

## Research Article

# Short-Term Load Forecasting Based on Frequency Domain Decomposition and Deep Learning

Qian Zhang <sup>1,2</sup>, Yuan Ma <sup>1</sup>, Guoli Li,<sup>2</sup> Jinhui Ma,<sup>3</sup> and Jinjin Ding <sup>4</sup>

<sup>1</sup>School of Electrical Engineering and Automation, Anhui University, Hefei, China

<sup>2</sup>Collaborative Innovation Centre of Industrial Energy-Saving and Power Quality Control, Anhui University, Hefei, China

<sup>3</sup>State Grid Anhui Electric Power Company, Hefei, China

<sup>4</sup>Anhui Electric Power Research Institute, Hefei, China

Correspondence should be addressed to Yuan Ma; 664328543@qq.com

Received 27 August 2019; Revised 22 December 2019; Accepted 4 February 2020; Published 25 February 2020

Academic Editor: Ioannis T. Christou

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

In this paper, we focus on the accuracy improvement of short-term load forecasting, which is useful in the reasonable planning and stable operation of the system in advance. For this purpose, a short-term load forecasting model based on frequency domain decomposition and deep learning is proposed. The original load data are decomposed into four parts as the daily and weekly periodic components and the low- and high-frequency components. Long short-term memory (LSTM) neural network is applied in the forecasting for the daily periodic, weekly periodic, and low-frequency components. The combination of isolation forest (iForest) and Mallat with the LSTM method is constructed in forecasting the high-frequency part. Finally, the four parts of the forecasting results are added together. The actual load data of a Chinese city are researched. Compared with the forecasting results of empirical mode decomposition- (EMD-) LSTM, LSTM, and recurrent neural network (RNN) methods, the proposed method can effectively improve the accuracy and reduce the degree of dispersion of forecasting and actual values.

## 1. Introduction

In order to meet the needs of rapid social development, the power system has gradually turned into a self-healing, large-scale renewable energy access, economic, and efficient smart grid. Load forecasting becomes the basis of smart grid planning and operation [1–3]. Short-term load forecasting is important for scheduling plan arrangement, unit combination optimization, and so on [4, 5].

At present, various load forecasting methods have been studied, such as the time series method [6], neural network method [7–11], support vector machine algorithm [12–14], local-weighted linear regression method [15], and RF method [16, 17]. Among them, neural network method is widely used in load forecasting research. In [11], the LSTM neural network method is used in 48 nonresidential consumers' energy consumption data in China. The

problem of trapped in local minimum and disappeared gradient in the traditional neural network method during training is effectively solved by the LSTM neural network, which is beneficial to improving the accuracy of load forecasting.

In recent years, methods for decomposing and reforecasting load sequences have become hot topics, such as EMD [18, 19], sparse decomposition [20], local mean decomposition [21], and frequency domain decomposition [22, 23]. Among them, the empirical mode decomposition method has strong adaptability, but there are problems such as modal mixing, end effects, and overenvelope. Sparse decomposition method is widely used in signal processing, but this method uses a double dictionary set decomposition algorithm, leading to the large calculation amount. Compared with empirical mode decomposition, local mean decomposition has less end effects and less iterations, but its

oversmoothness has some influence on forecasting accuracy. Frequency domain decomposition method is introduced based on the periodicity of the load sequence. Using this method, the periodicity of the load is mined and extracted from the perspective of the frequency domain. An appropriate forecasting method is established according to the law of load component. However, the forecasting accuracy of the high-frequency component still did not achieve the ideal forecasting precision.

In this paper, the field data of a Chinese city are decomposed based on the frequency domain decomposition algorithm to obtain four parts as the daily and weekly periodic components and the low- and the high-frequency components. The LSTM method is used to train and predict highly regular daily and weekly periodic components and smooth low-frequency components. The high-frequency component periodicity is not obvious, and the fluctuation is relatively large. For the purpose of cleaning abnormal data and removing high-frequency burrs, iForest algorithm is used for the data cleaning of high-frequency component training samples. Mallat algorithm is also used to decompose cleaned data, and then the high-frequency signal and low-frequency signal are obtained. Low-frequency signal is used as training samples, combined with the LSTM algorithm to forecasting the original high-frequency component. Finally, the forecasting results of the four components are reconstructed according to the time points. The overall forecasting method is shown in Figure 1.

## 2. Frequency Domain Decomposition

**2.1. Description of Decomposition.** After the four parts are obtained as the daily periodic, the weekly periodic, the low-frequency and the high-frequency components, different forecasting methods will be adopted according to the characteristics of different component sequences.

The load sequence is random, and after the frequency domain decomposition, the partial load sequence shows certain regularity. In this way, the forecasting accuracy will be improved.

**2.2. Frequency Domain Decomposition Algorithm.** Other frequency domain decomposition algorithms (EMD and VMD) are generally decomposed according to scale functions or filters. The frequency domain decomposition algorithms used in this paper are based on the preset angular frequency decomposition and then used the characteristics of periodic changes in the load to reconstruct the sequence and decompose.

