

# Spatiotemporal PM<sub>2.5</sub> forecasting via dynamic geographical Graph Neural Network

Qin Zhao <sup>a,b,c</sup>, Jiajun Liu <sup>a</sup>, Xinwen Yang <sup>b</sup>, Hongda Qi <sup>a,b</sup>, Jie Lian <sup>a,b,\*</sup>

<sup>a</sup> Department of Computer Science and Technology, Shanghai Normal University, 100 Haisi Road, Fengxian District, Shanghai, 201418, China

<sup>b</sup> Shanghai Engineering Research Center of Intelligent Education and Big Data, Shanghai Normal University, 100 Guilin Road, Xuhui District, Shanghai, 200234, China

<sup>c</sup> Key Laboratory of Embedded Systems and Service Computing of Ministry of Education, Tongji University, 4800 Caoan Highway, Jiading District, Shanghai, 201804, China

## ARTICLE INFO

### Keywords:

Air pollution  
PM<sub>2.5</sub> forecasting  
Graph Convolutional Network  
Dynamic graph  
Deep learning

## ABSTRACT

With the growing interest in data-driven methods, Graph Neural Networks (GNNs) have demonstrated strong performance in PM<sub>2.5</sub> forecasting as a deep learning architecture. However, GNN-based methods typically construct the graph based solely on the distance between stations, and few methods introduce geographical factors that significantly affect the spatial dispersion of PM<sub>2.5</sub>, leading to performance bottlenecks. Additionally, these methods often fail to process the dynamic wind-field data comprehensively, resulting in inaccurate PM<sub>2.5</sub> dispersion graph construction. These shortcomings greatly limit the interpretability of GNN models in forecasting air pollution. To address these issues, we propose a deep learning method that combines Graph Convolution Network (GCN) with Long Short-Term Memory (LSTM), leveraging geographical information within a dynamic graph. The model captures spatial dependencies between PM<sub>2.5</sub> monitoring stations using a dynamic directional graph derived from the wind-field data and a static graph to represent inherent geographical relationships. The combination of GCN and LSTM enables the extraction of both spatial and temporal correlations. The results of experiments suggest that our proposed model, which offers great interpretability, outperforms state-of-the-art methods, especially in 24, 30, and 36 hours forecasts.

## 1. Introduction

PM<sub>2.5</sub> is a significant component of atmospheric pollutants. Similar to other meteorological forecasts, PM<sub>2.5</sub> forecasting has mainly been dominated by model-based methods over the past few decades, achieving certain accomplishments. Representative models include the Nested Air Quality Forecasting Modeling System (NAQPMS) (Wang et al., 2001), the Community Multiscale Air Quality (CMAQ) model (Foley et al., 2010), and the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem) (Saide et al., 2011; Goldberg et al., 2019). For these models, accuracy largely depends on the meteorological and physical knowledge incorporated during modeling, which makes them highly interpretable. However, the computational complexity becomes a severe limitation, and the scalability struggles to keep pace with the increasing dimensions and volume of data (Lam et al., 2022).

In recent years, data-driven Machine Learning methods have demonstrated advantages in various physics-related domains, including solid earth science (Bergen et al., 2019), material science (Himanen et al.,

2019), and meteorology (Lam et al., 2022; Ben Bouallègue et al., 2024). These methods have also been applied to PM<sub>2.5</sub> forecasting (Karimian et al., 2019; Wu et al., 2023). Machine Learning approaches generally exhibit better scalability with data, potentially achieving higher accuracy at lower computational costs. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have been employed as proficient methods for time series problems to capture the periodicity and tendency of PM<sub>2.5</sub> concentrations (Zhao et al., 2019; Tsai et al., 2018). Recognizing that PM<sub>2.5</sub> dispersion involves both temporal and spatial correlations, some studies have attempted to incorporate Convolutional Neural Networks (CNNs) or Transformers (Li et al., 2020; Yu et al., 2023). The combination of CNN and LSTM has shown improved results in PM<sub>2.5</sub> forecasting problems (Qin et al., 2019; Wu and Li, 2022).

However, the relationship between monitoring stations cannot be fully captured in Euclidean space, as it involves not only distance but also factors such as terrain, landform, and wind direction. As a result,

\* Corresponding author at: Shanghai Engineering Research Center of Intelligent Education and Big Data, Shanghai Normal University, 100 Guilin Road, Xuhui District, Shanghai, 200234, China.

E-mail address: [lianjie@shnu.edu.cn](mailto:lianjie@shnu.edu.cn) (J. Lian).

<https://doi.org/10.1016/j.envsoft.2025.106351>

Received 4 November 2024; Received in revised form 19 January 2025; Accepted 26 January 2025

Available online 6 February 2025

1364-8152/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

these data are more appropriately represented in non-Euclidean space, which is incompatible with CNNs. Additionally, CNNs tend to focus on local spatial features, often neglecting long-range spatial relationships. Although Transformers do not have these limitations, they face significant computational challenges as input window sizes increase.

Graph Neural Networks (GNNs), as a method commonly used for handling non-Euclidean spatial relationships, have garnered wide spread attention in the fields of social networks (Li et al., 2023), biotechnology (Zhang et al., 2021), and knowledge graphs (Ye et al., 2022). Compared to other methods, GNNs can select which parts of representation interact with one another by the input graph, and allow the interaction over any range. This capability enables them to achieve good performance in fields closely related to real-world physics, such as weather forecasting, as demonstrated in several studies (Keisler, 2022; Ma et al., 2023). Some researchers have applied GNNs to  $PM_{2.5}$  forecasting as well and achieved good results (Qi et al., 2019). Nevertheless, GNN-based models need to obtain the correlations between stations, making it essential to establish accurate graphs containing sufficient information. To enhance the information encoded in the edges between stations, Zhou et al. utilize dynamic wind-field to establish a directional graph (Zhou et al., 2021), which has been adopted in some graph-based  $PM_{2.5}$  forecasting models with different structures (Zhou et al., 2022). This approach obtains the edge weights by calculating the wind-field distance (refers to the method of Li et al. (2014)) and finding the shortest distance, potentially overlooking non-primary dispersion paths. Xiao et al. propose to calculate the wind-field relationship by the wind speed and direction between stations (Xiao et al., 2022). The downside is that it fails to mention how to obtain wind vectors between stations using gridded wind-field data. Besides the wind-field information, meteorological factors have been increasingly incorporated into the models (Muthukumar et al., 2021; Wang et al., 2023). However, most geographical factors apart from distance have generally been neglected. Some geographical factors, such as mountains, forests, and buildings, can also have a considerable impact on  $PM_{2.5}$  dispersion (Xu et al., 2021; Yousefi et al., 2023).

To address the aforementioned limitations, we propose a novel approach DGGNN (Dynamic Geographical Graph Neural Network) to extract geographical dependencies in  $PM_{2.5}$  forecasting. As important factors in  $PM_{2.5}$  dispersion, wind-field and geographic relationships are embedded into dynamic and static directional graphs, respectively. The spatial dependencies among stations are learned by a graph convolution network layer. The continuous change wind-field information is embedded as a dynamic map and input into the neural network. Compared to previous methods, the most notable advancement of our approach lies in the introduction of a static graph to represent the inherent relationships between nodes (i.e., the geographical relationships between stations). By learning this graph directly from the model, the accuracy of edge embeddings can be significantly enhanced. We then use LSTM for the output sequence of graph convolution to extract temporal information and obtain forecasting results through a fully-connected layer. Finally, we conduct a set of experiments to compare our model with baseline models (LSTM, GC-LSTM, and STGCN). The results demonstrate the high effectiveness of the proposed approach.

