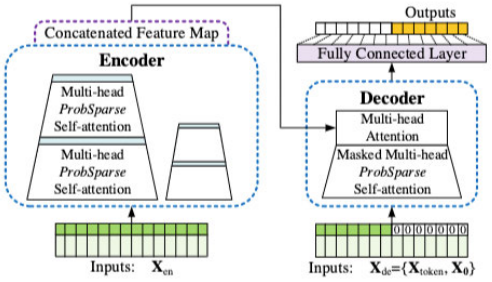



**FIGURE 3.** Informer structure diagram.

In the feature extraction part, the original power load data are decomposed using VMD, and in the decomposition process, the number of modes K and the penalty parameter $\alpha$ in VMD are optimized using FSA, and multiple sub-sequences will be obtained for the original power load data through FSA-VMD, and then MIC is used to compute the relevant influence features of each sub-sequence respectively, and the input matrix of the model is constructed.

MIC is used to evaluate the correlation between two variables. It is assumed that $X$, $Y$ are two different variables in the dataset, where $X = \{x_1, \cdots, x_n\}$, $Y = \{y_1, \cdots, y_n\}$, $n$ is the sample size. The mutual information between $X$ and $Y$ is defined as,

$$I(x; y) = \int p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \mathrm{d}x\mathrm{d}y \tag{1}$$

where $p(x, y)$ is the joint probability density between $X$ and $Y$; $p(x)$ and $p(y)$ denote the marginal probability density of $X$ and $Y$, respectively.

Grids are plotted on the scatterplot of the data consisting of variables $X$ and $Y$ the magnitude of the mutual information between the grids is calculated. The maximum value of mutual information $M(x; y)$ is selected using different gridding criteria and is calculated as,

$$M(x; y) = \max_{a*b<B} \frac{I(x; y)}{\log_2 \min(a, b)} \tag{2}$$

where $a$ and $b$ are the number of meshes divided in the $X$ and $Y$ directions, respectively, and $B$ is the maximum value of the mesh.

## III. FORECASTING METHOD BASED ON LSTM-INFORMER

After the original load sequence is feature filtered, a number of IMF input matrices containing different impact features are obtained. These IMF input matrices have different timing information. In the field of electromagnetism, the over-zero rate is often used as the main feature to classify the signal. In this regard, the over-zero rate is used as an evaluation index, the high-frequency IMF produces a higher over-zero rate due to its fast rate of change, and the low-frequency IMF produces a lower over-zero rate due to its slower rate of change. The forecasting error is calculated by calculating the over-zero rate for each IMF, sorting them from smallest to largest, and using the point separating them as the demarcation point between high-frequency IMFs and low-frequency IMFs. By obtaining the optimal cut-off point between high-frequency IMFs and low-frequency IMFs, those with an over-zero rate lower than the cut-off point are regarded as low-frequency IMFs and are forecasted using the LSTM model, and those with an over-zero rate higher than the cut-off point are regarded as high-frequency IMFs and are forecasted using the Informer model.

The low frequency signal components are typically slow and irregular, with varying period lengths. To address this, the LSTM model is utilized in this paper for forecasting. The LSTM forecasting model is well-suited for processing irregular time series data and adapting to changing trends in low-frequency signals. Its memory unit structure allows it to store previous information and recall it in subsequent calculations, enabling it to capture historical information and

