Solar Energy 300 (2025) 113770

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

PV power forecasting method using a dynamic spatio-temporal attention graph convolutional network with error correction

Zhao Zhen [a,b,d,e], Yufei Yang [a,b], Fei Wang [a,b,c,d,*], Nanpeng Yu [e],
Gang Huang [f], Xiqiang Chang [f], Guoqing Li [f]

[a] Yanzhao Electric Power Laboratory of North China Electric Power University, Baoding 071003, China
[b] Department of Electrical Engineering, North China Electric Power University, Baoding 071003, China
[c] State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (North China Electric Power University), Beijing 102206, China
[d] Hebei Key Laboratory of Distributed Energy Storage and Microgrid (North China Electric Power University), Baoding 071003, China
[e] Department of Electrical and Computer Engineering, University of California, Riverside, Riverside, CA 92521, USA
[f] State Grid Xinjiang Electric Power Co., Ltd, Urumqi 830018, China

ARTICLE INFO

Keywords:
Short-term PV power forecasting
Feature extraction
Multi-source NWP
STAGCN
Error correction

ABSTRACT

Superior short-term photovoltaic (PV) power forecasting is essential for formulating PV generation plans and implementing economic dispatch in power systems. However, existing short-term forecasting methods do not delve deeply into the interpretability of input features. They rely on static correlation analysis methods for processing numerical weather prediction (NWP), and often overlook the crucial step of correcting power forecasting errors. Addressing these three research deficiencies, this study develops a short-term PV power forecasting method employing a dynamic spatiotemporal attention graph convolutional network (STAGCN) with decomposition-based error correction. Initially, the spatio-temporal importance model explanation method is employed for feature extraction from multiple NWP sources, identifying key features for model forecasting. Furthermore, the spatiotemporal feature transformation and fusion technique is incorporated to construct a dynamic adjacency matrix and spatiotemporal dynamic graph structure using derived multi-source features. This structure feeds into the STAGCN model, yielding initial forecasting results. Ultimately, the improved complementary ensemble empirical mode decomposition with adaptive noise and variational mode decomposition, long short-term memory network and mixed kernel density estimation are utilized for the decomposition and forecasting of error sequences, subsequently refining the preliminary forecasting results. Simulation results demonstrate that the proposed forecasting approach enhances model interpretability, reduces forecasting errors, and achieves improved forecasting accuracy.

1. Introduction

The transition to renewable energy is reshaping the global energy landscape, holding significant strategic importance for achieving carbon neutrality, ensuring energy security, and combating climate change [1,2]. Among clean energy sources such as wind, hydro, and solar power, photovoltaic (PV) generation has emerged as the leading contributor to newly installed capacity due to its distinctive advantages. [3–5]. According to IRENA statistics, solar and wind energy remain the dominant forces in renewable energy expansion through 2025, jointly accounting for 96.6 % of new renewable capacity installations. Notably, solar power contributed over three-quarters of this new capacity. Solar installations surged by 32.2 %, reaching 1,865 GW. Wind power followed, growing by 11.1 %. China alone added 278 GW, with India contributing a further 24.5 GW [6].

The abundant availability of clean PV electricity offers benefits extending far beyond its inherent cleanliness and low-carbon attributes. It actively drives the development of new power systems, green certificate trading markets, carbon accounting mechanisms, and solar cell technologies [7–10]. Nevertheless, the volatility and intermittency of PV output remain core constraints to enhancing integration capacity in power grids with high renewable energy penetration. [11]. Spatio-temporal mismatches between PV output and electricity demand create risks of power curtailment and strain grid stability. Furthermore,

---

* Corresponding author at: Department of Electrical Engineering, North China Electric Power University, Baoding 071003, China.
E-mail address: feiwang@ncepu.edu.cn (F. Wang).

https://doi.org/10.1016/j.solener.2025.113770
Received 2 June 2025; Received in revised form 5 July 2025; Accepted 5 July 2025

the complex nature of changing meteorological conditions makes PV output challenging to manage. Accurate forecasting of PV generation over future periods is crucial; without it, the expanding application scope of PV power faces limitations. To address this challenge, researchers are increasingly exploring more advanced power forecasting techniques. Short-term PV power forecasting serves as a critical enabling technology that enhances grid integration capacity, drives power system transformation, optimizes energy storage dispatch, empowers electricity markets, and facilitates the achievement of dual-carbon goals [12,13].

Feature extraction is a critical step prior to undertaking forecasting tasks. It streamlines input data by identifying and transforming key variables and features, uncovering inherent data patterns to achieve efficient data compression and deeper information mining. This process enhances the forecasting model expressive power, reduces model redundancy, and improves computational efficiency. To address feature extraction in forecasting tasks, reference [14] calculated the Pearson correlation coefficient between input data and measured power. Their analysis revealed that for PV power generation, the most relevant numerical weather prediction (NWP) input variables are 15-minute resolution, hourly average irradiance, and solar elevation angle. Wind speed showed a lower correlation with power output and was consequently excluded from the feature set. Reference [15], considering the nonlinear correlation between meteorological features and wind power sequences, employed mutual information coefficients for feature analysis. Reference [16] applied grey relational analysis to compute correlations between meteorological factors, selecting the most relevant features for data dimensionality reduction. The process simplified the PV forecasting model and reduced computational complexity. These traditional feature analysis methods rely primarily on manual screening and basic correlation analysis. While capable of determining statistical associations between input and output feature sequences, they exhibit significant limitations. Specifically, it is difficult to validate the causality between the selected features and the model forecasting accuracy, resulting in opaque decision-making mechanisms. Current research has intensified systematic multidimensional investigations into advancing the interpretability of solar power forecasting models. When processing all-sky imager images and local meteorological data, reference [17] incorporates long short-term memory (LSTM) network to construct future representations of sky evolution. However, this approach lacks depth in spatial feature extraction, resulting in incomplete characterization of feature information. Reference [18] applied an artificial neural network (ANN) to extract features from NWP data, simulating the model application in operational PV forecasting scenarios. Reference [19] employed gaussian support vector regression to process NWP and satellite data, but provided limited explanation for the final forecasting results. Reference [20] compared seven models, including stacked ensemble learning, for feature extraction from sky images. Yet, the extraction process overlooked the dynamic evolution characteristics of cloud maps. Conversely, reference [21] developed a novel local variable selection network to integrate all input features for load forecasting, effectively improving the interpretability of deep learning models. Further, reference [22] conducted interpretability analysis on model outputs by calculating shapley additive explanation values to gain deeper insights into the causes of forecasting. Reference [23] proposed an attention-based temporal fusion transformer model. This model extracts statistical features from solar irradiance, temperature, and historical PV data and provides interpretable insights into the influence of different features. Reference [24] leveraged a domain-adversarial graph neural network (DAGNN), combining graph embedding with domain-adversarial algorithms, to extract domain-invariant forecasting features from graph structures. This enabled accurate power forecasting in data-scarce regions. The aforementioned works primarily attempt to explain the model forecasting process through single dimensions, such as input feature extraction, analysis of computational weight parameters, or evaluation of result metrics. A common limitation is the lack of joint validation methods that systematically link feature extraction with model forecasting to provide explanations for the information learned by the model. Therefore, effectively identifying and validating the key feature variables that truly drive model performance improvement, and subsequently enabling traceable analysis of forecasting results, merits further investigation in power forecasting.

In the field of PV power forecasting, researchers aim to develop model architectures, manage computational costs, and improve forecasting accuracy. To achieve this, various deep learning architectures have been widely adopted. Representative examples include recurrent neural networks (RNN) and their variants LSTM and gated recurrent unit (GRU), as well as convolutional neural networks (CNN) which excel at spatial feature extraction. At a practical level, reference [25] fused CNN and LSTM networks for solar radiation forecasting. Their results effectively supported the dynamic control and optimization of power output in PV power plants. Similarly, reference [26] focused on sky image data and utilized CNN technology to achieve short-term surface solar irradiance forecasting. This approach reduced forecasting bias. However, a key limitation of existing methods lies in their failure to explicitly model the spatial correlation characteristics of time series in non-Euclidean spaces, constraining model expressiveness [27]. To find suitable solutions, researchers are turning to approaches like graph convolutional neural networks (GCN). Currently, graph neural networks (GNN), represented by GCN, have primarily advanced in fields such as traffic flow forecasting [28–31], pedestrian trajectory forecasting [32], and power load forecasting [33]. Their application exploration in PV power forecasting remains relatively limited. Reference [34] proposed a sub-region division method incorporating spatio-temporal correlations for regional distributed PV short-term forecasting. This method used GCN to capture spatial dependency features and leveraged LSTM to model temporal evolution patterns. It built a power plant forecasting model adaptable to different weather scenarios, leading to improved forecasting performance. Nevertheless, this method relies on predefined static graph structures to characterize node relationships. It lacks support for the model to adaptively screen neighborhood information and also lacks a dynamic mechanism to capture time-varying interaction patterns between nodes. To optimize neighborhood information aggregation, some researchers have attempted to integrate attention mechanisms into GNN architectures. Reference [35] proposed the graph attention network (GAT). This architecture implicitly learns the weight distribution of neighboring nodes, enabling differentiated information aggregation. GAT was first applied to traffic flow forecasting and later gradually extended to the field of PV power forecasting [36,37]. Subsequent studies include reference [38] developing an improved time series dense encoder and graph attention network forecasting model to enhance the generalization capability of regional PV forecasting by extracting spatial features. Reference [39] combined amplitude-aware permutation entropy component reconstruction with GAT to build a graph sequence analysis framework for medium-to-long-term PV forecasting. Reference [40] constructed a parameter correlation adjacency matrix based on graph attention mechanisms, improving forecasting accuracy through dynamic weight optimization. Reference [41] designed a spatio-temporal attention-driven graph convolutional network, achieving accurate ultra-short-term forecasting for distributed PV via a dynamic adjacency matrix update mechanism. Reference [42] validated the effectiveness of an improved GAT framework in multi-source meteorological node modeling and dynamic topology modeling, confirming that dynamic adjacency matrices can converge forecasting errors. In summary, integrating spatio-temporal graph structures with dynamic adjacency matrices can significantly enhance model representational capacity. Furthermore, graph attention networks incorporating attention mechanisms have demonstrated considerable advantages in multi-domain time series forecasting tasks. However, while such GNN methods are widely applied in areas like traffic flow and pedestrian trajectory forecasting, they have yet to deeply penetrate PV power forecasting scenarios and exhibit notable limitations.

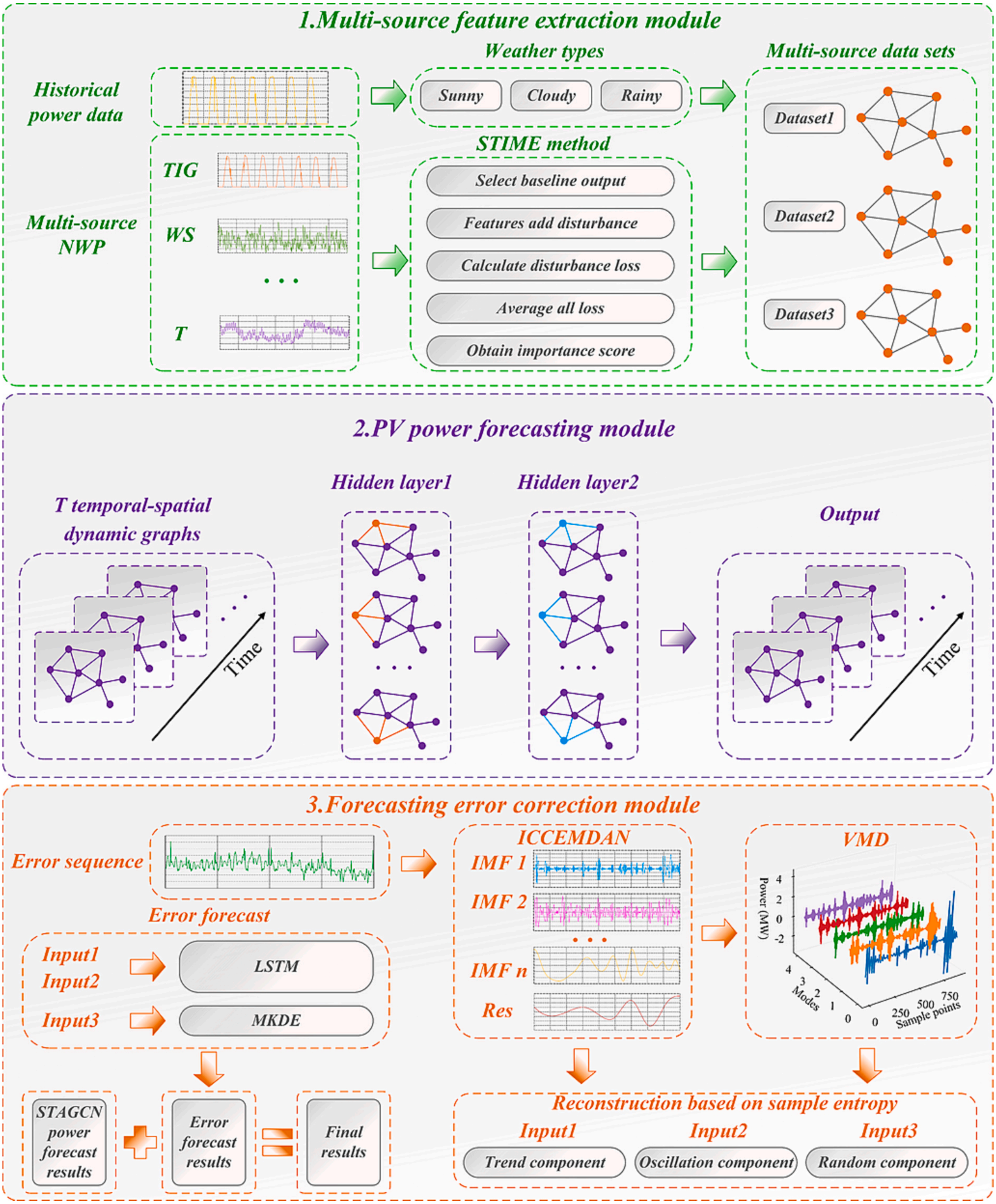


Fig. 1. The framework of the proposed forecasting method.

