

Spatial-Temporal Wind Power Probabilistic Forecasting Based on Time-Aware Graph Convolutional Network

Jingwei Tang , Zhi Liu , and Jianming Hu

Abstract—Spatial-temporal wind power prediction is of enormous importance to the grid-connected operation of multiple wind farms in the wind power system. However, most of the conventional methods are usually limited to predicting an individual wind farm’s power, and thus lack enough effectiveness of wind power forecasting of multiple adjacent wind farms. This paper proposes a novel spatial-temporal wind power probabilistic prediction approach, named ZF-GCN-MHTQF, based on time zigzags and flexible convolution at graph convolutional network, point-wise loss function and the heavy-tailed quantile function. The proposed framework combines the advantages of time zigzags and flexible convolution at graph convolutional networks that can extract temporally conditioned topological information from multiple wind farms efficiently and incorporate the extracted topological information to predict wind power. At the same time, the proposed method incorporates the strengths of point-wise loss functions and heavy-tailed quantile functions which can effectively tackle the problem of the traditional multi-quantile regression and accurately capture the full conditional distribution information of wind power. In our experiments, two real-world wind power datasets from Australia are utilized to validate the proposed model. Numerical experiments demonstrate the effectiveness and robustness of the proposed method compared to the state-of-the-art spatial-temporal models.

Index Terms—Time zigzag, flexible convolution, graph convolutional network, heavy-tail quantile function.

NOMENCLATURE

Indices

$\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_T$ Sequence of graph.

Manuscript received 12 April 2023; revised 21 September 2023 and 12 March 2024; accepted 6 April 2024. Date of publication 16 April 2024; date of current version 21 June 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 72071053 and in part by the Joint Foundation of Guangzhou City and Guangzhou University under Grant 202201020228. Paper no. TSTE-00362-2023. (Corresponding author: Jianming Hu.)

Jingwei Tang is with the Department of Mathematics, Faculty of Science and Technology, University of Macau, Macau 999078, China (e-mail: tongging-wai@163.com).

Zhi Liu is with the Department of Mathematics, Faculty of Science and Technology, University of Macau, Macau 999078, China, and also with Zhuhai-UM Science and Technology Research Institute, Zhuhai 519031, China (e-mail: liuzhi@um.edu.mo).

Jianming Hu is with the School of Mathematics and Statistics, Hainan University, Haikou 570228, China, and also with the College of Economics and Statistics, Guangzhou University, Guangzhou 510006, China (e-mail: hujm163@163.com).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TSTE.2024.3389023>.

Digital Object Identifier 10.1109/TSTE.2024.3389023

V, E, A_t	Nodes, the edges and the adjacency matrix of the graph.
N	Number of the nodes.
F	Number of features.
T	Time steps.
P	Window size.
t_b, t_d	Birth and death time of topological feature.
\hat{P}	Laplacian link.
$\theta_T^{(\ell)}$	Node embedding.
τ	Quantile level.
A	Positive constant.
Q	Length of the quantiles.
$\nabla \cdot_M f(X_t; \theta)$	Divergence with respect to monotonic variables.
$X^t \in \mathbb{R}^{N \times F}$	Feature matrix at time step t.
y	True value.
q_τ	τ -quantile prediction.
<i>Function</i>	
$G(x; \mu_X, \sigma_X^2)$	Gaussian density function with the mean and variance (μ_X, σ_X^2).
$w_{gauss}(x, y)$	Anisotropic Gaussian mask.
$\phi(x, y)$	Multiplicative filter network.
$g_j((x, y), \theta^{(i)})$	Anisotropic Gabor function.
$FlexConv(x, y)$	Flexible-size continuous kernel convolution.
$ZPI(\varphi_t)$	Zigzag persistence image.
ZPD'	Transformed multi-set in zigzag persistence diagram.
ZPD	Zigzag persistence diagram. diagram (ZPD).
$h(\omega)$	Weight function with mean ω .
$S_t^{(\ell)}$	ℓ th layer of the spatial graph convolution.
M	Power level of Laplacian.
P^m	Power series of normalized Laplacian.
$\theta_S^{(\ell)}$	Node embedding.
$T_t^{(\ell)}$	ℓ th layer of the temporal graph convolution.
$C_t^{(\ell)}$	ℓ th layer of spatial-temporal graph convolution.
u_t	Update gate.
r_t	Reset gate.
h_t	Output of the GRU.
$Q(\tau)$	Heavy-tailed quantile function.
Z_τ	τ -quantile of the standard normal distribution.

$\mathcal{L}_{\text{MHTQF}}$	Monotonic multi-quantile loss function of the HTQF.
$L_\tau(y, q_\tau)$	Quantile loss function.
$\mathcal{L}_{\text{HTQF}}$	Multi-quantile loss function of the HTQF.
<i>Parameters</i>	
$\theta^{(i)}, \mathbf{W}^{(i)}, \mathbf{b}^{(i)}$	Parameters, weight and bias of the multiplicative filter network.bm []'.
ω, σ^2	Mean and variance of the zigzag persistence diagram.
$\mathbf{W}_S^{(\ell)}$	Weight of the spatial graph convolution.
$\mathbf{W}_T^{(\ell)}$	Weight of the temporal graph convolution.
$\mathbf{W}_u, \mathbf{W}_r, \mathbf{W}_h$	Weight of the GRU.
$\mathbf{b}_u, \mathbf{b}_r, \mathbf{b}_h$	Bias of the GRU.
μ	Location parameter of the HTQF.
σ	Scale parameter of the HTQF.
u, v	Shape parameters of the HTQF.

I. INTRODUCTION

WIND energy is hailed as a clean renewable energy resource with no pollution in its application. The Global Wind Energy Council (GWEC) reports that 78 GW of onshore and offshore wind turbines are installed in 2022, bringing the total capacity up to 906 GW [1]. However, the high stochastic and intermittent of wind energy resources may lead to large power fluctuations and an immense challenge to the secure and efficient operation of the power grid. Wind power prediction is a critical step in finding the most cost-effective solution for the power grid operation [2]. In order to alleviate the adverse effects of wind power integration on the power grid, exact and efficient wind power predictions are imperative for the reliable management of wind farms and power systems, which could effectively mitigate the associated risks, thereby enhancing economic and social benefits [3], [4].

In the last few years, with the rapid development of machine learning, scholars have paid more and more attention to machine learning methods. These methods can effectively fit nonlinear relationships and are widely applied in time series forecasting. The common models include artificial neural networks (ANNs) and support vector machines (SVMs), etc. For instance, Li et al. [5] proposed a novel model combining SVM and an improved dragonfly algorithm to predict short-term wind power, which achieved better performance than existing models such as back propagation neural network and Gaussian process regression. Shahid et al. [6] utilized long short term memory (LSTM) to automatically learn temporal features from wind power and applied the genetic algorithm to optimize the hyperparameter in LSTM for wind power prediction. Li et al. [7] applied a hybrid improved cuckoo search arithmetic to optimize the hyper-parameters of the support vector regression machine which effectively improves the accuracy of short-term wind power prediction. Hong et al. [8] applied the convolutional neural Network (CNN) to learn local temporal characteristics of wind power, considering the uncertainty of wind speed for increasing the accuracy of 24-hour-ahead wind power forecasting. While the aforementioned methods effectively capture certain

temporal characteristics from historical data and enhance wind power prediction performance, they overlook the consideration of spatial-temporal correlations among multiple wind farms. Furthermore, these methods lack the capability to simultaneously extract spatial-temporal information from multiple wind farms, limiting their effectiveness in predicting spatial-temporal wind power.

In order to simultaneously extract spatial and temporal information for multiple wind farms' historical data, some spatial-temporal models based on graph neural network are proposed for predicting wind power. Khodayar and Wang [9] build a scalable graph convolutional deep learning architecture to extract the spatial-temporal information from wind energy data for predicting wind speed prediction. Yu et al. [10] proposed a new fast graph convolutional model based on fast graph convolutional kernel and particular pooling operation for short-term wind power prediction. Song et al. [11] combined graph convolutional network and multi-resolution convolution neural network to forecast wind power. Liao et al. [12] integrated the graph convolutional network and long short-term memory to extract the spatial-temporal features simultaneously for wind power forecasting. Although these GCN methods demonstrate outstanding performance in wind energy prediction, the structure of the spatial-temporal graph convolutional network relies primarily on the spatial-temporal correlation of explanatory variables and a pre-defined graph architecture, neglecting the inclusion of time-conditioned topological information. Additionally, there is a deficiency in methods for effectively extracting and learning time-conditioned topological information.