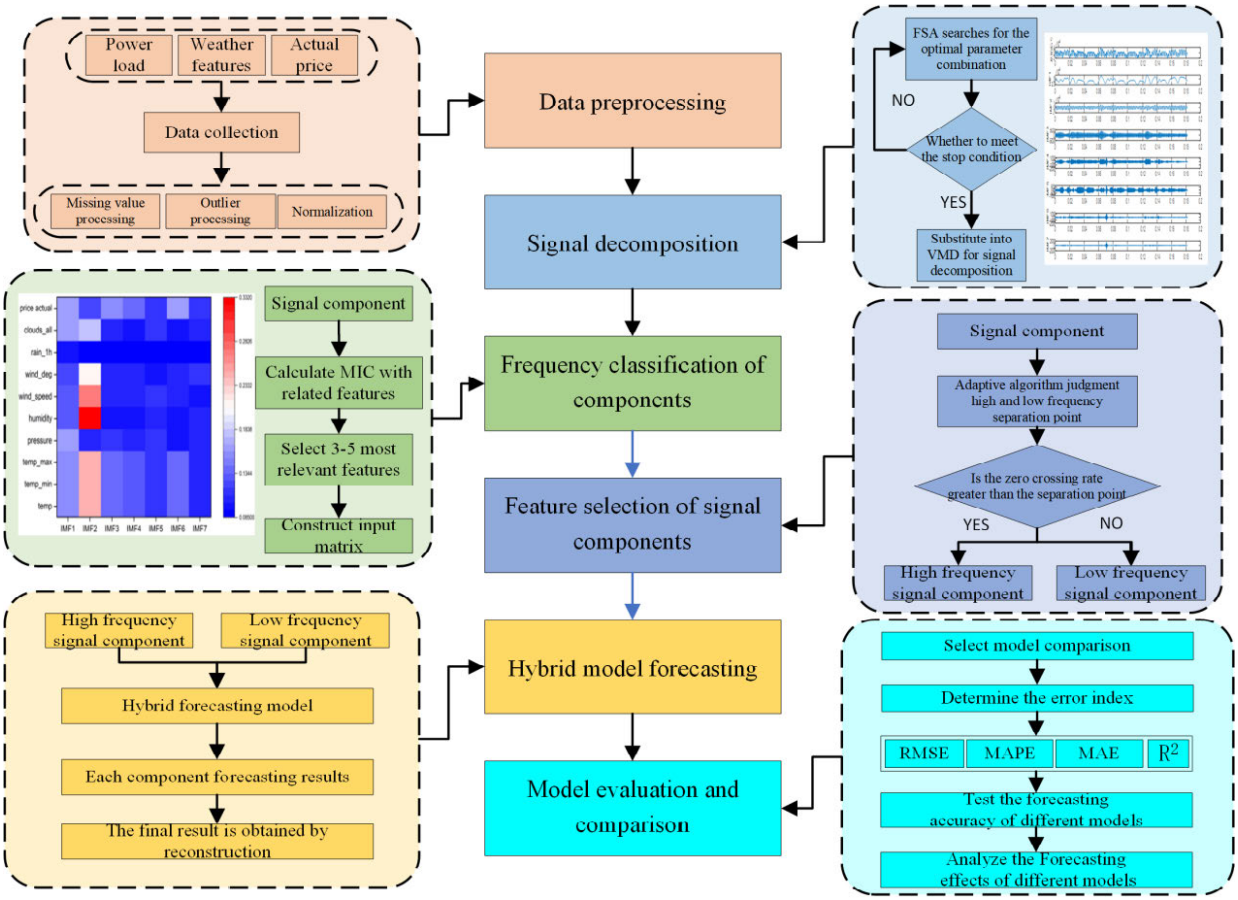



**FIGURE 4.** The overall flow chart of the model.

trends in low frequency signals efficiently. Additionally, the model's gating structure helps to avoid the problem of a long-term dependence on time series data, making it particularly suitable for forecasting low frequency time series components with smooth, low complexity, and obvious periodicity in signal components. The basic unit of LSTM is shown in Fig. 2.

The period of the high frequency signal component is very short, and its amplitude and frequency often vary in response to external factors or events, which is short and transient. The spectrum entropy energy distribution tends to have a concentration of high frequency signal component. In this regard, this paper employs the Informer model for forecasting because it adopts multi-scale convolution operation, time series attention mechanism, and adaptive feature selection mechanism. These features enable the effective extracting and utilizing of the multi-scale features from high-frequency signal components and automatically learn the spatial-temporal dependence of time series data. Furthermore, the Transformer structure in the Informer model can capture the long-term trends and periodic variation features in the high frequency signal components, thereby better adapting to the characteristics of these components. The Informer structure is shown in Fig. 3.

## IV. OVERALL FORECASTING MODEL FRAMEWORK

This paper proposes a hybrid model framework for electricity load forecasting, as shown in Fig. 4.

The specific steps taken to optimize the algorithm are as follows,

1) Data preprocessing: Collect relevant power load and weather data, process missing and outlier values, and normalize the data.

2) Signal decomposition: The optimal parameter combination in VMD is determined using FSA, and the electricity load data is decomposed into signal components using VMD.

3) Frequency classification of components: The coefficient sizes of each signal component are calculated using the MIC. The most relevant features are selected to construct the input matrix.

4) Feature selection of signal components: The adaptive algorithm is used to determine the zero-crossing rate separation point of the high and low frequencies of the signal components. Values above the point are considered as high frequency signal components, while values below it is considered as low frequency signal components.

5) Hybrid model forecasting: Time series forecasting models are applied to each input matrix based on the high and low frequency signal components, and the final forecasting results are obtained through a reconstruction operation.

6) Model evaluation and comparison: Evaluation metrics are selected to perform comparative experiments and analyze the forecasting performance of different models.

In order to assess the forecasting accuracy of the combined feature selection model and the time series forecasting model, this paper utilizes several evaluation metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Coefficient of Determination ($R^2$) as evaluation metrics.

## V. EXAMPLE ANALYSIS

### A. DATA PREPARATION AND PREPROCESSING

The electricity load data used in this paper comes from the Spanish electricity consumption data, which records the electricity load data from January 1, 2015 to December 31, 2018, with a data sampling interval of 1 hour, and its average daily load is shown in Fig. 5. The weather data was obtained from Kaggle open-source data. The dataset is divided into training set, validation set and test set according to the ratio of 8:1:1.

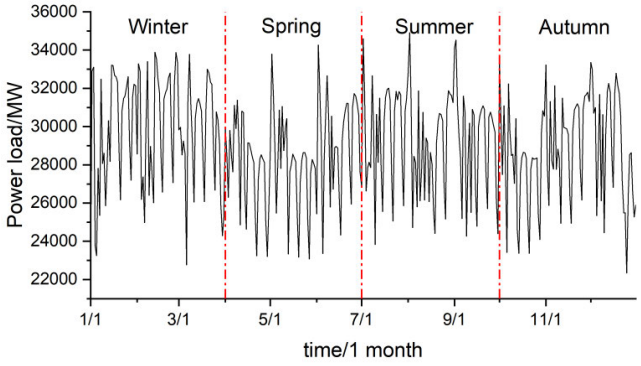



**FIGURE 5.** Average daily load curve of a location in spain.

Electricity load data may have data missing values and outliers during the collection and transmission process, which reduces the forecasting accuracy of the forecasting model, for this reason, a usability interpolation method is used to deal with the outliers and missing values, and the load data is corrected. At the same time, in order to avoid the different magnitude of each eigenvalue, which affects the forecasting accuracy and speeds up the gradient descent. Therefore, the data are normalized.