(1) Static Graph Limitations: Fixed adjacency matrices exhibit inadequate capability in processing complex multi-source feature data, leading to insufficient feature extraction in latent dimensions.

(2) Insufficient dynamic graph modeling: Current construction methods for dynamic adjacency matrices remain simplistic and offer limited feature representation capability. This hinders the extraction of deep spatio-temporal dependencies.

(3) Attention Technique Constraints: Existing attention techniques fail to uncover complex implicit relational topologies when processing irregular feature patterns.

A comprehensive PV power forecasting framework should encompass three steps including input feature extraction, forecasting model construction, and error correction mechanisms. While current research continuously introduces deep learning to enhance short-term forecasting accuracy, the focus largely lies on data quality optimization and model structure innovation. The need for post- forecasting error correction is often overlooked. PV power forecasting deviation is due to two factors. The first category is endogenous errors. These include model parameter configurations, inherent system biases, and measurement noise from data acquisition equipment. The second category involves exogenous disturbances. These encompass fluctuations in meteorological elements like irradiance, temperature, and humidity, which often exhibit predictable patterns. The interaction between these two categories leads to a combination of systematic errors and random deviations, necessitating the implementation of correction modules. Existing research on error correction predominantly adopts signal decomposition techniques [43–47]. For example, reference [48] used LSTM to correct forecasting based on similar-day samples but did not analyze the intrinsic characteristics of the error sequence itself. Reference [49] decomposed the error sequence using wavelet transform and coupled it with KNN subsequence forecasting for error compensation. Reference [47] applied both complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and variational mode decomposition (VMD) to decompose the original wind speed and error sequences. Reference [50] proposed a method based on improved adaptive noise complementary ensemble empirical mode decomposition (ICEEMDAN) and autoregressive integrated moving average (ARIMA) for error decomposition and forecasting. Comparisons with methods like empirical mode decomposition (EMD) and CEEMDAN demonstrated the effectiveness of error correction in reducing forecasting errors. It is important to note that the aforementioned methods are primarily applied in wind speed and wind power forecasting. Research on modal decomposition specifically for PV power errors remains largely unexplored. Therefore, conducting an in-depth analysis of the influencing factors, fundamental composition, and physical meaning of PV power errors is necessary to effectively correct errors in solar power forecasting results sequence. To achieve a breakthrough in PV power forecasting technology, it is urgently required to focus on the following dimensions based on a comprehensive analysis of the limitations in existing research:

1. The research on the interpretability of input variables is not considered in the feature extraction stage of input data.
2. During the model forecasting phase, few studies construct graph structures using multiple NWP sources, and the adoption of static, fixed adjacency matrices fails to adapt to changing meteorological features accordingly. Even when dynamic adjacency matrices are considered, they are often constructed using linear correlation coefficients. Such methods are limited in their ability to capture the complex, nonlinear relationship that commonly exist among meteorological variables, thereby restricting the model capacity to fully represent spatiotemporal dependencies.
3. In the stage of result analysis, the error correction of power forecasting results is often neglected. Even when error correction is incorporated, the underlying mechanisms contributing to forecasting errors are often not thoroughly analyzed or accurately understood.

Aiming at the above three problems, a comprehensive PV power forecasting method is proposed that integrates a dynamic spatiotemporal attention graph convolutional network (STAGCN) with iterative decomposition-based error correction. This method offers a novel approach that effectively improves forecasting accuracy by addressing three aspects of the forecasting tasks.

1. A spatiotemporal importance model explanation (STIME) that can identify the important features of the forecasting process is proposed. The spatio-temporal attention coefficient of the STAGCN model is combined with the STIME method to extract features and the influence of multi-source features on the model is further explored.
2. A STAGCN model based on spatiotemporal attention feature transformation and fusion technique is developed. By combining spatiotemporal transformation fusion vectors to construct a spatiotemporal correlation coefficient tensor and a dynamic graph, the model facilitates the transition of traditional approaches from static to dynamic, leading to modest improvements in forecasting accuracy.
3. A new error correction strategy based on ICEEMDAN-VMD secondary decomposition is proposed. This approach conducts a deep analysis of the error correction mechanism, decouples and reconstructs the error time series, and applies targeted correction methods to PV error components – effectively reducing overall forecasting error.
4. Using a variety of comparison methods, the reliability of feature extraction strategy, dynamic space–time diagram, multiple error correction and other ideas in forecasting work is verified.

The subsequent sections of this paper are structured as theoretical exposition, case simulation, and concluding synthesis. Section II elaborates the fundamental theoretical framework encompassing multi-source NWP feature extraction, the STAGCN-based forecasting architecture, and error correction techniques. Section III comprehensively details experimental dataset composition and computational environment configuration. Concretely, comparative validation against six advanced graph neural network benchmarks demonstrates the superiority and engineering viability of the proposed PV power forecasting approach. Concluding remarks are presented in Section IV along with prospective research directions addressing data volume limitations and extreme weather adaptation mechanisms.

2. Methodology

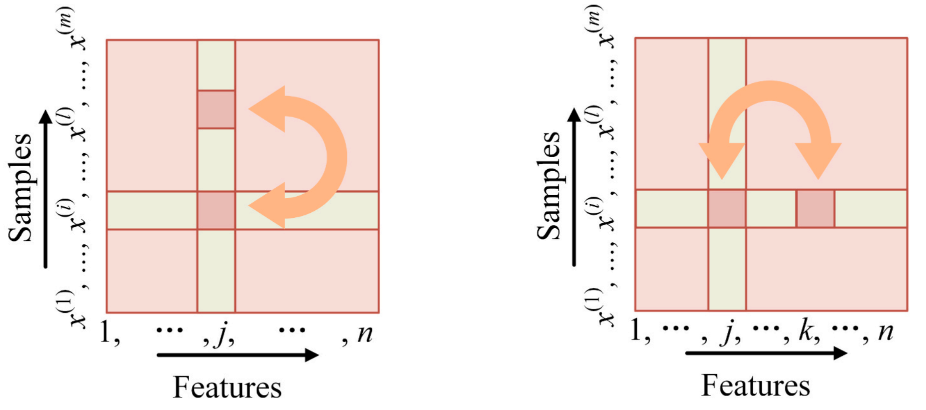
The forecasting approach proposed in this paper comprises three essential components: a feature extraction unit, a power forecasting module, and an error correction technique. The complete procedural workflow is illustrated in Fig. 1. Subsequent content will elucidate the operational principles and interactive dynamics of these integrated modules.

During the feature extraction phase, historical PV power data initially facilitate weather pattern classification. Key information extraction from multi-source NWP meteorological parameters encompassing irradiation, wind speed, and temperature employs the STIME technique, establishing dedicated datasets for three distinct weather patterns: clear, cloudy, and rainy conditions. The power forecasting module integrates spatiotemporal transformation fusion to process input features and construct dynamically evolving graphs. These graphs feed into the first two hidden layers of the STAGCN model, generating initial power forecasting outcomes. Construction specifics regarding STAGCN dynamic adjacency matrices and spatiotemporal graphs receive detailed elaboration in subsequent sections. The concluding error correction module implements a dual-decomposition strategy combining ICEEMDAN and VMD methodologies on forecasting residuals. Sample entropy reorganizes the derived components into trend, oscillatory, and stochastic elements. Employing LSTM and MKDE techniques respectively forecasts these constituent elements. Superposition of these forecasted components onto the STAGCN initial outcomes produces refined high-accuracy power forecasting.

2.1. Multi-source feature extraction module

2.1.1. The effectiveness of multi-source NWP

PV power plants require power data based on meteorological features from NWP for formulating power generation plans across different time scales. A strong physical linkage exists between variations in meteorological elements and the power output from solar PV panels. System operators depend on power forecasts for participation in



(a) Temporal dimension sample perturbation

(b)Spatial dimension feature perturbation

Fig. 2. The schematic diagram of STIME method.

electricity markets, developing bidding strategies, and power system dispatch, highlighting the significance of NWP [18]. Nevertheless, some research omits NWP in short-term forecasting. Reference [34] analyzed and forecasted power for a specific station by employing forecast information from neighboring stations. While this approach achieved favorable forecasting performance, it does not explain the physical causes behind the forecast results. Given the lack of direct causal relationships in mapping input power sequences to output power sequences, disregarding NWP in short-term PV power forecasting appears less reasonable [51].

The inherent difficulty in forecasting with a single NWP source often leads to substantial forecast errors under complex and changing weather conditions [52]. Integrating satellite cloud images with multi-source meteorological data, [53] effectively captured two-dimensional and three-dimensional spatial and temporal features within the data, improving the PV power forecasting capability of their model. Reference [54] demonstrated that forecasting methods incorporating ensemble forecasting systems exhibit significant advantages in enhancing the economic value of PV production. Reference [55] proves that the ensemble NWP obtained by running the same NWP model multiple times is effective for the probabilistic forecasting of PV power. Therefore, extracting meteorological feature information from multi-source NWP is a powerful means to increase the reliability of the model.

2.1.2. Spatio-temporal importance model explanation method

STIME, the novel approach proposed in this study, is purposefully designed to quantitatively evaluate the spatiotemporal importance of various features within input samples during the model forecasting process. The mechanism of this method involves strategically perturbing the distribution of two-dimensional input samples along both the lateral (feature space dimension) and longitudinal (temporal dimension) axes. This is achieved by generating a set of input samples with varying permutations and combinations of feature orders, thereby enabling the observation and analysis of corresponding changes in model performance. Following the perturbation applied to samples of a specific feature, the forecasting accuracy of the model output exhibits distinct fluctuation patterns. Based on these fluctuation patterns, the spatiotemporal contribution score of that feature to the model overall forecasting capability can be quantified and computed. STIME extracts and integrates meteorological feature information from multiple NWP sources pertaining to both the target forecasting station and neighboring stations. The application of the STIME method not only facilitates an in-depth analysis of the model internal decision logic, identifying meteorological factors critically influencing forecasting accuracy and thereby enhancing model interpretability, but also leverages its quantitative results to effectively fuse the complementary information inherent in multi-source NWP data. This synergistic fusion ultimately leads to improved forecasting performance for the target station.

Given an n-dimensional meteorological feature space provided by multi-source NWP data, where the feature dimensionality $n \in \mathbb{N}^+$. Let the pre-trained PV power forecasting model be represented as the mapping function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. For an input sample feature vector $x \in \mathbb{R}^n$, the model outputs the forecasting PV power value $\hat{y} = f(x)$. Our feature extraction task is to compute the spatiotemporal importance contribution score $I(f, j)$ of a specific feature ϕ_j, where $j \in \{1, 2, ..., n\}$, within the feature set towards the final forecasting value $\hat{y}$. To rigorously evaluate the model generalization performance, a test set $D = \left\{\left(x^{(i)}, y^{(i)}\right) | i = 1, 2, ...m\right\}$ containing m independent samples is constructed, where $x^{(i)}$ denotes the feature vector of the i_{th} sample and $y^{(i)}$ represents its corresponding true power observation. Forecasting deviation is quantified using a loss function $L(\hat{y}, y)$. Prior to initiating the spatiotemporal importance analysis, all features undergo min–max normalization to eliminate numerical scale bias introduced by differing units. When calculating feature importance across the spatiotemporal dimensions, a dynamic reference output can be established. For any sample i, its reference output is defined as:

$$f\left(x^{(i)}\right) = f\left(x_1^{(i)}, x_2^{(i)}, ..., x_j^{(i)}, ..., x_n^{(i)}\right) \tag{1}$$

