A Spatio-Temporal Neural Network for Long-Term Load Profile Forecasting with Topological Information

Fengbin An, Nanpeng Yu, Zhentong Shao, Yuzhou Chen
Department of Electrical and Computer Engineering
University of California, Riverside
Riverside, CA, USA
{fan006, nyu, zhentons, yuzhouc}@ucr.edu

Aowabin Rahman, Casey Burleyson
Konstantinos Oikonomou, Osten Anderson
Pacific Northwest National Laboratory
{aowabin.rahman, casey.burleyson}@pnnl.gov
{konstantinos.oikonomou, osten.anderson}@pnnl.gov

***Abstract*—Load profile forecasting plays a critical role in long-term power system planning. Traditional load profile forecasting approaches primarily focus on temporal patterns in historical demand, while often neglecting spatial correlations. As electricity demand becomes increasingly influenced by weather and socioeconomic demand indicators, accurate modeling requires the integration of both temporal and spatial information. This paper proposes a spatio-temporal graph convolutional network (ST-GCN) with multi-head attention for load profile forecasting across multiple balancing authorities (BAs). The proposed framework captures temporal dynamics, spatial dependencies among neighboring regions, and explanatory variables including temperature, humidity, solar radiation, wind speed, and socioeconomic demand indicators. Furthermore, zigzag persistence is employed to extract topological information from time-varying adjacency matrices, revealing the underlying structural evolution of the regional weather patterns. Experimental results based on real-world data from the Western Electricity Coordinating Council (WECC) region demonstrate that the proposed method improves forecasting accuracy by 23% compared with baseline methods.**

***Index Terms*—Load forecasting, socioeconomic demand, spatio-temporal correlation, zigzag persistence.**

I. Introduction

Accurate hourly load profiles that capture weather and socioeconomic demand influences are essential for reliable power system planning, including investment, capacity expansion, and reliability assessment [1]. Accurate spatio-temporal load profiles enable planners to evaluate system resilience under stressed conditions such as extreme weather events, regional demand surges, and electrification-driven variability [2]. These needs motivate the development of advanced data-driven and graph-based modeling frameworks capable of jointly capturing the spatial, temporal, and socioeconomic demand dependencies of large-scale electricity demand [3].

Electrical load is primarily driven by two categories of factors: meteorological and socioeconomic demand. Weather variables–such as temperature, humidity, solar radiation, and wind speed–strongly influence heating and cooling demand, yet their nonlinear and spatially heterogeneous effects remain challenging to model accurately [4]. Meanwhile, socioeconomic demand factors such as population growth, economic development, and sectoral electrification can also affect demand patterns [5]. With the recent rapid deployment of large electric loads, like data centers, and the electrification of other loads, increasing load magnification and variability introduce additional spatial–temporal complexities [6]. These emerging conditions underscore the need for advanced spatio-temporal load profile forecasting methods capable of adapting to evolving grid dynamics.

Most load forecasting studies have focused on single-region modeling, where the demand within each region was predicted independently from its own historical load series. For instance, Lin et al. [7] proposed a temporal graph model that captures intra-regional temporal dependencies but does not consider inter-regional interactions. To address this, multi-region frameworks have been introduced to exploit spatial correlations across areas. Liu et al. [8] jointly modeled multiple districts through shared learning structures; however their model lacks explicit representation of regional topology. Similarly, Ye et al. [9] integrated geographic and socioeconomic demand features to capture regional diversity but relies on static clustering, limiting its ability to reflect evolving dependencies. While these efforts mark progress toward multi-region load profile forecasting, there is a gap in existing work in modeling the time-varying spatial relationships arising from shared weather patterns, grid coupling, and socioeconomic connections. This gap motivates the need for a graph-based spatio-temporal framework that explicitly encodes dynamic spatial weather patterns and captures the time-varying inter-regional correlations essential for system-level forecasting and planning.

