POLITECNICO DI TORINO
Repository ISTITUZIONALE

Short-Term Prediction of Wind Farm Cluster Power Based on Bidirectional Interactive Fusion of Homogeneous and Heterogeneous Information

Original
Short-Term Prediction of Wind Farm Cluster Power Based on Bidirectional Interactive Fusion of Homogeneous and Heterogeneous Information / Zhang, Z., Wang, B., Yang, M., Wang, Z., Zhang, W., Huang, T., Su, X.. - In: PROTECTION AND CONTROL OF MODERN POWER SYSTEMS. - ISSN 2367-0983. - ELETTRONICO. - 11:2(2026), pp. 107-125. [10.23919/PCMP.2025.000058]

Availability:
This version is available at: 11583/3017421 since: 2026-11-03T10:28:52Z

Publisher:
Springer

Published
DOI:10.23919/PCMP.2025.000058

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

15 November 2026

Short-term Prediction of Wind Farm Cluster Power Based on Bidirectional Interactive Fusion of Homogeneous and Heterogeneous Information

Zhe Zhang, Bo Wang, Mao Yang, Zhao Wang, Wei Zhang, Tao Huang, *Senior Member*, *IEEE*, and Xin Su

***Abstract*—Recognizing the intricate spatiotemporal correlation (STC) among wind farms (WFs) is critical to achieving better predictions for wind farm clusters (WFCs). To describe the STC accurately, this paper employs the wind angular field method to transform the wind series of WFs into different 2-D feature maps, and then construct homogeneous and heterogeneous STC graphs from these maps. The graphs are dynamically updated at the frequency of data update to capture time-varying STC among WFs. Finally, a dynamic graph attention network, designed according to the STC graphs, is established for WFC prediction. Through the above process, dynamic and accurate descriptions of STC are realized in WFC prediction. From the case study of a large-scale WFC with a capacity over 5800 MW in Northeast China, the proposed method reduces the root mean square error of the prediction in the next 24 hours by 2.67%.**

***Index Terms*—Spatiotemporal correlation, wind farm cluster, wind angular field, homogeneous and heterogeneous, dynamic graph attention network.**

Received: April 10, 2025
Accepted: December 25, 2025
Published Online: March 1, 2026

Zhe Zhang is with the Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education, Northeast Electric Power University, Jilin 132012, China; and also with the National Key Laboratory of Renewable Energy Grid-Integration, China Electric Power Research Institute, Beijing 100192, China (e-mail: zhangzhe0607@outlook.com).

Bo Wang and Zhao Wang are with the National Key Laboratory of Renewable Energy Grid-Integration, China Electric Power Research Institute, Beijing 100192, China (e-mail: wangbo@epri.sgcc.com.cn; wangzhao@epri.sgcc.com.cn).

Mao Yang (corresponding author), Wei Zhang, and Xin Su are with the Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education, Northeast Electric Power University, Jilin 132012, China (e-mail: 962905724@qq.com; 1918119493@qq.com; suxin126@126.com).

Tao Huang is with the Department of Energy, Politecnico di Torino, Torino 10129, Italy (e-mail: tao.huang@polito.it).

DOI: 10.23919/PCMP.2025.000058

I. Introduction

With the continuous growth of global energy demand and the increasing severity of environmental problems, wind energy has become an increasingly important role in the adjustment of the energy structure. By 2023, the total installed capacity of wind power worldwide had reached 117 GW, playing an important role in expanding the share of renewable energy [1]. However, the inherent randomness, volatility, and uncertainty of wind energy pose increasingly serious challenges to the safe and stable operation of power systems [2]. Accurate short-term wind power prediction (1–3 days) is crucial to promote wind power consumption and reduce power supply risks [3], [4].

With the large-scale development of wind power, the distribution pattern of wind farms (WFs) is gradually shifting from a scattered distribution of small-scale low permeability to a cluster distribution of large-scale high permeability [5]. The development of large-scale weather systems often causes significant fluctuations in the output of large-scale wind farm clusters (WFCs). At the same time, the time-space mismatch and amplitude deviation of the numerical weather prediction (NWP) model for weather system prediction cause power prediction errors. Compared with scattered WFs, clustered WFs exhibit complex spatial and temporal correlations. Power prediction models that rely solely on data from individual WFs struggle to capture the impact of global meteorological prediction error on wind power, rendering traditional single-station power prediction methods inadequate. Effectively utilizing complex, multidimensional NWP data to construct the spatiotemporal correlation (STC) of large-scale WFC remains a major challenge for cluster prediction [6], [7].

Previous studies have extensively investigated the integration of spatiotemporal information in large-scale WFCs. Among the most common approaches are the whole method, accumulation method, and statistical upscaling (SU) prediction. The whole method [8] directly predicts the total output power of the WFC by means of feature averaging. After modeling each WF in the cluster by the accumulation method [9], WFC prediction is realized by summing the predicted power of multiple WFs. SU [10] maps the total power of the region by selecting

the benchmark WF to represent the characteristics of all WFs in the region. The single-to-single mapping mode of the global method and the cumulative method ignore the spatiotemporal information between WFs. Although SU can represent the collection of spatiotemporal information in the region by selecting the form of benchmark WF, it is difficult to form a global understanding of the spatiotemporal evolution of regional weather systems only on sparsely distributed single-point meteorological information. The multi-input multi-output (MIMO) method integrates multiple WFs into a joint WFC that can interact internally. By fully incorporating prior spatiotemporal dependency information, it achieves accurate global prediction performance and has therefore become one of the mainstream prediction methods [11]–[13]. However, convolutional neural networks (CNN) [14] usually need preset a fixed size, and can only consider part of the WF information in the receptive field, thereby limiting the captured global information and long-term dependence. Multi-layer perceptron (MLP) flattens high-dimensional input features into vectors, which induces the loss of some feature information. Attention mechanisms (AM) [15] can associate multi-targets and long temporal distances by calculating attention coefficients, and has become a popular method for constructing STC in MIMO.

With the rise of graph theory, researchers have increasingly sought to describe the transmission of spatiotemporal information between WFs by constructing diverse graph structures. Combining the advantages of the above modeling methods, graph neural networks (GNN) can make full use of the prior spatiotemporal dependence information to construct the correlation between nodes to form a large-scale WFC topology, and calculate the influence weights between nodes to construct the nonlinear relationship between input and output features of nodes [16]–[18]. In the construction of graph structure, most studies only consider the sequence correlation between adjacent nodes [19]. Reference [20] uses weighted Euclidean distance to establish the adjacency relationship between wind turbines, while reference [21] introduces mutual information (MI) to construct adjacency matrix to improve the accuracy of WF power prediction. There are also studies using cosine similarity to construct the correlation between WFs [22], thought it only describes the connection relationship between WFs through fixed sequence correlation and lacks the full utilization of spatiotemporal evolution information. In the calculation of neighbor weights, graph convolutional networks (GCN) combine the structural information of the graph and the degree information of the nodes to obtain the weight coefficient with a normalized adjacency matrix [23]. However, this approach has limitations in handling the dynamic changes of edge weights and the distribution of learning weights. Graph attention networks (GAT) cannot dynamically reflect the interaction of actual WFs through a predefined graph structure [24], while the update of graph structure and weight is of great significance for predicting WF under new weather characteristics.

With the gradual expansion of WFC scale [25], the evolution of regional meteorological conditions becomes complicated. The STC between WFs under different meteorological conditions, different output characteristics and scheduling characteristics is complex and difficult to construct [26]. There are a variety of WFs affected by different meteorological conditions in large-scale WFCs. The different output characteristics and the different mapping trends of NWP and power cannot be classified as the same type of node attributes. At the same time, the edge relationships affected by the interaction between different types of nodes also have a variety of heterogeneous attributes, and the edge relationships between all nodes cannot be set to the same attribute. In graph theory, a graph with all nodes sharing the same attribute set and all edges sharing the same type of relationship is defined as a homogeneous graph, which can construct the internal spatiotemporal relationship of a small-scale WFC under a single meteorological condition and output characteristics. The graph containing a variety of different types of nodes and edges is defined as a heterogeneous graph, which can construct a large-scale WFC internal spatiotemporal relationship including a variety of meteorological conditions and output characteristics. However, the most widely used homogeneous graph network in the field of wind power prediction cannot appropriately characterize the heterogeneous spatiotemporal information within a large-scale WFC. In the homogeneous graph, there is only one type of nodes and one type of edge relationship between nodes, making it suitable for the construction of small-scale WFC internal relations with similar output characteristics. In the heterogeneous graph, the edge relationship between nodes is diverse, and such graph is suitable for the construction of the relationship between WFs with different output characteristics.

In summary, the existing short-term power prediction methods for large-scale WFCs have the following issues.

1) In the construction of graph structure of a large-scale WFC, most studies weaken the spatiotemporal evolution relationship between adjacent nodes, and tend to use some shallow correlations to describe the relationship between adjacent nodes. Thus, it is difficult to capture the complex prior spatiotemporal dependence information within a WFC.

2) There are many types of WFs in a large-scale WFC, and there are also many interaction relationships between different types of WFs as heterogeneous nodes. Most of the existing prediction methods only consider the STC between nodes of the same type, and lack the integration and utilization of the STC between heterogeneous nodes.

3) The existing research on the heterogeneity of WFCs has no clear definition, and there is no suitable system to construct the heterogeneous network of WFC power.

In order to fill the above research gaps, by comprehensively considering the STC of large-scale WFCs, a short-term WFC power prediction method based on bidirectional interactive fusion of homogeneous and heterogeneous information constructed by STC is proposed. The main contributions are as follows.

1) A multi-feature Gramian angle field (GAF) feature transformation method is proposed to construct wind angular field (WAF). By incorporating wind speed, wind direction, and time series gradient of the target day, the scale-invariant feature transform (SIFT) correlation recognition method from image processing is used to determine whether STC exists between WFs, and the graph structure is constructed accordingly.

2) A large-scale WFC is divided into clusters, and a number of regional WFCs with different output characteristics and scheduling characteristics are obtained. The information of the spatiotemporal interaction within the WFC is refined into the homogeneous information within the regional WFCs and the heterogeneous information between the regional WFCs. Based on these two types of information, the prior spatiotemporal dependence information within the large-scale WFC is constructed.

3) A bidirectional interactive network of homogeneous and heterogeneous information based on dynamic graph attention network (DGAT) is constructed. The predefined graph structure is updated synchronously according to the update frequency of meteorological information to dynamically reflect the mutual influence of actual WFs. The two-module interactive network can extract the local correlation features of homogeneous information and the wide-area correlation features of heterogeneous information.

The structure of the rest of the paper is as follows. Section II introduces the overall process of the proposed method, while Section III introduces the principle of the proposed method in detail. Section IV demonstrates the effectiveness of the proposed method from various perspectives, and Section V concludes this study.

II. The Proposed Framework

Figure 1 shows the overall research framework.