As shown in Fig. 5, the average daily load curve exhibits significant seasonality. To enhance the accuracy of our forecasting evaluation, we partitioned the sample set into four distinct seasons: spring (April-June), summer (July-September), autumn (October-December), and winter (January-March), by the prevailing weather conditions in the city.

### B. FEATURE SELECTION

Using FSA-VMD to decompose power load can avoid the randomness caused by VMD setting parameters by experience. In order to verify the superiority of FSA for VMD parameter optimization, Particle Swarm Optimization (PSO) and Grey Wolf Optimization (GWO) are used to optimize the parameter K and $\alpha$ of VMD and in a reasonable range, where the value range of K is [2] and [10], and the value range of $\alpha$ is [200,2000]. In the VMD process, the envelope entropy is used as the objective function. Envelope entropy represents the sparse characteristics of the original signal, when there is more noise and less feature information in the component, the value of envelope entropy is larger, and vice versa, the value of envelope entropy is smaller. The specific formula for envelope entropy is as follows,

$$\begin{cases} E_p = -\sum_{j=1}^{N} p_j \lg p_j \\ p_j = a(j) \Big/ \sum_{j=1}^{N} a(j) \\ a(j) = \sqrt{[x(j)]^2 + \{H[x(j)]\}^2} \end{cases} \tag{3}$$

where $E_p$ is the envelope entropy; $p_j$ is the normalized form of $a(j)$; $a(j)$ is the sequence of envelope signals obtained after Hilbert mediation of the signal $x(j)$ $(j = 1, 2, \cdots, N)$; $H$ is the Hilbert transform of the signal.

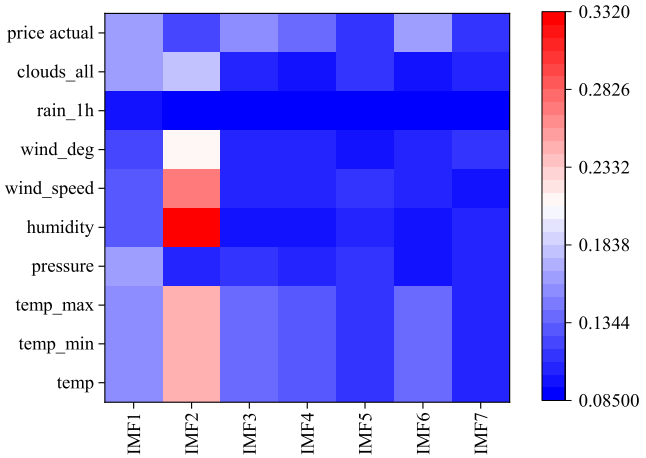



**FIGURE 6.** Heat map of MIC coefficient in spring.

In order to better compare the optimization effect of different optimization algorithms on VMD decomposition, modal loss is adopted as the evaluation index. By reconstructing the signal components after VMD decomposition, the difference between the reconstructed data and the original data is calculated, and the smaller the difference is indicates that the decomposition effect is better. The calculation results are shown in Table 1.

It can be seen from Table 1 that the decomposition loss of FSA-VMD is the smallest, and the optimal parameters of VMD can be determined adaptively to improve the decomposition effect. At the same time, compared to VMD, FSA-VMD will reduce the modal loss generated by decomposition. The optimization results of FSA-VMD decomposition parameters by FSA in each season are shown in Table 2.

