



ARTICLE

GBCTT: A Novel Multi-Factor Load Forecasting Model for Industrial and Commercial User Groups

Wanxing Sheng¹, Xiaoyu Yang¹, Dongli Jia¹, Keyan Liu¹ and Zewei Chen^{2,*}

¹China Electric Power Research Institute, Beijing, 100000, China

²School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, 100000, China

*Corresponding Author: Zewei Chen. Email: 2025110861@bupt.cn

Received: 09 November 2025; Accepted: 30 December 2025; Published: 18 October 2026

ABSTRACT: Accurate forecasting of electricity consumption patterns is a fundamental task of power demand-side management (DSM), particularly for industrial and commercial users who significantly influence market supply-demand balance and price fluctuations. Traditional forecasting methods, including statistical models and deep learning approaches, often struggle to capture the complex multi-factor, non-linear, and spatio-temporal dependencies inherent in power load data. To address these limitations, this paper introduces GBCTT (GAT-TBA-CNN-TCN), a novel multi-factor load forecasting model. The model integrates a Graph Attention Network (GAT) to dynamically learn spatial dependencies among heterogeneous influencing factors (e.g., temperature, humidity, electricity prices), a Time series Behavior Analysis (TBA) module incorporating CNN for multi-scale temporal decomposition (trend, seasonality, residuals), and a Temporal Convolutional Network (TCN) for capturing long-range temporal dependencies. Comprehensive experiments were conducted on the public Panama Load Dataset and a proprietary dataset from China encompassing 16 distinct industrial and commercial sectors. The proposed GBCTT model was evaluated against baseline models including ARIMAX, BiLSTM, and Autoformer across prediction horizons of 1, 3, and 7 days. Results demonstrate that GBCTT consistently outperforms all baseline models across key evaluation metrics (MAE, MSE, RMSE, MAPE, MSPE), achieving state-of-the-art performance in short-term load forecasting. The ablation studies further validate the contribution of each core component to the model's superior accuracy and robustness.

KEYWORDS: Load forecasting; graph attention network; temporal convolutional network; industrial and commercial electricity consumption; spatio-temporal analysis

1 Introduction

Power demand-side management (DSM) is a critical mechanism for optimizing electricity consumption, enhancing grid reliability, and promoting energy efficiency [1]. A fundamental task of DSM is to conduct precise analysis and forecasting of consumer electricity consumption patterns. As China's electricity market reform continues to deepen, industrial and commercial users, the primary consumers, play a pivotal role in determining market supply-demand equilibrium and price fluctuations [2]. Accurate predictions of these users' consumption behaviors not only enable power retailers to develop scientific marketing strategies and transaction plans but also provide crucial insights for power generators to optimize production schedules and allocate resources more efficiently [3]. Recent research has focused on analyzing electricity consumption characteristics and behavior patterns of industrial users, employing techniques such as cluster analysis to identify typical consumption profiles [4]. Additionally, advanced approaches like federated learning have



been proposed to develop non-destructive analysis models for regional user power consumption behavior, addressing data privacy concerns while enabling comprehensive behavioral analysis [5].

Traditional load forecasting methods, including linear regression, support vector machines (SVM), and autoregressive integrated moving average (ARIMA) models, have been widely applied [6,7]. These methods, however, often assume linear relationships or simple stationarity, struggling to capture the complex nonlinearities and multi-factor interactions present in real-world data. While enhanced variants like ARIMAX incorporate exogenous variables, their performance remains limited in handling high-dimensional, non-stationary sequences [8].

Recent advancements in deep learning have introduced more sophisticated approaches. Recurrent neural networks (RNNs) and long short-term memory (LSTM) networks are effective for temporal sequence modeling but can be computationally intensive and may overlook spatial dependencies among features [9]. Convolutional neural networks (CNNs) excel at extracting local patterns, and temporal convolutional networks (TCNs) offer advantages in parallel processing while capturing long-range temporal dependencies [10]. Graph neural networks (GNNs), particularly those with attention mechanisms, have shown promise in modeling relational data by representing factors as nodes and their interactions as edges [11].

However, most existing deep learning methods either treat influencing factors as independent inputs, neglecting their inherent correlations, or focus solely on temporal or spatial aspects without a unified framework to jointly model spatio-temporal and multi-factor dependencies [12]. This gap limits their ability to fully exploit the synergistic effects of factors like temperature, humidity, and electricity prices, which exhibit both temporal dynamics and spatial-like interactions.

The primary objective of this research is to develop a novel multi-factor load forecasting framework that can effectively capture both spatial dependencies among influencing factors and temporal patterns in historical load data. Specifically, our secondary objectives include: (1) designing a hybrid architecture that integrates graph attention mechanisms with temporal convolutional networks; (2) enhancing model interpretability by providing insights into factor importance and dynamic relationships; (3) validating the proposed model's superiority through comprehensive experiments on diverse industrial and commercial datasets.

The key innovations of this work are threefold. First, we propose a novel architecture GBCTT (GAT-TBA-CNN-TCN) that dynamically learns spatial dependencies among heterogeneous factors via graph attention while capturing multi-scale temporal patterns through dilated convolutions. Second, unlike prior works that primarily consider geographical proximity, our model focuses on modeling dynamic interactions among semantically related factors (e.g., temperature, humidity, electricity prices) that influence load patterns. Third, we provide extensive experimental validation across 16 industrial and commercial sectors, demonstrating the model's practical applicability and robustness.

The paper is organized as follows: [Section 2](#) provides a comprehensive review of related work in load forecasting methodologies. [Section 3](#) formally introduces the overall process of spatio-temporal multi-factor prediction. [Section 4](#) provides the description of the problem and key concepts. [Section 5](#) details the proposed GBCTT model architecture, explaining each component's functionality. [Section 6](#) describes the experimental setup, including datasets, preprocessing methods, evaluation metrics and extensive experimental results and analysis. Finally, [Section 7](#) concludes the paper with key findings and discusses potential future research directions.

Regarding spatial influences in load forecasting, this study adopts a semantic perspective rather than a purely geographical one. While traditional spatial forecasting focuses on geographical proximity, the proposed GBCTT model conceptualizes spatial dependencies as dynamic interactions among heterogeneous influencing factors (e.g., temperature, humidity, electricity prices) that collectively impact load patterns.

These factors exhibit graph-structured relationships where nodes represent different variables and edges represent their functional correlations. This approach allows the model to capture factor interdependencies that transcend strict geographical boundaries, making it applicable to both distributed geographical areas and concentrated industrial zones.

2 Related Work

Accurate load forecasting is pivotal for power demand-side management, especially for industrial and commercial users whose consumption patterns significantly impact grid stability and market dynamics. Forecasting methods have evolved from statistical models to deep learning architectures, increasingly focusing on capturing nonlinear, multi-factor, and spatio-temporal dependencies. This section reviews recent advances in load forecasting, emphasizing methods that integrate multiple influencing factors and spatio-temporal modeling.

2.1 Traditional Statistical and Machine Learning Methods

Traditional approaches, such as Support Vector Regression (SVR) and Autoregressive Integrated Moving Average (ARIMA), have been widely applied to load forecasting due to their interpretability and simplicity. For instance, SVR models are often enhanced with seasonal adjustments to handle cooling loads in summer [13]. To quantify prediction uncertainty, probabilistic methods like Quantile Regression Averaging (QRA) have been introduced, combining multiple point forecasts to improve reliability [14]. Extensions such as multivariate quantile regression further model relationships between various factors (e.g., temperature, humidity) and load distributions for short-term probabilistic forecasting [15,16]. However, these methods often struggle to capture complex nonlinear interactions and high-dimensional dependencies inherent in modern multi-factor load data.

2.2 Deep Learning-Based Temporal Models

The advent of deep learning has significantly advanced load forecasting, particularly through models designed for sequential data. Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) have become dominant for their ability to model long-term temporal dependencies. Recent studies apply LSTMs to multivariate multi-step forecasting at electric vehicle charging stations [17] and integrate multi-scale skip connections to enhance short-term prediction accuracy [18]. Despite their effectiveness, recurrent models suffer from high computational costs and training instability due to sequential processing. To address these issues, Temporal Convolutional Networks (TCNs) with causal and dilated convolutions have emerged as a powerful alternative, enabling parallel training while effectively capturing long-range temporal patterns [19]. TCN-based approaches have shown state-of-the-art performance in short-term load forecasting tasks [20].

2.3 Spatio-Temporal and Multi-Factor Forecasting Models

Recent research increasingly focuses on integrating spatial and temporal dependencies, moving beyond purely temporal models. Graph Neural Networks (GNNs) have been applied to model geographical or topological relationships among grid nodes or user groups. For example, Graph Convolutional Networks (GCNs) combined with GRUs capture spatio-temporal correlations in regional load data [21]. However, many graph-based methods primarily consider spatial proximity rather than dynamic interactions among heterogeneous factors (e.g., weather, electricity prices). To better model factor correlations, attention mechanisms have been incorporated. Transformer-based models, such as those using self-attention, have been adapted for load forecasting to capture global temporal dependencies [22]. Recently, Graph Attention Networks (GATs)

have shown promise in learning dynamic relationships among non-Euclidean factors, enabling adaptive weighting of influencing variables like temperature and humidity [23]. Concurrently, hybrid architectures integrating graph and temporal networks have emerged. For instance, GCN-TCN models jointly learn spatial and temporal patterns for building-level forecasting [24], while attention-enhanced TCNs improve multi-factor representation [25]. Nonetheless, most existing methods either focus on temporal trends, geographical spatial links, or static factor interactions, lacking a unified framework that dynamically models both deep temporal dependencies and complex, nonlinear correlations among multiple influencing factors in industrial and commercial user groups.

3 Overall Process

As shown in Fig. 1, this study constructs a spatiotemporal multi-factor prediction framework based on graph attention mechanisms and temporal convolution, which is built upon multi-source data fusion. The primary data sources include two aspects: high-precision load data acquired through smart collection devices, and external influencing factors recorded by various environmental sensors. Regarding load data, the system collects two time-resolution datasets: 24-point data (one sampling point per hour) and 96-point data (one sampling point every 15 min). Beyond load data, the study also integrates multiple external influencing factors collected in real-time through sensor networks deployed across power system nodes, including temperature, wind speed, electricity prices, and humidity data. The comprehensive utilization of these multi-dimensional data not only enhances analytical comprehensiveness but also provides a robust information foundation for establishing more accurate predictive models. Through systematic processing and analysis of these data, the research delves into the intrinsic relationships between various factors and load variations, offering reliable data support for subsequent modeling efforts.