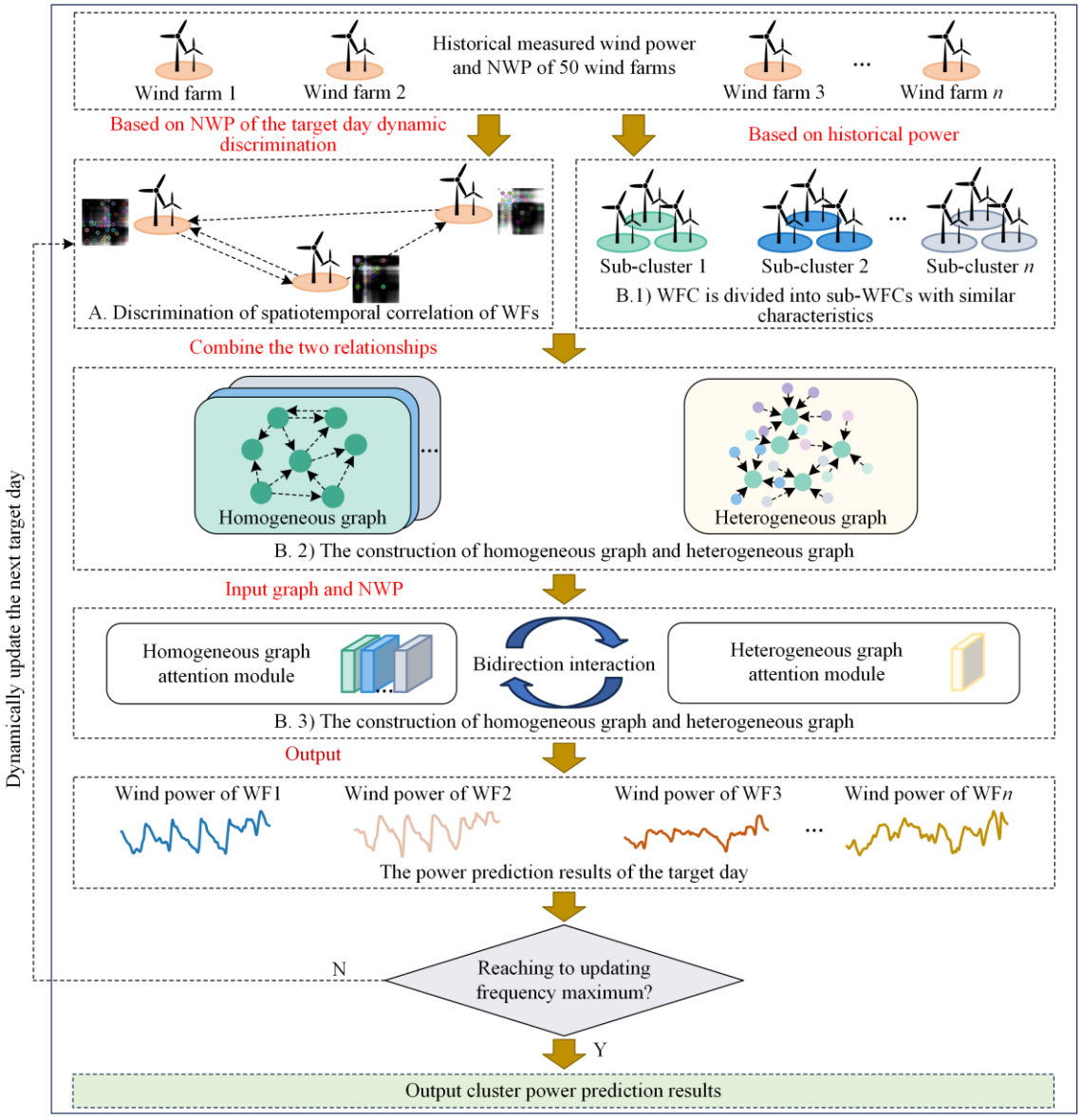


Fig. 1. Overall research framework.

As shown in Fig. 1, four main parts are included: 1) First, based on the historical wind power sequence, the unsupervised clustering evaluation index is used to determine the optimal grouping and generate multiple sub-clusters with different output; 2) Second, the WAF matrix of WFs is constructed to determine whether STC exists between WFs; 3) Then, based on different STC attributes, the homogeneous graph and heterogeneous graph are dynamically constructed according to the data update frequency to represent the construction of the STC within the large-scale WFC; 4) Finally, using the process of bidirectional interaction information between homogeneous modules and heterogeneous modules, the information within and between sub-clusters is extracted and integrated into node characteristics to predict the future power of nodes, thereby achieving large-scale WFC short-term power prediction.

III. Methodology

A. WFs Relation Recognition Method Based on WAF

The STC of WFs is typically determined using MI or the Pearson correlation coefficient (PCC) to measure the relationship between historical wind power sequence and NWP. However, due to the limited update frequency of historical power data, the NWP sequence may deviate from actual wind speed and direction. A feasible scheme is to match the STC of WFs based on the similarity of abrupt changes in the NWP during wind events. By considering the similarity of wind speed and direction series, correlated WFs under the same wind conditions can be identified more accurately by observing the trend of correlation coefficient over time.

Although the cross-correlation function in the time series analysis method can construct the change trend of the correlation coefficient between the above observation sequence points, the time series gradient relationship is inevitably lost during the construction process. This study proposes a STC-based WF relation recognition method. The WAF is obtained by improving the GAF, while the temporal correlation and temporal dependence of NWP are introduced into the STC characterization of WFs, thereby determining whether STC exists between WFs, as shown in Fig. 2.

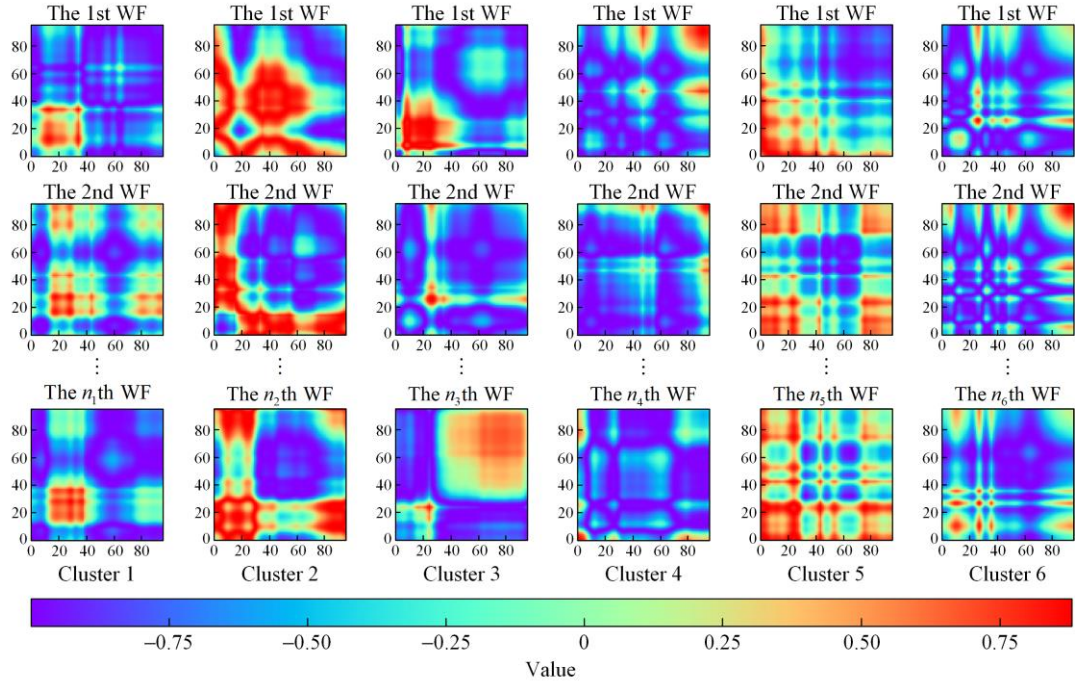


Fig. 2. The WAF of a part of WF in each cluster in the first predicting daily.

1) Construction of WAF Matrix

The time series value is processed as the radius, and the sum of the normalized inverse cosine value of the wind speed series and wind direction value is used as the angle to generate the polar coordinates. Then, the coordinates in a scheduling period are taken via a point-wise inner product to obtain the WAF that characterizes the WF scheduling period.

$$\begin{cases} \phi_{i,\mathrm{s}} = \arccos x_{i,\mathrm{s}} \\ \phi_{i,\mathrm{d}} = x_{i,\mathrm{d}} \times \dfrac{\pi}{180^\circ} \end{cases} \tag{1}$$

$$r_i = \frac{t_i}{N} \tag{2}$$

$$\cos\left(\left(\phi_{i,\mathrm{s}} + \phi_{i,\mathrm{d}}\right) + \left(\phi_{j,\mathrm{s}} + \phi_{j,\mathrm{d}}\right)\right) = \cos\left(\phi_i + \phi_j\right) \tag{3}$$

$$\boldsymbol{A}_{\mathrm{WAF}} = \begin{bmatrix} r_1^2\cos(\phi_1+\phi_1) & r_1r_2\cos(\phi_1+\phi_2)\cdots r_1r_n\cos(\phi_1+\phi_n) \\ r_2r_1\cos(\phi_2+\phi_1) & r_2^2\cos(\phi_2+\phi_2)\cdots r_2r_n\cos(\phi_2+\phi_n) \\ \vdots & \cdots \quad \vdots \\ r_nr_1\cos(\phi_n+\phi_1) & r_nr_2\cos(\phi_n+\phi_2)\cdots r_n^2\cos(\phi_n+\phi_n) \end{bmatrix} \tag{4}$$

where $\boldsymbol{A}_{\mathrm{WAF}}$ is the WAF matrix; the input is the normalized wind speed sequence $x_{i,\mathrm{s}}$ and wind direction sequence $x_{i,\mathrm{d}}$ of the target day for the *i*th WF; $\phi_{i,\mathrm{s}}$ is the inverse cosine value of the wind speed sequence $x_{i,\mathrm{s}}$; $\phi_{i,\mathrm{d}}$ is to convert the wind direction sequence of the *i*th WF into radian; ϕ_i and ϕ_j are the feature values at time steps t_i and t_j, respectively; and r_i is the radius in the polar coordinate system obtained by time series transformation.

The construction of the WAF matrix incorporates specific design choices aimed at ensuring its computational robustness and the effective representation of wind characteristics. Input wind speed data is normalized prior to its use in calculating the angular component, while wind direction angles are directly converted to radians due to their inherent range. The periodicity of the cosine function naturally addresses potential overflow in angular sums. The radius r_i, derived from the ratio of the time point's sequence number t_i to the total number of points N as t_i/N is thus normalized to the (0, 1) interval, representing the relative temporal position within the window. Consequently, the magnitude of the WAF matrix elements is effectively bounded, ensuring numerical stability. The one-day time window for WAF matrix construction aligns with the NWP update cycle. It ensures that the NWP data used to characterize wind conditions for a given day is the most recently available forecast, thereby maximizing its relevance to that day's actual conditions and power output. The data is provided at a 15 min temporal resolution, which defines the minimum time granularity for the analyses conducted in this study.

While wind power output is influenced by multiple factors (e.g., air density, operational status), the WAF construction in this study focuses on wind speed and direction. This is because they are the most direct drivers of the spatiotemporal wind flow patterns that govern inter-farm correlations, ensuring a robust and physically interpretable graph structure. The integration of additional variables to refine this structure is considered a valuable direction for future work.

2) STC Discrimination Between WFs

To identify the state mutation points in the image matrix, this study introduces the SIFT method to transform the WAF matrix into scale space to identify continuous wind process mutation points and time series correlation points. The specific process is shown in Fig. 3. First, the WAF is processed by Gaussian blur to obtain the key points of the WAF matrix. The Gaussian fuzzy representation of the matrix is as follows:

$$\boldsymbol{L}(x,y,\sigma) = \boldsymbol{G}(x,y,\sigma) * \boldsymbol{I}(x,y) \tag{5}$$

$$\boldsymbol{G}(x,y,\sigma) = \frac{1}{2\pi\sigma^2} \mathrm{e}^{-\frac{(x-m/2)^2+(y-n/2)^2}{2\sigma^2}} \tag{6}$$

where $\boldsymbol{L}(x,y,\sigma)$ is the matrix transformed into scale space; $\boldsymbol{G}(x,y,\sigma)$ is a Gaussian kernel of scale σ; $\boldsymbol{I}(x,y)$ is a primitive matrix; x and y are the spatial coordinates within the matrix; and m and n represent the dimensions of the Gaussian kernel template.

Then, the key points are detected by Gaussian difference calculation between different scales as:

$$\boldsymbol{D}(x,y,\sigma) = \left[\boldsymbol{G}(x,y,k\sigma) - \boldsymbol{G}(x,y,\sigma)\right] * \boldsymbol{I}(x,y) \tag{7}$$

where k is the scale change factor. This process can effectively detect stable feature points in the WAF matrix. After determining the key points, a feature descriptor based on the gradient direction and amplitude distribution around the key points is generated. The gradient direction and amplitude are calculated as:

$$\boldsymbol{m}(x,y) = \sqrt{\left[\boldsymbol{L}(x+1,y) - \boldsymbol{L}(x-1,y)\right]^2 + \left[\boldsymbol{L}(x,y+1) - \boldsymbol{L}(x,y-1)\right]^2} \tag{8}$$

$$\boldsymbol{\theta}(x,y) = \tan^{-1}\left[\frac{\boldsymbol{L}(x,y+1) - \boldsymbol{L}(x,y-1)}{\boldsymbol{L}(x+1,y) - \boldsymbol{L}(x-1,y)}\right] \tag{9}$$

where $\boldsymbol{m}(x,y)$ is the gradient amplitude; and $\boldsymbol{\theta}(x,y)$ is the gradient direction. The gradient direction and amplitude jointly construct the descriptor vector $\boldsymbol{f}$ of the key points, expressed as $\boldsymbol{f} = (\boldsymbol{m}, \boldsymbol{\theta})$ It can not only represent the local characteristics of the mutation points in the WAF matrix, but also show the gradient direction of the mutation points. When judging the relationship between the WAF matrices, the Euclidean distance between the mutation point descriptor vectors is calculated. Using David Lowe's distance ratio test method, for the descriptor of each feature point, it identifies the two closest matching points in the WAF to be matched, and calculates the nearest neighbor distance d_1 and the next nearest neighbor distance d_2. If the ratio of the nearest neighbor distance d_1 to the next nearest neighbor distance d_2 is less than the threshold, it is considered there STC exists between the two WFs.

$$d(\boldsymbol{f}_1,\boldsymbol{f}_2)=\sqrt{\sum_{n}^{i=1}\left(\boldsymbol{f}_{1,i}-\boldsymbol{f}_{2,i}\right)^2} \tag{10}$$

$$\lambda_{\mathrm{STC}_{i,j}}=\begin{cases}1, & \text{if } \dfrac{d_1}{d_2}\leqslant v_{\mathrm{Threshold}}\\ 0, & \text{else}\end{cases} \tag{11}$$

where $d(\boldsymbol{f}_1,\boldsymbol{f}_2)$ represents the Euclidean distance between the descriptor vectors $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$; $\lambda_{\mathrm{STC}_{i,j}}$ is the binary indicator representing the existence of STC between the ith and jth WFs, $\lambda_{\mathrm{STC}_{i,j}}=1$ indicates existence, otherwise, $\lambda_{\mathrm{STC}_{i,j}}=0$; $\boldsymbol{f}_{1,i}$ and $\boldsymbol{f}_{2,i}$ are the ith elements of the two descriptors, respectively. In general, the empirical value $v_{\mathrm{Threshold}}=0.75$ is a commonly used threshold in David Lowe's distance ratio test method. If the ratio of the nearest neighbor distance to the next nearest neighbor distance is less than 0.75, it is considered that STC exists between WFi and WFj.

