AEM JOURNAL .ORG

Short-Term Electricity Load Forecasting in Power Systems Using BO-VMD-STGCN

Zeshuo Sun[1], Tong Qu[2,*], Zeyu Zhao[1], Xiaodan Sun[1], Zhou Shu[1]
[1]Harbin University of Science and Technology Rongcheng College, Weihai, 264200, China
[2]Weihai Zhongyi Hydraulic Technology Co., Ltd. Weihai 264210 China

Corresponding author: Tong Qu (email: qutong19860527@163.com).

ABSTRACT To address the strong nonlinearity, multi-fluctuation coupling, and spatio-temporal correlation characteristics inherent in short-term power load, a forecasting model based on BO-VMD-STGCN (bayesian optimization, variational mode decomposition and spatio-temporal graph convolutional networks) is proposed. Firstly, BO is utilized for the adaptive tuning of key parameters in VMD, thereby achieving an optimal multi-scale decomposition of the original load sequence and mitigating data non-stationarity. Subsequently, the decomposed modal components along with the reconstructed sequences are fed into the STGCN, wherein a stacking mechanism with residual connections is incorporated to synergistically extract temporal dependencies and spatial correlations via spatio-temporal convolutions. Experimental results demonstrate that the proposed model achieves an MAPE, RMSE, and MAE of 2.84%, 62.68, and 38.36, respectively, along with an R2 of 0.92. Compared with LSTM, STGCN, STM-GCN, and GCN-LSTM, the R2 metric is improved by 26.39%, 19.73%, 18.42%, and 35.29%, respectively. Exhibiting marked superiority over benchmark approaches in both goodness-of-fit and predictive fidelity, the proposed model underscores its intrinsic strengths in jointly capturing spatio-temporal dependencies and adaptively disentangling non-stationary temporal dynamics. It thus establishes a reliable, accurate, and interpretable paradigm for short-term electricity load forecasting.

INDEX TERMS Power load forecasting, Spatio-temporal correlation, Variational Mode Decomposition, Spatio-Temporal Graph Convolutional Network

I. INTRODUCTION

MULTI-NODE spatio-temporal power load forecasting constitutes a core enabling technology for the planning and operation of new power systems [1], and it possesses irreplaceable application value for promoting energy structure transformation, optimizing resource allocation, and achieving the strategic objectives of “Carbon peaking and carbon neutrality” [2]. Distinct from conventional single-point forecasting, multi-node prediction necessitates the deep integration of power grid topology with inter-regional spatio-temporal correlation characteristics, thereby enabling collaborative modeling of load variations across individual nodes within the system [3]. Particularly in the context of high-proportion renewable energy integration, accurate short-term load forecasting can effectively mitigate the intermittency and volatility inherent on both the source and load sides, optimize energy storage dispatch strategies, and ensure the reliable operation of distribution networks [4].

In recent years, short-term load forecasting methodologies have progressed from traditional statistical models to machine learning, and eventually to deep learning. Traditional statistical models are grounded in time series analysis theory and achieve prediction through the mathematical modeling of historical load regularities. However, they exhibit inherent limitations in capturing the nonlinear, non-stationary, and multi-seasonal characteristics of load sequences, rendering them inadequate for meeting the high-precision forecasting demands of complex power system scenarios. Reference [5] employed a dual-seasonal Holt-Winters exponential smoothing technique designed to capture the coexisting daily and weekly periodicities in load behavior; nevertheless, it proves inadequate in capturing nonlinear and non-stationary characteristics. Reference [6] constructed a linear regression model based on daily periodic patterns; however, its linear framework constrains the capacity to represent complex nonlinear relationships. Reference [7] employed the ARMA (autoregressive moving average) model to achieve parameter estimation through deseasonalization and order determination; however, this approach presupposes series stationarity, whereas power load data frequently exhibit non-stationarity and multiple seasonality, rendering the preprocessing stage susceptible to information loss. Reference [8] utilized Kalman filtering to realize adaptive forecasting; nevertheless, its linear Gaussian assumption proves inadequate in capturing the nonlinear and non-Gaussian characteristics inherent in load data.

Traditional statistical models, constrained by assumptions of linearity and stationarity, exhibit limited capacity to characterize the nonlinear and non-stationary features of load profiles. In contrast, machine learning methods achieve precise power load forecasting through nonlinear mapping and adaptive learning. Reference [9] proposed an ES-dRNN (exponential smoothing and dilated recurrent neural network) hybrid model, which integrates exponential smoothing for extracting trend and seasonal components with the long-range dependency capturing capability of dilated RNNs. Nevertheless, its shallow architecture hinders the characterization of pronounced nonlinearity and multi-scale dynamics inherent in load profiles, and spatial coupling relationships are not taken into consideration. Reference [10] adopted a method combining feature extraction with an improved GRNN(general regression neural network), enhancing fitting capacity via optimization algorithms. However, as GRNN constitutes a shallow architecture, it similarly encounters difficulty in characterizing pronounced nonlinearity and multi-scale dynamic features. Reference [11] employed a method combining SSA (singular spectrum analysis) with a Kalman smoother for the task of short-term load prediction; however, the linear decomposition inherent in SSA is insufficient for capturing nonlinear dynamics. Reference [12] employed a random forest approach combined with permutation importance-based feature selection. Although this method is capable of reducing feature redundancy, it exhibits relatively weak capacity in modeling temporal dependencies.

Machine learning methods encounter difficulty in capturing the pronounced nonlinearity and multi-scale dynamic features of load profiles, while simultaneously neglecting inter-regional spatial coupling relationships. In contrast, deep learning, through its deep network architecture, achieves a technological leap from feature engineering-driven approaches to end-to-end spatio-temporal representation learning, thereby providing enhanced modeling capacity for load forecasting in complex power systems. Reference [13], employed a CNN (convolutional neural network) to extract local fluctuation features from load sequences; however, its restricted receptive field limits the capacity to model long-span temporal correlations. Reference [14] adopted a CEEMDAN (complete ensemble empirical mode decomposition with adaptive noise)–Transformer hybrid model, wherein decomposition is performed to mitigate sequence non-stationarity and the Transformer architecture is leveraged to capture long-range temporal dependencies. Nevertheless, CEEMDAN is subject to boundary effects and mode mixing issues, the Transformer incurs relatively high computational complexity, and the multi-stage pipeline is prone to error accumulation. Literature [15] employed a hybrid model combining VMD with GRU-TCN (gated recurrent unit and temporal convolutional network), wherein the load sequence is first decomposed via VMD, and the resulting components are subsequently fed into GRU-TCN for individual prediction and final reconstruction. However, the selection of key VMD parameters relies heavily on empirical expertise, rendering the decomposition susceptible to mode mixing or over-decomposition; moreover, inter-regional spatial coupling relationships are not taken into account. Reference [16] adopted a hybrid model integrating EMD (empirical mode decomposition) with GRU, wherein EMD decomposition is applied to mitigate non-stationarity and feature selection is incorporated to optimize the input representation. However, EMD is inherently prone to mode mixing and endpoint effects, resulting in limited decomposition accuracy; furthermore, inter-regional spatial coupling relationships are not taken into consideration. Reference [17] employed a TFT (temporal fusion transformer) model, incorporating meteorological factors for short-term load forecasting. Through its attention mechanisms and gating structures, TFT enables the effective integration of long-range temporal dependencies with exogenous variables. Nevertheless, this approach models each node independently, thereby neglecting inter-regional spatial coupling relationships and proving inadequate in capturing the influence of power network topology on load evolution dynamics.

