International Journal of Electric Power and Energy Studies
ISSN: 3006-2004 (Print), ISSN: 3006-0826 (Online) | Volume 5, Number 2, Year 2026
DOI: https://doi.org/10.62051/ijepes.v5n2.03
Journal homepage: https://wepub.org/index.php/IJEPES/index

Research on Power Load Forecasting Technology based on Time Series

Xiangyu Wang

Northeast Petroleum University, Qinhuangdao 066004, China

ABSTRACT

Multivariate, multi-step power load forecasting represents a classic complex time series modeling challenge, with its core difficulty lying in simultaneously capturing long-term dependencies and local non-stationary fluctuations. In this paper, the calculations of MSE and MAE are performed using standardized data scales. Consequently, the experimental results presented in subsequent chapters represent dimensionless errors on a standardized scale, rather than absolute errors measured in physical units such as megawatts. Thus, subtle numerical differences among models under this evaluation protocol remain clearly comparable.

KEYWORDS

Power Load Forecasting Technology; Power Emergency Repair Vehicles; Time Series.

1. INTRODUCTION

Multivariate, multi-step power load forecasting represents a classic complex time series modeling challenge, with its core difficulty lying in simultaneously capturing long-term dependencies and local non-stationary fluctuations. Under long-range forecasting conditions, the model must not only exhibit robust global trend modeling capabilities but also effectively mitigate error accumulation while enhancing responsiveness to local disturbances. Therefore, developing a forecasting model that combines efficient long-range sequence modeling capabilities with accurate local error characterization is pivotal for addressing such problems [1–2].

In light of this, this paper adopts a modeling approach that combines deep backbone prediction with residual learning correction. Centering on this framework, this chapter provides a theoretical analysis of the definition of multivariate, multi-step prediction problems, deep sequence modeling techniques, and ensemble learning models, with a focus on elucidating the role of these methods in long sequence modeling and residual compensation.

2. TIME SERIES PREDICTION OF ELECTRIC POWER LOAD

2.1. Multivariate Multi-step Prediction Problem

The time series is first sampled at equal intervals, with indexing performed at L integer time steps. Given a historical H input window of t length h and a prediction step length of l, the joint multivariate historical observation sequence at the reference time point can be expressed as in Formula (1):

$$\boldsymbol{X}_{t-L+1:t} = \{\boldsymbol{x}_{t-L+1}, \boldsymbol{x}_{t-L+2}, \ldots, \boldsymbol{x}_t\} \tag{1}$$

The formula $x_i \in \mathbb{R}^d$ $x_i \in \mathbb{R}^d$ i denotes i the input feature vector at the ith time step. This vector typically consists of multi-source information including historical load values, meteorological $X_{t-L+1:t}$ $X_{t-L+1:t}$ variables, and time f encoding f H. The H objective of the multivariate multi-step forecasting task is to establish a parametric mapping function given a historical input sequence, producing predictions for consecutive future time steps, as expressed in Formula (2).

$$\hat{Y}_{t+1:t+H} = f\left(\boldsymbol{X}_{t-L+1:t}\right) \tag{2}$$

The formula $\hat{Y}_{t+1:t+H} \in \mathbb{R}^{H \times m}$ represents H the joint prediction m result of the system's multi-dimensional $m > 1$ target variables over the next number of time steps. At this stage, the model must simultaneously characterize the evolutionary patterns of multiple correlated nodes or target components within a unified framework.

Multivariate multi-step forecasting poses greater challenges than univariate single-step forecasting. Firstly, the input comprises heterogeneous information from multiple sources; the model must learn dynamic correlations among variables over extended historical windows and extract stable temporal dependency structures. Secondly, the output spans consecutive future time steps that exhibit both temporal constraints and strong interdependencies. Consequently, this task cannot be simply decomposed into independent regression problems but requires a unified model that balances long-term evolutionary coherence with consistent multi-step outputs [3]. Figure 1 illustrates the input-output relationship for the multivariate multi-step power load forecasting task.