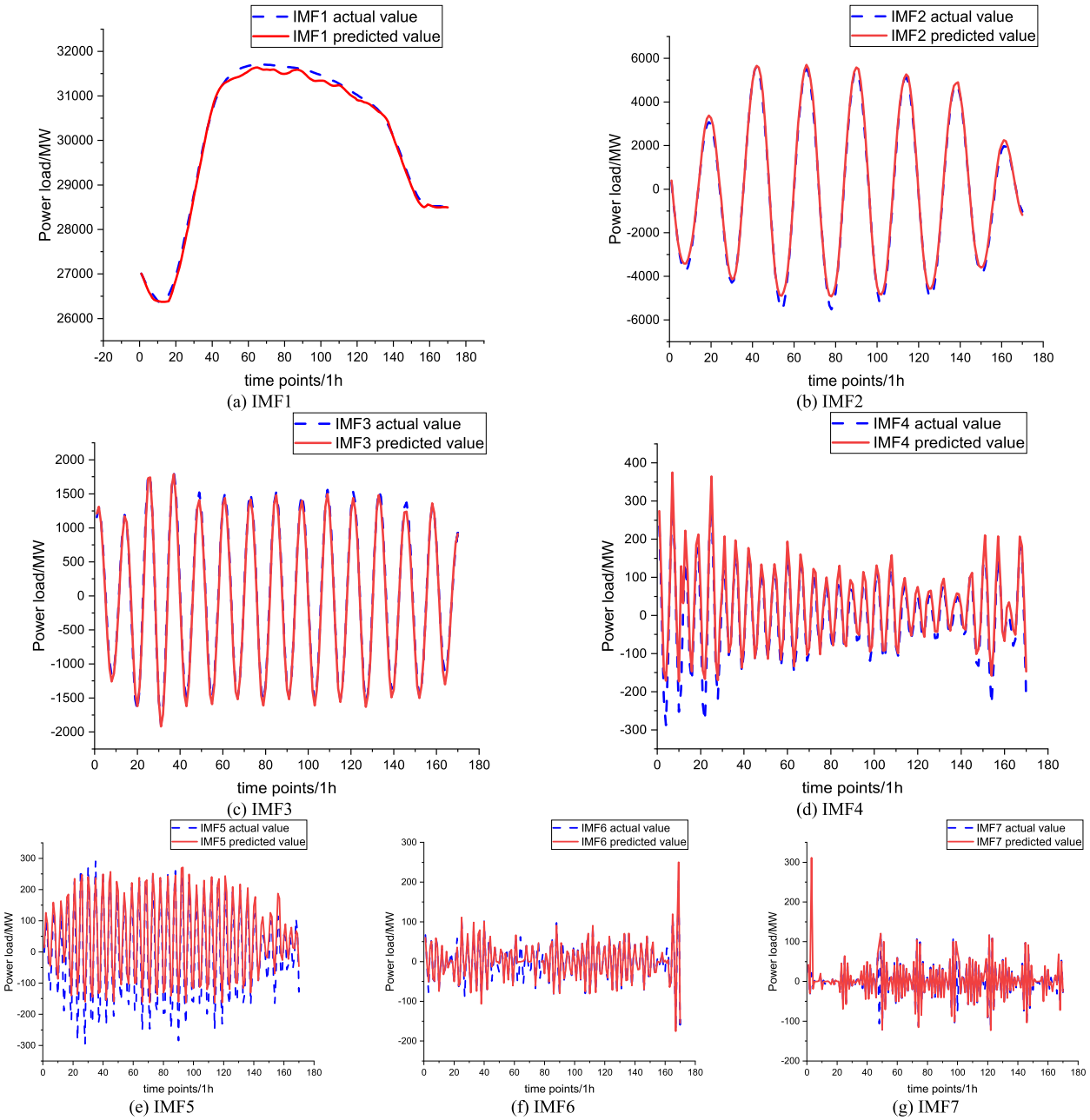



**FIGURE 7.** Forecasting chart of each signal component in spring.

**TABLE 1.** Decomposition results of different algorithms.

| Methods | K | α | Modal loss |
|---|---|---|---|
| VMD | 8 | 2000 | 5207 |
| PSO-VMD | 9 | 1870 | 3576 |
| GWO-VMD | 9 | 1950 | 3061 |
| FSA-VMD | 8 | 1530 | 2604 |

**TABLE 2.** Parameters of the best combination for all seasons.

| Season | K | α |
|---|---|---|
| Spring | 7 | 1374 |
| Summer | 8 | 1175 |
| Autumn | 8 | 1223 |
| Winter | 7 | 1175 |

The optimal [K, $\alpha$] combination parameters for each season are brought into the VMD, and the raw power data are decomposed to produce a number of IMF. In order to improve the forecasting accuracy, this paper adopts the over-zero rate