Traditional weather dependent machine learning methods model load variations by capturing nonlinear relationships between electricity demand and meteorological factors [10], [11]. Ding et al. [10] proposed an adaptive denoising hybrid model combining TDeepTCN and attention-enhanced BiLSTM for short-term load forecasting in cyber-physical energy systems, explicitly leveraging temperature-related features. Similarly, Wang et al. [11] employed Temporal Convolutional Network (TCN) to extract temporal dependencies in load and temperature data, followed by LightGBM for nonlinear feature fusion and residual correction. However, most existing approaches focus primarily on meteorological inputs, overlooking socioeconomic demand drivers that shape long-term

This material is based upon work supported by the U.S. Department of Energy (DOE). Pacific Northwest National Laboratory is operated for the DOE by Battelle Memorial Institute under Contract DE-AC05-76RL01830.

demand evolution. A robust load profile forecasting framework should therefore integrate meteorological and and long-term socioeconomic demand influences within a unified spatio-temporal representation capable of learning their nonlinear dependencies and dynamic inter-regional correlations.

The topological dependencies among BAs are inherently time-varying, shaped by seasonal and meteorological factors that influence inter-regional correlations. Modeling the system as a sequence of time-varying graphs enables explicit representation of these evolving topological properties, while Graph Neural Networks (GNNs) provide a flexible framework to jointly encode spatial and temporal dynamics. Chen et al. [12] introduced Z-GCNETs, the first framework to integrate zigzag persistent homology into GNNs for time-series forecasting. Subsequently, Tinarrage et al. [13] proposed ZigzagNetVis, which leverages zigzag persistent homology to identify meaningful temporal resolutions for evolution analysis. These advances motivate us to integrate zigzag persistent homology to enhance spatio-temporal load profile forecasting for large-scale, multi-region forecasting under evolving socioeconomic and environmental conditions.

Based on the existing research gap, the main contributions of this paper are summarized as follows:

- We proposed a spatio-temporal graph convolutional network (ST-GCN) with an attention mechanism that jointly captures weather dependencies and socioeconomic demand influences on regional electricity demand.
- We incorporated zigzag persistence to extract and encode dynamic topological relationships among time-varying weather correlation graphs, enabling the ST-GCN to model evolving spatial correlations over time.

The organization for the rest of the paper is as follows. Section II gives the framework of the proposed ST-GCN. Section III elaborates the technical methods of the proposed ST-GCN. Numerical study results are given in Section IV. Finally, the conclusions are presented in Section V.

II. Problem Formulation

The load profile forecasting problem is formulated as a spatio-temporal learning task over a network of N BAs. A dynamic graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is constructed to represent time-varying dependencies among BAs, where $\mathcal{V}$ denotes the set of BAs and $\mathcal{E}$ the set of edges capturing inter-regional relationships. The adjacency matrix $\mathbf{A}$ of size $N \times N$ encodes temperature-based correlations among the BAs, where $\mathbf{A}_{ij} = 1$ if $e_{ij} \in \mathcal{E}$ and $\mathbf{A}_{ij} = 0$ otherwise.

For each BA $i \in \{1, \ldots, N\}$, we observe an hourly time series of electricity demand along with meteorological variables such as temperature, solar radiation, and wind speed. Spatial dependencies are represented by a set of adjacency matrices $\{\mathbf{A}_m\}_{m=1}^{M}$, where each $\mathbf{A}_m \in \mathbb{R}^{N\times N}$ encodes the inter-BA connectivity during the m-th time period. Temporal dependencies arise naturally from the sequential nature of load data. At each time step t, the feature vector of BA i is defined as $\mathbf{x}_i^{(t)} = [x_{i,1}^{(t)}, x_{i,2}^{(t)}, \ldots, x_{i,F}^{(t)}]^\top \in \mathbb{R}^F$, where F denotes the number of input features. For model training, a sliding window of length L hours is used to construct input sequence $\mathbf{X}_{1:L} = \{\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(L)}\}$, where $\mathbf{X}^{(l)} = [\mathbf{x}_1^{(l)}, \ldots, \mathbf{x}_N^{(l)}]^\top \in \mathbb{R}^{N\times D}$ represents the feature matrix of all N BAs at time step l. The model learns to map the input sequence $\mathbf{X}_{1:L}$ to the corresponding target load sequence $\mathbf{Y}_{1:L} = [y_1^{(1:L)}, \ldots, y_N^{(1:L)}]^\top \in \mathbb{R}^{N\times L}$, where $y_i^{(1:L)}$ denotes the target load sequence of BA i.