In the prediction research discussed above, most of the research focuses on the method of point prediction, but in order to better quantify the uncertainty of wind power, improving the robustness and reliability of wind power prediction, probabilistic prediction (uncertainty prediction) urgently needs research and exploration. Quantile regression is a common method of probabilistic prediction and has been widely applied in recent years. He and Wang [13] implemented LASSO-quantile regression neural network to predict and analyze wind power. Zhang et al. [14] used the quantile regression gated recurrent unit network to predict and quantify the uncertainty of wind energy. He et al. [15] build a novel prediction method of probability density for wind and solar power which combines cubic spline interpolation and support vector quantile regression. Hu et al. [16] combined a conformalized quantile regression method with the temporal convolutional network to construct conformal quantile prediction for wind power without any distributional assumptions.

Furthermore, compared with single-quantile forecasts, multi-quantile forecasts can better describe the distribution of wind power and quantify the uncertainty of the forecasts. However, most of the existing studies lack in-depth exploration of the quantile monotonicity and mainly utilize the form of separate prediction, which causes the issue of quantile crossing, that is, for some multi-quantile regression, the quantile regression values do not increase with increasing the quantile levels. This phenomenon violates the strictly monotonic property of the conditional quantile function. Also, the more predictive quantile

levels required, the more problematic quantile crossing will be, since more crossover space will be provided for the multiple quantile regression curves [17]. The simultaneous estimation method can effectively solve the quantile crossing problem. In simultaneous estimation [18], the multi-quantile regression is constructed and estimated at the same time, and additional constraints are added to the parameter optimization process to avoid crossing problems. This method not only does not depend on the estimation order of quantiles but also achieves the effect of borrowing power from multiple quantile regression, which can improve the overall performance of the model and achieve better forecasting results than separate estimation [19]. Thence, Muggeo et al. [20] introduced monotonic constraints in addition to non-cross constraints in the quantile regression, since the growth curve should increase monotonically with the age of the organism. Likewise, Roth et al. [21] developed a nonlinear monotonic quantile regression model to predict the rainfall extremes that have non-decreasing trends. Hu et al. [22] newly proposed a joint quantile model based on the composite kernel to simultaneously estimate the multi-quantile regression of wind power. Wang et al. [37] introduced a novel deep learning model for wind power forecasting, addressing the challenges of effective feature extraction and loss function design by using maximal information coefficient for input selection, a multi-convolutional neural network with multi-scale information fusion, and an asymmetric Laplace-based loss function to handle uncertainty. Liu et al. [38] presented a transfer learning-based approach for probabilistic wind power forecasting, combining multi-layer extreme learning machines, particle swarm optimization, and joint distribution adaptation to improve accuracy and nonlinear fitting compared to other methods, demonstrated through testing on Chinese wind farms. Takeuchi et al. [23] proposed a nonparametric kernel quantile regression model similar to SVM, which considered constraints such as non-crossing and monotonic. Whereas, the implementation of the kernel quantile regression model [24] is computationally expensive, with the complexity of cubic of sample size. The above methods effectively alleviate the crossover problem of multi-quantile prediction, but these methods still have some issues, including an increasing number of parameters required for estimating additional quantiles and the lack of interpretability.

To solve the above problems, this paper proposes a novel spatial-temporal probabilistic prediction method for simultaneously predicting multiple quantiles of wind power. The proposed method is based on the novel time zigzags and flexible convolution at graph convolutional network, and the monotonic heavy-tailed quantile function. Firstly, the proposed model uses a flexible convolutional layer to extract the temporally conditioned topological information and incorporate this information into the graph convolutional network. Then, the spatial and temporal graph convolutional network is applied to learn spatial-temporal information between multiple wind farms. Finally, the pointwise loss function and the heavy-tailed quantile function are combined as the loss function for enforcing model monotonicity and adjustably controlling the complex distribution property of wind power. The main contributions of this paper are presented as follows:

- 1) As far as we know, this is the first time that the flexible size continuous kernel convolution is applied to extract and learn time-aware persistent homology information from spatial-temporal wind power data.
- 2) For the complex properties of wind power, such as time-varyingness and asymmetric heavy-tail, a novel heavy-tailed quantile function that has a limited number of parameters to be estimated is used to quantify and analyze uncertainty about wind power predictions.
- 3) To address the quantile crossing problem, quantile monotonic information is incorporated into the input structure, and the monotonic quantile loss function is utilized to change the learning process of the model. By applying soft constraints to the loss function, the model can account for quantile monotonicity during the learning process.
- 4) A novel time-aware graph convolutional network multi-quantile regression model for spatial-temporal wind power is proposed. This model allows for the quantification and analysis of uncertainty in spatial-temporal wind power predictions and overcomes the disadvantages of traditional quantile regression.

The rest of this paper is organized as follows: Section II briefly describes the principle of the proposed spatial-temporal model, Section III introduces the principle of the monotonic heavy-tailed quantile function, Section IV shows the results of the experiments and the corresponding analyzes, and the conclusion is presented in Section V.

II. TIME ZIGZAGS AND FLEXIBLE CONVOLUTION AT GRAPH CONVOLUTIONAL NETWORK

A. Preliminaries

The spatial-temporal networks of the wind farms can be denoted as a sequence of graphs, $\{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_T\}$, where $\mathcal{G}_t = \{V, E, A_t\}$, denotes the graph at time step t , $t = 1, 2, \dots, T$. In the graph \mathcal{G}_t , V is a set of N nodes, and E is a set of edges that indicate the connectivity between nodes, $A_t \in \mathbb{R}^{N \times N}$ represent the adjacency matrix of \mathcal{G}_t , that is $A_{ij}^t > 0$ for any $e_{ij} \in E$ and $A_{ij}^t = 0$, otherwise. Hence, let $\mathbf{X}^t \in \mathbb{R}^{N \times F}$ be the feature matrix at time step t , where F is the number of features associated with each node. In this paper, the purpose is to predict the wind power of the future T time steps, given the past wind power data and graph with window size P , as follows:

$$\hat{\mathbf{X}}^{(t+1):(t+T)} = \mathcal{F}(\mathbf{X}^{(t-P+1):t}; \mathcal{G}_{(t-P+1):t}), \quad (1)$$

where, $\mathcal{F}(\cdot)$ is a trainable model.

B. Flexible Size Continuous Kernel Convolution (FLEXCONV)

In this section, we introduce the FlexConv [39] which is defined as the multiplicative of a continuous convolutional kernel with an anisotropic Gaussian mask of local support. Moreover, the continuous convolutional kernel is parameterized by the Multiplicative Anisotropic Gabor networks and the anisotropic Gaussian mask parameterizes its size.

Anisotropic Gaussian Mask: The Gaussian density function with the mean and variance (μ_X, σ_X^2) is defined as:

$$G(x; \mu_X, \sigma_X^2) := \exp\left(-\frac{(x - \mu_X)^2}{2\sigma_X^2}\right). \quad (2)$$

The anisotropic Gaussian mask is defined as follows:

$$\begin{aligned} w_{gauss}(x, y; \{\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2\}) &= G(x; \mu_X, \sigma_X^2) \\ &\cdot G(y; \mu_Y, \sigma_Y^2). \end{aligned} \quad (3)$$

So the anisotropic non-centered windows can be trained by learning (μ_X, σ_X^2) and (μ_Y, σ_Y^2) .

Multiplicative Anisotropic Gabor Networks: The multiplicative anisotropic Gabor networks is based on the multiplicative filter network which is proposed by Fathony et al. [25]. Specifically, the multiplicative filter network (MFN) is defined as follows:

$$\begin{aligned} \mathbf{z}^{(1)} &= \mathbf{g}([x, y]; \boldsymbol{\theta}^{(1)}), \\ \mathbf{z}^{(i)} &= (\mathbf{W}^{(i)} \mathbf{z}^{(i-1)} + \mathbf{b}^{(i)}) \circ \mathbf{g}([x, y]; \boldsymbol{\theta}^{(i)}), \\ \phi(x, y) = \mathbf{z}^{(L)} &= \mathbf{W}^{(L)} \mathbf{z}^{(L-1)} + \mathbf{b}^{(L)}, \end{aligned} \quad (4)$$

where, $i = 2, \dots, L-1$, \circ is the Hadamard product, $\boldsymbol{\theta}^{(i)}, \mathbf{W}^{(i)} \in \mathbb{R}^{d(i) \times d(i-1)}$, $\mathbf{b}^{(i)} \in \mathbb{R}^{d(i)}$ denote parameters, weight and bias of the i th layer, respectively. $\mathbf{g}(\cdot; \boldsymbol{\theta}^{(i)}) = [g_1, g_2, \dots, g_{d(i)}]^T$ is the nonlinear function and the most common choice is the isotropic Gabor function:

$$\begin{aligned} g_j([x, y]; \boldsymbol{\theta}^{(i)}) &= \exp\left(-\frac{\gamma_j^{(i)}}{2} \left[(x - \mu_j^{(i)})^2 + (y - \mu_j^{(i)})^2 \right]\right) \\ &\cdot \sin\left(\omega_j^{(i)} \cdot [x, y] + \beta_j^{(i)}\right), \end{aligned} \quad (5)$$

with the parameters $\boldsymbol{\theta}^{(i)} = \{\boldsymbol{\gamma}^{(i)} \in \mathbb{R}^{d(i)}, \boldsymbol{\mu}^{(i)} \in \mathbb{R}^{d(i)}, \boldsymbol{\omega}^{(i)} \in \mathbb{R}^{d(i) \times 2}, \boldsymbol{\beta}^{(i)} \in \mathbb{R}^{d(i)}\}$, where, $\boldsymbol{\gamma}_j$ denotes the scale term of the j th Gabor function, $\mu_j^{(i)}$ denotes the mean of the j th Gabor function, and $j = 1, 2, \dots, d(i)$.

