Spatio-temporal Convolutional Network Based Power Forecasting of Multiple Wind Farms

Xiaochong Dong, Yingyun Sun, Ye Li, Xinying Wang, and Tianjiao Pu

***Abstract*—The rapidly increasing wind power penetration presents new challenges to the operation of power systems. Improving the accuracy of wind power forecasting is a possible solution under this circumstance. In the power forecasting of multiple wind farms, determining the spatio-temporal correlation of multiple wind farms is critical for improving the forecasting accuracy. This paper proposes a spatio-temporal convolutional network (STCN) that utilizes a directed graph convolutional structure. A temporal convolutional network is also adopted to characterize the temporal features of wind power. Historical data from 15 wind farms in Australia are used in the case study. The forecasting results show that the proposed model has higher accuracy than the existing methods. Based on the structure of the STCN, asymmetric spatial correlation at different temporal scales can be observed, which shows the effectiveness of the proposed model.**

***Index Terms*—Deep learning, spatio-temporal correlation, wind power forecasting, graph conventional network (GCN).**

I. Introduction

As a clean source of renewable energy, wind power generation has been rapidly developed and globally applied in recent years [1]. Because of its intermittent and stochastic nature, the rapid increase of wind power penetration presents new challenges to the dispatching and planning of power systems. Improving the accuracy of wind power forecasting can contribute to improving the power system stability and increasing the consumption rates of wind power [2].

Wind power forecasting technology has made great advances over the last several decades, but most existing methods have focused merely on single site using temporal features. However, power generation from multiple wind farms has not only temporal correlation but also spatial correlation. The existing research has also shown that wind power forecasting errors are reduced significantly when spatio-temporal correlation is considered [3].

Wind power forecasting with spatio-temporal correlation can be used not only in multiple wind farms in an area but also in the wind turbines of a wind farm [4]. Atmospheric motion produces similar patterns of wind speed and wind direction between different sites in an area. Therefore, the spatio-temporal correlation of wind power is related to wind speed, wind direction, and the distance between sites. A method for combining the historical data of multiple sites with prior knowledge of spatio-temporal correlation is the main goal of this study.

The existing wind power forecasting methods with spatio-temporal correlation can be divided into the following categories.

1) Physical methods: they are based on numerical weather prediction and geographical information. Although physical methods describe the atmospheric motion accurately, the high computational complexity limits the wide application of short-term wind power forecasting [5], [6].

2) Statistical methods: they use historical measurement data to establish explicit models between forecasts and measurements. These methods include Markov chains [7], copula functions [8], and regressive models [9], which have much simpler computational complexity as compared with physical methods. In [10], a conditional parametric model is proposed, which is later used as the basis for probabilistic forecasting [11]. In [12], a wind power forecasting framework is proposed by using off-site information for the purposes of privacy preservation. However, these approaches do not employ wind power data with high-dimensional nonlinear features.

3) Artificial intelligence (AI) methods: they use historical measurements to train deep learning models and establish implicit models [13]. Convolutional neural networks (CNNs) are widely used to learn spatial features of wind power data and are often combined with long short-term memory (LSTM) [14] or gated recurrent units (GRUs) [15] to learn temporal features. These methods enable implicit models to realize wind power forecasting when considering spatio-temporal correlation.

Although deep learning approaches have obtained better forecasting results than statistical models, most AI methods reported previously have several drawbacks as follows.

1) Many deep learning models represent the data in the Euclidean space. Although the sequential data can be expressed effectively, geographical data of multiple sites clear-

Manuscript received: December 4, 2020; revised: March 26, 2021; accepted: October 11, 2021. Date of CrossCheck: October 11, 2021. Date of online publication: December 8, 2021.

This work was supported in part by National Key Research and Development Program (No. 2020YFB0905900), and in part by National Natural Science Foundation of China (No. 51777065).

X. Dong and Y. Sun (corresponding author) are with the School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China (e-mail: dream_dxc@163.com; sunyy@ncepu.edu.cn).

Y. Li, X. Wang, and T. Pu are with the China Electric Power Research Institute, Beijing 100192, China (e-mail: liye@epri.sgcc.com.cn; wangxinying@geiri.sgcc.com.cn; tjpu@epri.sgcc.com.cn).

DOI: 10.35833/MPCE.2020.000849

ly belong to non-Euclidean domains. Describing geographical data in the Euclidean space will destroy prior information of the relative positions between sites [16].

2) Interpretability is a major problem for deep learning models that is difficult to combine with prior information (e.g., negative correlation between distance and spatial correlation, and spatial correlation is related to wind direction and speed).

Geographical data can be represented by a graph model. However, the structure complexity of graphical data imposes significant challenges on the existing deep learning models. The graph neural network (GNN) was first introduced in [17], and since then, many studies on extending GNN models have emerged. Inspired by the great success of CNN, the graph convolutional network (GCN) has shown excellent capabilities in spatial feature extraction. In the fields of transportation [18] and meteorology [19], the GCN model has become a hot research topic. Reference [20] uses a GCN model to describe the spatial correlation for wind speed forecasting. In addition, a learnable adjacency matrix for the GCN model to forecast wind speed is proposed in [21], and a GCN model is used for fault location in distribution systems in [22]. Finally, in [23], a GCN model is combined with the autoencoder for spatio-temporal solar irradiance forecasting.

Although GCNs have proven to be effective in dealing with non-Euclidean data, the challenges of wind power forecasting remain. The traditional GCN is a general graph learning structure of a GNN, and there is room for improvement in the application of wind power forecasting. In this paper, a spatio-temporal convolutional network (STCN) is proposed based on a spectral graph convolution and temporal convolutional network (TCN). The main contributions of this paper are described as follows.

1) A directed graph structure is redesigned for multiple wind farms, and a directed graph convolutional structure is proposed. The proposed directed graph convolutional structure can better extract asymmetric spatial features from the data of multiple wind farms, making the model more interpretable.

2) A GCN is combined with a TCN to form an STCN model that can transmit spatial features at different temporal scales. Simulation results show that the STCN has higher accuracy in spatio-temporal wind power forecasting as compared with traditional methods.

The rest of this paper is organized as follows. Section II describes the problem formulation of power forecasting for multiple wind farms. Section III presents the background theories on GCN and TCN. Section IV introduces the power forecasting model of STCN for multiple wind farms. Section V discusses the numerical results and Section VI concludes this study.

II. Problem Formulation

The objective of power forecasting for multiple wind farms is to capture the correlation among historical wind power data, geographical information, and wind power forecasting values in an area.

When the geographical scale of an area is tens to hundreds of kilometers, a mesoscale atmospheric physical model can be used to describe atmospheric motion [24]. Based on the atmospheric physical model, the power generation of multiple wind farms can be formulated by the nonlinear dynamic equation as [25]:

$$\begin{cases} x_{t+1} = f(x_t, \chi_t, t) + \xi_t \\ g_{t+1} = h(x_{t+1}, t+1) + \eta_{t+1} \end{cases} \tag{1}$$

where $f(\cdot)$ and $h(\cdot)$ are nonlinear functions; x is the state of the atmospheric system; χ is the state of the atmospheric system boundary, and the atmospheric states include wind speed, wind direction, temperature, and geopotential height; ξ and η are the white Gaussian noises; and t is the time index. The site state of the next moment x_{t+1} is related to the system boundary state χ_t and previous site state x_t, and the site wind power value g_{t+1} is related to the site state x_{t+1}.

Suppose a set of N wind farms is presented in an area, where the power of wind farm i at time t is denoted by P_t^i. The power of another wind farm $P_t^j (j \neq i)$ can then be used as an incomplete boundary value as input for the target power forecasting of wind farms.