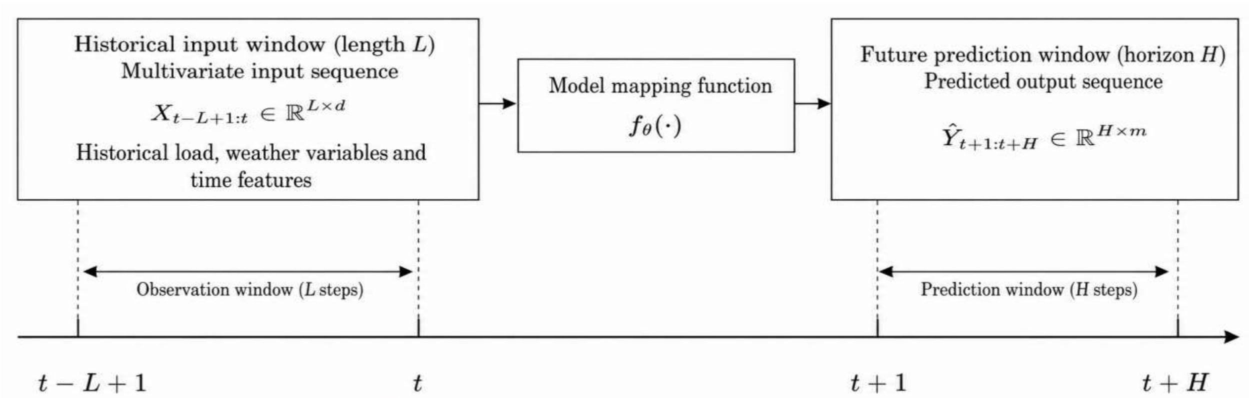

Figure 1. Schematic illustration of the input-output relationship in multivariate multi-step power load forecasting

Therefore, multivariate multi-step prediction tasks require models not only to capture global temporal dependencies but also to collaboratively model future multi-step outputs within a unified framework, imposing higher demands on both model architecture design and error control mechanisms [4].

2.2. Multi-step Inference Strategy and Error Cumulation Analysis

To achieve the mathematical mapping objective of the aforementioned multivariate, multi-step forecasting, existing research typically employs three fundamental strategies at the output end:

recursive forecasting, direct forecasting, and multiple-input multiple-output (MIMO).Recursive forecasting involves training a single-step forecasting model and iteratively using the prediction from the previous time step as input for the subsequent time step during inference; its formulation can be summarized in Equation (3):

$$\hat{y}_{t+h} = f_1\left(\boldsymbol{X}_{t-L+h:t+h-1}\right), h = 1,2,\ldots,H \tag{3}$$

This strategy features a simple structure and low training cost, but its primary limitation lies in the fact that prediction errors are progressively transmitted and accumulated during the recursive process—particularly under long-term forecasting conditions, which can easily cause the predicted trajectory to deviate from the true distribution. In contrast, the Direct Forecasting strategy establishes independent mapping functions for each future prediction step, as represented by Formula (4):

$$\hat{y}_{t+h} = f_h\left(\boldsymbol{X}_{t-L+1:t}\right), h = 1,2,\ldots,H \tag{4}$$

This strategy avoids explicit error propagation in recursive inference, thereby mitigating the error accumulation problem to some extent. However, since different prediction steps are typically modeled independently, the method struggles to fully leverage the correlations between future time steps within a unified framework; moreover, as the prediction step length increases, the model training cost also rises correspondingly. Multi-input multi-output (MIMO) prediction requires the model to directly output a complete future prediction sequence, which can be expressed as Formula (5):

$$\hat{Y}_{t+1:t+H} = f\left(\boldsymbol{X}_{t-L+1:t}\right) \tag{5}$$

The MIMO strategy avoids the explicit error accumulation inherent in recursive prediction while enabling the joint modeling of correlations across multiple future time steps within a unified $\boldsymbol{H}$ $\boldsymbol{H}$ framework, making it more suitable for multivariate, multi-step prediction tasks. It is important to note that when the prediction step size is large, even with the MIMO architecture, the model's ability to accurately characterize distant targets may diminish, leading to phenomena such as increased prediction errors in later stages, phase shifts, or local distortions. To facilitate a more intuitive comparison of the characteristics of different multi-step prediction strategies,

3. DEEP LEARNING SEQUENCE MODELING FOR POWER LOAD PREDICTION

The task of multi-variable, multi-step power load forecasting typically exhibits characteristics such as long-term dependencies, multivariable coupling, and local non-stationary fluctuations. Compared to traditional statistical models, deep learning approaches leverage multi-layer nonlinear mappings and high-dimensional feature representation capabilities, demonstrating greater adaptability in complex time series modeling. From a technological development perspective, deep sequence modeling has evolved through stages ranging from recurrent neural networks and attention mechanisms to efficient long sequence modeling architectures.

