Automation, Control and Intelligent Systems
2025, Vol. 13, No. 3, pp. 131-139
https://doi.org/10.11648/j.acis.20251303.16

Research Article

Research on Short-Term Electricity Consumption Forecasting Based on VMD-xLSTM

Chen Zhi[1], Han Zhong-kuan[1], Xun Chao[1], Shen Yu[1], Liu Lin[2], Chen Yan-tao[2], Zeng Shuai[3,*]

[1]State Grid Fujian Electric Power Co., Ltd., Fuzhou, China
[2]Institute of Economic and Technology, State Grid Fujian Electric Power Co., Ltd., Fuzhou, China
[3]Department of Electronic and Communication Engineering, North China Electric Power University, Baoding, China

Abstract

Enhancing the accuracy of electricity consumption forecasting is essential for the safe and stable operation of power systems and the strategic planning of grid companies. To achieve high-precision short-term electricity consumption forecasting, this paper proposes a VMD-xLSTM forecasting model that combines Variational Modal Decomposition (VMD) with an extended Long Short-Term Memory network (xLSTM). This novel hybird model firstly uses the VMD decomposition algorithm to decompose the original electricity consumption time series into multiple modal components and a residual component. Subsequently, leveraging the powerful long-term dependency capturing capabilities of the xLSTM neural network, each decomposed component is independently established the model and forecasted. Finally, the prediction results of each component are superimposed to reconstruct the overall predicted value of electricity consumption. Experimental results demonstrate that the proposed VMD-xLSTM model performs outstandingly, with a Symmetric Mean Absolute Percentage Error (SMAPE) of only 0.86%. This is notably lower than that of the traditional Long Short-Term Memory network (LSTM) at 2.33%, the standalone xLSTM model at 2.15%, and the VMD-LSTM model at 0.90%. Additionally, all other prediction error evaluation metrics of this model are comprehensively superior to the compared models, clearly verifying the VMD-xLSTM model's effectiveness in short-term electricity consumption forecasting tasks through comparative experiments.

Keywords

Electricity Consumption Forecasting, Extended Long Short-Term Memory Network, Variational Modal Decomposition

1. Introduction

Rapid economic growth is driving a continuous rise in electricity demand, challenging the safety and stability of our expanding power systems [1]. Accurate electricity forecasting is crucial for optimizing grid operations and ensuring system security. Therefore, enhancing prediction accuracy is essential for both refined power system management and its sustainable development.

The inherent high randomness and strong non-linear characteristics of short-term electricity consumption data pose a severe challenge to accurate forecasting. To address this, the

*Corresponding author: 2095058117@qq.com (Zeng Shuai)

Received: 27 August 2025; **Accepted:** 13 September 2025; **Published:** 25 September 2025

"decompose-forecast-reconstruct" strategy, inspired by digital signal processing, is widely applied. This method works by decomposing the complex original series into multiple, more stationary sub-series, forecasting each individually, and then reconstructing the results, thereby effectively reducing data complexity and enhancing prediction accuracy. Among the various decomposition algorithms, while Empirical Mode Decomposition (EMD) has its applications [2], it suffers from inherent drawbacks such as mode mixing and an uncontrollable number of decomposed modes. In contrast, Variational Mode Decomposition (VMD) enables active control over the decomposition process by presetting parameters like the number of modes (K) and the penalty factor (α) [3], effectively overcoming the shortcomings of EMD. VMD yields more stable and physically meaningful modal components. Given its superior decomposition performance and controllability, VMD has become a leading and effective time-series decomposition technique in the field of electricity forecasting.

Currently, short-term electricity forecasting methods are primarily classified into three categories: statistical models [4, 5], machine learning models [6, 8], and deep learning models [9-11]. Statistical models, represented by Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA) [4], and Seasonal ARIMA (SARIMA) [5], are characterized by their simple structures and computational efficiency. However, their underlying linear assumption hinders their ability to process non-linear electricity data effectively, thus limiting their scope of application. Machine learning models, such as Random Forest (RF) [6], Support Vector Machine (SVM) [7], and Least Squares Support Vector Regression (LSSVR) [8], can effectively handle non-linear data by leveraging their inherent parameter self-learning and non-linear adaptive capabilities. Nevertheless, these models commonly face bottlenecks, including a tendency to ignore the temporal nature of the data, slow training speeds, limited generalization capabilities, and difficulties in hyperparameter optimization.