The distance between wind farms i and j is denoted by d_{ij}, and the symmetric distance matrix $\boldsymbol{M}^{N \times N}$ is used to store the distance information. If L-step historical data are used as the forecasting input information to predict the wind power after K steps, the implicit formulation of the power forecasting problem for multiple wind farms can be written as:

$$H(P_{t-L+1:t}^i, \boldsymbol{M}, \theta) = P_{t+K}^i \tag{2}$$

where $P_{t-L+1:t}^i = \{P_{t-L+1}^i, P_{t-L+2}^i, ..., P_t^i\}$; $i \in \{1, 2, ..., N\}$; θ is the learnable parameter; and $H(\cdot)$ is the mapping function of the deep learning model.

III. Background Theories

A. Spectral Graph Convolution

First, an undirected graph model $G = (E, V, \boldsymbol{A})$ is used to describe the geographical information of wind farms, where E is the set of graph edges; V is the set of N nodes that corresponds to wind farms; and $\boldsymbol{A}$ is the adjacent matrix that includes the connection weight between nodes. In the geographical graph model, $\boldsymbol{A}$ can be calculated by the distance information matrix $\boldsymbol{M}$ [26].

Because it is difficult to define a GCN in a spatial domain, the spectral graph convolution in the Fourier domain is proposed, which transforms graph data x and convolution kernel g to the spectral domain for convolution. It then performs the inverse Fourier transform on the convolution result to obtain the spatial domain result [27].

An essential matrix in spectral graph convolution is the graph Laplacian, the normalized Laplacian matrix of which is defined as: $\boldsymbol{L} = \boldsymbol{I}_N - \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}}$, where $\boldsymbol{I}_N$ is the identity matrix; $\boldsymbol{\Lambda}$ is the eigenvalues diagonal matrix of $\boldsymbol{L}$; $\boldsymbol{U}$ is the eigenvector matrix of $\boldsymbol{L}$; and $\boldsymbol{D}$ is the degree matrix, the element of which is $D_{ii} = \sum_j A_{ij}$, and A_{ij} is the connection weight of nodes i and j. The spectral convolutional operation can be written as:

$$g*x = F^{-1}(F(g)\circ F(x)) = \boldsymbol{U}g(\boldsymbol{\Lambda})\boldsymbol{U}^{\mathrm{T}}x \tag{3}$$

where g is the convolution kernel; x is the input graph data; "$\circ$" is the Hadamard product; $F(x)=\boldsymbol{U}^{\mathrm{T}}x$ and $F^{-1}(x)=\boldsymbol{U}x$ are the graph Fourier transform of x and its inverse, respectively; and $g(\boldsymbol{\Lambda})=\mathrm{diag}(\theta)$ is defined as filter parameters to derive the GCN, but this design has $O(N^2)$ computational complexity. The solution of this problem is used to transform g_θ into a Chebyshev polynomial function as [28]:

$$g_\theta(\boldsymbol{\Lambda})*x = \boldsymbol{U}\left(\sum_{k=0}^{K}\theta_k T_k(\hat{\boldsymbol{\Lambda}})\right)\boldsymbol{U}^{\mathrm{T}}x = \sum_{k=0}^{K}\theta_k T_k(\hat{\boldsymbol{L}})x \tag{4}$$

where $\hat{\boldsymbol{\Lambda}}=2\boldsymbol{\Lambda}/\lambda_{\max}-\boldsymbol{I}_N$, and $\lambda_{\max}$ is the largest eigenvalue of $\boldsymbol{L}$; and $\hat{L}=2L/\lambda_{\max}-I_n$. Reference [29] proposes an approximate method as: $\lambda_{\max}\approx 2$, $K=1$, $\theta=\theta_0=-\theta_1$. Under these conditions, (4) can be simplified as:

$$g_\theta(\boldsymbol{\Lambda})*x \approx \theta_0 x + \theta_1(\boldsymbol{L}-\boldsymbol{I}_N)x = \theta(\boldsymbol{I}_N+\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2})x \tag{5}$$

To avoid the explosion/disappearance of gradients in the deep network, a renormalization trick can be used as:

$$\boldsymbol{L}=\boldsymbol{I}_N+\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}=\hat{\boldsymbol{D}}^{-1/2}\hat{\boldsymbol{A}}\hat{\boldsymbol{D}}^{-1/2},\hat{\boldsymbol{A}}=\boldsymbol{A}+\boldsymbol{I}_N,\hat{D}_{ii}=\sum_j \hat{A}_{ij} \tag{6}$$

Thus, (5) can be rewritten as:

$$g_\theta(\boldsymbol{\Lambda})*x = \theta\hat{\boldsymbol{D}}^{-1/2}\hat{\boldsymbol{A}}\hat{\boldsymbol{D}}^{-1/2}x \tag{7}$$

Most existing deep learning models use (7) to construct the first-order approximate graph convolution to avoid the calculation of high-order Chebyshev polynomials. This yields a clear geometric meaning, namely, the weighted summation of each node eigenvalues in the first-order neighborhood of the spatial domain.

B. Temporal Convolution

The spectral graph convolution can extract only spatial domain features of graph data. Therefore, it is necessary to choose a network structure to extract temporal features. Recurrent neural networks (RNNs) and the improved models (e.g., LSTM and GRU) are the common research methods to handle the sequential data. However, the recursive structures of these models make it difficult to compute in parallel, which introduces the problem of gradient explosion or disappearance.

The causal convolution of the TCN model is used to extract temporal features [30]. Unlike RNN-based approaches, the TCN model can handle the sequential data in a non-recursive manner, thus providing improved parallel computational capabilities and alleviating the gradient problem. When a sequential input data x and filter q are used, the causal convolution operation is represented as:

$$q(t)*x = \sum_{s=0}^{S-1} q(s)x(t-ds) \tag{8}$$

where S is the size of the convolution kernel; and d is the dilation factor that controls the interval of input convolution data. The output of causal convolution depends only on the previous sequential data, and therefore has a strictly causal relationship. In addition, the causal convolution network allows for an exponentially large receptive field by increasing the layer depth [31].

The structure of TCN based on causal convolution with kernel size of 2 is presented in Fig. 1. The output of different layers can transmit the relations between the historical sequential data and the forecasting values at different temporal scales. As shown in Fig. 1, layer 1 transmits the features of $[t-1, t]$ and layer 2 transmits the features of $[t-3, t]$, etc.

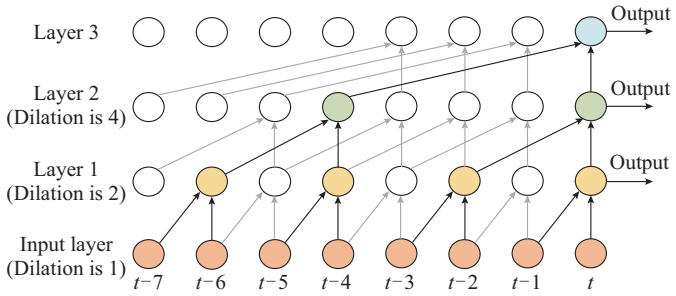

Fig. 1. Structure of TCN based on casual convolution with kernel size of 2.

IV. Methodology

In this section, the original graph and directed graph convolutional models are proposed. Then, the STCN power forecasting model for multiple wind farms is presented.

A. Graph Structure of Multiple Wind Farms

Section III provides background theories on traditional spectral graph convolution. It is worth noting that the traditional GCN model is based on the undirected graph with a definite connection. Using the traditional graph model for spatio-temporal forecasting has the following two drawbacks.

