WDE: Elegantly Forecasting Power Load through Advanced Signal Refinement and Graceful Normalization Approach

Yun Su[1], Junyi Sha[2], Yingjie Tian[1], Mi Wen[2], Naiwang Guo[1], Fan Li[1]*

[1]Shanghai Electric Power Company, Shanghai, China
[2]Shanghai University of Electric Power, Shanghai, China
*13083521760@163.com

Corresponding author: Junyi Sha

Keywords: Signal decomposition, Load forecasting, Data processing, Feature engineering, Normalization method

1 Introduction

Load forecasting is a vital element in the orchestration of urban energy management and is especially critical during periods of electricity paucity. According to the United Nations Human Settlements Programme, urban centers are responsible for consuming more than half of the global energy resources. Thus, efficient planning and management of urban energy are vital for energy conservation. Reaching carbon neutrality is a shared goal of humanity [1], necessitating accurate energy forecasting, which plays a crucial role in achieving these energy goals.

Recent developments in energy prediction have incorporated both deep learning and traditional machine learning. A novel deep learning structure employing an attention mechanism has been proposed by the Transformer model [2]. This model allows the system to access historical data irrespective of distance and is more adept at handling repetitive patterns with long-term dependencies than RNN-based approaches [3]. Several variants, including Autoformer [4], Pyformer [5], Fedformer [6], and Informer [7], have been developed based on this model. The Transformer model primarily utilizes the self-attention technique to extract the semantic interdependence of item pairs [2], reducing the time and space complexity. Various strategies have been suggested to enhance its effectiveness, including pyramidal attention [5] and loose masks [8]. The Dlinear model, a unique approach based on the Transformer model [2], combines the positional encoding strategy used in Autoformer [4] and Fedformer [6] with a linear layer. However, there is seldom any semantic relationship resembling a point between the basic numerical data in the time series.

Current research demonstrates that combining decomposition methods with certain models can enhance the prediction accuracy. For example, Peng's research delved into load forecasting utilizing the CEEMDAN and transformer methodologies [9], whereas Changchun explored the synergy hybrid network [10]. Han's approach centered on the EMD-Isomap-AdaBoost model [11], while Wang implemented a model based on VMD-CISSA-LSSVM for electricity load prediction [12]. Commonly, these studies lack robust feature selection methodologies, struggle to yield satisfactory outcomes when addressing nonlinear time series data, and lack an effective solution for the noise issue inherent in power load data. In response to these challenges, this paper introduces a comprehensive prediction framework based on feature processing and hybrid modeling. The contributions of this research are delineated as follows:

- To minimize the error in power consumption prediction, this paper employs a random forest method for feature analysis and selection, uses variational mode decomposition (VMD) to decompose power consumption, optimizes the IMF component with the WOA algorithm, and utilizes a metaheuristic Kalman filter for data reconstruction to avoid the impact of data noise.
- Given the limited improvement in prediction accuracy achieved by a single model, this paper introduces a hybrid model. This model optimizes input and output using efficient normalization and anti-normalization methods, narrows feature differences, accelerates convergence, and enhances prediction accuracy and efficiency. The TIDE model effectively addresses the limitations of linear models in modeling nonlinear relationships and external variables.
- To substantiate the efficacy of the proposed methodology, this study employs a dataset of power consumption from Shanghai's Jinshan district. The findings reveal that our predictive approach surpasses current methodologies in terms of performance.

The rest of this article is structured as follows:

In Section 2, the proposed data preprocessing method is discussed. Section 3 describes the proposed prediction approach. In section 4, the experimental setup and data analysis are given. The fifth part summarizes the entire paper.