III. Technical Methods

A. Overall Framework

As illustrated in Fig. 1, the proposed spatio-temporal Graph Convolutional Network (ST-GCN)-based load profile forecasting framework consists of four main stages. Unlike conventional spatio–temporal forecasting approaches that depend on static or implicitly learned graph structures, the proposed ST-GCN explicitly incorporates spatial correlation, topological evolution, and socioeconomic demand heterogeneity to capture the dynamic and multi-scale nature of regional electricity demand. The framework employs a Seq2Seq structure to map weather and temporal inputs to BA load sequences.

First, weather variables and temporal indicators are processed using a GCN. For each BA, the input consists of an L-hour sequence of historical features, including temperature, humidity, longwave and shortwave radiation, wind speed, and temporal indices such as year, month, day, hour and binary indicators for weekdays and holidays. The GCN layer transforms node inputs into spatially informed representations that capture inter-regional correlations among BAs, while the framework adopts a local design to preserve spatial locality. Specifically, each ST-GCN model is trained to forecast the target BA's load using its neighbors' information, thus preserving localized spatial dependencies and mitigating over-smoothing [14].

Next, zigzag persistence is applied to extract temporal topological signatures from time-varying temperature correlation graphs, as detailed in the following subsection. The resulting zigzag persistence diagrams capture the birth and death of each local topological feature over time and then are subsequently converted into persistence images which allow to encode the density distribution of these topological features across different lifespans. These topology-aware images can be processed by a CNN to learn high-level topological embeddings, which are concatenated with each BA's GCN output to construct a unified spatio-temporal representation.

The concatenated spatio-temporal embeddings are subsequently fed into a Long Short-Term Memory (LSTM) network augmented with a multi-head attention mechanism. The LSTM captures temporal dependencies, while the attention mechanism adaptively emphasizes the most influential time steps that drive load fluctuations. Finally, we combine the LSTM outputs and socioeconomic demand features (i.e., customer count and energy consumption by sector) which are encoded by an multi-layer perceptron (MLP). The fused representation is then fed to a final MLP to predict the load profile of the target BA, effectively integrating temporal dynamics with socioeconomic demand influences to enhance forecasting accuracy.

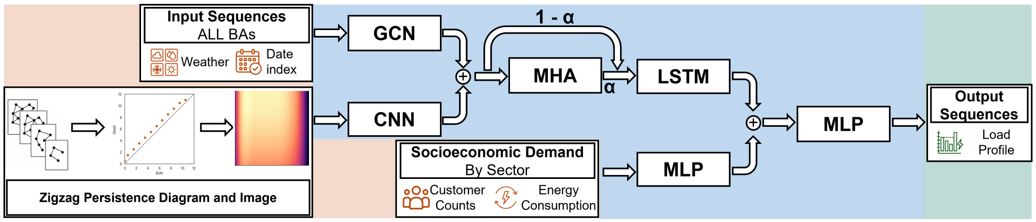

Fig. 1: Illustration of the whole framework of ST-GCN for load profile forecasting.

B. Graph Construction

Although static topologies manage to capture overall structural relationships among BAs, they fail to capture evolving inter-BA dependencies induced by seasonal and meteorological variations. To address this, we construct a dynamic k-nearest neighbour (kNN) graph based on weather similarity among BAs [15]. We consider the characteristic time period to span one month and as such, we generate monthly adjacency matrices $\{\mathbf{A}^{(m)}\}_{m=1}^{12}$ using the Pearson correlation coefficients of the hourly temperature series across all BAs.

Let Ω_i denote the index set of the top five nodes most correlated with node i. For each month m, the adjacency matrix $A_{ij}^{(m)}$ is defined as 1 if either node j is among the top-k most correlated nodes with node i, or node i is among the top-k most correlated nodes with node j; otherwise, it is set to 0. This procedure produces a set of temperature-driven, adaptive symmetric graphs that more accurately capture the evolving spatial dependencies among regions.

C. Socioeconomic Demand Features

To capture distinct growth patterns across load sectors, customer counts and energy consumption data at the BA level are collected for residential, commercial, industrial, and transportation users. Since these data are typically reported annually, linear interpolation is applied to generate smooth hourly-level variations and prevent abrupt intra-year changes.