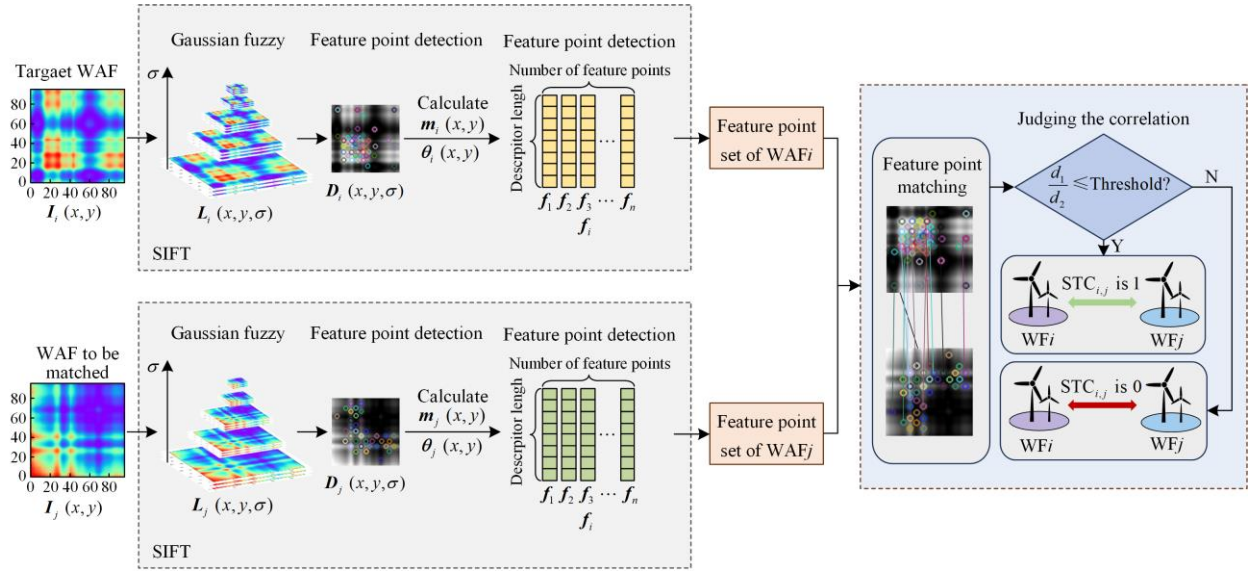


Fig. 3. The correlation discrimination process between WAFs.

B. Homogeneous and Heterogeneous Bidirectional Interactive Networks Based on DGAT Network

1) Construction of Homogeneous Graph and Heterogeneous Graph

Traditional large-scale WFC power prediction usually divides a WFC into several sub-clusters with similar output features according to its historical output characteristics, and then models and predicts each sub-cluster. In order to fully construct the STC topology of a large-scale WFC, this study continuously refines the different node attributes and relationship attributes of WFC after dividing the large-scale WFC. Two types of STC for large-scale WFC and sub-clusters are defined. All WFs in the same sub-cluster are nodes of the same output type, and the STC between the same nodes is Homogeneous STC. The WFs of different sub-clusters are nodes with different output types, and the STC between different types of nodes is Heterogeneous STC. Homogeneous STC and Heterogeneous STC interaction diagrams are shown in Fig. 4. The large-scale WFC topology diagrams based on these two associations are homogeneous and heterogeneous, respectively. Specific definitions and processes are described as follows.

Definition1: homogeneous graph. The homogeneous graph of WFC is expressed as $G_{\mathrm{Ho},i}=(V,E)$, where V and E are the sets of nodes and edges, respectively. The homogeneous graph constructed in parallel by multiple sub-clusters is $G_{\mathrm{Ho}}=\{G_{\mathrm{Ho},1},G_{\mathrm{Ho},2},\cdots,G_{\mathrm{Ho},C}\}$, where C is the total number of sub-clusters. The V interior of $G_{\mathrm{Ho},i}$ can be described as $\{V_1,V_2,\cdots,V_i\}$, where $i\in[1,C_i]$, i.e., i nodes within the same sub-cluster. Similarly, $E=\{E_{ij}\}_{i\in C_n,j\in C_n}$, where E_{ij} denotes the directed edge relationship between the field stations i and j in the same sub-cluster.

Definition2: heterogeneous graph. The heterogeneous graph of WFC is expressed as $G_{\mathrm{He}}=(\Phi_C,V,E)$, where Φ_C represents the meta-path; while V and E are the sets of nodes and edges, respectively. The nodes in V are described as $\{V_1,\cdots,V_c,\cdots,V_C\}$, where V_c is the set of nodes inside the C sub-cluster; $i\in[1,N_c]$ is the ith node of V_c, including N_c nodes, and the sub-cluster contains n nodes. Similarly, $E=\{E_{ij}\}_{i\in C_i,j\in C_j}$, where

E_{ij} denotes the directed edge relation between two different sub-clusters internal stations i and j. Different types of nodes are used to construct different edge relationships between WFs with bipartite graphs.

The attributes of nodes and edges in the homogeneous graph are relatively uniform, and the homogeneous GNN is used for feature extraction and relationship modeling. In contrast, the attributes of nodes and edges in heterogeneous graphs have different output characteristics and error characteristics, and the processing of heterogeneous graphs requires heterogeneous GNNs to consider the characteristics of each type of nodes and edges and the interaction between them.

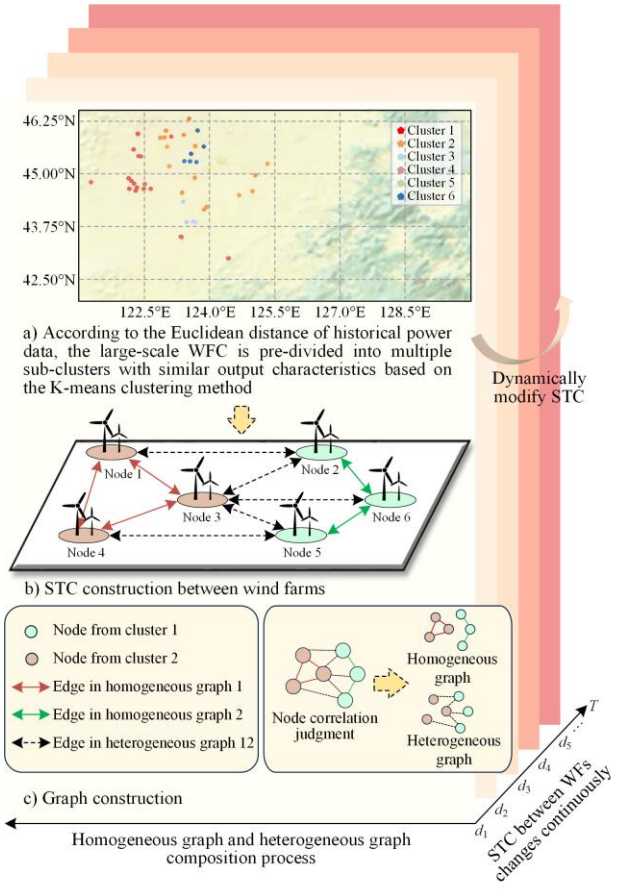


Fig. 4. The construction process of homogeneous graph and heterogeneous graph.

2) Construction of the Adjacency Matrix Considering STC, Homogeneous Graph and Heterogeneous Graph

The adjacency matrix is used to describe the relationship between any two connected nodes. It should be noted that if the result of graph relation recognition is $\lambda_{\mathrm{STC}_{i,j}}=1$ between WFi and WFj, then there is an edge between the two nodes, expressed as $\lambda_{\mathrm{Adjw}_{i,j}}=1$. Otherwise, $\lambda_{\mathrm{Adjw}_{i,j}}=0$ and it indicates that there is no connection between them. This ensures the sparsity of the adjacency matrix and helps to reduce the interference of unnecessary noise.

$$\lambda_{\mathrm{Adjw}_{i,j}}=\lambda_{\mathrm{STC}_{i,j}},\quad i=1,2,\cdots,n \text{ and } j=1,2,\cdots,n-1 \tag{12}$$

For the proposed homogeneous graph and heterogeneous graph, "Adj" determines the spatial connection relationship of nodes in multiple modes, which is the key to the homogeneous and heterogeneous graph attention network (HHGAT) learning STC. At the same time, the dynamics of graph structures require more attention, rather than being highly dependent on predefined structures, especially for the highly intermittent nature of wind power generation. Since the spatiotemporal dependence between different sites evolves over time, to enable the constructed graph to dynamically capture this time-varying relationship, this study updates the adjacency matrix by the actual update frequency (daily update) of the NWP to adapt to dynamics and practical engineering applications. Finally, the formulation of the adjacency matrix is given in the following, where $\boldsymbol{u}_i$ represents cluster i, i.e., the homogeneous graph of the ith sub-cluster; $\boldsymbol{u}_{i\text{-}j}$ is the heterogeneous graph with edge attribute i–j between the ith and jth sub-clusters.

$$\boldsymbol{A}_{\mathrm{Ho}}=\begin{bmatrix}\boldsymbol{u}_1&0&0&0&0&0\\0&\boldsymbol{u}_2&0&0&0&0\\0&0&\boldsymbol{u}_3&0&0&0\\0&0&0&\boldsymbol{u}_4&0&0\\0&0&0&0&\boldsymbol{u}_5&0\\0&0&0&0&0&\boldsymbol{u}_6\end{bmatrix} \tag{13}$$

$$\boldsymbol{A}_{\mathrm{He}}=\begin{bmatrix}E&\boldsymbol{u}_{1\text{-}2}&\boldsymbol{u}_{1\text{-}3}&\boldsymbol{u}_{1\text{-}4}&\boldsymbol{u}_{1\text{-}5}&\boldsymbol{u}_{1\text{-}6}\\\boldsymbol{u}_{2\text{-}1}&E&\boldsymbol{u}_{2\text{-}3}&\boldsymbol{u}_{2\text{-}4}&\boldsymbol{u}_{2\text{-}5}&\boldsymbol{u}_{2\text{-}6}\\\boldsymbol{u}_{3\text{-}1}&\boldsymbol{u}_{3\text{-}2}&E&\boldsymbol{u}_{3\text{-}4}&\boldsymbol{u}_{3\text{-}5}&\boldsymbol{u}_{3\text{-}6}\\\boldsymbol{u}_{4\text{-}1}&\boldsymbol{u}_{4\text{-}2}&\boldsymbol{u}_{4\text{-}3}&E&\boldsymbol{u}_{4\text{-}5}&\boldsymbol{u}_{4\text{-}6}\\\boldsymbol{u}_{5\text{-}1}&\boldsymbol{u}_{5\text{-}2}&\boldsymbol{u}_{5\text{-}3}&\boldsymbol{u}_{5\text{-}4}&E&\boldsymbol{u}_{5\text{-}6}\\\boldsymbol{u}_{6\text{-}1}&\boldsymbol{u}_{6\text{-}2}&\boldsymbol{u}_{6\text{-}3}&\boldsymbol{u}_{6\text{-}4}&\boldsymbol{u}_{6\text{-}5}&E\end{bmatrix} \tag{14}$$

3) The Construction of Homogeneous and Heterogeneous Modules

To fully extract the interaction between nodes with STC. In this study, a DGAT model is used to construct respective homogeneous modules and heterogeneous modules to fit the change process of weights between WFs at different times. The network framework is shown in Fig. 5.