While holding the feature dimension ϕ_j constant, perturbed samples $x_j^{(i,l)} (i \neq l)$ can be constructed to quantify the model temporal response characteristics across the evolutionary sequence. Specifically, by contrasting the model output differences arising from the permutation of features between sample i and sample l while preserving ϕ_j, the resulting state variation along the longitudinal dimension can be computed. When feature ϕ_j undergoes migration across different samples, the resultant change in the model forecasting output can be characterized by:

$$f\left(x_j^{(i,l)}\right) = f\left(x_1^{(i)}, x_j^{(i)}, ..., x_k^{(l)}, ..., x_n^{(i)}\right) \tag{2}$$

where the value of feature ϕ_j in sample i is replaced by the corresponding value from sample l, as illustrated in Fig. 2(a). Furthermore, feature interaction effects within the spatial dimension can be quantified through lateral permutation. This involves replacing the value of feature ϕ_j in sample $x^{(i)}$ with the observed value of a different feature $\phi_k (k \neq j)$ within the same sample i, as depicted in Fig. 2(b). The resultant model output for this laterally perturbed sample is defined by equation (3):

$$f\left(x_{(j,k)}\right) = f\left(x_1^{(i)}, x_k^{(i)}, ..., x_j^{(i)}, ..., x_n^{(i)}\right) \tag{3}$$

Considering the calculation method of the reference output is defined above, the space–time loss variation is measured by two loss functions:

$$\Delta L_j^{(i,l)} = L\left[y^{(i)}, f\left(x_j^{(i,l)}\right)\right] - L\left[y^{(i)}, f\left(x_j^{(i)}\right)\right] \tag{4}$$

$$\Delta L_{(j,k)}^{(i)} = L\left[y^{(i)}, f\left(x_{(j,k)}^{(i)}\right)\right] - L\left[y^{(i)}, f\left(x_j^{(i)}\right)\right] \tag{5}$$

Let $S_i = \{1, 2, \dots, m\}\backslash\{i\}$ denote the index subset excluding sample i. By iterating over all permutations of feature ϕ_j using samples within S_i, and computing the loss variation $m-1$ times, an expected estimate of the sensitivity change along the temporal dimension is obtained. Similarly, defining the feature exclusion set $\Gamma_j = \{1, .., k, \dots, n\}\backslash\{j\}$ to represent the complementary feature space of ϕ_j, allows for the computation of the expected estimate of the sensitivity change along the spatial dimension. The integrated calculation of spatiotemporal sensitivity change is formulated as:

$$\Delta \overline{L}_j^{(i)} = \frac{1}{|S_i|}\sum_{l\in S_i}\Delta L_j^{(i,l)} \tag{6}$$

$$\Delta \overline{L}_{(j,k)}^{(i)} = \frac{1}{|\Gamma_j|}\sum_{k\in \Gamma_j}\Delta L_{(j,k)}^{(i)} \tag{7}$$

The average loss variation has now been computed, based on the reference output, for sample i with respect to feature ϕ_j and for all features within sample i. Extending this computational framework to the entire sample space yields the spatiotemporal importance score $I(f,j)$ for feature ϕ_j:

$$I(f,j) = \frac{1}{m}\sum_{i=1}^{m}\left(\Delta \overline{L}_j^{(i)} + \Delta \overline{L}_{(j,k)}^{(i)}\right) \tag{8}$$

2.2. PV power forecasting module

2.2.1. PV power forecasting problem description

Forecasting models require capturing the multidimensional coupling mechanisms between meteorological elements and power generation. Dynamic processes of cloud movement induce coordinated changes in complex meteorological parameters among neighboring stations, forming nonlinear correlation networks spanning multiple spatial and temporal scales.Building upon this physical foundation, a dynamic spatio-temporal graph attention model is constructed where a time-varying edge set characterizes interrelationships among multi-source meteorological features and dynamic adjacency matrices enhance feature variable interpretability through dynamic encoding of spatio-temporal attention coefficients between nodes. This architecture establishes spatio-temporal topological mapping relationships by integrating multi-source NWP historical meteorological data with the power station time-series output. Multi-source NWP meteorological variables and the historical power sequence are jointly represented as graph signals defined on this dynamic graph structure.

The dynamic graph sequence $\{\mathscr{G}^t | t = 1, 2, \dots, T\}$ comprises a series of time-varying weighted undirected graphs with adjacent graph instances $\mathscr{G}^t$ and $\mathscr{G}^{t+1}$ separated by a fixed time interval τ along the temporal dimension. Each graph structure $\mathscr{G}^t = (\mathscr{V}, \mathscr{E}^t, A^t)$ contains three components including a fixed node set $\mathscr{V} = \{v_i\}_{i=1}^N$ corresponding to multi-source feature vectors, a time-varying edge set $\mathscr{E}^t$ representing dynamic spatial dependency connections between nodes where an element $e_{ij}^t = (v_i, v_j)$ denotes a connection existing at time t, and a dynamic adjacency matrix $A^t \in \mathbb{R}^{N\times N}$ fully characterizing the graph time-varying topology. The elements A_{ij}^t quantify the association strength between nodes v_i and v_j at time t with a value of zero assigned if no connection exists. This formulation defines the graph structure as a mapping $x^t : \mathscr{V} \rightarrow \mathbb{R}$. The forecasting task is formally expressed as utilizing the historical feature node state tensor for the past M time steps $X_{\tau-M:\tau-1} = [x_{\tau-M}, x_{\tau-M+1}, \dots, x_{\tau-1}]$ where $x_\tau \in \mathbb{R}^N$ represents the full node feature vector at time τ to infer the future H-step power output sequence. The forecasting process is expressed through equations (9) and (10).

$$[X_{\tau-M:\tau-1}, G^t] \xrightarrow{f} \widehat{P}_{\tau:\tau+H-1} \tag{9}$$

$$X_{\tau-M:\tau-1} = X_{\tau-M}, X_{\tau-M+1}, \dots, X_{\tau-1} \tag{10}$$

2.2.2. Attention-based spatiotemporal feature transformation and fusion technique

Spatial attributes exist among meteorological feature nodes. Weighted graph edge connections are often represented using the Pearson correlation coefficient or distance-based gaussian radial basis functions and inverse functions [56]. Such descriptions may not fully capture the underlying correlation characteristics between meteorological features. This paper introduces an attention-based spatiotemporal feature transformation and fusion technique to construct a dynamic adjacency matrix, thereby characterizing the complex interrelationships among meteorological features.

The STAGCN model employs a spatiotemporal co-attention architecture that integrates temporal and spatial transformation fusion structures through a frame segmentation mechanism. Within this framework, the temporal attention feature transformation and fusion technique first transforms multi-source features along non-stationary time dimensions to capture nonlinear dependencies, then combines them with a dynamic spatial attention technique to identify spatial correlations within time-varying implicit features. Node signal representations are defined as $x_i = \left[x_i^1, x_i^2, \dots, x_i^f\right] \in \mathbb{R}^f$ for $i \in \{1,\ 2,\ \dots,\ N\}$, where f denotes the node feature dimensionality. A multi-dimensional attention technique quantifies the feature relevance and topological influence of node j on target node i by retaining implicit information from preceding and subsequent sequences [35]. The feature space embedding process utilizes a learnable linear transformation kernel $W \in \mathbb{R}^{f' \times f}$ to project original features into an f'-dimensional latent space, facilitating feature fusion, followed by the derivation of normalized attention coefficients α_{ij}.

$$e_{ij} = LeakyReLU\left(a \cdot \left[Wx_i \| Wx_j\right]\right), j \in \mathscr{N}_i \tag{11}$$

$$\alpha_{ij} = softmax_j\left(e_{ij}\right) = \frac{\exp\left(e_{ij}\right)}{\sum_{k\in\mathscr{N}_i}\exp(e_{ik})} \tag{12}$$

The original association strength e_{ij} characterizes the unnormalized degree of node interaction. Here, the modified $LeakyReLU(\cdot)$ unit serves as the nonlinear transformation function, with its advantage lying in mitigating the vanishing gradient problem. The parametric vector $a \in \mathbb{R}^{2f' \times 1}$ acts as a trainable adaptive weight regulator. Feature interaction detection is achieved via the vector dot product operation $\cdot$. The node feature fusion employs the tensor concatenation operator $\|$, mapping high-dimensional features into a unified representation space. The initial association degree e_{ij} obtained through the above operations requires probabilistic normalization. This transformation ensures the attention weights within the neighborhood satisfy the probability distribution law $\sum_j \alpha_{ij} = 1$. Finally, the semantically enhanced representation h_i for node v_i is generated through attention-guided neighborhood aggregation as expressed in Equation (13), where $\mathscr{N}_i$ represents the neighborhood of node i and $\sigma(\cdot)$ denotes the activation function.

$$h_i = \sigma\left(\sum_{j\in\mathscr{N}_i}\alpha_{ij} W x_j\right) \tag{13}$$

The temporal attention feature transformation and fusion technique demonstrates advantages in fully leveraging nonlinear dependencies and evolving characteristics embedded across different time frames of

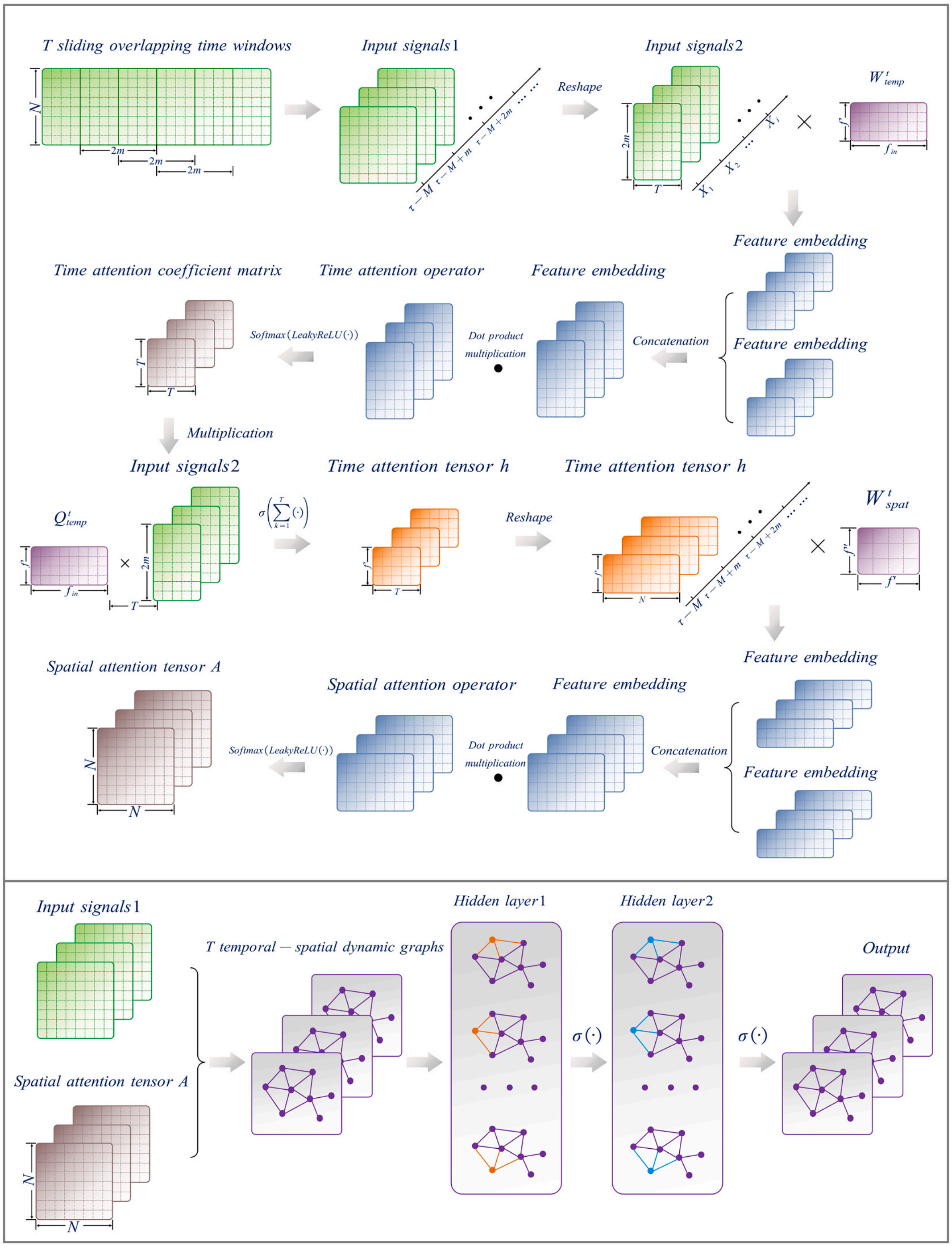


Fig. 3. STAGCN forecasting model construction process.

sequential data. Its input corresponds to the feature vector $x_i^t \in \mathbb{R}^{f_{in}\times 1}$ of node i at time frame t. Considering the notable time-varying nature of influence weights from different time frames on the current state, this technique dynamically adjusts its computational strategy. Specifically, it introduces a learnable temporal weight matrix $W_{temp}^t \in \mathbb{R}^{f'\times f_{in}}$ for each time frame t to comprehensively extract cross-timestep feature correlations, achieving feature transformation of multi-source features along the longitudinal dimension. By applying an adaptively varying temporal feature fusion vector $a_{temp}^t \in \mathbb{R}^{2f'\times 1}$ to all nodes and time frames, combined with the tensor concatenation operator $\|$ for feature fusion, this technique processes temporal characteristics. For target node i and reference time frame t, the computation of temporal attention transformation fusion coefficient α_i^{kt} integrates the nonlinear transformation of the *LeakyReLU*$(\cdot)$ activation function with the normalization of the *Softmax*$(\cdot)$ function, enhancing the retention of temporal transformation and fusion information during model training.