Power load is a time series with a strong periodical characteristic. Fourier transform is used to decompose the load time series  $P(t)$  as follows [24]:

$$P(t) = a_0 + \sum_{i=1}^{N-1} (a_i \cos w_i t + b_i \sin w_i t), \quad (1)$$

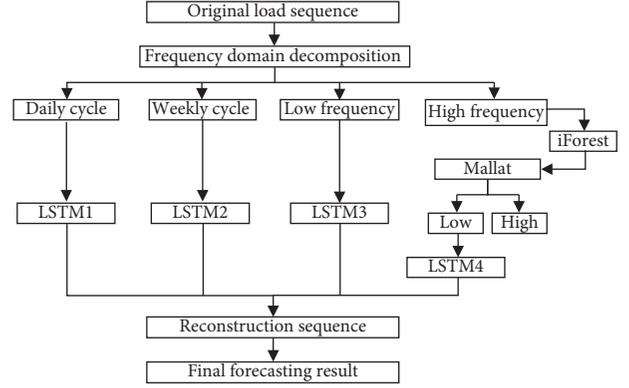


FIGURE 1: Overall forecasting model.

where  $N$  is the number of original load data and  $a_0$  is a direct current component. In this way,  $P(t)$  is decomposed into components in which angular frequency is  $w_i = 2\pi \times i/N$ , ( $i = 1, 2, \dots, N - 1$ ). Taking advantage of the periodicity of load variation, according to the size of angular frequency  $w_i$  and coefficients  $a_i$  and  $b_i$ ,  $P(t)$  can be reconstructed as follows:

$$P(t) = a_0 + D(t) + W(t) + L(t) + H(t), \quad (2)$$

where  $a_0 + D(t)$  is the daily periodic component and  $W(t)$  is the weekly periodic component; the two above periodic components vary according to a fixed cycle.  $L(t)$  and  $H(t)$  are low- and high-frequency component, respectively.

The reconstruction method of decomposition results is introduced by modular operation; use  $\text{mod}(x, y)$  to represent the remainder of  $x$  divided by  $y$ . In this paper, 96 points of power load daily sampling are taken as an example.

- (1) The period of  $a_0 + D(t)$  is 96 sampling time, which is a load component that varies within a period of one day, and its angular frequency set is

$$\Omega_{\text{day}} = \{w_0\} \cup \left\{ w_i \mid \text{mod}\left(96, \frac{2\pi}{w_i}\right) = 0 \right\}. \quad (3)$$

- (2) The period of  $D(t)$  is  $7 \times 96$  sampling time, which is a load component that varies with a period of one week, and its angular frequency set is

$$\Omega_{\text{week}} = \left\{ w_i \mid \text{mod}\left(7 \times 96, \frac{2\pi}{w_i}\right) = 0 \cap \text{mod}\left(96, \frac{2\pi}{w_i}\right) \neq 0 \right\}. \quad (4)$$

- (3) The rest parts are divided into  $L(t)$  and  $H(t)$  after deducting  $a_0 + D(t)$  and  $W(t)$ .  $L(t)$  reflects the influence of slow-change-related factors such as meteorological factors on load, and  $H(t)$  mainly reflects the randomness of load change. The angular

frequency sets of the two are shown in the following formula:

$$\begin{cases} \Omega_{\text{low}} = \left\{ w_i \mid \frac{2\pi}{w_i} \geq 96 \cap w_i \notin \Omega_{\text{day}} \cup \Omega_{\text{week}} \right\}, \\ \Omega_{\text{high}} = \left\{ w_i \mid \frac{2\pi}{w_i} < 96 \cap w_i \notin \Omega_{\text{day}} \cup \Omega_{\text{week}} \right\}. \end{cases} \quad (5)$$

Taking 28 days and 96 sampling points per day as an example, the set of all possible values of frequency subscripts  $i$  of  $a_0 + D(t)$ ,  $D(t)$ ,  $L(t)$ , and  $H(t)$  components is as follows:

$$\begin{cases} \{i_{a_0+D(t)} \mid \text{mod}(i, 28) = 0, \quad i = 0, 1, \dots, N-1\} \cup \{0\}, \\ \{i_{W(t)} \mid \text{mod}(i, 4) = 0 \cap \text{mod}(i, 28) \neq 0, \quad i = 0, 1, \dots, N-1\}, \\ \{i_{a_0+D(t)} \mid \text{mod}(i, 4) \neq 0, \quad i = 1, \dots, 28\}, \\ \{i_{a_0+D(t)} \mid \text{mod}(i, 4) \neq 0, \quad 28 < i < N\}, \end{cases} \quad (6)$$

where the angular frequency equals to 0 which corresponds to the direct component.

Discrete Fourier transform and an inverse discrete Fourier transform are listed as the following formulas:

$$\begin{aligned} X(w_k) &= \sum_{n=0}^{N-1} x(n) e^{-j \left(\frac{2\pi}{N}\right) nk}, \\ x(n) &= \frac{1}{N} \sum_{k=0}^{N-1} X(w_k) e^{j \left(\frac{2\pi}{N}\right) nk}. \end{aligned} \quad (7)$$

According to the Fourier transform decomposition coefficients  $a_i$  and  $b_i$  and spectrum  $X(w_i)$ , the following relationship is obtained:

$$X(w_i) = N(a_i - jb_i). \quad (8)$$

Therefore, after the discrete Fourier transform load sequence, the coefficients  $a_i$  and  $b_i$  can be obtained from the spectrum values.