## 2. Related work

### 2.1. Physical statistical methods

Based on air dynamic theory, meteorology theory, and atmospheric physical chemistry knowledge, physical statistical methods can achieve relatively accurate results on a small area and dataset. Recent studies have made improvements on the classic models, such as CMAQ, WR-Chem, and NAQPMS. For example, Li et al. combine the meteorology field generated by WRF with CMAQ model for  $PM_{2.5}$  and  $O_3$  prediction in Xianghe and Taizhou (Li et al., 2022).

In addition, with the development of new-generation geostationary satellites, optical data based on satellite retrieval promoted the improvement of models. Yeganeh proposes a satellite-based model to estimate the ground-level  $PM_{2.5}$  concentration with soft computing methods (Yeganeh et al., 2017). Park et al. propose to estimate the spatially continuous daytime particulate matter concentrations through the synergistic use of satellite-based AOD and numerical models (Park et al., 2020). Hong et al. develop an empirical and statistics-based scheme for improving the  $PM_{2.5}$  prediction of WRF-Chem using aerosol optical depth (AOD) data assimilation methods (Hong et al., 2020). However, the data of higher precision and resolution are not easily utilized by the model-driven physical statistical methods. Therefore, people are beginning to seek methods for better handling large amounts of data, such as machine learning and deep learning methods.

### 2.2. Machine learning methods

Machine learning methods can complete classification or regression tasks without extensive professional theoretical knowledge, but rely on a large amount of real historical data. Some researchers have been applying universal and effective machine learning methods on  $PM_{2.5}$  prediction. Ma et al. establish a modified XGBoost model to conduct  $PM_{2.5}$  forecasting in Shanghai, and achieved better results in comparison with WRF-Chem (Ma et al., 2020). Zhan et al. develop a novel machine learning algorithm, Geographically-Weighted Gradient Boosting Machine (GW-GBM) to address the spatial non-stationarity of the relationships between  $PM_{2.5}$  concentrations and predictor variables (Zhan et al., 2017). Udristioiu et al. propose to combine the models like Input Variable Selection (IVS) and Machine Learning to forecast the daily concentrations of  $PM_{2.5}$  and Air Quality Index (AQI) (Udristioiu et al., 2023). Zhao et al. propose a Hybrid Integration (HIG) algorithm to optimize RNN and LSTM. The results show that the HIG-RNN and HIG-LSTM are more advantageous than the ordinary method in terms of reasonable weight assignment (Zhao et al., 2024). Dai et al. propose a hybrid model combining XGBoost, four GARCH models and MLP model (XGBoost-GARCH-MLP) to predict  $PM_{2.5}$  concentration values and volatility (Dai et al., 2022). Aman et al. use AOD as the main predictor variable to predict  $PM_{2.5}$  in Greater Bangkok by four individual machine learning models (Aman et al., 2024). Although it is possible to learn the potential patterns and rules from historical data to achieve a better performance, machine learning methods have significant deficiencies in the processing of complex data and feature extraction capabilities. Therefore, deep learning methods to address this issue have become a trend in recent years.

### 2.3. Deep learning methods

As the computing power bottleneck is broken through, deep learning methods have become popular in air pollution prediction. Zhang et al. propose a deep learning model based on an auto-encoder and bidirectional long short-term memory (Bi-LSTM) to forecast  $PM_{2.5}$  concentrations to reveal the correlation between  $PM_{2.5}$  and multiple climate variables (Zhang et al., 2020b). Zhang et al. propose a  $PM_{2.5}$  prediction model MTD-CNN-GRU model for intensive stations and utilize multi-task deep learning to effectively mine deeper air quality information (Zhang et al., 2020a). Faraji et al. combine a three-dimensional convolutional neural network and gated recurrent unit (3DCNN-GRU) to learn spatial patterns from similar air quality stations (Faraji et al., 2022).

Noting the advantages of graphs in modeling stations in non-Euclidean space, an increasing number of studies are using graphs to model the spatial relationships of pollutants. Tan et al. propose an ensemble Graph Attention Reinforcement Learning Recursive Network to aggregate spatial-temporal correlation (Tan et al., 2022). Teng et al. apply a hybrid graph deep neural network (GNN-LSTM) to represent the physical mechanism of pollutant transport across the space (Teng et al.,

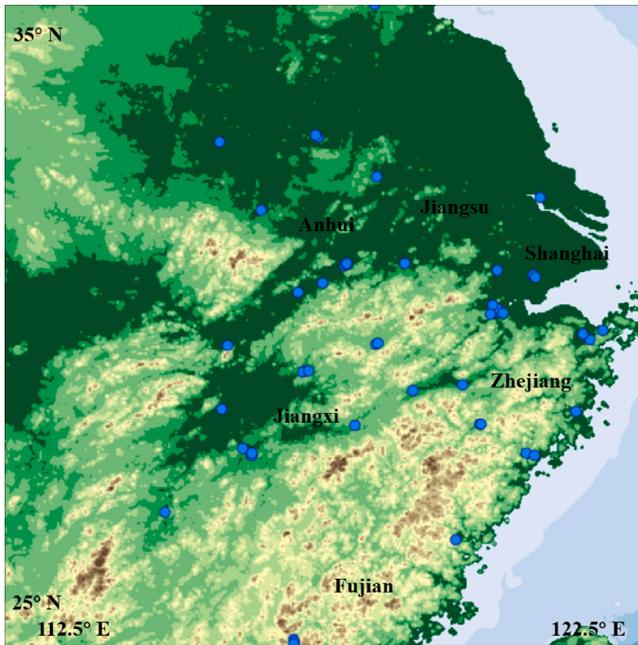


Fig. 1. Study area and stations.

2023), and use AOD as a feature to improve performance. Kim et al. develop a novel framework  $PM_{2.5}$  prediction that utilizes multiple edges for feature extraction and employs a multi-gated graph neural network for feature calculations (Kim et al., 2023). To model the complicated dispersion process of  $PM_{2.5}$ , Ouyang et al. propose a Dual-Channel Spatial-Temporal Difference Graph Neural network (DC-STDGN) to capture more comprehensive spatio-temporal correlations by introducing spatial differences with domain knowledge (Ouyang et al., 2023). Zhang et al. propose to capture the spatial dependencies by employing the dynamic multi-graph attention (MGAtt) module, which achieved superior performance in  $PM_{2.5}$  concentration prediction (Zhang et al., 2024).

Compared to CNN-based models, which ignore the non-Euclidean characteristics of spatial-temporal data, graph-based methods adeptly address this issue. However, the existing graph-based methods do not consider geographical factors comprehensively when constructing the graphs, leading to inaccurate modeling of interrelationships between stations. Therefore, we propose a novel method to address this issue by introducing a geographical relationship graph, and using the GNN model to learn spatial and temporal features. This approach allows for a more comprehensive consideration of the relationships between stations.