The isotropic Gabor functions with the same γ for the horizontal and vertical directions are improper as the basis function for MFN. When frequency only needs to be along a certain direction, it will be along both directions at the same time. So the ability of the MFN is not optimally utilized. This question can be solved by applying the anisotropic Gabor functions instead:

$$\begin{aligned} g_j([x, y]; \boldsymbol{\theta}^{(i)}) &= \exp\left\{-\frac{1}{2} \left[\left(\gamma_{X_j}^{(i)} (x - \mu_{X_j}^{(i)})\right)^2 \right.\right. \\ &\quad \left.\left. + \left(\gamma_{Y_j}^{(i)} (y - \mu_{Y_j}^{(i)})\right)^2 \right]\right\} \\ &\cdot \sin\left(\omega_j^{(i)} [x, y] + \beta_j^{(i)}\right), \end{aligned} \quad (6)$$

$$\begin{aligned} \boldsymbol{\theta}^{(i)} &= \left\{ \boldsymbol{\gamma}_X^{(i)} \in \mathbb{R}^{d(i)}, \boldsymbol{\gamma}_Y^{(i)} \in \mathbb{R}^{d(i)}, \boldsymbol{\mu}_X^{(i)} \in \mathbb{R}^{d(i)}, \right. \\ &\quad \left. \boldsymbol{\mu}_Y^{(i)} \in \mathbb{R}^{d(i)}, \boldsymbol{\omega}^{(i)} \in \mathbb{R}^{d(i)}, \boldsymbol{\beta}^{(i)} \in \mathbb{R}^{d(i)} \right\}. \end{aligned}$$

Based on the multiplicative anisotropic Gabor Networks and anisotropic Gaussian mask, we can construe the flexible-size continuous kernel convolution as follows:

$$FlexConv(x, y) = w_{gauss}(x, y; \theta_{mask}) \circ \phi(x, y), \quad (7)$$

where, $w_{gauss}(x, y; \theta_{mask})$ is the anisotropic gaussian mask (2), $\phi(x, y)$ is the multiplicative anisotropic Gabor networks which are obtained from the combination of the MFN (4) and the anisotropic Gabor functions (6). The Flexconv provides better control over the frequency components of the introduced approximation and exhibits important improvements in descriptive power and convergence speed.

C. Time Zigzags and FlexConv At Graph Convolutional Network (ZF-GCN)

Our proposed model is based on the time zigzags at graph convolutional networks [26]. The ZF-GCN easily exacts and learns the long-term topological information of the zigzag persistence image, using FlexConv instead of the CNN and pooling. The proposed ZF-GCN is illustrated in Fig. 1, which is composed of zigzag persistence image, FlexConv, Spatial-Temporal Graph Convolutional (STGC) layer and gated recurrent unit (GRU) layer. The following outlines the details of the architecture.

First, we learn the topological features of graphs through Zigzag Persistence Image (ZPI) and FlexConv,

$$ZPI(\varphi_t) = \iint_{\varphi} \sum_{\omega \in ZPD'} h(\omega) \exp\left\{-\frac{\|\varphi_t - \omega\|^2}{2\sigma^2}\right\} d\varphi_{x_t} d\varphi_{y_t}, \quad (8)$$

where, ZPD' is the transformed multi-set in zigzag persistence diagram (ZPD), i.e., $ZPD'(x, y) = (x, y - x)$; $ZPD = \{(t_b, t_d) \in \mathbb{R}^2 | t_b < t_d\}$, t_b and t_d are the birth and death time of each topological feature, respectively. $h(\omega)$ is weight function with mean $\omega = (\omega_x, \omega_y) \in \mathbb{R}^2$ and variance σ^2 . Let

$$\mathbf{F}^{(\ell)} = FlexConv^{(\ell)}(ZPI(\varphi_t)), \quad (9)$$

where, $ZPI(\varphi_t)$ denotes the ZPI based on the feature matrix sequence $\mathbf{X}^{(t-P+1):t}$ with window size P . $FlexConv^{(\ell)}$ is the flexible size continuous kernel convolution in the ℓ -th layer, $\mathbf{F}^{(\ell)}$ represents the extracted topological features from ZPI by FlexConv.

Second, we apply the spatial graph convolution and temporal graph convolution to extract the spatial-temporal information of the feature matrix sequence, and then combine them together with the topological features. The Spatial Graph Convolution (Spatial GConv) can be summarized as follows:

$$\mathbf{S}_t^{(\ell)} = \left(\tilde{\mathbf{P}} \mathbf{S}_t^{(\ell-1)}\right)^{\top} \boldsymbol{\theta}_{\mathbf{S}}^{(\ell)} \mathbf{W}_{\mathbf{S}}^{(\ell)}, \quad (10)$$

where, $\tilde{\mathbf{P}} = [\mathbf{I}, \mathbf{P}, \mathbf{P}^2, \dots, \mathbf{P}^M] \in \mathbb{R}^{N \times N \times (M+1)}$ is Laplacianlink, M is the power level, \mathbf{I} denotes the identity matrix, \mathbf{P} represents the self-adaptive adjacency matrix [27], \mathbf{P}^m is the power series of normalized Laplacian, $\boldsymbol{\theta}_{\mathbf{S}}^{(\ell)} \in \mathbb{R}^{N \times e}$ is the node embedding, $\mathbf{W}_{\mathbf{S}}^{(\ell)} \in \mathbb{R}^{e \times (M+1) \times d_{in} \times d_{out}/2}$ is the learnable weight parameter (d_{in} and $d_{out}/2$ are the dimension of the input and output, separately). $\mathbf{S}_t^{(\ell-1)} \in \mathbb{R}^{N \times d_{in}}$ is the input of

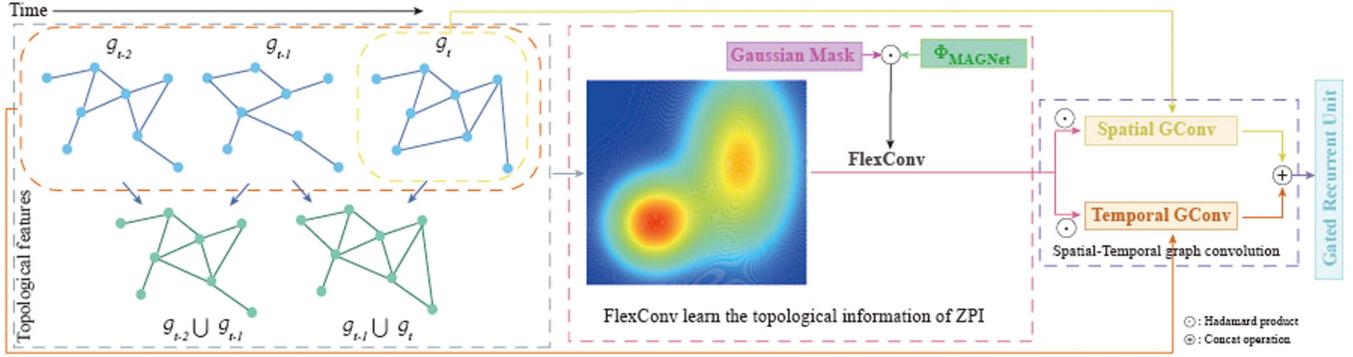


Fig. 1. Structure of time zigzags and FlexConv at graph convolutional network.