$$\alpha_i^{kt} = softmax\left(LeakyReLU\left(a_{temp}^t\cdot\left[W_{temp}^t x_i^k \middle\| W_{temp}^t x_i^t\right]\right)\right) \tag{14}$$

The coefficient α_i^{kt} quantifies the temporal influence of multi-source features from historical time frame k on feature node i within the context of time frame t. Based on the computed normalized temporal attention transformation fusion coefficient α_i^{kt}, weighted aggregation of multi-source features across all time frames k generates the time-aware embedded feature vector $h_i^t \in \mathbb{R}^{f' \times 1}$ for node i at current frame t. Subsequently, stacking embedded vectors of all nodes N across all time frames T forms the temporal attention transformation fusion coefficient matrix $h^t \in \mathbb{R}^{N \times f'}$ and the complete temporal attention transformation fusion feature tensor $h \in \mathbb{R}^{N \times T \times f'}$. And the time transform weight matrix is $Q_{temp}^t \in \mathbb{R}^{f' \times f_{in}}$.

$$h_i^t = \sigma\left(\sum_{k=1}^{T} \alpha_i^{kt} Q_{temp}^t x_i^k\right) \tag{15}$$

To capture complex spatial correlations among different nodes within the same time frame t, we introduce an attention-based spatial feature transformation and fusion technique. This technique utilizes the temporal embedded feature tensor h, obtained from the preceding temporal transformation fusion step, as its foundational input. The spatial transformation fusion technique constructs a dynamically updated spatial relation graph. For each time frame t, it computes the spatial association strength A_{ij}^t between any node pair (i, j) according to equation (16).

$$A_{ij}^t = softmax\left(l\left(a_{spat}^i \cdot \left[W_{spat}^t h_i^t \middle\| W_{spat}^t h_j^t\right]\right)\right) \tag{16}$$

where A_{ij}^t characterizes the importance of multi-source feature information from node j to node i at specific time frame t. The spatial feature fusion vector $a_{temp}^t \in \mathbb{R}^{2f' \times 1}$, employed to compute association strength, remains consistent across different time frames, ensuring model parameter efficiency. Association strengths of all multi-source node pairs at time frame t constitute the spatial feature fusion coefficient matrix $A^t \in \mathbb{R}^{N \times N}$ for that moment, collectively forming the spatial feature fusion tensor A for the entire sequence. Further processing is typically required before computing spatial associations, where node features undergo spatial domain feature mapping. For this purpose, the technique equips each time frame t with a dedicated spatial transformation weight matrix $W_{temp}^t \in \mathbb{R}^{f' \times f_{in}}$. This matrix projects high-dimensional features from the temporal fusion output (dimension f') into a lower-dimensional latent feature space f'', refining more discriminative spatial interaction features while reducing model complexity.

2.2.3. Dynamic STAGCN model

This study proposes a dynamic STAGCN model incorporating multi-source NWP meteorological features. By leveraging time-varying information between multi-source NWP meteorological characteristics and PV power, the model adaptively captures spatiotemporal correlations. It integrates a spatiotemporal attention transformation and fusion technique with convolutional layers to perform PV power forecasting for single power stations, with the forecasting workflow illustrated in Fig. 3.

To effectively process the N-node feature matrix composed of multi-source NWP and PV power data, we employ an attention-based spatio-temporal feature transformation and fusion technique to partition the multivariate feature matrix into overlapping time frames. Specifically, continuous input sequences are segmented into T spatiotemporal analysis frames with a 50 % overlap ratio. This half-frame overlap strategy draws inspiration from reference [57], which utilizes overlap-add techniques to reconstruct original signals from windowed frames, aiming to preserve temporal continuity while reducing modeling complexity. These segments are designed to overlap rather than remain independent, with each analysis unit sharing half its data with adjacent units. When the input sequence contains M historical observation points, setting the total frame count T yields a single-frame scale of $2m$, satisfying $m(T+1) = M \in \mathbb{N}$. This partitioning mechanism effectively mitigates two pitfalls of long-term temporal modeling: direct capture of full-node feature correlations over M steps causes computational redundancy, while fine-grained temporal segmentation may induce forecasting oscillations.

Following this overlap design, the original multi-source feature matrix is defined as $X = [x_\lambda, x_{\lambda+1}, \dots, x_{\lambda+2m}] \in \mathbb{R}^{N \times T \times f_{in}}$, where feature dimension f_{in} equals frame length $2m$. The set of traversal for the frame start index λ is $\{\tau - M + k \cdot m | k = 0, 1, \dots, T\}$, covering the historical period from $\tau - M$ to $\tau - 1$. This initial structure (Input Signal 1) undergoes reshaping to vertically integrate identical feature sequences across all frames, generating a reorganized multi-source feature tensor (Input Signal 2). For deep temporal feature extraction, the attention-based temporal feature transformation and fusion technique first processes the reshaped tensor. This technique captures dynamic correlations among embedded features across time points, outputting matrices containing N temporal transformation fusion coefficients that quantify influence weights of each feature node within specific time frames. Subsequently, these coefficient matrices form a higher-dimensional temporal fusion tensor, enabling dynamic weight allocation for temporal analysis. Following temporal processing, spatial transformation fusion operates on the temporally embedded features. Spatial fusion vectors dynamically generate spatial transformation fusion tensor A. Critically, these vectors are computed independently per time frame, rendering tensor A a time-evolving dynamic adjacency matrix that precisely characterizes shifting spatial relationships among multi-source feature nodes across time. Finally, integrating this dynamic adjacency matrix A with spatiotemporally transformed features constructs graph-structured data encapsulating dynamic spatiotemporal dependencies. This graph structure feeds into graph convolutional layers that perform feature propagation and aggregation on the spatiotemporal dynamic graph, comprehensively integrating cross-dimensional dependencies to generate preliminary PV power forecasting. The propagation pattern between graph convolutional layers follows [51]:

$$X^{l+1} = \sigma\left(\widehat{D}^{-\frac{1}{2}} \widehat{A} \widehat{D}^{-\frac{1}{2}} X^{(l)} W^{(l)}\right) \tag{17}$$

where the node feature matrix is X , the self-loop augmented adjacency matrix is $\widehat{A} = A + A_n$, A_n is the identity matrix, and symmetric normalization is applied based on the degree matrix $\widehat{D}$. The trainable parameter set is denoted as the weight matrix W, and the feature transformation process is implemented by the nonlinear activation function $\sigma(\cdot)$.

2.3. Forecasting error correction module

PV power forecasting deviation constitutes a critical bottleneck constraining further improvements in forecasting accuracy. As discussed in the literature review, existing research commonly overlooks the error correction stage in the forecasting backend or inadequately analyzes the internal composition mechanisms of errors. Traditional single correction methods face difficulties in fully characterizing and eliminating such composite errors. Therefore, constructing an error correction module based on ICEEMDAN-VMD secondary decomposition is essential for enhancing final forecasting accuracy. This module conducts an in-depth analysis of the intrinsic composition and physical meaning of the original forecasting error sequence P^{err}, separating error subcomponents dominated by different physical mechanisms including endogenous noise, exogenous meteorological disturbances, and latent system biases. It then performs targeted reconstruction of each subcomponent, ultimately yielding a more accurate post- forecasting result through correction. The error sequence is formally defined as the actual power minus the forecasting power, as expressed in equation (18).

$$P^{err} = P^{act} - P^{for} \tag{18}$$

The module first employs ICEEMDAN to perform primary decomposition on the original error sequence P^{err}. This step effectively mitigates mode mixing issues and reduces residual noise, preliminarily separating error oscillation patterns at different time scales along with the residual. To address the complex mixed information, particularly within high-frequency modes, that may persist after primary decomposition, VMD is subsequently applied for secondary fine decomposition. Leveraging its strong adaptability and precise frequency band segmentation capability, VMD decouples potential subsequences within high-frequency error components driven by distinct physical sources. Finally, for the subsequences obtained through decomposition, the sample entropy method is utilized for reconstruction. The reconstructed components are then forecasted using LSTM networks and MKDE, achieving effective error compensation. This secondary decomposition framework offers a novel approach for deepening the understanding of PV forecasting error composition and enabling precise correction.

2.3.1. *Primary decomposition of ICEEMDAN*

Existing error correction methods based on EMD and its variants Ensemble EMD (EEMD) and CEEMDAN often encounter issues of mode mixing and residual noise interference when processing non-stationary, nonlinear signals like PV forecasting errors, hindering the effective extraction of intrinsic error modes [50]. To address these limitations, ICEEMDAN is selected as the primary error decomposition tool. ICEEMDAN significantly suppresses mode mixing by adaptively injecting uniquely modulated white noise at specific stages of the decomposition process and computing the ensemble average of local means. This approach concurrently reduces the residual noise level within the ultimately extracted intrinsic mode functions (IMFs), thereby offering superior decomposition robustness and feature extraction capability, making it particularly suitable for dissecting the complex structure of PV forecasting errors.

Add special noise $E_k\left(w^{(i)}\right)$ to P^{err}, calculate the k_{th} residual and modal component P_k^{err} by (19) and (20).

$$r_k = \frac{1}{I}\sum_{i=1}^{I} M\left(r_{k-1} + \beta_k E_k\left(w^{(i)}\right)\right) \tag{19}$$

$$P_k^{err} = r_{k-1} - r_k \tag{20}$$

where $M(\cdot)$ is the operator that generates the local mean of the signal, and I is the number of white noise groups. $w^{(i)}$ denotes the i_{th} group of white noise used for auxiliary decomposition. β_k is the iterative expected signal-to-noise ratio for the k_{th} decomposition and r_k is the residual of the k_{th} decomposition.

2.3.2. *Secondary decomposition of VMD*

While ICEEMDAN primary decomposition effectively mitigates mode mixing, the high-frequency IMF components it yields typically remain complex non-stationary signals potentially containing mixed error fluctuations originating from multiple physical sources. Directly correcting such mixed components may yield limited effectiveness. To more precisely distinguish these complex high-frequency error constituents, VMD is introduced to perform secondary decomposition on the selected high-frequency IMFs [58].

$$\begin{cases} \min\limits_{\{u_k\},\{w_k\}}\left\{\sum_{k=1}^{K}\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-jw_kt}\right\|_2^2\right\} \\ s.t. \sum_{k=1}^{K} u_k = P_1^{err} \end{cases} \tag{21}$$

where K is the number of decomposed modes; $*$ is the convolution operator, $\|\cdot\|_2$ is the two norm, ∂_t is the partial derivative, and $\delta(t)$ is the Dirac function.

2.3.3. *Error correction strategy*

The error is decomposed into a series of components P_k^{err} by ICEEMDAN-VMD twice decomposition method, and then combined with sample entropy. The error component is reconstructed into trend component P_T^{err}, oscillation component P_O^{err} and random component P_R^{err}. The sample entropy formula is shown in (22).

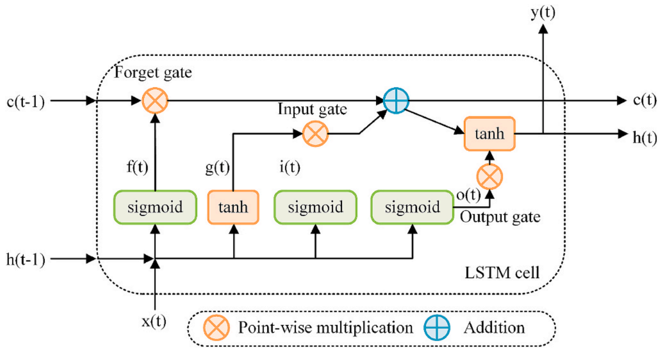


Fig. 4. The structure of LSTM cell.

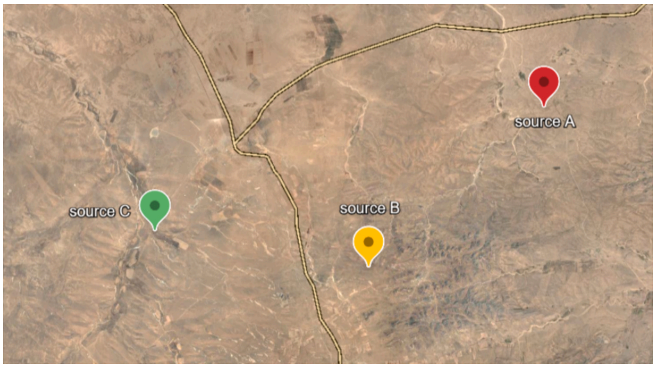


Fig. 5. Geographical location of multi-source NWP.

$$Sampen(m, r, N) = -\ln\left(\frac{B^{m+1}(r)}{B^m(r)}\right) \tag{22}$$

where $Sampen(\cdot)$ is calculates sample entropy; N is the length of the modal component; $B^m(r)$ and $B^{m+1}(r)$ are the probabilities that the sequence matches m points and $m+1$ points under tolerance r, respectively. For P_T^{err} and P_O^{err}, LSTM is used for error forecasting. LSTM is a mainstream deep learning method [59], and its cell structure is shown in the Fig. 4.