1) In the undirected graph spectral convolution, the spatial correlation between nodes is equal. The adjacency matrix $\boldsymbol{A}$ of the undirected graph is a symmetric matrix, thus $A_{ij}=A_{ji}$. However, the atmospheric motion is directional, which causes an unequal spatial correlation between wind farms. This characteristic can be extracted by means of a wind rose diagram [32].

2) The weight of edge in graph is used to describe the correlation between wind farms. However, the atmospheric motion is everywhere in an area. Therefore, the correlation of wind power is not only limited to the edges in the graph but also exists between and outside the graph nodes.

In summary, we need to construct a directed graph to describe the different correlation among wind farms. The graph structure of directed wind farms is designed using a fully-connected directed graph, which contains external interaction edges on the nodes. As an example, Fig. 2 presents a directed graph structure with three wind farms.

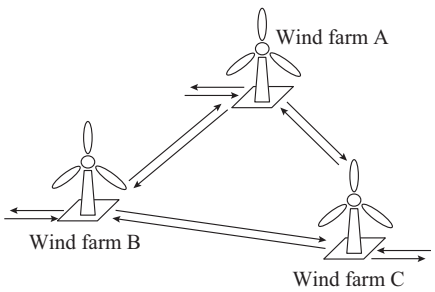

Fig. 2. Directed graph structure with three wind farms.

B. Directed Graph Convolutional Structure

A directed graph convolutional structure provides the foundation for describing the spatial correlation of different directions between graph nodes. The existing studies typically use matrix $\boldsymbol{M}$ to calculate matrix $\boldsymbol{A}$, where A_{ij} and A_{ji} correspond to the edges of nodes i and j in different directions. Using (7) produces the spectral graph convolution of the directed graph (Laplacian matrix $\boldsymbol{L}$ remains a symmetric constant matrix). Therefore, this study proposes a directed spectral graph convolutional structure to match the novel graph structure. The first-order directed spectral graph convolution can be transformed into [33]:

$$\begin{cases} g_\theta(\boldsymbol{\Lambda}) * x = [\theta_1(\boldsymbol{D}_O^{-1}\boldsymbol{A}) + \theta_2(\boldsymbol{D}_I^{-1}\boldsymbol{A}^{\mathrm{T}})]x \\ \boldsymbol{D}_O = \mathrm{diag}\left(\sum_i A_{ij}\right) \\ \boldsymbol{D}_I = \mathrm{diag}\left(\sum_j A_{ij}\right) \end{cases} \tag{9}$$

where $\boldsymbol{D}_O^{-1}\boldsymbol{A}$ and $\boldsymbol{D}_I^{-1}\boldsymbol{A}^{\mathrm{T}}$ are the Laplacian matrices of the output and input processes, respectively. Note that although a directed graph is used, no real two-way transmission process exists between wind farms. As an example, A_{ij} represents only the spatial correlation of wind farm i to wind farm j. When forecasting the power of wind farm j, there is no need to consider the spatial correlation of wind farm j to wind farm i. Therefore, in the power forecasting problem for multiple wind farms, the graph convolution needs to consider only the input process.

In the graph structure, the nodes have a connection with the outside. Therefore, we propose two self-adaptive diagonal matrices to capture the hidden spatial correlation with the outside. In summary, the first-order graph convolution is:

$$\begin{cases} g_\theta(\boldsymbol{\Lambda}) * x = [\theta(\boldsymbol{D}_O^{-1}\boldsymbol{A} + \boldsymbol{\theta}_I)]x \\ \boldsymbol{D}_O = \mathrm{diag}\left(\sum_i A_{ij} + \theta_O(i,i)\right) \end{cases} \tag{10}$$

where $\boldsymbol{\theta}_O = \mathrm{diag}(\theta')$ and $\boldsymbol{\theta}_I = \mathrm{diag}(\theta'')$ describe the correlation of each node to the outside and the correlation of the outside to each node, respectively; and $\theta', \theta'' \in \mathbf{R}^N$. One self-adaptive matrix has N learnable parameters. Off-diagonal elements of matrix $\boldsymbol{A}$ are represented by $1/d_{ij}$ and $A_{ii} = 0$, and therefore $\boldsymbol{D}_O^{-1}\boldsymbol{A} + \boldsymbol{\theta}_I$ is the Laplacian matrix of the proposed directed graph convolution. When the element $\theta_O(i,i)$ increases, it can effectively describe the characteristics that node i has a high spatial correlation with the outside and a weak correlation with the other nodes. When the element $\theta_I(i,i)$ increases, it can effectively describe the characteristics that the outside has a strong correlation with node i but the other nodes have a weak correlation with node i.

The directed graph convolutional structure is based on the directed graph structure. Because the two adaptive matrices have learnable parameters, the spatial correlation between multiple wind farms is not completely limited by prior distance information. In addition, the graph structure of multiple wind farms is a fully-connected graph so that the spatial correlation learning can be realized using the first-order graph convolution. There is no need to rely on high-order GCNs to transmit non-adjacent node information.

C. STCN Structure

Although the first-order graph convolutional layer can realize the transmission of spatial correlation features, the spatio-temporal features are not similar at different temporal scales. Therefore, this paper uses multiple first-order GCN layers to extract spatial features at different temporal scales. Considering the multi-layer TCN can extract the temporal features at different temporal scales, the GCN layer is combined with the TCN layer to form a GCN-TCN layer that can realize the spatial feature transmission at the corresponding temporal scales.

The STCN mainly consists of three stacked GCN-TCN layers, as shown in Fig. 3. With the stacking of multiple GCN-TCN layers, the GCN layer transmits short- and long-term spatial features at the bottom and top layers, respectively. To extract the spatio-temporal correlation features of various temporal scales to forecast wind power, STCN adds a skip connection to transmit output values of different GCN-TCN layers.

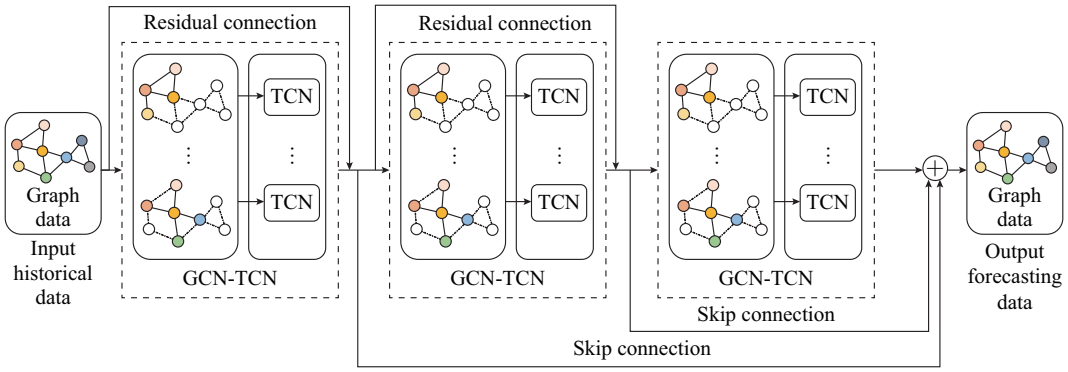

Fig. 3. Structure of STCN.

It should be noted that the TCN layer only contains one layer of the TCN network. This is due to the reasonable design of the dilation and kernel sizes of multiple TCN layers. Finally, the input data temporal scale can be compressed to one dimension in output, and overall features of all sequential data can be extracted effectively.

In addition to the output layer, each GCN-TCN layer has a residual connection to the next layer. A residual connection can be integrated into various layers of neural networks to prevent network degradation resulting from the deep layer [34]. In the STCN, a residual connection is used to avoid the spatial correlation of the underlying network that affects the feature extraction of the upper network. Therefore, using a residual connection can improve the fitting ability of the STCN. The prior information is fully integrated into the STCN design to improve the interpretability of this model.