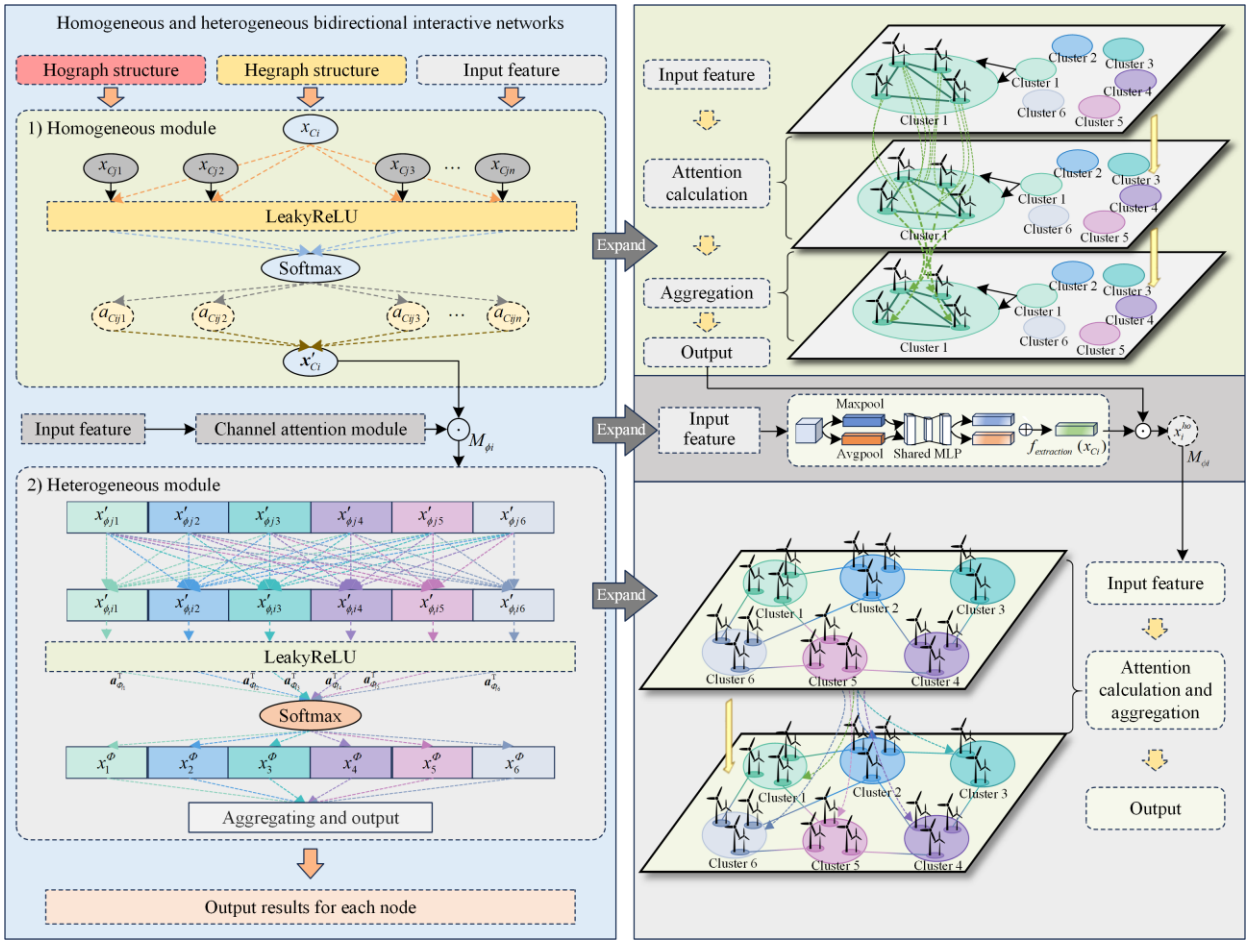


Fig. 5. Homogeneous module and heterogeneous module frameworks.

Homogeneous modules extract local correlation features. In order to construct the STC within the sub-cluster, DGAT is used to calculate the local correlation characteristics of each sub-cluster based on the constructed homogeneous graph. The specific process can be described as:

$$\text{LeakyReLU}(x)=\begin{cases}x, & x>0\\ -\gamma x, & x\leqslant 0\end{cases} \tag{15}$$

$$e(x_{Ci},x_{Cj})=\boldsymbol{a}^{\mathrm{T}}\left[\boldsymbol{W}x_{Ci}\,\|\,\boldsymbol{W}x_{Cj}\right],\ j\in N_{Ci} \tag{16}$$

where $e(x_{Ci},x_{Cj})$ is expressed as the nonlinearly transformed attention coefficient of node j (the neighbor node of node i) to node i in the sub-cluster C; LeakyReLU$(\cdot)$ is a nonlinear transformation function; $\boldsymbol{a}^{\mathrm{T}}[\cdot]$ is the attention coefficient; $\|$ is the feature splicing operation; while x_{Ci} and x_{Cj} are the input feature of node i and node j, respectively; $\boldsymbol{W}$ is a shared parameter matrix of linear transformation; N_{Ci} is the total number of neighbor nodes of node i; and γ is the positive slope of the negative region of the function.

The attention weight α_{Cij} is obtained by normalizing the obtained attention coefficient, and the output vector $\boldsymbol{x}'_{Ci}$ is obtained by weighted summation, as:

$$\alpha_{Cij}=\text{softmax}_j(e_{Cij})=\frac{\exp\left(\text{LeakyReLU}\left(e\left(x_{Ci},x_{Cj}\right)\right)\right)}{\sum_{n=1}^{N_i}\exp\left(\text{LeakyReLU}\left(e\left(x_{Ci},x_{Cjn}\right)\right)\right)} \tag{17}$$

$$\boldsymbol{x}'_{Ci}=\Big\|_{k=1}^{K}\sigma\left(\sum_{j\in N_{Ci}}\alpha_{Cij}\boldsymbol{W}_{\text{ho}}x_{Cj}\right) \tag{18}$$

where $\text{softmax}_j(\cdot)$ is the function of normalizing the attention coefficient e_{Cij} of the neighbor node j; x_{Cjn} represents the input feature vector of the neighbor node j within the sub-cluster C; $\boldsymbol{W}_{\text{ho}}$ shared parameter matrix for homogeneous features; $\sigma(\cdot)$ is a sigmoid activation function; and the number of heads of the multi-head attention mechanism is K. At the same time, the spatial and temporal features of the WFC are extracted uniformly by channel attention module (CAM), which is

represent by $f_{\mathrm{CAM}}(\cdot)$. The important node features of the original WFs are obtained by average pooling and maximum pooling of the features of WFs. After the MLP and sigmoid activation operation, the feature vector set of $f_{\mathrm{CAM}}(x_{Ci})$ representing the original WFs is obtained, which is fused with the spatiotemporal vector $\boldsymbol{x}'_{Ci}$ extracted by DGAT to obtain the spatiotemporal feature vector of each WF through the homogeneous module, as:

$$f_{\mathrm{CAM}}(x_{Ci})=\sigma\left(f_{\mathrm{MLP}}\left(\left[\mathrm{AvgPool}\left(x_{Ci}\right);\mathrm{MaxPool}\left(x_{Ci}\right)\right]\right)\right)= \sigma\left(f_{\mathrm{MLP}}\left(\left[x_{Ci}^{\mathrm{avg}};x_{Ci}^{\max}\right]\right)\right) \tag{19}$$

$$x_i^{\mathrm{ho}}=f_{\mathrm{CAM}}\left(x_{Ci}\right)\odot\boldsymbol{x}'_{Ci} \tag{20}$$

where $\odot$ is the Hadamard product operator; $f_{\mathrm{MLP}}(\cdot)$ is the non-linear transformation function of the multi-layer perceptron; x_{Ci}^{avg} and $x_{Ci}^{\max}$ are the feature vectors obtained by average pooling and maximum pooling, respectively.

Heterogeneous module extracts global correlation features: HDGAT is used to construct the STC within the WFC based on the constructed heterogeneous graph. The calculation process of HDGAT's attention mechanism is composed of node-level attention mechanism and type-level attention mechanism.

1) The Calculation Process of Node-level Attention

After processing different sub-cluster WFs into the same feature space, the self-attention mechanism is used to learn weights for different types of nodes. Given a node pair (i, j) and a meta-path Φ connecting them, the node-level attention is calculated as follows:

$$x'_{\phi i}=\boldsymbol{M}_{\phi_i}x_i^{\mathrm{ho}} \tag{21}$$

$$e_{ij}^{\Phi_i}\left(x'_{\phi i},x'_{\phi j};\boldsymbol{\Phi}_i\right)=\boldsymbol{a}_{\Phi_i}^{\mathrm{T}}\left[\boldsymbol{W}x'_{\phi i}\,\|\,\boldsymbol{W}x'_{\phi j}\right] \tag{22}$$

where $\boldsymbol{M}_{\phi_i}$ is a transformation matrix that processes the ith sub-cluster into the same feature space; $\boldsymbol{\Phi}_i$ is the ith meta-path; e_{ij}^{Φ} is the asymmetric node attention coefficient; and using softmax(·) to normalize, the neighbor weight parameter α_{ij}^{Φ} based on the meta-path is obtained as:

$$\alpha_{ij}^{\Phi_i}=\mathrm{softmax}_j\left(e_{ij}^{\Phi_i}\right)=\frac{\exp\left(\mathrm{LeakyReLU}\left(e_{ij}^{\Phi_i}\right)\right)}{\sum_{n\in C^{\Phi_i}}\exp\left(\mathrm{LeakyReLU}\left(e_{in}^{\Phi_i}\right)\right)} \tag{23}$$

$$x_i^{\Phi_i}=\Big\|_{k=1}^{K}\sigma\left(\sum_{j\in\mathcal{N}^{\Phi_i}}\alpha_{ij}^{\Phi_i}\boldsymbol{W}_{\mathrm{he}}x'_{\phi_i j}\right) \tag{24}$$

where $\alpha_{ij}^{\Phi_i}$ is a node-level attention coefficient; and $x_i^{\Phi_i}$ is the output feature by node i through meta-path $\boldsymbol{\Phi}_i$; and $\boldsymbol{W}_{\mathrm{he}}$ is a shared parameter matrix for heterogeneous features. Node-level attention is used to capture the importance of WFs with different output characteristics and reduce the weight of noise nodes.

2) The Calculation Process of Type-level Attention

Type-level attention aggregates the attention vector $\left\{x_1^{\Phi},x_2^{\Phi},\cdots,x_C^{\Phi}\right\}$ of group C nodes obtained based on multiple meta-paths in the previous step as input, using nonlinear functions to transform the node embedding of different output feature sub-clusters. Then these edges are aggregated according to the weights to obtain the final representation x_{pre}.

$$w_{\Phi_i}=\frac{1}{|C|}\sum_{i\in C}\boldsymbol{q}^{\mathrm{T}}\tanh\left(\boldsymbol{M}x_i^{\Phi_i}+\boldsymbol{b}\right) \tag{25}$$

$$\lambda_{\Phi_i}=\frac{\exp(w_{\Phi_i})}{\sum_{i=1}^{P}\exp(w_{\Phi_i})} \tag{26}$$

$$x_{\mathrm{pre}}=\sum_{i=1}^{P}\lambda_{\Phi_i}x_{\Phi_i} \tag{27}$$

where $\boldsymbol{M}$ is the weight matrix; $\boldsymbol{b}$ is the biased vector; and $\boldsymbol{q}$ is the type-level attention vector. Type-level attention learns the weights of adjacent sub-clusters of different output feature types, and weights the feature vectors. In the training process, both homogeneous attention vectors and heterogeneous attention vectors are learnable through gradient back propagation. So far, STC modeling has integrated the spatiotemporal features within the entire large-scale WFC, representing the future state of all nodes in the WFC.

To provide a more intuitive understanding of how node-level attention aggregates information from different meta-paths, and how the resulting representations are further fused through type-level attention, a conceptual illustration is presented in Fig. 6.