The MKDE method is used to forecast the error for P_R^{err}. The expression is shown in the following equation.

$$f_{m_0}(x) = \frac{1}{Nh_{m_0}}\sum_{i=1}^{N} K\left(\frac{x-x_i}{h_{m_0}}\right) \tag{23}$$

$$f_{MKDE}(x) = \sum_{m_0=1}^{M} w_{m_0} f_{m_0}(x) \tag{24}$$

$$K(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \tag{25}$$

In the formula, h_{m_0} is the width of the window m_0; N is the total number of samples; $K(\cdot)$ is the kernel function; x is a random variable, $f_{m_0}(x)$ is the kernel density expression, w_{m_0} is the weight coefficient of $f_{m_0}(x)$, $f_{MKDE}(x)$ is the expression of mixed kernel density, and M is the total number of bandwidths.

Table 1
Features and geographic information of multi-source NWP.

Multi-source NWP	Multi-source features	PV plants	Longitude	Latitude
source A	A_TIG, A_TEM, A_HUM, A_CA, A_WS, A_WD, A_PRE	#1	111°23′40″	41°58′45″
source B	B_TIG, B_TEM, B_HUM, B_CA, B_WS	#2	111°15′38′	41°51′48″
source C	C_TIG, C_TEM, C_HUM, C_CA, C_WS	#3	111°5′20″	41°53′17″

3. Case study

3.1. Data description & implementation details

The dataset originates from three geographically distinct data sources in the Ulanqab region of northern China, containing measured power data, NWP data for one target PV power station, and NWP data from two neighboring stations. The three groups of NWP sources contain 17 meteorological characteristics. The NWP of the target station is defined as source A, and other adjacent stations are defined as source B and source C. Fig. 5 shows the geographical locations of the three NWP sources.

If the seven meteorological features of total irradiance ground (TIG), temperature (TEM), humidity (HUM), cloud amount (CA), wind direction (WD), pressure (PRE), wind speed (WS) of the source A are expressed as A_TIG, A_TEM, A_HUM, A_CA, A_WS, A_WD, and A_PRE. Then the five meteorological features of source B and source C can be expressed as B_TIG, B_TEM, B_HUM, B_CA, B_WS, C_TIG, C_TEM, C_HUM, C_CA, C_WS, respectively. Table 1 describes the basic information of the three NWP sources. The time span of the data set used is from September 1, 2022 to October 1, 2024. After removing outliers and null values, the time step of the data set is 68448.

The method operates within the following hardware and software environment: the computer is equipped with NVIDIA GeForce GTX 1650 GPU, and the host memory is 8 GB. PyTorch 1.13.1 deep learning framework is used in python3.8 software to support CUDA 11.6 acceleration. This configuration can meet the computational requirements of deep learning tasks required for experiments.

3.2. Accuracy metrics

In this study, we employ the root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination R^2 as metrics to assess the forecasting accuracy of the proposed method. The formulas for calculating each metric are as follows:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(p_t - \widehat{p}_t)^2} \tag{26}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|p_t - \widehat{p}_t| \tag{27}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(p_t - \widehat{p}_t)^2}{\sum_{t=1}^{T}(p_t - \overline{p}_t)^2} \tag{28}$$

where p_t and $\widehat{p}_t$ represent the real power value and the forecasting power value at time t, respectively. T represents the forecasting time scale, and $\overline{p}_t$ is the average power value.

3.3. Simulation design

In the simulation design phase, the entire dataset was stratified by weather type employing accumulated historical PV measurements, yielding sunny, cloudy, and rainy categories accounting for 61 %, 21 %, and 18 % of the total dataset respectively. This study established six benchmark models for initial verification of the method leading capability. Subsequent comparative experiments confirmed the feature extraction superiority of the STIME approach. For the power forecasting component, multi-source NWP data derived from feature extraction alongside historical PV power data formed dynamic graph structures incorporating diverse graphical information. Joint interpretable analysis of input features was conducted via an attention mechanism for spatiotemporal feature transformation and fusion technique combined with importance scores generated by the STIME method, clarifying the critical role of dynamic spatiotemporal graphs. Regarding error correction, error sequences from different weather types underwent primary ICEEMDAN decomposition followed by secondary VMD decomposition. Sample entropy reconstructed the decomposed sequences into trend, oscillatory, and stochastic elements. LSTM network forecasted the trend and oscillatory elements, while MKDE forecasted the stochastic element. The preliminary power forecast was aggregated with forecasting of the three error components to yield the PV power forecast.

Table 2
Experimental configuration based on forecasting model and error correction mode.

Methods	Neural network	Dynamic graph	Error correction strategy		
			ICEEMDAN	VMD	ICEEMDAN-VMD
Proposed	STAGCN	Yes	No	No	Yes
Benchmark-1	STAGCN	Yes	No	Yes	No
Benchmark-2	STAGCN	Yes	Yes	No	No
Benchmark-3	STAGCN	Yes	No	No	No
Benchmark-4	STGCN	Yes	No	No	Yes
Benchmark-5	GAT	No	No	No	Yes
Benchmark-6	GCN	No	No	No	Yes

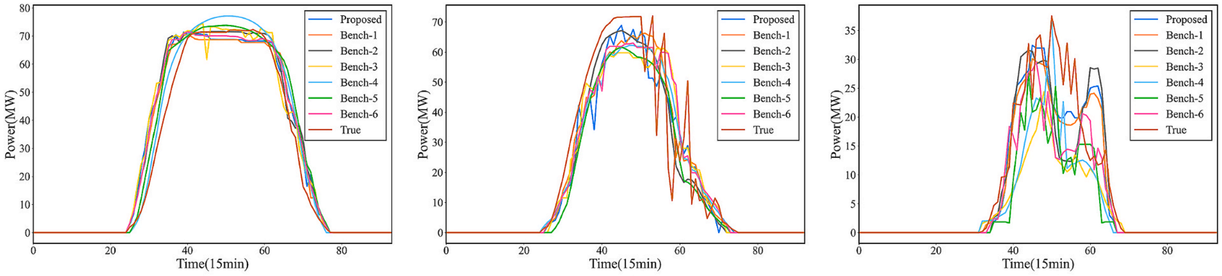
3.4. The results and discussion of simulation experiment

3.4.1. Comparisons with other setting models

To thoroughly investigate the forecasting capability of the method, ablation experiments examined three dimensions: selection of neural network architectures, integration of dynamic graph structures, and implementation of the ICEEMDAN-VMD secondary decomposition error correction strategy. Six supplementary benchmark methods pertaining to graph neural networks were introduced for comparative assessment. Fundamental configurations of these approaches appear in Table 2. Outcomes across distinct weather conditions are visualized in Fig. 6.

Table 3 presents the RMSE, MAE, and R^2 metrics for the proposed method and baseline methods under various weather conditions. When the weather type is sunny, comparing the three metrics between the proposed method and Bench-1 and Bench-2 reveals that different error correction strategies yield similar effects. A comparison between the proposed method and Bench-3 indicates that the twice decomposition error correction strategy enhances the RMSE, MAE, and R^2 of the STAGCN model by 2.71 %, 2.70 %, and 4.53 %, respectively. The STGCN model in benchmark 4 does not consider the attention mechanism. The GAT and GCN models of benchmark 5 and benchmark 6 adopt the network of static graph structure. Their performance is close, but they are not as good as the proposed method.

When the weather type is classified as cloudy, a similar approach to the analysis of power forecasting results for clear skies is employed. Comparing the proposed method with Bench-3, it is observed that incorporating an error decomposition correction strategy leads to a

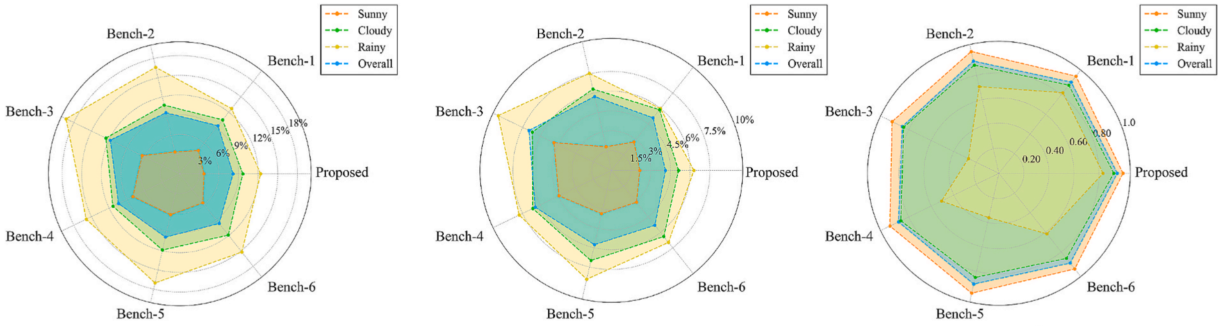


(a) Sunny (b) Cloudy (c) Rainy

Fig. 6. Power forecasting results for different weather types.

Table 3
Model performance under different weather types.

Sky Type	Metric	Forecasting methods						
		Proposed	Bench-1	Bench-2	Bench-3	Bench-4	Bench-5	Bench-6
Sunny	RMSE	3.70 %	4.55 %	3.42 %	6.41 %	7.96 %	6.38 %	5.64 %
	MAE	2.16 %	2.76 %	1.84 %	4.86 %	4.43 %	3.36 %	3.08 %
	R^2	0.9919	0.9876	0.9934	0.9466	0.9665	0.9775	0.9723
Cloudy	RMSE	9.64 %	10.50 %	10.72 %	12.55 %	11.35 %	11.89 %	11.91 %
	MAE	5.11 %	5.88 %	6.32 %	6.68 %	6.44 %	7.00 %	6.40 %
	R^2	0.9182	0.8962	0.8823	0.8446	0.8699	0.8492	0.8655
Rainy	RMSE	12.36 %	12.70 %	16.63 %	19.30 %	15.88 %	17.02 %	15.22 %
	MAE	6.28 %	5.99 %	7.54 %	9.56 %	7.79 %	8.45 %	6.97 %
	R^2	0.8326	0.8201	0.7067	0.2701	0.5090	0.3616	0.6157
Overall	RMSE	8.15 %	9.36 %	9.54 %	11.80 %	10.41 %	9.87 %	9.66 %
	MAE	4.12 %	5.09 %	5.74 %	6.96 %	6.34 %	5.77 %	5.31 %
	R^2	0.9462	0.9271	0.9145	0.8563	0.8902	0.9032	0.9133



(a) RMSE (b) MAE (c) R^2

Fig. 7. Indicator radar charts of all benchmark methods under different conditions.

reduction of 1.91 % in RMSE, 0.57 % in MAE, and an enhancement of 4.36 % in R^2; when contrasted with Bench-1 and Bench-2, it is found that the twice error decomposition correction strategy can maximally decrease RMSE by 1.08 % and MAE by 1.12 %. Moreover, the R^2 indicator of the proposed method stands at 0.9182, surpassing that of other methods.

During rainy conditions, meteorological features such as irradiance, temperature, and humidity exhibit significant fluctuations, leading to increased errors. Nevertheless, the present approach maintains lower RMSE and MAE values than all alternative methods. Its R^2 value of 0.8326 surpasses other benchmarks, suggesting strong agreement between forecast results and measured power values even under rainy circumstances.

To facilitate a comprehensive comparison and analysis of the metric results, Fig. 7 presents a comparative analysis. Overall, the RMSE and MAE of all methods are minimized under clear weather conditions, larger on the overall dataset and cloudy dataset, and maximized under rainy weather conditions, while the R^2 is generally lower in rainy conditions. To pursue lower errors, one could select methods corresponding to the flatter heptagonal areas in the radar chart for forecasting. For instance, during rainy weather, one could opt for the proposed method, Bench-1, Bench-2, and Bench-6. Collectively, the results from all three metrics demonstrate robust performance of the methodology in PV power forecasting.

The preceding analysis indicates that incorporating error correction enhances PV power forecasting accuracy across all weather types, with measurable improvements observed. Under adverse meteorological conditions, the methodology achieves superior forecasting precision, yielding substantially enhanced forecasting outcomes.