2. Data Preorocess Method

2.1 Data Preprocessing

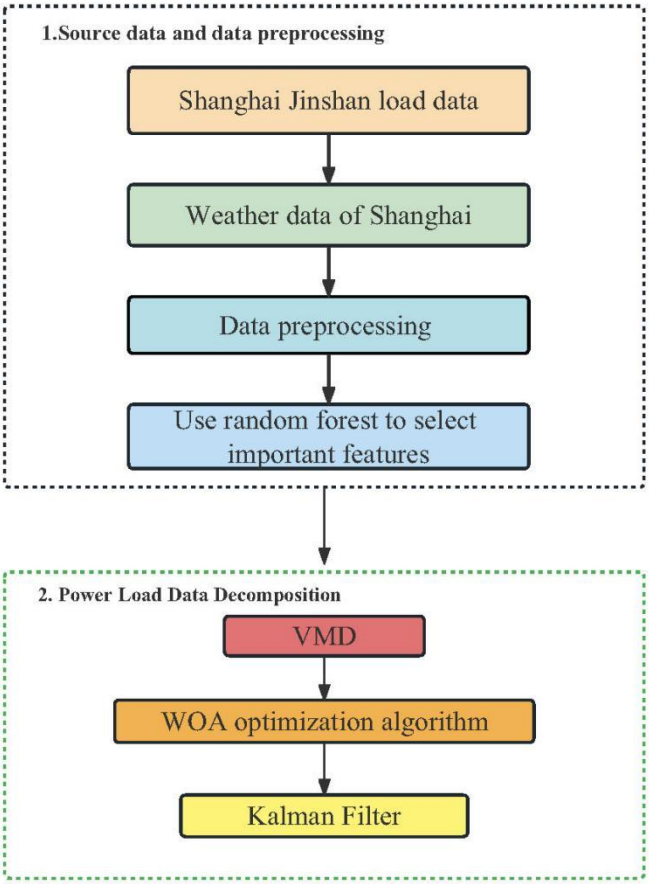

Fig. 1. The data preprocessing procedure.

The overall preprocessing procedure is shown in Fig. 1. First, in the case of missing values in the dataset, this paper uses linear interpolation to fill the missing data, including filtering some dirty data.

Feature engineering is a simpler way to represent uncertain data and improve its accuracy. Random forests are used in this process because they allow for a more accurate assessment of correlations between features. In a random forest, each decision tree is based on a different sample, and each node is randomly divided in the feature selection process to compare errors in different backgrounds.

In this paper, a random forest is used as a feature engineering method for feature selection, and the features not closely related to the power load are eliminated to improve the prediction accuracy. Using the dataset of the JinShan region in Shanghai, this paper selected five positive correlation features, namely, the maximum temperature and minimum temperature.

2.2 Power Load Data Decomposition

WOA-VMD decomposes a signal sequence into its natural modal components and residuals. This is achieved by adapting to actual changes in the signal sequence, iteratively finding the ideal frequency center through a whale optimization algorithm, and applying a limited bandwidth to each mode. The power load time series signal is separated into its frequency domain. VMD[13] is capable of significantly reducing the number of decompositions compared to EMD [14], and the stationary properties and frequency scales of the subsequences differ. The two key components of VMD[13] technology are the formulation and optimization of variational problems (VPs).

- VMD Module. The center frequency is obtained by constantly updating the center frequencies and gradually adjusting the frequency band of each mode. The variational challenge may be formulated as:

$$\min_{\{v_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \otimes \right. \right. \right.$$

$$\text{s.t.} \sum_{k=1}^{K} u_k = f$$

and can be described as:

$$L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|^2 + \|f(t) - \sum_k u_k(t)\|_2^2 + < \lambda(t), f(t) - \sum_k u_k(t) > \quad (2)$$

and the new modal components are denoted by:

$$\grave{u}_k^{n+1}(\omega) = \frac{\grave{f}(\omega) - \sum_{i \neq k} \grave{u}_i(\omega) + (\grave{\lambda}(\omega)/2)}{1 + 2\alpha(\omega - \omega_k)^2} \quad (3)$$

where $\grave{f}(\omega), \grave{u}_i(\omega), \grave{\lambda}(\omega)$ and $\grave{u}_k^{n+1}$ are the estimated values of $f(t), u_t(t), \lambda(t)$ and u_k^{n+1} after the Fourier transform, n is the number of iterations, and ω is the frequency. Figure 4 depicts the VMD [13] calculation procedure, and the parameter estimate can be expressed as:

$$\grave{\omega}_k^{n+1} = \frac{\int_0^\infty \omega |\grave{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\grave{u}_k(\omega)|^2 d\omega} \quad (4)$$