However, the aforementioned methods predominantly model individual nodes independently or simply concatenate multivariate inputs, thereby overlooking the intrinsic spatio-temporal coupling characteristics inherent in load evolution—such as nodal interactions, grid topology influences, and meteorological propagation effects. Consequently, a notable degradation in forecasting accuracy is incurred under scenarios involving regional or cascading load fluctuations. To overcome the limitations confined to a single temporal dimension, spatio-temporal joint modeling based on graph neural networks has emerged as a burgeoning research focus. Reference [18] systematically investigated the potential of GCN and TCN in mining implicit multivariate relationships and extracting features across multiple time scales. In particular, the GCN model incorporating LDTW (local dynamic time warping) [19] and the STGCN [20] significantly enhance the capacity to capture spatial correlations through the explicit construction of spatio-temporal information graphs. Nevertheless, existing deep spatio-temporal models still confront significant challenges. On the one hand, the selection of key parameters in signal decomposition techniques such as VMD heavily relies on empirical expertise and lacks an adaptive mechanism, rendering the process susceptible to mode mixing or excessive decomposition. On the other hand, graph-based models including STGCN, when stacking deep layers to extract high-order features, are highly prone to the over-smoothing phenomenon—whereby node representations tend toward homogeneity and the discriminative capacity of deeper layers is progressively lost, thus constraining further improvements in predictive performance.

In summary, to address the dual challenges of pronounced nonlinearity and spatio-temporal coupling inherent in short-term load forecasting, this paper proposes the BO-VMD-STGCN model. Within this framework, Bayesian Optimization is employed to achieve adaptive selection of VMD parameters, thereby enhancing decomposition accuracy. Furthermore, a residual stacking mechanism is incorporated into the STGCN architecture to mitigate the over-smoothing issue in deep networks and to strengthen the capacity for spatio-temporal feature extraction. Experimental results demonstrate the superiority of the proposed model over prevailing alternatives with respect to prediction accuracy, resilience to noise, and cross-regional generalizability, thereby establishing a viable and efficient framework for the task of multi-node load forecasting.

II. BASIC MODEL THEORY

A. SPATIO-TEMPORAL GRAPH CONVOLUTIONAL NETWORK MODEL

To simultaneously characterize the evolutionary dynamics of power load sequences along the temporal dimension and the coupling dependencies across the spatial dimension, the STGCN is adopted as the core forecasting architecture in this study.

The power network topology is modeled as a graph $G = (V, E)$ [21]. where V denotes the set of nodes and E represents the electrical or geographical adjacency relationships between nodes. Let $S \in \mathbb{R}^{N\times T\times F}$ denote the input feature tensor, with N representing the total number of vertices in the graph, T the horizon of considered time steps, and F the size of the per-node feature space [22]. The fundamental building block of STGCN is the S-T-Conv Block, which can be formulated as:

$$S^{l+1} = \sigma\left(T_{\text{conv}}\left(G_{\text{conv}}\left(S^l\right) W^{(l)}\right)\right) \tag{1}$$

where S^l denotes the input feature tensor of the l-th spatio-temporal convolutional block; $T_{\text{conv}}(\cdot)$ and $G_{\text{conv}}(\cdot)$ represent the gated temporal convolution and graph convolution operators, respectively; $W^{(l)}$ denotes the trainable parameter matrix associated with layer l; $\sigma(\cdot)$ denotes the Sigmoid activation function.

1) GATED TEMPORAL CONVOLUTION

The GTCN module employs one-dimensional causal convolution [23] to preserve the temporal causality of the sequence. Given an input S^l, the module generates two parallel mappings, P and Q, and applies nonlinear modulation via a GLU (Gated Linear Unit), which can be expressed as:

$$T_{\text{conv}}\left(S^l\right) = P \odot \sigma(Q) \in \mathbb{R}^{N\times T'\times C'} \tag{2}$$

where P and Q are obtained by convolving S^l with a common causal kernel; $\sigma(\cdot)$ denotes the Sigmoid activation function; and $\odot$ represents element-wise multiplication.

2) GRAPH CONVOLUTIONAL NETWORK

Building upon the output of the temporal convolution, spatial information from neighboring nodes is aggregated via the graph Laplacian operator [24]. A normalized adjacency matrix with self-loops is adopted, expressed as:

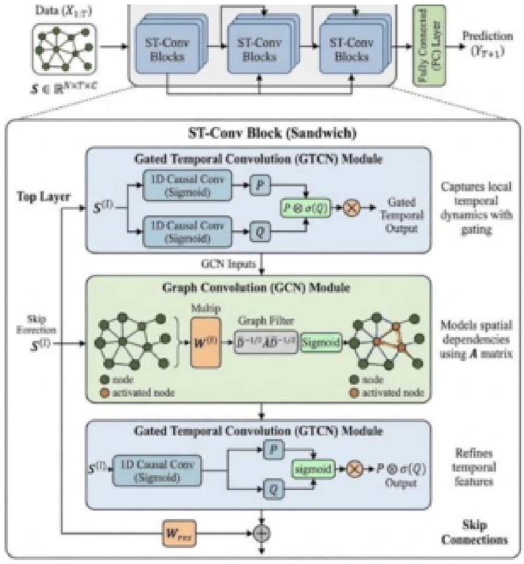


FIGURE 1. Network structure diagram of spatio-temporal graph convolutional network

$$G_{\text{conv}}(V) = \sigma\left(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}VW^{(l)}\right) \tag{3}$$

where $\widetilde{A} = A + I_N$ represents the self-augmented adjacency matrix, $\widetilde{D}$ denotes the associated degree matrix, and $W^{(l)} \in \mathbb{R}^{C'\times C''}$ is the trainable weight matrix of layer l. thin each ST-Conv Block, skip connections are introduced between the two GTCN modules to mitigate the vanishing gradient problem in deep networks and to facilitate feature reuse. By stacking multiple ST-Conv Blocks, the model is capable of abstracting spatio-temporal features in a hierarchical manner, ultimately yielding the predicted load values for the subsequent H time steps through a fully connected layer. The aforementioned architecture is illustrated in Figure 1.

B. VARIATIONAL MODE DECOMPOSITION

VMD, introduced by Konstantin Dragomiretskiy and Dmitry Zosso [25], is a signal decomposition method grounded in a variational optimization framework. Built upon a rigorous variational optimization paradigm, VMD adaptively separates the raw load series $f(t)$ into K band-limited intrinsic oscillatory modes (IMFs), thereby effectively separating the trend, periodic components, and stochastic noise embedded within the load sequence. This process yields input features with an enhanced signal-to-noise ratio for the subsequent STGCN. The overall computational workflow of VMD is illustrated in Figure 2.

Assuming that each mode $u_k(t)$ is compactly distributed around its central frequency ω_k,VMD seeks to reduce the aggregate spectral width of the extracted modes by addressing a constrained variational formulation, subject to the requirement that the summation of all decomposed

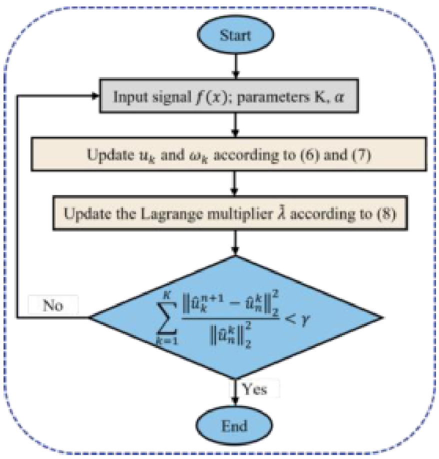


FIGURE 2. VMD calculation process

constituents exactly recovers the original load sequence $f(t)$:

$$\begin{aligned} &\min_{\{u_k\},\{\omega_k\}} \quad \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ &\text{s.t.} \quad \sum_{k=1}^{K} u_k(t) = f(t) \end{aligned} \tag{4}$$

where $\{u_k\} = \{u_1, u_2, \ldots, u_K\}$ denotes the set of K modal components obtained after decomposition; $\{\omega_k\} = \{\omega_1, \omega_2, \ldots, \omega_K\}$ represents the central frequencies corresponding to u_k; ∂_t is the operator for partial temporal derivatives; $\delta(t)$ stands for the Dirac impulse distribution; and the symbol $*$ denotes convolution.