3.4.2. Feature extraction based on STIME method

In the case of limited computing resources, if all the multi-source NWP data close to the two-year time scale are fed into the model, it will inevitably lead to the increase of the complexity of the forecasting model, the learning of too many redundant features, the difficulty of convergence, and the decrease of forecasting accuracy. Based on this, ensuring that the input data of the model is useful information for forecasting and that the model can be interpreted according to the input

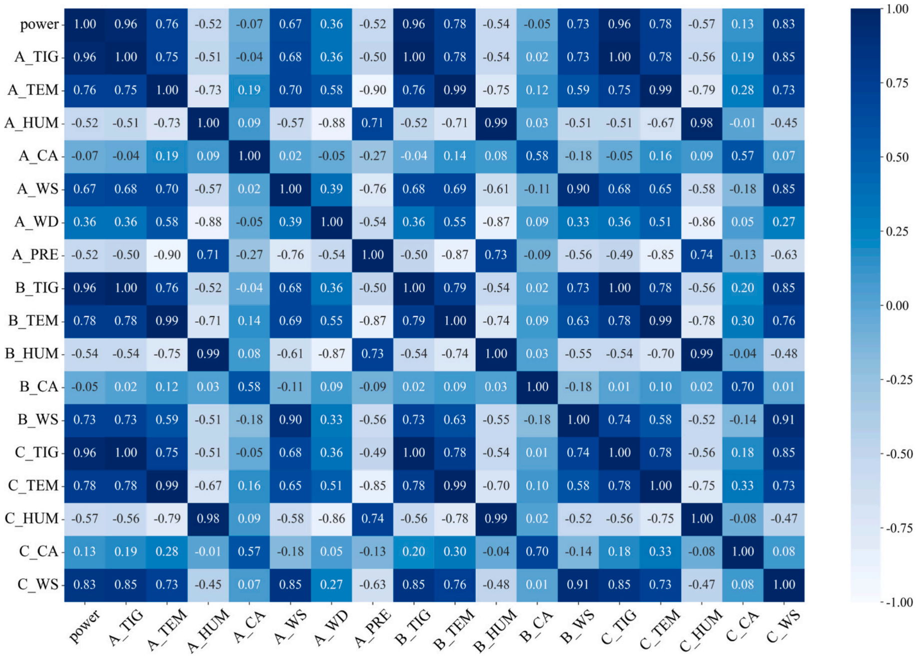


Fig. 8. Gradient map illustrating the Kendall correlation coefficient between meteorological characteristics and PV power.

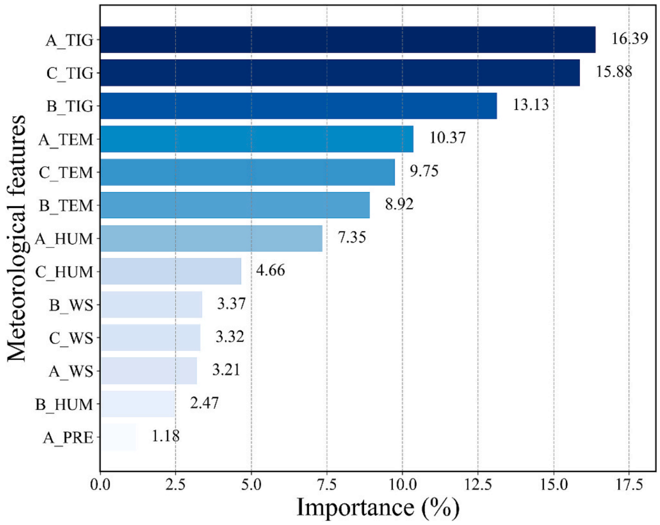


Fig. 9. The feature importance score calculated by the STIME method.

Table 4
Comparison of the effects of various feature extraction strategies on the model.

Extraction method	Neural network	Multi-source feature results	RMSE	MAE	R^2
STIME	STAGCN	A_TIG, A_TEM, A_HUM, B_TIG, B_TEM, C_TIG, C_TEM	11.80 %	6.96 %	0.8563
Pearson	STAGCN	A_TIG, A_TEM, B_TEM, B_TIG, C_TIG	12.34 %	7.05 %	0.8465
Kendall	STAGCN	A_TIG, A_PRE, A_TEM, B_TIG, B_TEM, C_TIG	16.24 %	8.85 %	0.7242
Spearman	STAGCN	A_TIG, A_TEM, B_TEM, B_TIG, B_HUM, C_TIG	12.65 %	7.72 %	0.6466
NMI	STAGCN	A_TIG, A_WS, A_TEM, B_TIG, B_TEM, C_TIG, C_HUM	14.95 %	7.13 %	0.7839

features is one of the problems that have been studied in the forecasting field. The traditional correlation analysis can only ensure that the selected features are highly correlated with the forecasting target sequence. Whether it can really affect the forecasting effect of the model needs further discussion.

The Kendall correlation coefficient assesses linear and nonlinear relationships within complex datasets while remaining independent of data distribution patterns [60]. This section employs the Kendall correlation coefficient to identify features exhibiting the strongest statistical associations with PV power generation among 17 meteorological characteristics, followed by application of the STIME technique to evaluate these features' contribution levels to the forecasting model.

As shown in Fig. 8, there are 13 meteorological features closely related to PV power, which are A_TIG, A_TEM, A_HUM, A_WS, A_PRE, B_TIG, B_TEM, B_HUM, B_WS, C_TIG, C_TEM, C_HUM, C_WS. There are more 0 values in the cloudiness feature samples, so the correlation coefficient of cloudiness in the three NWP sources is very low. The target PV power station is situated in a region abundant with Gobi deserts and grasslands, characterized by a mid-latitude, arid climate with scarce cloud formation. Additionally, the wind direction exhibits seasonal patterns and changes gradually, resulting in minimal correlation with the fluctuating PV power output. These feature importance scores have been normalized before using the STIME method. As can be seen in Fig. 9, in addition to the six features of C_HUM, B_WS, C_WS, A_WS, B_HUM and A_PRE, the importance of other features is higher than 5 %, which is marked with a darker blue in the figure.

Consequently, the selected meteorological characteristics include A_TIG, C_TIG, B_TIG, A_TEM, C_TEM, B_TEM, and A_HUM. These features demonstrate both strong correlations with PV power generation and enhanced model performance, thereby improving model interpretability. It merits attention that features not selected do not necessarily indicate irrelevance to PV output; rather, excluding features with minor contributions reduces model complexity and facilitates clearer interpretation of forecasting mechanisms.

To evaluate the effectiveness of this feature extraction technique, alternative approaches were examined: Pearson correlation coefficient,

Table 5
The influence of spatio-temporal dynamic graph in the model.

Neural network	Dynamic graph	RMSE	MAE	R^2
STAGCN	Yes	11.80 %	6.96 %	0.8563
STGCN	No	12.28 %	7.24 %	0.8344
GAT	No	13.24 %	7.33 %	0.7914
GCN	No	12.61 %	7.11 %	0.8412

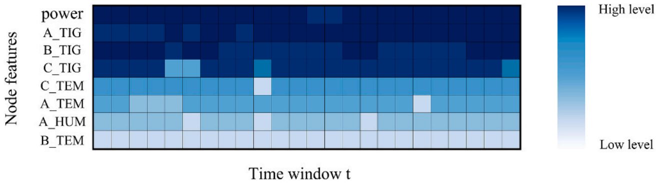


Fig. 10. Spatio-temporal transformation and fusion coefficients under different time frames.

Kendall correlation coefficient, Spearman correlation coefficient, and normalized mutual information. Through these extraction methods, the accuracy indicators are compared using the same model. Table 5 shows the feature results extracted by these methods and the accuracy index results of forecasting. It can be found from the table that all feature extractions contain the feature of irradiance, and atmospheric pressure and wind speed will cause large errors in forecasting. The RMSE of the method based on Kendall and NMI reached 16.24 % and 14.95 %, respectively. The Pearson correlation coefficient does not extract the key feature of A_HUM, resulting in an error 0.54 % higher than the STIME method. The RMSE, MAE and R^2 of feature extraction based on STIME have good performance.

3.4.3. Analysis of forecasting results of dynamic STAGCN model

In the domain of power forecasting, the task of short-term PV power forecasting is accomplished by inputting three-day multi-source NWP data and historical power data to output PV power data for the forthcoming day. Based on the test results outlined in section 3.4.2, seven meteorological features extracted from the multi-source NWP are combined with historical PV power data to form eight feature nodes, which serve as input signals for the STAGCN model. The training dataset, with a time step length M of approximately 47913, can be divided into $T = 248$ sliding overlapping frames, with each time frame having an overlap of $m = 192$ time steps. The input signals, with dimensions of $8 \times 248 \times 384$, are fed into the STAGCN model for training and forecasting, and the attention-based spatiotemporal feature transformation and fusion technique is employed to calculate the time and space attention coefficient matrices. In practice, conducting adjacency analysis on all dynamic graphs is challenging. Thus, representative coefficient matrices from clear-sky data are selected to demonstrate the model efficacy and undergo a joint interpretability test.

As illustrated in Fig. 10, the spatiotemporal transformation and fusion coefficients generated across different time frames during dynamic adjacency matrix construction are presented. Observations indicate that PV power and T_IG receive consistently higher attention weights during forecasting, whereas temperature and humidity exhibit relatively lower values. These attention weights dynamically adjust within time frames as temporal steps progress, demonstrating the model capacity to selectively adapt focus toward more critical features under varying spatiotemporal conditions for enhanced forecasting accuracy. Although direct ranking of node features by attention coefficients is infeasible, all features maintain substantial influence throughout different time frames. Analysis of evolving spatiotemporal transformation and fusion coefficients provides interpretability for identifying primary features affecting model accuracy. This mechanism, combined with the feature extraction module, collectively validates the methodological significance and interpretability of the proposed forecasting framework.

To determine whether spatiotemporal dynamic graphs genuinely enhance accuracy within graph networks, comparative experiments were conducted without error correction. Test accuracy outcomes are summarized in Table 4, with forecasting curves displayed in Fig. 11. The table reveals that incorporating spatiotemporal dynamic graphs effectively improves precision: the STAGCN model achieves RMSE reductions of 0.48 %, 1.44 %, and 0.81 % relative to STGCN, GAT, and GCN respectively, alongside MAE decreases of 0.28 %, 0.37 %, and 0.15 %, and R^2 improvements of 2.19 %, 6.49 %, and 1.51 %. Consequently, integrating spatiotemporal dynamic graphs in graph neural networks enables highlighting beneficial features across varying time frames during forecasting, thereby elevating forecasting accuracy.

Weather classification yields forecasting error sequences across distinct weather patterns. Fig. 12 displays sunny day forecasting outcomes and error sequences from the STAGCN model. Red zones exhibit predominantly positive errors while yellow areas show mainly negative deviations. The error distribution demonstrates higher magnitudes at daily extremes and lower values during midday periods, revealing discernible patterns. Consequently, Section 3.4.4 explores forecasting error components to refine forecasting outputs, constituting a valuable research endeavor.

3.4.4. Analysis of forecasting results considering error correction

To further investigate the underlying mechanisms of error generation, this study employs the ICEEMDAN-VMD twice decomposition and error correction strategy, using clear-sky weather type data as an illustrative example. Prior to decomposition, invalid information with zero errors during nighttime has been removed. Initially, the clear-sky error sequence is subjected to ICEEMDAN decomposition. When the decomposition reaches the seventh iteration, IMF becomes sufficiently smooth, indicating that further decomposition is unnecessary. This process yields seven components: IMF1, IMF2, IMF3, IMF4, IMF5, IMF6, and Residual1, as shown in Fig. 13(a). It can be observed that IMF1, generated by ICEEMDAN, is a high-frequency, strongly non-stationary component with numerous spikes and strong volatility. Directly forecasting this component would result in significant errors. Hence, the high-frequency, strongly non-stationary component IMF1 is further decomposed using VMD.

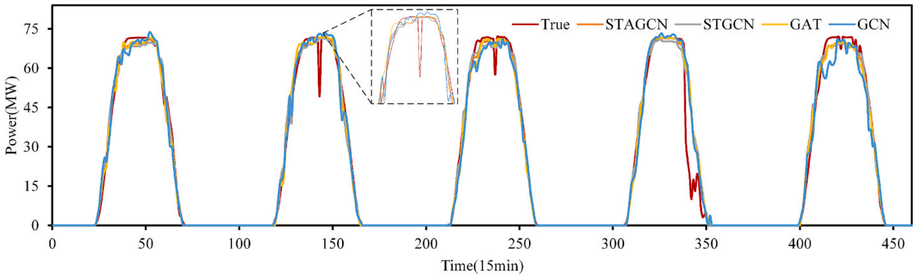


Fig. 11. Forecasting results of different graph network models.