V. Numerical Results

A. Dataset

In this case study, historical power data of 15 wind farms are considered to train and evaluate the proposed power forecasting model for multiple wind farms. The 15 wind farms (wind farms A-O) are located in Australia, are shown in Fig. 4. The historical wind power data are obtained from the Australian Energy Market over 3 years from January 2018 to Febuary 2021 in 5-min intervals [35]. Each set of day data contains 288 data samples. Wind power data from 2018 to 2019 are treated as a training set. Wind power data from 2020 to 2021 are treated as a testing set to evaluate the model performance. All data are normalized during the training process.

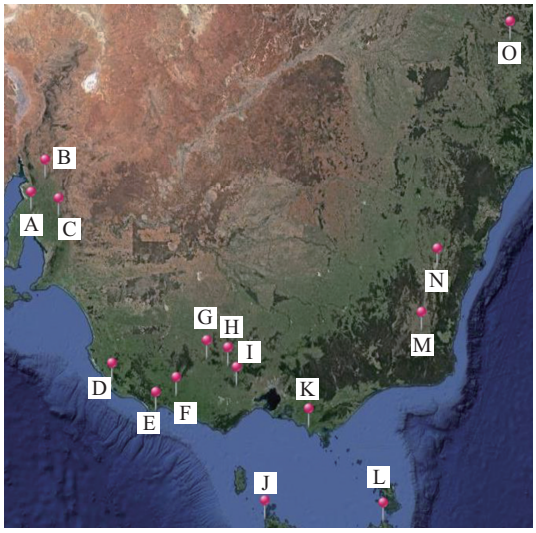

Fig. 4. Locations of 15 wind farms in Australia.

B. Evaluation Indices

Root mean square error (RMSE) and mean absolute error (MAE) are widely employed as evaluation indices of wind power forecasting.

$$RMSE = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{1}{C}\sum_{t=1}^{C}(\hat{P}_t^i - P_t^i)^2} \tag{11}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{C}\sum_{t=1}^{C}|\hat{P}_t^i - P_t^i|\right) \tag{12}$$

where $\hat{P}_t^i$ is the forecasting value of wind power at site i at time t; and C is the total number of test samples.

MAE and RMSE can be used as the indices of prediction accuracy and precision for short-term wind power forecasting, respectively. Accuracy refers to the difference between the forecasting and measured values. MAE can better reflect the actual situation of the forecasting error. Precision describes the stability of the forecasting error, and thus RMSE gives more weight to large residual samples. With similar MAEs, a smaller RMSE represents a more stable forecasting result [36], [37].

C. Model Architecture and Details of Training

The basic STCN model in this study uses three GCN-TCN layers, as shown in Fig. 3. The temporal dimension of input historical wind power data involves 12 steps to predict the wind power after one step. The kernel sizes and dilation coefficients of the three TCN layers are given as $\{(2\times 1,1),(2\times 1,2),(3\times 1,4)\}$, as shown in Fig. 1. Leaky-ReLU activation function is employed after the GCN and TCN layers with a coefficient of 0.2. ReLU activation is used in the output layer to ensure that the wind power forecasting value is greater than zero. Off-diagonal and diagonal elements of adjacent matrix $\boldsymbol{A}$ are represented by $1/d_{ij}$ and 0, respectively.

The model is implemented on a computer with an NVIDIA TITAN V graphics card and Intel Core i9-7900X 3.30 GHz. The proposed STCN model is designed using Python 3 and Pytorch 1.6.0 backend, CUDA 10.1, and cuDNN 7.5 libraries. The learning rate is set to be 10^{-3} and the number of training epochs is 100 with a batch size of 50. All models trained in this study use the Adam optimizer algorithm [38]. The autograd mechanism of PyTorch is used for gradient derivation. The detailed training process of the Adam optimizer algorithm is shown in Algorithm 1.

Algorithm 1: Adam optimizer algorithm

Input: input objective function with parameters $loss(\theta)$; initialize learning rate lr; model parameters θ_0; smoothing constant $\beta_1=0.9$ and $\beta_2=0.999$; the first-moment coefficient $m_0=0$; the second-moment coefficient $v_0=0$; $t=0$; $\varepsilon=10^{-8}$

While θ_t is not converged **do**

$t\leftarrow t+1$

Get gradients at timestep t: $g_t\leftarrow\nabla loss_t(\theta_{t-1})$

Update biased first-moment estimate: $m_t\leftarrow\beta_1 m_{t-1}+(1-\beta_1)g_t$

Update biased second-moment estimate: $v_t\leftarrow\beta_2 v_{t-1}+(1-\beta_2)g_t^2$

Correct the first-moment estimate: $\hat{m}_t\leftarrow m_t/(1-\beta_1)$

Correct the second-moment estimate: $\hat{v}_t\leftarrow v_t/(1-\beta_2)$

Update model parameters: $\theta_t\leftarrow\theta_{t-1}-lr\cdot\hat{m}_t/\left(\sqrt{\hat{v}_t}+\varepsilon\right)$

end

return θ_t

D. Comparison with State-of-the-art Methods

To verify the superiority of the proposed STCN, four models including the persistence (PSS) method, fully-connected network (FCN), RNN, and LSTM are adopted as baselines. Tables I and II show the average RMSEs and MAEs of 15 wind farms with 5-min-, 15-min-, 30-min-, 1-hour-, 2-hour-, and 3-hour-ahead power forecasting tasks, respectively. Thirty samples with 1-hour-ahead forecasting values and forecasting absolute errors for wind farm A on July 2-3, 2020, are shown in Fig. 5.

TABLE I
RMSEs OF FORECASTING MODELS

Model	RMSE with forecasting intervals (MW)					
	5-min	15-min	30-min	1-hour	2-hour	3-hour
PSS	4.835	9.696	13.519	17.902	24.557	29.198
FCN	5.407	9.652	13.384	17.806	23.972	27.819
RNN	6.576	10.783	14.057	17.743	23.219	27.067
LSTM	4.901	9.901	13.546	17.896	23.543	27.427
STCN	4.709	9.600	13.263	17.417	23.382	27.184

TABLE II
MAEs OF FORECASTING MODELS

Model	MAE with forecasting intervals (MW)					
	5-min	15-min	30-min	1-hour	2-hour	3-hour
PSS	2.653	5.343	7.856	11.216	16.194	19.420
FCN	3.351	5.367	7.941	11.205	16.088	19.116
RNN	4.471	6.750	9.130	12.422	16.692	19.391
LSTM	2.853	5.662	8.306	12.049	16.481	19.314
STCN	2.604	5.298	7.812	11.096	15.731	18.716

The PSS only considers the smoothness assumption in the time series and thus does not consider the spatial correlation. When the forecasting temporal interval is extended, the correlation of future wind power values with current values is decreased. As shown in Tables I and II, PSS produces accurate short-term wind power forecasting values under the 5-min-ahead condition. However, the corresponding RMSE and MAE increase at very fast rates as the forecasting temporal interval is extended.

The FCN only considers the temporal correlation but does not consider the spatial correlation. Because the temporal features are considered, the MAE and RMSE of FCN are smaller than those of PSS in 1- to 3-hour-ahead wind power forecasting.

For the 5-, 15-, and 30-min-ahead wind power forecasting of multiple wind farms, the MAEs and RMSEs of the RNN and LSTM are higher than those of the PSS and FCN. This is due to the recurrent network structure, which can extract the temporal features but cannot effectively process non-Euclidean data of multiple wind farms. Although the RMSE of RNN is smaller than that of any other test model in the 2-hour and 3-hour-ahead wind power forecasting, the corresponding MAE of RNN shows unsatisfactory results. This shows that RNN has a higher precision in this case, but is less accurate. As shown in Fig. 5, RNN and LSTM have higher absolute errors. Overall, LSTM is more stable and accurate than RNN.

