International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 341-349

https://doi.org/10.70517/ijhsa46429

Research on power load forecasting model based on the combination of time-frequency domain analysis and STL decomposition

Tianyi Yang[1,*]
[1] School of Mechanical Engineering, Wuhan University of Science and Technology, Wuhan, Hubei, 430081, China
Corresponding authors: (e-mail: poloyang233@wust.edu.cn).

Abstract This paper preprocesses missing and anomalous power load data to reduce the error of power load forecasting. Loess smoothing-based time series decomposition algorithm (STL) is introduced to initially decompose the data into trend component, period component and residual component according to the inner and outer loop process. Aiming at the noise limitation of the residual component, the method of reconstruction is carried out by improving the STL-ICEEMDAN quadratic decomposition. Combined with sample entropy and maximum information number, it is quadratically divided into high-complexity component and low-complexity component to improve the prediction accuracy. The results show that the trend component autocorrelation coefficients are 0.786 and 0.729, the correlation is high, there is periodicity, and the decomposition method is effective. The root mean square error, average absolute error, average absolute percentage error, and relative absolute error of power load prediction of this paper's model are 91.66kW, 77.91kW, 0.88%, and 0.98%, respectively, and the prediction error is smaller than the comparison model.

Index Terms electricity load forecasting, STL decomposition, inner and outer loops, quadratic decomposition reconstruction, residual components

I. Introduction

With the continuous development of power systems and smart grids, the marketization of electric energy has received increasing attention internationally in recent years. Electricity load forecasting is crucial for scientific electricity management, rational planning of energy construction, and improving the economic and social benefits of the power system [1], [2]. For power sales companies, the results of power load forecasting are the basis of contracting strategy, offer strategy, trading strategy, and individual user economic measurement and other behaviors [3], [4]. Therefore, improving the accuracy of load forecasting is a key issue in the field of electric energy.

Currently, power load forecasting can be divided into two major categories. One is the traditional time series power load forecasting method, which has the advantages of simple structure, fast forecasting speed, etc., but it is difficult to highly fit the nonlinear and multifactor data [5], [6]. The second is power load forecasting with machine learning, whose goal is to mine the implied laws from a large amount of historical data, which not only can highly fit the nonlinear data, but also has the advantages of good adaptive ability and robustness [7]-[9]. The above data-driven time series and machine learning approaches are based on fitting forecasting models to historical data, so the nature of data features and dynamic changes have a greater impact on the accuracy of model forecasting [10]-[12].

However, in the practical application of power load forecasting, the uncertainty of domestic and international power load demand and the dynamic changes in power consumption patterns hinder the accuracy of machine learning algorithm forecasting [13], [14]. In contrast, the STL time-series decomposition algorithm can decompose the load data so that the components with different uncertainties and dynamic change patterns can be separated [15], [16]. Therefore, constructing a combination method combining the time-frequency domain analysis method and STL decomposition can effectively improve the prediction accuracy of uncertain power loads [17].

Accurate forecasting of loads can improve the emergency response capability of power systems. This paper focuses on the problem of normalized preprocessing of power load data and proposes a decomposition and reconstruction method. For anomalous data, the inner and outer loop technique of STL decomposition is used to smooth the processing time series, introduce robust weights, and realize the decomposition of trend component, cycle component and residual component of the data. Construct the secondary decomposition reconstruction model based on STL-ICEEMDAN. Measure the complexity of the decomposed component data by sample entropy, and calculate the data correlation by using the maximum information number to complete the re-decomposition reconstruction of the residual component. Compare the actual power load prediction effect of this paper's model and the same type of prediction model to verify the practical value of this paper's method.

II. Implementation of power load forecasting based on the combination of time-frequency domain analysis and STL decomposition

II. A. Power load data preprocessing

II. A. 1) Missing data processing

Missing items in power load data are values that are missing or zero in the power load data. These missing items may be caused by factors such as interference in data transmission, errors in staff operation, or problems with the power equipment itself. These missing items do not represent the load situation in this period, and will cause some interference in load forecasting, so it is necessary to effectively fill in the missing data in advance. In the field of load forecasting, the principle of dealing with missing data is to maintain the integrity of the data as much as possible, and the current common methods include deletion method and filling method.