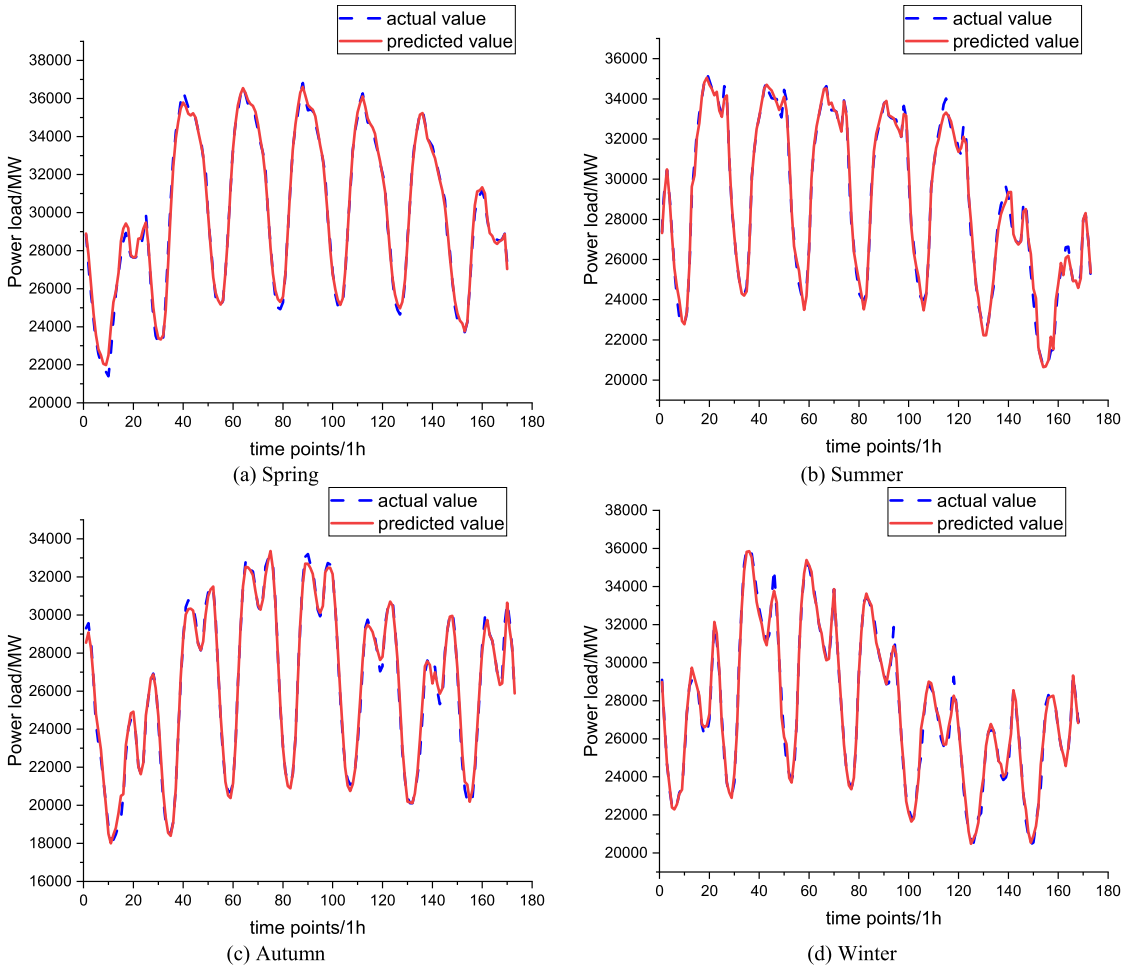



**FIGURE 8.** The forecasting results of model in each season.

as the evaluation index, and sets the over-zero rate higher than 0.01 as the high-frequency IMF, and sets the over-zero rate lower than 0.01 as the low-frequency IMF. The cut-off points for the over-zero rate of the spring subsequence are 0.01, 0.1, 0.2, 0.4, 0.5, 0.7, for the summer subsequence are 0.01, 0.1, 0.2, 0.4, 0.5, 0.55, 0.7, for the fall subsequence are 0.01, 0.1, 0.2, 0.4, 0.5, 0.7, 0.8, and for the winter subsequence The cut-off points of over-zero rate are 0.01, 0.1, 0.2, 0.4, 0.5, 0.7. For different seasons, the high and low-frequency components are divided with these cut-off points, respectively, and the forecasting and calculation of errors are made, and through the error analysis, the optimal cut-off point for the high and low-frequency components in each season is 0.01. The over-zero rate of the various signal components for spring, summer, autumn, and winter seasons are shown in Table 3.

Meanwhile, in order to better characterize the impact feature information contained in each IMF after decomposition, this paper adopts MIC to filter the features and enhance the forecasting capability. The heat map visualization of the MIC coefficient matrix for the spring season is shown in Fig. 6. The darker the red color indicates a larger MIC coefficient and a higher degree of correlation, while the darker the blue color indicates a smaller MIC coefficient, i.e., a higher degree of correlation is lower.

**TABLE 3.** Zero crossing rate of each signal component in four seasons.

| Season | Spring | Summer | Autumn | Winter |
|---|---|---|---|---|
| IMF1 | 0 | 0 | 0 | 0 |
| IMF2 | 0.0833 | 0.0838 | 0.0833 | 0.0838 |
| IMF3 | 0.1667 | 0.1667 | 0.1667 | 0.1667 |
| IMF4 | 0.3310 | 0.3293 | 0.3265 | 0.3301 |
| IMF5 | 0.4199 | 0.4144 | 0.4162 | 0.4208 |
| IMF6 | 0.5746 | 0.5023 | 0.5756 | 0.5755 |
| IMF7 | 0.8997 | 0.5879 | 0.7405 | 0.7262 |
| IMF8 | | 0.9017 | 0.9040 | |