By introducing a quadratic penalty term α [26] to enhance robustness in the presence of noisy load data, and a Lagrange multiplier $\lambda(t)$ introduced to guarantee exact fulfillment of the equality condition, the augmented Lagrangian reads:

$$\begin{aligned} L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_{k=1}^{K} &\left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] \right. \\ &\left. \times e^{-j\omega_k t} \right\|_2^2 \\ &+ \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 \\ &+ \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle. \end{aligned} \tag{5}$$

The saddle point of the augmented Lagrangian is obtained by iteratively updating $\hat{u}_k^{n+1}$, ω_k^{n+1}, and λ via the ADMM (alternating direction method of multipliers). The iterative procedure is formulated as follows:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{6}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \tag{7}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right) \tag{8}$$

where $\hat{u}_k(\omega)$, $\hat{\lambda}^n(\omega)$, and $\hat{f}(\omega)$ denote the Fourier transforms of $u_k(t)$, $\lambda(t)$, and $f(t)$, respectively; n represents the iteration index; and τ is the iteration step size. This iterative updating proceeds until either the prescribed convergence tolerance is met or the pre-defined iteration limit is attained:

$$\sum_{k=1}^{K} \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \gamma \tag{9}$$

C. BAYESIAN OPTIMIZATION

Short-term load forecasting data in emerging power systems are characterized by marked nonlinearity, non-stationarity, and multi-scale coupling characteristics. The mode count K and the regularization coefficient α within the VMD framework, together with the multi-dimensional architectural hyperparameters of STGCN, collectively constitute a high-dimensional, black-box combinatorial optimization problem. In this paper, BO is introduced as the core adaptive optimization engine.

First, the hyperparameter combination is defined as a decision vector $x \in \mathcal{X}$, and the objective function $f(x)$ is taken as the validation error of the model, and the optimization aims to locate the hyperparameter set x^* yielding the lowest $f(x)$ subject to a total evaluation limit of N trials.

$$x^* = \arg\min_{x \in \mathcal{X}} f(x) \tag{10}$$

where $f(x)$ denotes a non-differentiable black-box function. BO approximates $f(x)$ by constructing a probabilistic surrogate model and employs an acquisition function to guide the sampling process. Given the stochastic fluctuations inherent in power load data, a GP(gaussian process) [27] is utilized as the surrogate model in this paper. At each step t, the subsequent evaluation candidate $\alpha(x)$ is chosen as the maximizer of the acquisition function x_{t+1}.

$$x_{t+1} = \arg\max_{x \in \mathcal{X}} a(x; \mathcal{D}_{1:t}) \tag{11}$$

Considering that the hyperparameter space may contain multiple local optima, the EI (expected improvement) [28] is chosen as the acquisition function in this paper. EI can analytically compute the expected performance gain from a new sample point, given the current best observation $f(x^+)$ and the posterior distribution.

$$\alpha_{\mathrm{EI}}(x) = \mathbb{E}\left[\max\left(0, f(x^+) - f(x)\right)\right] \tag{12}$$

The BO algorithm identifies near-optimal parameters for VMD and STGCN within few iterations. It overcomes tuning difficulties arising from noisy load data and costly model training. Consequently, both predictive accuracy and model robustness are markedly improved.

III. CONSTRUCTION OF THE BO-VMD-STGCN FORECASTING MODEL

To effectively extract the complex spatio-temporal correlations and multi-scale fluctuation characteristics embedded in power load data, the BO-VMD-STGCN model is constructed in this paper. Within the proposed pipeline, the initial step involves a VMD-based decomposition of the raw load data into multiple intrinsic modes, thereby achieving the decoupling of multi-scale fluctuations. Subsequently, BO is introduced to adaptively tune the key parameters of VMD, ensuring optimal decomposition quality. Finally, the decomposed modes, together with temporal features, are jointly fed into the STGCN, where a stacked architecture comprising graph convolutions and gated temporal convolutions is employed to capture spatial dependencies and temporal dynamics. Additionally, a trend decomposition and residual prediction strategy is incorporated to further enhance forecasting accuracy. The detailed derivation process is presented as follows.

Step 1: Data Acquisition. The experimental data utilized here comprises power load data from eight regions of Bangladesh, spanning the period from November 21, 2019 to December 30, 2024, with a daily sampling frequency. A total of 1,850 load data points were acquired. The raw data constitute a multi-region load matrix $S_{\text{raw}} \in \mathbb{R}^{N\times T}$, where $N = 8$ denotes the number of regional nodes and $T = 1850$ represents the temporal step length.

Step 2: Data Preprocessing. To address potential missing values or abnormal interruptions within the original load sequences, a combined approach of bidirectional interpolation and boundary padding is employed for data restoration, thereby ensuring the completeness and continuity of the sequences. Specifically, linear interpolation is performed for internal missing entries, while forward filling and backward filling are applied to consecutive missing values at the beginning and end of the sequence, respectively. Following preprocessing, a clean and continuous multi-regional load time series is obtained, thereby establishing a solid data foundation for subsequent decomposition and modeling.

Step 3: Data Decomposition and Feature Enhancement. First, BO is introduced to adaptively search for the optimal parameter pair K, α of VMD. With the objective of minimizing spectral aliasing and reconstruction error, a Gaussian Process surrogate model is constructed to obtain the globally optimal solution within a limited number of iterations. By applying the optimized parameters, VMD is performed independently on each region, yielding a set of modal components $IMF_1, \ldots, IMF_K$. These modes are subsequently concatenated with temporal encodings to form an enhanced feature tensor $S_{\text{enh}} \in \mathbb{R}^{N\times T\times(C_{\text{IMF}}+C_{\text{time}})}$, which simultaneously preserves the multi-scale time-frequency characteristics and injects temporal positional information. To enable the deep model to focus on fitting high-frequency fluctuations and local details, thereby preventing trend components from interfering with spatio-temporal modeling. Moving Average Trend Extraction. The MA(moving average) is employed as a non-learnable branch to extract the low-band trend component of the original sequence, denoted as $S_{\text{trend}} \in \mathbb{R}^{N\times T}$. The residual component is subsequently defined as $S_{\text{res}} = S_{\text{raw}} - S_{\text{trend}}$, where S_{raw} denotes the original multi-regional load matrix.

Step 4: Construction of the Short-Term Load Prediction Framework. The enhanced features S_{enh} and the residual component S_{res} are fused to form an augmented spatio-temporal input $S_{\text{in}} \in \mathbb{R}^{N\times T\times D}$, which is subsequently fed into the enhanced STGCN network. Building upon the standard STGCN architecture, this network integrates a spatio-temporal attention mechanism, wherein spatial attention dynamically adjusts the weights for information aggregation among nodes, and temporal attention highlights the critical time steps within the historical sequence. Through this architecture, the model simultaneously captures temporal evolution patterns and spatial topological dependencies, ultimately yielding the residual prediction values for the subsequent H steps, denoted as $\hat{S}_{\text{res}} \in \mathbb{R}^{N\times H}$.

Step 5: Model Reconstruction and Fusion. The final load forecast is structurally reconstructed through the linear summation of the trend component and the residual prediction values:

$$\hat{S}_{\text{final}} = S_{\text{trend}} + \hat{S}_{\text{res}} \in \mathbb{R}^{N\times H} \tag{13}$$

where S_{trend} denotes the trend component extracted via moving average, and the trend values for future time steps are extrapolated using the last available trend value. This reconstruction fusion organically integrates the stable trend extracted by the non-learnable branch with the fluctuating residual learned by the deep network, thereby ensuring trend consistency in the forecast results while simultaneously enhancing the fitting accuracy of local fluctuations.

Step 6: Multi-Metric Performance Evaluation. The MAE (mean absolute error), RMSE (root mean square error), MAPE (mean absolute percentage error), and R2 (coefficient of determination) are adopted as evaluation metrics to quantitatively assess the forecasting performance of each model. By benchmarking against several reference models (e.g., LSTM, STGCN, STM-GCN, and GCN-LSTM), the superiority of the proposed BO-VMD-STGCN model is validated in terms of forecasting accuracy, stability, and generalization capability.