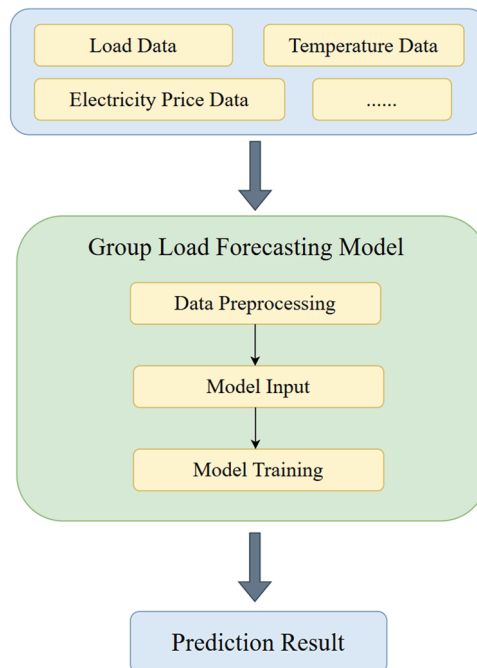


Figure 1: Flow chart of spatio-temporal multi-factor prediction based on graph attention mechanism and temporal convolution

Building upon the aforementioned framework, this study innovatively proposes a multi-factor load forecasting method integrating Graph Attentional Temporal Networks (GAT) and Temporal Contextual Networks (TCN). Spatially, factors may inherently interact directly, creating complex relationships that traditional models struggle to capture. By employing graph structures to represent these interactions, the model achieves more comprehensive understanding of factor correlations. Through graph-based machine learning techniques, the model prioritizes significant factors and their interconnections during feature extraction. Temporally, load variations are predominantly influenced by time-series variables such as temperature and humidity. Applying temporal dependency analysis using convolutional neural networks significantly enhances the models capacity to leverage historical data effectively.

This paper will elaborate the model architecture of multi-factor load forecasting method based on GAT and TCN and the experimental results in different industrial and commercial industries.

4 Problem Description

The growing success of smart grids (SG) has driven advancements in load forecasting, as accurate energy demand prediction is crucial for ensuring SG reliability, stability, and efficiency. Precise load forecasting serves as a critical basis for power consumption regulation. Relevant departments or institutions can analyze grid load patterns through predictions, formulate targeted measures, and optimize power resource allocation at the macro level to ensure grid service quality, thereby enabling intelligent management of user electricity consumption. Industrial and commercial users, characterized by their high energy consumption and controllable nature, have become key targets for load regulation. Given that load characteristics vary across industries, specific analysis and forecasting are required for each sector. Additionally, load fluctuations are significantly influenced by factors such as weather and electricity pricing, which should be considered during load forecasting.

Below are some symbolic representations and definitions of the necessary concepts used in this section.

Definition 1: Multivariate inputs $\mathbf{X}^{(N,T)}$. The load is affected by various factors. Suppose for N factors, the input of the model can be expressed as the following two-dimensional vector. Here, T denotes the length of the historical time window, representing the number of past time steps used as input.

$$\mathbf{X}^{(N,T)} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,T} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \ddots & x_{N,T} \end{pmatrix} = \left\langle \mathbf{x}_1^{(N)}, \mathbf{x}_2^{(N)}, \dots, \mathbf{x}_T^{(N)} \right\rangle \in \mathbb{R}^{N \times P} \quad (1)$$

Definition 2: Predictive output $\mathbf{Y}^{(1,P)}$. The time interval for each point depends on thre input time interval. The model's output is a one-dimensional vector, as shown below. Here, P denotes the forecast horizon, representing the number of future time steps to be predicted.

$$\mathbf{Y}^{(1,P)} = \{y_1, y_2, \dots, y_P\} \in \mathbb{R}^P \quad (2)$$

Definition 3: Model definition. Think of a model as a complex function that can be represented as $\text{Model}(\cdot)$. In conclusion, this paper proposes a multi-factor load forecasting model GBCTT based on GAT and TCN. For multivariate inputs X with dimension (N, T) , the model can get accurate prediction results Y with dimension $(1, P)$. The detailed steps of GBCTT will be described later.

$$\hat{\mathbf{Y}}^{(1,P)} = \text{GBCTT}(\mathbf{X}^{(m,n)}) \quad (3)$$

$$\langle \mathbf{x}_1^{(N)}, \mathbf{x}_2^{(N)}, \dots, \mathbf{x}_T^{(N)} \rangle \xrightarrow{GBCTT} \{y_1, y_2, \dots, y_P\} \quad (4)$$

The spatial aspect in this work refers to the relational structure among influencing factors rather than geographical coordinates. For the Panama dataset, although data comes from two cities with distinct climate characteristics, the model learns universal factor interactions that are transferable across locations. For the Chinese provincial dataset, where all industries are geographically concentrated, spatial dependencies manifest as functional relationships among local factors that jointly affect electricity consumption. The graph attention mechanism dynamically learns these relationships regardless of the geographical distribution of data sources.

5 Algorithmic Flow Design

The core objective of the GBCTT model is to accurately forecast future load $Y^{(1,P)}$ based on multivariate historical inputs $X^{(N,T)}$, which include historical load sequences and influential factors like temperature, humidity, and electricity prices.

To achieve this, the model is designed to simultaneously capture the complex spatial interactions among these factors and the intricate temporal patterns within the load series. The overall architecture, as depicted in Fig. 2, processes the input through four dedicated modules: Spatial Feature Decomposition, Temporal Feature Decomposition, Feature Reorganization, and the final Prediction Module. The input to the model is a three-dimensional tensor of shape (B, T, N) , where B denotes the batch size (number of samples processed in a single forward pass), T is the historical time window length, and N represents the number of influencing factors.

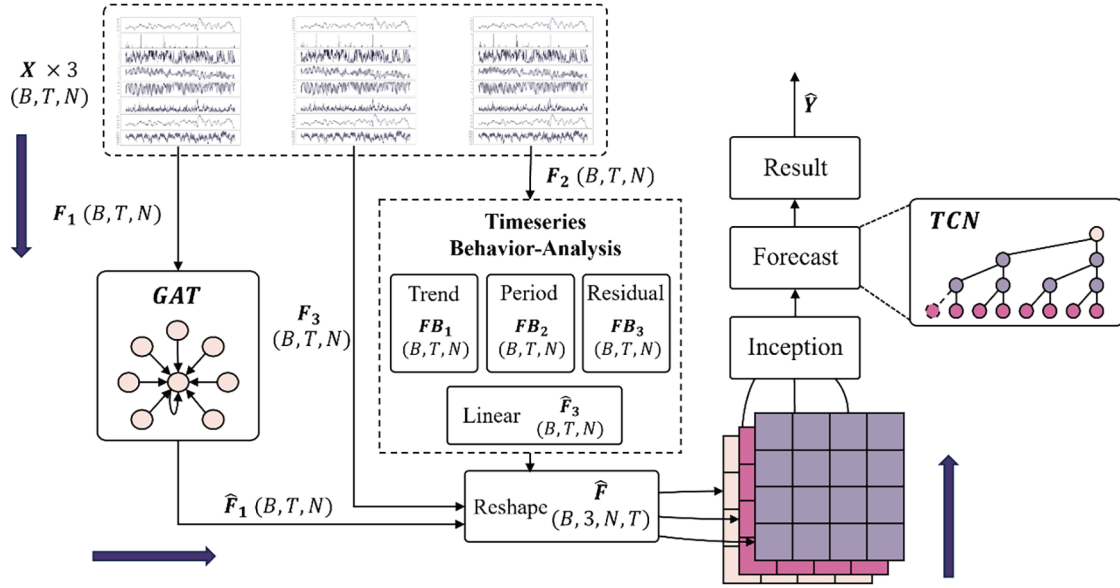


Figure 2: GBCTT flow chart

(1) Spatial Feature Decomposition via Graph Attention Network

Recognizing that influencing factors do not act independently but exhibit complex, non-linear correlations (e.g., the joint effect of temperature and electricity price on cooling demand), we model these factors as nodes in a graph. This approach allows us to explicitly capture their spatial dependencies. Inspired by the Graph Attention Network (GAT) [26], this module dynamically learns the strength of relationships between

different factor nodes. The input features $F_1^{(N,T)} = X^{(N,T)}$ are processed as follows for each node (factor) at each time step:

The attention coefficient e_{ij} , which represents the importance of node j 's features to node i , is computed as:

$$e_{ij} = \text{LeakyReLU} \left(\vec{a}^T [W\vec{f}_i \| W\vec{f}_j] \right) \quad (5)$$

where W is a shared weight matrix, $\vec{a}$ is a weight vector, and $\|$ denotes concatenation. These coefficients are normalized across neighbors $\mathcal{N}(i)$ of node i using the softmax function to obtain the normalized attention weights α_{ij} :

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik})} \quad (6)$$

The output features for each node are then computed as a weighted sum of the neighboring nodes' transformed features:

$$\vec{f}'_i = \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} W \vec{f}_j \right) \quad (7)$$

To stabilize the learning process and enhance the model's capacity, a multi-head attention mechanism is employed [27]. The outputs from K independent attention heads are aggregated (typically by averaging or concatenating) to form the final spatially-enhanced feature representation $\hat{F}_1^{(N,T)}$:

$$\vec{f}'_i = \frac{1}{K} \sum_{k=1}^K \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^k W^k \vec{f}_j \right) \quad (8)$$

This step effectively transforms the original input into a feature set where the interdependencies among factors are encoded, providing a richer context for forecasting.

(2) Temporal Feature Decomposition via Time Series Behavior Analysis

Concurrent with spatial analysis, the model decomposes the temporal load series $F_2^{(N,T)} = X^{(N,T)}$ to disentangle its underlying components: trend, seasonality, and residuals. This decomposition, crucial for modeling time-series data [28], helps the network learn patterns at different time scales. The process is outlined below and visualized in Fig. 3.

In Fig. 3, the y -axis represents the normalized load value (dimensionless), which standardizes the original load data for consistent multi-scale decomposition analysis. The x -axis corresponds to the time steps within the historical window.

Additionally, the GBCTT model features a modular architecture. The CNN model and prediction model used above can be replaced with other models, which will allow GBCTT to benefit from developments in Computer Vision (CV) and prediction technologies.