Given the limitations of the aforementioned methods, deep learning has emerged as the mainstream approach in current research. By constructing deep and complex neural networks, these models can effectively mine profound features from data, exhibiting superior capabilities in non-linear fitting and data representation, particularly when processing long-sequence data. Among these, Recurrent Neural Networks (RNN) [9] and their variants—Long Short-Term Memory (LSTM) [10] and Gated Recurrent Unit (GRU) [11]—are the most widely applied representatives. However, these models are still hampered by drawbacks such as limited memory capacity and insufficient sensitivity in capturing long-term dependencies [12]. To address this challenge, the newly proposed Extended Long Short-Term Memory (xLSTM) [13] introduces innovative mechanisms like exponential gating, state normalization, and memory mixing. These innovations significantly enhance the model's ability to capture long-term dependencies and temporal correlations, demonstrating more powerful performance than traditional models in time-series forecasting tasks.

Based on the preceding research, this paper constructs a VMD-xLSTM model for short-term electricity consumption forecasting. The model firstly employs VMD algorithm to decompose the complex original electricity time series into multiple, more stationary modal components and a residual component. Subsequently, the xLSTM network is used to model and forecast each decomposed component independently. The final prediction is then reconstructed by summing these individual forecasts. Experimental results demonstrate that the proposed method achieves higher prediction accuracy compared to other baseline models, and ablation studies confirm the effectiveness of each of its modules.

2. Theoretical Basis of the Model

2.1. Variational Mode Decomposition (VMD)

The VMD algorithm is designed to overcome the inherent drawbacks of Empirical Mode Decomposition (EMD), such as mode mixing and endpoint effects.[14] As a completely non-recursive signal decomposition technique, the core idea of VMD is to construct and solve a constrained variational problem. It adaptively decomposes an original signal $f(t)$ into a predetermined number, K, of Intrinsic Mode Functions (IMFs), denoted as $u_k(t), k=1,2,\ldots,K$. Each IMF has a specific center frequency and a finite bandwidth, thereby effectively extracting key features of the signal and reducing the complexity of time series analysis. Unlike EMD, VMD addresses the issue of an uncontrollable number of generated modes found in algorithms like EMD by pre-setting the number of modes, K, and a penalty factor, α, to adjust the decomposition result.

The specific computational process of the VMD algorithm is as follows:

Step 1: Construct the constrained variational optimization problem as shown below:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad \text{s. t.} \sum_{k=1}^{K} u_k(t) = f(t) \quad (1)$$

Where, K is the preset number of decomposed modes; $u_k(t)$ is the k-th IMF; ω_k is the center frequency corresponding to the k-th IMF; $\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)$ is the one-sided spectrum of the mode $u_k(t)$ calculated using the Hilbert transform; and $e^{-j\omega_k t}$ is used to adjust the center frequency of each mode.

Step 2: Transform the original constrained problem into an unconstrained optimization problem for iterative solution by introducing a quadratic penalty factor α and a Lagrange multiplier $\lambda(t)$. The augmented Lagrangian function is expressed as:

$$L = \underbrace{\alpha\sum_{k=1}^{K}\left\|\partial_t[\cdots]e^{-j\omega_k t}\right\|_2^2}_{} + \underbrace{\left\|f(t)-\sum_k u_k(t)\right\|_2^2}_{} + \underbrace{\left\langle \lambda(t), f(t)-\sum_k u_k(t)\right\rangle}_{} \quad (2)$$

Step 3: Use the Alternate Direction Method of Multipliers (ADMM) to iteratively update the modes u_k, center frequencies ω_k, and the Lagrange multiplier λ in the frequency domain. The iterative update expressions for each variable are as follows:

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

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

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau\left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)\right) \quad (5)$$

Where: n is the iteration number, ω is the frequency, and $\hat{u}_k(\omega)$, $\hat{\lambda}^{n+1}(\omega)$, and $\hat{f}(\omega)$ are the Fourier transforms of $u_k(t)$, $\lambda(t)$, and $f(t)$, respectively.

Step 4: Set the termination condition for the iteration in Step 3 as follows:

$$\sum_{k=1}^{K}\frac{\|\hat{u}_k^{n+1}-\hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \epsilon \quad (6)$$

Where: the threshold ϵ is typically set to 10^{-6}. If this condition is met, the algorithm is considered to have converged; otherwise, it returns to Step 3 to continue iterating.