Figure 3 provides a visual summary of the entire BO-VMD-STGCN forecasting pipeline, which cohesively combines all the previously described stages. By leveraging the underlying regularities present in electricity consumption data, this unified framework delivers a forecasting approach that is both computationally efficient and readily interpretable for next-generation power systems.

IV. RESULTS ANALYSIS AND DISCUSSION

A. DATASET CONSTRUCTION AND PREPROCESSING

1) DATASET DESCRIPTION

The dataset employed in this study comprises power load data collected from eight regions of Bangladesh, covering

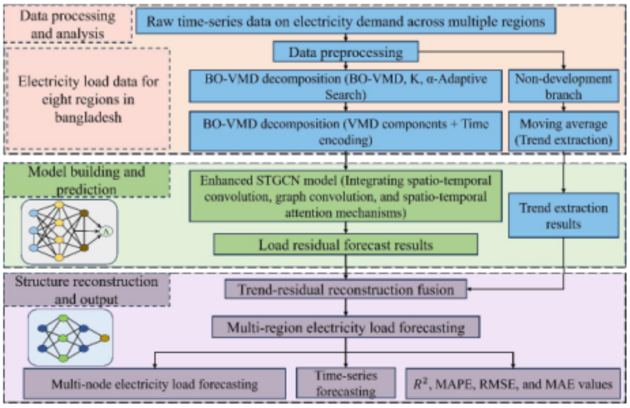


FIGURE 3. Short-term power load forecasting procedure based on BO-VMD-STGCN

FIGURE 4. Spatial relationships among eight regions (https://baike.baidu.com/item/%E5%AD%9F%E5%8A%A0%E6%8B%89%E5%9B%BD%E8%A1%8C%E6%94%BF%E5%8C%BA%E5%88%92/12585436)

the period from November 21, 2019, to December 30, 2024, with a daily sampling frequency. A total of 1,850 load data points were obtained. The spatial relationships among the eight regions are constructed and visualized based on their longitude and latitude coordinates, see Figure 4. The initial 1,295 load data points are allocated as training samples, while the subsequent 555 points are designated as testing samples, which are subsequently compared against the actual load values. The original load data are illustrated in Figure 5.

2) DATA PREPROCESSING

To address potential missing values or abnormal interruptions arising during the acquisition of power load data, a combined approach of bidirectional interpolation and boundary padding is employed for data restoration in this study [29]. Let the original sequence be denoted as $\{x_t\}_{t=1}^{T}$, and let $\mathcal{M}$ represent the index set of missing positions. Bidirectional linear interpolation is first performed as follows: for each $t \in \mathcal{M}$, the nearest known points x_{t_0} and x_{t_1} preceding and succeeding t are identified:

$$x_t = x_{t_0} + \frac{x_{t_1} - x_{t_0}}{t_1 - t_0}(t - t_0) \tag{14}$$

For consecutive missing values occurring at the beginning or the end of the sequence, forward filling and backward filling are applied, respectively [30]. This strategy ensures the completeness and continuity of the sequence, thereby providing a reliable data foundation for subsequent modeling.

To eliminate discrepancies stemming from varying feature scales, the load observations are normalized using the Standard Scaler technique [31]. This transformation ensures the processed data conform to a zero-mean, unit-variance distribution. The corresponding operation is given by:

$$x' = \frac{x - \mu}{\sigma} \tag{15}$$

where x corresponds to the raw load observation, μ and σ are the mean and standard deviation computed from the training partition, respectively, and x' stands for the resulting scaled value.

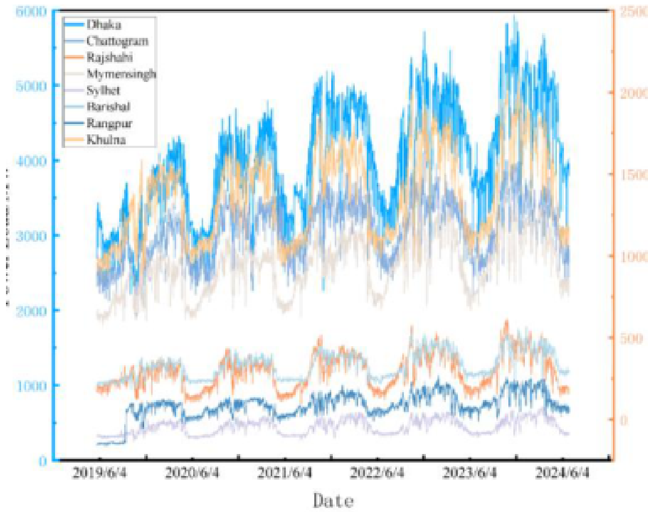


FIGURE 5. Raw load data

3) BO-VMD PARAMETER OPTIMIZATION AND DECONVOLUTION RESULTS

The load data are decomposed via the BO-VMD pipeline, an operation that improves both the stationarity of the derived representation and the predictive fidelity of the subsequent model. Bayesian Optimization drives an automated, error-informed search for the most suitable VMD hyperparameters, thereby ensuring that the resulting modal partition is well-suited to the forecasting objective. The specific configurations adopted for BO and VMD are summarized in Table 1 and Table 2, respectively.

TABLE 1. BO PARAMETER SETTINGS

α	K	Maximum number of iterations	Number of training rounds	Random seed
$[100, 3000]$	$[2, 8]$	12	12	42

TABLE 2. VMD PARAMETER SETTINGS

Number of factors K	Penalty factor α	Optimize the number of variables	Convergence criterion r
4	343	2	$1e^{-7}$

The BO-VMD optimization procedure ultimately yielded an optimal decomposition number of $K = 4$ and an optimal penalty parameter of $\alpha = 344$.

Following the adaptive decomposition via BO-VMD, the original load sequence is decomposed into four IMFs, designated as IMF1 through IMF4, along with a residual component, as illustrated in Figure 6. Specifically, the high-frequency component IMF1 exhibits pronounced short-period oscillatory behavior, characterized by high frequency and low amplitude, primarily representing the stochastic perturbations and measurement noise embedded within the load sequence. The intermediate-frequency components, IMF2 and IMF3, exhibit stable periodic fluctuations, reflecting the cyclical variation patterns of the load at both intra-day and multi-day scales. The low-frequency component IMF4 exhibits gradual variations with pronounced trend characteristics, thereby delineating the long-term evolution behavior of the load. The residual component exhibits relatively small overall amplitude and lacks discernible structural features, indicating that the vast majority of useful information has been adequately extracted by the modal components. Benefiting from the adaptive optimization of the decomposition parameters K, α via BO, the degree of spectral aliasing among the modal components is substantially mitigated, endowing the decomposition results with enhanced physical interpretability and an improved signal-to-noise ratio. This decomposition step substantially reduces the non-stationarity and intricacy present in the raw load series, yielding input representations that are structurally cleaner and more stationary for the downstream enhanced STGCN architecture. This, in turn, contributes to improving the accuracy and generalization capability of multi-region short-term load forecasting.