The inverse Fourier transform is combined with Euler's formula  $e^{j\theta} = \cos\theta + j\sin\theta$ , and the decomposed sequence is obtained as

$$\begin{aligned} X(w_k) e^{j(2\pi nk)/N} &= N(a_k \cos w_k + b_k \sin w_k) \\ &\quad + jN(a_k \sin w_k - b_k \cos w_k). \end{aligned} \quad (9)$$

Considering the length of the article, we only list the process of calculating  $a_0 + D(t)$  sequence as follows:

- (1)  $P(t)$  is processed by discrete Fourier transform, and  $X(w_i)$ ,  $w_i = 2\pi \times i/N$ , ( $i = 1, 2, \dots, N-1$ ) is obtained.
- (2) For all  $w_i \notin \Omega_{\text{day}}$ , let  $X(w_i) = 0$ , and a new spectrum sequence  $X'(w_i)$  is obtained, as shown in the following formula:

$$X'(w_i) = \begin{cases} X(w_i), & w_i \in \Omega_{\text{high}}, \\ 0, & \text{other values.} \end{cases} \quad (10)$$

$X'(w_i)$  is inverted by Fourier transform as follows:

$$\begin{aligned} \text{IDFT}[X'(w_i)] &= \sum_{w_i \in \Omega} (a_i \cos w_i t + b_i \sin w_i t) \\ &\quad + j \sum_{w_i \in \Omega} (a_i \cos w_i t - b_i \sin w_i t). \end{aligned} \quad (11)$$

After  $X'(w_i)$  is inverted by Fourier transform,  $a_0 + D(t)$  is obtained as the real part of the sequence  $\text{IDFT}[X'(w_i)]$ . In addition, the other component sequences can also be obtained according to the aforementioned method.

### 3. Forecasting of the Daily Periodic, Weekly Periodic, and Low-Frequency Components

**3.1. Description of the Daily Periodic, Weekly Periodic, and Low-Frequency Component Forecasting.** At present, LSTM has been researched in the field of load forecasting and has achieved good forecasting accuracy [25]. The regularity of the daily periodic component and the periodic component is very obvious, and the low-frequency component is smooth and flat. LSTM neural network is used to forecasting the daily periodic, weekly periodic, and low-frequency components.

**3.2. LSTM Neural Network.** LSTM architecture was proposed by Hochreiter and Schmidhuber and further improved by Gers et al. [26]. The computing node of the LSTM is composed of the input gate, output gate, forgetting gate, and a cell. LSTM controls the drop or increase information through gates and what new information is added to the input gate control; the output gate is used to control how much of the current cell state is filtered out. The output value of the forgetting gate is between [0, 1]: 0 means that all information is discarded, and 1 means that all information is retained. Figure 2 is the compute node of LSTM.

Cell is used to record the current status, which can be defined using the following equation:

$$\begin{cases} i_t = g(W_{xi}x_t + W_{hi}h_{t-1} + b_i), \\ f_t = g(W_{xf}x_t + W_{hf}h_{t-1} + b_f), \\ o_t = g(W_{xo}x_t + W_{ho}h_{t-1} + b_o). \end{cases} \quad (12)$$

Input transformation is defined as

$$c_{\text{-in}_t} = \tan h(W_{xc}x_t + W_{hc}h_{t-1} + b_{c_{\text{-in}}}). \quad (13)$$

The status is updated as

$$\begin{cases} c_t = f_t \cdot c_{t-1} + i_t \cdot c_{\text{-in}_t}, \\ h_t = o_t \cdot \tan h(c_t), \end{cases} \quad (14)$$

where  $x_t$  and  $h_t$  are the input and the output of the computing node,  $i_t$  is the output of the input gate,  $f_t$  is the output of the forgetting gate,  $c_t$  is the cell unit state, and  $o_t$  is the

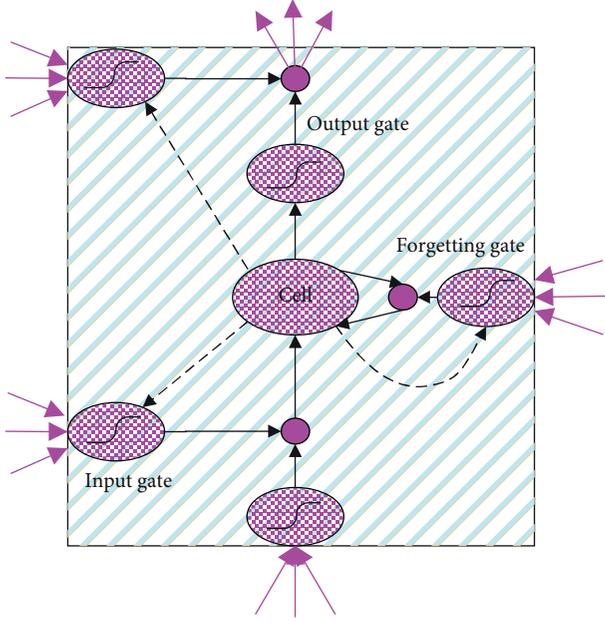


FIGURE 2: LSTM compute node.

output of the output gate,  $t$  is the current time, and  $W$  and  $b$  are the parameter matrices.

In formula (14), it can be seen that not only  $o_t$  works but  $h_t$  is a special feature of the LSTM network, and the problem of low training efficiency of parameters is solved.

#### 4. Forecasting of the High-Frequency Component

**4.1. Description of the High-Frequency Component Forecasting.** The fluctuation of the high-frequency component is gentler than the original load sequence but still exists. The sequence after the recombination of the high-frequency component accounts for a large proportion of the original load sequence. Therefore, the high-frequency component forecasting results directly determine the quality of the load forecasting model proposed in this paper.

The high-frequency component reflects the randomness of the load fluctuation. During the process of power load data measurement, the abnormal values are unavoidable because of power failure, machine breakdown, poor communication, and so on. It makes an obstacle to precise load forecasting. For the possibility of abnormal data in the measured load data, the iForest algorithm is used to clean the training samples.

Next, for the purpose of removing burrs in the high-frequency component, Mallat algorithm is used to decompose the cleaned high-frequency component sequence. The low-frequency and high-frequency signals are obtained. Among them, the low-frequency signal is an approximate signal which can reflect the trend and characteristics of the high-frequency component sequence. Therefore, the reconstructed low-frequency signal sequence is used to forecasting the original high-frequency component in combination with the LSTM neural network.