Finally, perform an inverse Fourier transform on the converged $\hat{u}_k(\omega)$ to obtain the final output of each mode as follows:

$$u_k(t) = \mathcal{F}^{-1}\left[\hat{u}_k^{\text{final}}(\omega)\right], k=1,2,\ldots,K \quad (7)$$

2.2. Extended Long Short-Term Memory Network (xLSTM)

Long Short-Term Memory network (LSTM) was proposed by Hochreiter and Schmidhuber in the 1990s. This neural network was designed with three gating mechanisms—the input gate (i_t), the forget gate (f_t), and the output gate (o_t)—along with a cell state (c_t), to address the problems of gradient vanishing and explosion in traditional Recurrent Neural Networks (RNNs) when modeling long sequences [15]. The core operational mechanisms of LSTM are described by the following equations:

$$i_t = \sigma\left(W_i\cdot[h_{t-1},x_t]+b_i\right) \quad (8)$$

$$f_t = \sigma\left(W_f\cdot[h_{t-1},x_t]+b_f\right) \quad (9)$$

$$o_t = \sigma\left(W_o\cdot[h_{t-1},x_t]+b_o\right) \quad (10)$$

$$c_t = f_t\odot c_{t-1}+i_t\odot\tilde{c}_t,\quad \tilde{c}_t = \tanh\left(W_C\cdot[h_{t-1},x_t]+b_C\right) \quad (11)$$

$$h_t = o_t\odot\tanh(c_t) \quad (12)$$

There, σ is the Sigmoid activation function. The cell state c_t is used to store long-term information. The input gate i_t, forget gate f_t, and output gate o_t are responsible for controlling the storage of new information, the retention of historical information, and the amount of information output from c_t at the current state, respectively. The hidden state h_t is the output at the current time step. The internal structure of the LSTM hidden layer is shown in Figure 1.

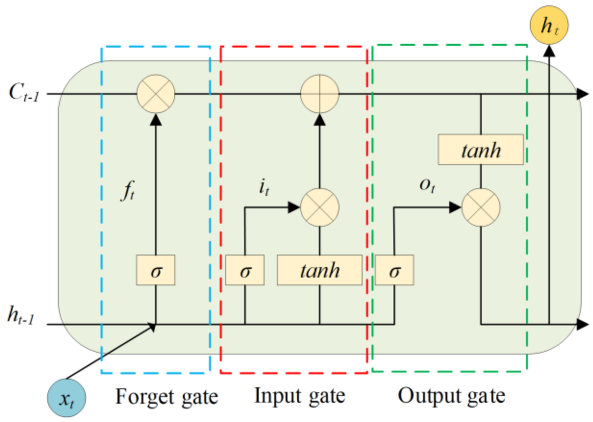

Figure 1. *Internal Structure of the LSTM Hidden Layer.*

Based on the mechanisms described above, LSTM can effectively capture long-term dependencies, making it one of the classic models in the field of time series forecasting. However, as research has deepened and application scenarios have become more complex, traditional LSTM has revealed several limitations when processing large-scale long-sequence data:

1) The value range of the Sigmoid gating function in LSTM is restricted to, which limits the gate's modulation capability. Consequently, when the input gate i_t controls the information stored into the cell state c_t, it is difficult

to revise historical information dynamically based on new information.

2) LSTM uses a scalar memory storage unit, which restricts its information storage capacity.

3) LSTM lacks parallel computing capabilities.

To overcome these bottlenecks, Maximilian Beck's team proposed xLSTM. The xLSTM architecture includes two novel types of LSTM memory units: scalar Long Short-Term Memory (sLSTM), which employs scalar memory and a memory mixing mechanism, and matrix Long Short-Term Memory (mLSTM), which uses matrix memory and covariance updates and supports parallelization. The key improvements in the sLSTM unit compared to LSTM are as follows:

1) Introduce the exponential gating to replace the sigmoid gating function of the input gate (optional for forgetting gate) in the original LSTM network, giving the gating function a larger output range and greater gradient. This not only alleviates the gradient disappearance problem effectively, but also makes the input gate and forget gate more sensitive to changes in input information, enabling more accurate capture of trend changes and sudden signals in time series, thereby more accurately correcting memory decisions.

2) To address potential numerical instability caused by exponential gating, sLSTM designs a normalizer state n_t and a stabilizer state m_t. These are used to constrain the range of gate values and prevent numerical overflow caused by excessively large exponential gate values, thereby avoiding the gradient explosion problem and stabilizing the gating mechanism.

The expressions for the gates and states within the sLSTM unit are as follows:

$$i_t = \exp\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{13}$$

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \text{ or } exp\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{14}$$

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{15}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot z_t, \quad z_t = \tanh\left(W_z \cdot \left[h_{t-1}, x_t\right] + b_z\right) \tag{16}$$

$$n_t = f_t \odot n_{t-1} + i_t \tag{17}$$

$$h_t = o_t \odot c_t \odot n_t^{-1} \tag{18}$$

There, σ is the Sigmoid activation function. The gating function for the input gate i_t is changed to an exponential gate. The forget gate f_t can optionally use the exponential gate to adapt to different application scenarios. The cell state c_t and output gate o_t remain unchanged. The normalizer state n_t is used to normalize the cell state within the hidden state, and $c_t \odot n_t^{-1}$ can be regarded as the normalized cell state.