the ℓ th layer and $S_t^{(0)} = \mathbf{X}^t$. The Temporal Graph Convolution (Temporal GConv) can be summarized as follows:

$$\mathbf{T}_t^{(\ell)} = \left(\left(\tilde{\mathbf{P}} \mathbf{T}_t^{(\ell-1)} \right)^{\top} \boldsymbol{\theta}_T^{(\ell)} \mathbf{W}_T^{(\ell)} \right) \mathbf{L}^{(\ell)}, \quad (11)$$

where, $\mathbf{W}_T^{(\ell)} \in \mathbb{R}^{e \times d_{in} \times d_{out}/2}$ is the learnable weight, $\mathbf{L} \in \mathbb{R}^{P \times 1}$ is the learnable projection vector, $\boldsymbol{\theta}_T^{(\ell)} \in \mathbb{R}^{N \times e}$ is the node embedding, and $\mathbf{T}_t^{(\ell-1)} \in \mathbb{R}^{P \times N \times d_{in}}$ is the input of the ℓ -th layer and $\mathbf{T}_t^{(0)} = \mathbf{X}^{(t-P+1):t} \in \mathbb{R}^{P \times N \times F}$. Let

$$\mathbf{C}_t^{(\ell)} = \text{CONCAT}^{(\ell)} \left(\mathbf{S}_t^{(\ell)} \mathbf{F}^{(\ell)}, \mathbf{T}_t^{(\ell)} \mathbf{F}^{(\ell)} \right), \quad (12)$$

where, $\mathbf{S}_t^{(\ell)} \in \mathbb{R}^{N \times (d_{out}/2)}$ represents the output of the spatial graph convolutional layer, $\mathbf{T}_t^{(\ell)} \in \mathbb{R}^{N \times (d_{out}/2)}$ represents the output of the temporal graph convolutional layer, and the overall output $\mathbf{C}_t^{(\ell)} \in \mathbb{R}^{N \times d_{out}}$ combines $\mathbf{S}_t^{(\ell)}$ and $\mathbf{T}_t^{(\ell)}$ at time t .

Finally, we utilize GRU layer to extract the temporal dependence of spatial-temporal topological information and output predictions. The GRU, which is a simplified variant of long short-term memory, is proposed by Cho et al. [28]. Compared with LSTM, GRU has only two doors reset gate and update gate, and is easier to ameliorate the problem of vanishing and exploding gradients. The detailed structure of GRU is shown as follows:

$$\begin{aligned} \mathbf{u}_t &= f(\mathbf{W}_u [\mathbf{h}_{t-1}, \mathbf{C}_t] + \mathbf{b}_u), \\ \mathbf{r}_t &= f(\mathbf{W}_r [\mathbf{h}_{t-1}, \mathbf{C}_t] + \mathbf{b}_r), \\ \tilde{\mathbf{h}}_t &= g(\mathbf{W}_h [\mathbf{r}_t \circ \mathbf{h}_{t-1}, \mathbf{C}_t] + \mathbf{b}_h), \\ \mathbf{h}_t &= \mathbf{u}_t \circ \mathbf{h}_{t-1} + (1 - \mathbf{u}_t) \circ \tilde{\mathbf{h}}_t, \end{aligned} \quad (13)$$

where, $f(\cdot)$ is the activation function, $g(\cdot)$ is the tanh function, \circ is the Hadamard product, \mathbf{u}_t and \mathbf{r}_t are the update gate and reset gate respectively, $\mathbf{W}_u, \mathbf{b}_u, \mathbf{W}_r, \mathbf{b}_r, \mathbf{W}_h$ and \mathbf{b}_h are learnable weight, \mathbf{C}_t is the output of the Spatial-Temporal graph convolutional layer and the input of the GRU, \mathbf{h}_t is the output of the GRU.

III. THE MONOTONIC HEAVY-TAILED QUANTILE FUNCTION

A. Heavy-Tailed Quantile Function

Now, time series frequently demonstrate the properties of non-stationary, nonlinearity, and high dimensionality, accompanied by increasingly complex distribution such as asymmetric heavy tail and time-varyingness. Heavy-tailed quantile function (HTQF) is a novel parametric quantile function proposed by Yan [29]. According to properties of the normal distribution, we can know that $\mu + \sigma Z_\tau$ is the τ -quantile of the normal distribution $N(\mu, \sigma^2)$, where Z_τ is the τ -quantile of the standard normal distribution. Furthermore, the Q-Q plot of the normal distribution is a straight line, while the Q-Q plot of the heavy tail distribution is an inverted S shape. Therefore, the proposed HTQF utilizes two parameters to control the upper and lower tails of the inverted S shape and has a controllable shape in the Q-Q plot. The HTQF is defined as follows:

$$Q(\tau | \mu, \sigma, u, v) = \mu + \sigma Z_\tau \left(\frac{e^{uZ_\tau}}{A} + 1 \right) \left(\frac{e^{-vZ_\tau}}{A} + 1 \right), \quad (14)$$

where, μ is the location parameter, σ is the scale parameter, A is a positive constant, u and v are the shape parameters that control the up and down tails of the inverted S graph respectively. When $u = v = 0$, the HTQF is the quantile function of the normal distribution. It is evident that $\frac{e^{uZ_\tau}}{A} + 1$ is a monotonically increasing function of z_τ and, as $z_\tau \rightarrow -\infty$, $\frac{e^{uZ_\tau}}{A} + 1 \rightarrow 1$, so that $\frac{e^{uZ_\tau}}{A} + 1$ can control the up tail of the inverted S shape. The conclusion is similar for $\frac{e^{-vZ_\tau}}{A} + 1$. Furthermore, since the inverse function of HTQF exists and is a CDF, it has a unique probability distribution associated with it. In conclusion, the HTQF has the superiority to model complex distributions including asymmetry and time-varyingness.

B. Monotonic Quantile Loss Function

Changing the model's training process is a valid method to endue the model with monotonicity, which is more appropriate and convenient than changing the model architecture. These methods generally refer to increasing constraints during the back-propagation process, which can be 'soft constraints' that regularize the loss function, or 'hard constraints' similar to

TABLE I
DESCRIPTION OF DATASETS

	Dataset A	Dataset B
#Nodes	9	20
#Edges	32	81
Time interval	1 hour	1 hour
Time steps	37 224	37 224
Installation capacity	2224 MW	2351 MW

Lagrange multipliers. Research illustrates that soft constrain outperforms hard constrain, since hard constraints may generate a sub-optimal result [30]. Inspired by the point-wise loss function proposed by Gupta et al. [31], we construct a monotonic quantile loss function. Considering quantiles of length Q ($0 < \tau_1 < \tau_2 < \dots < \tau_Q < 1$) that have a monotonically increasing relationship with the response variable, the quantiles can be used as a part of the explanatory variables with monotonicity for multi-quantile regression. The monotonic multi-quantile loss function of the HTQF is defined as:

$$\mathcal{L}_{\text{MHTQF}} = \sum_{t=1}^T \max(0, -\nabla \cdot_M f(X_t; \theta)) + \mathcal{L}_{\text{HTQF}}, \quad (15)$$

where, $\nabla \cdot_M f(X_t; \theta)$ denotes the divergence with respect to monotonic variables, $f(\cdot; \theta)$ is the function with parameters θ . $f(\cdot; \theta)$ is the ZF-GCN model with the graph structure and the quantile loss function is as follows:

$$L_\tau(y, q_\tau) = \begin{cases} \tau|y - q_\tau|, & \text{if } y > q_\tau \\ (1-\tau)|y - q_\tau|, & \text{if } y \leq q_\tau \end{cases}, \quad (16)$$

where, y is the true value, q_τ is τ -quantile prediction based on HTQF. Moreover, the empirical heavy-tailed multi-quantile loss function can be expressed as follows:

$$\mathcal{L}_{\text{HTQF}} = \frac{1}{Q} \frac{1}{T} \sum_{q=1}^Q \sum_{t=1}^T L_{\tau_q}(y_t, Q(\tau_q | f(X_t; \theta))), \quad (17)$$

where, $Q(\tau_q | f(X_t; \theta)) = Q(\tau_q | \mu_t, \sigma_t, u_t, v_t)$, since $\mu_t, \sigma_t, u_t, v_t$ is the output of $f(X_t; \theta)$, Q is the number of the quantiles, T is the number of the sample size.

IV. EXPERIMENTS AND DISCUSSIONS

A. Datasets and Preprocessing

In order to demonstrate the validity and practicability of the proposed method, we conduct comparative experiments on two real spatial-temporal wind energy public datasets collected from New South Wales (Dataset A) and South Australia (Dataset B) at <https://anero.id/energy/wind-energy>, respectively. Both Datasets A and B are collected with 1-hour intervals, therefore there are overall 37224 samples selected from January 1st, 2018 to March 31st, 2022 without missing values. The edges are determined based on the geographical distance between each wind farm and the distance is less than the average distance between all wind farms. Table I describes the details of Datasets A and B. Fig. 2 depicts the geographic location of wind farms in Datasets A and B.