**4.2. iForest Algorithm.** iForest algorithm is a fast anomaly detection method based on ensemble, which has linear time complexity and high accuracy [27]. iForest algorithm is based on two characteristics of abnormal data: (i) abnormal data only account for a small amount and (ii) abnormal data eigenvalues are very different from normal data. iForest is composed of a large number of binary trees. A single binary tree is called isolation tree (iTree). The iTree building process is a completely random process. The implementation steps are as follows [28]:

- (1) Suppose the training data set is  $X$ , the current tree height is  $h$ , and the height limit of the tree is  $l$
- (2)  $X$  is placed in the root node, a dimension  $q$  is randomly selected in  $X$ , and value  $p$  is randomly selected between the maximum and minimum values on  $q$
- (3) Samples in  $X$  that are larger than  $p$  on  $q$  are placed in the right child node, and other samples are placed in the left child node
- (4) Then, repeat (2) and (3) until there are more identical samples or only one sample in each child node or the height of the tree reaches  $l$

The average path length of point  $x$  to be judged on each tree is counted as  $E(h(x))$ . The path length of  $x$  is counted as  $h(x)$ , which is the number of edges experienced by the root node to the external node.

The formula for calculating the anomaly score of each data to be measured is

$$s(x, n) = 2^{-E(h(x))/c(n)}, \quad (15)$$

where  $c(n) = 2H(n-1) - (2(n-1)/n)$  is the average path length of the binary search tree for the purpose of the result normalization.  $H(k)$  can be estimated by the formula  $H(k) = \ln(k) + \xi$ , where  $\xi$  is Euler's constant.

According to formula (15), it is shown that when the score is closer to 1, the probability that it is an abnormal point is higher. When the scores are smaller than 0.5, then the basic data can be determined as normal. When all the scores are near 0.5, then the data do not contain obvious abnormal samples.

**4.3. Mallat Algorithm.** Mallat algorithm, which is a kind of fast discrete orthogonal wavelet transform [29], is used to decompose and reconstruct high-frequency component sequences.

$$\begin{cases} c_j(k) = \langle X, \varphi_{jk}(t) \rangle, \\ d_j(k) = \langle X, \psi_{jk}(t) \rangle, \end{cases} \quad (16)$$

where  $c_j(k)$  is an approximate signal,  $d_j(k)$  is the detail signal,  $j$  and  $k$  are integers,  $\langle *, * \rangle$  is the inner product, and  $\varphi_{jk}(t)$  and  $\psi_{jk}(t)$  are the scale functions.  $\varphi_{jk}(t)$  is obtained by the mother wavelet  $\varphi(t)$  by translation and expansion:

$$\begin{aligned} \varphi_{jk}(t) &= 2^{-j/2} \varphi(2^{-j}t - k), \\ a_{j+1} &= H_f a_j, \quad d_{j+1} = G_f d_j, \quad j = 0, 1, 2, \dots, N, \end{aligned} \quad (17)$$

where  $a_j$  and  $d_j$  are sequences waiting for decomposition according to the decomposition algorithm.  $H_f$  and  $G_f$  are low-pass and high-pass filters, respectively.  $\varphi_j(t)$  is a low-pass filter. The signal is decomposed into an approximate signal of a large time scale and sets the detail signal at different small scales by wavelet transform.

Mallat algorithm [30, 31] is used to decompose the original high-frequency component sequence. Considering that Daubechies wavelet has compact support and regularity, the information energy of the nonstationary random load sequence can be relatively concentrated after decomposition. The approximate low-frequency signal and the high-frequency detail signal are obtained. The approximate signal can reflect the trend and characteristics of the original high-frequency component sequence, while the detail signal reflects the dynamic factors such as disturbance. The decomposed sequence can be reconstructed by Mallat algorithm according to the following formula:

$$a_{j+1} = a_j H_f^* + d_j G_f^*, \quad j = 0, 1, 2, \dots, N, \quad (18)$$

where  $H_j^*$  and  $G_j^*$  are the dual operators of  $H_f$  and  $G_f$ , respectively. In order to keep the length of the decomposition reconstruction sequence consistent, the quadratic interpolation method is used by the Mallat reconstruction algorithm via adding zero between each adjacent sequence of the input sequence.

## 5. Case Study

The training and test results of the model are strongly related to the size of the selected data set. However, the quality of the results is also strongly related to the quality of the selected data. The load data are inherently unstable. After the frequency domain decomposition, part of the load data shows certain regularity. It is more conducive to the training and prediction of the model, while reducing the number of data sets. From June 1, 2017, to July 5, 2017, the sampling time was 15 minutes, and a total of 3,360 sampling points were researched. The first 2,688-time sampling points in the sample are selected for the training, and the final 672-time sampling points are used for the test. The training samples are decomposed by EMD and the frequency domain decomposition algorithm, and the decomposition results are shown in Figure 3 and 4, respectively.

From Figure 4, we can see that the law of the daily and weekly periodic components is obvious, the low-frequency component is a smooth curve, and the high-frequency component has strong randomness. The frequency domain decomposition method is based on the day of the week and the regularity of these factors to reconstruct the daily and weekly periodic components, and the residual components are reconstructed based on the randomness of meteorological factors. Compared with Figures 3 and 4 shows that (i) the number of components decomposed by the frequency domain decomposition method is less, which can reduce the calculation scale; (ii) the number of high-frequency components decomposed by EMD is more, which is not

conducive to prediction; and (iii) as a whole, the frequency domain decomposition method is more regular.