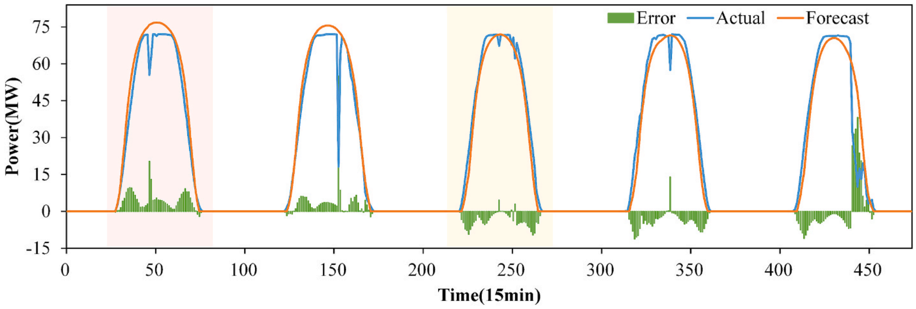
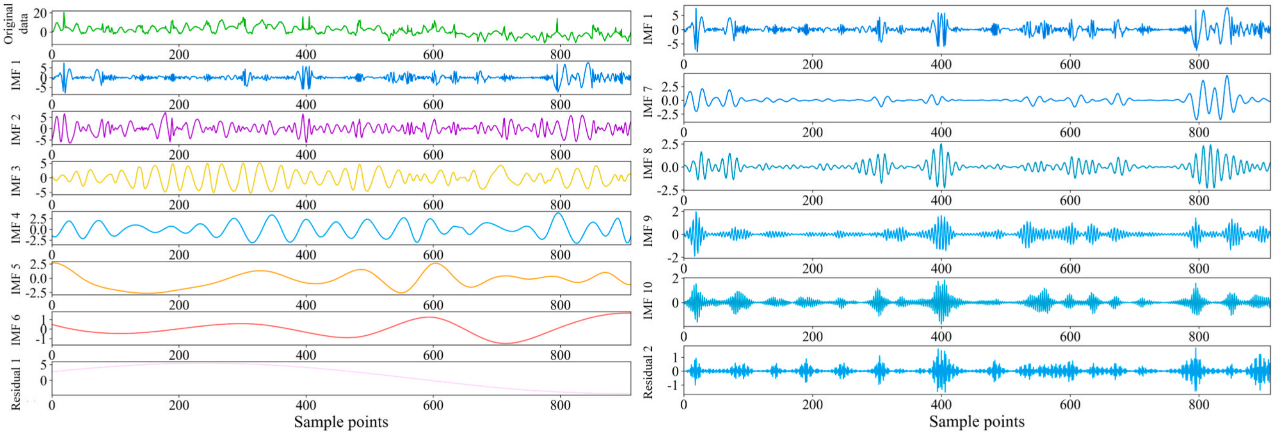


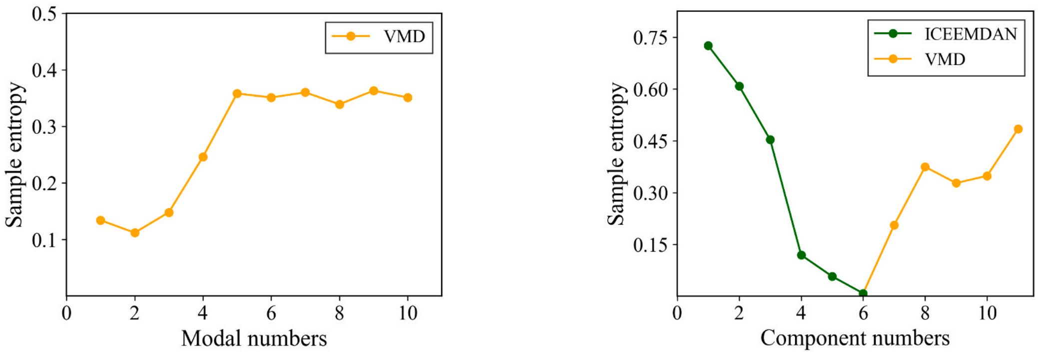
Fig. 12. STAGCN sunny forecasting results and error results.



(a) ICEEMDAN sub-component decomposition results

(b) VMD sub-component decomposition results

Fig. 13. Twice decomposition results of ICEEMDAN-VMD based on sample entropy.



(a) VMD mode number selection

(b) Sample entropy of ICEEMDAN-VMD twice decomposition

Fig. 14. The selection of mode number K and the results of twice decomposition sample entropy.

Determining the number of sub-modes in VMD decomposition requires preset parameters. Excessive sub-modes introduce extraneous noise, whereas insufficient quantities cause incomplete decomposition. To mitigate subjectivity in selecting decomposition numbers, the mean sample entropy value guides the VMD sub-mode determination, with outcomes illustrated in Fig. 14(a). At decomposition number $K = 5$, the mean sample entropy fluctuates near 0.35 without substantial increments. Consequently, $K = 5$ was selected, yielding decomposition results shown in Fig. 13(b) and producing five components: IMF7, IMF8, IMF9, IMF10, and Residual2.

Forecasting each modal component acquired through two-stage modal decomposition would significantly increase computational costs while overlooking correlations among the sub-components. Grouping correlated components for collective processing can shorten computation time and highlight the characteristics shared within each group. Therefore, we employ SE to assess the complexity of components derived from both decomposition stages. Sub-components from ICEEMDAN decomposition, excluding IMF1, are sequentially labeled components 1 to 6. Sub-components resulting from VMD decomposition of IMF1 are labeled components 7 to 11; the assessment outcome is visualized in Fig. 14(b).

Components 4 to 7 display SE values between 0 and 0.3. These are reconstructed into a new component representing the primary trend of the original error sequence, designated the trend component. Similarly,

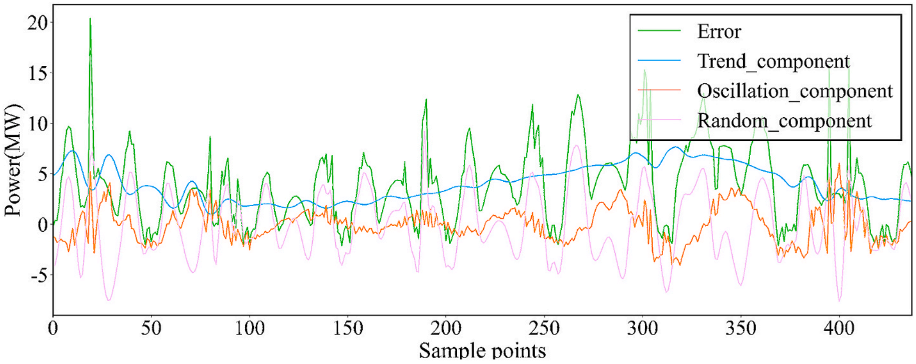


Fig. 15. Error sequence sub-component reconstruction waveform.

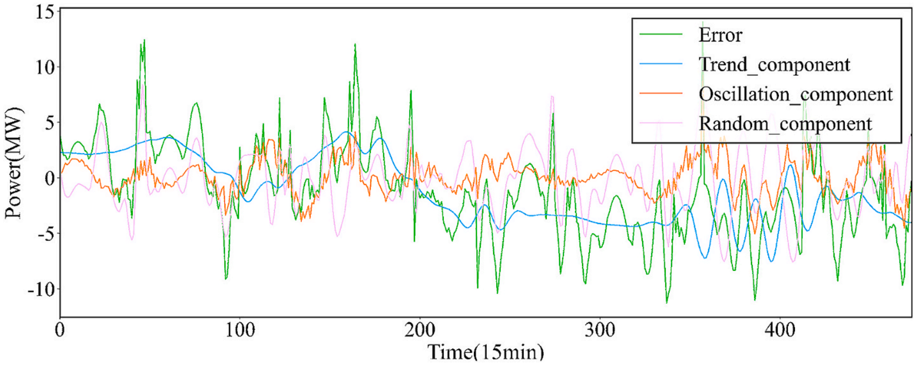


Fig. 16. The forecasting results of three error components.

components 3, 8, 9, 10, and 11 exhibit SE values between 0.3 and 0.6 and are reconstructed as the oscillatory component. Components 1 and 2, with SE values between 0.6 and 0.9, form the random component; the reconstructed waveforms are shown in Fig. 15. The trend component, containing less physical information and exhibiting smaller fluctuations, demonstrates strong predictability, leading to LSTM forecasting. The oscillatory component, holding more physical information and displaying a degree of periodicity, also lends itself to LSTM forecasting. Conversely, the random component exhibits intense fluctuations and lacks distinct patterns, making direct forecasting challenging; MKDE sampling forecasting is applied here. The resulting error forecasts appear in Fig. 16. After the error is decomposed and reorganized, the trend error describes the overall trend of the error over a period of time, and reveals the time influence of multiple meteorological features of NWP on the power forecasting error. The periodicity of the oscillation error illustrates the system error caused by the internal parameter setting of the model. The random error fluctuates violently, which characterizes the influence caused by the fluctuation and instability of power acquisition equipment and complex geographical and meteorological factors. Sampling forecasting is more suitable for dealing with such errors.

4. Conclusion

This work employs a short-term PV power forecasting approach integrating error decomposition correction with a spatiotemporal attention graph convolutional network. The forecasting performance undergoes rigorous evaluation across diverse weather conditions. Comparative analysis involves six benchmark forecasting methods, examining the presence or absence of error correction strategies and dynamic spatiotemporal graphs. Results demonstrate the presented model achieves superior performance across all evaluation metrics—RMSE, MAE, and R—compared to every benchmark model.

(1) The integration of nonlinear correlation analysis with the STIME method enables the selection and extraction of features that not only exhibit a high degree of correlation with PV power but also genuinely benefit the forecasting model. The approach of combining attention-based spatiotemporal feature transformation and fusion technique to assess feature consistency enhances the interpretability of the model.
(2) The incorporation of spatiotemporal attention mechanisms and dynamic adjacency matrices in the model constructs dynamic graph signal means, which effectively increase forecasting accuracy. The superior forecasting performance of STAGCN over models such as STGCN, GAT, and GCN validates this point. Considering the ICEEMDAN-VMD twice decomposition error correction strategy significantly reduces forecasting errors.
(3) This study presents a method incorporating secondary decomposition error correction, which clarifies the composition and generation mechanisms of errors and contributes significantly to enhancing power forecasting accuracy. Under cloudy and rainy conditions particularly, the approach captures temporal variations in power output and forecasting errors more effectively.

The methodology offers guidance for selecting features essential to model construction across varied forecasting scenarios. For power stations lacking sufficient historical data and facing highly variable weather conditions, this approach may prove unsuitable. Future research could explore forecasting and error correction strategies tailored to limited data samples and extreme weather events.

CRediT authorship contribution statement

Zhao Zhen: Writing – review & editing. **Yufei Yang:** Writing – original draft. **Fei Wang:** Supervision. **Nanpeng Yu:** Investigation. **Gang Huang:** Data curation. **Xiqiang Chang:** Validation. **Guoqing Li:**

Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This paper is supported by the National Key R&D Program of China (2024YFB4204800).

References

[1] X. Liu, H. Shi, D. Yang, X. Chen, X. Xia, Y. Xie, Reducing the impact of climate change on renewable energy systems through wind-solar blending: a worldwide study with CMIP6, Sol. Energy 290 (2025) 113365, https://doi.org/10.1016/j.solener.2025.113365.

[2] Z. Ge, Z. Xu, J. Li, J. Xu, J. Xie, F. Yang, Technical-economic evaluation of various photovoltaic tracking systems considering carbon emission trading, Sol. Energy 71 (2024) 112451, https://doi.org/10.1016/j.solener.2024.112451.

[3] L.D. Santos, T. Alskaif, G.C. Barroso, P.C.M.D. Carvalho, Photovoltaic power estimation and forecast models integrating physics and machine learning: a review on hybrid techniques, Sol. Energy 284 (2024) 113044, https://doi.org/10.1016/j.solener.2024.113044.

[4] I. López, N. Goitia-Zabaleta, A. Milo, J. Gómez-Cornejo, I. Aranzabal, H. Gaztañaga, E. Fernandez, European energy communities: characteristics, trends, business models and legal framework, Renew. Sustain. Energy Rev. 197 (2024) 114403, https://doi.org/10.1016/j.rser.2024.114403.

[5] P.K. Kesavan, U. Subramaniam, D.J. Almakhles, S. Selvam, Modelling and Coordinated Control of Grid Connected Photovoltaic, Wind Turbine Driven PMSG, and Energy Storage Device for a Hybrid DC/AC Microgrid, Protection and Control of Modern Power Systems 9 (1) (2024) 154–167, https://doi.org/10.23919/PCMP.2023.000272.

[6] International Renewable Energy Agency. Accessed: Mar. 26 2025. [Online]. Available, https://www.irena.org/News/pressreleases/2025/Mar/Record-Breaking-Annual-Growth-in-Renewable-Power-Capacity-ZH.

[7] M. Farhoumandi, S. Bahramirad, M. Shahidehpour, A. Alabdulwahab, Blockchain for Peer-to-Peer Energy Trading in Electric Vehicle Charging Stations With Constrained Power Distribution and Urban Transportation Networks, Energy Internet 2 (1) (2025) 28–44, https://doi.org/10.1049/ein2.12026.

[8] National Energy Administration. Accessed: Mar 17 2025. [Online]. Available, https://www.nea.gov.cn/20250318/78dbc76b24434bd9ae9db87c3e1f18eb/c.html.

[9] L. Liu, X. Hu, J. Chen, R. Wu, F. Chen, Embedded Scenario Clustering for Wind and Photovoltaic Power, and Load Based on Multi-Head Self-Attention, Protection and Control of, Modern Power Systems 9 (1) (2024) 122–132, https://doi.org/10.23919/PCMP.2023.000296.

[10] A. Morales-Acevedo, Fundamentals of solar cell physics revisited: Common pitfalls when reporting calculated and measured photocurrent density, open-circuit voltage, and efficiency of solar cells, Sol. Energy 262 (2023) 111774, https://doi.org/10.1016/j.solener.2023.05.051.

[11] D.S. Kumar, O. Gandhi, C.D. Rodríguez-Gallegos, D. Srinivasan, Review of power system impacts at high PV penetration Part II: potential solutions and the way forward, Sol. Energy 210 (2020) 202–221, https://doi.org/10.1016/j.solener.2020.08.047.

[12] Z. Liu, X. Chen, Y. Huang, X. Luo, S. Zhang, G. You, X. Qiang, Q. Kang, A novel bimodal feature fusion network-based deep learning model with intelligent fusion gate mechanism for short-term photovoltaic power point-interval forecasting, Energy 303 (2024) 131947, https://doi.org/10.1016/j.energy.2024.131947.