Furthermore, the stabilizer function m_t is used to stabilize the exponential gates in the input and forget gates, preventing potential numerical overflow issues that might arise from the exponential activation function. The relevant calculation formulas are as follows:

$$m_t = \max\left(\log\left(f_t\right) + m_{t-1}, \log\left(i_t\right)\right) \tag{19}$$

$$i_t^{'} = \exp\left(\log\left(i_t\right) - m_t\right) \tag{20}$$

$$f_t^{'} = \exp\left(\log\left(f_t\right) + m_{t-1} - m_t\right) \tag{21}$$

Here, $f_t^{'}$ and $i_t^{'}$ are used to replace the original forget and input gate functions in sLSTM during the forward pass. It has been mathematically proven that the updated forget and input gate functions can stabilize the gating calculations without changing the overall network output or affecting the gradient computation. The xLSTM architecture discussed in this paper uses sLSTM as its internal unit. The internal structure of the xLSTM hidden layer is shown in Figure 2.

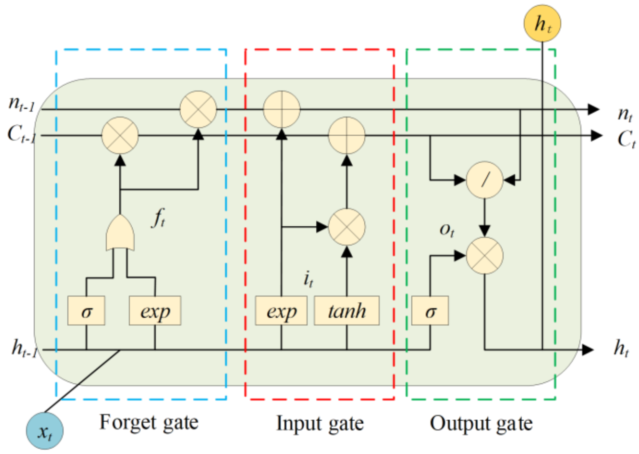

Figure 2. *Internal Structure of the xLSTM Hidden Layer.*

3. Short-term Electricity Consumption Forecasting Process Based on VMD-xLSTM

Based on the theoretical foundation of the above algorithms, this paper proposes a short-term electricity consumption prediction model based on VMD-xLSTM. The overall prediction process of this model is shown in Figure 3.

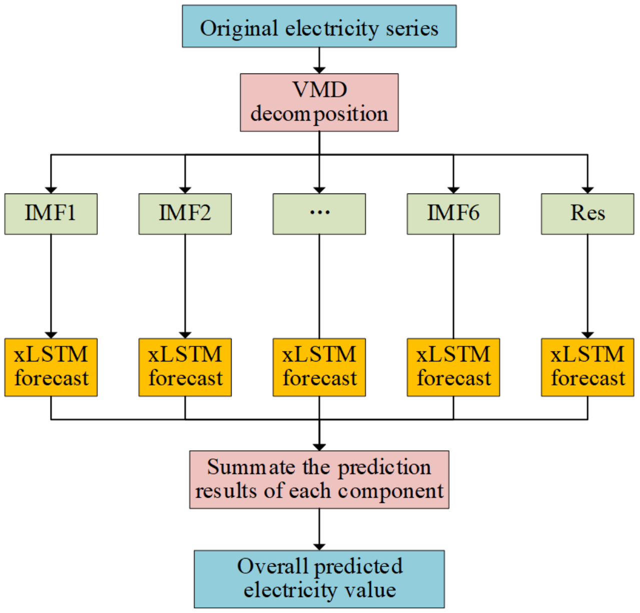

Figure 3. *The overall prediction process of the model.*

The specific prediction steps of this model are as follows:

Step 1: Use the VMD decomposition algorithm to decompose the original electricity time series. Based on experimental experience, set the number of modal components (K) to 6, and then decompose the original sequence into a combination of simple modal components (IMF1 to IMF6) and a residual component (Res) to reduce the complexity of the electricity data.

Step 2: Use the xLSTM model to establish separate forecasting models for the different decomposed components and complete the prediction work for each component.

Step 3: Combine and reconstruct the prediction results of each modal component to obtain the overall prediction value for the total electricity consumption data.

4. Case Analysis

The electricity consumption dataset used in this experiment is sourced from daily electricity consumption data for the entire province of a southern province in China, covering the period from January 1, 2019, to July 31, 2024, with a total of 2,038 data points. The dataset contains no outliers or missing values. To validate the predictive performance of the model proposed in this paper, the dataset was divided into a training set and a test set, with a ratio of 8:2.

In the experiment, the sliding window width was set to 8, meaning that data from the previous 8 days was used to predict data for the next day. The number of training epochs was set to 200, and the initial learning rate was set to 0.01. A fixed random seed was designed for each group of experiments to ensure fairness.