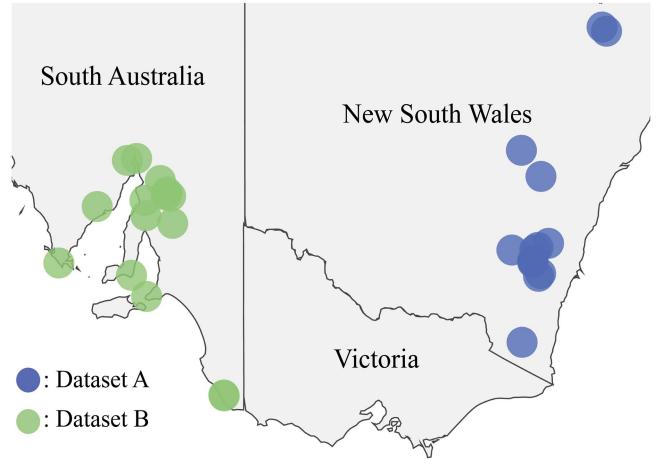


Fig. 2. Map with the wind farms: Blue represents Dataset A from New South Wales and Green represents Dataset B from South Australia.

All features of datasets are normalized in the range of $[0, 1]$ which eliminates the influence of the base unit and speeds up the algorithm's convergence. The min-max normalization method is used:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (18)$$

where, x is the feature, x_{\max} and x_{\min} are the maximum and minimum of x . Additionally, so as to capture spatial-temporal relationship, the adjacency matrix A_t of the graph \mathcal{G}_t at time t is constructed as:

$$A_{t,uv}^\lambda = \begin{cases} A_{t,ij} & (i,j) \in E \text{ and } A_{t,ij} \leq \lambda \\ 0 & (i,j) \in E \text{ and } A_{t,ij} > \lambda \\ 0 & (i,j) \notin E, \end{cases} \quad (19)$$

where, $A_{t,ij} = \exp(-\frac{\|x_{t,i} - x_{t,j}\|^2}{2\sigma^2})$ is the Radial Basis Function kernel which measures the similarity between the feature $x_{t,i}$ of the i -th node and the feature $x_{t,j}$ of the j -th node at time t . We set the parameter $\sigma^2 = 1$ and the thresholds $\lambda = 0.5$.

B. Experiment Settings

We divide the dataset into a train set and a test set using a 3:1 ratio. Simultaneously, we use the historical data of 24 hours for multi-quantile ($\tau = [0.1, 0.2, \dots, 0.9]$) wind power forecasting in the next 24 hours. To validate the effectiveness of time zigzags and flexible convolutional at graph convolutional network with monotonic heavy-tailed quantile function (ZF-GCN-MHTQF), we evaluate the median quantile prediction and probabilistic prediction performances of ZF-GCN-MHTQF on two wind power datasets versus 6 regression methods and 4 state-of-the-art benchmark models and integrate these methods and benchmark models to generate comparison models including:

- ZF-GCN-LR: time zigzags and flexible convolutional at graph convolutional network with linear regression;
- ZF-GCN-QR: time zigzags and flexible convolutional at graph convolutional network with quantile regression;

- ZF-GCN-MQR: time zigzags and flexible convolutional at graph convolutional network with monotonic quantile regression;
- ZF-GCN-HQR: time zigzags and flexible convolutional at graph convolutional network with Huber quantile regression;
- ZF-GCN-HTQF: time zigzags and flexible convolutional at graph convolutional network with heavy-tailed quantile function;
- ZF-GCN-MHQR: time zigzags and flexible convolutional at graph convolutional network with monotonic Huber quantile regression;
- Z-GCN-LR: time zigzags at graph convolutional network with linear regression;
- Z-GCN-QR: time zigzags at graph convolutional network with quantile regression;
- Z-GCN-MQR: time zigzags at graph convolutional network with monotonic quantile regression;
- Z-GCN-HQR: time zigzags at graph convolutional network with Huber quantile regression;
- Z-GCN-MHQR: time zigzags at graph convolutional network with monotonic Huber quantile regression;
- Z-GCN-HTQF: time zigzags at graph convolutional network with heavy-tailed quantile function;
- Z-GCN-MHTQF: time zigzags at graph convolutional network with monotonic heavy-tailed quantile function;
- GCN-QR: graph convolutional network with quantile regression;
- GCN-MHTQF: graph convolutional network with monotonic heavy-tailed quantile function;
- AGCRN-QR: adaptive graph convolutional recurrent network [32] with quantile regression;
- AGCRN-MHTQF: adaptive graph convolutional recurrent network with monotonic heavy-tailed quantile function;
- DAAGCN-QR: dynamic adaptive and adversarial graph convolutional network [33] with monotonic heavy-tailed quantile function;
- DAAGCN-MHTQF: dynamic adaptive and adversarial graph convolutional network with monotonic heavy-tailed quantile function;
- DSTAGNN-QR: dynamic spatial-temporal aware graph neural network [34] with monotonic heavy-tailed quantile function;
- DSTAGNN-MHTQF: dynamic spatial-temporal aware graph neural network with monotonic heavy-tailed quantile function;
- STDGCNN-MHTQF: spatiotemporal directed graph convolution network [35] with monotonic heavy-tailed quantile function;
- GGCN: gated graph convolutional network [36];
- AL-MCNN-BiLSTM: asymmetric Laplace multi-convolutional neural network and bidirectional long-short-term memory [37];
- MLELM: multi-layer extreme learning machine [38].

For median quantile prediction, the evaluation indicators are root mean square error (RMSE) and mean absolute error (MAE). And for probabilistic prediction, the evaluation

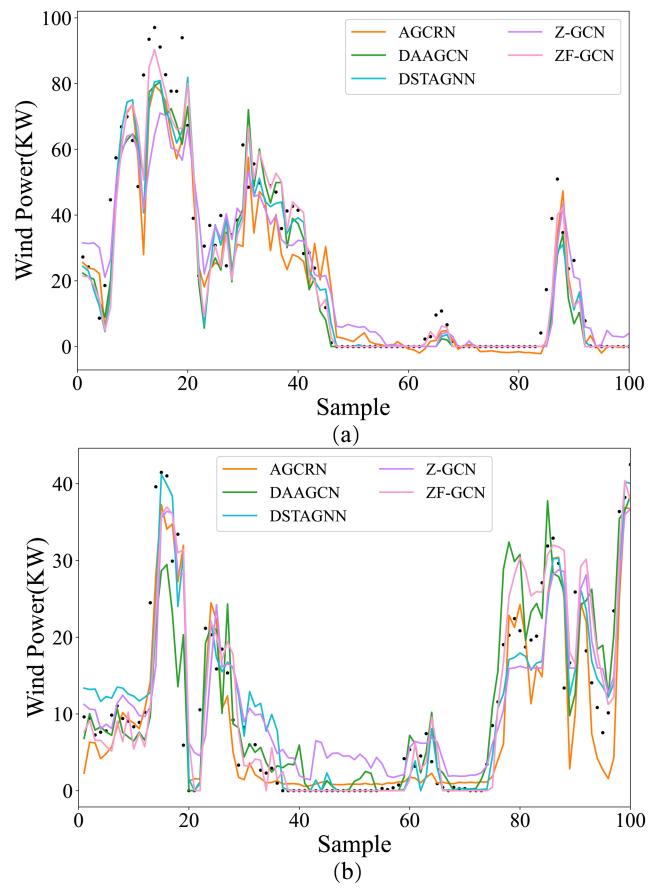


Fig. 3. Comparison of wind power median quantile prediction between benchmark models and the proposed model ZF-GCN with monotonic heavy-tailed quantile function on a snapshot of Datasets A and B: (a) the one-step-ahead forecast in the 24-step-ahead forecast for the 1st wind farm in Dataset A, and (b) the one-step-ahead forecast in the 24-step-ahead forecast for the 3rd wind farm in Dataset B.

indicators are cross-loss (CL), continuous ranked probability score (CRPS), prediction interval coverage probability (PICP) and mean prediction interval width (MPIW). Moreover, Diebold-Mariano test [40] (DM test) is a hypothesis testing method used to assess variations in prediction accuracy among different methods, relying on computed DM statistics to determine the significance of differences between these prediction methods. The above evaluation indicators are expressed as follows:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2}, \quad (20)$$

$$MAE = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t|, \quad (21)$$

$$CL = \frac{1}{T} \sum_{t=1}^T \sum_{\tau=2}^Q \max(0, \hat{y}_{t,\tau-1} - \hat{y}_{t,\tau}), \quad (22)$$

$$PICP = \frac{1}{T} \sum_{t=1}^T \mathbf{1}(\hat{y}_t \geq \hat{y}_{t,L}) \mathbf{1}(\hat{y}_t \leq \hat{y}_{t,U}), \quad (23)$$