For a given month $m \in \{1, 2, \ldots, 12\}$, the interpolation weight is defined as: $w = \frac{m-1}{12}$. The interpolated feature vector $\tilde{f}_i^{(year,m)}$ for BA i at month m of year ($year$) is given by $\tilde{f}_i^{(year,m)} = (1-w)\, f_i^{(year)} + w\, f_i^{(year+1)}$.
This formulation preserves the smooth temporal evolution of customer characteristics between annual reporting periods.

D. ST-GCN

To capture both spatial and temporal dependencies in load profile forecasting, we adopt the ST-GCN framework [16]. The hybrid ST-GCN architectures with residual connections and multi-step recursive forecasting are further explored to enhance stability and extended prediction horizons.

Specifically, a GCN is used to capture spatial correlations among BAs. At each time step t, the input feature matrix $X^{(t)} \in \mathbb{R}^{N \times F}$ comprises F weather and calendar features for N BAs. The spatial dependencies among regions are modeled through a graph convolutional layer defined as:

$$h_i^{(t)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} \tilde{A}_{ij} W_g x_j^{(t)}\right), \tag{1}$$

where $\tilde{A} = D^{-\frac{1}{2}}(A+I)D^{-\frac{1}{2}}$ is the normalized adjacency matrix, D is the the diagonal node degree matrix of $\tilde{A}$, and W_g denotes the learnable weight matrix.

A residual connection is introduced after the graph convolution to enhance training stability and preserve node-level information: $\tilde{h}_i^{(t)} = \alpha\, h_i^{(t)} + (1-\alpha)\, W_r x_i^{(t)}$, where $\alpha = \sigma(\theta)$ is a learnable gating scalar that tunes importance of the current graph features and original input representation, and W_r is a linear projection corresponding to the residual connection.

Next, a multi-head attention (MHA) mechanism is applied to the residual-enhanced representations $\{\tilde{h}_1^{(t)}, \ldots, \tilde{h}_T^{(t)}\}$ to capture long-range temporal dependencies and emphasize informative time steps:

$$\mathrm{MHA}(H') = \mathrm{Concat}(\mathrm{head}_1, \ldots, \mathrm{head}_K)\, W^O, \tag{2}$$

$$\mathrm{head}_k = \mathrm{Softmax}\left(\frac{Q_k K_k^\top}{\sqrt{d_k}}\right) V_k, \tag{3}$$

where $Q_k = H'W_k^Q$, $K_k = H'W_k^K$, and $V_k = H'W_k^V$ denote the query, key, and value projections for the k-th head, d_k is the head dimension, and W^O is the output projection matrix.

The attention-refined temporal representation $H^{\mathrm{att}} = \mathrm{MHA}(H')$ is then passed to an LSTM to capture short-term sequential dependencies $(h_t, c_t) = \mathrm{LSTM}(H_t^{\mathrm{att}}, h_{t-1}, c_{t-1})$, where h_t and c_t represent the hidden and cell states at time t.

Finally, socioeconomic demand features are incorporated to capture long-term socioeconomic influences on electricity demand. A static embedding $\mathbf{s}_i = f_{\mathrm{static}}(\mathbf{e}_i)$ is generated using a MLP that projects customer count and energy consumption data into the same latent space as the dynamic representation for BA i. The final load prediction is obtained by combining the temporal hidden state and the static embedding $\hat{y}_i^{(t+1)} = f_{\mathrm{MLP}}(\mathrm{Concat}(h_t, \mathbf{s}_i))$, where $f_{\mathrm{MLP}}(\cdot)$ denotes the output regression layer producing the next-step load forecast.

E. Zigzag Persistence

For time-varying graphs, topological information captures higher-order structural properties that extend beyond edge weights. As described in Section III.B, the graph evolves with weather-driven correlations, causing its connectivity patterns and topological signatures to dynamically emerge or vanish over time. To tackle this challenge, zigzag persistence [17] provides a principled framework for quantifying these temporal topological changes by tracking the birth and death of local topological features across evolving graph structures [12].