4.1. Error Evaluation Index

To assess the accuracy of the model's prediction results, this experiment selected five error evaluation metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE) to evaluate the model's performance. The lower the error calculated by these evaluation metrics, the higher the accuracy of the model's predictions and the better its performance. The formulas for each metric are as follows:

$$E_{\mathrm{RMSE}} = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2\right)} \tag{22}$$

$$E_{\mathrm{MAE}} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \tag{23}$$

$$E_{\mathrm{MAPE}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - y_i\right|}{\left|y_i\right|}\times 100\% \tag{24}$$

$$E_{\mathrm{SMAPE}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - y_i\right|}{\left(\left|y_i\right| + \left|y_i\right|\right)/2}\times 100\% \tag{25}$$

Where, y_i is the actual value, $\hat{y}_i$ is the model prediction value, and n is the number of observation points in the test set. Among the various metrics, the units for RMSE, MAE are the same as those used for electricity consumption data in the dataset, namely 1000*KWh, while the units for MAPE and SMAPE are percentages.

4.2. VMD Decomposition Results

In this experiment, the VMD algorithm was used to decompose the original electricity time series. Based on experimental experience, the number of modal components (K) was set to 6, and the penalty factor (α) was set to 2000. The original series was decomposed into a combination of simple modal components (IMF1 to IMF6) and the residual component (Res). The decomposition results of each component are shown in Figure 4:

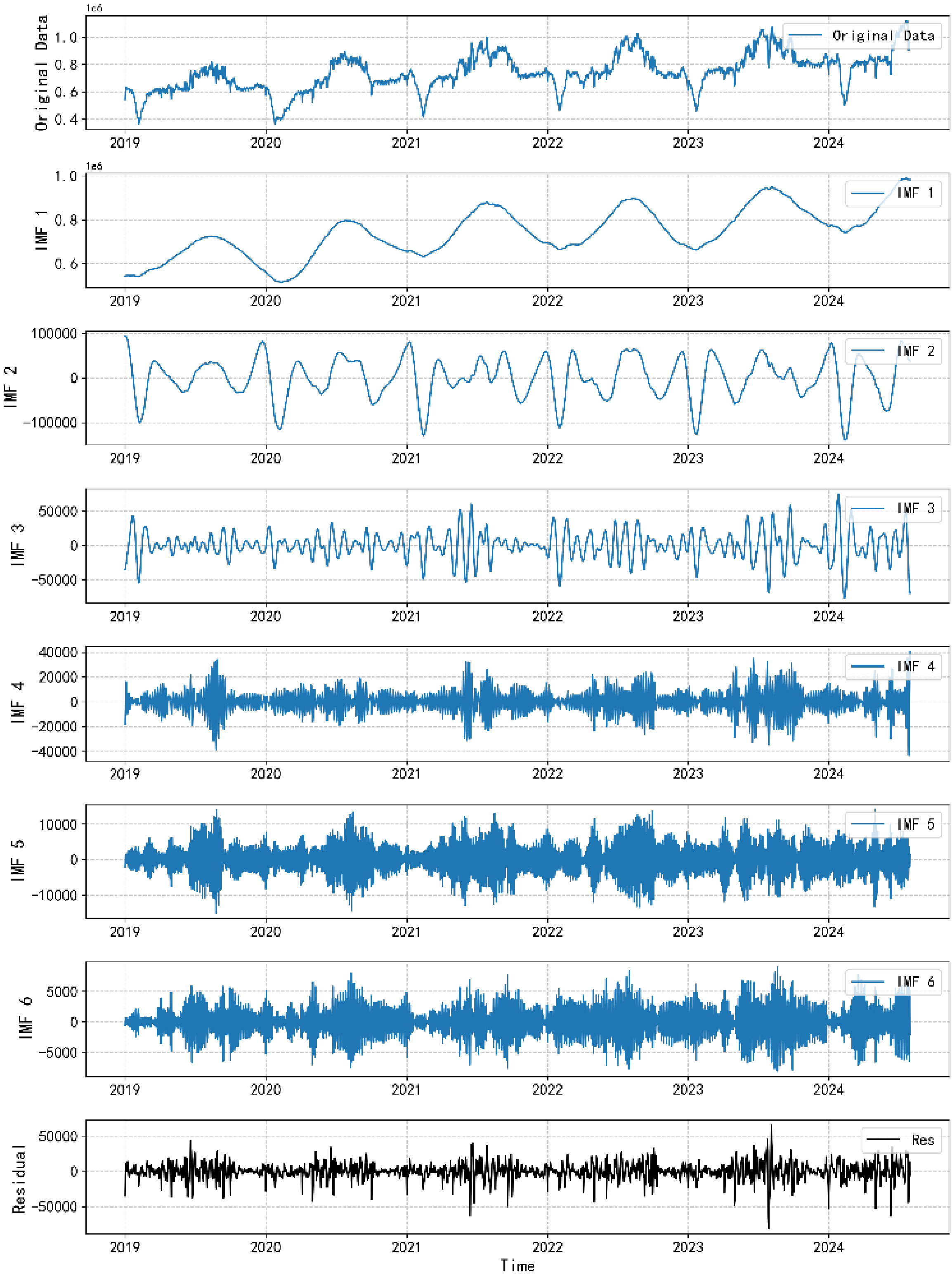

Figure 4. *VMD decomposition results diagram of original data.*

As shown in Figure 4, the original data is decomposed into six modal components (IMF1-IMF6), arranged sequentially from the lowest to the highest frequency, and one residual component (Res). The low-frequency components (IMF1, IMF2) clearly capture the macroscopic trend and long-term periodic features of the original sequence. In contrast, the high-frequency components (IMF3-IMF6) meticulously delineate the short-term fluctuations, local abrupt changes, and other high-frequency details within the series. The residual component (Res) exhibits the characteristics of random noise with no discernible pattern, representing the random disturbances in the signal that are difficult to model. This decomposition method successfully disentangles the signal's trend, periodicity, and randomness, thereby reducing the data's complexity and laying the foundation for the subsequent work.

4.3. Comparison and Analysis of Experimental Results

In order to systematically evaluate the effectiveness of the VMD-xLSTM model proposed in this paper, we designed a series of comparative experiments and ablation experiments.

Overall comparison experiment

The comparative experiment compared the VMD-xLSTM model with the LSTM, xLSTM, and VMD-LSTM models, all of which performed short-term forecasting tasks on the experimental dataset mentioned above. The prediction results of each model were compared with the actual data values. Figure 5 shows a comparison of the prediction results of each model with the actual values.

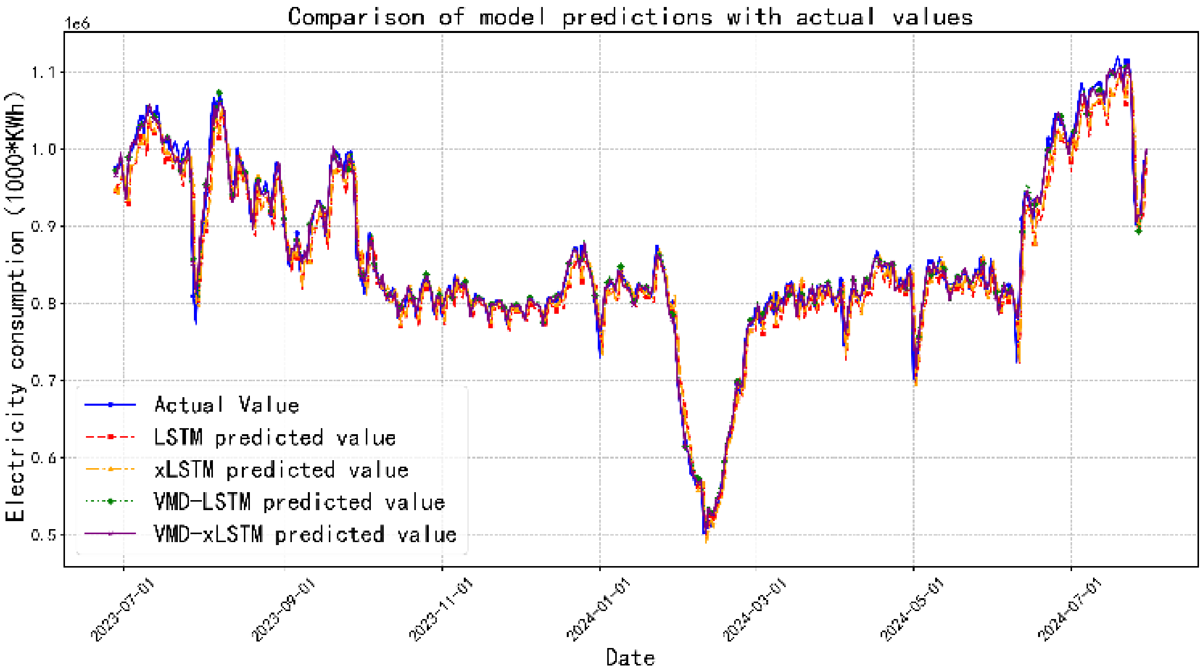

Figure 5. *Comparison of model predictions with actual value.*

As can be seen from the prediction results curve in Figure 5, the VMD-xLSTM model's prediction curve closely matches the actual value curve in terms of overall trend and local details, intuitively demonstrating its superior prediction capabilities compared to the other three models.