Node-level attention depicts how a central cluster interacts with nodes in other different types of clusters via various meta-paths. For each meta-path $\boldsymbol{\Phi}_{ij}$, the node-level attention mechanism learns attention weights for the target node over its neighbors guided by this meta-path, and aggregates their information to generate a meta-path-specific embedding for the target node. Thus, the output of this stage is a set of meta-path-specific embeddings. Type-level attention layer takes the set of meta-path-specific embeddings from the node-level attention stage as input, learns importance weights for these different meta-path-specific embeddings, and performs a weighted fusion of them. This

fusion process allows the model to integrate information from diverse semantic contexts, ultimately producing updated and richer representations for the relevant clusters. This two-stage process clearly demonstrates how the model distinguishes and fuses information from different sets of meta-paths.

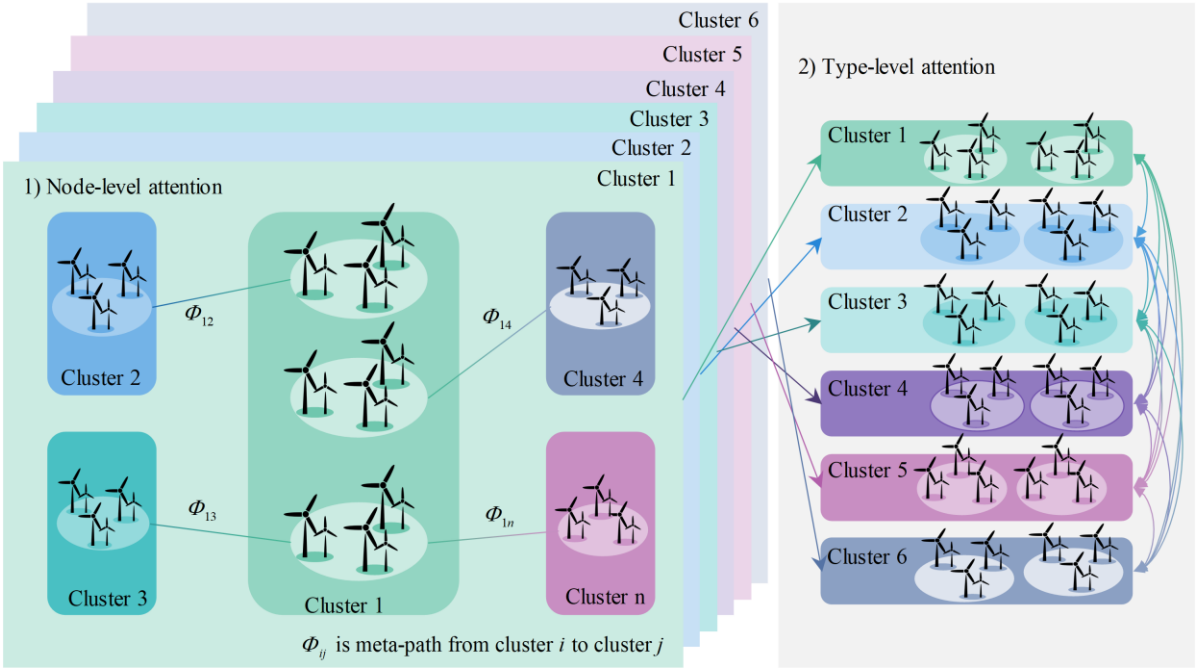


Fig. 6. Conceptual illustration of the two-stage attention mechanism for meta-path based information fusion in heterogeneous graphs.

IV. Application and Case Study

A. Datasets and Evaluation Index Selection

1) Datasets Introduction

To evaluate the performance of WAF-HHGAT under different scenarios, this study utilizes wind power data from 50 WFs in a Chinese province collected from January to August 2021 to conduct short-term WFC power prediction. The total installed capacity of this WFC is 5867.91 MW, with a time resolution of 15 min for both power and meteorological data. Meanwhile, 60% of the power and NWP sequence data in the dataset are used to train the model, 20% are used to validate the model parameters, and the remaining 20% are used for testing. The NWP sequence data include weather variables such as wind speed, wind direction, and other relevant weather elements. K-means is used to cluster these 50 WFs in the WFC according to the power of the training set. For the three commonly used unsupervised clustering indicators: Silhouette score, Davies Bouldin score, and Calinski Harabasz score [27], the normalized clustering results are shown in Fig. 7. From the results, the optimal clustering number is 6.

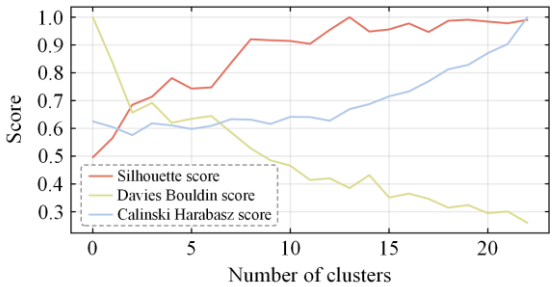


Fig.7. The clustering results of the three clustering indexes after clustering 50 WFs.

2) Evaluation Metrics

To evaluate the prediction performance of different models, this study adopts the following evaluation indicators: root mean square error (RMSE), mean absolute error (MAE), and R-squared (R^2) given as:

$$X_{\mathrm{RMSE}}=\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{\mathrm{p}}-y_{\mathrm{t}}\right)^2}}{p_{\mathrm{Cap}}} \tag{28}$$

$$X_{\mathrm{MAE}}=\frac{\frac{1}{n}\sum_{i=1}^{n}\left|y_{\mathrm{p}}-y_{\mathrm{t}}\right|}{p_{\mathrm{Cap}}} \tag{29}$$

$$X_{R^2}=1-\frac{\sum_{i=1}^{n}\left(y_p-y_t\right)^2}{\sum_{i=1}^{n}\left(y_t-\bar{y}_t\right)^2} \tag{30}$$

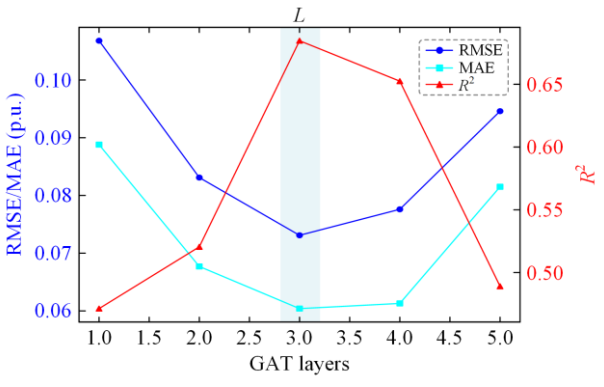
where y_p is the predicted value; y_t is the true value; $\bar{y}_t$ is the mean value of the true value; n is the number of predicted time points; and p_{Cap} is the installed capacity of the WFC.

B. Experimental Details

The code is implemented in PyTorch, and the model is trained for 100 epochs with a learning rate of 10^{-4} and a batch size of 64. All datasets are trained on NVIDIA GeForce RTX3070.

In this study, several classical and advanced methods are used to compare the performance of WAF-HHGAT. In the linear regression method, least absolute shrinkage and selection operator (LASSO) [28] are used as the baseline model. In the non-GNN regression methods, advanced algorithms that can extract the spatial and temporal characteristics of WFs including: Transformer, temporal convolutional network-long short-term memory (TCN-LSTM) [29], and convolutional block attention module-long short-term memory (CBAM-LSTM) [30], are used for comparison. In the GNN regression method, GCN is used as the comparison algorithm.

To ascertain the optimal hyperparameter settings for the WAF-HHDGAT model, a sensitivity analysis is performed, are the results are presented in Fig. 8. The analysis varies the number of GAT layers (L), attention heads (H), MLP channel dimensions (d_{MLP}), and the graph update frequency to observe their impact on prediction accuracy (f). It is found that 3 GAT layers achieve an optimal RMSE of 0.0731. Similarly, employing 3 attention heads result in the best RMSE of 0.0592. For MLP channel dimensions, the configuration [128, 256] yields the lowest RMSE at 0.0577. Finally, a graph update frequency of 96 is identified as optimal, providing an RMSE of 0.0577.



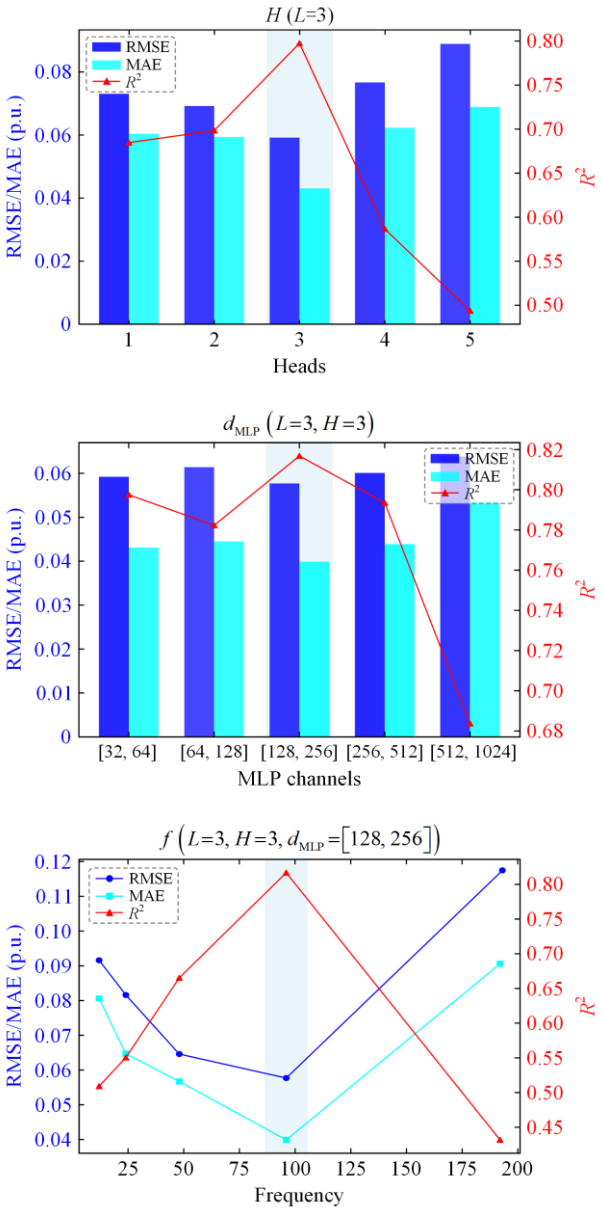


Fig. 8. Sensitivity analysis of hyperparameters for the WAF-HHDGAT model.

This optimal frequency aligns with the typical update cycle of NWP data, which often covers 96 time points. Updating the graph structure less frequently than 96 points may prevent the model from capturing the continuous characteristic changes present in the NWP. Conversely, updating more frequently than 96 points could introduce noise or become constrained by the inherent prediction accuracy limitations of the NWP itself over extended horizons. Consequently, these empirically determined optimal values: 3 GAT layers, 3 attention heads, MLP channels of [128, 256], and an update frequency of 96, are adopted for all subsequent experiments to ensure robust model performance. The hyperparameters used in the final models are shown in Table I.

TABLE I
HYPERPARAMETER CONFIGURATION

	Function	Structure	Configurations
Hyperparameter configuration of the WAF-HH structure model	WAF-HHGAT	Δ WAF Δ homogeneous GAT network Δ heterogeneous GAT network Δ MLP	# Graph update frequency: 96 (/15 min) # Layers of GAT: 3 # Number of heads for GAT:3 # Number of homogeneous GAT network: 6 # Number of homogeneous GAT network: 1 # Output channels of MLP [128, 256]
	WAF-HHGCN	Δ WAF Δ Homogeneous GCN network Δ Heterogeneous GCNT network Δ MLP	# Graph update frequency: 96 (/15 min) # Layers of GCN: 3 # Number of homogeneous GCN network: 6 # Number of homogeneous GCN network: 1 # Output channels of MLP [128, 256]
Hyperparameter configuration of the graph network regression model	G-GAT	Δ The adjacency matrix is constructed by geographic position. Δ MLP	# Graph update frequency: 96(15 min); # Layers of GAT: 2 # Number of heads for GAT: 3 # Output channels of MLP [128, 256]
Hyperparameter configuration of other advanced wind power forecasting methods	Transformer	Δ Multi-head attention Δ Encoder-decoder	# Number of encoder layers: 4 # Number of decoder layers: 4 # Model dimension: 256 # Number of attention heads: 8 # Feed-forward dimension: 1024 # Dropout rate: 0.1
	CBAM-LSTM	Δ Convolutional block attention module Δ LSTM	# Kernel size of Conv2d: 1×7 # reduction ratio: 16 # Layers of LSTM: 2 # Hidden units of LSTM: 64 # Dropout: 0.5 # Active functions of fully connected network: ReLU
	TCN-LSTM	Δ Temporal convolutional networks Δ LSTM	# Channels of TCN: [32,32,32] # Layers of LSTM: 2 # Hidden units of LSTM: 64 # Dropout: 0.5 # Active functions of fully connected network: ReLU

C. Ablation Experiment

In order to verify the validity of the model structure of WAF-HHDGAT constructed in this study, this method is compared with the following methods.