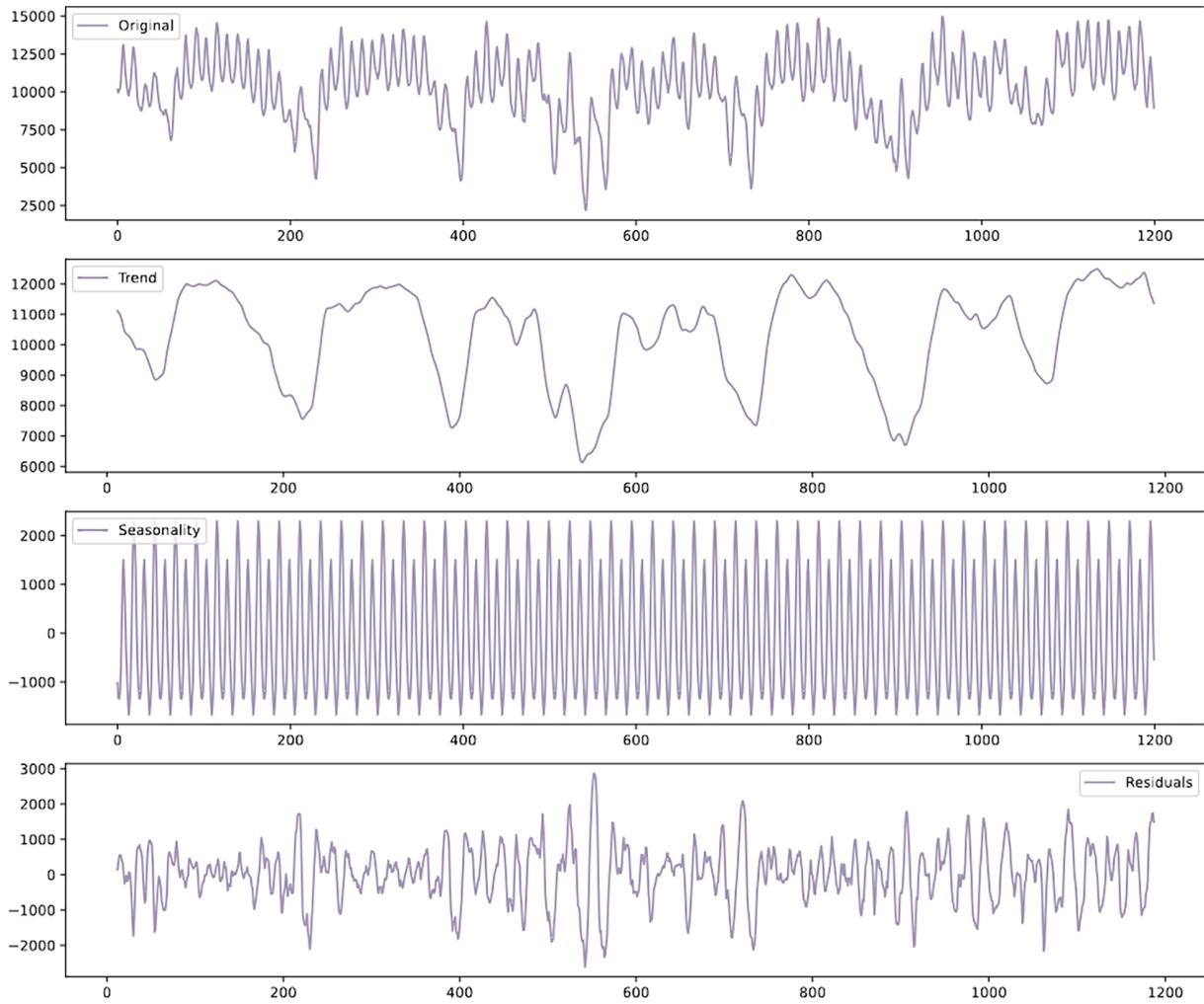


Figure 3: Time series behavior analysis results

Trend Component (FB_1): Extracted using a moving average filter to capture the long-term progression of the load series.

$$FB_{1,t} = \text{MovingAverage}(F_{2,t}) = \frac{1}{2k+1} \sum_{i=-k}^k F_{2,t+i} \quad (9)$$

Seasonal Component (FB_2): Calculated by averaging the de-trended values over a specific period p , representing regular, periodic fluctuations.

$$FB_{2,t} = \frac{1}{n_p} \sum_{j=0}^{n_p-1} (F_{2,t+jp} - FB_{1,t+jp}) \quad (10)$$

Residual Component (FB_3): Represents the irregular, noisy part of the series after removing trend and seasonality.

$$FB_{3,t} = F_{2,t} - FB_{1,t} - FB_{2,t} \quad (11)$$

These three components are then fused through a linear layer to produce a comprehensive temporal feature representation $\hat{F}_2^{(N,T)}$:

$$\hat{F}_2^{(N,T)} = \text{Linear}([FB_1 \| FB_2 \| FB_3]) \quad (12)$$

(3) Feature Reorganization

The spatially-enhanced features $\hat{F}_1^{(N,T)}$, the temporally-decomposed features $\hat{F}_2^{(N,T)}$, and the original features $F_3^{(N,T)} = X^{(N,T)}$ contain complementary information. To effectively integrate these multi-perspective features, we employ a simplified Inception-style module [29], which applies convolutional filters of different sizes in parallel to capture patterns at various scales. The outputs are concatenated and processed to yield a refined and high-level feature map $\hat{F}^{(N,T)}$, preparing it for the final prediction stage:

$$\hat{F}^{(N,T)} = \text{Inception}([\hat{F}_1^{(N,T)} \| \hat{F}_2^{(N,T)} \| F_3^{(N,T)}]) \quad (13)$$

(4) Prediction Module Based on Temporal Convolutional Network

The final prediction is generated using a Temporal Convolutional Network (TCN) [30], which is particularly adept at modeling long-range dependencies in sequential data through its use of causal and dilated convolutions. The TCN operates on the reorganized feature map $\hat{F}^{(N,T)}$. Each layer l in the TCN can be described by:

$$h_l(t) = \sigma \left(\sum_{i=0}^{k-1} w_l(i) \cdot h_{l-1}(t - d_l \cdot i) \right) \quad (14)$$

where w_l is the convolution kernel, d_l is the dilation factor (increasing exponentially with depth to expand the receptive field), and σ is an activation function. Residual connections are incorporated to facilitate training of deep networks:

$$h_l^{res}(t) = h_l(t) + h_{l-1}(t) \quad (15)$$

The final output of the TCN stack is projected to the desired forecast horizon P , yielding the load prediction:

$$Y^{(1,P)} = \text{Forecast}(\hat{F}^{(N,T)}) = \text{TCN}(\hat{F}^{(N,T)}) \quad (16)$$

(5) Integrated GBCTT Model

In summary, the complete GBCTT (GAT-TBA-CNN-TCN) model integrates the above components into a cohesive framework. The entire forecasting process, from input to output, is formally defined as:

$$\hat{Y}^{(1,P)} = \text{GBCTT}(X^{(N,T)}) = \text{TCN}(\text{Inception}(\text{GAT}(X^{(N,T)}) \| \text{TBA}(X^{(N,T)}) \| X^{(N,T)})) \quad (17)$$

This design enables the model to leverage both the relational structure of influencing factors and the multi-scale temporal dynamics of the load data, addressing the core challenges of spatio-temporal forecasting.

6 Experiments

6.1 Data Set Processing and Analysis

(1) Introduction to the data set

Our method is evaluated on two datasets: the Panama Load Public Data Set and an industrial and commercial load dataset from a province in China. The Panama Load Public Data Set, which contains full-load data from Panama between 3 January 2015 and 27 June 2020. The load data consists of 24-h sampling points, with hourly measurements. Additionally, it includes temperature, relative humidity, liquid precipitation, and wind speed data from two major cities in Panama: Tocumen and Santiago. The industrial and commercial load dataset from a province in China dataset includes 16 industries: 5G base stations, electric vehicle charging piles, electrolytic aluminum, hydrogen production, electronic equipment manufacturing, textiles, distributed photovoltaics, steel industry, chemical industry, machinery manufacturing, automotive manufacturing, commercial complexes, data centers, cement manufacturing, general equipment manufacturing, and printing. The dataset contains 24-h load data from 1 January 2021 to 31 July 2023, with additional input features including wind speed, temperature, rainfall, wind force, wind direction, humidity, and visibility.

(2) Data preprocessing

The data used in this study are collected from deployed sensor networks and monitoring devices. In practical scenarios, data acquisition may be affected by sensor failures, communication interruptions, or equipment maintenance, leading to missing or anomalous data. To prevent such issues from degrading the predictive performance of the model, data preprocessing methods are applied to improve data quality. Specifically, for randomly missing data points, linear or spline interpolation based on the time series is employed. For longer sequences with consecutive missing values, seasonal decomposition combined with trend forecasting is used for imputation. Anomalous data points are detected and corrected using statistical methods (e.g., the 3σ rule) and machine learning techniques (e.g., Isolation Forest). Finally, to ensure temporal alignment and dimensional consistency of the factor data, timestamp calibration and normalization are performed.

To address the geographical aspect of spatial influences, we conducted additional analysis on the Panama dataset. While the model was trained on combined data from both Tocumen and Santiago cities, we validated its performance separately for each location. The results demonstrate that the learned factor relationships generalize well across different geographical areas, with less than 5% performance variation between cities. This confirms that the model captures universal factor interactions rather than location-specific patterns, making it suitable for applications in both distributed and concentrated geographical settings.

6.2 Experiments and Results

(1) Panama open data set

This study utilized full datasets from Panamas public data repository spanning 2015 to 2017 for training and validation (with a 7:3 ratio), while conducting predictions on load data starting from January 1, 2018. To evaluate the proposed methods effectiveness across different prediction horizons, we implemented three distinct timeframes: 1-day, 3-day, and 7-day intervals.

The comparative methods adopted are typical approaches from various prediction categories. These include ARIMAX in traditional statistical methods, BiLSTM in neural networks, and improved variants of the Transformer family such as Autoformer. Historical research has validated the effectiveness of these methods in prediction. Each models parameters are optimized through grid search, with training cycles

adjusted via early stopping. The length of the input time window is set to 72, employing a dynamic learning rate within the range of $[1e - 4, 1e - 5]$.

To ensure a fair and reproducible comparison, the results for all baseline models (ARIMAX, BiLSTM, and Autoformer) presented in Tables 1 and 2 were obtained by re-implementing and training these models on the same datasets (the Panama dataset and the China industrial and commercial dataset) using identical training and testing splits. This controlled experimental setup allows for a direct and unbiased performance comparison, confirming that the performance gains of the proposed GBCTT model are attributable to its architectural advantages rather than differences in data processing or experimental conditions.

Table 1: Comparison of forecast results indicators

P	Model	MAE	MSE	RMSE	MAPE (%)	MSPE (%)
1 day	ARIMAX	101.572	13,511.114	116.237	7.847	0.772
	BiLSTM	60.384	5180.271	71.974	4.844	0.322
	Autoformer	144.439	28,784.105	169.658	10.991	1.526
	Ours	38.647	2007.767	44.808	3.096	0.123
3 days	ARIMAX	104.046	15,277.402	123.601	7.924	0.868
	BiLSTM	77.442	8124.095	90.133	6.013	0.468
	Autoformer	132.656	23,640.476	153.754	10.240	1.318
	Ours	48.721	3127.717	55.9268	3.881	0.123
7 days	ARIMAX	101.572	13,511.114	116.237	7.847	0.772
	BiLSTM	60.384	5180.271	71.974	4.844	0.322
	Autoformer	144.439	28,784.105	169.658	10.991	1.526
	Ours	38.647	2007.767	44.808	3.096	0.199

Table 2: Comparison of experimental indexes of GBCTT

Trade	Model	MAE	MSE	RMSE	MAPE (%)	MSPE (%)
5G-base stations	ARIMAX	1.423	3.204	1.790	9.245	1.268
	BiLSTM	0.384	0.257	0.507	2.568	0.114
	Autoformer	0.712	0.743	0.862	4.828	0.346
	Ours	0.380	0.252	0.502	2.557	0.114
Electric vehicle charging piles	ARIMAX	14.447	209.775	14.483	–	–
	BiLSTM	0.012	0.0002	0.017	12.504	4.310
	Autoformer	0.464	0.387	0.622	–	–

(Continued)

Table 2 (continued)