1) Deletion method

The deletion method is to directly delete the missing items of the historical power load data directly, in order to purify the data and improve the performance of the model. Deletion method is a direct and effective method in data preprocessing, but its applicability is limited by the number of missing items. When there are a large number of missing items in the dataset, the deletion method is not the best choice. In the scenario of continuous data such as load forecasting, although the deletion of missing items can simplify the dataset, too many deleted missing items may damage the integrity and representativeness of the data, leading to a reduction in the sample size, which reduces the model's generalization ability. In addition, if the deletion is of missing terms in specific features, it may lead to the loss of important information provided by these features, affecting the final prediction results. Therefore, before using the deletion method, there is a need to weigh the integrity of the data and the accuracy of the prediction results.

2) Filling method

The purpose of the filling method is to estimate or fill in the missing items in a number of ways that can maximize the retention of information from the original data, while reducing the prediction error due to missing data. The power load data used in this paper is sampled at 10-minute intervals, the time interval between two adjacent samples is short, and the load data is time series data, so when the missing term is a single one, which has a large correlation with the two load values before and after it, it can be filled directly by linear interpolation, and the formula is shown in equation (1):

$$x_k = \frac{x_{k-1} + x_{k+1}}{2} \tag{1}$$

where x_k is the k th sampled data, and x_{k-1} and x_{k+1} are the previous and next data of x_k, respectively.

Simple linear interpolation may fail in the face of continuous missing data. In this case, an effective filling method is to use the neighboring point average method based on the same type of day, which takes into account the similarity characteristics of the power load data and is used to fill by using the average value of the same type of day (e.g., the same day of the week or other day of the same month) adjacent to the date on which the missing item is located, as shown in the calculation formula in Equation (2):

$$x = \frac{x_1^i + x_2^i + \cdots + x_m^i}{m} \tag{2}$$

where x is the missing value, i is some date of the same type, and m is the number of dates of the same type selected.

II. A. 2) Abnormal data handling

Since the power load data belongs to the time series data, its change has a certain inertia and shows a continuous smooth trend, and there are relatively few cases of sudden sharp increase or decrease, therefore, if there is a sharp increase or decrease in a data point, it is an abnormal term. If these data are not processed before the model training, it will affect the training effect of the model and lead to a decrease in training accuracy. The following two methods can be used to correct the anomalous terms:

1) Horizontal processing method

The power load data presents smooth and continuous characteristics, and the load data changes in adjacent moments are more stable. However, when the difference between the load value at a certain moment and the load value at adjacent moments exceeds the preset threshold, it can be judged as an abnormal term. When dealing with the abnormal term, the load data of the moments before and after the abnormal term can be used to analyze and deal with the specific formula as follows:

Firstly, the abnormal term is determined:

$$\begin{cases} |y(i,t)-y(i,t-1)| > \alpha(t) \\ |y(i,t)-y(i,t+1)| > \beta(t) \end{cases} \tag{3}$$

Calculate the average of the moments $y_1(i,t)$ before and after the moment t, as shown in equation (4):

$$y_1(i,t)=\frac{y(i,t-1)+y(i,t+1)}{2} \tag{4}$$

Calculate the average of the four neighboring moments $y_2(i,t)$ at moment t, as shown in equation (5):

$$y_2(i,t)=\frac{y(i,t-2)+y(i,t-1)+y(i,t)+y(i,t+1)}{4} \tag{5}$$

Calculate the average of the four neighboring moments $y_3(i,t)$ at the moment $y_2(i,t)$, as shown in equation (6):

$$y_3(i,t)=\frac{y_2(i,t-1)+y_2(i,t)+y_2(i,t+1)+y_2(i,t+2)}{4} \tag{6}$$

The final weighting is obtained:

$$y(i,t)=a_1 y_1(i,t)+a_2 y_2(i,t)+a_3 y_3(i,t) \tag{7}$$

Among them, $\alpha(t)$ and $\beta(t)$ are expressed as the thresholds of 15% and -15%, respectively, $y(i,t)$ is the load value at the time of t on day i, $y(i,t-1)$ is the load value at the time $t-1$ on day i, and $y(i,t+1)$ is expressed as the load value at time $t+1$ on day i.

2) Vertical processing method

Power load shows obvious periodicity, in the absence of emergencies, the load changes in the neighboring days show similarity. When the difference between the load value at a certain moment and the load value at the same moment of the neighboring days exceeds the preset threshold, it can be judged as an anomalous term, and in order to correct these anomalous terms, it can be processed by using the average value of the load at the same moment of the neighboring days, i.e., the vertical processing method.