Although the VMD [13] method can effectively remove noise, the corresponding hyperparameters need to be determined in advance. This paper uses the WOA [15] to optimize VMD [13] and determine the hyperparameters in advance. In later experiments, it was proven that by comparing different IMF quantities, the WOA [15] is indeed effective

- WOA Module. The whale optimization algorithm (WOA) [15], utilizing the minimal envelope entropy value as its fitness function, has been employed for the optimization of VMD [13] parameters. First, the position vector [K,a] of the whale swarm is initialized. Then, the fitness of each whale is calculated using the envelope entropy as the fitness function. Finally, the iterative formula is used to update the position vector iteratively until the optimal VMD [13] parameters are obtained. After VMD [13] is used, the noise of each IMF is still not smooth enough, which is not conducive to our prediction. Therefore,

Kalman [16] was used to smooth the IMF data and reduce the cause of inaccurate prediction accuracy caused by overdryness.

- Kalman Filter. The Kalman filter [16] is a method that employs the governing equation of a linear system to approximate the best state of the input and output observation information. It is an effective approach for noise reduction and data recovery in data processing.
  The Kalman filter [16] consists of two stages: prediction and correction. The process of prediction and correction is as follows:
  - Prediction: The initial estimate of the current time step k is determined, and the state of the present time step k is predicted based on the posterior estimate of the preceding time step k-1.

$$\begin{aligned} \dot{x}_k^- &= A\dot{x}_{k-1} + Bu_k \\ P_k^- &= AP_{k-1}A^T + Q \end{aligned} \tag{5}$$

    where $\dot{x}_k^-$represents the prior state estimate value at time k, and the result at time k predicted according to the optimal estimate of the previous time $k-1$, A signifies that the state is essentially a hypothesized model for the target state transition; B represents the input control matrix, how does external influence translate into state influence; $\dot{x}_{k-1}$ and $\dot{x}_k$, respectively, a posteriori state estimates of the $k-1$ moment and the k moment, are the results of filtering, that is, the updated result, also known as the optimal estimate; and Q represents a covariance matrix for predicting the state
  - Measurements: To adjust estimations made during the prediction stage and provide an a posteriori estimate for the present.

In the prediction stage, the Kalman filter [16] computes the prior estimates of state variables and error covariance based on the state estimates from the previous instant. In the correction stage, an improved posterior estimate and additional measurement variables are incorporated into the a priori estimations to refine the state estimates for the present moment.

3 The Proposed Forecast Approach

The components of the WDE model, including TimesBlock and the normalized anti-normalized model, are described below.

3.1 TimesBlock

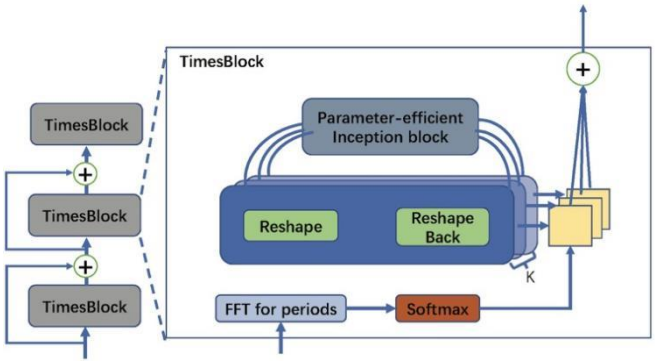

Fig. 2. The Times Bolck Model.