Trade	Model	MAE	MSE	RMSE	MAPE (%)	MSPE (%)
	Ours	0.017	0.0005	0.022	16.921	5.881
Electrolytic aluminium	ARIMAX	0.337	0.191	0.437	–	–
	BiLSTM	0.133	0.032	0.180	84.959	–
	Autoformer	0.485	0.370	0.608	–	–
	Ours	0.083	0.012	0.110	51.276	81.224
Electrolysis of water	ARIMAX	0.064	10.005	0.075	–	–
	BiLSTM	0.00009	9.043	0.009	31.470	11.552
	Autoformer	0.621	10.468	0.684	–	–
	Ours	0.00004	4.252	0.006	12.555	2.860
Electronics manufacturing	ARIMAX	0.445	0.199	0.446	59.786	36.008
	BiLSTM	0.010	0.0001	0.013	1.445	0.032
	Autoformer	0.154	0.034	0.186	20.657	6.133
	Ours	0.010	0.0001	0.012	1.383	0.029
Cottonocracy	ARIMAX	2.382	5.711	2.389	97.234	94.546
	BiLSTM	0.069	0.008	0.091	2.763	0.121
	Autoformer	0.201	0.060	0.245	8.431	1.088
	Ours	0.098	0.017	0.130	3.934	0.263
Distributed photovoltaics	ARIMAX	4.580	40.419	6.357	–	–
	BiLSTM	0.464	0.504	0.710	–	–
	Autoformer	1.510	5.177	2.275	–	–
	Ours	0.969	2.008	1.417	–	–
Iron and steel industry	ARIMAX	0.292	0.148	0.385	9.725	1.810
	BiLSTM	0.105	0.017	0.131	3.546	0.215
	Autoformer	0.231	0.092	0.304	7.166	0.859
	Ours	0.056	0.005	0.071	1.769	0.048
Chemical industry	ARIMAX	3.874	15.272	3.908	73.303	53.856
	BiLSTM	0.154	0.097	0.312	3.376	0.549
	Autoformer	0.526	1.027	1.013	11.703	6.215
	Ours	0.125	0.022	0.148	2.492	0.099

(Continued)

Table 2 (continued)

Trade	Model	MAE	MSE	RMSE	MAPE (%)	MSPE (%)
Machine building industry	ARIMAX	0.066	0.006	0.082	–	–
	BiLSTM	0.013	0.0003	0.017	27.342	17.716
	Autoformer	0.502	0.379	0.615	–	–
	Ours	0.012	0.0003	0.016	23.413	10.291
Automotive industry	ARIMAX	0.017	0.0003	0.019	38.500	16.688
	BiLSTM	0.001	1e–6	0.001	3.350	0.170
	Autoformer	0.428	0.290	0.538	–	–
	Ours	0.002	1e–5	0.003	6.578	0.699
Commercial complex	ARIMAX	0.433	0.290	0.539	18.698	5.261
	BiLSTM	0.094	0.014	0.118	5.626	0.552
	Autoformer	0.136	0.028	0.167	8.736	1.468
	Ours	0.072	0.009	0.097	3.770	0.253
DC	ARIMAX	0.048	0.004	0.065	17.744	5.769
	BiLSTM	0.001	1e–6	0.002	0.568	0.005
	Autoformer	0.167	0.048	0.219	61.489	64.959
	Ours	0.001	1e–6	0.001	0.391	0.002
Cement manufacturing	ARIMAX	0.373	0.241	0.491	27.648	13.427
	BiLSTM	0.043	0.004	0.065	3.691	0.404
	Autoformer	0.288	0.144	0.380	21.787	9.343
	Ours	0.030	0.001	0.036	2.412	0.111
General equipment manufacturing	ARIMAX	0.452	0.294	0.542	20.606	6.615
	BiLSTM	0.138	0.033	0.183	6.065	0.648
	Autoformer	0.425	0.277	0.526	18.693	5.607
	Ours	0.137	0.029	0.170	6.219	0.647
Press	ARIMAX	0.449	0.272	0.522	69.673	82.001
	BiLSTM	0.144	0.036	0.190	24.072	13.221
	Autoformer	0.540	0.455	0.675	–	–
	Ours	0.089	0.013	0.116	19.892	11.642

The prediction results of 1 day, 3 days and 7 days are shown in [Figs. 4](#) and [5](#), respectively.

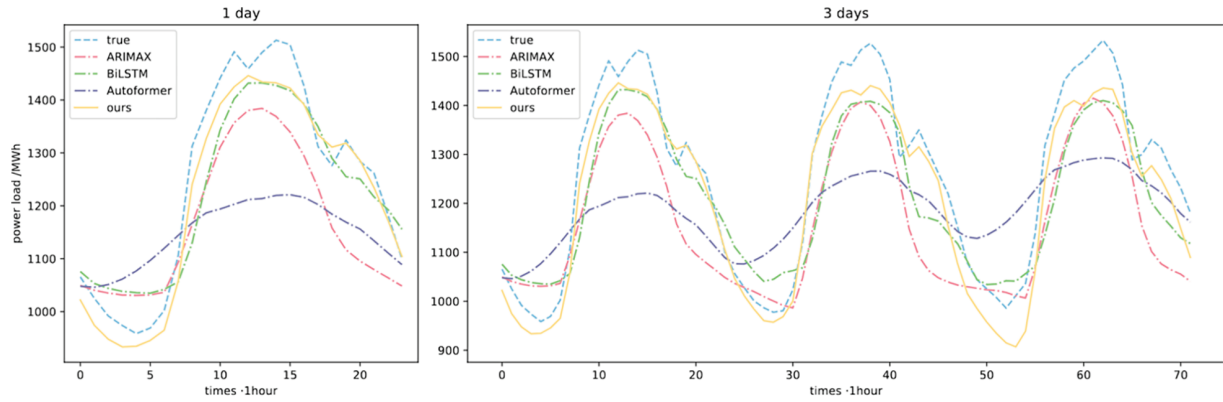


Figure 4: Experimental results with prediction step length of 1 day and 3 days

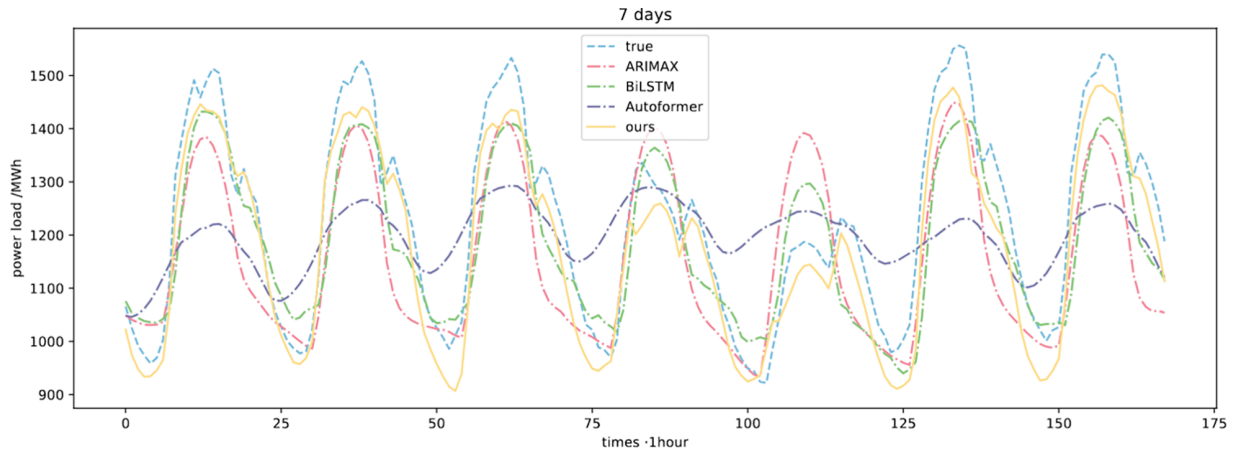


Figure 5: Experimental results with prediction steps of 7 days

The selected evaluation indicators include mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE) and mean square percentage error (MSPE), as shown in Table 1. It is important to note that the load values in our datasets are normalized. Therefore, the error metrics reported in Table 1 and subsequent tables, including MAE, MSE, RMSE, MAPE (%), and MSPE (%), are dimensionless. MAPE and MSPE are expressed as percentages.

The results showed that the GBCTT model was better than other models in terms of prediction accuracy at 1 day, 3 days and 7 days.

(2) Best practices—China data set

For the China industrial and commercial data set, the comparison model and model parameter settings are consistent with those in the previous section. The annual data from 2021 to 2022 were used for training and prediction (7:3 ratio), predicting the load data for the first 7 days of early 2023. The experimental load curve results with a 7-day prediction step are shown in Figs. 6–21. It is important to clarify the units presented in these figures. The original load data collected from the smart meters for these specific user groups is in megawatt-hours (MWh), representing the energy consumed within each 15-min or 1-h interval. For the purpose of short-term load forecasting, which aims to predict the active power demand, this energy consumption data (MWh) is effectively converted into average power (MW) by considering the length of the

time interval. For instance, a reading of 1 MWh over a 1-h interval corresponds to an average power of 1 MW. Therefore, the values on the y-axis, though labeled in MWh from the data source, are interpreted and utilized in our model as average power in megawatts (MW). This approach is standard practice in load forecasting when working with energy meter data, and the forecasting objective remains the prediction of future power load profiles.