Each topological point p is characterized by a birth-death pair (b_p, d_p), indicating the time interval over which a topological feature exists. The collection of these points forms the persistence diagram $\mathrm{Dgm} = \{(b_p, d_p) \in \mathbb{R}^2 | b_p < d_p\}$, which compactly represents the evolution of structural features in the time-varying graph. The persistence $b_p - d_p$, quantifies the stability of each feature, with longer lifetimes reflecting features that persist longer across scales and thus possess higher topological significance. For computational purpose, we encode these topological points (i.e., features) in a differentiable form, i.e., the Zigzag Persistence Image (ZPI), which is formulated as:

$$\mathrm{ZPI} = \iint_z \sum_{\mu \in \mathrm{Dgm}} g(\mu) \exp\left(-\frac{\|z-\mu\|^2}{2\sigma^2}\right) dz_{\mathrm{x}}\, dz_{\mathrm{y}}, \quad (4)$$

where $z = (x, y)$ denotes a point in the birth–death plane. Each Gaussian kernel centered at μ corresponds to a persistence pair in the diagram, with high-intensity regions in the ZPI highlighting structural features that remain stable over extended periods. The resulting topological embedding $\mathbf{z}_i$ is extracted from the ZPI representation of each region's evolving topology $\mathbf{z}_i = f_{\mathrm{ZPI}}(\pi_i)$, which encodes temporal variations in the graph's topological structure.

Finally, the concatenated feature $[\tilde{h}_i^{(t)}, \mathbf{z}_i]$ is fed into a recurrent neural network (RNN)-based model to capture temporal dependencies and sequential evolution of load-related patterns $(h_t, c_t) = \mathrm{LSTM}\Big([\tilde{h}_i^{(t)}, \mathbf{z}_i], h_{t-1}, c_{t-1}\Big)$.

F. Model Training

The model parameters θ are trained by minimizing the mean squared error (MSE) between the predicted and actual loads across all BAs and time steps:

$$\mathcal{L}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_i^{(t)} - \hat{y}_i^{(t)}\right)^2, \quad (5)$$

where $y_i^{(t)}$ and $\hat{y}_i^{(t)}$ denote the observed and predicted loads of BA i at time t, respectively.

Optimization is performed using the Adam optimizer with an initial learning rate of 1×10^{-3} and a batch size of 32. Training proceeds for up to 150 epochs, with early stopping triggered if the validation loss fails to improve for 20 consecutive epochs, ensuring convergence and preventing overfitting.

IV. Numerical Study

A. Data Description and Preprocessing

In this study, we use hourly data from 28 BAs in the Western Electricity Coordinating Council (WECC) spanning 2016–2024 [18] to evaluate our proposed model. The dataset integrates multiple information, including electric load, meteorological variables derived from historical reanalysis data (temperature, humidity, shortwave and longwave radiation, and wind speed), and BA level socioeconomic demand indicators such as electricity consumption and customer counts across four sectors–residential, commercial, industrial, and transportation [19]. To ensure high historical load data quality, a seasonal outlier filtering procedure is applied. Specifically, interquartile range (IQR) filtering is used to remove erroneous demand values within each month based on the 25%–75% percentile range. We select the multiplier associated with the filter leniently (i.e., $3 \leq k \leq 7$, depending on each BA's load outlier severity), so as to remove these erraneous values while retaining the temporal dynamics of the demand profile.

The dataset is chronologically divided into training (January 2016–December 2022), validation (January 2023-December 2023), and testing (January 2024-December 2024) subsets to prevent data leakage and maintain temporal consistency. All variables across the three datasets are normalized using min-max values computed from the training set, ensuring consistent scaling and comparability during model evaluation.

B. Forecasting Accuracy of the Proposed and Baseline Models

The hyper-parameters of the proposed (ST-GCN and baseline models (MLP, GCN, LSTM) are set as follows.

1) MLP: A two-layer fully connected network with 256 and 128 hidden neurons.

2) GCN: Two graph convolutional layers with 256 hidden neurons, followed by three fully connected layers with 128, 64, and 1 neurons.

3) LSTM: Two LSTM layers with 128 hidden neurons and 4 attention heads.

4) ST-GCN: The GCN component uses 32 neurons, while the LSTM module contains 128 neurons and 4 attention heads. The MLP for processing socioeconomic demand features includes two fully connected layers with 4 and 16 hidden neurons, respectively. The final MLP for hourly load prediction consists two fully connected layers with 64 and 1 neuron(s).