### 3. Materials

#### 3.1. Observed $PM_{2.5}$ concentration datasets

The real-time observed air quality data used in this study is released hourly by the China National Environmental Monitoring Centre (CNEMC) on its platform. This data includes the Air Quality Index (AQI) and the concentrations of  $PM_{2.5}$ ,  $PM_{10}$ ,  $SO_2$ ,  $NO_2$ ,  $CO$ , and  $O_3$ . As shown in Fig. 1, our study area is primarily in East China, spanning  $110^\circ E$  to  $130^\circ E$  and  $20^\circ N$  to  $40^\circ N$ . We select 50 stations out of 200 in the study area, ensuring that stations with less than 5% missing data were chosen as experimental subjects. Our dataset contains hourly  $PM_{2.5}$  concentration data from 2020 to 2022, and the missing data was filled using linear interpolation.

#### 3.2. Continuous wind-field data

The wind data is obtained from ERA5, the fifth-generation ECMWF reanalysis for the global climate and weather (Hersbach et al., 2020). We collect 10-m wind data from 2020 to 2022 for our study area, with a spatial resolution of  $0.25^\circ$  in both longitude and latitude. The temporal resolution is 1 h to match the  $PM_{2.5}$  data.

## 4. Proposed method

### 4.1. Problem definition

**Definition 1.** The  $PM_{2.5}$  forecasting is a typical time series problem, where the future concentration of  $PM_{2.5}$  is forecasted based on the concentration and meteorological information over a past period:

$$X_{t+1}^n \leftarrow (X_{t-r}^n, \dots, X_t^n, W_{t-r}, \dots, W_t), \quad (1)$$

where  $X_t^n$  is the vector which represents the  $PM_{2.5}$  concentrations of  $n$  number of stations at time  $t$ , while  $W_t$  represents the meteorological variables at time  $t$ .

**Definition 2.** We propose a directional graph  $G = (V, E)$  to describe the geographical topological structure of stations. Due to the long-term invariance of geographical relationships, we define it as a static graph, in contrast to the dynamic wind-field graph.  $V = \{v_1, v_2, \dots, v_n\}$  denotes the  $PM_{2.5}$  monitoring stations in our study.  $E = \{e_1, e_2, \dots, e_m\}$  represents the impact of geographical conditions on the dispersion of  $PM_{2.5}$  between stations. The edges are represented by adjacency matrix  $A_g \in R^{N \times N}$ . The elements of  $A_g$  are obtained through learning.  $a_{ij}$  is initialized as the reciprocal of distance and will be zeroing if less than a certain value. Notably,  $a_{ij}$  is not equal to  $a_{ji}$  because of the asymmetry of geography, which is the reason for using the directional graph.

**Definition 3.** The wind-field varies over time in the prediction process, so we introduce  $G^t = (V, E^t)$  to represent the dynamic wind-field graph, where  $E^t = \{e_1^t, e_2^t, \dots, e_m^t\}$  represents the impact factor of wind-field between stations at time  $t$ .  $A_w^t \in R^{N \times N}$  is the adjacency matrix of wind-field at time  $t$ . Then we propose a semi-dynamic directed graph by combining the two graphs. The edges of the graph are obtained by combining static and dynamic edges to represent the invariant and changing parts of the spatial relationship between stations.

### 4.2. The overall framework

Since the spatiotemporal graph contains the inherent and dynamic relationship between nodes, we propose a Graph Convolution Long Short-Term Memory using Dynamic Geographical Information. The structure of the model is shown in Fig. 2. The dynamic graph convolutional layer captures the spatial correlation of the connected nodes by combining two graphs. The static graph learned from the model represents the invariant inherent relationship and the dynamic graph from inputs represents the constantly changing external relationship between nodes (i.e., the constantly changing wind-field between stations). Conducting dynamic graph convolution on historical data, the hidden layer containing time series information can be obtained. Then, using the temporal aggregation layer, time dependencies are extracted from the hidden layer by LSTM, and the final prediction results are generated by a fully-connected layer.

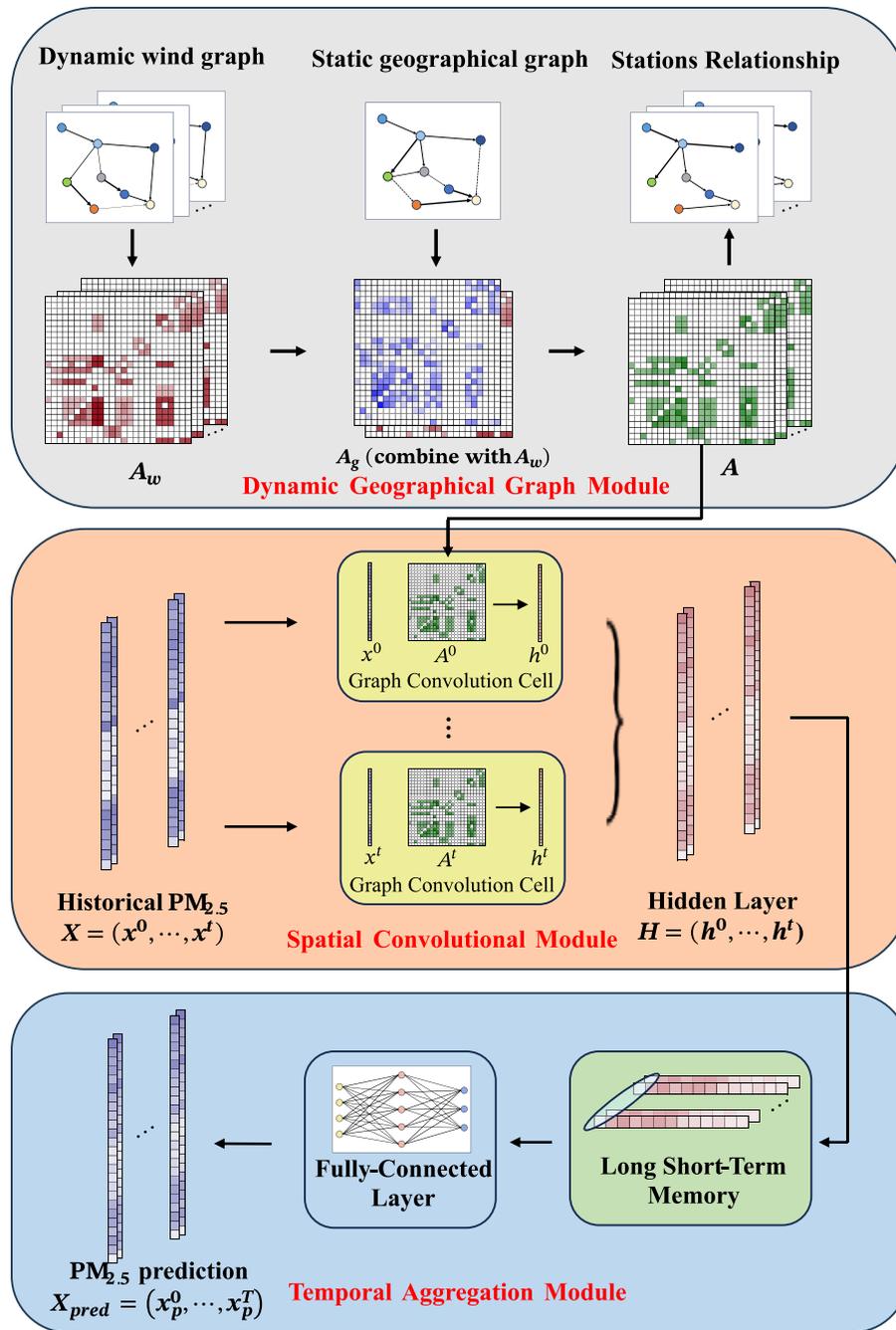


Fig. 2. The structure of DGGNN model.