To meet the needs of subsequent model design in this paper, this section does not provide a comprehensive review of various models but instead summarizes the technical foundations directly relevant to Chapters 3 and 4. It focuses on introducing recurrent neural networks and gating architectures, sequence modeling methods based on attention mechanisms, representative architectures for long sequence modeling, as well as the underlying concepts of state-space models and xLSTM, thereby providing a theoretical foundation for constructing subsequent hybrid prediction models.

3.1. Recurrent Neural Networks and Gate Structures

Recurrent Neural Networks (RNNs) are a class of recursive networks designed for processing sequential data. Their fundamental principle involves sharing parameters across time dimensions and $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T\}$ transmitting historical information between adjacent time steps through hidden states. Assuming the input sequence is represented by, the hidden state update of a standard recurrent unit can be expressed as Formula (6).

$$\boldsymbol{h}_t = \phi\left(\boldsymbol{W}_{xh}\boldsymbol{x}_t + \boldsymbol{W}_{hh}\boldsymbol{h}_{t-1} + \boldsymbol{b}_h\right) \tag{6}$$

In the formula $\boldsymbol{x}_t$, represents the input features $\boldsymbol{h}_t$ at the current $\boldsymbol{W}_{xh}$ sampling $\boldsymbol{W}_{hh}$ time, denotes the system's $\phi(\cdot)$ hidden state, and $\phi(\cdot)$ are the weight matrices corresponding to feature mapping and state recurrence, respectively; is the nonlinear activation function. Based on this architecture, the model enables dynamic modeling of sequence data across the time dimension.

However, conventional RNNs are prone to gradient disappearance or gradient explosion during long sequence modeling, resulting in unstable learning of long-range dependencies [5]. In power load forecasting scenarios, load sequences typically contain multi-scale periodic information spanning daily and weekly cycles; thus, inadequate long-range dependency modeling directly impacts prediction performance.

To address this issue, the Long Short-Term Memory (LSTM) architecture incorporates cell states and gating mechanisms into traditional recurrent structures. By employing forgetting gates, input gates, and output gates to regulate information retention, updating, and output, it enhances the model's capacity to retain long-term dependent information [5]. Compared to conventional RNNs, LSTM demonstrates superior stability in modeling long sequences, making it particularly well-suited for complex time-series tasks such as power load forecasting.

As shown in Figure 2, RNN primarily relies on the recursive transmission of information across hidden states in the temporal dimension, whereas LSTM introduces cell states and gating mechanisms to provide a more stable pathway for long-term information retention. Relevant studies indicate that deep models based on recurrent gating structures achieve superior performance in short-term load prediction tasks. Moreover, the fundamental concept of LSTM's long-term memory modeling offers valuable insights for the subsequent development of extended long-short-term memory networks (xLSTM).

Although models based on recurrent architectures perform well in short-term load forecasting, they remain susceptible to gradient decay and error accumulation in long sequence modeling, limiting their effectiveness for long-term prediction tasks. This limitation has driven the subsequent development of attention mechanisms and efficient long-sequence modeling methods.