Implementation details of several main algorithms in this paper are as follows:

- (1) LSTM is provided by the python-based keras package. The training of the model is optimized by adjusting parameters, such as Layers, Units, and Dropout. After testing, the main parameters are set as follows: Layers = 2 (the number of LSTM stacked layers) and Units = 50 (the number of hidden neurons). To prevent overfitting, Dropout is set as 0.5, batch\_size = 16 (if the value is too small, it is easy to not converge. If the value is too large, it is easy to fall into the local optimal solution).
- (2) The “emsemble.IsolationForest” module provided by the sklearn package in Python is used for the iForest algorithm. The 2,688 high-frequency component training samples are cleaned with anomaly points. After cleaning, the number of remaining samples is 2,553. The main parameters are set as follows: n\_estimators = 100 and contamination = 0.05.
- (3) Db 4 wavelet is selected as the wavelet function in Mallat algorithm [32].

According to the description in Section 3, the prediction results, which consist of daily periodic, weekly periodic, and low-frequency components, are obtained.

There are 2,553 sampling points in the training samples of the high-frequency component after deburring using iForest. Figure 5 shows the result of the high-frequency component is decomposed by Mallat. The forecasting results of the four parts are reconstructed according to the time point. Figure 6 demonstrates the comparisons of HFDA-LSTM, EMD-LSTM, LSTM, and RNN. Mean absolute percentage error (MAPE) is used to evaluate the quality of the model, and root mean square error (RMSE) is used to reflect the precision of the forecasting. They are defined in formula (19). In Table 1, the error statistics of the overall forecasting results are listed.

$$\left\{ \begin{array}{l} \text{MAPE} = \frac{\sum_{i=1}^n (|y_i^* - y_i| * 100 / y_i)}{n}, \\ \text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i^* - y_i)^2}{n}}, \end{array} \right. \quad (19)$$

where  $y_i^*$  is the forecasting value,  $y_i$  is the actual value, and  $n$  is the number of samples.

In order to present the proposed method in graphs and tables, the proposed method is defined as the hybrid frequency domain analysis- (HFDA-) LSTM method. In order to highlight the advantages of the method proposed in this paper, the prediction results of HFDA-LSTM are compared with those of EMD-LSTM, single LSTM network, and RNN, respectively.

In the amplification area in Figure 6, it can be seen that the forecasting error of the single LSTM method and the

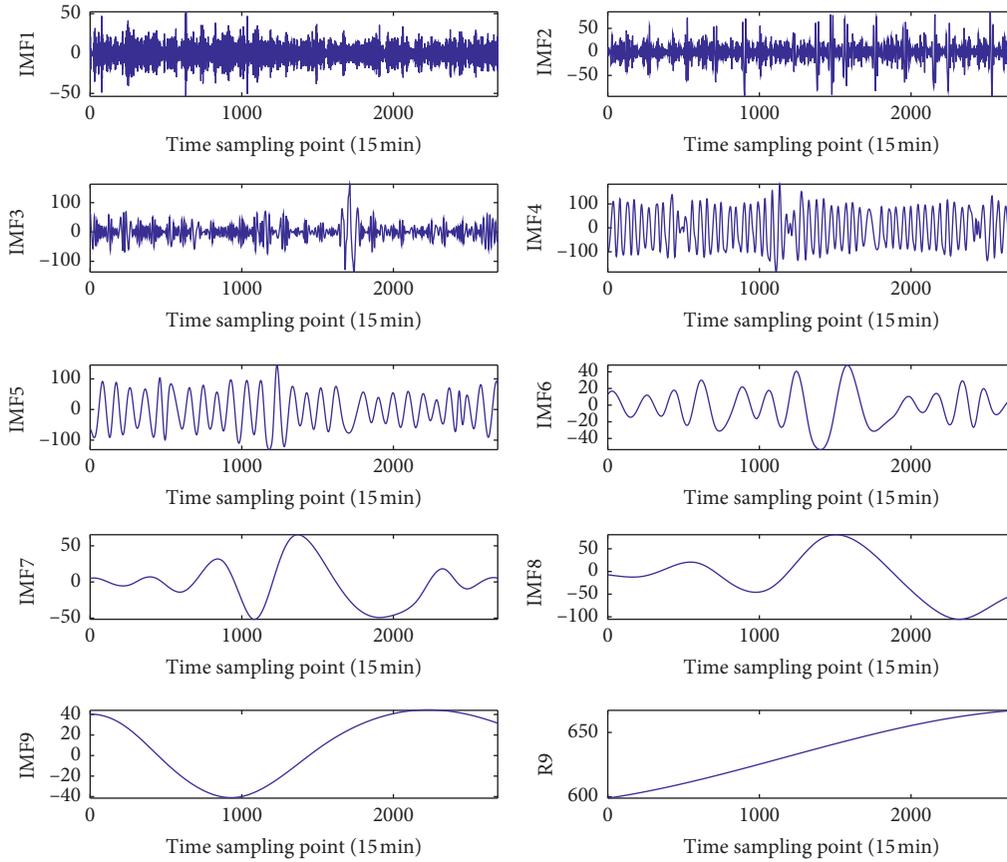


FIGURE 3: Results of EMD decomposition.