#### 4.2.1. Calculation of wind-field

To extract the wind-field relationships between stations, we propose a method to compute this relationship from the original gridded wind-field data. The algorithm comprises two steps. The first step involves selecting the grid points deemed influential. This is achieved by defining the line between two stations as the midline and then drawing a square based on this midline. All grid points that fall within this square are considered influential and are selected for further analysis, as illustrated in Fig. 3.

The second step is to aggregate the wind data from the selected grid points. The wind data consists of two components: the  $u$ -component, which represents the horizontal wind moving towards the east, and the  $v$ -component, which represents the horizontal wind moving towards the north. Since PM<sub>2.5</sub> does not always disperse directly from one station to another and may be influenced by various wind paths, we use

two directed edges ( $w_{AB}$  and  $w_{BA}$ ) to describe the relationship between the stations, accounting for the bi-directional influence of the wind:

$$\begin{cases} w_{AB} = \frac{1}{n} \sum (\bar{w}_{pi}, \bar{AB}) \times d_i, \bar{w}_{pi} \times \bar{AB} > 0, \\ w_{BA} = \frac{1}{n} \sum (\bar{w}_{pi}, \bar{AB}) \times d_i, \bar{w}_{pi} \times \bar{AB} < 0, \end{cases} \quad (2)$$

where  $\bar{w}_{pi}$  is the wind-field of grid point  $p_i$  and  $\bar{AB}$  is the direction vector from station  $A$  to station  $B$ , with  $n$  equals the number of grid points included in the calculation. By performing the vector dot product, we obtain the component of the wind in the direction of two nodes, and it can be calculated from the  $u$ ,  $v$  components:

$$(\bar{w}_{pi}, \bar{AB}) = u \cdot \cos \frac{D_{AB} \times \pi}{180} + v \cdot \cos \frac{(D_{AB} - 90) \times \pi}{180}, \quad (3)$$

where  $D_{AB}$  represents the azimuth from  $A$  to  $B$ .

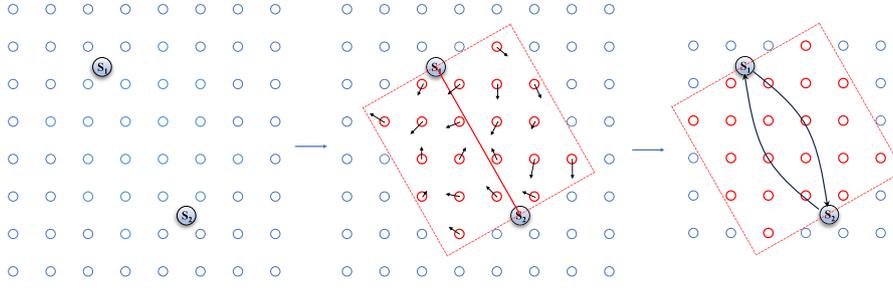


Fig. 3. Select region and calculate wind-field relationship.

The wind far from the direct path of  $A$  and  $B$  has less impact. Therefore, we introduce  $d_i$  as a distance penalty coefficient to assign weights for the point at different positions:

$$d_i = \frac{d_{AB}}{d_{Ai} + d_{Bi}}, \quad (4)$$

where  $d_{AB}$ ,  $d_{Ai}$  and  $d_{Bi}$  represent the distance of  $(A, B)$ ,  $(A, i)$  and  $(B, i)$ .

Notably, we calculate the wind from  $A$  to  $B$  and from  $B$  to  $A$  separately. Thus the component will only be added to  $w_{AB}$  if it is from  $A$  to  $B$  and vice versa.

#### 4.2.2. Spatial convolutional module for capturing spatial information

Graph Convolution Network (GCN) is an approach to aggregate spatial information of non-Euclidean space, which leverages the advantages of both GNN and CNN (Kipf and Welling, 2016). In GCN, the propagation between layers can be shown as follows:

$$H^{l+1} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^l W^l), \quad (5)$$

where  $H^l$  denotes the input from the last layer.  $W^l$  is the trainable weight of this layer.  $\tilde{A}$  is derived from the following:

$$\tilde{A} = A + I, \quad (6)$$

where  $A$  is the adjacency matrix of the graph to add the node relationships information to model, and  $I$  represents the identity matrix.

The normalized adjacency matrix can be represented as  $\hat{A}$ :

$$\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}, \quad (7)$$

where  $\tilde{D}$  is the degree matrix of  $\tilde{A}$ , and  $\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$  is to perform matrix normalization with the degree. It is a strategy of weighted averaging, which ensures nodes with different degrees having different magnitudes of impact.

In the  $PM_{2.5}$  forecasting problem, understanding the impact of geographical and meteorological factors on spatial dispersion is crucial for both prediction accuracy and model interpretability. The wind-field is the main factor driving the dispersion of  $PM_{2.5}$ , dictating how pollutants are transported across regions. Its dynamic nature is represented by a dynamic graph, capturing temporal variations in wind speed and direction. Meanwhile, geographical factors, which influence wind patterns through features such as terrain and land use, are captured in a static graph that represents the inherent and invariant spatial relationships between stations.

Specifically, if  $e_{ij}^t$  represents the impact of node  $v_j$  obtained from  $v_i$  at time  $t$ , we considered it will be determined by three parts. That is to say that the  $PM_{2.5}$  concentration of  $v_i$ , the wind-field state and the geography between the nodes jointly have an impact on the dispersion. As a result, the  $e_{ij}^t$  is obtained as follows:

$$e_{ij}^t = a_{w(ij)}^t a_{g(ij)}^t x_i^t, \quad (8)$$

where  $a_{w(ij)}^t$  represents the wind-field from  $v_i$  to  $v_j$  at time  $t$ , and  $a_{g(ij)}^t$  represents the weaken or strengthen effect on the wind-field by geographical relationship from  $v_i$  to  $v_j$ . Then the impact  $e_{ij}^t$  to  $v_j$  can

be represented as:

$$e_j^t = \sum_{i=1}^n a_{w(ij)}^t a_{g(ij)}^t x_i^t, \quad (9)$$

$$e_j^t = (A_{w(j)}^t \circ A_{g(j)}^t) X^t, \quad (10)$$

where the Hadamard product ( $\circ$ ) represents multiplying the elements of the corresponding position in the wind matrix  $A_{w(j)}^t$  and geography matrix  $A_{g(j)}$ .

Then we structure a model based on the GCN, which is more tailored to  $PM_{2.5}$  forecasting. First, we replace the binary adjacency matrix with a numerical matrix to introduce weights information of edges:

$$A^t = A_{w}^t \circ A_g, \quad (11)$$

$A^t$  is the adjacency matrix at time  $t$ , which is derived from the combination of the wind-field matrix  $A_{w}^t$  and the geographical matrix  $A_g$ .  $A_{w}^t$  is an input computed by real-world wind data.  $A_g$  is obtained through training and the elements measure the weaken or strengthen effect on the wind-field of the geography between stations.

$$\tilde{A}^t = A^t + \lambda I, \quad (12)$$

$\lambda I$  is the self-loop term. In this scenario, it can be explained as the natural dissipation of the station's  $PM_{2.5}$  concentration proportional to itself.  $I$  is the identity matrix and the coefficient  $\lambda$  is a hyperparameter to control the weight of the node's own influence to itself.