As shown in Fig. 2, TimesNet [17] is composed of layered TimesBlocks. The input first passes through the embedding layer to extract the deep features. $\mathbf{X}_{1\mathrm{D}}^0 \in \mathbb{R}^{T\times d_{\text{model}}}$, for layer L TimesBlock, its input is $\mathbf{X}_{1\mathrm{D}}^{l-1} \in \mathbb{R}^{T\times d_{\text{model}}}$ After that, 2D convolution is used to extract 2D timing changes:

$$\mathbf{X}_{1\mathrm{D}}^{l} = \mathrm{TimesBlock}\left(\mathbf{X}_{1\mathrm{D}}^{l-1}\right) + \mathbf{X}_{1\mathrm{D}}^{l-1} \tag{6}$$

Specifically, TimesBlock consists of the following subprocedures:

- Transform: First, the input one-dimensional timing features $\mathbf{X}_{1\mathrm{D}}^{l-1}$ are extracted, and the period is transformed into a two-dimensional tensor to represent a two-dimensional timing change. The highest intensity k frequency $\{f_1, \cdots, f_k\}$ corresponds to the most significant k period length $\{p_1, \cdots, p_k\}$.
- The two-dimensional tensor $\left\{\mathbf{X}_{2D}^{l,1}, \mathbf{X}_{2D}^{l,2}, \ldots, \mathbf{X}_{2D}^{l,k}\right\}$ is extracted since it has a two-dimensional locality. Therefore, the information is extracted via 2D convolution. Here, using the classical Inception model, namely:

$$\widehat{\mathbf{X}}_{2\mathrm{D}}^{l,i} = \mathrm{Inception}\left(\boldsymbol{X}_{2D}^{l,i}\right) \tag{7}$$

- Dimensionality reduction: For the extracted temporal features, they are converted back to one dimension

$$\widehat{\mathbf{X}}_{1\mathrm{D}}^{l,i} = \mathrm{Trunc}\left(\mathrm{Reshape}_{1,(p_i\times f_i)}\left(\widehat{\mathbf{X}}_{2\mathrm{D}}^{l,i}\right)\right), i \in \{1, \cdots, k\} \tag{8}$$

  Among them, $\widehat{\mathbf{X}}_{1\mathrm{D}}^{l,i} \in \mathbb{R}^{T\times d_{\text{model}}}$, $\mathrm{Trunc}(\cdot)$ means removing the 0 added by the padding.)operation in the above.
- Adaptive Fusion: In the following step, an onedimensional representation of $\left\{\widehat{\mathbf{X}}^{l,1}, \cdots, \widehat{\mathbf{X}}^{l,k}\right\}$ is weighted to sum the intensity of the response frequency to obtain the final output.

3.2 The Forecast Model

The overall model architecture is shown in Fig. 3.

DataSet	Time-Series	Time-Points	Frequency
Electricity-Jinshan	7	8760	1 Hour

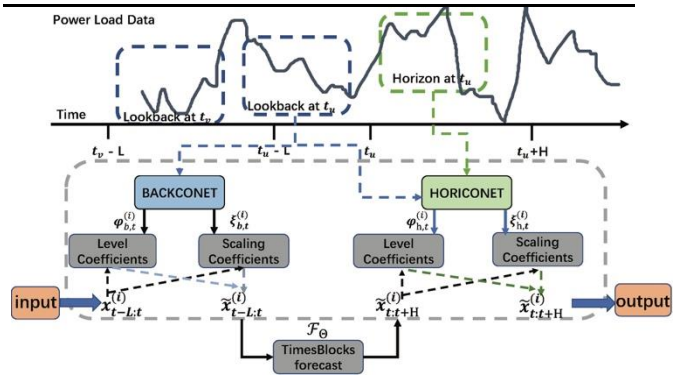

Fig. 3. The Forcast Model.

In the prediction model, this paper uses the Dish-Ts [18] architecture to integrate TimeBlocks into the model through a normalization-anti-normalization process.

- Conet: unsteady time series are difficult to predict accurately. Pilot works measure the distribution and its variation by means of statistics (usually means and standard deviations) or a distance function. However, these operations are not reliably quantifiable and have limited expressive power. In this regard, the general statement is

$$\varphi, \xi = \mathrm{CONET}(x) \tag{9}$$

  $\varphi \in R^1$ represents the horizontal coefficient, indicating the total length of the input sequence within the window $x \in R^L; \xi \in R^1$ denotes the factor, representing the variance scale of x. Typically, the model can be configured to any neural architecture for linear or nonlinear mapping, endowing it with considerable modeling ability and adaptability.