To comprehensively evaluate forecasting performance across spatial scales, two sets of metrics are employed:

(i) *Unweighted* RMSE and MAPE, computed as the simple average of RMSE and MAPE across all BAs, assigning equal importance to each region;

(ii) *Weighted* RMSE and MAPE, calculated by weighting each BA according to its mean annual load, thereby giving greater emphasis to high-demand BA.

A comparison of model accuracy using these metrics is presented in Table I. The boldface indicates the best results.

TABLE I: Total and Weighted RMSE and MAPE of the Proposed and Baseline Methods

Method	**Avg. RMSE** (MWh)	**Avg. wtd. RMSE** (MWh)	**Avg. MAPE** (%)	**Avg. wtd. MAPE** (%)
MLP	322.50	879.08	11.25	9.08
GCN	313.05	868.35	10.80	8.85
LSTM	304.50	806.56	11.07	8.78
ST-GCN	**238.84**	**676.07**	**8.23**	**6.53**

As shown in Table I, we observe that the proposed ST-GCN model consistently achieves the highest accuracy across both RMSE and MAPE metrics, outperforming conventional MLP, GCN and LSTM baselines. In Fig. 2, the ST-GCN model exhibits smaller interquartile ranges and fewer extreme outliers

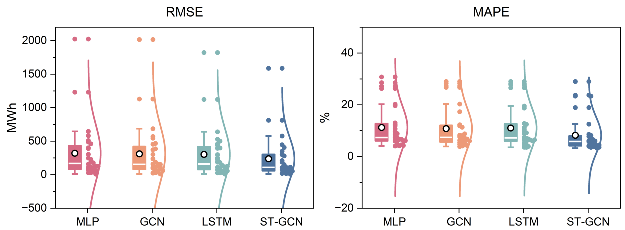

Fig. 2: RMSE and MAPE distributions across 28 BAs.

across the 28 BAs, indicating more stable and consistent forecasting performance. Specifically, relative to the MLP baseline, the ST-GCN reduces the average weighted RMSE by 23.09%, while achieving additional reductions of 1.22% and 8.25% compared with the GCN and LSTM models, respectively. These results highlight the superior ability of the proposed ST-GCN to jointly capture spatial dependencies and temporal dynamics across BAs, leading to more accurate and robust load forecasts under varying system conditions.

C. Ablation Study

To further assess the contribution of individual components, an ablation study is performed to quantify the impact of socioeconomic demand features and Zigzag persistence through 3 model configurations: ST-GCN, ST-GCN without Customer Features, and ST-GCN without Zigzag Persistence.

TABLE II: Total and Weighted RMSE and MAPE of the Proposed Model in Ablation Study

Method	**Avg. RMSE** (MWh)	**Avg. wtd. RMSE** (MWh)	**Avg. MAPE** (%)	**Avg. wtd. MAPE** (%)
ST-GCN	**238.84**	**676.07**	**8.23**	**6.53**
ST-GCN w/o CF	266.16	714.55	9.48	7.48
ST-GCN w/o ZP	256.05	699.81	8.56	7.02

Table II shows that the ST-GCN method achieved the best performance, with a weighted RMSE/MAPE of 676.07 MWh/6.53% outperforming both ST-GCN w/o CF (714.55 MWh/7.48%) and ST-GCN w/o ZP (699.81 MWh/7.02%). The noticeable increase in errors when either customer features (CF) or zigzag persistence (ZP) is removed demonstrates that socioeconomic demand and topological information provide complementary benefits, jointly enhancing the model's representational power and overall forecasting accuracy.

The 0-dimension ZPD shows two persistence pairs, [0.5, 1.5] and [0, 11], indicating that the overall BA network forms early and remains connected throughout the year. In contrast, the 1-dimension ZPD contains 12 pairs distributed slightly above the diagonal, representing the short-lived local loops in the BA network. This pattern reveals monthly variations in inter-BAs correlations driven by temperature dynamics. These results confirm that Zigzag persistence effectively captures both stable and transitional spatial dependencies induced by evolving weather conditions.

V. Conclusion

This paper proposes a novel spatial-temporal load profile forecasting approach that integrates Zigzag persistence with a ST-GCN to capture dynamic topological relationships among BAs. Leveraging multi-year load, weather, and socioeconomic demand data from the WECC region, the proposed framework significantly reduces both RMSE and MAPE compared with baseline models. The results underscore the critical role of incorporating topological evolution and socioeconomic demand factors in improving regional load profile forecasting accuracy. Future work will explore adaptive graph representations to enhance the model's ability to capture complex, evolving spatial dependencies across large-scale power systems.