STCN obtains the minimum MAEs and RMSEs in 5-min-, 15-min-, 30-min-, and 1-hour-ahead wind power forecasting of multiple wind farms. Although the RMSE of STCN is slightly more than RNN under the 2-hour and 3-hour-ahead conditions, when the MAE and RMSE are combined, the STCN performs the best. In the case of 3-hour-ahead forecasting, STCN is the only model with MAE less than 19 MW, which is much lower than those of other models. In summary, the STCN model achieves better results in power forecasting for multiple wind farms than the traditional models.

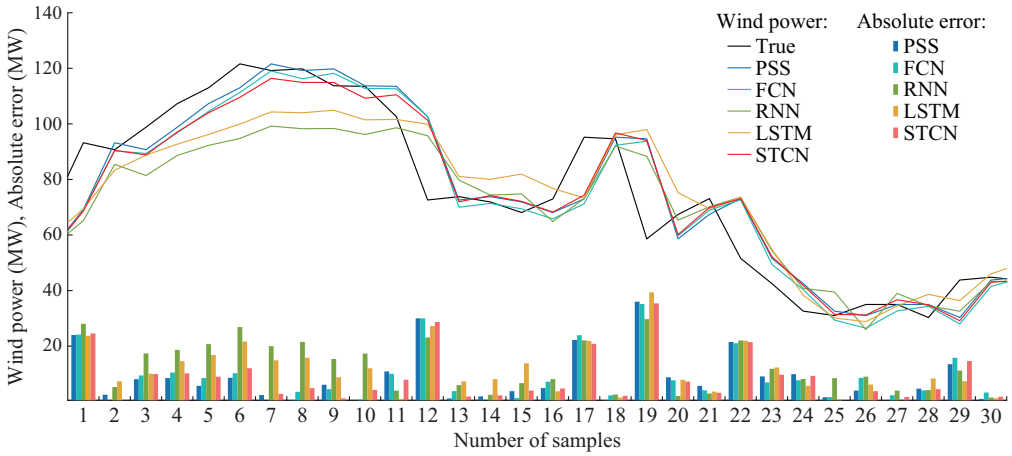

Fig. 5. Thirty samples with 1-hour-ahead forecasting values and forecasting absolute errors for wind farm A on July 2-3, 2020.

E. STCN Model Analysis

This section analyzes the effects of STCN on the accuracy and precision by adjusting the graph convolutional structure, different input predictors, and different season data.

To verify the efficacy of the directed graph structure and directed graph convolution, the following four graph convolutional models are integrated into the STCN for comparison: ① a constant Laplacian matrix with an undirected graph structure using a Chebyshev polynomial first-order approximation GCN as given in (7); ② a Laplacian matrix using an output matrix $\boldsymbol{\theta}_O$; ③ a Laplacian matrix using an input matrix $\boldsymbol{\theta}_I$; ④ a Laplacian matrix using an output matrix $\boldsymbol{\theta}_O$ and input matrix $\boldsymbol{\theta}_I$. Table III shows the average scores of RMSE and MAE in the 3-hour-ahead forecasting tasks.

The constant Laplacian matrix is unable to learn spatial correlation dynamically at various temporal scales. The model that only uses a single self-adaptive matrix performs better than the constant Laplacian matrix model in terms of

RMSE and MAE. It is shown that adding a self-adaptive matrix can improve the learning ability of spatio-temporal correlation. The STCN model includes two self-adaptive matrices, which achieve the lowest scores on two evaluation metrics.

TABLE III
AVERAGE RMSES AND MAES OF DIFFERENT LAPLACIAN MATRICES

Form	Laplacian matrix configuration	RMSE (MW)	MAE (MW)
Constant	$\boldsymbol{A}$	27.986	19.275
Output	$\boldsymbol{A}, \boldsymbol{\theta}_I$	27.523	18.811
Input	$\boldsymbol{A}, \boldsymbol{\theta}_O$	27.199	19.005
Output-input	$\boldsymbol{A}, \boldsymbol{\theta}_I, \boldsymbol{\theta}_O$	27.184	18.716

To analyze the impact of different wind farm predictors on the accuracy of wind power forecasting, the data from different numbers of wind farms are selected for model training and comparison with the forecasting results of three target wind farms (wind farms C, G, and M). Table IV shows the average RMSEs and MAEs in the testset for 3-hour-ahead wind power forecasting of target wind farms.

TABLE IV
AVERAGE RMSES AND MAES WITH DIFFERENT NUMBERS OF TARGET WIND FARM

Number	Wind farms	RMSE (MW)	MAE (MW)
3	C, G, M	28.330	19.464
5	D, J, C, G, M	28.215	19.460
10	A, D, F, J, K, L, O, C, G, M	27.648	19.203
15	All	27.513	18.847

As shown in Table IV, the increase in the number of observable wind farms reduces the MAEs and RMSEs with three target wind farms. When 15 observable wind farms are used as the input, the STCN model improves the MAEs and RMSEs by 3.17% and 2.88%, respectively. It is also verified that the power data of the other wind farms could be used as effective spatial input for target wind power forecasting to improve the forecasting accuracy.

The increase in spatial information can improve the accuracy of wind power forecasting. In this study, the wind power data with different time steps are input into the STCN model for analysis. Table V shows the kernel sizes and dilation coefficients of the TCN layers as well as the average RMSEs and MAEs of 1-hour-ahead wind power forecasting with different input time steps.

As shown in Table V, with four input time series, the forecasting accuracy is unsatisfactory. Fewer input time steps cause the STCN model to learn temporal features insufficiently, which in turn affects the forecasting results. With 12 and 6 input time series, the minimum values of RMSE and MAE are achieved, respectively. When the number of time steps increases, the RMSE and MAE become stable. Wind power data have a positive correlation in a short-term time series, but overlong input time steps have insignificant effects on forecasting results. With an increase in the number of input temporal predictors, the results of STCN tends to be stable and does not deteriorate, thus verifying the effectiveness of the skip and residual connections in the STCN model. The results thus show that the proposed model can flexibly extract wind power features at various temporal scales without being affected by overlong temporal predictors.

TABLE V
AVERAGE RMSES AND MAES WITH DIFFERENT INPUT TIME STEPS

Time steps	(Kernel size, dilation coefficient)	RMSE (MW)	MAE (MW)
4	(2×1,1), (2×1,2)	18.231	11.310
6	(2×1,1), (3×1,2)	17.616	11.018
8	(2×1,1), (2×1,2), (2×1,4)	17.606	11.012
12	(2×1,1), (2×1,2), (3×1,4)	17.417	11.096
18	(2×1,1), (3×1,2), (3×1,4)	17.829	11.030
24	(2×1,1), (3×1,2), (4×1,4)	17.739	11.014

Finally, wind power forecasting in different seasons is analyzed. PSS is used as a benchmark of wind power volatility to compare with STCN forecasting. Table VI shows the average RMSEs and MAEs of 1-hour-ahead wind power forecasting in different seasons.

TABLE VI
AVERAGE RMSES AND MAES OF 1-HOUR-AHEAD WIND POWER FORECASTING IN DIFFERENT SEASONS

Season	STCN		PSS	
	RMSE (MW)	MAE (MW)	RMSE (MW)	MAE (MW)
Spring	20.627	13.109	21.049	13.222
Summer	19.581	12.644	20.800	13.303
Autumn	17.660	11.236	18.486	11.667
Winter	16.757	10.375	17.821	11.073

As shown in Table VI, the STCN model has the smallest RMSEs and MAEs in winter. Spring and summer have larger forecasting errors. It can be shown that the wind power fluctuations of the PSS in spring and summer are significantly greater than those in autumn and winter. However, compared with the PSS, the STCN can improve the prediction accuracy in each season.