- Dual-Conet: To mitigate internal space shifts and interval space shifts in the aforementioned time series. The BACKCONET is specifically designed for comprehending the spatial distribution within the input $\left\{x_{t-L:t}^{(i)}\right\}_{t=L}^{T-H} \in x_{\text{input}}^{(i)}$, HORICONET for the output and spatial distribution of $\left\{x_{t:t+H}^{(i)}\right\}_{t=L}^{T-H} \in x_{\text{output}}^{(i)}$. In multivariate forecasting, the two CONEts are represented as:

$$\varphi_{b,t}^{(i)}, \xi_{b,t}^{(i)} = BACKCONET\left(x_{t-L:t}^{(i)}\right), i = 1, \cdots, N$$
$$\varphi_{h,t}^{(i)}, \xi_{h,t}^{(i)} = BACKCONET\left(x_{t-L:t}^{(i)}\right), i = 1, \cdots, N \tag{10}$$

  where $\varphi_{b,t}^{(i)}, \xi_{b,t}^{(i)} \in R^1$ is the regressive coefficient of the window and $\varphi_{h,t}^{(i)}, \xi_{h,t}^{(i)} \in R^1$ is the coefficient of the horizontal line at time step number t given a single ith variable sequence. Although the same input is $x_{t-L:t}^{(i)}$, the two CONETS have different goals.

4 EXPERIMENTAL SETUP AND DATA ANALYSIS

In this chapter, experimental data and several different experiments are used to support the findings presented in this paper. In the first section, this paper introduces the experimental environment and data. In the second section, this paper will present the comparison results with the transformer class model [2], DLinear model [19], and TimesNet [17]. In the third section, this paper will present the results of comparisons between different modal decomposition techniques.

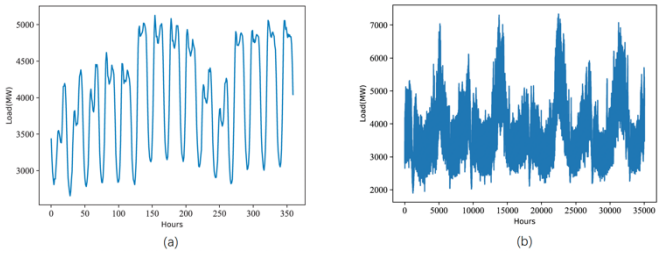

Fig. 4. (a) 15-day power load data. Yearly seasonality load profiles.

4.1 Experimental environment and load data

This paper assessed how well the suggested prediction techniques performed through a number of experiments. The experiment was implemented in Python 3.8, using PyTorch version 1.16 and MATLAB 2018 to write the WOA [15] and VMD [13] programs, with a computer configured with a Core (TM) i7-9700 CPU and 16.00 GB of RAM and a GTX2060 GPU. **Data Set.** The dataset is the actual electricity consumption of the Jinshan area in Shanghai, as shown in Table 1.

Table 1 SUMMARY OF DATASETS.

The purpose of this paper is to predict the power consumption of 12 h − 336 h in the future, so this paper uses direct multistep prediction.

The power load fluctuates due to the unstable output of renewable energy. Additionally, power load data are heavily polluted by random noise, which is attributed to users' unique

usage patterns. Fig. 4 presents the annual cycle and daily cycle data, wherein the noise is visibly significant.

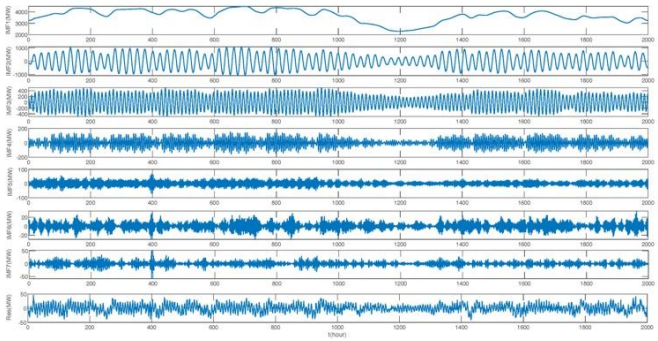

Fig. 5. Modal component of VMD.