If,

$$|y(i,t)-M(t)| > \varepsilon \tag{8}$$

Then,

$$y(i,t)=\begin{cases} M(t)-\varepsilon, & y(i,t) < M(t) \\ M(t)+\varepsilon, & y(i,t) > M(t) \end{cases} \tag{9}$$

where ε is denoted as the threshold value and $M(t)$ is denoted as the average value of the week corresponding to moment t.

II. B.STL decomposition

STL is a time series decomposition algorithm based on Loess smoothing. This algorithm can decompose the time series into trend component, period component, and residual component as shown in equation (10):

$$Y_t = T_t + S_t + R_t \tag{10}$$

where Y_t is the observation at moment t, and T_t, S_t, and R_t are the trend component, the period component, and the residual component at moment t, respectively.

The STL decomposition consists of two recursive processes, the inner loop and the outer loop. The inner loop is used to update the trend and period components of the time series, and the outer loop is used to compute the robust weights required for the next round of inner loops.

II. B. 1) Internal circulation

Let $S_v^{(k)}$ and $T_v^{(k)}$ be the results of the k th iteration, then the iteration process of the $k+1$ th inner loop is as follows:

1) Detrending. Compute the detrended sequence $Y_v - T_v^{(k)}$.

2) Smoothing the periodic subsequence. Perform Loess smoothing on the periodic subsequence of the sequence in step 1) to obtain the temporary periodic sequence $C_v^{(k+1)}$.

3) Apply a low-pass filter to $C_v^{(k+1)}$ to obtain $L_v^{(k+1)}$. The low-pass filter consists of three moving average processes and a Loess process.

4) Calculate the periodic component. The expression for calculating the period component for the $k+1$ th iteration is $S_v^{(k+1)} = C_v^{(k+1)} - L_v^{(k+1)}$. where $L_v^{(k+1)}$ is subtracted from $C_v^{(k+1)}$ to prevent low-frequency power from entering the periodic component.

5) Removing the periodic component. The expression for removing the periodic component is $Y_v - S_v^{(k+1)}$.

6) Calculate the trend component. The time series obtained in step 5) is smoothed using the Loess smoothing method to obtain the trend component for $k+1$ iterations as $T_v^{(k+1)}$.

II. B. 2) External circulation

Let the period component and trend component obtained after one inner loop be S_v and T_v, respectively. The residual component is calculated as follows:

$$R_v = Y_v - T_v - S_v \tag{11}$$

Robust weights are then introduced for each time point. The magnitude of the robust weights reflects the extremes of the residual component R_v. The robust weights are calculated as follows:

$$h = 6median\left(\left|R_v\right|\right) \tag{12}$$

$$B(u) = \begin{cases} \left(1-u^2\right)^2, & 0 \le u < 1 \\ 0, & u \ge 1 \end{cases} \tag{13}$$

$$\rho_v = B\left(\left|R_v\right| / h\right) \tag{14}$$

where ρ_v is the robust weight corresponding to time point v and R_v is the residual component corresponding to time point v. $median(\ )$ is the median function. $B(u)$ is the bi-square weight function.

In the subsequent inner loop process, when performing Loess smoothing in steps 2) and 6), the weight values should be multiplied by the robust weights ρ_v of the corresponding moments on top of the original weight values to minimize the effect of outliers on the decomposition.

II. C.STL-ICEEMDAN based quadratic decomposition reconstruction models

After the decomposition of power load data, the complexity and relevance of the components need to be analyzed and reconstructed, this paper chooses two indicators, sample entropy and maximum information number, to analyze the complexity of the decomposed components, and the principle is introduced next.

II. C. 1) Sample entropy

Sample entropy (SE) is a time series complexity measure that determines the complexity of data by measuring the size of the probability of generating new patterns in the data. The greater its probability, the greater the complexity of the data.SE is similar to approximate entropy, but the calculation of SE does not depend on the length of the data and has better robustness and consistency.

Let the load sequence consist of N data with time series $x(n) = x(1), x(2), \cdots x(N)$, and first form a one-dimensional array of m vector sequences, $X_m(1), \cdots, X_m(N-M+1)$; where $X_m(i)$ stands for the value of x for m consecutive x starting from the i th one.