In summary, the STCN model can effectively process non-Euclidean structural data. The directed GCN with self-adaptive diagonal matrices can increase the accuracy and precision of forecasting results, and the GCN-TCN layers combined with skip and residual connections can identify spatio-temporal features at different temporal scales. The STCN model has high accuracy in power forecasting of multiple wind farms.

F. Model Complexity

The model complexity of deep learning models is a critical consideration when a new model is devised for real-world problems. In this section, we discuss two main aspects of model complexity, which are presented in Table VII.

In terms of the number of model parameters, FCN has the highest number as it fully connects all units between layers. The number of the parameters of RNN is significantly less

than that in the FCN. Regardless of the length of the sequential data, the RNN only requires a set of network parameters. In addition, the LSTM contains multiple gate units, and therefore the number of the parameters is much greater than that of the RNN. The proposed GCN contains two self-adaptive matrices, and therefore the number of parameters of the STCN is slightly more than that of the STCN with constant Laplacian matrix. More parameters mean the model is more complex. However, the FCN and LSTM do not show better results than the STCN. It should be noted that more parameters usually can learn more accurate forecasting information in model training, however, STCN obtains more accurate forecasting results with relatively fewer model parameters. Therefore, the structure of STCN is confirmed to be excellent.

TABLE VII
NUMBERS OF PARAMETERS AND COMPUTATION TIME OF DIFFERENT MODELS

Model	Number	Time (s)
FCN	3025	23.67
RNN	480	22.72
LSTM	1920	23.72
STCN (constant L)	567	69.54
STCN	657	88.33

For the forecasting task, the STCN requires longer training time than the other models. The STCN contains skip and residual connections that result in an increase in network structure complexity. The training time of the RNN and LSTM is longer than the FCN because a recurrent structure is not suitable for parallel acceleration. Compared with other models, the STCN improves the forecasting accuracy of multiple wind farm within the acceptable training time range.

G. Spatio-temporal Analysis

The directed graph convolutional structure represents the major contribution of the proposed model. Although the accuracy of the STCN model is higher than those of existing models, to show the effectiveness of spatial feature extraction for directed graph models, we use the Laplacian matrix of STCN for qualitative analysis of the spatial correlation. The non-diagonal elements in the matrix represent the spatial correlation. To show the non-diagonal elements more clearly, diagonal elements are set to be zero in this case. The resulting value could be normalized as:

$$\hat{L} = \frac{L - L_{\min}}{L_{\max} - L_{\min}} \tag{13}$$

where $L_{\max}$ and $L_{\min}$ are the maximum and minimum values of L, respectively.

Geographical distance is the prior knowledge necessary to reflect the spatial correlation of multiple wind farms. Geographical distance data between wind farms can be used to construct the adjacency matrix $\boldsymbol{A}$. However, the constant Laplacian matrix obtained by Chebyshev polynomial first-order approximation is symmetric, as shown in Fig. 6.

The Laplacian matrix is a major factor affecting spatial correlation in the GCN model. The constant Laplacian matrix is a symmetric matrix, which means that the spatial correlation between wind farms is equal. In addition, the constant Laplacian matrix does not change with the various temporal scales.

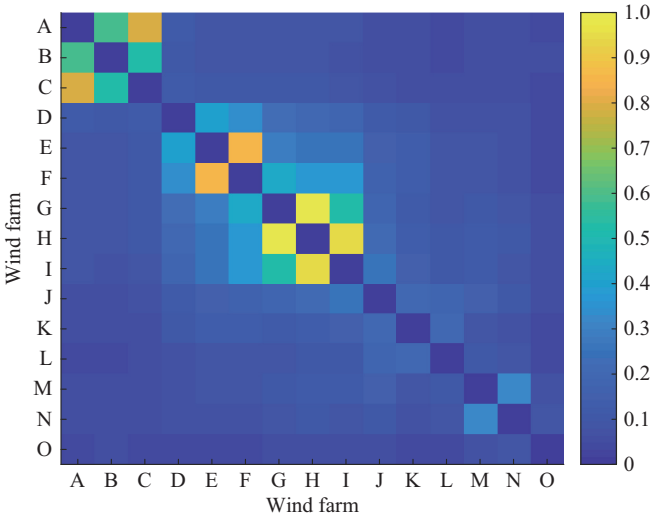

Fig. 6. Constant Laplacian matrix calculated by (7).

To qualitatively illustrate the asymmetry of the spatial correlation of multiple wind farms, the cross-correlation function (CCF) is introduced. CCF measures the similarity between time series and lagged versions of another time series, which is a function of the lag. Here, considering two time series $x(t)$ and $y(t)$, CCF at lag d is defined as:

$$CCF = \frac{\sum_t (x(t)-\bar{x})(y(t-d)-\bar{y})}{\sqrt{\sum_t (x(t)-\bar{x})^2}\sqrt{\sum_t (y(t-d)-\bar{y})^2}} \tag{14}$$

where $\bar{x}$ and $\bar{y}$ are the average values of the two time series.

CCF can be used to analyze the correlation of two wind farms at different time lags. Historical data of wind farms A, B, and C are used to calculate the CCFs, as shown in Fig. 7. The unit lag temporal interval is 3 hours. The correlation of A to C is higher than C to A. Therefore, in spatio-temporal forecasting, the historical data of A play a more important role in the forecasting of wind farm C. Similarly, the spatial correlation of A to B is higher than B to A at lag temporal interval 1-4. However, B and C have similar correlations at lag temporal interval 1-2.

It can be observed from Fig. 7 that the spatial correlation among wind farms is not completely symmetrical and equal and may change over time. The proposed directed graph convolution can describe the asymmetric spatial features. The Laplacian matrix is extracted for analysis, as shown in Fig. 8.

In the asymmetric Laplacian matrix, $L(i,j)$ denotes the spatial correlation of farms i to j. As shown in Fig. 8(a), L(A,B) is higher than L(B,A), L(A,C) is higher than L(C,A), and L(B,C) is similar to L(C,B). Qualitative spatial features are consistent with the CCF analysis. Note that the wind farms G, H, and I have weak spatial correlations in the 3-hour-ahead forecasting. However, Fig. 6 shows that they have a

close geographical distance. Considering that the temporal scale can affect the spatial correlation, the Laplacian matrix is extracted in the 1-hour-ahead wind power forecasting, as shown in Fig. 8(b). The spatial correlation of wind farms G, H, and I can be compared with Fig. 8(a). It can be considered that the wind farms are too close to each other, and neighboring wind farms cannot provide a more effective spatial correlation for 3-hour-ahead forecasting. However, in 1-hour-ahead forecasting, the neighboring wind farms have a high spatial correlation.