In this paper, the WOA-VMD module is employed to reconstruct the data. Fig. 5 displays the modal component diagram, where the X-axis represents hours, and the Y-axis represents the specific load. Here, 2000 hours are taken as an example. This indicates that the modal component of VMD [13] has a smoother distribution of values than does the distribution of the original data.

Table 2 displays the core frequency of each IMF component, delineating a breakdown. When $K < 7$, the central frequency decreases, indicating that the IMF model might not be adequately decomposed. When $K = 8$, IMF5 and IMF6 exhibit similar modes, demonstrating that $K = 7$ is appropriate for the experiment. Additionally, Fig. 6 illustrates the calculation process of VMD [13] and the number of IMFs calculated by the WOA [15].

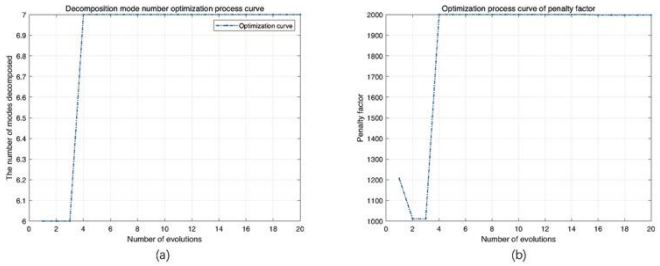

Fig. 6. (a) The calculation process of the optimal penalty factor of VMD [13] and (b) the number of IMFs calculated by the WOA [15]

After obtaining each IMF component, the data are reconstructed, and the Kalman filter [16] is used for noise reduction and smoothing. During the prediction phase, a prior estimate of the current state variable, a prior estimate of the error covariance, and the estimate obtained from the state at the previous moment are calculated. In the revision stage, the prior estimate is integrated with the new measurement variable to refine the posterior estimate. The final data are illustrated in Fig. 7.

TABLE 2 FUNDAMENTAL FREQUENCY OF EACH-IMF COMPONENTS AT VARIOUS K

K	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	IMF10
5	0.02	88.31	46.23	293.41	170.71					
6	0.02	88.30	46.23	213.28	295.42	169.82				
7	0.01	46.43	89.41	169.31	213.44	169.76	379.00			
8	0.01	46.43	89.40	169.32	213.40	295.29	378.77	461.80		
9	0.01	46.45	89.41	169.34	213.38	295.19	333.33	374.47	461.43	
10	0.01	46.44	89.42	169.34	213.29	253.33	296.26	329.93	378.61	461.53

TABLE 3 COMPARISON WITH TRANSFOMERS AND OTHER METHODS

TimesNet [17] models are selected for experimental

Method		WDE		TimesNet [17]		DLinear [19]		Transformer [2]		Autoformer [4]		Fedformer [6]		Pyformer [5]		Informer [7]	
Metric		MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
Electricity-Jinshan	12 h	0.131	0.037	0.141	0.05	0.149	0.056	0.145	0.044	0.227	0.112	0.356	0.252	0.133	0.039	0.134	0.037
	24 h	0.166	0.063	0.18	0.075	0.183	0.078	0.176	0.067	0.241	0.116	0.364	0.259	0.167	0.055	0.159	0.05
	36 h	0.191	0.077	0.212	0.18	0.223	0.22	0.247	0.114	0.292	0.159	0.372	0.265	0.216	0.089	0.202	0.08
	48 h	0.106	0.22	0.122	0.238	0.134	0.245	0.235	0.113	0.3	0.17	0.374	0.266	0.219	0.101	0.223	0.097
	72 h	0.253	0.142	0.268	0.153	0.278	0.161	0.295	0.166	0.339	0.217	0.373	0.264	0.276	0.147	0.272	0.143
	96 h	0.269	0.157	0.282	0.168	0.296	0.179	0.282	0.172	0.33	0.207	0.377	0.268	0.291	0.172	0.301	0.174
	128 h	0.281	0.176	0.298	0.191	0.311	0.202	0.327	0.209	0.381	0.279	0.389	0.285	0.306	0.187	0.309	0.189
	256 h	0.336	0.251	0.349	0.264	0.362	0.268	0.349	0.25	0.387	0.29	0	0.37	0.353	0.284	0.38	0.29
	336 h	0.355	0.284	0.369	0.298	0.378	0.302	0.401	0.325	0.402	0.331	0.441	0.352	0.39	0.306	0.389	0.304