1) The proposed method WAF-HHDGAT.

2) WAF-HoDGAT (homogeneous dynamic graph attention network). In this proposed method, the heterogeneous module is removed, leaving only the homogeneous module to predict each node of the WF.

3) WAF-HeDGAT (heterogeneous dynamic graph attention network). In this proposed method, the homogeneous module is removed, leaving only the heterogeneous module to predict each node of the WF.

4) HHGAT. Instead of using WAF to construct graph structure, Pearson correlation coefficient is used to measure the correlation between WFs.

5) WAF-HHGAT. In this proposed method, the graph structure is no longer dynamically changed, but instead, it uses a predefined graph structure.

The final prediction results are shown in Fig. 9. As seen, when any structure is removed, the prediction accuracy of the proposed method decreases in the short-term prediction task. The statistical results of ablation experiment errors are shown in Table II.

Compared with WAF-HHDGAT, WAF-HoDGAT without heterogeneous module results in a 2% increase in RMSE, a 1.47% increase in MAE, and a 14.87% reduction in R^2. It can be concluded that the heterogeneous STC between clusters established in the heterogeneous modules is effective. WFs between different feature clusters learn the same or contradictory trends of different types of edge relationships in heterogeneous modules. Compared with the proposed method, WAF-HeDGAT without the homogeneous module leads to a 0.26% increase in RMSE, a 0.28% increase in MAE, and a 1.69% decrease in R^2. Compared with the MIMO model, the accurate construction of homogeneous STC can greatly improve the accuracy of wind power prediction. The edge relationship of the homogeneous module ensures that the spatiotemporal relationship between WFs within similar clusters can be reflected in the mapping between NWP and power. Compared with this proposed method, RMSE and MAE of HHDGAT without WAF are increased by 0.88% and 1.03%, respectively, while R^2 is decreased by 5.96%. This is because the traditional distance measurement method between features cannot mine deep spatiotemporal relationships, and the introduction of noise edges

results in wrong information during the feature fusion process. Compared with the proposed method, RMSE of WAF-HHGAT using a predefined graph structure is increased by 0.79%, MAE is increased by 0.14%, and R^2 is decreased by 1.81%. The dynamic change of the graph structure significantly enhances the model's ability to fit and capture severe fluctuations in wind power. Periodically updating the graph structure according to the prior information provided by the meteorological center avoids the rigidity associated with a predefined graph structure.

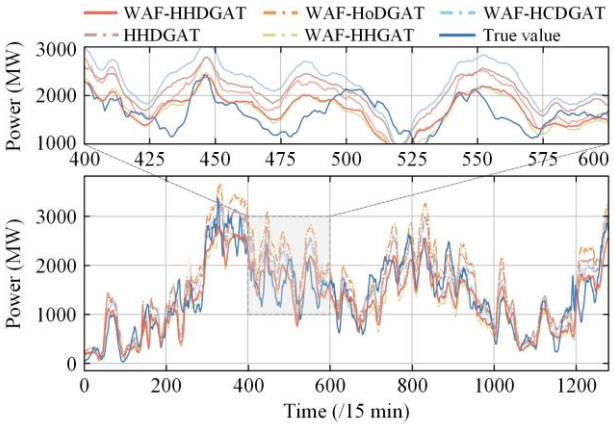


Fig. 9. The necessity verification of each part of the model proposed in this paper.

TABLE II
ABLATION EXPERIMENTAL RESULTS

Method	RMSE	MAE	R^2
WAF-HHDGAT	0.0577	0.0399	0.8168
WAF-HoDGAT	0.0777	0.0556	0.6681
WAF-HeDGAT	0.0603	0.0427	0.7999
HHDGAT	0.0665	0.0502	0.7572
WAF-HHGAT	0.0656	0.0413	0.7987

D. Comparative Analysis with Different Graph Construction

In order to verify the effectiveness of the adjacency matrix constructed by the wind characteristic angle in the proposed method, it is compared with the following other graph construction methods.

Method1: The proposed method.

Method2: The method for constructing adjacency matrix is modified to utilize wind speed-based GASF for adjacency matrix construction.

Method3: The method of constructing adjacency matrix is modified to utilize wind direction-based GASF for adjacency matrix construction.

Method4: The method of constructing adjacency matrix is modified to utilize wind speed-based Pearson correlation coefficient for adjacency matrix construction.

Method5: The method of constructing adjacency matrix is modified to utilize geographic location for adjacency matrix construction.

With different adjacency matrix construction methods, the final prediction results are shown in Fig. 10, and the error statistics of each adjacency matrix construction method are shown in Table III. Compared with Method1, RMSE of Method2 and Method3 are increased by 0.29% and 0.57%, MAE are increased by 0.3% and 1.02%, and R^2 are decreased by 1.87% and 3.79%, respectively. This indicates that it is beneficial to introduce the wind direction feature to provide the model more prior information to actively distinguish wind speed and power fitting process under different trends. The mapping process of the proposed method is adaptively adjusted according to the trend of wind speed and wind direction, hereby more effectively reflecting the STC of WFs. Method3 only takes into account the adjacency matrix constructed based on wind direction, which results in the loss of wind speed information in the edges of the graph—information that is closely related to changes in power trends. Compared with Method1, Method4 shows an increase of 1.65% in RMSE, 1.33% in MAE, and a decrease of 11.92% in R^2. Similarly, Method5 exhibits an increase of 2.09% in RMSE, 2.16% in MAE, and a decrease of 15.59% in R^2 compared with Method1. The example demonstrates the limitations of predefining the STC between WFs based solely on the distance between time series and the geographical location of WF for wind power prediction. It confirms that the graph structure constructed by the proposed method can better characterize the STC characteristics of WFs.

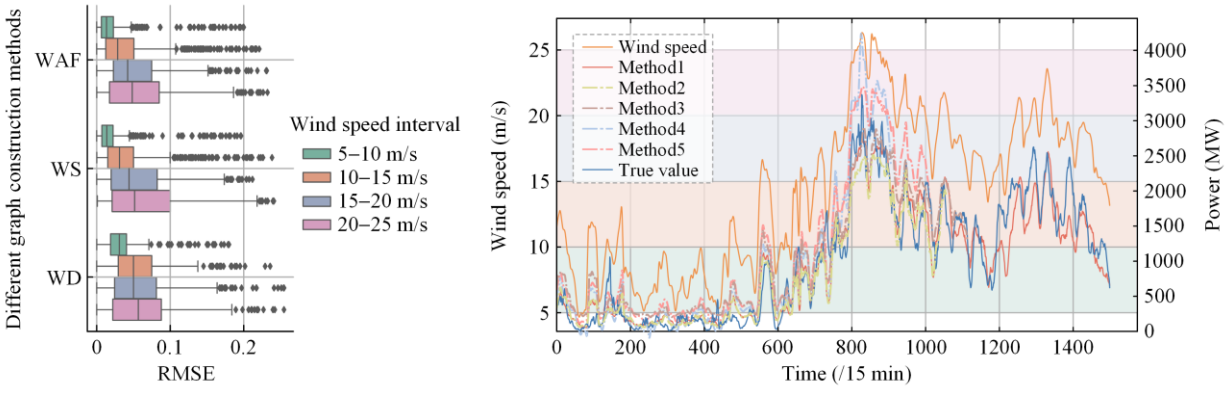


Fig. 10. The proposed adjacency matrix construction method can adapt to the power change in different wind speed ranges.

TABLE III
THE ERROR STATISTICAL RESULTS OF DIFFERENT ADJACENCY MATRIX CONSTRUCTION METHODS

Method	RMSE	MAE	R^2
1	0.0577	0.0399	0.8168
2	0.0606	0.0429	0.7981
3	0.0634	0.0501	0.7789
4	0.0742	0.0532	0.6976
5	0.0786	0.0615	0.6609

E. Comparative Analysis with Other Advanced Wind Power Forecasting Methods

In order to further verify the superiority of the constructed WAF-HHGAT and the extraction effectiveness of spatial and temporal characteristics of wind power, the proposed method is compared with other advanced wind power prediction methods under the same dataset, including:

1) WAF-HHDGAT—the proposed method;

2) WAF-HHGCN—replacing GAT in the proposed method by GCN, with other structures unchanged;

3) Transformer—an advanced model with a large number of parameters, widely used for its multi-head attention mechanism;

4) G-GAT—the adjacency matrix of GAT based on geographical location;

5) CBAM-LSTM—the NWP is converted into a four-dimensional matrix and input into CBAM-LSTM to extract the temporal and spatial features between WFs for wind power prediction;

6) TCN-LSTM—the NWP is converted into a two-dimensional matrix and input into the TCN-LSTM to extract the spatiotemporal features between WFs for wind power prediction;

7) LASSO—using the linear model LASSO to train and predict wind power.

Figure 11 specifically visualizes the wind power prediction results against the actual observed wind power for the proposed WAF-HHDGAT and baseline models. First, WAF-HHDGAT is compared with WAF-HHGCN. In this study, the DGAT within the designed structure addresses the limitations of traditional GCNs in handling dynamic changes in edge weights between different nodes and in assigning varying learning weights to different neighboring nodes in a directed graph, thus more effectively capturing the influence relationships among WFs. Comparing WAF-HHDGAT with WAF-HHGCN and G-GAT, it can be seen that the proposed framework is applicable to other GNNs. Compared with the GCN in wind power prediction, WAF-HHDGAT can greatly improve the model's ability to extract the spatial and temporal characteristics of each WF. Next, when compared with the Transformer model, which is an important large-scale parametric model, WAF-HHDGAT demonstrates the advantages of its specifically designed spatiotemporal feature extraction strategy. Although Transformer excels at capturing long-range dependencies in sequences, WAF-HHDGAT integrates the WAF matrix and a heterogeneous graph attention network, can more explicitly and finely model the dynamic spatial correlations and influencing factors among geographically distributed wind farms, which is crucial for precise wind power forecasting. The proposed method is then compared with CBAM-LSTM and TCN-LSTM. The proposed method extracts the spatial correlation between WFs by considering the characteristics of each WF and the topological structure between WFs. Consequently, it addresses the problem that TCN and CBAM cannot construct the correlation between WFs based on prior knowledge. Finally, WAF-HHDGAT is compared with the linear benchmark model LASSO.

The error statistics of the advanced WFC prediction methods are shown in Table IV. Compared with WAF-HHGCN and G-GAT, WAF-HHDGAT results in 0.28% and 2.66% reductions in RMSE, 0.14% and 1.88% reductions in MAE, and 1.81% and 20.7% increases in R^2. Similarly, compared with Transformer, its RMSE is decreased by 0.96%, MAE is decreased by 1.22%, and R^2 is increased by 10.52%. Compared with CBAM-LSTM, its RMSE is reduced by 1.65%, MAE is reduced by 1.33%, and R^2 is increased by 11.92%. Compared with TCN-LSTM and LASSO, its RMSE is decreased by 2.64% and 3.71%, MAE is decreased by 2.43% and 3.38%, and R^2 is increased by 20.53% and 25.15%, respectively. Therefore, it can be seen from the evaluation indexes that the prediction results of WAF-HHDGAT are the best, and the proposed prediction method has the best prediction effect.

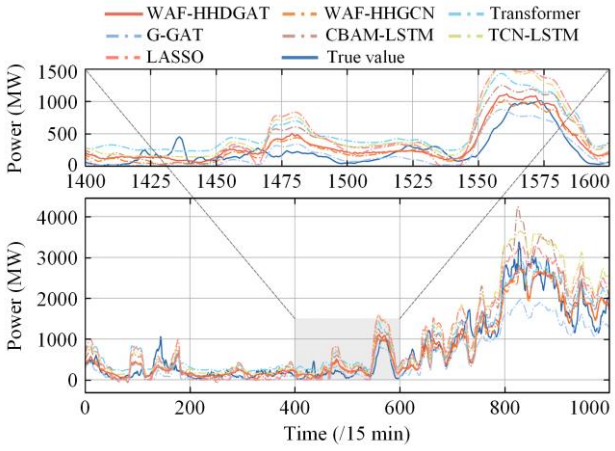


Fig. 11. Forecasting results of different wind power forecasting methods.