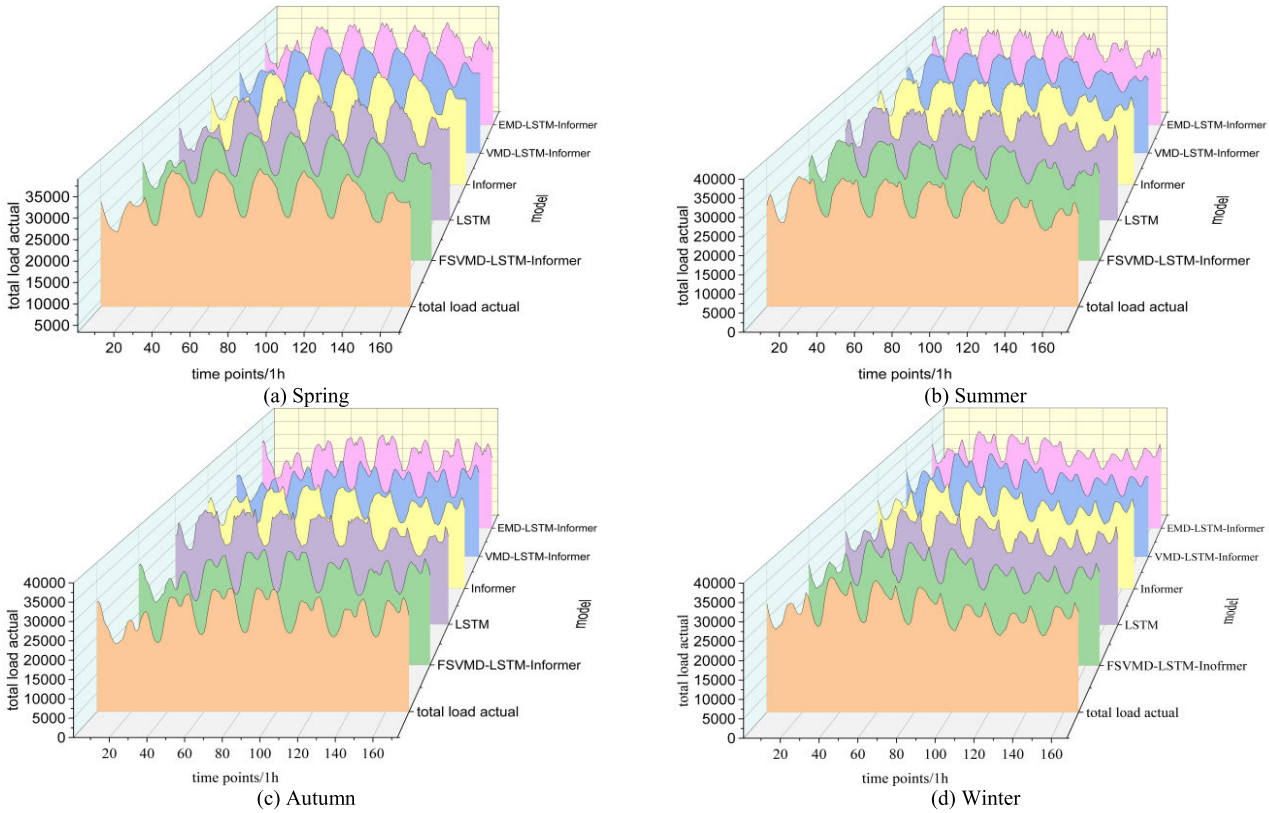



**FIGURE 9.** Forecasting results of each model.

From Fig. 6, the influence characteristics of each signal component are different, combined with the MIC coefficients of each IMF corresponding to the influence factors, this paper selects the three to five features with the highest correlation to construct the feature matrix. In order to facilitate the analysis, each influencing factor is replaced by a number. Among them, ① stands for Pressure, ② stands for Clouds_all, ③ stands for Price actual, ④ stands for Temp, ⑤ stands for Temp_min, ⑥ stands for Temp_max, ⑦ stands for Humidity, ⑧ stands for Wind_deg, and ⑨ stands for Wind_speed. The results of the feature screening in different seasons are shown in Table 4.

**TABLE 4.** Feature screening results.

| Season | Spring | Summer | Autumn | Winter |
|---|---|---|---|---|
| IMF1 | ①②③ | ①③④⑤⑥ | ②⑤⑦ | ④⑤⑥⑦⑨ |
| IMF2 | ④⑤⑥⑦⑨ | ④⑤⑥ | ③④⑤ | ③④⑤⑥ |
| IMF3 | ③④⑤⑥ | ①②③⑧ | ③④⑤ | ①③⑨ |
| IMF4 | ③④⑤⑥ | ①③⑨ | ①④⑤ | ①②③ |
| IMF5 | ③④⑤⑥ | ①③④⑤⑥ | ③⑥⑦ | ④⑤⑥ |
| IMF6 | ③④⑤⑥ | ①②③ | ②④⑤⑥⑧ | ④⑤⑥ |
| IMF7 | ①③⑧ | ①③④⑤⑥ | ③④⑥⑧ | ④⑤⑥ |
| IMF8 | | ④⑤⑥⑧ | ③④⑥ | |

### C. POWER LOAD FORECASTING RESULTS BASED ON HYBRID MODEL

The high and low frequency signal components exhibit different feature characteristics. Therefore, this paper applies the LSTM model to forecast the signal component input matrix with zero crossing rates less than 0.01, and the Informer model to forecast the input matrix with zero crossing rates greater than 0.01. The forecasting outcomes of each signal component in spring are presented in Fig. 7.

Based on Fig. 7, it is observed that the forecasting results for IMF1, IMF2 and IMF3 signal components are relatively accurate, and exhibit a smooth trend over time. However, the forecasting results for IMF4 and IMF5 signal components are less accurate, showing large fluctuations and significant differences between each peak. The forecasting of IMF6 and IMF7 signal components, which belong to high frequency components is challenging, but the amplitude is relatively small, and their forecasting errors have a limited impact. By reconstructing the forecasting results of each signal component in Fig 7, we can obtain the spring power load forecasting results. In this paper, power load forecasting is divided into four seasons. We construct

**TABLE 5.** Forecasting error statistics of each model for four seasons.

| Season | Models | MAE | RMSE | MAPE | $R^2$ |
|---|---|---|---|---|---|
| Spring | FSVMD-LSTM-Informer | **309.65** | **410.03** | **1.03** | **0.9905** |
| | VMD-LSTM-Informer | 394.13 | 509.93 | 1.35 | 0.9854 |
| | EMD-LSTM-Informer | 764.83 | 963.06 | 2.57 | 0.9480 |
| | LSTM | 716.34 | 900.16 | 2.35 | 0.9760 |
| | Informer | 439.55 | 698.57 | 1.74 | 0.9546 |
| Summer | FSVMD-LSTM-Informer | **315.49** | **475.26** | **1.11** | **0.9927** |
| | VMD-LSTM-Informer | 402.14 | 509.79 | 1.43 | 0.9848 |
| | EMD-LSTM-Informer | 728.32 | 925.48 | 2.55 | 0.9499 |
| | LSTM | 644.78 | 813.67 | 2.25 | 0.9613 |
| | Informer | 431.03 | 627.05 | 1.65 | 0.9793 |
| Autumn | FSVMD-LSTM-Informer | **289.53** | **380.27** | **1.14** | **0.9920** |
| | VMD-LSTM-Informer | 472.29 | 573.55 | 1.63 | 0.9832 |
| | EMD-LSTM-Informer | 675.70 | 878.92 | 2.54 | 0.9570 |
| | LSTM | 829.38 | 1028.72 | 2.94 | 0.9381 |
| | Informer | 459.92 | 678.30 | 1.72 | 0.9740 |
| Winter | FSVMD-LSTM-Informer | **344.91** | **456.32** | **1.21** | **0.9917** |
| | VMD-LSTM-Informer | 425.96 | 555.12 | 1.64 | 0.9719 |
| | EMD-LSTM-Informer | 705.32 | 884.30 | 2.60 | 0.9489 |
| | LSTM | 620.69 | 806.85 | 2.29 | 0.9575 |
| | Informer | 439.51 | 651.48 | 1.78 | 0.9766 |

a hybrid model for the spring season and forecast the remaining three seasons using the same methodology. The forecasting results of the four seasons are shown in Fig. 8. Meanwhile, in the low-frequency components, the eigenmatrices of IMF1 are Pressure, Clouds_all, Price actual. in the high-frequency components, the eigenmatrices of IMF2 are Temp, Temp_min, Temp_max, Humidity, Wind_speed; the eigenmatrices of IMF3 are Price actual, Temp, Temp_min, Temp_max; the feature matrices of IMF4 are Price actual, Temp, Temp_min, Temp_max; the feature matrices of IMF5 are Price actual, Temp, Temp_min, Temp_max; the feature matrices of IMF6 are Price actual, Temp, Temp_min, Temp_max; and the characteristic matrix of IMF7 is Pressure, Price actual, Wind_deg.