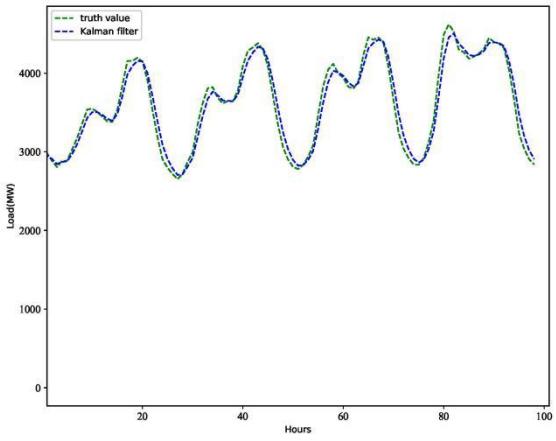

Fig. 7. The reconstructed data represent the data after VMD, and the Kalman filter represents the smooth data after Kalman filtering.

4.2 Prediction comparison based on Transformer architecture models

In this section, the Transformer [2], Autoformer [4], Informer [7], Fedformer [6], Pyformer [5], DLinear [19] and comparison.

models with direct multistep output to predict data at step sizes ranging from 12 h to 336 h.

The experimental results demonstrate that this model surpasses
the Transformer model and other algorithmic models in terms of prediction accuracy, with the smallest MAE and MSE errors. Notably, the Informer and Pyformer models outperform the WDE model when the output step size is between 48 h and 72 h. However, as the output step size increases, the WDE model exhibits superior performance.

4.3 Prediction comparison based on modal decomposition models

As depicted in Table 4, this section provides a comparison of
with other existing modal decomposition models. By comparing the MAE and MSE evaluation indices, it is demonstrated that WDE model prediction, following WOA-VMD and Kalman preprocessing, can address the nonlinear characteristics of power loads and enhance prediction accuracy.

TABLE 4 COMPARED WITH MODAL METHOD

Method		WDE		VMD-Transformer		VMD-GRU-TCN		VMD-Isomap-AdaBoost [11]		VMD-CISSA-LSSVM [12]	
Metric		MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
Electricity-JinShan	12 h	0.131	0.037	0.281	0.176	0.324	0.179	0.162	0.053	0.291	0.187
	24 h	0.166	0.063	0.306	0.201	0.311	0.202	0.183	0.081	0.299	0.194
	36 h	0.191	0.077	0.338	0.226	0.382	0.249	0.227	0.094	0.307	0.2
	48 h	0.106	0.22	0.248	0.364	0.37	0.248	0.235	0.105	0.309	0.201
	72 h	0.253	0.142	0.394	0.279	0.43	0.301	0.274	0.152	0.308	0.199
	96 h	0.269	0.157	0.408	0.296	0.417	0.307	0.265	0.142	0.312	0.203
	128 h	0.281	0.176	0.424	0.317	0.462	0.344	0.316	0.214	0.324	0.22
	256 h	0.336	0.251	0.475	0.39	0.484	0.385	0.322	0.225	0.371	0.305
	336 h	0.355	0.284	0.495	0.424	0.536	0.46	0.337	0.266	0.376	0.287

5 CONCLUSION

This paper presents a method that employs feature processing and hybrid modeling to enhance the prediction efficiency and accuracy. Initially, a random forest was applied for feature selection, followed by the use of the WOA-VMD and Kalman filter methods for data noise reduction. In the subsequent prediction phase, the method employs the WDE model, which demonstrates superior predictive performance. The effectiveness of this methodology is validated through evaluations on various datasets, showing its potential as an auxiliary tool in power grid operations.

Nevertheless, the method presented herein is not without its limitations, primarily due to the imperative of processing data in real time within the test set prior to executing predictions. Our future endeavors will focus on amalgamating advanced decomposition techniques with cutting-edge prediction models.