The standard GCN uses a symmetric normalization method based on degrees, which does not fit in the directional and weighted graphs. Therefore, we use random walk normalization to normalize the adjacency matrix.

$$\hat{A}^t = D^{-1} \tilde{A}^t, \quad (13)$$

where  $D^{-1}$  represents the inverse matrix of the degree matrix  $D$ . The sum of the weight in each row of the adjacency matrix becomes 1 after normalization. Then the next layer can be obtained by multiplying the processed adjacency matrix with the input:

$$h^t = \hat{A}^t X^t W^t, \quad (14)$$

$$h^t = D^{-1} (A_{w}^t \circ A_g + \lambda I) X^t W^t, \quad (15)$$

For an sequence input  $(X_{t-l+1}, \dots, X_t)$  of length  $l$ , there is also a dynamic graph of length  $l$  along with its corresponding adjacency matrices  $(A_{t-l+1}, \dots, A_t)$ . The inputs and adjacency matrices are matched one-to-one based on the time steps, individually processed through the graph convolution module and concatenated into a complete output of the same length  $l$ .

$$H = [h^{t-l+1}, \dots, h^t]. \quad (16)$$

#### 4.2.3. Temporal aggregation module for capturing time information

The spread of  $PM_{2.5}$  is limited by the speed of the wind-field. The impact between stations may not always manifest immediately, especially during long-range effects. Therefore, not just the previous time stamp, but also the wind-field and  $PM_{2.5}$  concentration from the previous period should be considered in the dispersion of  $PM_{2.5}$ .

Therefore, we introduce a temporal aggregation layer to capture the long-term dependencies in the dispersion of  $PM_{2.5}$ . The module is mainly composed of LSTM for its ability to dynamically capture short- and long-term dependencies. Long Short-Term Memory (LSTM) is a type of artificial neural network model commonly used for processing sequential data (Hochreiter and Schmidhuber, 1997). It is a specialized form of Recurrent Neural Networks (RNNs) designed to address issues such as vanishing gradients or exploding gradients encountered by traditional RNNs when dealing with long sequences. Compared to Transformer, LSTM has a simpler architecture and higher computational efficiency for moderate sequence lengths. As another common method, temporal convolution is less effective than LSTM in capturing long-term dependencies.

The core structure of a LSTM consists of units, each comprising an input gate, a forget gate, an output gate, and a cell state. These gate units control the flow of information through learning, allowing the model to selectively ignore or store information from input data and determine when to output the internal state of the model.

Specifically, the input gate ( $z^i$ ) controls the updating of the cell state with new input data; the forget gate ( $z^f$ ) regulates the influence of past cell states on the current state; the output gate ( $z^o$ ) determines what output is produced from the current state:

$$z = \tanh(W[x^t, h^{t-1}]), \quad (17)$$

$$z^i = \text{sigmoid}(W^i[x^t, h^{t-1}]), \quad (18)$$

$$z^f = \text{sigmoid}(W^f[x^t, h^{t-1}]), \quad (19)$$

$$z^o = \text{sigmoid}(W^o[x^t, h^{t-1}]), \quad (20)$$

where  $x^t$  and  $h^{t-1}$  represent the current input and the previous hidden state.

Then the cell state will be updated. The input gate determines how much of the new information should be added to the cell state, while the forget gate decides how much of the previous cell state should be retained.

$$c^t = z^f \circ c^{t-1} + z^i \circ z, \quad (21)$$

At last, the output ( $y^t$ ) and next hidden state ( $h^t$ ) can be obtained as follows:

$$h^t = z^o \circ \tanh(c^t). \quad (22)$$

#### 4.2.4. Prediction output module

The result of spatial convolution will be input into the LSTM layer to integrate the temporal information from the time series. By capturing the long-term dependencies, the delay of propagation between nodes will not be ignored. Finally, we use a fully connected layer as a decoder to transform the hidden layer to the prediction output aggregating the spatial and temporal information. The output can be expressed as:

$$X^{l+1} = \phi_{fc}(\text{ReLU}(\phi_t(\text{ReLU}(\phi_s(X^l))))), \quad (23)$$

where  $\phi_s$  is the spatial convolutional module, and represents the semi-dynamic directional graph convolutional module here.  $\text{ReLU}(\cdot)$  is the rectified linear unit as an activation function.  $\phi_t$  is the temporal aggregation module and  $\phi_{fc}$  is the fully connected layer. Despite its uncomplicated structure, the most distinct improvement of our model is the introduction of combining a learnable static graph with a dynamic graph. This is an improvement to edge embedding, allowing edge weights to better reflect the various relationships between nodes.

## 5. Experiment

In this section, we conducted a series of experiments to evaluate the performance of the proposed model. We also compared the performance of our model with some existing methods such as LSTM, GC-LSTM and STGCN.

### 5.1. Experiment settings

We used 3-year (2020–2022) hourly  $PM_{2.5}$  and wind-field data as our experimental dataset. The entire dataset was randomly divided into a training set (80%), a validation set (10%) and a test set (10%). This division was based on time rather than stations, as the forecasting process relies on the unique relationships between stations learned during training. The input comprises historical hourly data for  $PM_{2.5}$  over a 12-h period, while the output entails forecasting the  $PM_{2.5}$  data for the next 36 h.

To evaluate the  $PM_{2.5}$  forecasting performance of these models, we used Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) as performance metrics. The calculations are shown below:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad (24)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (25)$$

where  $\hat{y}_i$  and  $y_i$  denote the predicted value and the true value of the  $i$ th sample.

### 5.2. Baselines

We selected five baselines for comparison:

(1) Long Short Term Memory (LSTM). LSTM is a type of Recurrent Neural Network (RNN) architecture designed to overcome the vanishing gradient problem in standard RNNs. We used a typical LSTM to compare with our model. The comparison aims to demonstrate the importance of spatial information in  $PM_{2.5}$  forecasting (Zhao et al., 2019).

(2) Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM). CNN-LSTM utilizes CNN to learn features from both temporal and spatial dimensions, and process the long-term dependencies of time series data by LSTM. There have been some studies on its performance in  $PM_{2.5}$  prediction problem (Li et al., 2020).

(3) Graph Convolutional Long Short-Term Memory (GC-LSTM). GC-LSTM is a model that combines GCN with LSTM. The model leverages the spatial information captured by GCNs to enhance the predictive capabilities of LSTM for time series forecasting tasks on graph-structured data. By incorporating both spatial and temporal dependencies, GC-LSTM can effectively model the complex dynamics present in such data (Qi et al., 2019).

(4) Graph Attention Networks Long Short-Term Memory (GAT-LSTM). GAT is a commonly used graph network architecture, which aggregates the nodes' features by global or mask graph attention. Similar to GC-LSTM, GAT-LSTM can be applied in graph-based time series prediction as well (Han et al., 2021).

(5) Spatial-Temporal Graph Convolutional Network (STGCN). STGCN is a model used for handling spatio-temporal sequence data. STGCN is widely applied in fields such as traffic flow forecasting, pedestrian flow forecasting, and meteorological data analysis. It is designed to tackle complex relationships in both spatial and temporal dimensions, leveraging the advantages of Graph Convolutional Networks (GCNs) in dealing with graph-structured data (Yu et al., 2017).