Fig. 8. illustrates that the load sequence exhibits poor regularity due to rapid temperature changes during summer and winter, making forecasting more challenging. However, the models proposed in this paper demonstrate the ability to accurately track the actual load trend, resulting in high overall forecasting accuracy.

### D. ANALYSIS OF FORECASTING RESULTS

To evaluate the accuracy of the proposed model, this paper compares it with two single models (LSTM and Informer) and two combined models (EMD-LSTM-Informer and VMD-LSTM-Informer) using several evaluation indices, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Determinable Coefficient ($R^2$). The mean values of load in spring, summer, autumn and winter are 28384.47, 29278.07, 28814.73517 and 29790.36 MW, respectively, and the standard deviations are 4251.97, 4569.60, 4662.41 and 4887.93 MW, respectively. The forecasting error statistics of each model in four seasons are shown in Table 5, with the minimum errors highlighted in bold. The forecasting results of each model in four seasons are shown in Fig. 9.

The overall model of this paper is integrated by five methods, FSA, VMD, MIC, LSTM and Informer, and the running process between different steps can be disassembled to be carried out without generating a large running load, and at the same time, in the field of power load forecasting, there is

enough time and space to make forecasting in order to obtain more accurate results.

Fig. 9. demonstrates that the forecasting curves of the model proposed in this paper accurately track the actual load trends across all seasons, particularly in areas with strong load fluctuations. When combined with the results presented in Table 5, it is clear that the MAE, RMSE, MAPE and $R^2$ evaluation indices of the hybrid model are significantly superior to those of the single models in each season. The results of the comparison between the proposed feature selection model and the Informer model indicate a significant improvement in the performance of the former. Specifically, the MAPE of the proposed model is reduced by 0.71%, 0.5%, 0.58% and 0.57% in each season, respectively. These results indicate that the feature selection model can effectively extract the relevant feature information from the load time series, resulting in reduced forecasting errors and improved overall model fitting. In addition, the proposed model outperformed other combined models in all four seasons, as demonstrated by the significantly better performance in the four evaluation indexes. This superior performance can be attributed to the use of VMD, which effectively addresses the modal confounding phenomenon in EMD. The combined model based on VMD decomposition generally achieved lower forecasting errors and better model fitting in all seasons. In comparison to the VMD decomposition combination forecasting model with an artificial empirical setting of parameter combinations [K, $\alpha$], the model proposed in this paper uses the FSA to search for the best parameter combinations [K, $\alpha$] in VMD. As a result, the proposed model achieved a reduction in MAPE by 0.32%, 0.32%, 0.49% and 0.43% in each season, respectively. These results suggest that the preferred strategy of FSA-VMD for parameter combinations [K, $\alpha$] improves the decomposition quality, resulting in better forecasting accuracy of the load in each season.

## VI. CONCLUSION

To meet the increasing need for precise forecasting of power system load, this paper proposes a short-term power load forecasting method that utilizes feature selection and hybrid models. The advantages of the model in this paper are as follows,

1) The model in this paper can significantly improve the forecasting accuracy by combining the characteristics of each of the five methods, namely FSA, VMD, MIC, LSTM and Informer, and giving full play to their respective advantages.

2) The forecasting accuracy is further improved by optimizing the number of modes and K and the penalty parameter $\alpha$ in VMD by FSA and screening the influencing factors by MIC.

3) By determining the optimal over-zero rate decomposition point and MIC to construct the feature matrix of each signal component, the forecasting accuracy can be effectively improved by selecting different models for different input matrices with respect to their characteristics.

The research content of this paper still has some shortcomings, such as not considering the weekend and holidays. In the follow-up work, we can further consider the holidays and the user demand side to further improve the accuracy of power load forecasting.