Compute the distance between two sequences of vectors.

$$d\left[X_m(i), X_m(j)\right] = \max_{k=0,\cdots,m-1}\left(\left|x(i+k) - x(j+k)\right|\right) \tag{15}$$

Count the number $A^m(r)$ of distances between sequences less than or equal to r, which is the probability that two sequences match at m points. r is the similarity tolerance, which represents a measure of similarity.

$$A_i^m(r) = \frac{1}{N-m-1} A_i \tag{16}$$

When N is finite data, its sample entropy is calculated as shown in equation (17).

$$SE_{(m,r,N)} = -\ln\left[\frac{A^{m+1}(r)}{A^m(r)}\right] \tag{17}$$

where m is the embedding dimension of the moving window of the reconstructed time series.

II. C. 2) Maximum number of messages

The Maximum Information Coefficient (MIC) has been proposed to describe the degree of correlation between any two sets of random variables under big data, which is not susceptible to outliers and is characterized by universal and fair data measurement. If there is some relationship between two variables $X = \{1, 2, \cdots, n\}$ and $Y = \{1, 2, \cdots, n\}$, it can be treated as an ordered set $D(x,y)$, and then x and y are taken in some way in the data X and Y directions to be divided in a grid This makes most of the points fall centrally in the cells, from which the MIC values of the two sets of variables can be calculated as shown in Eq. (18).

$$MIC(D) = \max(D)_{X,Y} = \max \frac{I(D,X,Y)}{lb\min(X,Y)} \tag{18}$$

where $I(D,X,Y)$ is the maximum mutual information of all the distributions generated by D on the X and Y lattices, if the stronger the correlation between the two sets of variables X and Y is, the closer $MIC(D)$ tends to be to 1, and vice versa the closer it is to 0, which means that it is more likely that the two sets of variables are independent variables.

II. C. 3) STL-ICEEMDAN based quadratic decomposition reconstruction model construction

In this paper, the STL-ICEEMDAN quadratic decomposition reconstruction strategy is used to process the original power load data. Figure 1 shows the improvement strategy flow.

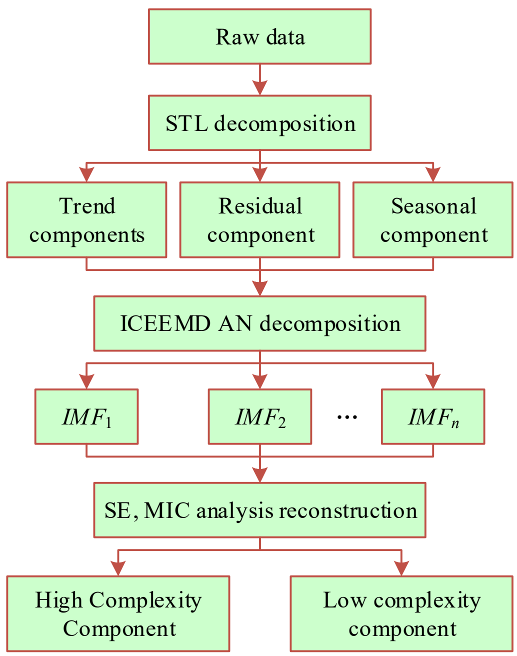

Figure 1: Shows the process of the secondary decomposition and reconstruction model

The STL primary decomposition is first performed to obtain the trend, seasonal and residual components. However, the residual component still contains complex nonlinearities and noise, and ICEEMDAN decomposition is performed on the residual component to obtain multiple IMF_n (intrinsic modal components), so as to obtain useful information not found in the residual component. Due to the excessive number of components obtained from the decomposition, direct prediction will take more time. Meanwhile, there are still some correlations between these components, which can be further analyzed by reconstructing the related components to reduce the inputs for prediction and improve the efficiency of prediction while reducing the non-stationarity of the data.

In this paper, sample entropy and maximum information number are taken to analyze the components. The complexity of each intrinsic modal component is calculated by sample entropy, and the maximum information count calculates the correlation between each intrinsic modal component and the original data. The components with high complexity and small MIC value are reconstructed as high complexity components, and the rest are reconstructed as low complexity components, and finally four parts of the components are obtained, i.e., trend component, seasonal component, high complexity component and low complexity component.