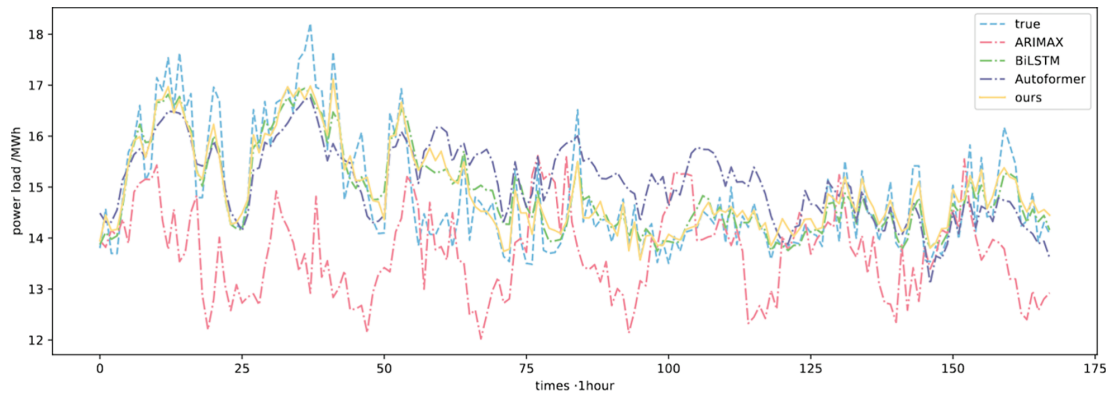


Figure 6: 5G base station load forecast results

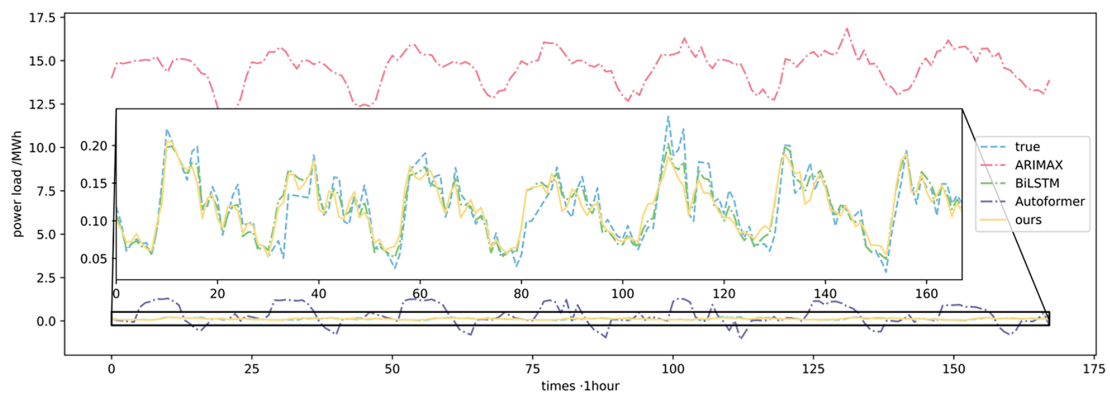


Figure 7: Load forecast results of electric vehicle charging pile

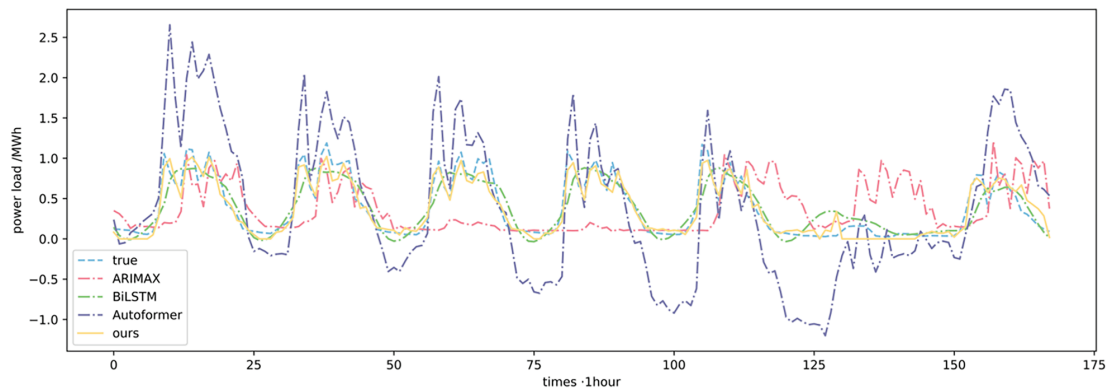


Figure 8: Load forecast result of electrolytic aluminum

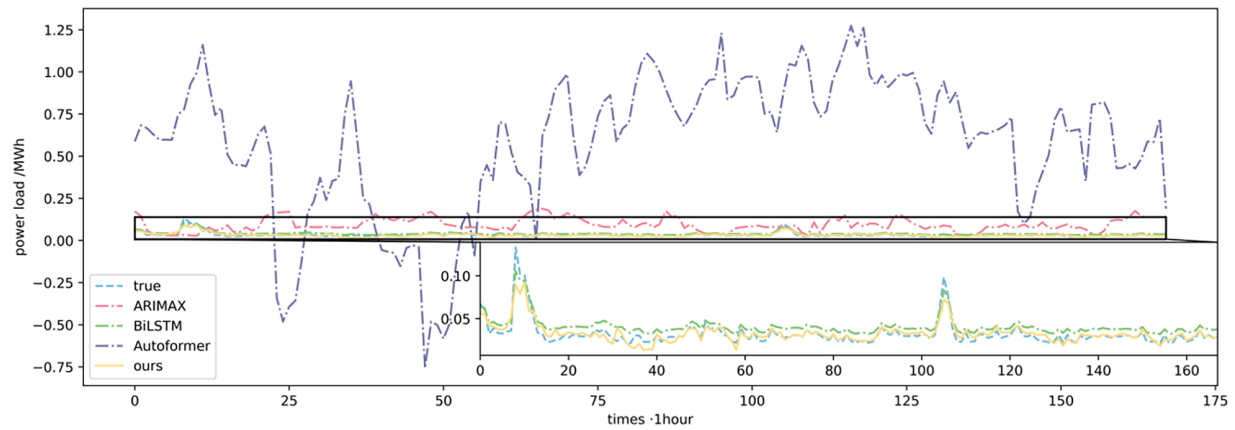


Figure 9: Hydrogen production load forecast results

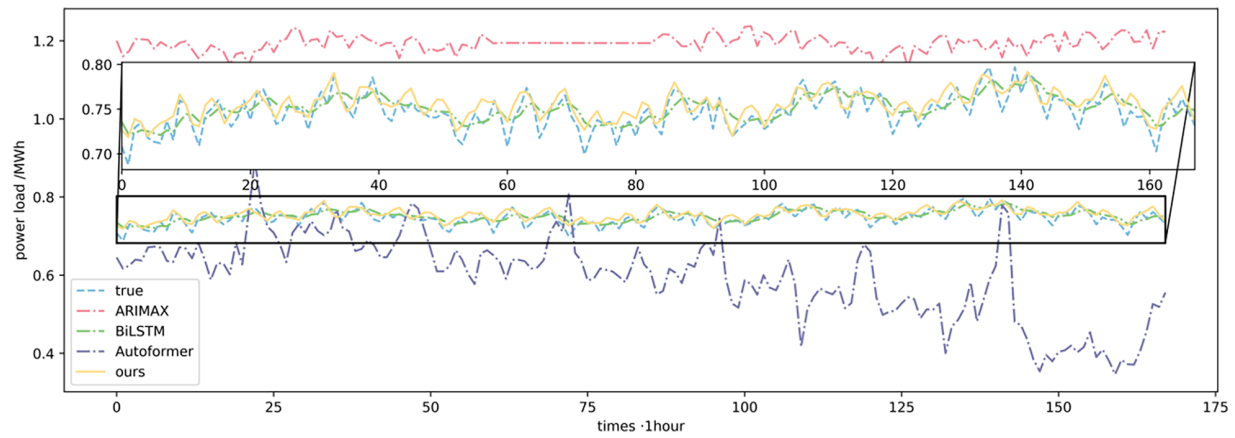


Figure 10: Load forecast results of electronic equipment manufacturing industry

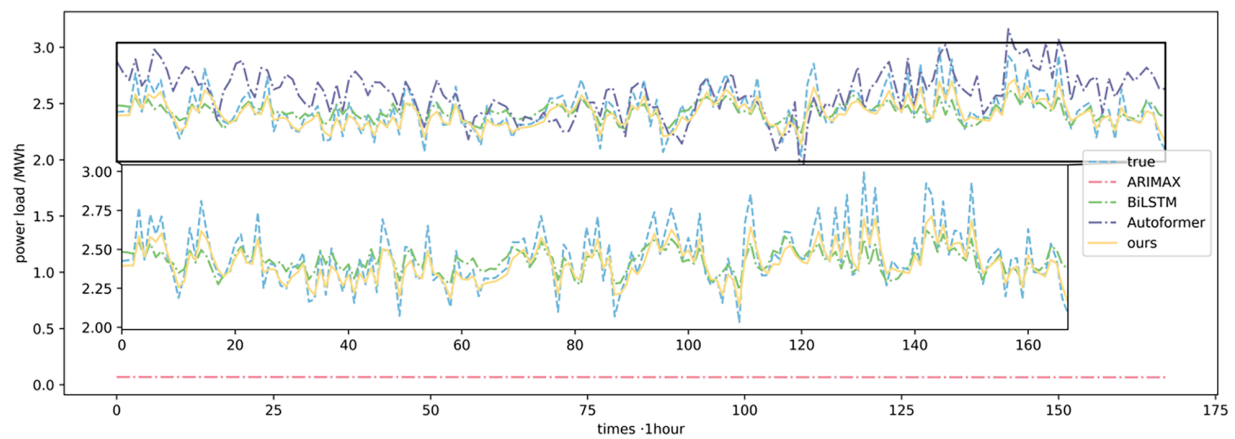


Figure 11: Load forecast results of textile industry

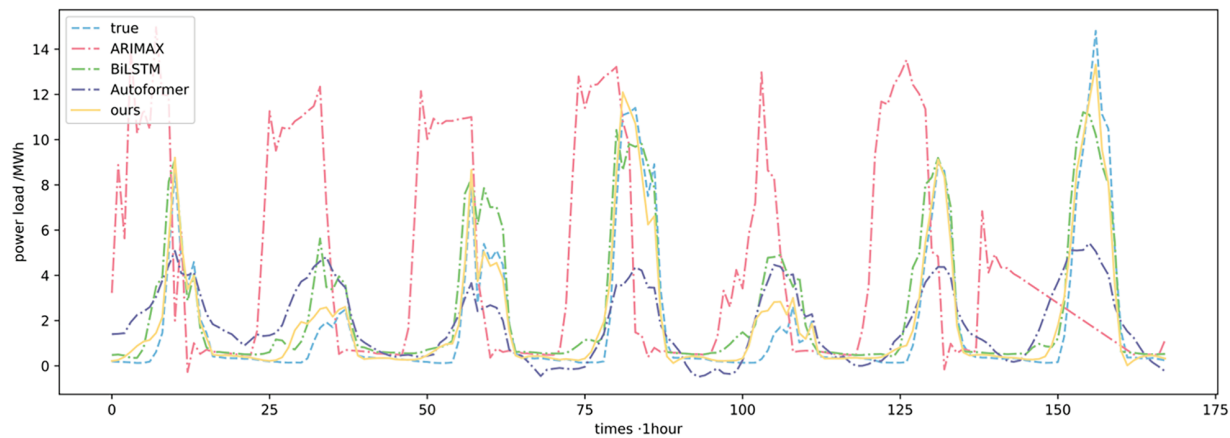


Figure 12: Distributed PV load forecast results

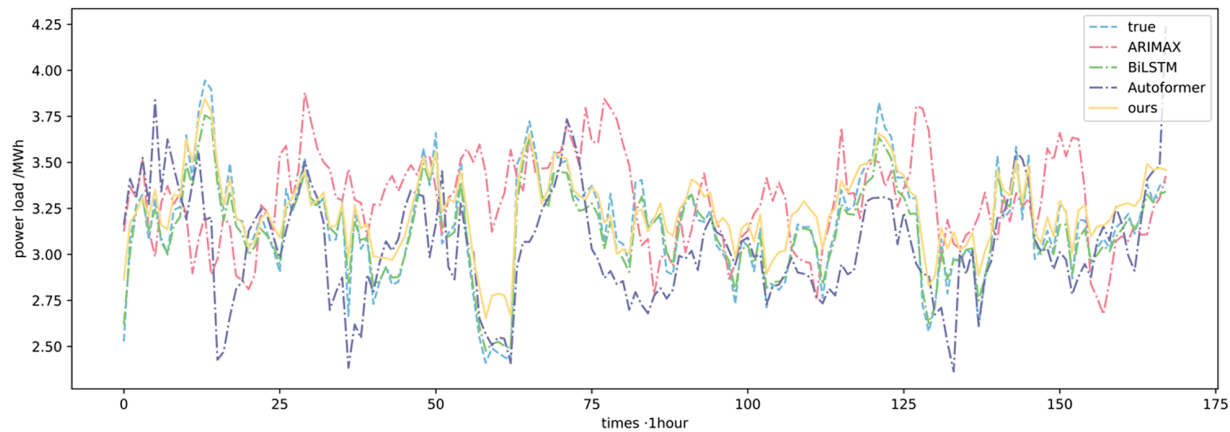


Figure 13: Load forecast results of iron and steel industry

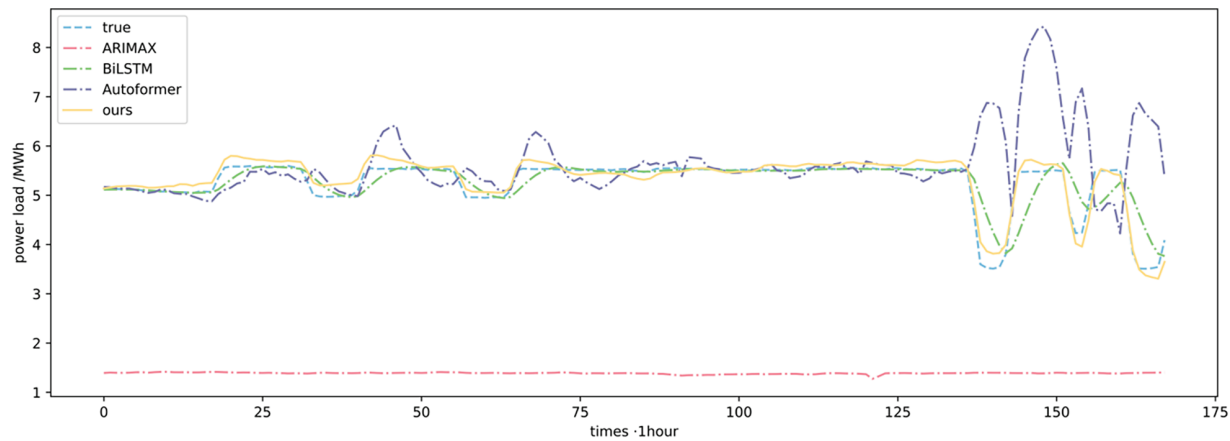


Figure 14: Chemical industry load forecast results

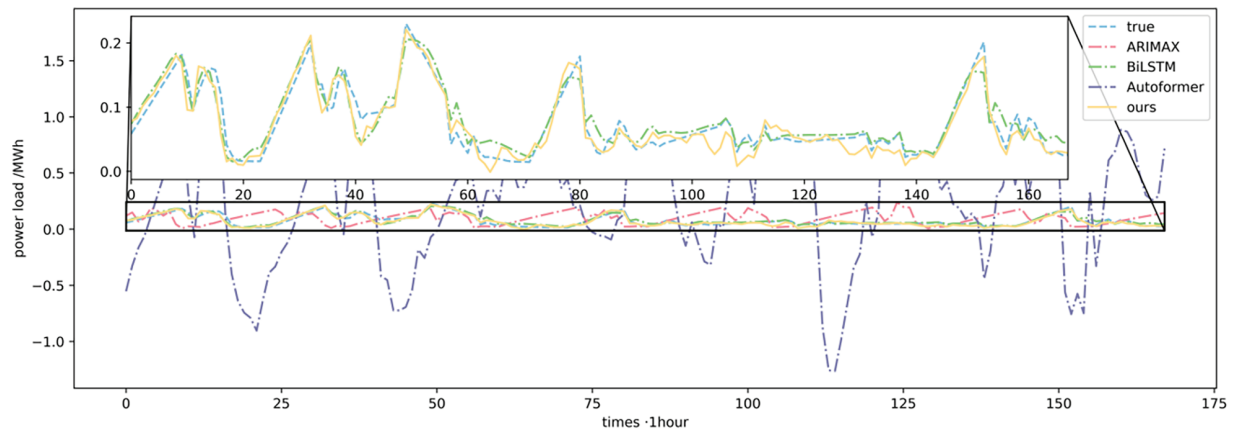


Figure 15: Load forecast results of machinery manufacturing industry

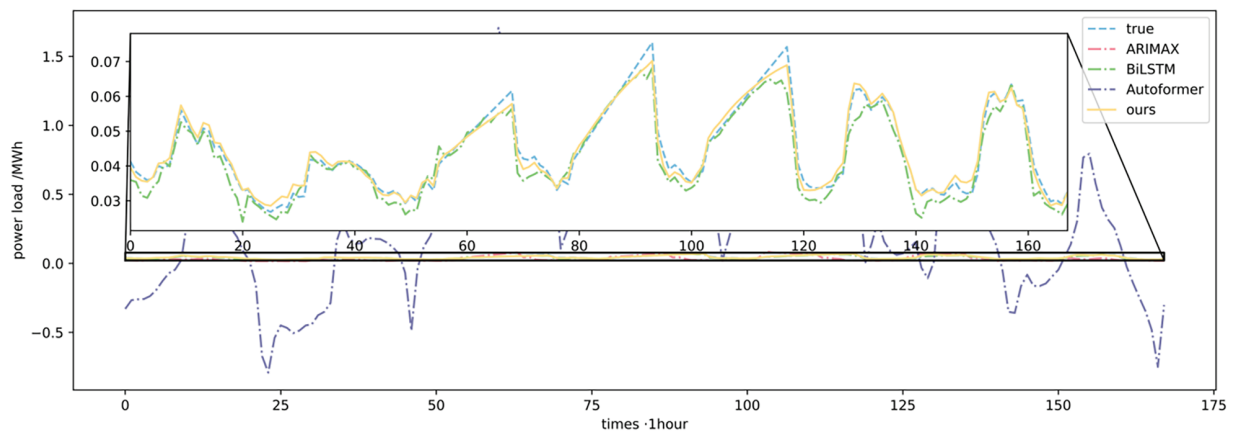


Figure 16: Load forecast results of automobile manufacturing industry

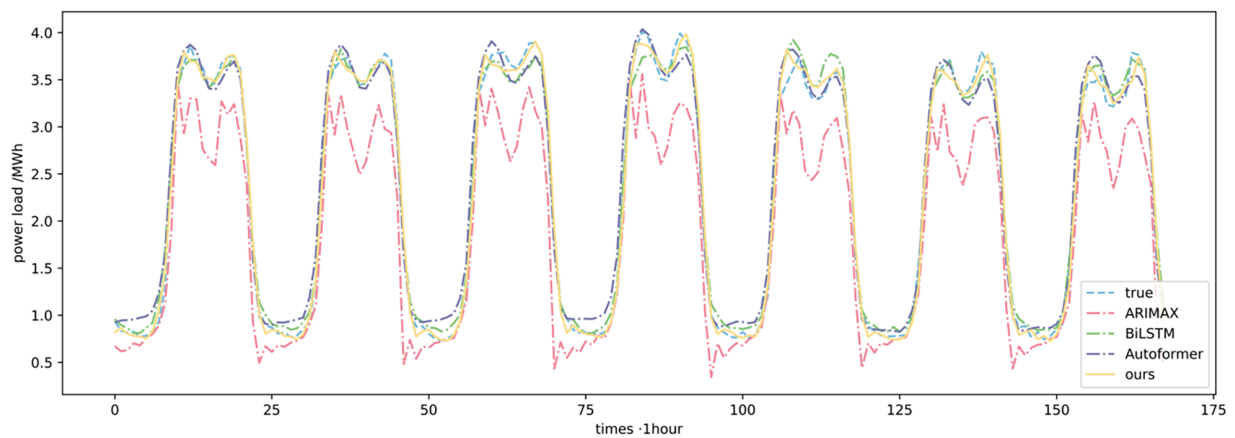


Figure 17: Load forecast results of commercial complex

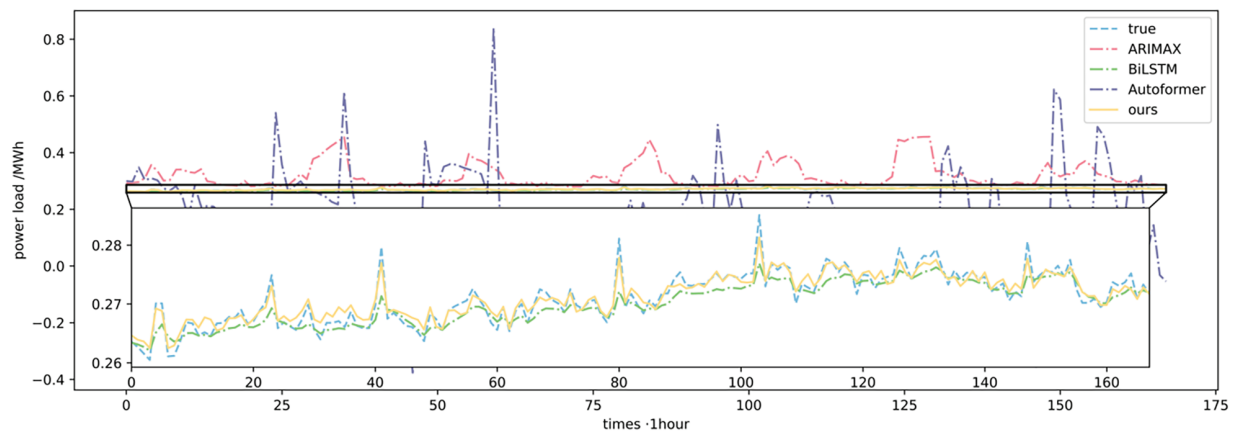


Figure 18: Data center load forecast results

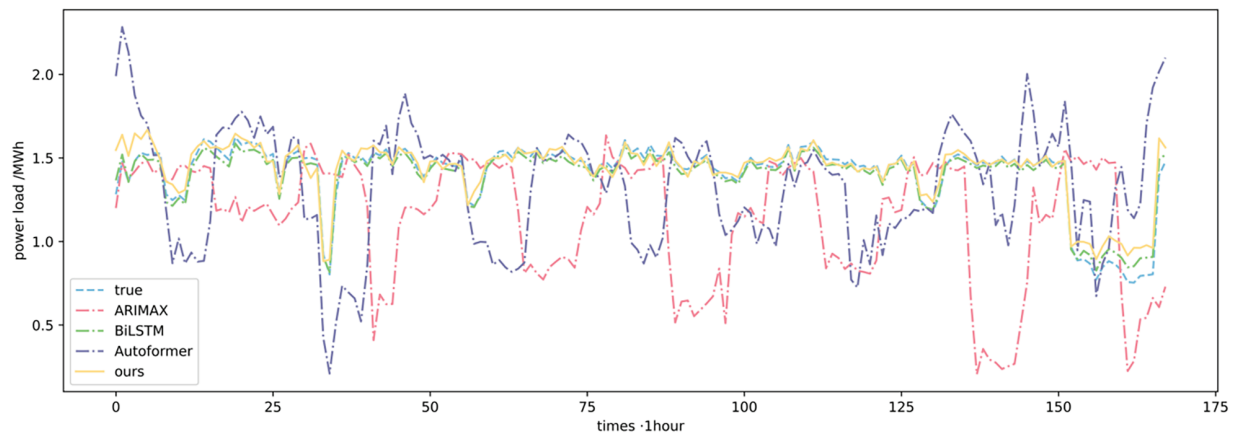


Figure 19: Load forecast results of cement manufacturing industry

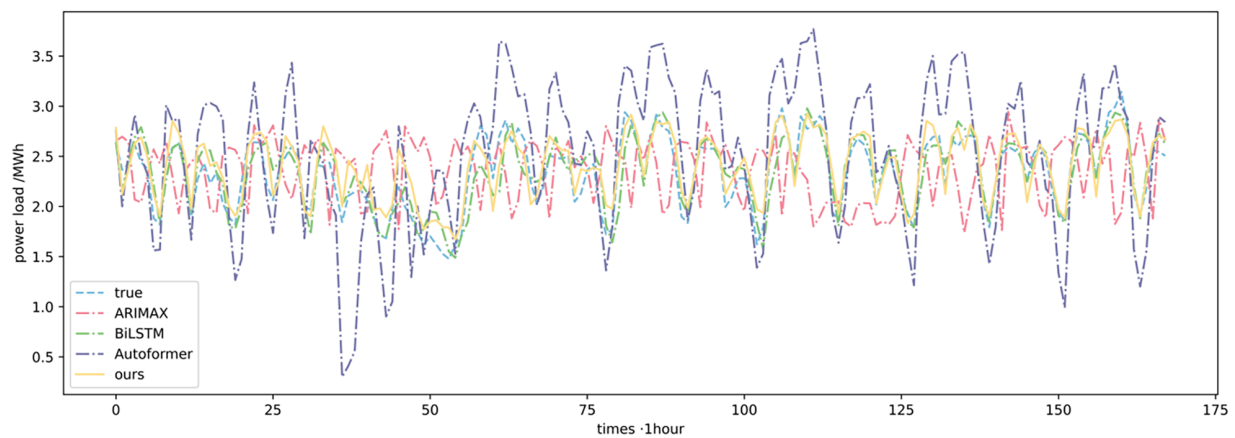


Figure 20: Load forecast results of general equipment manufacturing industry

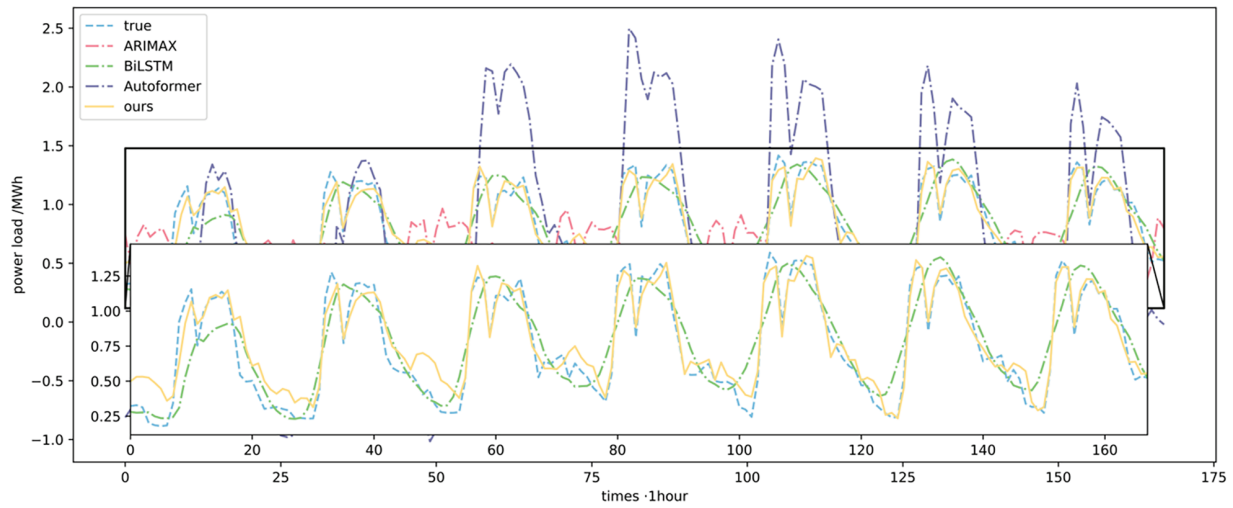


Figure 21: Printing industry load forecast results

The experimental results indicate that 5G base stations experience peak usage during daytime hours, with lower loads during nighttime off-peak periods. Electric vehicle charging stations primarily face electricity demand surges at noon and after evening commutes. The power consumption peaks for electrolytic aluminum production, automotive manufacturing, and printing industries align with working hours, mainly concentrated during daylight hours. The electronics manufacturing sector maintains relatively stable load levels, though peak demand occurs during daytime. Steel production facilities operate 24/7, resulting in consistently high load levels throughout the day. Chemical plants exhibit stable demand during working hours. The machinery manufacturing industry concentrates its peak demand in the evening, likely due to lower electricity prices during nighttime. Commercial complexes load levels correlate with foot traffic, peaking mainly during daytime and early morning hours. Cement production facilities can implement flexible shutdowns, with peak demand concentrated at night to reduce energy costs. General equipment manufacturers experience fluctuating loads tied to working hours, with minor peaks occurring both day and night due to three-shift work schedules. By understanding these industrial users consumption patterns, power grid companies can optimize generation and transmission plans while ensuring timely equipment procurement for grid safety. For electricity retailers, this data enables them to develop targeted marketing strategies and sales packages, while also facilitating demand response programs to encourage users use electricity in the valley section to reduce the consumption of electricity during peak periods and reduce the pressure on the power grid. Users can also understand the electricity consumption situation in time according to the forecast results, and make production scheduling plans in advance to reduce the electricity cost.