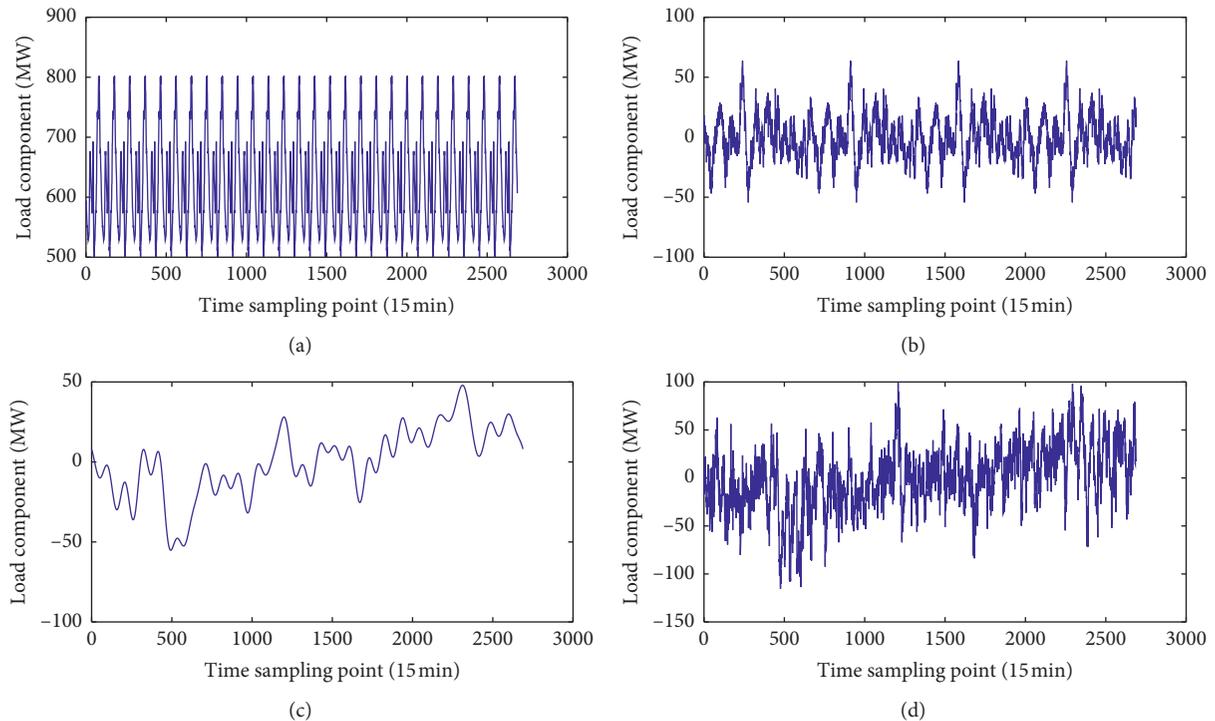


FIGURE 4: Results of frequency domain decomposition. (a) Daily periodic component. (b) Weekly periodic component. (c) Low-frequency component. (d) High-frequency component.

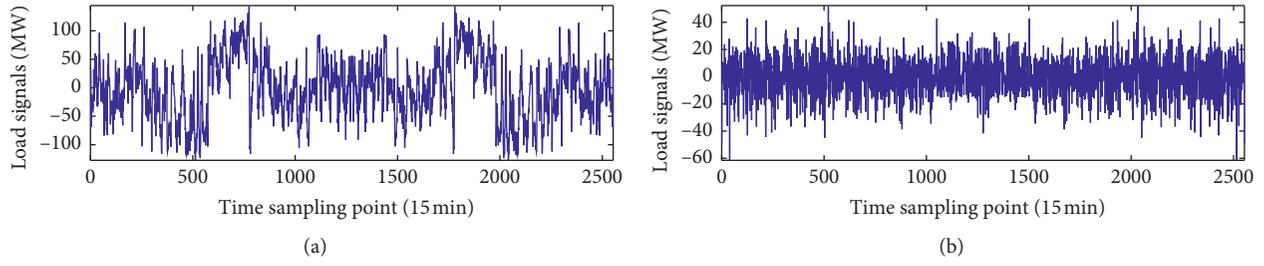


FIGURE 5: Comparison of high-frequency components after Mallat decomposition. (a) Sublow frequency. (b) Subhigh frequency.