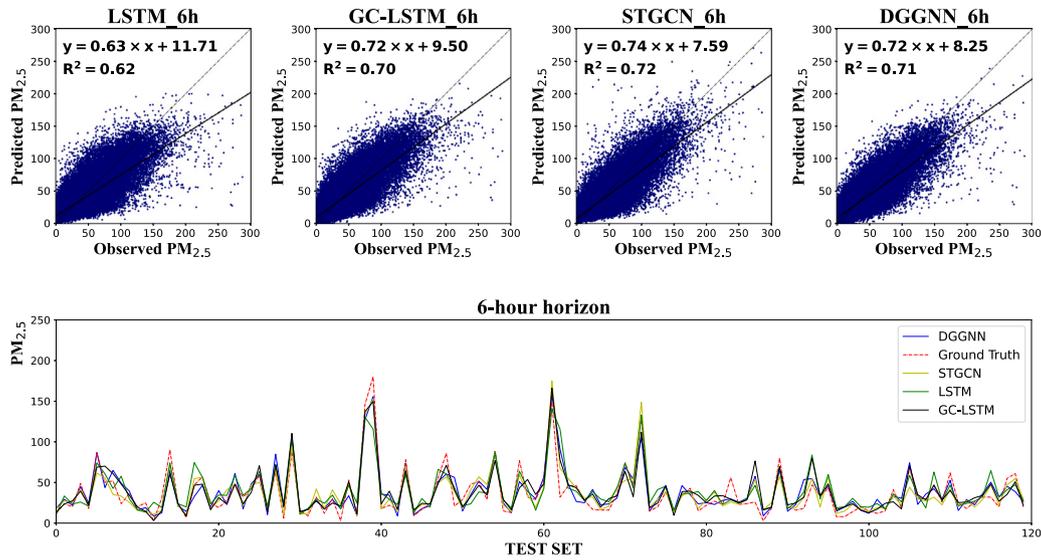


Fig. 4. Line and scatter plots on 6-h horizon.

**Table 1**  
The results of models.

Model	Metric	6 h	12 h	18 h	24 h	30 h	36 h
LSTM	RMSE	13.59	14.23	14.56	14.68	15.03	15.50
	MAE	9.89	10.30	10.48	10.58	10.85	11.21
CNN-LSTM	RMSE	<b>11.66</b>	13.86	14.55	14.53	14.75	15.27
	MAE	<b>8.04</b>	9.79	10.24	10.34	10.60	11.02
GC-LSTM	RMSE	11.74	13.02	<b>13.57</b>	<b>13.65</b>	<b>13.91</b>	14.42
	MAE	8.35	9.26	<b>9.66</b>	<b>9.75</b>	<b>9.97</b>	10.38
GAT-LSTM	RMSE	13.68	13.65	13.75	13.77	13.90	<b>14.32</b>
	MAE	9.71	9.73	9.74	9.82	9.94	<b>10.29</b>
STGCN	RMSE	11.83	<b>12.90</b>	13.93	14.19	14.64	14.91
	MAE	8.21	<b>8.99</b>	9.73	10.01	10.45	10.78
DGGNN	RMSE	<b>11.63</b>	<b>12.21</b>	<b>12.45</b>	<b>12.51</b>	<b>12.77</b>	<b>13.36</b>
	MAE	<b>8.15</b>	<b>8.63</b>	<b>8.77</b>	<b>8.87</b>	<b>9.08</b>	<b>9.54</b>

### 5.3. Comparison of model performance

We conducted a series of experiments to compare the performance of DGGNN with the baseline models (LSTM, CNN-LSTM, GC-LSTM, GAT-LSTM and STGCN). The specific results are shown in Table 1. The best result for each metric is highlighted in bold font. From the results, we can observe that the spatial-temporal convolutional models outperform the simple temporal sequence model (LSTM) significantly, which is expected. It demonstrates the crucial role of spatial convolution in PM<sub>2.5</sub> forecasting.

Overall, our proposed model achieved the best performance across almost all metrics. CNN-LSTM achieves better in the short-term forecast, it is due to the more complicated temporal convolution. Then the performance of CNN-LSTM tends to be similar to LSTM, and starts to widen the gap by the graph-based models. We believe that in short-term forecasting of PM<sub>2.5</sub>, the concentration is highly correlated with the historical states of the nodes themselves, as the spatial dispersion of PM<sub>2.5</sub> between nodes takes time. And the spatial correlation may become more pronounced in long-term PM<sub>2.5</sub> forecasting. That explains the performance of CNN-LSTM, and proves that CNN is unable to capture spatial information in non-Euclidean space well compared to GNN.

The long-term performance of GAT-LSTM and GC-LSTM is similar for the same graph structure used in both models, and it proves that

GAT and GCN have similar spatial information learning abilities. Different from other models that directly generate the whole time series predictions, the results of STGCN are generated hourly in our experiments. Therefore, the comparison does not have special significance. Through the results, we can find that the advantages of our model become more evident with increasing time spans. That is because the advantages are provided by the improvement of the spatial convolution module, which originally introduced the combination of wind-field and geographical information.

To further compare the performance of these models, we used the line charts and scatter plots showing the observed values and predicted values for the 6 h, 12 h, 24 h and 36 h forecasts in Figs. 4–7, respectively. The scatter plots contained the results of all stations on the entire test set, and the line charts used the results of one station (the 28th station located at 118.29°E, 29.70°N). Through the line charts, we can observe that DGGNN exhibits fewer faults in estimating certain high and low points of PM<sub>2.5</sub> compared with other models. By fitting the predicted values and observed values of different models, our model also achieves the best result. The correlation coefficients ( $R^2$ ) of 6 h, 12 h, 24 h and 36 h horizons are 0.71, 0.70, 0.69, and 0.63, respectively. It is significantly better than LSTM (0.62, 0.60, 0.57, 0.51) and GC-LSTM (0.70, 0.66, 0.63, 0.57). Although slightly inferior to STGCN on 6 h forecast (0.72), the results of 12 h, 24 h and 36 h still outperform STGCN (0.65, 0.60 and 0.56) a lot. The enhancement of our model is particularly evident on the 24 h and 36 h forecasts.

One of the main objectives of PM<sub>2.5</sub> forecasting is to enable people to make a healthy travel plan. Therefore, it is crucial to predict days with severe PM<sub>2.5</sub> pollution accurately. In Fig. 8, we compared the average recall rates of different models in predicting severe pollution. According to air quality standards, we selected thresholds of 75  $\mu\text{g}/\text{m}^3$ , 115  $\mu\text{g}/\text{m}^3$ , and 150  $\mu\text{g}/\text{m}^3$  to classify pollution levels. For weather conditions with mild pollution or above ( $\geq 75 \mu\text{g}/\text{m}^3$ ), our model achieved a recall of 74.55%, higher than LSTM (69.35%), GC-LSTM (69.47%) and STGCN (71.38%). For moderate pollution or above ( $\geq 115 \mu\text{g}/\text{m}^3$ ), the result in comparison are 72.37% and 66.08%, 66.18%, 66.33%. For heavily pollution ( $\geq 150 \mu\text{g}/\text{m}^3$ ), the results are 65.41% and 63.30%, 59.78%, 55.11%.