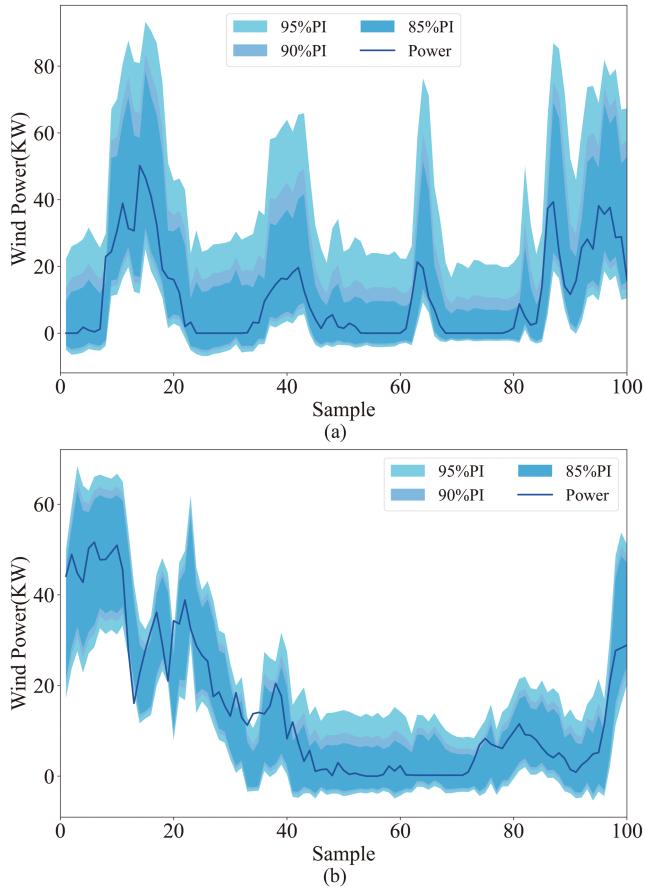


Fig. 4. Multiple prediction intervals of our ZF-GCN-MHTQF on a snapshot of Datasets A and B: (a) One-step-ahead multiple prediction intervals in the 24-step-ahead forecast of the first wind farm in Dataset A, and (b) one-step-ahead multiple prediction intervals in the 24-step-ahead forecast of the third wind farm in Dataset B.

$$MPIW = \frac{1}{T} \sum_{t=1}^T (\hat{y}_{t,U} - \hat{y}_{t,L}), \quad (24)$$

where, y_t is the true value at time t , \hat{y}_t is predictive value, T denotes the length of the time series, $\hat{y}_{t,\tau}$ represents τ quantile predicted value at time t , $\hat{y}_{t,U}$ and $\hat{y}_{t,L}$ are the upper bound and the lower bound of the prediction interval, respectively.

All experimental programs are edited with Python 3.9 and carried out on a Ubuntu server with a 2.30 GHz CPU, 64.0 GB RAM, and RTX 2080Ti GPU. Identify the optimal values for various hyperparameters through cross-validation on the training dataset. Hyper-parameters setting of the proposed model is indicated in Table II.

C. Comparison With the Benchmark Models

Table III shows the median quantile prediction evaluation metrics of our proposed model and benchmark models for two wind power 24-step-ahead forecasting tasks, where the best results are bold. From Table III, it is axiomatic that the proposed model ZF-GCN-MHTQF consistently outperforms all the benchmark models on Datasets A and B. The improvement gain of the

TABLE II
HYPER-PARAMETERS SETTING

Hyper-parameter	Quantity
Optimization	Adam
Batch size	256
Learning rate	0.05
Decay rate	0.98
Number of hidden layers	2
Hidden layer size	64
Epoch	1200
Embedding dimension	16

ZF-GCN-MHTQF over the next most accurate models with the same MHTQF methods ranges from 1.14% to 2.07% in MAE for Dataset A and Dataset B. It can be speculated that the ZF-GCN makes full use of the topological information by utilizing FlexConv to incorporate zigzag topological features on topological space. At the same time, in the comparison of the six regression methods, the proposed method MHTQF outperforms the baseline methods on Dataset A and Dataset B.

Table III also presents the results of the DM tests for various forecasting models, with this study encompassing a comparison of 24 forecasting models against the proposed model. In comparison to the comparison models, it is evident that the proposed model performs significantly better, as indicated by DM statistics surpassing the critical value at the 1% significance level. The results reveal that the proposed method predicts the median quantiles with a smaller error.

Fig. 3 illustrates a snapshot of the wind power median prediction of the comparison models and the proposed model, where the black points are the actual values. It is evident that the median prediction of our model ZF-GCN is closer to the actual value and has less volatility where the true value is smooth, compared with benchmark models, on Datasets A and B.

An outstanding prediction interval is as narrow as possible while meeting the coverage level. Moreover, when PICP is larger, the prediction interval has a greater probability of covering the true value. To evaluate the performance of interval prediction forecasted by different models, the prediction interval nominal confidence levels are specified as 85%, 90% and 95%. Table IV presents the results of MPIW and PICP on Datasets A and B, where the best MPIW are bold when the PICP meets the coverage level. For all prediction interval nominal confidence levels, most PICP values are larger than the corresponding prediction interval nominal confidence values, and only a few PICP values are slightly lower. These demonstrate that the prediction intervals generated by all models almost satisfy the effective coverage rate of the datasets. Moreover, it is obvious that the proposed ZF-GCN-MHTQF has the narrowest MPIW values satisfying the corresponding coverage level on both datasets and the prediction interval nominal confidence levels. On Dataset A, it has the best MPIW values of 75.96, 91.14 and 119.45 at 85%, 90% and 95% prediction interval nominal confidence levels respectively, and on Dataset B, it also has the best values of 58.29, 66.15 and 83.34 respectively. These show that the proposed method attains an excellent balance between MPIW and PICP, and fully satisfies the specified coverage level, which

TABLE III
MAE AND RMSE OF DIFFERENT MODELS ON DATASETS A AND B

Model	Dataset A			Dataset B		
	MAE	RMSE	DM test	MAE	RMSE	DM test
AGCRN-QR	26.67	41.05	10.50***	18.14	24.23	16.32***
AGCRN-MHTQF	25.63	39.83	7.77***	17.13	23.43	5.86***
DAAGCN-QR	26.7	40.14	13.57***	18.06	24.35	12.25***
DAAGCN-MHTQF	25.32	39.3	8.75***	17.74	24.03	8.26***
DSTAGNN-QR	26.88	39.81	17.71***	17.17	23.75	6.38***
DSTAGNN-MHTQF	23.76	36.74	4.15***	16.81	22.97	3.29***
STDGCNN-QR	26.39	39.09	9.99***	17.41	23.84	5.48***
STDGCNN-MHTQF	25.88	37.48	7.15***	17.05	23.08	3.30***
GGCN	26.39	37.96	7.89***	17.99	23.74	13.38***
AL-MCNN-BiLSTM	26.47	39.15	12.14***	17.39	23.94	7.29***
MLELM	26.63	38.4	8.21***	17.9	23.7	10.34***
GCN-QR	26.53	40.71	17.35***	17.49	23.55	9.14***
GCN-MHTQF	25.11	39.03	6.19***	17.32	23.08	5.64***
Z-GCN-LR	27.78	38.97	16.26***	17.75	23.44	8.27***
Z-GCN-QR	24.4	38.8	3.96***	17.24	23.79	6.80***
Z-GCN-MQR	24.33	38.73	5.93***	16.95	23.34	3.69***
Z-GCN-HQR	24.34	38	4.94***	16.97	23.26	4.24***
Z-GCN-MHQR	24.43	37.97	5.31***	17.52	23.72	9.80***
Z-GCN-HTQF	24.54	38.33	6.33***	17.26	23.38	6.98***
Z-GCN-MHTQF	24.85	38.48	9.38***	16.93	23.1	5.51***
ZF-GCN-LR	29.07	40.45	18.64***	20.81	25.84	19.52***
ZF-GCN-QR	31.48	47.97	23.84***	16.96	23.73	4.21***
ZF-GCN-MQR	24.13	38.68	4.54***	17.13	23.53	9.80***
ZF-GCN-HQR	25.22	38.4	6.59***	17.1	23.39	5.22***
ZF-GCN-MHQR	23.63	36.57	2.66***	17.08	23.39	4.70***
ZF-GCN-HTQF	25.88	39.32	7.11***	16.64	23.15	2.77***
ZF-GCN-MHTQF	23.28	36.33	-	16.62	22.89	-

¹ *** indicates the 1% significance level $Z_{0.01/2} = 2.58$.

² The bold numbers indicate the optimal value of the indicators.