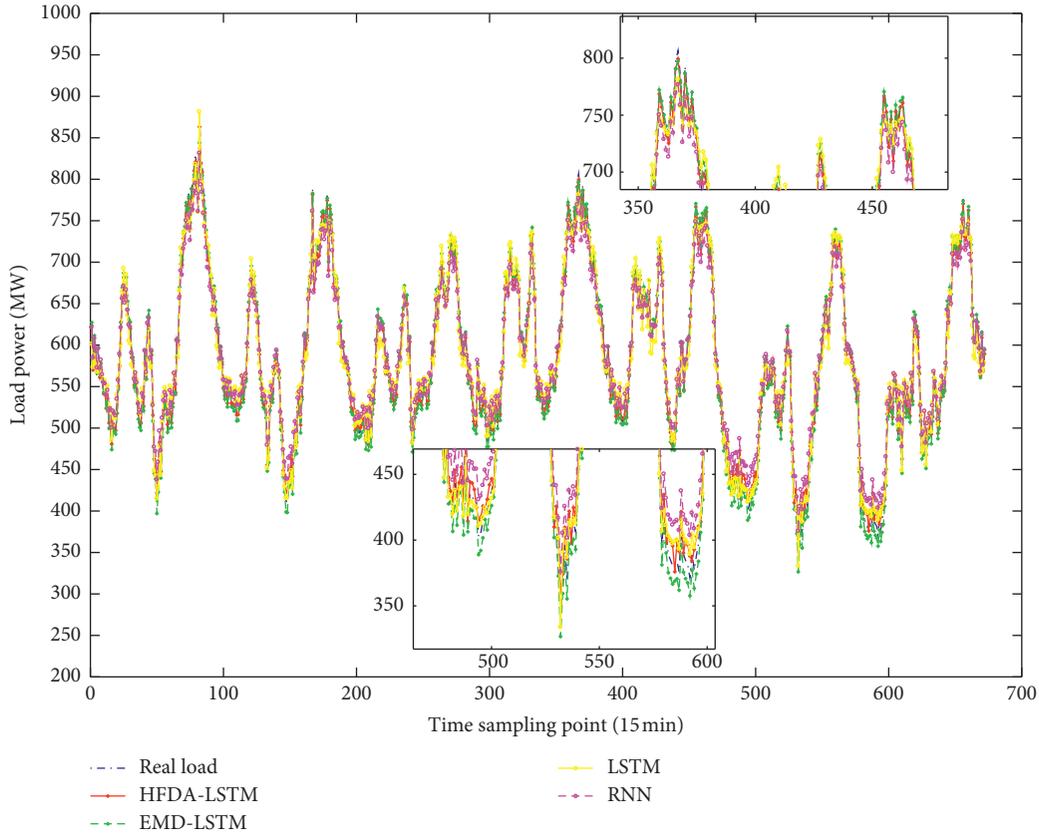


FIGURE 6: The forecasting error of four methods.

TABLE 1: The forecasting error calculation results of four methods.

Time sampling point (15 min)	HFDA-LSTM		EMD-LSTM		LSTM		RNN	
	MAPE (%)	RMSE (MW)	MAPE (%)	RMSE (MW)	MAPE (%)	RMSE (MW)	MAPE (%)	RMSE (MW)
1-96	0.95	7.38	2.23	14.93	2.36	21.36	3.01	22.23
97-192	0.92	6.27	2.09	13.72	2.55	22.31	3.24	23.18
193-288	0.82	5.68	2.26	14.02	2.13	20.06	2.56	19.73
289-384	0.80	5.88	2.24	15.12	2.40	21.47	2.85	22.97
385-480	0.88	6.08	2.06	13.59	2.11	19.98	2.78	21.56
481-576	1.55	8.47	1.77	10.99	3.77	29.66	3.22	21.37
577-672	1.35	7.81	2.43	13.87	3.56	28.30	3.54	24.89
1-672	1.04	6.87	2.15	13.81	2.70	23.31	3.03	22.33

RNN method in the peak load part is relatively large, while the method proposed in this paper is slightly better than the EMD-LSTM method, and the prediction result is relatively

ideal. In Table 1, from the point of view of the average value of time sampling points, the method proposed in this paper has the best prediction accuracy, MAPE is 0.62%, RMSE is

7.91 MW, which is 1.11% higher than that of the EMD-LSTM method, and RMSE is 6.94 MW higher than that of the other three methods. This indicates that the short-term load prediction method proposed in this paper has more accurate prediction results, and the discrete range of predicted value and actual value lower degrees.

## 6. Conclusions

In this paper, a short-term load forecasting model based on frequency domain decomposition and deep learning is proposed. The frequency domain decomposition algorithm is used to decompose the load data, and the four components of the daily periodic, the weekly periodic, the low-frequency and the high-frequency are obtained. The characteristics of different components are analysed and use different forecasting methods in a targeted manner.

LSTM algorithm is used to forecasting the daily periodic, weekly periodic, and low-frequency components. The combination of iForest and Mallat with the LSTM method is constructed in forecasting the high-frequency component. The forecasting results of the four components are added together and compared with the prediction results of EMD-LSTM, single LSTM network, and RNN, and the prediction accuracy of the proposed method is higher.

The contributions of this paper are as follows: (i) combining the frequency domain decomposition algorithm with the deep learning LSTM network for short-term load forecasting. (ii) For the high-frequency component with large fluctuations, iForest algorithm is used to clean the abnormal data of training samples. Then, Mallat algorithm is used to decompose the cleaned samples, the low-frequency signal and the high-frequency signal are obtained, and the low-frequency signal is selected as the training sample combined with the LSTM neural network to forecasting the original high-frequency component.

Next, we consider whether the transfer learning method can be added to the model to improve the sparse decomposition ability to further improve the accuracy of the forecasting model proposed in this paper.

## Data Availability

The data used to support the results of this study can be provided by the corresponding authors.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by research project of National Key Research and Development Program of China (2016YFB0900405), State Grid Corporation of China (521205160021), and the Doctoral research foundation of Anhui University under grant J01001929.