This performance superiority of the DGGNN model in severe pollution prediction can be attributed to its integration of wind-field data, which enables the dynamic graph to capture spatial relationships and pollutant dispersion patterns effectively. The wind-field integration allows the model to account for the directional flow of pollutants, which

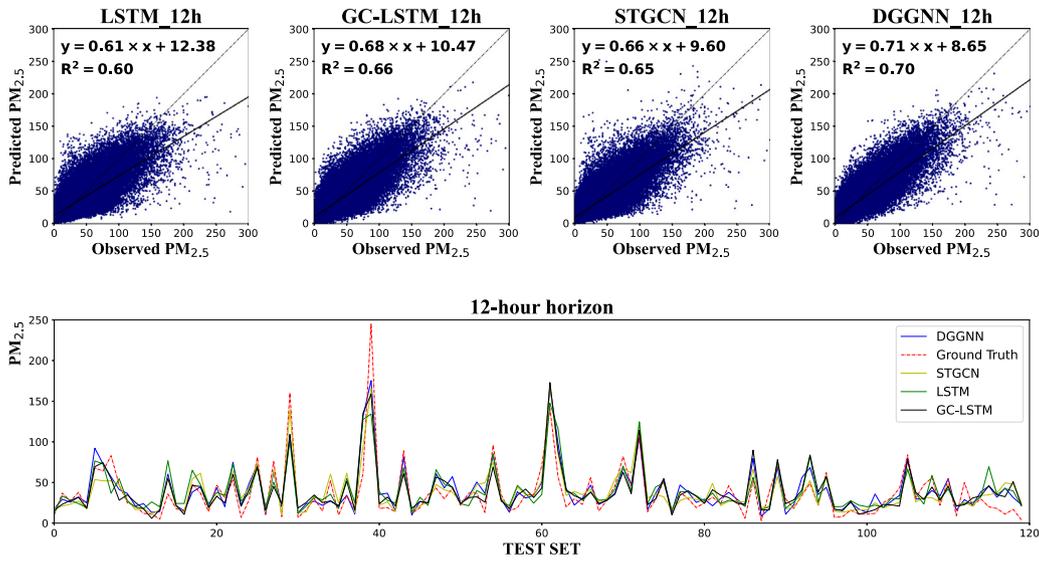


Fig. 5. Line and scatter plots on 12-h horizon.

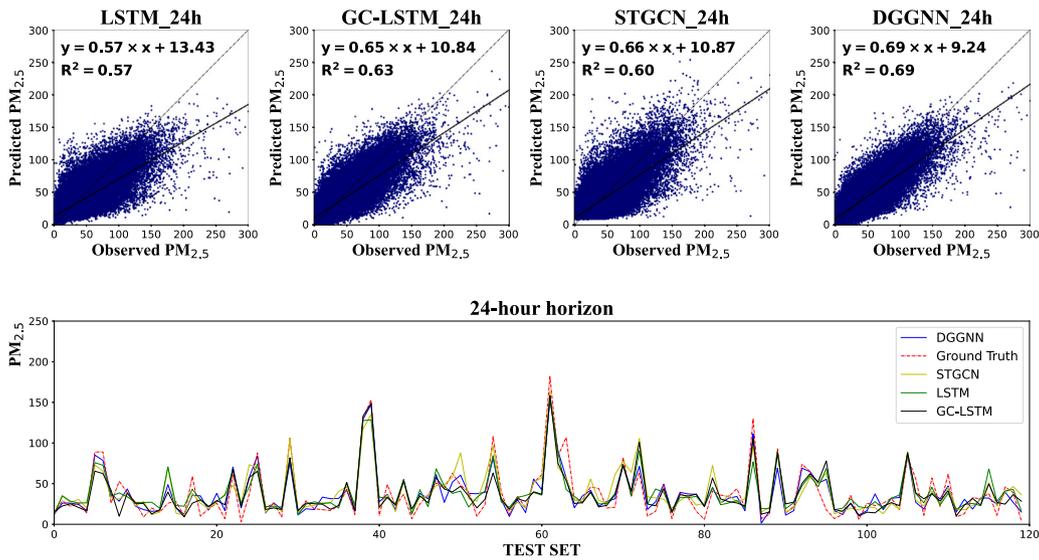


Fig. 6. Line and scatter plots on 24-h horizon.

plays a significant role during high-pollution weather events. Moreover, the combination of static and dynamic graphs in DGGNN enhances the model’s capability to represent both the inherent geographical relationships and the dynamic meteorological influences. These design components collectively enable the model to generate more accurate predictions for heavily polluted weather conditions.

To gain deeper insights, we performed an error analysis for all stations individually, as illustrated in Fig. 9. By comparing prediction errors across stations, we observed that stations with poorer performance (Stations 1, 19, 21, 22, and 31) tend to have relatively few neighboring stations (fewer than 5), which may indicate that limited spatial relationships contribute to lower accuracy. However, as the number of neighbors increases, no clear trend of further performance improvement was observed. This suggests that, due to significant differences between stations, the quantity of spatial information alone is not the primary determinant of prediction accuracy. The accuracy for

each station may also be influenced by various other factors, such as sources of air pollution, seasonal variations, and local wind patterns.

#### 5.4. Ablation study

In order to verify the impact of each module on the proposed model individually, we conducted a series of ablation experiments. The entire spatial convolutional module or the temporal aggregation module is removed to consider the spatiotemporal correlation. And we removed some information in the spatial module from the complete model to investigate their influence to the prediction of PM<sub>2.5</sub> dispersion, which include the geographical matrix  $A_g$  or the dynamic wind-field matrix  $A_w$ . The Fig. 10 shows the distinction in performance among these models.

According to the results, removing the entire spatial or temporal module leads to a significant decline in performance. Referring to

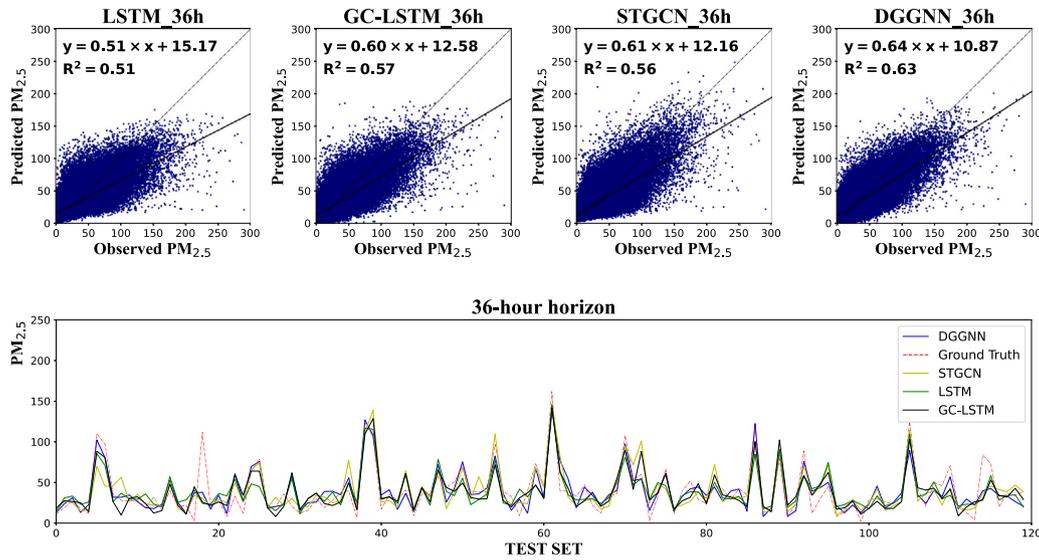


Fig. 7. Line and scatter plots on 36-h horizon.