TABLE IV
EVALUATION OF DIFFERENT MODELS FOR INTERVAL PREDICTION ON DATASETS A AND B

Model	Dataset A						Dataset B					
	85%		90%		95%		85%		90%		95%	
	PICP	MPIW	PICP	MPIW	PICP	MPIW		PICP	MPIW	PICP	MPIW	
AGCRN-QR	86.19%	99.56	92.51%	108.56	97.61%	149.87	86.04%	64.13	93.72%	74.51	98.29%	97.37
AGCRN-MHTQF	88.47%	85.15	95.19%	104.85	98.71%	144.57	88.55%	60.68	94.51%	72.29	98.69%	93.38
DAAGCN-QR	87.32%	101.22	94.24%	108.04	96.67%	148.69	85.77%	64.64	93.31%	77.10	97.54%	97.56
DAAGCN-MHTQF	88.54%	84.71	95.24%	103.57	99.00%	140.63	89.28%	63.01	95.30%	75.07	98.97%	96.99
DSTAGNN-QR	88.19%	98.43	92.55%	100.26	97.53%	137.93	87.13%	60.20	92.43%	70.51	98.21%	90.27
DSTAGNN-MHTQF	87.75%	80.32	93.88%	96.80	98.09%	128.06	87.81%	59.12	94.16%	70.07	98.33%	89.70
STDGCNN-QR	87.01%	97.21	94.51%	101.68	97.45%	141.64	86.47%	62.13	94.36%	70.09	97.75%	92.68
STDGCNN-MHTQF	87.33%	83.92	94.10%	96.31	97.41%	124.30	88.80%	59.42	94.47%	70.07	97.96%	90.50
GGCN	86.87%	88.02	94.81%	100.36	98.59%	128.82	86.53%	61.16	93.29%	69.19	97.24%	92.62
AL-MCNN-BiLSTM	88.45%	91.03	93.45%	99.53	97.15%	127.16	87.05%	62.34	92.71%	72.17	98.73%	91.89
MLELM	87.79%	93.31	93.26%	102.70	98.95%	138.84	87.45%	62.57	94.10%	71.64	96.30%	93.33
Z-GCN-QR	86.12%	98.84	90.75%	106.47	95.15%	134.57	88.06%	65.77	93.04%	71.52	93.75%	81.36
Z-GCN-MHTQF	87.51%	90.75	92.78%	105.78	98.09%	137.62	88.54%	62.69	94.32%	71.17	97.85%	98.15
Z-GCN-QR	85.85%	97.17	89.46%	105.32	95.78%	120.06	89.87%	64.70	92.00%	68.73	94.05%	74.21
Z-GCN-MQR	86.83%	87.73	90.32%	95.15	94.78%	103.86	83.23%	61.62	86.39%	66.50	88.79%	72.73
Z-GCN-HQR	83.84%	79.31	93.03%	96.00	98.46%	127.96	83.56%	57.08	91.55%	68.15	97.40%	88.43
Z-GCN-MHQR	86.89%	79.53	93.40%	95.93	98.17%	126.97	83.96%	59.14	92.47%	70.79	98.05%	92.27
Z-GCN-HTQF	88.18%	83.60	94.52%	100.53	98.84%	132.04	87.67%	60.14	94.22%	71.28	98.57%	91.10
Z-GCN-MHTQF	88.55%	84.56	95.07%	103.09	98.99%	139.47	87.55%	59.41	94.21%	70.58	98.38%	90.64
ZF-GCN-QR	91.27%	107.23	93.55%	120.35	95.72%	145.76	89.34%	64.23	91.68%	68.33	94.07%	74.74
ZF-GCN-MQR	85.35%	86.74	89.14%	94.32	94.53%	103.45	83.51%	61.50	86.41%	66.77	89.00%	73.28
ZF-GCN-HQR	84.52%	83.20	92.93%	100.11	98.53%	131.84	83.53%	56.28	93.94%	69.19	98.29%	88.81
ZF-GCN-MHQR	86.93%	79.00	93.87%	95.17	98.15%	125.71	84.89%	56.24	94.52%	70.38	98.57%	91.93
ZF-GCN-HTQF	89.16%	88.42	95.63%	106.37	99.45%	140.13	88.52%	58.77	91.25%	66.76	97.29%	85.57
ZF-GCN-MHTQF	86.68%	75.96	92.82%	91.14	97.83%	119.48	87.62%	58.29	91.60%	66.15	97.03%	83.34

TABLE V
CRPS AND CROSS-LOSS OF DIFFERENT MODELS ON DATASETS A AND B

Model	Dataset A		Dataset B	
	CRPS	CL	CRPS	CL
AGCRN-QR	20.30	3.28	12.61	0.84
AGCRN-MHTQF	18.08	0.00	12.05	0.00
DAAGCN-QR	18.99	6.52	12.68	0.64
DAAGCN-MHTQF	17.90	0.00	12.66	0.00
DSTAGNN-QR	17.16	1.12	12.52	0.45
DSTAGNN-MHTQF	16.88	0.00	12.30	0.00
STDGCNN-QR	17.52	2.19	12.25	0.28
STDGCNN-MHTQF	17.12	0	12.14	0
GGCN	18.19	5.04	12.30	0.41
AL-MCNN-BiLSTM	18.66	0	12.38	0
MLELM	19.70	0	12.50	0
GCN-QR	19.93	4.31	12.58	0
GCN-MHTQF	17.31	0	12.35	0
Z-GCN-QR	19.82	7.68	12.53	0.31
Z-GCN-MQR	17.45	1.76	12.24	0
Z-GCN-HQR	17.10	0	12.04	0
Z-GCN-MHQR	17.20	0	12.39	0
Z-GCN-HTQF	17.42	0	12.29	0
Z-GCN-MHTQF	17.62	0	12.11	0
ZF-GCN-QR	22.60	3.58	12.32	0.59
ZF-GCN-MQR	17.28	2.14	12.03	0
ZF-GCN-HQR	17.77	0	12.07	0
ZF-GCN-MHQR	16.69	0	12.08	0
ZF-GCN-HTQF	18.31	0	11.96	0
ZF-GCN-MHTQF	16.51	0	11.94	0

has enormous implications for practical wind power forecasting applications. Moreover, only the HTQF method and the MHTQF method can satisfy all the conditions that the PICP is greater than the preset confidence level, and the MPIW of the MHTQF method is significantly better than the QR method.

Table V demonstrates the evaluation metrics value of the proposed model and benchmark models for two wind power multi-quantile regression forecasting tasks, where the best results are bold. From Table V, it is observed that the proposed model ZF-GCN-MHTQF has the smallest CRPS and CL values on both Dataset A and Dataset B, which illustrates that the proposed model outperforms all benchmark models in the terms of the performance of probabilistic prediction. Concerning the problem of the crossing phenomenon, take Dataset A as an example, the CL of the Z-GCN-QR, Z-GCN-MQR, ZF-GCN-QR and ZF-GCN-MQR are 7.68, 1.76, 3.58 and 2.14, respectively. It can be shown that the classical multi-quantile regression method has the problem of crossing and the point-by-point monotonic loss function can effectively alleviate this problem. In addition, the CL of the HQR, MHQR, HTQF and MHTQF methods are both 0, which indicates that our proposed prediction model not only generates excellent probabilistic prediction but also prevents the crossing problem.

Fig. 4 draws the multiple prediction intervals of the proposed model ZF-GCN-MHTQF on a snapshot of Datasets A and B, where the blue linear represents the true wind power. It can be seen from Fig. 4(a) that the true value is covered by three prediction intervals with different significance levels. At the same time, for Fig. 4(b), most of the true values are covered by multiple prediction intervals.

As depicted in Tables III-V, Z-GCN consistently outperforms GCN, indicating that leveraging topological information

through Zigzags enhances the model's effectiveness. The complexity of Falcon may occasionally result in suboptimal performance. However, when FlexConv is combined with monotonicity, there is a significant performance improvement compared to models that do not incorporate these methods, as monotonicity enhances the learning process. Similar conclusions are drawn when monotonicity is combined with specific quantile loss functions. Similar conclusions can be drawn when monotonicity is combined with a specific quantile loss function, in which the loss function combining monotonicity with HTQF has the most outstanding performance. Therefore, it can be concluded that monotonicity plays a crucial role in the proposed model. In summary, our proposed model not only effectively predict wind power, but also can successfully quantify the uncertainty of wind power prediction.

V. CONCLUSION

This paper proposes a new deep learning model ZF-GCN-MHTQF for wind power forecasting. Our ZF-GCN-MHTQF model is based on time zigzags and flexible convolution at graph convolutional network framework, the point-wise monotonic loss function and the heavy-tailed quantile function. The novel time zigzags and flexible convolution at graph convolutional network framework not only exact the time zigzags topological information of the time series data but also simultaneously learn spatial-temporal features for probabilistic prediction. Additionally, the point-wise loss function and the heavy-tailed quantile function are successfully combined to handle the quantile crossing problem and generate the multi-quantile prediction of wind power. In the experiment, the proposed model reaches state-of-the-art performance on the two wind power datasets for probabilistic prediction, compared to the recent benchmark methods, demonstrating the effectiveness and superiority of the proposed model.