Figure 7 presents a comparative illustration of the original load sequences and the BO-VMD reconstructed sequences across multiple regions during the period from December 2019 to June 2024. The reconstructed curves for each region closely match the actual observations, indicating that the BO-VMD based signal decomposition and reconstruction process possesses robust information preservation capability. This enables a low-distortion reconstruction of the original sequence without compromising critical trend and fluctuation characteristics. However, at certain peak load points, the reconstructed curves exhibit a slight smoothing effect. This phenomenon is primarily attributable to the inherent suppression effect of VMD on high-frequency noise and

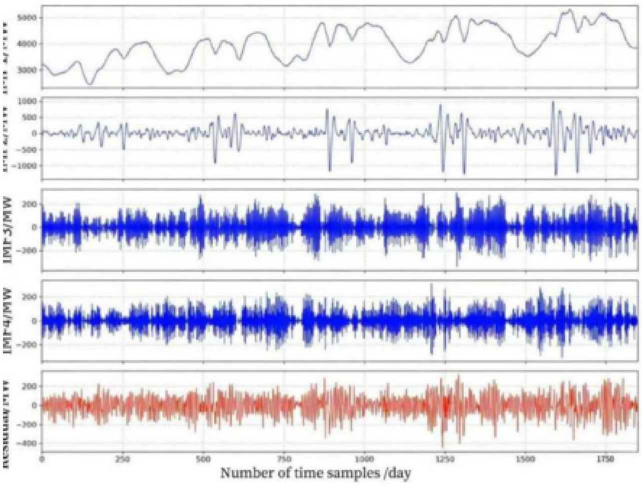


FIGURE 6. Modal components and residuals

stochastic perturbations during the decomposition process, which results in a certain degree of filtering of local extrema. Nonetheless, by adaptively selecting the decomposition parameters K, α through Bayesian Optimization, the BO-VMD approach achieves a reasonable balance between noise suppression and information preservation, effectively mitigating sequence non-stationarity while enhancing the signal-to-noise ratio. Therefore, BO-VMD not only provides input data with a clearer structure and enhanced stationarity for subsequent spatio-temporal modeling, but also preserves the critical dynamic characteristics of the load, thereby establishing a solid data foundation for high-precision forecasting with the enhanced STGCN model.

B. MODEL EVALUATION METRICS

To facilitate comparative analysis, four evaluation metrics are adopted in this experiment, namely MAE, MAPE, RMSE, and R2. The computational formulas for these metrics are defined as follows:

$$MAE = (1/m) \times \sum_{i=1}^{m} |y_i - \hat{y}_i| \tag{16}$$

$$MAPE = (1/m) \times \sum_{i=1}^{m} |(y_i - \hat{y}_i)/y_i| \times 100\% \tag{17}$$

$$RMSE = \sqrt{(1/m) \times \sum_{i=1}^{m} (y_i - \hat{y}_i)^2} \tag{18}$$

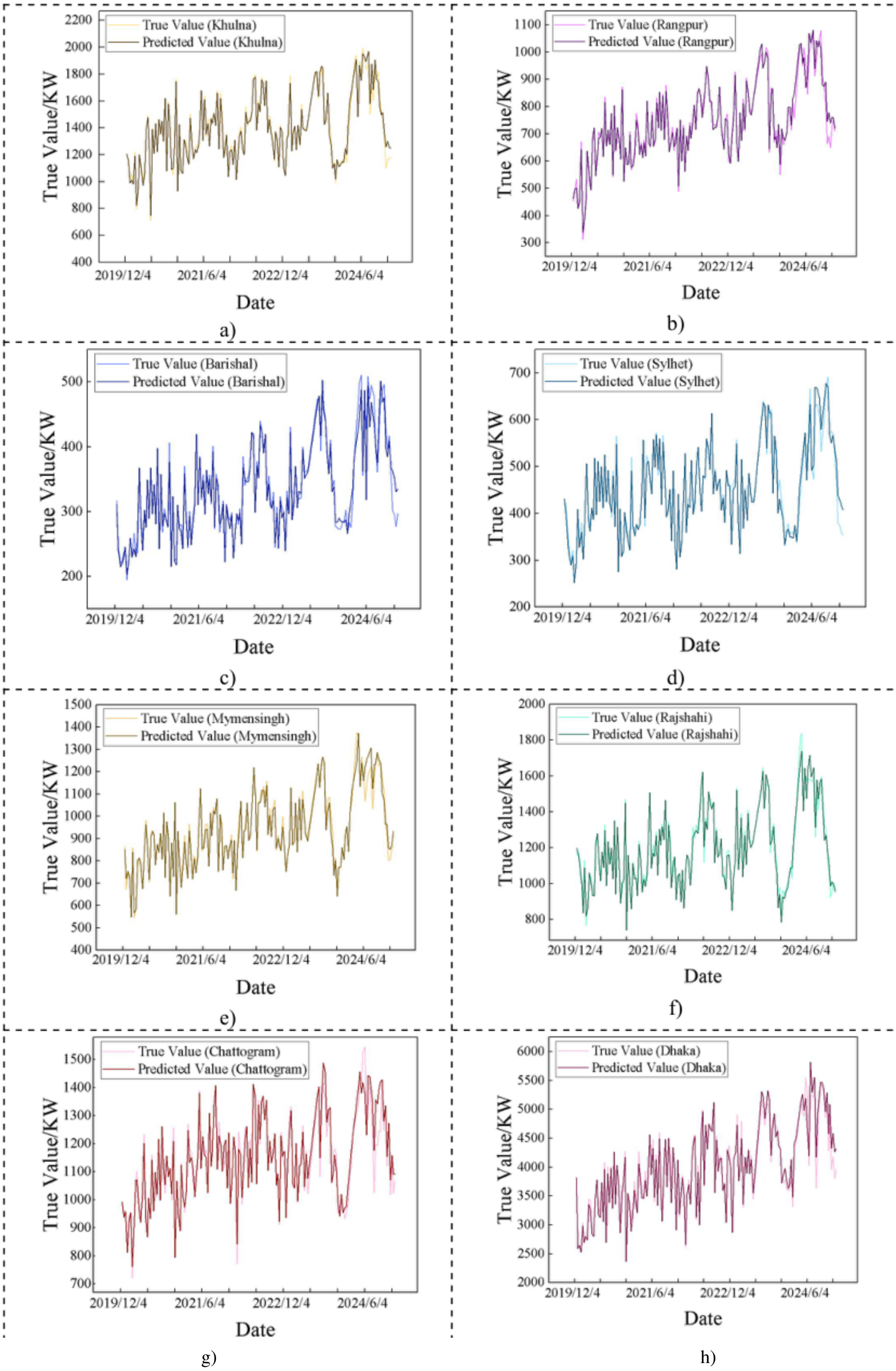


FIGURE 7. Contrasting Reconstructed and Actual Load Series. a)Region: Khulan; b)Region: Rangpur; c)Region: Barishal; d)Region: Sylhet; e)Region: Mymensingh; f)Region: Rajshahi; g)Region: Chattogram; h)Region: Dhaka

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y}_i)^2} \tag{19}$$

where y_i denotes the observed load value, $\hat{y}_i$ is the corresponding forecasted value produced by the model, and m