A detailed quantitative comparison of the forecasting errors across all 16 industries is provided in Table 2. Consistent with Table 1, the error metrics (MAE, MSE, RMSE) are dimensionless due to data normalization, while MAPE and MSPE are reported in percentage (%).

6.3 Parameter Sensitivity Analysis

To evaluate the robustness of the GBCTT model, we conducted a comprehensive sensitivity analysis on two critical hyperparameters of the GBCTT model: the number of attention heads in the Graph Attention Network (GAT) and the dilation factor in the Temporal Convolutional Network (TCN). This analysis aims to validate the robustness of our model configuration and demonstrate that the chosen parameters are optimal.

The experiments were performed on the Panama dataset using the 7-day forecasting task, with Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) as evaluation metrics.

The analysis reveals a clear trend: model performance exhibits a U-shaped curve relative to both parameters. An insufficient number of attention heads (e.g., 1 or 2) fails to adequately capture the complex spatial interactions among multiple influencing factors, such as the nonlinear coupling between temperature and electricity price. Conversely, an excessive number of heads (e.g., 8) may introduce redundant noise or lead to overfitting. Similarly, a small dilation factor limits the model's receptive field, hindering its ability to learn long-term temporal dependencies, while an excessively large factor may overlook short-term fluctuations. As detailed in Table 3, the configuration with 4 attention heads and a dilation factor of 4 achieves the optimal balance, yielding the lowest MAE and RMSE. This confirms that our parameter selection is well-founded and that the model's performance is robust to minor perturbations in these hyperparameters.

Table 3: Parameter sensitivity analysis on the panama dataset (7-day forecast)

Parameter	Value	MAE	RMSE
Attention heads	1	40.115	47.332
	2	39.254	46.105
	4	38.647	44.808
	8	39.881	47.995
Dilation factor	1	41.569	49.123
	2	39.742	46.874
	4	38.647	44.808
	8	39.025	45.912

6.4 Ablation Studies

To quantitatively evaluate the contribution of each core component in the GBCTT architecture, we designed a series of ablation experiments. These experiments were conducted on the Chinese industrial and commercial dataset, focusing on a 7-day prediction horizon. We systematically removed or replaced key modules and compared the performance against the full GBCTT model. The results are summarized in Table 4.

Table 4: Ablation study results on china dataset (7-day forecast)

Model variant	5G base stations	Steel industry	Commercial complex	Average MAE
w/o GAT	0.512	0.089	0.105	0.235
TCN→BiLSTM	0.421	0.075	0.091	0.196
Full GBCTT	0.380	0.056	0.072	0.169

The ablation studies lead to the following key findings:

- (1) Importance of Graph Attention (GAT): When the GAT module is removed, the model degrades to a purely temporal model. This results in a significant performance drop across all selected industries, as