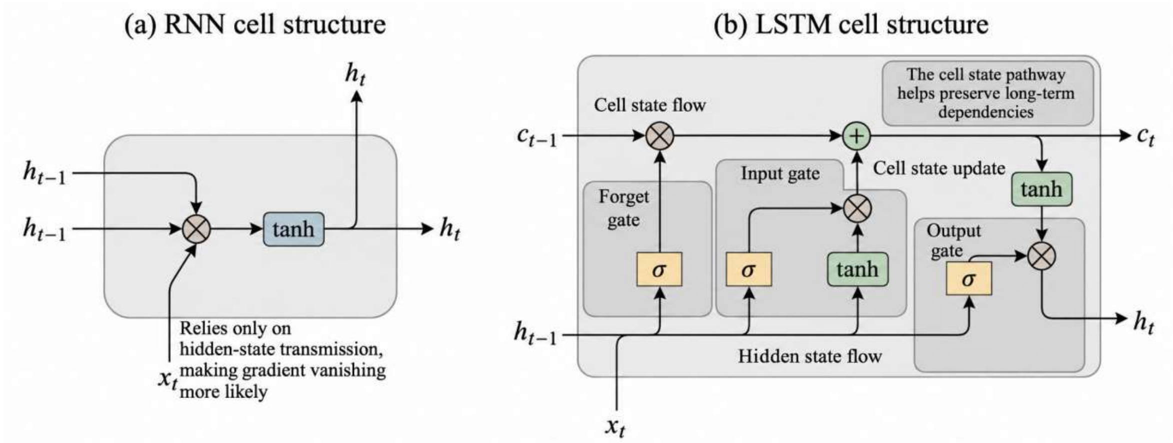

Figure 2. Schematic diagram of information flow mechanisms

3.2. Attention Mechanism and Transformer Sequence Modeling

The Attention Mechanism transforms the traditional sequential model's modeling approach, which primarily relies on recursive methods to transmit historical information. Its core principle involves assigning distinct weights to different positions in the input sequence based on the current task objective, thereby highlighting the most critical time segments for prediction. The Self-Attention mechanism further Q models K global V dependencies by explicitly calculating correlations between positions within the sequence. Assuming the input sequence undergoes a linear transformation to yield the query matrix, key matrix, and value matrix, the computational formulation for the standard scaled dot-product self-attention is given by Formula (7):

$$\mathrm{Attn}(Q,K,V)=\mathrm{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \tag{7}$$

In the formula d_k, denotes the dimension of the $\sqrt{d_k}$ bond vector. The scaling factor is introduced to prevent excessively large high-dimensional inner product results, thereby ensuring stability during the training process.

Building upon the self-attention mechanism, Vaswani proposed the Transformer model. This model establishes direct correlations between different time points within a single computational layer through multi-head attention and position encoding. Compared to recurrent neural networks, Transformers possess stronger parallel computing capabilities and are better suited for capturing long-range dependencies and intervariable coupling in multivariate time series. Consequently, attention-based models have garnered widespread attention in long-series load forecasting tasks.

However, Transformers still exhibit two limitations in long-series load forecasting. On one hand, the attention matrix scales rapidly with increasing input length, imposing substantial computational and storage costs - a problem particularly pronounced in modeling ultra-long historical windows. On the other hand, self-attention fundamentally relies on globally weighted aggregation, which tends to smooth out local high-frequency fluctuations. For short-term meteorological disturbances, sudden dispatch changes, or localized peak fluctuations in power load sequences, this insufficient local responsiveness may further compromise the stability of long-term forecasts.

Overall, while Transformers excel in global dependency modeling, they suffer from high computational complexity and are prone to smoothing artifacts when modeling local high-frequency fluctuations, thereby exhibiting limitations in complex non-stationary load forecasting tasks. This limitation has driven ongoing advancements in efficient modeling architectures for long sequences and provides a theoretical foundation for this paper's subsequent focus on state-space models and error compensation mechanisms.

3.3. Representative Methods for Long-Sequence Time Series Models

To address issues such as the high computational overhead of standard Transformers in long sequence modeling and their limited generalization capability for long-term predictions, researchers have proposed various improved models tailored for long sequence prediction. These approaches optimize the long sequence modeling architecture from different perspectives and serve as representative comparative baselines in current related research.

Table 1. Comparison of representative models for long-sequence time series forecasting and their core mechanism

Model Category	Representative Model	Core Mechanism	main features
Transformer Variant	Informer	Probabilistic Sparsity Attention	Reduce the complexity of long-sequence modeling
	Autoformer	Sequence Decomposition + Autocorrelation	Enhancing the modeling of trend and cyclical terms
	FEDformer	Frequency Domain Enhancement + Decomposition Mechanism	Using low-frequency information to maintain long-term trends
	PatchTST	Sequence Block Representation	Preserve local semantics and reduce high-frequency noise interference
	iTransformer	Variable Combination	Enhance multivariate relationship modeling capabilities
	Pyraformer	Hierarchical pyramid structure	Shorten the information transmission path dependent on long distances
Efficient Baseline Model	DLinear	Trend Decomposition + Linear Mapping	The structure is simple with small parameter scales.