REFERENCES

- [1] Global wind energy council. GWEC: Global wind report 2023. [Online]. Available: <https://gwec.net/globalwindreport2023/>
- [2] G. J. Osório, J. C. O. Matias, and J. P. S. Catalão, "Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information," *Renewable Energy*, vol. 75, pp. 301–307, 2015.
- [3] T. Ji et al., "Wind power forecast with error feedback and its economic benefit in power system dispatch," *IET Gener. Transmiss. Distrib.*, vol. 12, no. 21, pp. 5730–5738, 2018.
- [4] S. Goodarzi, H. N. Perera, and D. Bunn, "The impact of renewable energy forecast errors on imbalance volumes and electricity spot prices," *Energy Policy*, vol. 134, 2019, Art. no. 110827.
- [5] L. L. Li et al., "Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm," *J. Cleaner Prod.*, vol. 242, 2020, Art. no. 118447.
- [6] F. Shahid, A. Zameer, and M. Muneeb, "A novel genetic LSTM model for wind power forecast," *Energy*, vol. 223, 2021, Art. no. 120069.
- [7] L. Li et al., "Improving short-term wind power prediction using hybrid improved cuckoo search arithmetic-support vector regression machine," *J. Cleaner Prod.*, vol. 279, 2021, Art. no. 123739.
- [8] Y. Y. Hong and C. L. P. Rioflorido, "A hybrid deep learning-based neural network for 24-h ahead wind power forecasting," *Appl. Energy*, vol. 250, pp. 530–539, 2019.
- [9] M. Khodayar and J. Wang, "Spatio-temporal graph deep neural network for short-term wind speed forecasting," *IEEE Trans. Sustain. Energy*, vol. 10, no. 2, pp. 670–681, Apr. 2019.

- [10] X. Yu, B. Tang, and K. Zhang, "Fault diagnosis of wind turbine gearbox using a novel method of fast deep graph convolutional networks," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 6502714.
- [11] Y. Song, D. Tang, J. Yu, Z. Yu, and X. Li, "Short-term forecasting based on graph convolution networks and multiresolution convolution neural networks for wind power," *IEEE Trans. Ind. Inform.*, vol. 19, no. 2, pp. 1691–1702, Feb. 2023.
- [12] W. Liao et al., "Short-term power prediction for renewable energy using hybrid graph convolutional network and long short-term memory approach," *Electric Power Syst. Res.*, vol. 211, 2022, Art. no. 108614.
- [13] Y. He and Y. Wang, "Short-term wind power prediction based on EEMD-LASSO-QRNN model," *Appl. Soft Comput.*, vol. 105, 2021, Art. no. 107288.
- [14] C. Zhang et al., "Evolutionary quantile regression gated recurrent unit network based on variational mode decomposition, improved whale optimization algorithm for probabilistic short-term wind speed prediction," *Renewable Energy*, vol. 197, pp. 668–682, 2022.
- [15] Y. He et al., "Uncertainty analysis of wind power probability density forecasting based on cubic spline interpolation and support vector quantile regression," *Neurocomputing*, vol. 430, pp. 121–137, 2021.
- [16] J. Hu et al., "Conformalized temporal convolutional quantile regression networks for wind power interval forecasting," *Energy*, vol. 248, 2022, Art. no. 123497.
- [17] H. Dette and S. Volgushev, "Non-crossing non-parametric estimates of quantile curves," *J. Roy. Stat. Soc.: Ser. B. (Stat. Methodol.)*, vol. 70, no. 3, pp. 609–627, 2008.
- [18] Y. Liu and Y. Wu, "Simultaneous multiple non-crossing quantile regression estimation using kernel constraints," *J. Nonparametric Statist.*, vol. 23, no. 2, pp. 415–437, 2011.
- [19] S. Bang, H. J. Cho, and M. Jhun, "Simultaneous estimation for non-crossing multiple quantile regression with right censored data," *Statist. Comput.*, vol. 26, no. 1, pp. 131–147, 2016.
- [20] V. M. R. Muggeo et al., "Estimating growth charts via nonparametric quantile regression: A practical framework with application in ecology," *Environ. Ecolog. Statist.*, vol. 20, no. 4, pp. 519–531, 2013.
- [21] M. Roth, T. A. Buishand, and G. Jongbloed, "Trends in moderate rainfall extremes: A regional monotone regression approach," *J. Climate*, vol. 28, no. 22, pp. 8760–8769, 2015.
- [22] J. Hu, J. Tang, and Y. Lin, "A novel wind power probabilistic forecasting approach based on joint quantile regression and multi-objective optimization," *Renewable Energy*, vol. 149, pp. 141–164, 2020.
- [23] I. Takeuchi et al., "Nonparametric quantile estimation," *J. Mach. Learn. Res.*, vol. 7, no. 45, pp. 1231–1264, 2006.
- [24] T. Hofmeister, "QRSVM: SVM quantile regression with the pinball loss," R. Package Version 0.2, 2017, 1.
- [25] R. Fathony et al., "Multiplicative filter networks," in *Proc. Int. Conf. Learn. Representations*, 2020.
- [26] Y. Chen, I. Segovia, and Y. R. Gel, "Z-GCNETs: Time zigzags at graph convolutional networks for time series forecasting," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 1684–1694.
- [27] Z. Wu et al., "Graph wavenet for deep spatial-temporal graph modeling," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 1907–1913.
- [28] K. Cho et al., "On the properties of neural machine translation: Encoder-decoder approaches," in *Proc. SSST-8, Eighth Workshop Syntax, Semantics Struct. Stat. Transl.*, 2014, pp. 103–111.
- [29] X. Yan et al., "Parsimonious quantile regression of financial asset tail dynamics via sequential learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018.
- [30] P. Márquez-Neila, M. Salzmann, and P. Fua, "Imposing hard constraints on deep networks: Promises and limitations," *CVPR Workshop Negative Results Comput. Vis.*, 2017.
- [31] A. Gupta et al., "How to incorporate monotonicity in deep networks while preserving flexibility?" 2019, *arXiv:1909.10662*.
- [32] L. Bai et al., "Adaptive graph convolutional recurrent network for traffic forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 17804–17815.
- [33] J. Jiang et al., "Dynamic adaptive and adversarial graph convolutional network for traffic forecasting," 2022, *arXiv:2208.03063*.
- [34] S. Lan et al., "DSTAGNN: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting," in *Proc. Int. Conf. Mach. Learn.*, 2022, pp. 11906–11917.
- [35] Z. Li et al., "A Spatiotemporal directed graph convolution network for ultra-short-term wind power prediction," *IEEE Trans. Sustain. Energy*, vol. 14, no. 1, pp. 39–54, Jan. 2023.
- [36] L. Wang et al., "A gated graph convolutional network with multi-sensor signals for remaining useful life prediction," *Knowl.-Based Syst.*, vol. 252, 2022, Art. no. 109340.
- [37] Y. Wang et al., "A Deep asymmetric laplace neural network for deterministic and probabilistic wind power forecasting," *Renewable Energy*, vol. 196, pp. 497–517, 2022.
- [38] Y. Liu and J. Wang, "Transfer learning based multi-layer extreme learning machine for probabilistic wind power forecasting," *Appl. Energy*, vol. 312, 2022, Art. no. 118729.
- [39] D. W. Romero et al., "FlexConv: Continuous kernel convolutions with differentiable kernel sizes," in *Proc. Int. Conf. Learn. Representations*, 2021.
- [40] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Statist.*, vol. 13, pp. 253–263, 1995.

Jingwei Tang received the B.S. and M.S. degrees in statistics from the College of Economics and Statistics, Guangzhou University, Guangzhou, China, in 2017 and 2020, respectively. He is currently working toward the Ph.D. degree in mathematics with the Department of Mathematics, Faculty of Science and Technology, University of Macau, Macau, China. His research interests include statistical learning, probability prediction, and time series forecasting.

Zhi Liu received the B.S. degree in computational mathematics and the M.S. degree in statistics from the School of Mathematics and Statistics, Lanzhou University, Lanzhou, China, in 2003 and 2006, respectively, and the Ph.D. degree in statistics from the Department of Mathematics, The Hong Kong University of Science and Technology, Hong Kong, China, in 2011. He is currently an Associate Professor with the Department of Mathematics, also an Affiliate Researcher with the Center for Applied Mathematics, and the Center for Artificial Intelligence and Robotics, University of Macau, Zhuhai, China. His research interests include statistics for stochastic processes, financial statistics, and statistical learning for health science.

Jianming Hu received the B.S. degree in industrial engineering from the School of Management, Zhengzhou University, Zhengzhou, China, in 2007, and the M.S. degree in applied statistics and the Ph.D. degree in probability and mathematical statistics from the School of Mathematics and Statistics, Lanzhou University, Lanzhou, China, in 2013 and 2016, respectively. He is currently a Professor with the School of Mathematics and Statistics, Hainan University, Haikou, China. His research interests include Bayesian learning, applied statistics, and deep learning.