III. Electricity load data processing and forecasting practices

III. A. STL-ICEEMDAN based quadratic decomposition for reconstructing load data

III. A. 1) Analysis of raw load data

Taking the power load data of a city power grid as the basic research value, using the secondary decomposition and reconstruction model based on STL-ICEEMDAN to decompose and reconstruct the original load data, and taking the obtained trend component and its cycle change rule as an example to analyze the effectiveness of the data preprocessing method proposed in this paper. Figure 2 shows the raw power load data obtained with 10 minutes as the collection unit. It can be seen in the raw load data of 6000min consecutively, the change range of electric power load volume is between 8000MW-13000MW, and there exists some kind of law with 1500min as the differentiation, which meets the requirements of decomposition and reconstruction.

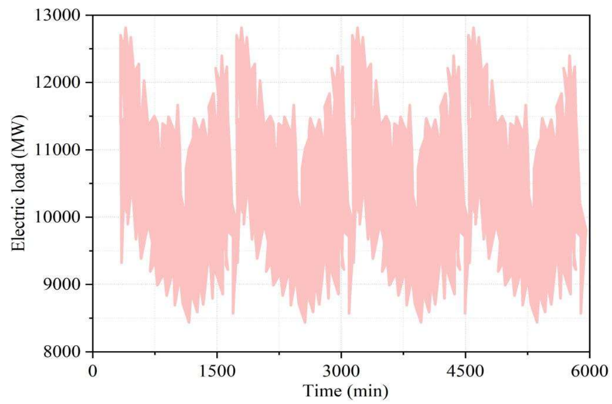

Figure 2: Original power load data

III. A. 2) Trend components and their autocorrelation analysis

Figure 3 shows the trend components obtained from the decomposition. Figure 4 is the result of autocorrelation calculation of the trend component. The maximum value of the trend component obtained by decomposition is 11,817.6 MW, the minimum value is 7,434.1 MW, and the mean value is 9,547.1 MW.The red dots in Fig. 4 indicate the correlation coefficients between the data at n*332 intervals, which are 0.786 and 0.729, respectively, and both of them are greater than 0.700.It shows that there is a high correlation between the trend component obtained by decomposition at n*332 intervals, proving that there is a periodicity in the trend component. There is a high correlation, which proves that the trend component contains periodicity. By processing the raw power load data with outliers through the secondary decomposition reconstruction model of STL-ICEEMDAN, the periodicity of the power load data can be found, so as to realize the normalization of the raw power load data and improve the accuracy of the subsequent power load forecast.