TABLE IV
ERROR STATISTICAL RESULTS OF DIFFERENT WIND POWER FORECASTING METHODS

Method	RMSE	MAE	R^2
WAF-HHDGAT	0.0577	0.0399	0.8168
WAF-HHGCN	0.0605	0.0413	0.7987
Transformer	0.0673	0.0521	0.7116
G-GAT	0.0843	0.0587	0.6098
CBAM-LSTM	0.0742	0.0532	0.6976
TCN-LSTM	0.0841	0.0642	0.6115
LASSO	0.0948	0.0737	0.5653

To further verify the performance of the proposed method in the large-scale WFC prediction, it is compared with three best performed CBAM-LSTM cluster prediction methods: the whole method, the accumulation method, and the statistical upscaling method. The comparison results are shown in Fig. 12, which demonstrate that the proposed model performs well in large-scale WFC power prediction tasks.

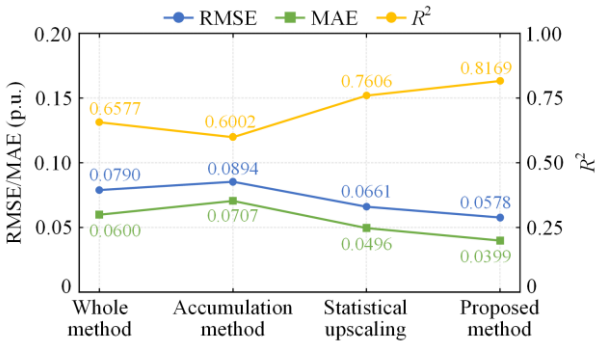


Fig. 12. Comparison with other cluster power prediction methods.

F. Complexity Analysis and Deployment Feasibility

To comprehensively evaluate the computational burden and practical applicability of the proposed model, a detailed complexity analysis is conducted and the key performance metrics are summarized in Table V.

The model has 6086 trainable parameters. It employs a dynamic graph strategy, updating the graph structure using the latest 96 steps of NWP. A single complete graph update takes 0.305 s, and the model is trained for 100 epochs. The average time per epoch, including dynamic graph updates and optimization, is 32.96 s. The estimated total graph construction time during training is 3053.22 s, while pure model training takes approximately 448.12 s. The average pure training time remains efficient at 0.47 ms per sample.

In the testing, the average pure inference time is 0.0631 ms per sample. For online prediction, the total prediction time for one window is approximately 0.311 s (0.305 s for graph update and 6.06 ms for inference).

These metrics support both offline and online deployments. The estimated total training duration for 100 epochs is 3515.36 s, suitable for offline model development and retraining. For online use, the 0.311 s response for a 96-sample window prediction is efficient for real-time tasks, especially when aligned with NWP update cycles. The low per-sample inference cost also benefits frequent prediction scenarios.

TABLE V
COMPUTATIONAL PERFORMANCE AND DYNAMIC GRAPH UPDATE COSTS OF THE PROPOSED MODEL

Metric	Value
Total trainable parameters	6086
Total training duration (s)	3515.36
Total graph construction time during training (s)	3053.22
Total pure model training time (s)	448.12
Average time per epoch (s)	32.96
Average time per graph update (s)	0.305
Average pure training time per sample (ms)	0.47
Total testing duration (s)	18.96
Total testing graph construction time (s)	18.59
Total pure model inference time (s)	0.3674
Total prediction time for 96 samples (s)	0.311
Inference graph update (s)	0.305
Inference time for 96 samples (ms)	6.06
Average pure inference time per sample (ms)	0.0631

G. Visualization of Model Mechanisms and Attention Analysis

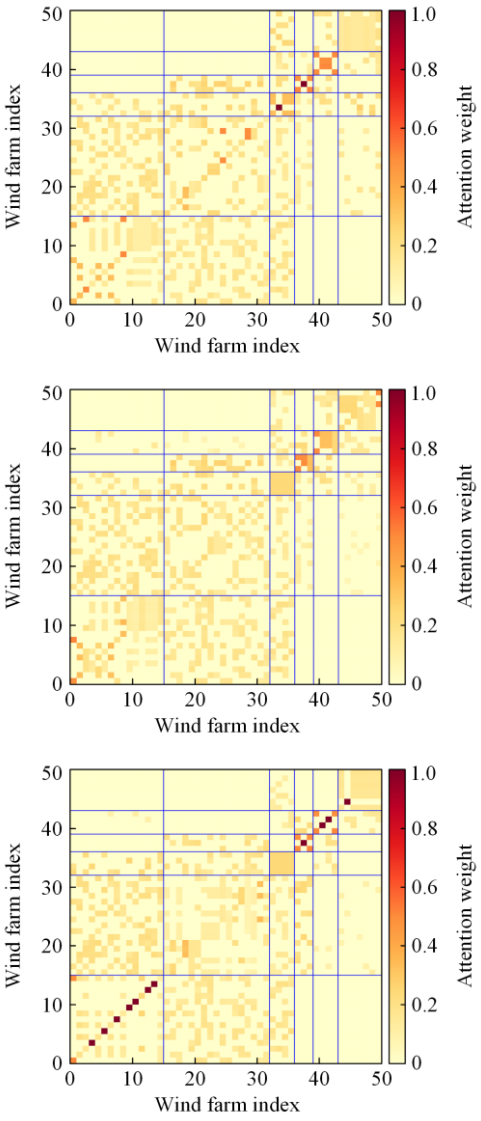
To further assess the model's ability to capture complex spatiotemporal dependencies by concurrently processing intra-cluster homogeneous and inter-cluster heterogeneous relationships, this section visually analyzes the learned comprehensive attention weights.

Figure 13 presents the normalized attention weight heatmaps for four random test days, where the color intensity of cell (i, j) indicates the attention target wind farm i assigns to source wind farm j for its prediction. The blue grid lines demarcate distinct wind farm clusters, derived from historical power data clustering.

First, the prominently highlighted diagonals confirm that each wind farm strongly prioritizes its own NWP state for prediction. Second, significant attention weights are visible within the clusters, demonstrating the model's effective learning of mutual influences among homogeneously grouped wind farms; while the daily dynamic shifts in these internal attentions further reflect its real-time adaptability. Furthermore, observable and dynamically changing attention weights between different clusters prove that the model successfully captures and quantifies heterogeneous associations beyond intra-cluster dependencies. This enables the

learning of influences between, for instance, distant or characteristically different clusters, potentially driven by large-scale meteorological events or grid-level correlations. Finally, clear daily variations in both intra-cluster and inter-cluster attention patterns underscore the model's overall capacity to adaptively allocate attention, thereby capturing time-varying homogeneous and heterogeneous spatiotemporal correlations.

In summary, the visualized attention weights intuitively demonstrate the model's unified framework for handling both homogeneous and heterogeneous information. The model effectively leverages individual wind farm data and intra-cluster influences while, crucially, identifying, quantifying, and dynamically adjusting inter-cluster dependencies. This enhances interpretability and mechanistically validates the proposed method's efficacy for power prediction in wind farm clusters characterized by complex internal and external relationships.



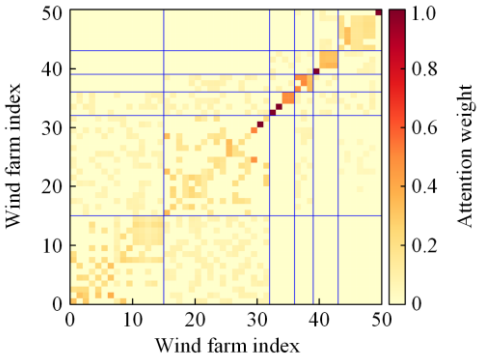
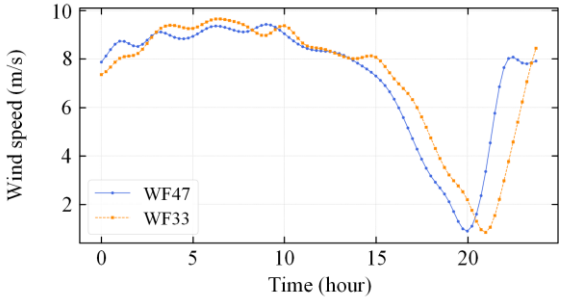
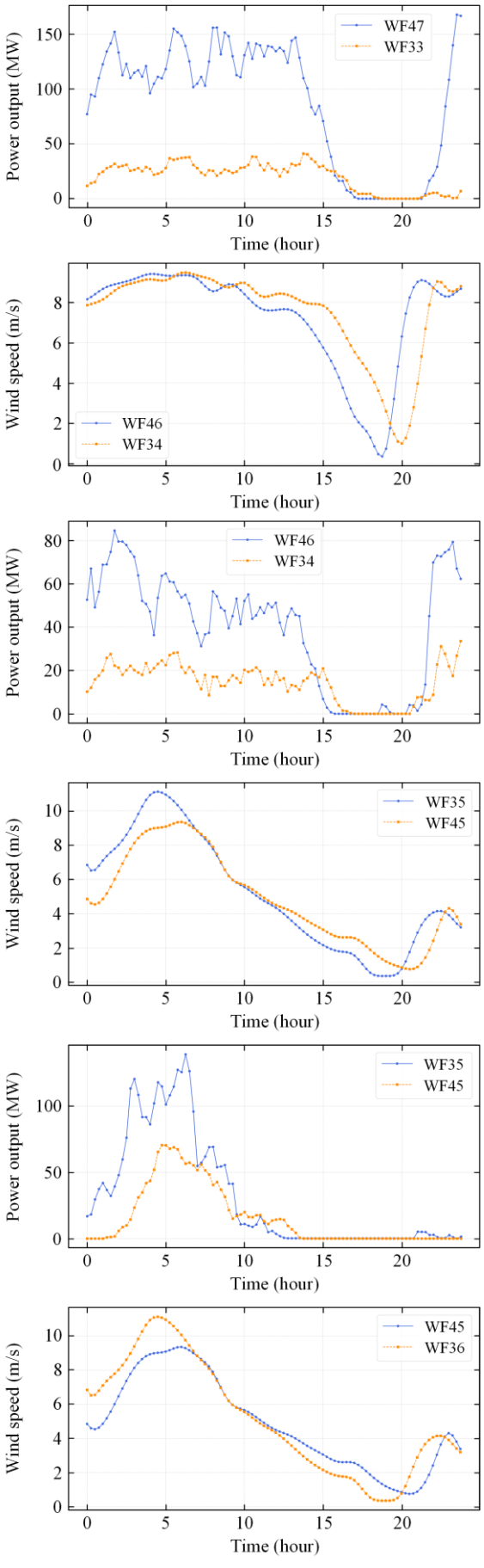


Fig. 13. The attention weight heat map on the 4th, 9th, 17th, and 24th days of the test set.

To further investigate the physical meaning behind the learned attention weights, a detailed case study is conducted. The fourth day of the test set (corresponding to the heatmap in Fig. 13) is selected and the analysis is focused on four pairs of wind farms with high attention weights between the third and sixth clusters. The NWP wind speed and actual power output data for these four pairs on that day are visualized in Fig. 14.

As seen, Fig. 14 clearly reveals the physical connections underlying these high attention weights. A prominent shared pattern across all four pairs is that the two curves within each pair exhibit highly similar fluctuation patterns, but with a distinct time lag. For instance, in the first pair (WF 47 and WF 33) and the third pair (WF 35 and WF 45), the overall shapes and trends of the two power profiles are nearly identical, yet one consistently peaks or changes direction several hours after the other. The second pair (WF 46 and WF 34) and the fourth pair (WF 45 and WF 36) exhibit similar lagged correlation characteristics. This phenomenon—where paired wind farms share similar patterns that are offset in time—provides strong evidence of a large-scale weather system with a clear directional movement sequentially affecting these geographically dispersed regions. Therefore, the high attention weights learned by the model are not mere statistical coincidences but an accurate capture of the non-local, time-lagged spatiotemporal correlations inherent in this physical process, intuitively validating the proposed method's capability in interpreting complex meteorological dynamics.





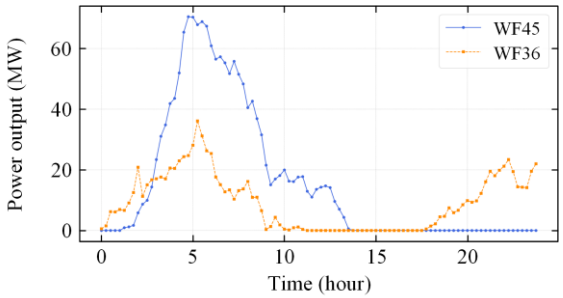


Fig. 14. Case study of the spatiotemporal correlation between WFs between the third cluster and the sixth cluster wind farms on the fourth day of the test set.