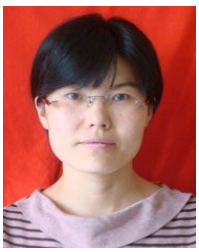
## REFERENCES

[1] I. S. Jahan, V. Snasel, and S. Misak, "Intelligent systems for power load forecasting: A study review," *Energies*, vol. 13, no. 22, p. 6105, Nov. 2020.
[2] Q. Ge, C. Guo, H. Jiang, Z. Lu, G. Yao, J. Zhang, and Q. Hua, "Industrial power load forecasting method based on reinforcement learning and PSO-LSSVM," *IEEE Trans. Cybern.*, vol. 52, no. 2, pp. 1112–1124, Feb. 2022.
[3] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.
[4] M. S. Daley, L. M. Bonacci, D. H. Gever, K. Diaz, and J. B. Bolkhovsky, "Cluster analysis for the separation of auditory scenes," *IEEE Access*, vol. 9, pp. 130959–130967, 2021.
[5] B. Yang, H. Ren, T. Zuo, and Z. Liu, "A stream function smoothing method for the design of MRI gradient coils on non-developable surfaces," *Sensors*, vol. 23, no. 18, p. 7912, Sep. 2023.
[6] W. Xiang and L. Zhou, "Planning research on electrically coupled integrated energy system based on ARIMA-LSTM model," *Appl. Math. Nonlinear Sci.*, vol. 9, no. 1, Jan. 2024.
[7] Y. Zhao and A. Altalbe, "Analysis of the causes of the influence of the industrial economy on the social economy based on multiple linear regression equation," *Appl. Math. Nonlinear Sci.*, vol. 275, no. 3, pp. 916–924, Dec. 2021.
[8] Y. Tian, Z. Lian, P. Wang, M. Wang, Z. Yue, and H. Chai, "Application of a long short-term memory neural network algorithm fused with Kalman filter in UWB indoor positioning," *Sci. Rep.*, vol. 14, no. 1, p. 1925, Jan. 2024.
[9] F. Meng, J. Gu, L. Wang, Z. Qin, M. Gao, J. Chen, and X. Li, "A quantitative model based on grey theory for sea surface temperature prediction," *Frontiers Environ. Sci.*, vol. 10, Nov. 2022, Art. no. 1014856.
[10] Z. Xu, A. Che, and H. Zhou, "Seismic landslide susceptibility assessment using principal component analysis and support vector machine," *Sci. Rep.*, vol. 14, no. 1, p. 3734, Feb. 2024.
[11] Z. Li, Y. Li, Y. Liu, P. Wang, R. Lu, and H. B. Gooi, "Deep learning based densely connected network for load forecasting," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 2829–2840, Jul. 2021.
[12] J. Lin, J. Ma, J. Zhu, and Y. Cui, "Short-term load forecasting based on LSTM networks considering attention mechanism," *Int. J. Electr. Power Energy Syst.*, vol. 137, May 2022, Art. no. 107818.
[13] J. Qin, Y. Zhang, S. Fan, X. Hu, Y. Huang, Z. Lu, and Y. Liu, "Multi-task short-term reactive and active load forecasting method based on attention-LSTM model," *Int. J. Electr. Power Energy Syst.*, vol. 135, Feb. 2022, Art. no. 107517.
[14] M. Gong, Y. Zhao, J. Sun, C. Han, G. Sun, and B. Yan, "Load forecasting of district heating system based on informer," *Energy*, vol. 253, Aug. 2022, Art. no. 124179.
[15] D. Li, Y. Tan, Y. Zhang, S. Miao, and S. He, "Probabilistic forecasting method for mid-term hourly load time series based on an improved temporal fusion transformer model," *Int. J. Electr. Power Energy Syst.*, vol. 146, Mar. 2023, Art. no. 108743.
[16] 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.
[17] P. Jia, H. Zhang, X. Liu, and X. Gong, "Short-term photovoltaic power forecasting based on VMD and ISSA-GRU," *IEEE Access*, vol. 9, pp. 105939–105950, 2021.
[18] F. A. Hashim and A. G. Hussien, "Snake optimizer: A novel meta-heuristic optimization algorithm," *Knowl.-Based Syst.*, vol. 242, Apr. 2022, Art. no. 108320.
[19] Y. Li, B. Tang, S. Jiao, and Q. Su, "Snake optimization-based variable-step multiscale single threshold slope entropy for complexity analysis of signals," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 6505313.
[20] Y. Li, B. Tang, S. Jiao, and Y. Zhou, "Optimized multivariate multiscale slope entropy for nonlinear dynamic analysis of mechanical signals," *Chaos, Solitons Fractals*, vol. 179, Feb. 2024, Art. no. 114436.

[21] M. Eni, I. Dinstein, M. Ilan, I. Menashe, G. Meiri, and Y. Zigel, "Estimating autism severity in young children from speech signals using a deep neural network," *IEEE Access*, vol. 8, pp. 139489–139500, 2020.
[22] X. Tang, J. Wang, J. Lu, G. Liu, and J. Chen, "Improving bearing fault diagnosis using maximum information coefficient based feature selection," *Appl. Sci.*, vol. 8, no. 11, p. 2143, Nov. 2018.
[23] P. Suebsombut, A. Sekhari, P. Sureephong, A. Belhi, and A. Bouras, "Field data forecasting using LSTM and bi-LSTM approaches," *Appl. Sci.*, vol. 11, no. 24, p. 11820, Dec. 2021.
[24] W. Zhang, L. Gu, Y. Shi, X. Luo, and H. Zhou, "A hybrid SVR with the firefly algorithm enhanced by a logarithmic spiral for electric load forecasting," *Frontiers Energy Res.*, vol. 10, Sep. 2022, Art. no. 977854.
[25] W. Sulandari, M. H. Lee, and P. C. Rodrigues, "Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks," *Energy*, vol. 190, Jan. 2020, Art. no. 116408.

**QINGBIN CHEN** received the B.S. degree in automation from Tianjin University of Technology and Education, in 2022. His main research interests include time series prediction and integrated energy systems.

**GENGHUANG YANG** received the D.S. degree in industrial and agricultural electrification and automation from China Agricultural University, in 2007. He is currently a Professor with Tianjin University of Technology and Education, the Dean of the School of Automation and Electrical Engineering, and the Director of Tianjin Key Laboratory of Information Sensing and Intelligent Control. His main research interests include power system information processing, intelligent control, and load forecasting.



**LIQING GENG** is currently an Associate Professor with Tianjin University of Technology and Education. Her main research interests include signal detection and processing and artificial intelligence.