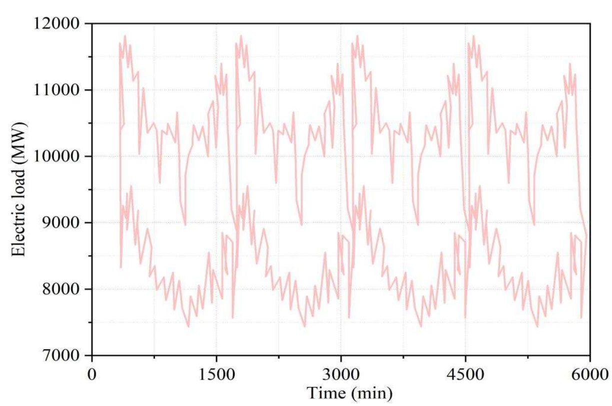

Figure 3: Decomposed trend components

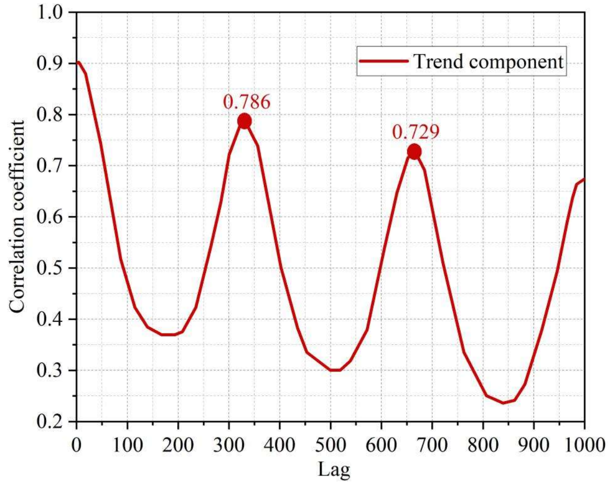

Figure 4: Calculation result of the autocorrelation of the trend component

III. B. Comparison of Power Load Forecasting Effectiveness

III. B. 1) Comparison of model predictions

In order to reflect the advantages of this paper's power load data preprocessing method based on time-frequency domain analysis and STL-ICEEMDAN decomposition, the single prediction models such as Extreme Learning Machine (ELM), Support Vector Regression (SVR), LSTM, and TCN, as well as the power load prediction model combining this paper's method, are selected for comparative analysis. Fig. 5 shows the prediction results of the five methods on a particular day in the test set. In a 24-hour period, the real power load has 2 peaks around 6h and 16h, and the overall power load variation ranges from 8411.5 MW to 12143.1 MW. Among the prediction results of the five models, this paper's model is the closest between the model and the real power load change. The model of this paper predicts that the power peaks at 6.5h and 16.2h, with the change range of 8499.2MW-12219.5MW, and the overall difference is not more than 1h and 100MW. The prediction effect of other single prediction models is not as good as this paper's model, and there is an obvious deviation from the real value.

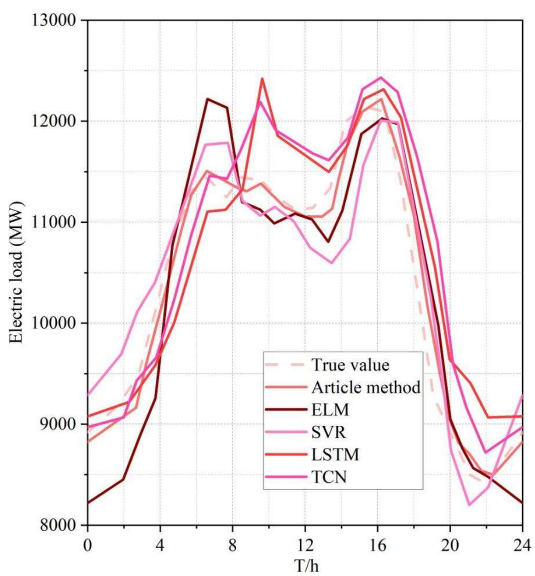

Figure 5: Comparison of the prediction results of each model

III. B. 2) Comparison of prediction errors of different models

The model prediction results were judged using three error judging metrics, i.e., judging criteria were used for judging, i.e., Root Mean Square Error (RRMSE), Mean Absolute Error (MMAE), and Mean Absolute Percentage Error (MMAPE). Table 1 shows the prediction error statistics of different models. The root mean square error of this paper's model is 91.66kW, the average absolute error is 77.91kW, and the average absolute percentage error is 0.88%. Compared with the three corresponding index data of the other four models, this paper's model has a smaller root mean square error, average absolute error and average absolute percentage error. Figure 6 shows the distribution of the relative absolute errors of the prediction results of each model at any moment. Compared with other models, this paper's method has a smaller relative absolute error at any moment, with a maximum of no more than 0.98%, and the maximum relative absolute error results are all smaller than the four comparison models of ELM (10.36%), SVR (11.06%), LSTM (10.71%), and TCN (10.84%). It indicates that after the time-frequency domain analysis and STL-ICEEMDAN decomposition, the power load data are more standardized, which improves the meticulousness and accuracy of the model power load prediction results.

Table 1: Prediction errors of different models

Model	R_{RMSE} (kW)	M_{MAE} (kW)	M_{MAPE} (%)
Article method	91.66	77.91	0.88
ELM	1100.69	908.58	10.19
SVR	1158.52	976.92	10.95
LSTM	1117.40	949.65	10.64
TCN	1118.98	953.98	10.69