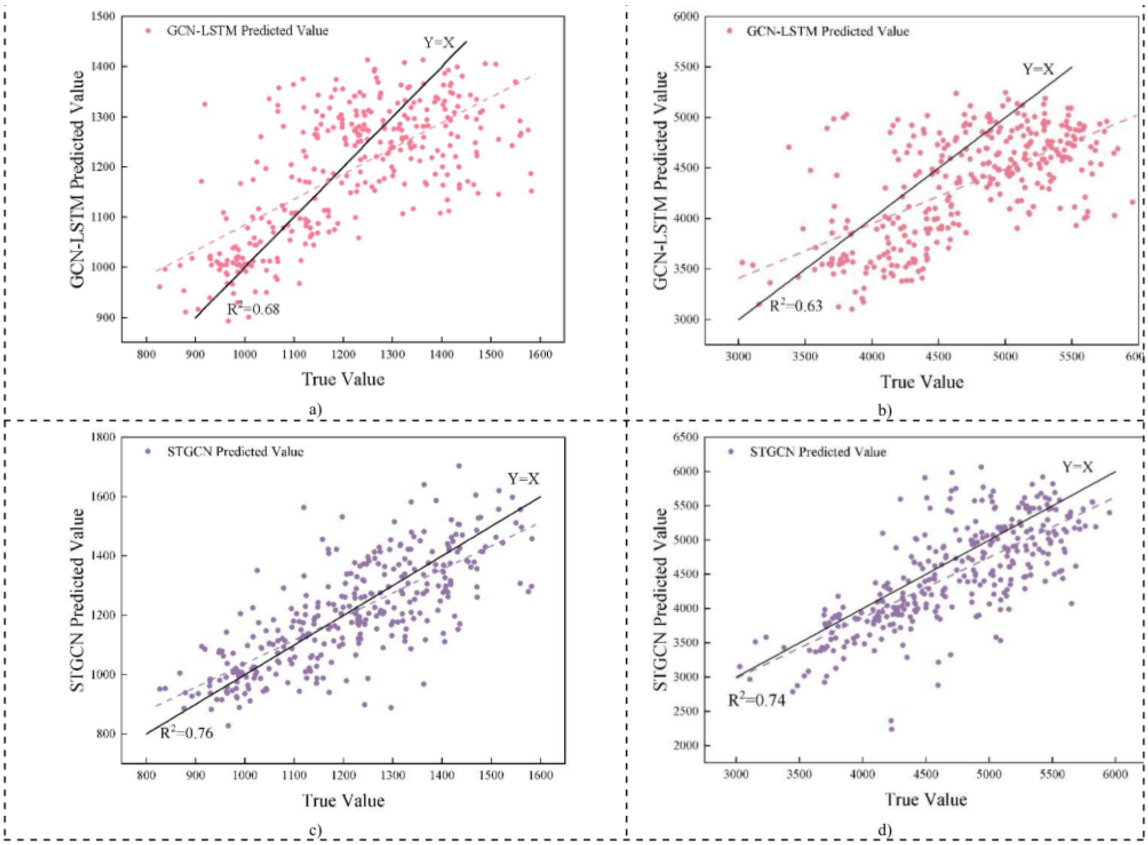
indicates the size of the test set.

C. EVALUATION AND INTERPRETATION OF EXPERIMENTAL FINDINGS

This section assesses the performance of the BO-VMD-STGCN model through the four designated evaluation metrics, with its results contrasted against those of four competing baseline architectures.

1) FIT ANALYSIS

To demonstrate the superiority of the BO-VMD-STGCN model, the established models in the field of power load forecasting—namely, GCN-LSTM [32], STM-GCN [33], STGCN [34], and LSTM [35] are adopted as comparative benchmarks, and their fitting curves are analyzed accordingly. The degree of fit is illustrated in Figure 8.



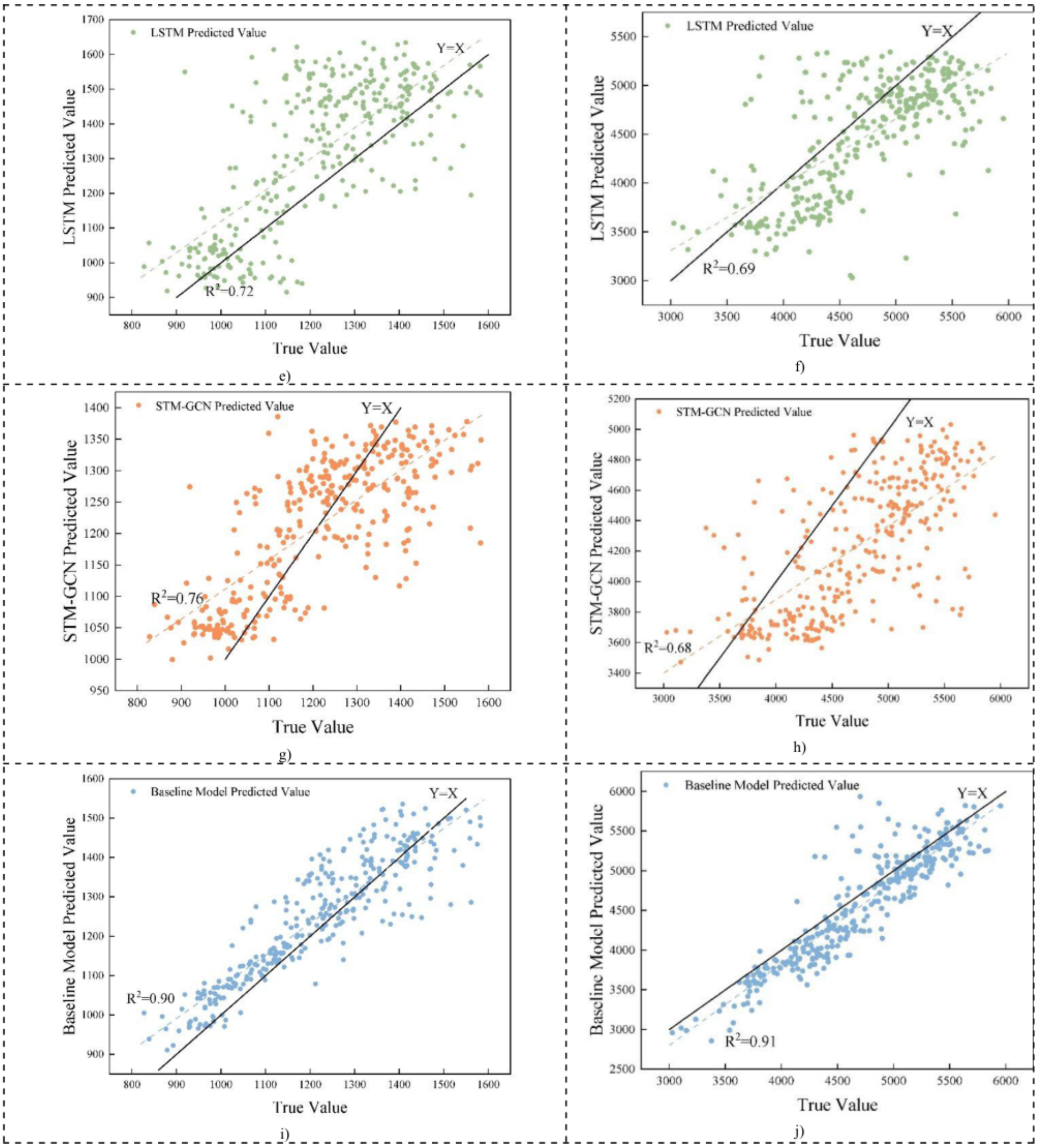


FIGURE 8. Comparison of the fit of different models. a)Model: GCN-LSTM, Region: Chattogram; b)Model: GCN-LSTM, Region: Dhaka; c)Model: STGCN, Region: Chattogram; d)Model: STGCN, Region: Dhaka; e)Model: LSTM, Region: Chattogram; f)Model: LSTM, Region: Dhaka; g)Model: STM-GCN, Region: Chattogram; h)Model: STM-GCN, Region: Dhaka; i)Model: BO-VMD-STGCN, Region: Chattogram; j)Model: BO-VMD-STGCN, Region: Dhaka

In short-term power load forecasting, the key challenge lies in simultaneously capturing the temporal dynamic evolution of the sequences and the spatial coupling dependencies among nodes, while effectively addressing the non-stationarity and high-frequency noise inherent in the raw data. Figure 8 presents a comparative overview of the R2 values achieved by the five forecasting models across different regions. Among them, the standalone LSTM model yields the lowest R2 value due to its lack of explicit modeling of spatial dependencies, indicating that this class of recurrent neural networks struggles to adequately capture the influence of power network topology on load dynamics. The standard STGCN model, through its joint architecture of graph convolution and gated temporal convolution, enhances the fitting capability to a certain extent; however, its capacity to handle the complex fluctuations and noise components embedded in the original load sequence remains insufficient. In contrast, the STGCN model augmented with BO-VMD for adaptive decomposition preprocessing significantly enhances the goodness-of-fit, with its R2 stably exceeding 0.92. This indicates that the decomposed and reconstructed input features align more closely with the spatial correlation patterns assumed by the graph convolutional architecture. Compared with the LSTM, STGCN, STM-GCN, and GCN-LSTM models, the proposed model achieves improvements in goodness-of-fit of 26.39%, 19.73%, 18.42%, and 35.29%, respectively. The above results substantiate the pronounced advantages of BO-VMD-STGCN in addressing the challenges of data non-stationarity, noise interference, and spatio-temporal dependency coupling inherent in short-term load forecasting.