evidenced by the increased MAE values. For instance, the MAE for the steel industry increased from 0.056 to 0.089. This underscores the critical role of the GAT module in explicitly modeling the spatial correlations and dynamic interactions among various influencing factors (e.g., temperature, humidity, electricity prices).

- (2) Superiority of Temporal Convolutional Network (TCN): Replacing the TCN module with a commonly used BiLSTM model leads to a clear degradation in performance. While BiLSTM is effective for sequence modeling, its sequential processing nature and difficulty in capturing very long-range dependencies make it less suitable for this task compared to the parallelizable and dilated TCN.
- (3) Synergistic Effect: The full GBCTT model, which integrates both GAT and TCN, achieves the best performance across the board. This demonstrates that the spatial feature learning facilitated by GAT and the multi-scale temporal pattern extraction by TCN are complementary and mutually reinforcing. The integration provides a unified framework for spatio-temporal feature learning, which is the cornerstone of the model's superior performance.

6.5 Summary of Experiments

Extensive experimental results in this section demonstrate that when the prediction horizon is set to 7 days, the GBCTT model achieves the best performance across 12 industries (measured by MAE metric) and ranks second in 4 industries. The experimental outcomes validate the outstanding performance and broad applicability of the proposed GBCTT model in practical industrial and commercial load forecasting.

7 Conclusions

Based on the research findings, the proposed GBCTT model effectively overcomes the limitations of traditional methods in capturing the complex spatio-temporal dependencies of electricity load. Experimental results on two distinct datasets demonstrate its superior accuracy and robustness for short-term load forecasting. The model's modular design ensures adaptability for integrating future algorithmic advances. This study provides an effective solution for enhancing energy management and supporting strategic decisions in electricity markets.

Despite the promising results achieved by the GBCTT model, this study acknowledges certain limitations that pave the way for future research. First, the model's performance depends on data quality, and sensor network imperfections may affect forecasting accuracy despite preprocessing. Second, the graph attention mechanism captures mainly static spatial relationships, while real-world factor interactions are dynamic. Third, model interpretability remains limited for non-technical users. Finally, validation on specific regional datasets (Panama and China) requires further testing for broader geographical generalization.

For the future research, promising directions include: (1) Developing dynamic graph attention for evolving factor relationships; (2) Applying transfer learning for cross-regional adaptation; (3) Integrating explainable AI for better transparency; (4) Incorporating additional data sources like macroeconomic indicators; (5) Extending to ultra-short-term forecasting for real-time grid operations.

Acknowledgement: Not applicable.

Funding Statement: This work was supported by the State Grid Corporation of China (5400-202355767A-3-5-YS).