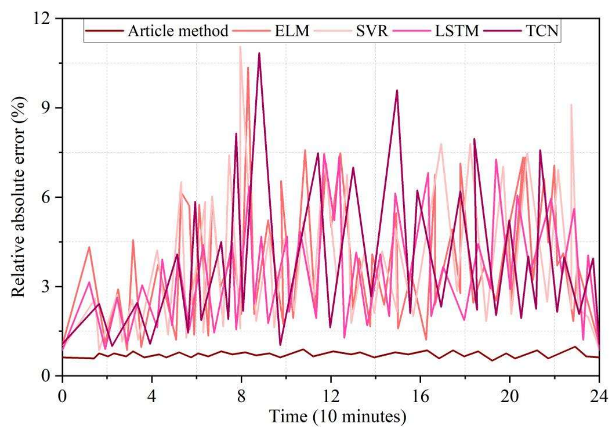

Figure 6: Distribution of relative and absolute errors

IV. Conclusion

In this paper, the abnormal power load data are processed through data time series analysis and STL-ICEEMDAN quadratic decomposition reconstruction model to improve the level of power load forecast refinement. The model in this paper predicts that the peak power load occurs at 6.5h and 16.2h, and the load changes between 8499.2MW and 12219.5MW, which is closest to the real value. And the root mean square error = 91.66kW, the average absolute error = 77.91kW, the average absolute percentage error = 0.88%, and the relative absolute error at any moment is <0.98%, which is better than the comparative model in different judging indexes. In the future, the time length of power load prediction can be appropriately increased to test the generalization ability of this paper's method.

References

[1] Yildiz, B., Bilbao, J. I., & Sproul, A. B. (2017). A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renewable and Sustainable Energy Reviews, 73, 1104-1122.

[2] Carvallo, J. P., Larsen, P. H., Sanstad, A. H., & Goldman, C. A. (2018). Long term load forecasting accuracy in electric utility integrated resource planning. Energy Policy, 119, 410-422.

[3] Zhu, J., Dong, H., Zheng, W., Li, S., Huang, Y., & Xi, L. (2022). Review and prospect of data-driven techniques for load forecasting in integrated energy systems. Applied Energy, 321, 119269.

[4] Oreshkin, B. N., Dudek, G., Pełka, P., & Turkina, E. (2021). N-BEATS neural network for mid-term electricity load forecasting. Applied Energy, 293, 116918.

[5] Gasparin, A., Lukovic, S., & Alippi, C. (2022). Deep learning for time series forecasting: The electric load case. CAAI Transactions on Intelligence Technology, 7(1), 1-25.

[6] Yaslan, Y., & Bican, B. (2017). Empirical mode decomposition based denoising method with support vector regression for time series prediction: A case study for electricity load forecasting. Measurement, 103, 52-61.

[7] Aguilar Madrid, E., & Antonio, N. (2021). Short-term electricity load forecasting with machine learning. Information, 12(2), 50.

[8] Cordeiro-Costas, M., Villanueva, D., Eguía-Oller, P., Martínez-Comesaña, M., & Ramos, S. (2023). Load forecasting with machine learning and deep learning methods. Applied Sciences, 13(13), 7933.

[9] Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. Energies, 11(7), 1636.

[10] Al Mamun, A., Sohel, M., Mohammad, N., Sunny, M. S. H., Dipta, D. R., & Hossain, E. (2020). A comprehensive review of the load forecasting techniques using single and hybrid predictive models. IEEE access, 8, 134911-134939.

[11] Sadaei, H. J., e Silva, P. C. D. L., Guimaraes, F. G., & Lee, M. H. (2019). Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series. Energy, 175, 365-377.

[12] Fallah, S. N., Deo, R. C., Shojafar, M., Conti, M., & Shamshirband, S. (2018). Computational intelligence approaches for energy load forecasting in smart energy management grids: state of the art, future challenges, and research directions. Energies, 11(3), 596.

[13] Zha, W., Ji, Y., & Liang, C. (2024). Short-term load forecasting method based on secondary decomposition and improved hierarchical clustering. Results in Engineering, 22, 101993.

[14] Fan, M., Hu, Y., Zhang, X., Yin, H., Yang, Q., & Fan, L. (2019, November). Short-term load forecasting for distribution network using decomposition with ensemble prediction. In 2019 Chinese Automation Congress (CAC) (pp. 152-157). IEEE.

[15] Sun, T., Zhang, T., Teng, Y., Chen, Z., & Fang, J. (2019). Monthly electricity consumption forecasting method based on X12 and STL decomposition model in an integrated energy system. Mathematical Problems in Engineering, 2019(1), 9012543.

[16] Tian, Y. J., Zhou, S. J., Wen, M., & Li, J. G. (2021, March). A short-term electricity forecasting scheme based on combined GRU model with STL decomposition. In IOP Conference Series: Earth and Environmental Science (Vol. 701, No. 1, p. 012008). IOP Publishing.

[17] Mohan, N., Soman, K. P., & Kumar, S. S. (2018). A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model. Applied energy, 232, 229-244.