2) ANALYSIS OF PREDICTION ERRORS

To further assess the forecasting capability of the proposed approach, a side-by-side comparison is carried out based on three error measures—MAPE, MAE, and RMSE—with the outcomes summarized in Table 3.

TABLE 3. COMPARISON OF PREDICTION RESULTS ACROSS MODELS

Predictive model	MAE	RMSE	MAPE
GCN-LSTM	129.9648	228.0	9.18%
STM-GCN	121.6	154.3	8.70%
LSTM	88.26	149.71	6.84%
STGCN	47.37	77.15	3.72%
BO-VMD-STGCN	38.36	62.68	2.84%

Table 3 indicates that the proposed BO-VMD-STGCN framework attains the best results across all three evaluation criteria, yielding MAE, RMSE, and MAPE scores of 38.36, 62.68, and 2.84%, respectively—markedly outperforming the competing approaches. Relative to the standalone STGCN, these figures correspond to declines of 19.0% in MAE, 18.8% in RMSE, and 23.7% in MAPE. This result convincingly demonstrates that the integration of the BO-VMD enhancement strategy effectively elevates the modeling capacity of STGCN with respect to the complex spatio-temporal characteristics of power load. The prediction error of the conventional time series model LSTM is notably higher than that of STGCN and BO-VMD-STGCN, both of which incorporate spatio-temporal information, indicating that the omission of spatial correlations leads to a degradation in forecasting accuracy. Although GCN-LSTM and STM-GCN integrate graph convolution with recurrent networks, their prediction errors remain higher than those of the STGCN family of models, suggesting that their capacity to capture spatio-temporal dependencies is limited.

To facilitate a clearer visual comparison of model performance across the MAE, RMSE, and MAPE metrics, the quantitative data listed in Table 3 are depicted as a bar chart in Figure 9. Overall, the evaluation metrics MAPE, MAE, and RMSE exhibit a consistent trend of variation across the different models. LSTM models only temporal dependencies and consequently yields relatively high prediction errors. The GCN-LSTM and STM-GCN models, which incorporate spatial information, enhance the forecasting performance to a certain extent; however, constrained by the effectiveness of their spatio-temporal fusion mechanisms, the degree of improvement remains relatively limited. The STGCN model is capable of simultaneously characterizing spatio-temporal correlations, resulting in a notable improvement in forecasting accuracy. The proposed BO-VMD-STGCN model exhibits the best performance, corresponding to the lowest values in the bar chart across all three evaluation metrics. This indicates that the model outperforms the other comparative methods in terms of both error control capability and forecasting stability.

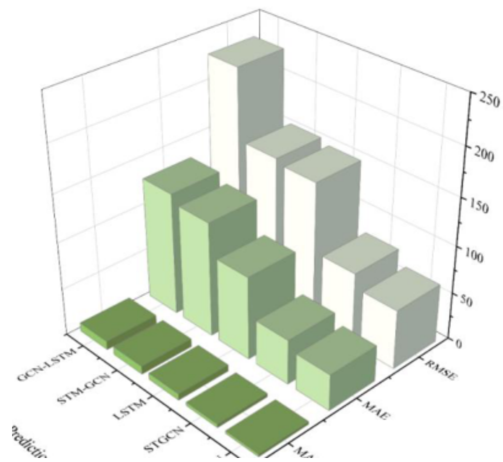


FIGURE 9. Bar chart comparing the prediction results of various models

3) TREND FORECAST ANALYSIS

To more intuitively compare the evolutionary trends of the proposed model, the baseline models, and the actual load values over a continuous time axis, Figure 10 is plotted with the trend colors of the proposed model and the actual values highlighted. It can be observed from the figure that the prediction curve of BO-VMD-STGCN almost coincides with the actual value curve, accurately tracking the trend of actual load variations even in regions characterized by substantial fluctuations. In contrast, the predictions of the other comparative models generally deviate from the actual curve, with more pronounced discrepancies observed particularly at load peaks and valleys.

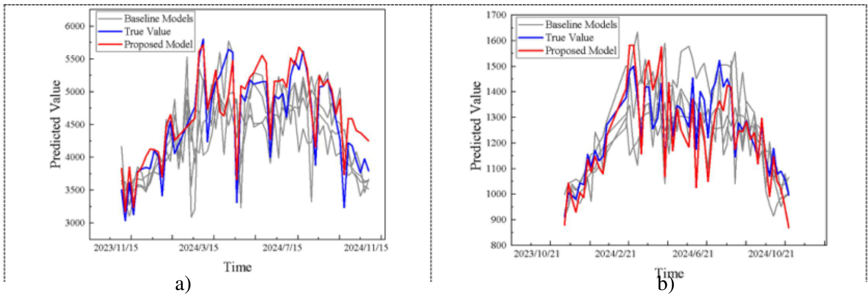


FIGURE 10. Prediction results of different methods on the dataset. a) Region: Dhaka; b) Region: Rajshahi.

In summary, the BO-VMD-STGCN model demonstrates superior performance in multi-region short-term load forecasting. Its coefficient of determination R2 exceeds 0.92, representing improvements of 26.39%, 19.73%, 18.42%, and 35.29% over LSTM, STGCN, STM-GCN, and GCN-LSTM, respectively. The MAE, RMSE, and MAPE are 38.36, 62.68, and 2.84%, respectively, which correspond to reductions of 19.0%, 18.8%, and 23.7% relative to the STGCN model. The predicted curve closely mirrors the observed data, faithfully capturing the underlying fluctuations, most notably around the load extrema. Through its chained fusion framework of "decomposition-separation-prediction-reconstruction," the model effectively enhances forecasting accuracy, robustness, and generalization capability, thus offering a practical and transparent approach to short-term load prediction in modern power grids.

V. CONCLUSION

Focusing on the marked nonlinearity, intertwined multi-scale fluctuations, and spatial–temporal interdependencies embedded in short-term power load data, this study arrives at the following findings:

(1) This study develops a BO-VMD-STGCN framework for short-term electricity load prediction. First, BO-VMD is employed to perform adaptive decomposition on the original multi-region load sequences, thereby obtaining the IMFs which are subsequently concatenated with temporal encodings to mitigate data non-stationarity. Second, the moving average is introduced as a non-learnable branch to extract the low-frequency trend component, while the residual component is designated as the prediction target, thereby decoupling the trend term. Subsequently, the decomposed modal features are fused with the residual component and fed into the STGCN to enable the modeling of spatio-temporal dependencies within the residual sequence. Finally, structural reconstruction is accomplished by summing the trend component and the residual predictions, thereby achieving the accurate capture of spatio-temporal coupling characteristics.

(2) Based on the actual load data collected from eight regions of Bangladesh, the proposed BO-VMD-STGCN model is benchmarked against four baseline models, namely GCN-LSTM, STM-GCN, LSTM, and STGCN, using MAE, RMSE, and MAPE as evaluation metrics. Experimental results indicate that the proposed model attains the best scores across all evaluation criteria, recording values of 38.36, 62.68, and 2.84%, respectively. Compared with the standard STGCN model, the MAPE is reduced by 23.7%, while the MAE and RMSE are reduced by 19.0% and 18.8%, respectively. The proposed model attains relatively high coefficients of determination across all regions, and the forecast curve accurately tracks the peaks and valleys, thereby effectively mitigating time lags and deviations while exhibiting favorable stability and generalization capability. The proposed approach successfully disentangles the multi-scale dynamics embedded in the load series while making comprehensive use of the spatial–temporal dependencies across different regions. It surpasses current state-of-the-art techniques across the dimensions of prediction precision, stability, and generalizability, thus presenting an effective and practical alternative for short-term power load forecasting.