H. Robustness and Applicability Analysis Under Varying NWP Conditions

To further validate the applicability and robustness of the proposed method under varying NWP conditions, an additional comparative experiment is conducted. In this experiment, three distinct NWP datasets with different update frequencies (once, twice, and four times a day) are utilized, which inherently contain accuracy variations corresponding to different forecast horizons. The performance of the proposed method is then compared with the benchmark method (TCN-LSTM) under these scenarios, with the results presented in Table VI.

As shown in Table VI, the proposed method significantly outperforms the benchmark method across all evaluation metrics in every tested scenario. Specifically, while the performances of both models are improved with more frequent NWP updates, the superiority of the proposed method remains consistent. Even if the NWP data quality varies, the performance advantage of the proposed model over the benchmark remains robust. These results strongly demonstrate that the proposed method exhibits excellent adaptability and robustness to variations in both NWP update frequency and accuracy.

TABLE VI
MODEL PERFORMANCE COMPARISON UNDER DIFFERENT NWP UPDATE FREQUENCIES AND ACCURACY

NWP update frequency	NWP accuracy	Evaluating indicator	Benchmark method	Proposed method
Once a day	0.0809	RMSE	0.0841	0.0577
		MAE	0.0642	0.0399
		R^2	0.6115	0.8168
Twice a day	0.0811	RMSE	0.0879	0.0593
		MAE	0.0712	0.0407
		R^2	0.5988	0.8099
Four times a day	0.0792	RMSE	0.0801	0.0572
		MAE	0.0575	0.0391
		R^2	0.6543	0.8224

V. CONCLUSION

In this paper, a graph structure learning and STC recognition and judgment process between WFs based on the interactive fusion of homogeneous and heterogeneous information is proposed for large-scale WFC short-term power prediction. The STC characteristics between WFs in a large-scale WFC are modeled in a refined manner. In this paper, analyzing and leveraging

the output heterogeneity of WFCs offers a novel approach for more accurately characterizing the complex relationship within a WFC.

The STC discrimination based on the WAF matrix considers the correlation of wind power and selects the WF nodes with similar characteristics at the same time, while the information obtained through interaction acts on the learning of the target WF nodes. The proposed model of homogeneous and heterogeneous information interaction effectively improves the power prediction accuracy. Compared with other advanced methods, RMSE and MAE are decreased respectively by 2.67% and 2.26% on average, while R^2 is increased by 19.63% on average.

Although the proposed method effectively improves prediction accuracy, there exist several limitations. Methodologically, the discrimination of STC is affected by factors beyond wind speed and direction, and the clustering method used is relatively simple. Furthermore, the study is validated based on a dataset from a single region in Northeast China covering an eight-month period. Consequently, the model's applicability and robustness across different geographical areas and seasonal conditions need be fully explored.

Future work will proceed in three directions: 1) refining the current model with additional factors and advanced clustering; 2) generalizing the WAF-SIFT framework to other domains, such as photovoltaic power forecasting; and 3) integrating the proposed spatiotemporal module with large-scale architectures to synergistically enhance the prediction accuracy and robustness.

ACKNOWLEDGMENT

Not applicable.

AUTHORS' CONTRIBUTIONS

Zhe Zhang: conceptualization, methodology, software, validation, formal analysis, and writing original draft preparation. Bo Wang: supervision, project administration, funding acquisition, writing review, and editing. Mao Yang: supervision, conceptualization, writing review, and editing. Zhao Wang: methodology, investigation, and data curation. Wei Zhang: validation and resources. Tao Huang: visualization, writing review, and editing. Xin Su: data curation and software. All authors read and approved the final manuscript.

FUNDING

This work is supported by the National Key R&D Program of China (No.2022YFB2403000).

AVAILABILITY OF DATA AND MATERIALS

Please contact the corresponding author for data material request.

DECLARATIONS

Competing interests: 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 article.

AUTHORS' INFORMATION

Zhe Zhang is studying for a master's degree in Northeast Electric Power University, Jilin, China. His research direction is wind power/photovoltaic power prediction.

Bo Wang is a professor-level senior engineer and director of the Electric Power Meteorology Forecasting Team at the China Electric Power Research Institute (CEPRI). His expertise lies in renewable energy power forecasting and electricity meteorology. He has led over 20 national and corporate research projects, authored 4 monographs, published more than 40 papers, and been granted 20 invention patents.

Mao Yang received the Ph.D. degree in control theory and control engineering from Jilin University, Changchun, China, in 2010. Since 2018, he has been a professor with the School of Electrical Engineering, Northeast Electric Power University, Jilin, China. His research interests include the areas of load forecasting and micro-grid operation management with an emphasis on wind power/photovoltaic power prediction.

Zhao Wang is a senior engineer and graduate supervisor at CEPRI's Renewable Energy Department. He obtained his Ph.D. degree from Tsinghua University in 2018. With extensive theoretical and practical experience, his research focuses on renewable energy forecasting, power supply capacity evaluation, and AI applications in meteorology.

Wei Zhang is an associate professor, her main research direction is renewable energy power prediction.

Tao Huang received the Ph.D. degree from Politecnico di Torino, Turin, Italy, in 2011. He is currently an associate professor with the Department of Energy, Politecnico di Torino. He is also an adjunct professor with Xihua University, Chengdu, China, and Polytechnic University of Henan, Jiaozuo, China. His research interests include vulnerability assessment, electricity markets, smart grids, active distribution network, and artificial intelligence in power systems.

Xin Su graduated from Jilin University in 2011. She is with the Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education. Her main focus is on renewable energy power prediction.

REFERENCES

[1] IRENA, (2024, Jul.), Renewable Energy Statistics 2024, [Online]. Available: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2024/Jul/IRENA_Renewable_Energy_Statistics_2024.pdf

[2] M. Yang, Y. Guo, and T. Huang *et al*., "Power prediction considering NWP wind speed error tolerability: a strategy to improve the accuracy of short-term wind power prediction under wind speed offset scenarios," *Applied Energy*, vol. 377, May 2024.
[3] S. Hanifi, X. Liu, and Z. Lin *et al*., "A critical review of wind power forecasting methods—past, present and future," *Energies*, vol. 13, no. 15, Jul. 2020.
[4] G. Mu, F. Wu, and X. Zhang *et al*., "Analytic mechanism for the cumulative effect of wind power fluctuations from single wind farm to wind farm cluster," *CSEE Journal of Power and Energy Systems*, vol. 8, no. 5, pp. 1290-1301, Sep. 2022.
[5] M. Yang, L. Zhang, and Y. Cui *et al*., "Investigating the wind power smoothing effect using set pair analysis," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 3, pp. 1161-1172, Jul. 2020.
[6] J. Simeunovic, B. Schubnel, and P.-J. Alet *et al*., "Spatio-temporal graph neural networks for multi-site PV power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 1210-1220, Apr. 2022.
[7] M. Yang, Y. Huang, and C. Xu *et al*., "Review of several key processes in wind power forecasting: Mathematical formulations, scientific problems, and logical relations," *Applied Energy*, vol. 377, May 2024.
[8] G. Hou, J. Wang, and Y. Fan, "Wind power forecasting method of large-scale wind turbine clusters based on DBSCAN clustering and an enhanced hunter-prey optimization algorithm," *Energy Conversion and Management*, vol. 307, May 2024.
[9] J. Zhang, D. Liu, and Z. Li *et al*., "Power prediction of a wind farm cluster based on spatiotemporal correlations," *Applied Energy*, vol. 302, Nov. 2021.
[10] A. Ahmadpour and S. Gholami Farkoush, "Gaussian models for probabilistic and deterministic wind power prediction: wind farm and regional," *International Journal of Hydrogen Energy*, vol. 45, no. 51, pp. 27779-27791, Oct. 2020.
[11] Y. Zhao, L. Ye, and P. Pinson *et al*., "Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5029-5040, Sep. 2018.
[12] Z. Tang, J. Liu, and J. Ni *et al*., "Power prediction of wind farm considering the wake effect and its boundary layer compensation," *Protection and Control of Modern Power Systems*, vol. 9, no. 6, pp. 19-29, Nov. 2024.
[13] X. Liu, Y. Pan, and Y. Wang *et al*., "Enhanced ultra-short-term wind power forecasting: a PatchTST-POA-VMD-iTransformer hybrid model," *Power System Protection and Control*, vol. 19, no. 53, pp. 68-78, Apr. 2025. (in Chinese)
[14] L. Yang and Z. Zhang, "A deep attention convolutional recurrent network assisted by K-shape clustering and enhanced memory for short term wind speed predictions," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 856-867, Apr. 2022.
[15] J. Zhao, Z. Liu, and Q. Sun *et al*., "Attention-based dynamic spatial-temporal graph convolutional networks for traffic speed forecasting," *Expert Systems with Applications*, vol. 204, Oct. 2022.
[16] S. Xia, C. Zhang, and Y. Li *et al*., "GCN-LSTM based transient angle stability assessment method for future power systems considering spatial-temporal disturbance response characteristics," *Protection and Control of Modern Power Systems*, vol. 9, no. 6, pp. 108-121, Nov. 2024.
[17] C. Pan, C. Wang, and H. Sun *et al*, "STAGN ultra-short-term wind power forecasting based on hyperparameter optimization and error correction," *Power System Protection and Control*, vol. 8, no. 53, pp. 117-129, Aug. 2025. (in Chinese)
[18] Y. Hu, L. Zhu, and J. Li *et al*, "Short-term wind power forecasting with the integration of a deep error feedback learning and attention mechanism," *Power System Protection and Control*, vol. 4, no. 52, pp. 100-108, Dec. 2024. (in Chinese)
[19] N. Hu, D. Zhang, and K. Xie *et al.*, "Graph learning-based spatial-temporal graph convolutional neural networks for traffic forecasting," *Connection Science*, vol. 34, no. 1, pp. 429-448, Nov. 2022.
[20] X. Xu, S. Hu, and H. Shao *et al*., "A spatio-temporal forecasting model using optimally weighted graph convolutional network and gated recurrent unit for wind speed of different sites distributed in an offshore wind farm," *Energy*, vol. 284, Dec. 2023.
[21] Y. Song, D. Tang, and J. Yu *et al*., "Short-term forecasting based on graph convolution networks and multiresolution convolution neural networks for wind power," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1691-1702, May 2022.
[22] W. Zhang, K. Zhu, and S. Zhang *et al*., "Dynamic graph convolutional networks based on spatiotemporal data embedding for traffic flow forecasting," *Knowledge-Based Systems*, vol. 250, Aug. 2022.
[23] Z. Li, L. Ye, and Y. Zhao *et al*., "A spatiotemporal directed graph convolution network for ultra-short-term wind power prediction," *IEEE Transactions on Sustainable Energy*, vol. 14, no. 1, pp. 39-54, Jan. 2023.
[24] M. Yang, Y. Guo, and F. Fan, "Ultra-short-term prediction of wind farm cluster power based on embedded graph structure learning with spatiotemporal information gain," *IEEE Transactions on Sustainable Energy*, vol. 16, no. 1, pp. 308-322, Jan. 2025.
[25] Y. Gu, J. D. McCalley, and M. Ni, "Coordinating large-scale wind integration and transmission planning," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 4, pp. 652-659, Oct. 2012.
[26] Z. Li, L. Ye, and X. Song *et al*., "Heterogeneous spatiotemporal graph convolution network for multi-modal wind-PV power collaborative prediction," *IEEE Transactions on Power Systems*, vol. 39, no. 4, pp. 5591-5608, Jul. 2024.
[27] K. Wang, X. Qi, and H. Liu *et al*., "Deep belief network based k-means cluster approach for short-term wind power forecasting," *Energy*, vol. 165, pp. 840-852, Dec. 2018.
[28] Q. Gao, "Multi-temporal scale wind power forecasting based on Lasso-CNN-LSTM-LightGBM," *EAI Endorsed Transactions on Energy Web*, vol. 11, pp. 1-8, Dec. 2024.
[29] X. Liu, L. Zhang, and J. Wang *et al*., "A unified multi-step wind speed forecasting framework based on numerical weather prediction grids and wind farm monitoring data," *Renewable Energy*, vol. 211, pp. 948-963, Jun. 2023.
[30] M. Yang, Y. Guo, and B. Wang *et al*., "A day-ahead wind speed correction method: enhancing wind speed forecasting accuracy using a strategy combining dynamic feature weighting with multi-source information and dynamic matching with improved similarity function," *Expert Systems With Applications*, vol. 263, no. 125724, Apr. 2024.