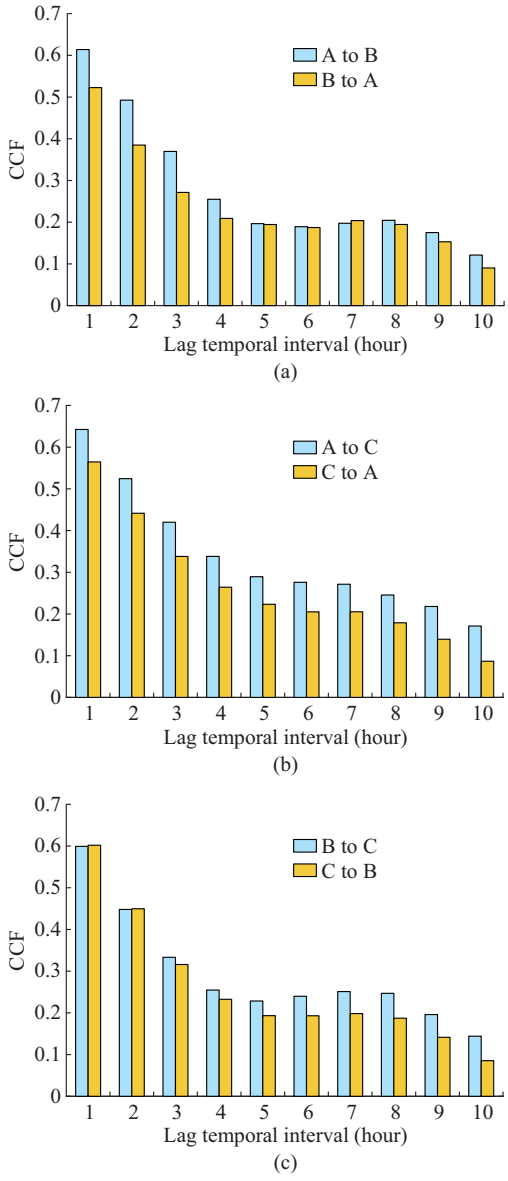

Fig. 7. CCFs among wind farms A, B, and C. (a) A and B. (b) A and C. (c) B and C.

The results demonstrate that the directed graph convolutional structure with adaptive parameters has more powerful extraction capabilities of spatio-temporal feature. Because of the symmetry of the adjacent matrix $\boldsymbol{A}$, the traditional spectral GCN cannot effectively describe the asymmetry features in spatial correlation. The constant Laplacian matrix cannot adapt to the spatial correlation of various temporal scales and may cause excessive smoothing after passing through the high-order GCN [39]. The STCN model can not only use adaptive parameters to extract asymmetric spatial correlation but also learn spatial correlation at various temporal scales.

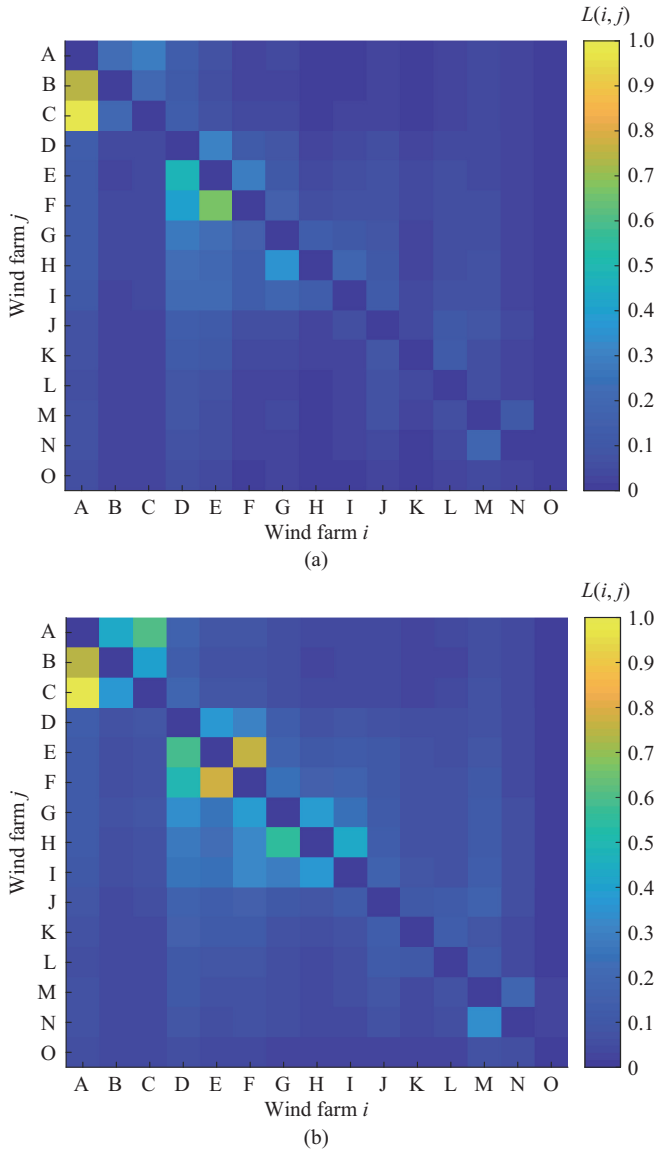

Fig. 8. Laplacian matrix extracts from STCN in the first GCN-TCN layer. (a) STCN is trained for 3-hour-ahead forecasting. (b) STCN is trained for 1-hour-ahead forecasting.

The STCN model can capture the complex nonlinearity spatio-temporal mapping of a time series of multiple wind farms. Compared with the undirected graph model, the directed graph model combines with adaptive parameters and thus has better adaptability and interpretability for spatio-temporal forecasting.

VI. Conclusion

In this study, a deep learning architecture STCN is proposed based on a graph model for spatio-temporal wind power forecasting. To effectively describe the asymmetry of the spatial correlation between multiple wind farms, a novel di-

rected graph model and the corresponding GCN structure are proposed. The directed GCN layer adds self-adaptive diagonal matrices for spatial feature extraction, and the TCN layer extracts temporal features of each wind farm. The multiple GCN-TCN layers extract spatio-temporal correlation of different temporal scales. Compared with the traditional GCN model, the novel model can not only describe the asymmetric spatial correlation but also adaptively learn the correlation characteristics at different temporal scales.

The superiority of the proposed model is demonstrated on real-world data in an example. Compared with the existing models, the proposed model improves the RMSEs and MAEs in power forecasting for multiple wind farms. In addition, it extracts asymmetric spatial correlation at various temporal scales more accurately. It is interesting to note that in this study, the STCN model exhibit certain interpretability, which is not available in traditional deep learning models.

The proposed model can improve the accuracy of power forecasting for multiple wind farms. However, wind farms in an area are usually owned by different companies and these companies are typically unwilling to share private data because of security considerations. In the follow-up study, a wind power forecasting framework will be investigated to ensure data security for multiple wind farms.

REFERENCES

[1] M. Bromer, H. Busch, T. Couture *et al.* (2021, Mar.). Renewables in cities: 2021 global status report. [Online]. Available: https://www.ren21.net/reports/global-status-report/

[2] H. Holttinen. (2018, Dec.). Design and operation of power systems with large amounts of wind power. [Online]. Available: https://iea-wind.org/wp-content/uploads/2021/08/T350.pdf

[3] G. Giebel and G. Kariniotakis, "Wind power forecasting–a review of the state of the art," *Renewable Energy Forecasting*, vol. 2017, pp. 59-109, Dec. 2017.

[4] M. Khalid and A. V. Savkin, "A method for short-term wind power prediction with multiple observation points," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1898-1899, May 2013.

[5] M. Lange and F. Ulrich, *Physical Approach to Short-term Wind Power Prediction*. New York: Springer, 2006.

[6] B. Candy, S. J. English, and S. J. Keogh, "A comparison of the impact of QuikScat and winds at wind vector products on met office analyses and forecasts," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 6, pp. 1632-1640, Jun. 2009.

[7] M. He, L. Yang, J. Zhang *et al.*, "A spatio-temporal analysis approach for short-term forecast of wind farm generation," *IEEE Transactions on Power Systems*, vol. 29, no. 4, pp. 1611-1622, Jul. 2014.

[8] M. Sun, C. Feng, and J. Zhang, "Conditional aggregated probabilistic wind power forecasting based on spatio-temporal correlation," *Applied Energy*, vol. 256, pp. 113842, Oct. 2019.