While the BO-VMD-STGCN framework yields strong predictive results in the context of multi-regional short-term load forecasting, several avenues for future enhancement remain open. The current formulation relies exclusively on historical load observations and does not yet account for auxiliary inputs such as weather patterns or calendar-related effects. Future work may integrate multi-source heterogeneous data to enhance forecasting accuracy under complex scenarios. Moreover, the computational cost of the model in large-scale systems is relatively high; therefore, subsequent efforts could focus on simplifying the network architecture or devising more efficient spatio-temporal feature extraction mechanisms to improve scalability and real-time applicability.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

DATA SHARING AGREEMENT

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

[1] M. Tan, S. Yuan, S. Li, et al., "Ultra-short-term industrial power demand forecasting using LSTM based hybrid ensemble learning," IEEE transactions on power systems, vol. 35, no. 4, pp. 2937–2948, 2019.

[2] S. Li, W. Kong, Y. Wang, et al., "Medium and long-term energy demand forecasts by sectors in China under the goal of "carbon peaking & carbon neutrality": Based on the LEAP-China model," Energy, vol. 310, Art. no. 133017, 2024.

[3] M. Tan, C. Hu, J. Chen, et al., "Multi-node load forecasting based on multi-task learning with modal feature extraction," Engineering applications of artificial intelligence, vol. 112, Art. no. 104856, 2022.

[4] Q. Dong, R. Huang, C. Cui, et al., "Short-Term Electricity-Load Forecasting by deep learning: A comprehensive survey," Engineering Applications of Artificial Intelligence, vol. 154, Art. no. 110980, 2025.

[5] J. W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," Journal of the Operational Research Society, vol. 54, no. 8, pp. 799–805, 2003.

[6] G. Dudek, "Pattern-based local linear regression models for short-term load forecasting," Electric power systems research, vol. 130, pp. 139–147, 2016.

[7] S. S. Pappas, L. Ekonomou, D. C. Karamousantas, et al., "Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models," Energy, vol. 33, no. 9, pp. 1353–1360, 2008.

[8] H. Al-Hamadi and S. Soliman, "Short-term electric load forecasting based on Kalman filtering algorithm with moving window weather and load model," Electric power systems research, vol. 68, no. 1, pp. 47–59, 2004.

[9] S. Smyl, G. Dudek, and P. Pełka, "ES-dRNN: A hybrid exponential smoothing and dilated recurrent neural network model for short-term load forecasting," IEEE transactions on neural networks and learning systems, vol. 35, no. 8, pp. 11346–11358, 2023.

[10] Y. Liang, D. Niu, and W.-C. Hong, "Short term load forecasting based on feature extraction and improved general regression neural network model," Energy, vol. 166, pp. 653–663, 2019.

[11] K. Alhashemi and O. T. Altinoz, "Power Demand Forecasting in Iraq Using Singular Spectrum Analysis and Kalman Filter-Smoother," IEEE Access, 2025.

[12] N. Huang, G. Lu, and D. Xu, "A permutation importance-based feature selection method for short-term electricity load forecasting using random forest," Energies, vol. 9, no. 10, Art. no. 767, 2016.

[13] M. Krichen, "Convolutional neural networks: A survey," Computers, vol. 12, no. 8, Art. no. 151, 2023.

[14] P. Ran, K. Dong, X. Liu, et al., "Short-term load forecasting based on CEEMDAN and Transformer," Electric Power Systems Research, vol. 214, Art. no. 108885, 2023.

[15] C. Cai, Y. Li, Z. Su, et al., "Short-term electrical load forecasting based on VMD and GRU-TCN hybrid network," Applied Sciences, vol. 12, no. 13, Art. no. 6647, 2022.

[16] X. Gao, X. Li, B. Zhao, et al., "Short-term electricity load forecasting model based on EMD-GRU with feature selection," Energies, vol. 12, no. 6, Art. no. 1140, 2019.

[17] P. C. Huy, N. Q. Minh, N. D. Tien, et al., "Short-term electricity load forecasting based on temporal fusion transformer model," Ieee Access, vol. 10, pp. 106296–106304, 2022.

[18] M. Liu, H. Qin, R. Cao, et al., "Short-term load forecasting based on improved TCN and DenseNet," IEEE Access, vol. 10, pp. 115945–115957, 2022.

[19] C. Zuo, J. Wang, M. Liu, et al., "An ensemble framework for short-term load forecasting based on TimesNet and TCN," Energies, vol. 16, no. 14, Art. no. 5330, 2023.

[20] M. Wei, M. Wen, and Y. Zhang, "A novel spatial electric load forecasting method based on LDTW and GCN," IET Generation, Transmission & Distribution, vol. 18, no. 3, pp. 491–505, 2024.

[21] Z. Tan, C. Ge, and Z. Pan, "Short-term spatial load forecasting method based on spatio-temporal graph convolutional network," in 2024 4th International Conference on New Energy and Power Engineering (ICNEPE), 2024:286–291.

[22] Z. Li, F. Liu, W. Yang, et al., "A survey of convolutional neural networks: analysis, applications, and prospects," IEEE transactions on neural networks and learning systems, vol. 33, no. 12, pp. 6999–7019, 2021.

[23] B. Wang, Y. Tang, Y. Huang, et al., "Power system emergency control strategy based on severely disturbed units identification and STGCN-DDQN," Electric Power Systems Research, vol. 226, Art. no. 109903, 2024.

[24] Y. Jiang, P. Dai, P. Fang, et al., "Electrical-STGCN: An electrical spatio-temporal graph convolutional network for intelligent predictive maintenance," IEEE Transactions on Industrial Informatics, vol. 18, no. 12, pp. 8509–8518, 2022.

[25] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," IEEE transactions on signal processing, vol. 62, no. 3, pp. 531–544, 2013.

[26] H. Li, T. Liu, X. Wu, et al., "An optimized VMD method and its applications in bearing fault diagnosis," Measurement, vol. 166, Art. no. 108185, 2020.

[27] C. K. Williams and D. Barber, "Bayesian classification with Gaussian processes," IEEE Transactions on pattern analysis and machine intelligence, vol. 20, no. 12, pp. 1342–1351, 1998.

[28] Z. Chen, S. Mak, and C. J. Wu, "A hierarchical expected improvement method for Bayesian optimization," Journal of the American Statistical Association, vol. 119, no. 546, pp. 1619–1632, 2024.

[29] C. Hall, "Natural cubic and bicubic spline interpolation," SIAM Journal on Numerical Analysis, vol. 10, no. 6, pp. 1055–1060, 1973.

[30] J.-H. Cho, W. Song, H. Choi, et al., "Hole filling method for depth image based rendering based on boundary decision," IEEE Signal Processing Letters, vol. 24, no. 3, pp. 329–333, 2017.

[31] I. Izonin, R. Tkachenko, N. Shakhovska, et al., "A two-step data normalization approach for improving classification accuracy in the medical diagnosis domain," Mathematics, vol. 10, no. 11, Art. no. 1942, 2022.

[32] H. Chen, M. Zhu, X. Hu, et al., "Research on short-term load forecasting of new-type power system based on GCN-LSTM considering multiple influencing factors," Energy Reports, vol. 9, pp. 1022–1031, 2023.

[33] Y. Hu, X. Cheng, S. Wang, et al., "Times series forecasting for urban building energy consumption based on graph convolutional network," Applied Energy, vol. 307, Art. no. 118231, 2022.

[34] S. Zhang, S. Zhang, J. James, et al., "ST-AGNet: Dynamic power system state prediction with spatial–temporal attention graph-based network," Applied Energy, vol. 365, Art. no. 123252, 2024.

[35] J. Lin, J. Ma, J. Zhu, et al., "Short-term load forecasting based on LSTM networks considering attention mechanism," International Journal of Electrical Power & Energy Systems, vol. 137, Art. no. 107818, 2022.