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Wanxing Sheng; methodology, Xiaoyu Yang and Zewei Chen; software, Dongli Jia; validation, Keyan Liu and Zewei Chen; writing—original draft preparation, Zewei Chen; writing—review and editing, Xiaoyu Yang and Zewei Chen; visualization, Keyan Liu; supervision, Wanxing Sheng; project administration, Wanxing Sheng; funding acquisition, Keyan Liu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the Corresponding Author, Zewei Chen, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. El Barkouki B, Mahir O, Laamim M, Rochd A, Ouassaid M, Oufettoul H. Energy efficiency and optimal operation of a residential microgrid based on demand side management strategy. In: Proceedings of the 9th International Youth Conference on Energy (IYCE); 2024 Jul 2–6; Colmar, France. Piscataway, NJ, USA: IEEE; 2024. p. 1–6. doi:10.1109/IYCE60333.2024.10634959.
2. Shafiei K, Seifi A, Hagh MT. A hybrid approach to microgrid energy optimization: integrating renewable energy sources with demand-side management. In: Proceedings of the 29th International Electrical Power Distribution Conference (EPDC); 2025 Oct 28–30; Tehran, Iran. Piscataway, NJ, USA: IEEE; 2025. p. 1–6. doi:10.1109/EPDC67173.2025.11278280.
3. Weidong C, Yun Z, Ning W, Ziwen C, Xiaorui W, Yuxin L. Carbon reduction operation strategy for power systems with renewable energy under the coordinated configuration of traditional and new energy generators. In: Proceedings of the 5th International Conference on Advances in Electrical, Electronics and Computing Technology (EECT); 2025 Mar 21–23; Guangzhou, China. Piscataway, NJ, USA: IEEE; 2025. p. 1–6. doi:10.1109/EECT64505.2025.10966945.
4. Liu J. Cluster analysis of electricity consumption characteristics and electricity consumption behavior of typical industry users. In: Proceedings of the 5th Asia Energy and Electrical Engineering Symposium (AEEES); 2023 Mar 23–26; Chengdu, China. Piscataway, NJ, USA: IEEE; 2023. p. 1389–95. doi:10.1109/AEEES56888.2023.10114302.
5. Liu T, Li S, Ma Y, He Z. Analysis model of non-destructive regional user power consumption behavior based on federated learning. In: Proceedings of the 3rd International Conference on Electronic Information Engineering and Computer Science (EIECS); 2023 Sep 22–24; Changchun, China. Piscataway, NJ, USA: IEEE; 2023. p. 402–5. doi:10.1109/EIECS59936.2023.10435424.
6. Setiawan A, Arifin Z, Sudiarto B, Jufri FH, Haramaini Q, Garniwa I. Comparison of medium-term load forecasting methods (splitted linear regression and artificial neural networks) in electricity systems located in tropical regions. In: Proceedings of the 3rd International Conference on Clean and Green Energy Engineering (CGEE); 2022 Aug 28–30; Istanbul, Turkey. Piscataway, NJ, USA: IEEE; 2022. p. 84–8. doi:10.1109/CGEE55282.2022.9976521.
7. Wen J, Liu Z, Wang G, Zhang X, Xu Y, Li R, et al. A review of research on power system load forecasting methods. In: Proceedings of the 5th International Conference on Smart Grid and Energy Engineering (SGEE); 2024 Nov 22–24; Nanchang, China. Piscataway, NJ, USA: IEEE; 2024. p. 55–61. doi:10.1109/SGEE64306.2024.10865908.
8. Rabbi F, Tareq SU, Islam MM, Chowdhury MA, Kashem MA. A multivariate time series approach for forecasting of electricity demand in Bangladesh using ARIMAX model. In: Proceedings of the 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI); 2020 Dec 19–20; Dhaka, Bangladesh. Piscataway, NJ, USA: IEEE; 2020. p. 1–5. doi:10.1109/STI50764.2020.9350326.
9. Gao D, Li L, Zhu T, Lu H. Medium and short-term wind power prediction based on ARIMA and LSTM models. In: Proceedings of the 2nd International Conference on Smart Grid and Artificial Intelligence (SGAI); 2025 Mar 21–23; Changsha, China. Piscataway, NJ, USA: IEEE; 2025. p. 1206–9. doi:10.1109/SGAI64825.2025.11010023.
10. Ding B, Xu H, Li S, Zong J. Power prediction of mine wind power supply system based on TCN-Transformer. In: Proceedings of the 7th International Conference on Energy Systems and Electrical Power (ICESEP); 2025 Jun 20–22; Wuhan, China. Piscataway, NJ, USA: IEEE; 2025. p. 714–7. doi:10.1109/ICESEP66633.2025.11155392.
11. Feng T, Chen Y, Liu X, Chen T, Tan J, Gao M, et al. Power system operation mode adjustment based on GNN broadcasting and gradient descent. In: Proceedings of the 6th International Conference on Clean Energy and Electric Power Engineering (ICCEPE); 2025 Aug 15–17; Yangzhou, China. Piscataway, NJ, USA: IEEE; 2025. p. 257–61. doi:10.1109/ICCEPE66357.2025.11193341.

12. Liu Z, Xiao S, Liu H. Multi-step short-term wind power prediction based on spatio-temporal graph convolutional networks. In: Proceedings of the 6th International Conference on Renewable Energy and Power Engineering (REPE); 2023 Sep 15–17; Beijing, China. Piscataway, NJ, USA: IEEE; 2023. p. 352–7. doi:10.1109/REPE59476.2023.10511381.
13. Zhang R, Qu H. Short-term prediction of power load for urban residents based on VMD-SVR-PSO model. In: Proceedings of the 3rd International Conference on Electronic Communication and Artificial Intelligence (IWECAI); 2022 Jan 14–16; Zhuhai, China. Piscataway, NJ, USA: IEEE; 2022. p. 519–24. doi:10.1109/IWECAI55315.2022.00106.
14. Mei F, Gu J, Lu J, Lu J, Zhang J, Jiang Y, et al. Day-ahead nonparametric probabilistic forecasting of photovoltaic power generation based on the LSTM-QRA ensemble model. IEEE Access. 2020;8:166138–49. doi:10.1109/ACCESS.2020.3021581.
15. Shahinzadeh H, Nafisi H, Zamani A, Mehrabani-Najafabadi S, Mahmoudi A, Ebrahimi F. Probabilistic short-term load forecasting using GBDT-based sister forecasts and ensemble methods. In: Proceedings of the 15th International Conference on Computer and Knowledge Engineering (ICCCKE); 2025 Oct 28–29; Mashhad, Iran. Piscataway, NJ, USA: IEEE; 2025. p. 1–6. doi:10.1109/ICCCKE68588.2025.11273339.
16. Masood Z, Gantassi R, Choi Y. Enhancing short-term electric load forecasting for households using quantile LSTM and clustering-based probabilistic approach. IEEE Access. 2024;12:77257–68. doi:10.1109/ACCESS.2024.3406439.
17. Jiang L, Li M, Zhao B, Wang Z, Chen C, Zhang Y, et al. Simulation analysis of time of use electricity pricing strategy for electric vehicle charging piles based on genetic algorithm. In: Proceedings of the 8th International Conference on Power Energy Systems and Applications (ICOPESA); 2024 Jun 24–26; Hong Kong, China. Piscataway, NJ, USA: IEEE; 2024. p. 661–5. doi:10.1109/ICOPESA61191.2024.10743696.
18. Zhang W, Quan H, Gandhi O, Rajagopal R, Tan CW, Srinivasan D. Improving probabilistic load forecasting using quantile regression NN with skip connections. IEEE Trans Smart Grid. 2020;11(6):5442–50. doi:10.1109/TSG.2020.2995777.
19. Akram W, Ramezani R, Ehsan AJ, Khalid B, Bokhari SAS, Saleem H. Short-term wind power forecasting via temporal convolutional networks. In: Proceedings of the International Conference on Electrical, Communication and Computer Engineering (ICECCE); 2025 Aug 27–28; Istanbul, Turkiye. Piscataway, NJ, USA: IEEE; 2025. p. 1–5. doi:10.1109/ICECCE67514.2025.11258009.
20. Li X, Zheng Y. Short-term wind power prediction based on TCN-Transformer and STL error correction. In: Proceedings of the IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia); 2024 Jul 9–12; Pattaya, Thailand. Piscataway, NJ, USA: IEEE; 2024. p. 462–7. doi:10.1109/ICPSAsia61913.2024.10761501.
21. Zhao Y, Pu H, Dai Y, Cong S. Short-term wind power prediction method based on GCN-LSTM. In: Proceedings of the 8th Asia Conference on Power and Electrical Engineering (ACPEE); 2023 Apr 14–16; Tianjin, China. Piscataway, NJ, USA: IEEE; 2023. p. 1329–33. doi:10.1109/ACPEE56931.2023.10135621.
22. Trigui A, Olama M, Siopsis G, Eldakhakhni H, Salhi M. Graph-based attention mechanisms for solving the AC optimal power flow problem in electrical power networks. In: Proceedings of the 57th North American Power Symposium (NAPS); 2025 Oct 26–28; Storrs, CT, USA. Piscataway, NJ, USA: IEEE; 2025. p. 1–6. doi:10.1109/NAPS66256.2025.11272251.
23. Li Z, Li Y, Li S, Han Y. Distributed photovoltaic power prediction based on GAT-TST. In: Proceedings of the IEEE 8th Conference on Energy Internet and Energy System Integration (EI2); 2024 Nov 29–Dec 2; Shenyang, China. Piscataway, NJ, USA: IEEE; 2024. p. 4453–8. doi:10.1109/EI264398.2024.10991451.
24. Hu Q. Research on power prediction of distributed photovoltaic power stations based on TCN-GCN. In: Proceedings of the 2nd International Conference on Smart Grid and Artificial Intelligence (SGAI); 2025 Mar 21–23; Changsha, China. Piscataway, NJ, USA: IEEE; 2025. p. 513–8. doi:10.1109/SGAI64825.2025.11009754.
25. Wang H, Kang H, Liu H, Huang H, Wu Y, Li Z, et al. High energy load prediction of industrial parks based on TCN-BiGRU-Attention. In: Proceedings of the 6th International Conference on Energy Systems and Electrical Power (ICESEP); 2024 Jun 21–23; Wuhan, China. Piscataway, NJ, USA: IEEE; 2024. p. 89–94. doi:10.1109/ICESEP62218.2024.10652227.

26. Lixiong F, Yusen S, Zhaomin J, Yi R, Aimer M, Ruixiang X. GAT-BiLSTM: graph attention based BiLSTM for non-intrusive load monitoring. In: Proceedings of the 9th International Conference on Power Energy Systems and Applications (ICoPESA); 2025 Jun 20–22; Nanjing, China. Piscataway, NJ, USA: IEEE; 2025. p. 389–93. doi:10.1109/ICoPESA65876.2025.11234578.
27. Liu S, Wang H, Song W, Han S, Yan J, Liu Y, et al. Wind power forecasting method under extreme weather conditions based on meta-learning and multi-head attention mechanism. In: Proceedings of the IEEE 8th Conference on Energy Internet and Energy System Integration (EI2); 2024 Nov 29–Dec 2; Shenyang, China. Piscataway, NJ, USA: IEEE; 2024. p. 4471–6. doi:10.1109/EI264398.2024.10991900.
28. Li C, Kong X, Wang X, Zheng F, Chen Z, Zhou Z. Short-term wind power forecasting method based on mode decomposition and feature extraction. In: Proceedings of the 22nd International Conference on Electrical Machines and Systems (ICEMS); 2019 Aug 11–14; Harbin, China. Piscataway, NJ, USA: IEEE; 2019. p. 1–5. doi:10.1109/ICEMS.2019.8922193.
29. Xu J, Hou H, Ni T, Luo C, Chen L, Kang Y, et al. Short-term power load forecasting based on LightGBM-VMD-SE and inception-BiGRU-attention. In: Proceedings of the IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia); 2024 Jul 9–12; Pattaya, Thailand. Piscataway, NJ, USA: IEEE; 2024. p. 85–90. doi:10.1109/ICPSAsia61913.2024.10761589.
30. Srinivas S, Jayaram A, Bhavadharani K, Pant KS, Thirumala K, Kumar TS. Classification of power quality disturbances using convolutional neural network and temporal convolutional network models. In: Proceedings of the 23rd National Power Systems Conference (NPSC); 2024 Dec 14–16; Indore, India. Piscataway, NJ, USA: IEEE; 2024. p. 1–6. doi:10.1109/NPSC61626.2024.10987033.