To evaluate the model's performance, this paper selects five error metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE). These metrics serve to quantitatively assess the model's predictive accuracy and provide a data-driven basis for performance comparison among the different models. Among these indicators, the units for RMSE, MAE, and MSE are the same as those of the electricity consumption data in the dataset (1000*kWh), while MAPE and SMAPE are expressed as percentages. The calculated results for each metric are presented in Table 1.

Table 1. *Error metrics values of different models.*

Model	RMSE	MAE	MAPE	SMAPE
LSTM	27471.67	19925.56	2.32%	2.32%
xLSTM	26229.34	18432.50	2.15%	2.15%
VMD-LSTM	10819.03	7565.13	0.90%	0.90%

Model	RMSE	MAE	MAPE	SMAPE
VMD-xLSTM	10463.48	7310.94	0.86%	0.86%

As the data in Table 1 indicates, the VMD-xLSTM model achieves the lowest values across all error metrics among the compared models. For instance, its SMAPE of 0.86% is significantly lower than all other models, proving its superior prediction accuracy.

To further validate the individual contributions of the VMD decomposition and the xLSTM network to the model's performance, this paper conducts the following ablation experiments based on the results in Table 1.

Ablation Experiment 1: To verify the effectiveness of the VMD decomposition algorithm, this study compared the performance of "decompose-forecast" hybrid counterparts against their standalone models by conducting two sets of comparisons: LSTM vs. VMD-LSTM, and xLSTM vs. VMD-xLSTM. The results clearly show that after integrating the VMD algorithm, the hybrid models exhibit lower error metrics across the board, indicating a significant improvement in prediction accuracy. The SMAPE of VMD-LSTM decreased by 1.43% compared to the standalone LSTM model, while VMD-xLSTM's SMAPE decreased by 1.29% compared to xLSTM. This fully demonstrates that by decomposing a complex series into more stationary sub-series, VMD effectively reduces data non-stationarity and noise interference, thereby significantly enhancing the model's predictive performance.

Ablation Experiment 2: To prove the superiority of xLSTM networks over LSTM, this study confirmed that replacing the traditional LSTM with xLSTM further boosts prediction accuracy by comparing the results of (LSTM vs. xLSTM) and (VMD-LSTM vs. VMD-xLSTM). The SMAPE of xLSTM was 0.18% lower than that of LSTM. Similarly, the SMAPE of VMD-xLSTM (0.86%) was also superior to that of VMD-LSTM (0.90%). This confirms that xLSTM, with its improved gating and memory mechanisms, more effectively captures long-range dependencies and critical abrupt changes within the time series.

In summary, whether considering the curve fitting of the predictions or the quantitative error metrics, the proposed VMD-xLSTM model consistently demonstrates optimal performance. The ablation studies also clearly prove the crucial roles of both the VMD decomposition and the xLSTM network in enhancing the model's overall prediction accuracy, validating the advanced nature and effectiveness of this hybrid model.

5. Conclusions

To achieve high-precision short-term electricity forecasting, this paper proposed a model based on VMD-xLSTM. The advantages of this model are twofold:

1) The model utilizes the VMD algorithm to decompose the original power load series, which effectively mitigates the data's non-linearity and high volatility, thereby reducing subsequent modeling complexity.
2) It leverages the superior performance of the xLSTM network in capturing long-term temporal dependencies and storing information to forecast each component accurately, ultimately enhancing overall prediction accuracy.

The experimental results fully demonstrate the effectiveness and advanced nature of the proposed model. The VMD-xLSTM model performed optimally across multiple error evaluation metrics, including RMSE, MAE, MAPE, and SMAPE. Furthermore, through ablation studies, we confirmed that both key modules—VMD decomposition and the xLSTM network—contributed to the enhancement of the model's overall performance.

Despite the promising results achieved by this model, there remains room for improvement:

1) The current study shows that the model's forecasting performance on the residual component is suboptimal. Further decomposition of the residual to extract more effective information could enhance the overall prediction accuracy.
2) The current research is based on a univariate time-series forecast using only historical electricity consumption data. In reality, electricity consumption is influenced by a combination of climatic factors (e.g., temperature) and socio-economic factors (e.g., holidays and policies). Therefore, future research can focus on developing a multivariate forecasting model. By incorporating external features such as meteorological data and socio-economic indicators, it may be possible to uncover deeper underlying correlations and achieve further improvements in prediction accuracy.