[13] Y. Li, J. Li, Y. Guo, Y. Wang, X. Yang, Y. Xu, Technology and Application of Multi-Energy System: An Engineering Study in China, Energy Internet 2 (2) (2025) 126–137, https://doi.org/10.1049/ein2.70002.

[14] D. Markovics, M.J. Mayer, Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction, Renew. Sustain. Energy Rev. 161 (2022) 112364, https://doi.org/10.1016/j.rser.2022.112364.

[15] H. Xu, Y. Zhang, Z. Zhen, F. Xu, F. Wang, Adaptive feature selection and GCN with optimal graph structure-based ultra-short-term wind farm cluster power forecasting method, IEEE Trans. Ind. Appl. 60 (1) (2024) 1804–1813, https://doi.org/10.1109/TIA.2023.3321863.

[16] L. Yang, X. Cui, W. Li, A method for predicting photovoltaic output power based on PCC-GRA-PCA meteorological elements dimensionality reduction method, Int. J. Green Energy 21 (10) (2024) 2327–2340, https://doi.org/10.1080/15435075.2024.2303357.

[17] N.Y. Hendrikx, K. Barhmi, L.R. Visser, T.A. de Bruin, M. Po, A.A. Salah, W.G.J.H.M. van Sark, All sky imaging-based short-term solar irradiance forecasting with Long Short-Term networks, Sol. Energy 272 (2024) 112463, https://doi.org/10.1016/j.solener.2024.112463.

[18] C. Brester, V. Kallio-Myers, A. Lindfors, M. Kolehmainen, H. Niska, Evaluating neural network models in site-specific solar PV forecasting using numerical weather prediction data and weather observations, Renew. Energy 207 (2023) 266–274, https://doi.org/10.1016/j.renene.2023.02.130.

[19] A. Catalina, C.M. Alaíz, J.R. Dorronsoro, Combining numerical weather predictions and satellite data for PV energy nowcasting, IEEE Trans Sustain Energy 11 (3) (2020) 1930–1937, https://doi.org/10.1109/TSTE.2019.2946621.

[20] G. Ruan, X. Chen, E.G. Lim, L. Fang, Q. Su, L. Jiang, Y. Du, On the use of sky images for intra-hour solar forecasting benchmarking: Comparison of indirect and direct approaches, Sol. Energy 276 (2024) 112649, https://doi.org/10.1016/j.solener.2024.112649.

[21] C. Xu, G. Chen, Interpretable transformer-based model for probabilistic short-term forecasting of residential net load, Int. J. Electr. Power Energy Syst. 155 (2024) 109515, https://doi.org/10.1016/j.ijepes.2023.109515.

[22] G. Mitrentsis, H. Lens, An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting, Appl. Energy 309 (2022) 118473, https://doi.org/10.1016/j.apenergy.2021.118473.

[23] M.M. Islam, S.S. Shuvo, M.J.A. Shohan, M.O. Faruque, Forecasting of PV plant output using interpretable temporal fusion transformer model, in: North American Power Symposium, NAPS, 2023, https://doi.org/10.1109/NAPS58826.2023.10318698.

[24] Y. Wang, W. Fu, J. Wang, Z. Zhen, F. Wang, Ultra-short-term distributed PV power forecasting for virtual power plant considering data-scarce scenarios, Appl. Energy 373 (2024) 123890, https://doi.org/10.1016/j.apenergy.2024.123890.

[25] B. Hu, H. Shao, C. Shao, W. Tang, A satellite-based novel method to forecast short-term (10 min-4 h) solar radiation by combining satellite-based cloud transmittance forecast and physical clear-sky radiation model, Sol. Energy 290 (2025) 113376, https://doi.org/10.1016/j.solener.2025.113376.

[26] V.A.M. Lopez, G. van Urk, P.J.F. Doodkorte, M. Zeman, O. Isabella, H. Ziar, Using sky-classification to improve the short-term prediction of irradiance with sky images and convolutional neural networks, Sol. Energy 269 (2024) 112320, https://doi.org/10.1016/j.solener.2024.112320.

[27] L. Cheng, H. Zang, T. Ding, Z. Wei, G. Sun, Multi-meteorological-factor-based graph modeling for photovoltaic power forecasting, IEEE Trans Sustain Energy 12 (3) (2021) 1593–1603, https://doi.org/10.1109/TSTE.2021.3057521.

[28] A. Ali, Y. Zhu, M. Zakarya, Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction, Neural Netw. 145 (2022) 233–247, https://doi.org/10.1016/j.neunet.2021.10.021.

[29] F. Li, J. Feng, H. Yan, G. Jin, F. Yang, F. Sun, D. Jin, Y. Li, Dynamic graph convolutional recurrent network for traffic prediction: benchmark and solution, ACM Trans. Knowl. Discovery Data 17 (1) (2023) 9, https://doi.org/10.1145/3532611.

[30] X. Tang, L. Zheng, J. Chen, Z. Chen, Y. Qin, Learning-based hierarchical cooperative eco-driving with traffic flow prediction for hybrid electric vehicles, Energy Convers. Manage. 321 (2024) 119000, https://doi.org/10.1016/j.enconman.2024.119000.

[31] B. Yu, H. Yin, Z. Zhu, J. Lang, Spatio-temporal hraph convolutional networks: a deep learning framework for traffic forecasting, in: 27th International Joint Conference on Artificial Intelligence, 2018, pp. 3634–3640.

[32] R. Wang, Z. Hu, X. Song, W. Li, Trajectory distribution aware graph convolutional network for trajectory prediction considering spatio-temporal interactions and scene information, IEEE Trans. Knowl. Data Eng. 36 (8) (2024) 4304–4316, https://doi.org/10.1109/TKDE.2023.3329676.

[33] J. Cao, C. Liu, C.-L. Chen, N. Qu, Y. Xi, Y. Dong, R. Feng, A short-term load forecasting method for integrated community energy system based on STGCN, Electr. Power Syst. Res. 232 (2024) 110265, https://doi.org/10.1016/j.epsr.2024.110265.

[34] W. Lai, Z. Zhen, F. Wang, W. Fu, J. Wang, X. Zhang, H. Ren, Sub-region division based short-term regional distributed PV power forecasting method considering spatio-temporal correlations, Energy 288 (2024) 129716, https://doi.org/10.1016/j.energy.2023.129716.

[35] Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. In: International conference on learning representations, 2018. https://doi.org/10.48550/arXiv.1710.10903.

[36] D. Wu, K. Peng, S. Wang, V.C.M. Leung, Spatial-temporal graph attention gated recurrent transformer network for traffic flow forecasting, IEEE Internet Things J. 11 (8) (2024) 14267–14281, https://doi.org/10.1109/JIOT.2023.3340182.

[37] Q. Zhang, C. Li, F. Su, Y. Li, Spatiotemporal residual graph attention network for traffic flow forecasting, IEEE Internet Things J. 10 (13) (2023) 11518–11532, https://doi.org/10.1109/JIOT.2023.3243122.

[38] J. Liu, T. Li, Multi-step power forecasting for regional photovoltaic plants based on ITDE-GAT model, Energy 293 (2024) 130468, https://doi.org/10.1016/j.energy.2024.130468.

[39] S. Shan, Z. Ding, K. Zhang, H. Wei, C. Li, Q. Zhao, ACGL-TR: a deep learning model for spatio-temporal short-term irradiance forecast, Energy Convers. Manage. 284 (2023) 116970, https://doi.org/10.1016/j.enconman.2023.116970.

[40] S.T. Thanh, H.D. Dinh, G.N.H. Minh, T.N. Trong, T.N. Duc, Spatial-temporal graph hybrid neural network for PV power forecast, in: 8th International Conference on Green Energy and Applications, 2024, pp. 317–322, 10.1109/ICGEA60749.2024.10560564.

[41] X. Liu, X. Xie, Q. Fang, F. Xu, C. Huang, Z. Zhen, F. Wang, Spatial-temporal attention mechanism and graph convolutional networks based distributed PV ultra-short-term power forecasting. International Conference on Energy Technologies for Future Grids, ETFG, 2024, 10.1109/ETFG55873.2023.10408590.

[42] H. Yin, M.S. Yang, H. Lan, Y.Y. Hong, D. Guo, F. Jin, A hybrid graph attention network based method for interval prediction of shipboard solar irradiation, Energy 298 (2024) 131131, https://doi.org/10.1016/j.energy.2024.131131.

[43] Y. Deng, B. Wang, Z. Lu, A hybrid model based on data preprocessing strategy and error correction system for wind speed forecasting, Energy Convers. Manage. 212 (2020) 112779, https://doi.org/10.1016/j.enconman.2020.112779.

[44] J. Tong, L. Xie, S. Fang, W. Yang, K. Zhang, Hourly solar irradiance forecasting based on encoder-decoder model using series decomposition and dynamic error compensation, Energy Convers. Manage. 270 (2022) 116049, https://doi.org/10.1016/j.enconman.2022.116049.

[45] J. Li, S. Zhang, Z. Yang, A wind power forecasting method based on optimized decomposition prediction and error correction, Electr. Power Syst. Res. 208 (2022) 107886, https://doi.org/10.1016/j.epsr.2022.107886.

[46] H. Liu, R. Yang, Z. Duan, Wind speed forecasting using a new multi-factor fusion and multi-resolution ensemble model with real-time decomposition and adaptive error correction, Energy Convers. Manage. 217 (2020) 112995, https://doi.org/10.1016/j.enconman.2020.112995.

[47] Z. Ma, H. Chen, J. Wang, X. Yang, R. Yan, J. Jia, W. Xu, Application of hybrid model based on double decomposition, error correction and deep learning in short-term wind speed prediction, Energy Convers. Manage. 205 (2020) 112345, https://doi.org/10.1016/j.enconman.2019.112345.

[48] G. Li, X. Wei, H. Yang, Decomposition integration and error correction method for photovoltaic power forecasting, Measurement 208 (2023) 112462, https://doi.org/10.1016/j.measurement.2023.112462.

[49] R. Zhang, G. Li, S. Bu, G. Kuang, W. He, Y. Zhu, S. Aziz, A hybrid deep learning model with error correction for photovoltaic power forecasting, Front. Energy Res. 10 (2022) 948308, https://doi.org/10.3389/fenrg.2022.948308.

[50] L. Wang, X. Li, Y. Bai, Short-term wind speed prediction using an extreme learning machine model with error correction, Energy Convers. Manage. 162 (2018) 239–250, https://doi.org/10.1016/j.enconman.2018.02.015.

[51] D. Yang, W. Wang, J.M. Bright, C. Voyant, G. Notton, G. Zhang, C. Lyu, Sol. Energy 236 (2022) 743–755, https://doi.org/10.1016/j.solener.2022.03.004.

[52] M.J. Mayer, D. Yang, B. Szintai, Comparing global and regional downscaled NWP models for irradiance and photovoltaic power forecasting: ECMWF versus AROME, Appl. Energy 352 (2023) 121958, https://doi.org/10.1016/j.apenergy.2023.121958.

[53] L. Tan, R. Kang, J. Xia, Y. Wang, Application of multi-source data fusion on intelligent prediction of photovoltaic power, Sol. Energy 277 (2024) 112706, https://doi.org/10.1016/j.solener.2024.112706.

[54] M. Marrocu, L. Massidda, Estimating the value of ECMWF EPS for photovoltaic power forecasting, Sol. Energy 279 (2024) 112801, https://doi.org/10.1016/j.solener.2024.112801.

[55] M.J. Mayer, D. Yang, Pairing ensemble numerical weather prediction with ensemble physical model chain for probabilistic photovoltaic power forecasting, Renew. Sustain. Energy Rev. 175 (2023) 113171, https://doi.org/10.1016/j.rser.2023.113171.

[56] G. Jin, Y. Liang, Y. Fang, Z. Shao, J. Huang, J. Zhang, Y. Zheng, Spatio-temporal graph neural networks for predictive learning in urban computing: a survey, IEEE Trans. Knowl. Data Eng. 36 (10) (2024) 5388–5408, https://doi.org/10.1109/TKDE.2023.3333824.

[57] M. Bahoura, Efficient FPGA-based architecture of the overlap-add method for short-time fourier analysis/synthesis, Electronics 8 (12) (2019) 1533, https://doi.org/10.3390/electronics8121533.

[58] Q. Liu, Y. Li, H. Jiang, Y. Chen, J. Zhang, Short-term photovoltaic power forecasting based on multiple mode decomposition and parallel bidirectional long short term combined with convolutional neural networks, Energy 286 (2024) 129580, https://doi.org/10.1016/j.energy.2023.129580.

[59] D. Kim, D. Kwon, L. Park, J. Kim, S. Cho, Multiscale LSTM-based deep learning for very-short-term photovoltaic power generation forecasting in smart city energy management, IEEE Syst. J. 15 (1) (2021) 346–354, https://doi.org/10.1109/JSYST.2020.3007184.

[60] M.T. Puth, M. Neuhäuser, G.D. Ruxton, Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits, Anim. Behav. 102 (2015) 77–84, https://doi.org/10.1016/j.anbehav.2015.01.010.