Overall, improvements to existing long order time-series models can be broadly categorized into two approaches. The first approach focuses on sparsification of attention mechanisms and decomposition in both the time and frequency domains. For instance, Informer reduces modeling complexity through probabilistic sparse attention; Autoformer and FEDformer enhance prediction performance by decomposing sequences in the time domain and boosting representations in the frequency domain, respectively, to better capture trend and cyclical components. The second approach emphasizes input organization and representation architectures: PatchTST employs block-based representations to preserve local semantics while mitigating high-frequency noise; iTransformer strengthens multivariate relationships through variable-dimensional restructuring; Pyraformer utilizes a hierarchical structure to shorten information propagation paths for long-range dependencies, improving modeling efficiency. Additionally, DLinear achieves predictions through trend decomposition and linear mappings, maintaining strong competitiveness even with a small parameter scale.

Overall, the aforementioned models address key challenges in long sequence modeling from perspectives such as complexity control, trend extraction, local semantic preservation, and multivariate relationship representation, while also providing representative references for comparative analysis in subsequent experiments. To facilitate comparison of the core mechanisms across different models, Table 1 summarizes the relevant representative approaches. In subsequent experiments, this study will select models most relevant to the research task as comparison baselines to objectively evaluate the performance of the proposed hybrid prediction method.

4. EVALUATION INDICATORS

To objectively and quantitatively evaluate the generalization performance of multivariate multi-step prediction models, this paper adopts Mean Squared Error (MSE) and Mean Absolute Error (MAE) as the primary evaluation metrics.

The mean square error is the expectation of the squares of prediction deviations, and its mathematical definition can be expressed by the following formula (8):

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{8}$$

The average absolute error measures the average absolute value of prediction deviations and can be expressed by Formula (9):

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{9}$$

In the formula$y_i\hat{y}_i n$, represents the actual observed value, represents the model's predicted value, and represents the total sample size. MSE is more sensitive to large errors, making it suitable for assessing the model's fitting robustness under complex fluctuation scenarios; MAE provides a more intuitive reflection of the overall prediction bias level. Combining both metrics allows for evaluating the model's predictive performance from different perspectives. For deep learning models involving random initialization, multi-random seed experiments can be conducted to conduct further statistical analysis of the mean and range of results.

5. SUMMARY

This study analyzes the theoretical foundations of multivariate, multi-step power load forecasting, outlining the definition of forecasting tasks and the multi-step error propagation patterns. By examining the load series' dual characteristics of macro-level cyclical dependence and micro-level non-stationary fluctuations, it identifies the limitations of single deep models in capturing local errors. Building on this, the paper highlights the long-range trend extraction capability of deep sequence modeling approaches and the local nonlinear compensation potential of tree model residual learning mechanisms, demonstrating the rationality of their structured synergy.

REFERENCES

[1] Zheng Haowen. Exploration of the Application of GPS Technology in Modern Logistics Transportation Management. Logistics Engineering and Management, 2023, 45(7): 45-47.

[2] Xu Jianjun, Huang Lida, Yan Limei, Yi Na. Insulator Self-Explosion Defect Detection Based on Hierarchical Multi-Task Deep Learning[J]. Transactions of China Electrotechnical Society, 2021,36(07):1407-1415.

[3] Yan Limei, LIU Yongqiang, XU Jianjun, et al. Broken string diagnosis of composite insulator based on Grabcut segmentation and filler area discrimination[J]. Power System Protection and Control,2021,49(22): 114-119

[4] Yi, Q. Wang, L. Yan, et al., A multi-stage game model for the false data injection attack from attacker's perspective. Sustainable Energy Grids & Networks 28 (2021).

[5] Na Yi, Jianjun Xu, Limei Yan, Lin Huang. Task Optimization and Scheduling of Distributed Cyber-physical System Based on Improved Ant Colony Algorithm. Future Generation Computer Systems, 109(Aug. 2020),134-148.