6 ACKNOWLEDGEMENTS

The spatiotemporal load characteristics of the power grid were modeled, and the new energy consumption potential was assessed based on the Graph Neural Network 52094022004C Research Project of the Shanghai Electric Power Artificial Intelligence Engineering Technology Research Center (19DZ2252800).

7 REFERENCES

[1] S. Qiu, T. Lei, J. Wu, and S. Bi, "Energy demand and supply planning of China through 2060," Energy, vol. 234, p. 121193, 2021.

[2] K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, and Y. Wang, "Transformer in transformer," Advances in Neural Information Processing Systems, vol. 34, pp. 15908-15919, 2021.

[3] C. Wang, Y. Wang, Z. Ding, T. Zheng, J. Hu, and K. Zhang, "A transformer-based method of multienergy load forecasting in integrated energy system," IEEE Transactions on Smart Grid, vol. 13, no. 4, pp. 2703-2714, 2022.

[4] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with autocorrelation for long-term series forecasting," Advances in Neural Information Processing Systems, vol. 34, pp. 22 419-22430, 2021.

[5] S. Liu, H. Yu, C. Liao, J. Li, W. Lin, A. X. Liu, and S. Dustdar, "Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting," in International conference on learning representations, 2021.

[6] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," in International Conference on Machine Learning. PMLR, 2022, pp. 27 268-27 286.

[7] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in Proceedings of the AAAI conference on artificial intelligence, vol. 35, no. 12, 2021, pp. 11 106-11 115.

[8] X. Nie, X. Zhou, Z. Li, L. Wang, X. Lin, and T. Tong, "Logtrans: Providing efficient local-global fusion with transformer and cnn parallel network for biomedical image segmentation," in 2022 IEEE 24th Int Conf on High Performance Computing & Communications; 8th Int Conf on Data Science & Systems; 20th Int Conf on Smart City; 8th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys). IEEE, 2022, pp. 769-776.

[9] P. Ran, K. Dong, X. Liu, and J. Wang, "Short-term load forecasting based on ceemdan and transformer," Electric Power Systems Research, vol. 214, p. 108885, 2023.

[10] C. Cai, Y. Li, Z. Su, T. Zhu, and Y. He, "Short-term electrical load forecasting based on vmd and gru-tcn hybrid network," Applied Sciences, vol. 12, no. 13, p. 6647, 2022.

[11] X. Han, J. Su, Y. Hong, P. Gong, and D. Zhu, "Mid-to long-term electric load forecasting based on the emd-isomap-AdaBoost model," Sustainability, vol. 14, no. 13, p. 7608, 2022.

[12] G. Wang, X. Wang, Z. Wang, C. Ma, and Z. Song, "A vmd-cissa-lssvm based electricity load forecasting model," Mathematics, vol. 10, no. 1, p. 28, 2021.

[13] W. Humphrey, A. Dalke, and K. Schulten, "Vmd: visual molecular dynamics," Journal of molecular graphics, vol. 14, no. 1, pp. 33-38, 1996.

[14] Q. Liu, Y. Shen, L. Wu, J. Li, L. Zhuang, and S. Wang, "A hybrid fcwemd and kf-ba-svm based model for short-term load forecasting," CSEE Journal of Power and Energy Systems, vol. 4, no. 2, pp. 226-237, 2018.

[15] J. Nasiri and F. M. Khiyabani, "A whale optimization algorithm (woa) approach for clustering," Cogent Mathematics & Statistics, vol. 5, no. 1, p. 1483565,2018.

[16] G. F. Welch, "Kalman filter," Computer Vision: A Reference Guide, pp. 1-3, 2020.

[17] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "Timesnet: Temporal 2d-variation modeling for general time series analysis," arXiv preprint arXiv:2210.02186, 2022.

[18] W. Fan, P. Wang, D. Wang, D. Wang, Y. Zhou, and Y. Fu, "Dish-ts: a general paradigm for alleviating distribution shift in time series forecasting," Proceedings of the AAAI Conference on Artificial

Intelligence, 2023.
[19] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" arXiv preprint arXiv:2205.13504, 2022.