[9] Y. Zhao, L. Ye, 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, Sept. 2018.

[10] J. Tastu, P. Pinson, E. Kotwa *et al*., "Spatio-temporal analysis and modeling of short-term wind power forecast errors," *Wind Energy*, vol. 14, no. 1, pp. 43-60, Jan. 2011.

[11] J. Tastu, P. Pinson, P. J. Trombe *et al.*, "Probabilistic forecasts of wind power generation accounting for geographically dispersed information," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 480-489, Jan. 2014.

[12] Y. Zhang and J. Wang, "A distributed approach for wind power probabilistic forecasting considering spatio-temporal correlation without direct access to off-site information," *IEEE Transactions on Power Systems* , vol. 33, no. 5, pp. 5714-5726, Sept. 2018.

[13] M. C. Alexiadis, P. S. Dokopoulos, and H. S. Sahsamanoglou, "Wind speed and power forecasting based on spatial correlation models," *IEEE Transactions on Energy Conversion*, vol. 14, no. 3, pp. 836-842, Sept. 1999.

[14] Y. Chen, S. Zhang, W. Zhang *et al.*, "Multifactor spatio-temporal correlation model based on a combination of convolutional neural network and long short-term memory neural network for wind speed forecasting," *Energy Conversion and Management*, vol. 185, pp. 783-799, Feb. 2019.

[15] Q. Zhu, J. Chen, D. Shi *et al.*, "Learning temporal and spatial correlations jointly: a unified framework for wind speed prediction," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 1, pp. 509-523, Jan. 2020.

[16] Z. Wu, S. Pan, F. Chen *et al.*, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4-24, Jan. 2021.

[17] F. Scarselli, M. Gori, A. C. Tsoi *et al.*, "The graph neural network model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61-80, Jan. 2009.

[18] J. Yu and J. Gu, "Real-time traffic speed estimation with graph convolutional generative autoencoder," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3940-3951, Oct. 2019.

[19] Y. Qi, Q. Li, H. Karimian *et al.*, "A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory," *Science of The Total Environment*, vol. 664, pp. 1-10, May 2019.

[20] M. Khodayar and J. Wang, "Spatio-temporal graph deep neural network for short-term wind speed forecasting," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 670-681, Apr. 2019.

[21] T. Stańczyk and S. Mehrkanoon. (2021, Jan.). Deep graph convolutional networks for wind speed prediction. [Online]. Available: https://arxiv.org/abs/2101.10041v1

[22] K. Chen, J. Hu, Y. Zhang *et al.*, "Fault location in power distribution systems via deep graph convolutional networks," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 1, pp. 119-131, Jan. 2020.

[23] M. Khodayar, S. Mohammadi, M. E. Khodayar *et al.*, "Convolutional graph autoencoder: a generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 571-583, Apr. 2020.

[24] P. Li, J. Wu, X. Guan *et al.*, "Probabilistic forecasting of aggregated generation for regional wind farms with geographical dynamic model," in *Proceedings of IEEE PES General Meeting*, San Diego, USA, Jul. 2012, pp. 1-8.

[25] P. Li, X. Guan, J. Wu *et al.*, "Modeling dynamic spatial correlations of geographically distributed wind farms and constructing ellipsoidal uncertainty sets for optimization-based generation scheduling," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1594-1605, Oct. 2015.

[26] D. Chai, L. Wang, and Q. Yang, "Bike flow prediction with multi-graph convolutional networks," in *Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, Seattle, USA, Nov. 2018, pp. 397-400.

[27] J. Bruna, W. Zaremba, A. Szlam *et al.*, "Spectral networks and locally connected networks on graphs," in *Proceedings of 2th International Conference on Learning Representations (ICLR)*, Banff, Canada, Dec. 2013, pp. 1-14.

[28] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proceedings of 2016 Neural Information Processing Systems (NIPS)*, Barcelona, Spain, Dec. 2016, pp. 3844-3852.

[29] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proceedings of 5th International Conference on Learning Representations (ICLR)*, Toulon, France, Apr. 2017, pp. 1-14.

[30] S. Bai, J. Z. Kolter, and V. Koltun. (2018, Apr.). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. [Online]. Available: https://arxiv.org/abs/1803.01271

[31] Z. Wu, S. Pan, G. Long *et al.*, "Graph wavenet for deep spatial-temporal graph modeling," in *Proceedings of 2019 International Joint Conference on Artificial Intelligence (IJCAI)*, Macao, China, Aug. 2019, pp. 1907-1913.

[32] L. Xie, Y. Gu, X. Zhu *et al.*, "Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 511-520, Jan. 2014.

[33] Y. Li, R. Yu, C. Shahabi *et al.*, "Diffusion convolutional recurrent neural network: data-driven traffic forecasting," in *Proceedings of 6th International Conference on Learning Representations (ICLR)*, Vancouver, Canada, Apr. 2018, pp. 1-14.

[34] K. He, X. Zhang, S. Ren *et al.*, "Deep residual learning for image rec-

ognition," in *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, Jun. 2016, pp. 770-778.
[35] Aneroid Energy. (2021, May). Wind power in the Australian energy market. [Online]. Available: https://anero.id/energy/wind-energy
[36] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79-82, Dec. 2009.
[37] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247-1250, Jun. 2014.
[38] D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization," in *Procedings of 3rd International Conference on Learning Representations (ICLR)*, San Diego, USA, May 2015, pp. 1-15.
[39] Q. Li, Z. Han, and X. Wu, "Deeper insights into graph convolutional networks for semi-supervised learning," in *Proceedings of 32th AAAI Conference on Artificial Intelligence (AAAI)*, Louisiana, USA, Feb. 2018, pp. 3538-3545.

Xiaochong Dong received the B.S. degree from China Three Gorges University, Yichang, China, in 2017, and the M.S. degree from North China Electric Power University, Beijing, China, in 2020, all in electrical engineering. He is currently pursuing the Ph.D. degree in electrical engineering at North China Electric Power University. His research interests include artificial intelligence and optimization of power system.

Yingyun Sun received the B.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 1996, the M.S. and Ph.D. degrees from Tsinghua University, Beijing, China, in 2004 and 2007, respectively, all in electrical engineering. Presently, he is a Professor in the School of Electrical Engineering at North China Electric Power University, Beijing, China. His research interests include artificial intelligence, renewable energy, and power system analysis.

Ye Li received the M.S. degree at North China Electric Power University in 2015, Beijing, China, and has worked in Artificial Intelligence Application Research Section of China Electric Power Research Institute (CEPRI), Beijing, China, until now. She is the Secretary of IEEE PES Sub-committee on Power System Communications and Artificial Intelligence. Her research interests include distributed energy optimization scheduling and electric power artificial intelligence.

Xinying Wang received the Ph.D. degree at Dalian University of Technology, Dalian, China, in 2015, and has worked in China Electric Power Research Institute (CEPRI), Beijing, China, until now. He is the Deputy Director of Artificial Intelligence Application Research Section of CEPRI and the member of CSEE. His research interest includes electric artificial intelligence.

Tianjiao Pu received the B.E. and M.Sc. degrees from Tianjin University, Tianjin, China, in 1992 and 1997, respectively. From 1997 to now, he is a Professor Senior Engineer of China Electric Power Research Institute (CEPRI), the Fellow of IET, the Senior Member of IEEE, the Senior Member of Chinese Society for Electrical Engineering, the CIGRE Member, Secretary General of the AI Committee of CSEE, the Deputy Editor-in-Chief of IET Smart Grid. His research interests include power dispatching automation, smart grid simulation, active distribution network, and artificial intelligence.