References

[1] T. Hong, J. Wilson, and J. Xie, "Long-term probabilistic load forecasting and normalization with hourly information," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 456–462, Jan. 2014.

[2] H. Yang, Y. Gao, Y. Ma, and D. Zhang, "Optimal modification of peak-valley period under multiple time-of-use schemes based on dynamic load point method considering reliability," *IEEE Trans. Power Syst.*, vol. 37, no. 5, pp. 3889–3901, Sep. 2022.

[3] C. Wang, Y. Zhou, Q. Wen, and Y. Wang, "Improving load forecasting performance via sample reweighting," *IEEE Trans. Smart Grid*, vol. 14, no. 4, pp. 3317–3320, Jul. 2023.

[4] Y. Ji, E. Buechler, and R. Rajagopal, "Data-driven load modeling and forecasting of residential appliances," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2652–2661, May 2020.

[5] X. Zhu and B. Mather, "Data-driven distribution system load modeling for quasi-static time-series simulation," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1556–1565, Mar. 2020.

[6] I. Zografopoulos, A. Srivastava, C. Konstantinou *et al.*, "Cyber-physical interdependence for power system operation and control," *IEEE Transactions on Smart Grid*, Feb. 2025.

[7] W. Lin, D. Wu, and B. Boulet, "Spatial-temporal residential short-term load forecasting via graph neural networks," *IEEE Trans. Smart Grid*, vol. 12, no. 6, pp. 5373–5384, Nov. 2021.

[8] H. Liu, X. Zhang, H. Sun, and M. Shahidehpour, "Boosted multi-task learning for inter-district collaborative load forecasting," *IEEE Trans. Smart Grid*, vol. 15, no. 1, pp. 973–986, Jan. 2024.

[9] C. Ye, Y. Ding, P. Wang, and Z. Lin, "A data-driven bottom-up approach for spatial and temporal electric load forecasting," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 1966–1979, May 2019.

[10] D. Ding, J. Li *et al.*, "Short-term load forecasting reliability in power plant of cyber-physical energy system considering adaptive denoising," *IEEE Syst. J.*, vol. 17, no. 4, pp. 5183–5194, Dec. 2023.

[11] Y. Wang, J. Chen, X. Chen *et al.*, "Short-term load forecasting for industrial customers based on TCN-LightGBM," *IEEE Trans. Power Syst.*, vol. 36, no. 3, pp. 1984–1997, May 2021.

[12] Y. Chen, I. Segovia-Dominguez, and Y. R. Gel, "Z-GCNETs: Time zigzags at graph convolutional networks for time series forecasting," *Proceedings of the 38 th International Conference on Machine Learning*, p. 1684–1694., 2021.

[13] R. Tinarrage, J. R. Ponciano, C. D. G. Linhares *et al.*, "ZigzagNetVis: suggesting temporal resolutions for graph visualization using zigzag persistence," *IEEE Trans. Visual. Comput. Graphics*, vol. 31, no. 10, pp. 6852–6869, Oct. 2025.

[14] T. K. Rusch, M. M. Bronstein, and S. Mishra, "A survey on oversmoothing in graph neural networks," *arXiv preprint arXiv:2303.10993*, 2023.

[15] L. Franceschi, M. Niepert, M. Pontil, and X. He, "Learning discrete structures for graph neural networks," *Proceedings of the 36 th International Conference on Machine Learning*, p. 1972–1982., 2019.

[16] L. Zhao, Y. Song, C. Zhang *et al.*, "T-GCN: a temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transport. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020.

[17] G. Carlsson and V. De Silva, "Zigzag persistence," *Foundations of computational mathematics*, vol. 10, no. 4, pp. 367–405, 2010.

[18] C. R. McGrath, C. D. Burleyson, Z. Khan *et al.*, "Tell: A Python package to model future total electricity loads in the United States," vol. 7, no. 79, p. 4472, Nov. 2022.

[19] Annual Electric Power Industry Report, Form EIA-861 detailed data files - U.S. Energy Information Administration (EIA). [Online]. Available: https://www.eia.gov/electricity/data/eia861/