Abbreviations

VMD	Variational Modal Decomposition
xLSTM	extended Long Short-Term Memory Network
SMAPE	Symmetric Mean Absolute Percentage Error
LSTM	Long Short-Term Memory Network
EMD	Empirical Mode Decomposition
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Moving Average
SARIMA	Seasonal ARIMA
RF	Random Forest
SVM	Support Vector Machine
LSSVR	Least Squares Support Vector Regression

RNN Recurrent Neural Networks
GRU Gated Recurrent Unit

Author Contributions

Chen Zhi: Conceptualization, Methodology, Writing – original draft
Han Zhong-kuan: Formal analysis, Methodology, Writing – review & editing
Xun Chao: Formal analysis, Funding acquisition, Writing – review & editing
Shen Yu: Conceptualization, Software, Supervision
Liu Lin: Supervision, Validation
Chen Yan-tao: Investigation, Data curation
Zeng Shuai: Data curation, Visualization

Funding

This work was financially supported by the consulting service commissioned project of State Grid Fujian Electric Power Co., Ltd (SGFJ0000FZWT2500219).

Data Availability Statement

The data is available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Seasonal Least Squares Support Vector Machine with Fruit Fly Optimization Algorithm in Electricity Consumption Forecasting [J]. Journal of Donghua University (English Edition), 2019, 36(1): 67-76. https://doi.org/10.3969/j.issn.1672-5220.2019.01. 010

[2] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis [J]. Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences, 1998, 454(1971): 903-995. https://doi.org/10.1098/rspa.1998.0193

[3] Dragomiretskiy K, Zosso D. Variational mode decomposition [J]. IEEE transactions on signal processing, 2013, 62(3): 531-544. https://doi.org/10.1109/TSP.2013.2288675

[4] RAMOS P, SANTOS N, RUI R. Performance of state space and ARIMA models for consumer retail sales forecasting [J]. Robotics and Computer-Integrated Manufacturing, 2015, 34: 151-163. https://doi.org/10.1016/j.rcim.2014.12.015

[5] Liu Xiaolei, Lin Zi, Feng Ziming. Short-term offshore wind speed forecast by seasonal ARIMA - A comparison against GRU and LSTM [J]. Energy, 2021, 227: 1-12. https://doi.org/10.1016/j.energy.2021.120492

[6] Huang Nantian, Wang Wenting, Wang Sining, Wang Jun, Cai Guowei, Zhang Liang. Incorporating Load Fluctuation in Feature Importance Profile Clustering for Day-Ahead Aggregated Residential Load Forecasting [J]. IEEE Access, 2020, 8: 25198-25209. https://doi.org/10.1109/ACCESS.2020.2971033

[7] PAPAIOANNOU, GEORGE P., DIKAIAKOS, CHRISTOS, DRAMOUNTANIS, ANARGYROS, et al. Analysis and Modeling for Short- to Medium-Term Load Forecasting Using a Hybrid Manifold Learning Principal Component Model and Comparison with Classical Statistical Models (SARIMAX, Exponential Smoothing) and Artificial Intelligence Models (ANN, SVM): The Case of Greek Electricity Market [J]. Energies, 2016, 9(8): 635. https://doi.org/10.3390/en9080635

[8] YU, LEAN, XU, HUIJUAN, TANG, LING. LSSVR ensemble learning with uncertain parameters for crude oil price forecasting [J]. Applied Soft Computing, 2017, 56692-701. https://doi.org/10.1016/j.asoc.2016.09.023

[9] G. Petneházi. Recurrent Neural Networks for Time Series Forecasting. arXiv, Dec. 31, 2018. https://doi.org/10.48550/arXiv.1901.00069

[10] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory [J]. Neural Computation, 1997, 9(8): 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[11] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder decoder for statistical machine translation [J]. arXiv: 1406. 1078, 2014. https://doi.org/10.48550/arXiv.1406.1078

[12] JIAWEN TIAN, SHUQIANG WANG. Tool wear prediction based on xLSTM combined with the ECA attention mechanism model [J]. Journal of Mechanical Science and Technology, 2025, 39(2): 519-530. https://doi.org/10.1007/s12206-024-1211-9

[13] Beck, M., Pöppel, K., Spanring, M., et al. xLSTM: Extended Long Short-Term Memory. 2024, arXiv preprint arXiv: 2405.04517. https://doi.org/10.48550/arXiv.2405.04517

[14] Lalitpat A, Warut P, Jirachai B, et al. A Hybrid Model of VMD-EMD-FFT, Similar Days Selection Method, Stepwise Regression, and Artificial Neural Network for Daily Electricity Peak Load Forecasting [J]. Energies, 2023, 16(4): 1860-1860. https://doi.org/10.3390/EN16041860

[15] Hu Z, Gao Y, Sun L, et al. A novel attention-enhanced LLM approach for accurate power demand and generation forecasting [J]. Renewable Energy, 2025, 252123465-123465. https://doi.org/10.1016/J.RENENE.2025.123465