## References

- [1] P. Li, D. Xu, Z. Y. Zhou, W. J. Lee, and B. Zhao, "Stochastic optimal operation of microgrid based on chaotic binary particle swarm optimization," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 66–73, 2016.
- [2] M. Zidar, N. D. Hatziaargyriou, D. Škrlec, P. S. Capuder, and Georgilakis, "Review of energy storage allocation in power distribution networks: applications, methods and future research," *IET Generation, Transmission & Distribution*, vol. 10, no. 3, pp. 645–652, 2016.
- [3] J. W. Mi, L. B. Fan, X. C. Duan, and Y. Y. Qiu, "Short-term power load forecasting method based on improved exponential smoothing grey model," *Mathematical Problems in Engineering*, vol. 2018, Article ID 3894723, 11 pages, 2018.
- [4] Z. S. Zhang and W. J. Gong, "Short-term load forecasting model based on quantum Elman neural networks," *Mathematical Problems in Engineering*, vol. 2016, Article ID 7910971, 8 pages, 2016.
- [5] G. Cerne, D. Dovzan, and I. Skrjanc, "Short-term load forecasting by separating daily profiles and using a single Fuzzy model across the entire domain," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 9, pp. 7406–7415, 2018.
- [6] M. Espinoza, C. Joye, R. Belmans, and B. DeMoor, "Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series," *IEEE Transactions on Power Systems*, vol. 20, no. 3, pp. 1622–1630, 2005.
- [7] B. Li, J. Zhang, Y. He, and Y. Wang, "Short-term load-forecasting method based on wavelet decomposition with second-order gray neural network model combined with ADF test," *IEEE Access*, vol. 5, pp. 16324–16331, 2017.
- [8] S. B. Zheng, Q. W. Zhong, L. L. Peng, and X. D. Chai, "A simple method of residential electricity load forecasting by improved Bayesian neural networks," *Mathematical Problems in Engineering*, vol. 2018, Article ID 4276176, 16 pages, 2018.
- [9] C. Guan, P. B. Luh, L. D. Michel, Y. Wang, and P. B. Friedland, "Very short-term load forecasting: wavelet neural networks with data pre-filtering," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 30–41, 2013.
- [10] S. Li, P. Wang, and L. Goel, "A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1788–1798, 2016.
- [11] R. Jiao, T. Zhang, Y. Jiang, and H. He, "Short-term non-residential load forecasting based on multiple sequences LSTM recurrent neural network," *IEEE Access*, vol. 6, pp. 59438–59448, 2018.
- [12] M. Capuno, J. S. Kim, and H. Song, "Very short-term load forecasting using hybrid algebraic prediction and support vector regression," *Mathematical Problems in Engineering*, vol. 2017, Article ID 8298531, 9 pages, 2017.
- [13] H. Jiang, Y. Zhang, E. Muljadi, J. J. Zhang, and D. W. Gao, "A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3341–3350, 2018.
- [14] S. Rashidi, M. Vafakhah, and E. K. Lafdani, "Evaluating the support vector machine for suspended sediment load forecasting based on gamma test," *Arabian Journal of Geosciences*, vol. 9, no. 11, pp. 77–84, 2016.
- [15] Y. He and Y. Zheng, "Short-term power load probability density forecasting based on Yeo-Johnson transformation

- quantile regression and Gaussian kernel function,” *Energy*, vol. 154, no. 2, pp. 143–156, 2018.
- [16] X. Y. Wu, J. X. He, and P. Zhan, “Analysis on characteristics of communication delay in wide area measurement system based on probability distribution,” *Automation of Electric Power Systems*, vol. 39, no. 12, pp. 50–55, 2015.
- [17] Q. P. Huang, Y. J. Li, and S. Lu, “Short-time load forecasting based on fuzzy clustering and random forest,” *Electrical Measurement & Instrumentation*, vol. 54, no. 23, pp. 41–46, 2017.
- [18] L. Ghelardoni, A. Ghio, and D. Anguita, “Energy load forecasting using empirical mode decomposition and support vector regression,” *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 549–556, 2013.
- [19] Y. Ren, P. N. Suganthan, and N. Srikanth, “A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods,” *IEEE Transactions on Sustainable Energy*, vol. 6, no. 1, pp. 236–244, 2013.
- [20] Z. W. Qu, K. Zhang, and Y. J. Wang, “Short-term load forecasting based on WD-ASD optimized by PSO-OMP,” *Advanced Technology of Electrical Engineering and Energy*, vol. 36, no. 12, pp. 39–45, 2017.
- [21] H. B. Zou, C. L. Fu, and S. Yu, “Short-term load forecasting based on Akima-LMD and GRNN,” *Advanced Technology of Electrical Engineering and Energy*, vol. 37, no. 1, pp. 51–56, 2018.
- [22] G. Mu, K. Y. Hou, and Y. H. Yang, “Studies on load regularity evaluating method for load forecasting,” *Proceedings of the CSEE*, vol. 21, no. 10, pp. 96–101, 2001.
- [23] J. B. Wen, X. Wang, and L. X. Li, “Short-term wind power load forecasting based on frequency domain decomposition,” *Transactions of China Electrotechnical Society*, vol. 28, no. 5, pp. 67–72, 2013.
- [24] C. Q. Kang, Q. Xia, and M. Liu, *Power System Load Forecasting*, China Electric Power Press, Beijing, China, 2007.
- [25] J. Bedi and D. Toshniwal, “Empirical mode decomposition based deep learning for electricity demand forecasting,” *IEEE Access*, vol. 6, pp. 49144–49156, 2018.
- [26] 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.
- [27] L. Puggini and S. McLoone, “An enhanced variable selection and Isolation Forest based methodology for anomaly detection with OES data,” *Engineering Applications of Artificial Intelligence*, vol. 67, no. 3, pp. 126–135, 2016.
- [28] X. Zhao, Y. Wu, D. L. Lee, and W. Cui, “iForest: interpreting random forests via visual analytics,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 407–416, 2019.
- [29] H. Shao, F. Cui, and X. Deng, “Short-term wind speed forecasting using the wavelet decomposition and AdaBoost technique in wind farm of east China,” *IET Generation, Transmission & Distribution*, vol. 10, no. 11, pp. 2585–2592, 2016.
- [30] Z. D. Tian, S. J. Li, Y. H. Wang, and X. W. Gao, “Short-term wind speed combined prediction for wind farms based on wavelet transform,” *IET Generation, Transmission & Distribution*, vol. 30, no. 9, pp. 112–120, 2015.
- [31] J. Li and L. P. Chang, “A SAR image compression algorithm based on Mallat tower-type wavelet decomposition,” *Optik*, vol. 126, no. 23, pp. 3982–3986, 2015.
- [32] G. Z. Li, W. Y. Liu, H. Z. Yun, and Y. H. Gao, “A new data pre-processing method for bus load forecasting,” *Power System Technology*, vol. 34, no. 2, pp. 149–154, 2010.