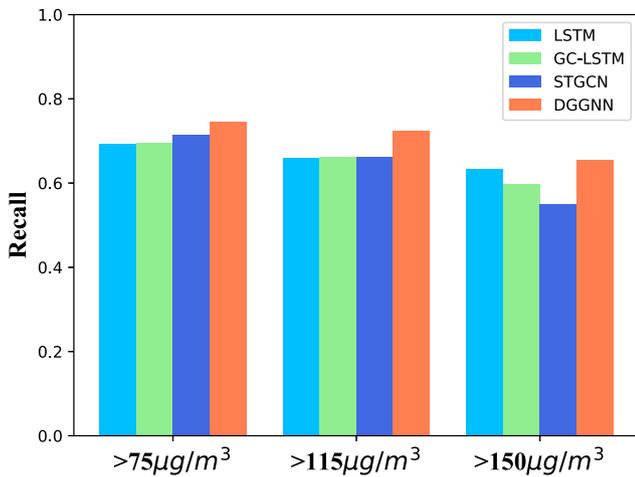


Fig. 8. The recall of different models.

the average across all horizons, there is a 18.1% performance decline without the spatial module and a 18.8% without the temporal module, which demonstrates the spatial and temporal information are both indispensable for the prediction of  $PM_{2.5}$  dispersion.

Then we focused on analyzing the role of each part of the spatial convolution module in extracting spatial information between stations. When spatial information including distance and wind-field data has been removed, only the adjacency matrix remains, which contains extremely weak geographic information (consisting solely of unweighted edges between stations), the model's performance dropped a lot due to the lack of spatial information. If specific distance information is introduced, the performance will be greatly improved, which proves the fundamental and decisive role of distance information in  $PM_{2.5}$  dispersion. On this basis, we compared the results when the wind-field matrix  $A_w$  or the learnable geographical matrix  $A_g$  is introduced. The former is similar to some previous methods, achieving better performance due to the introduction of graph dynamics, with only a 8.0% performance loss compared to the final model. While the latter unexpectedly achieved good results despite the lack of wind-field information, proving that introducing learnable matrices to capture intrinsic relationships is a feasible method, and it can, in fact, have different physical interpretations in models with varying structures. In

the end, the complete model which incorporates all of these spatial information achieved the best results, especially when the prediction length exceeds 12 h.

We also examined the model's performance in predicting individual stations, considering the actual geographic conditions. The Fig. 11 shows the performance gap of the models at two stations with different geographic conditions. The monitoring station A is located in Hangzhou, Zhejiang, situated in a plain region. The unobstructed  $PM_{2.5}$  dispersion in this area leads to only a modest improvement in our model's performance. However, for the monitoring station B in Wenzhou, Zhejiang, where is surrounded by mountains, the role of the matrix for geographic information is amplified. And the performance our model is much better than the model only with wind-field and distance information.

## 6. Conclusion

In our study, we propose a Dynamic Geographical Graph Neural Network model DGGNN to incorporate more comprehensive  $PM_{2.5}$  dispersion relationships between stations into  $PM_{2.5}$  forecasting. Unlike prior GNN-based methods that primarily construct graphs based on the explicit spatial distances between stations, our approach uniquely integrates the inherent relationships between stations into the graph representation. Specifically, we constructed a dynamic graph derived from wind-field data to capture the temporal dynamics of pollutant dispersion and a static graph to represent the invariant geographical relationships. Through the model, these hidden relationships are learned and leveraged to improve prediction accuracy and interpretability. The integration of static graph learning with spatio-temporal feature extraction, implemented through GCN and LSTM modules, allows for a more nuanced understanding of  $PM_{2.5}$  transport mechanisms. The results of our experiments indicated that the proposed model achieved better performance in fitting and forecasting than LSTM, GC-LSTM and STGCN using RMSE and MAE as metrics, with significant improvement in the longer-term forecasts (24-h, 30-h and 36-h). By comparing recall rates, our model could predict the occurrence of high-pollution weather more accurately.

Despite the promising results, several limitations remain. The substantial increase in the number of stations in real-world scenarios poses a significant challenge to the computational complexity of GNNs. Additionally, the dynamic graphs introduce dependencies that hinder computational parallelism, highlighting another area in need of optimization. Furthermore, while our approach demonstrates strong

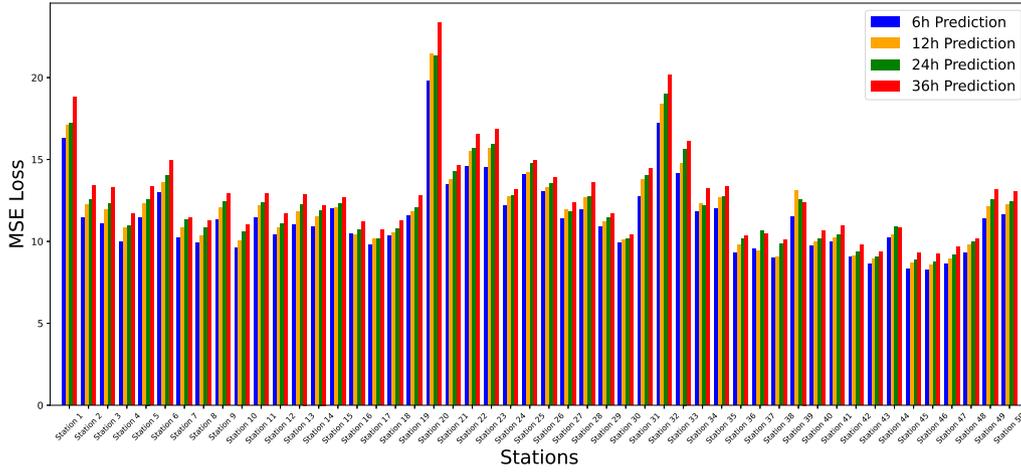


Fig. 9. Prediction error for different horizons of 50 stations.

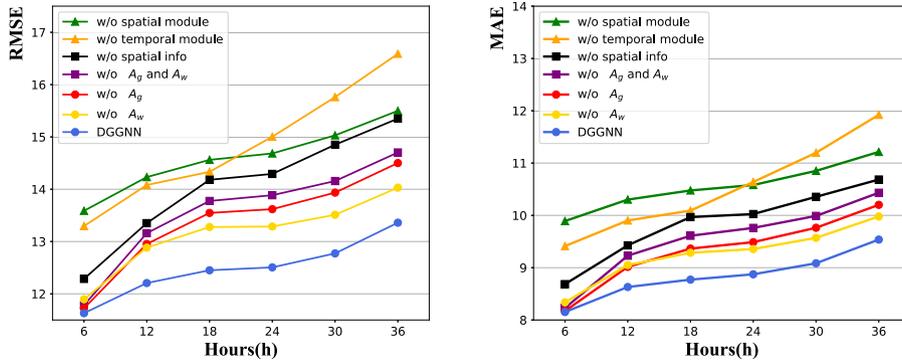


Fig. 10. Ablation experiment result.

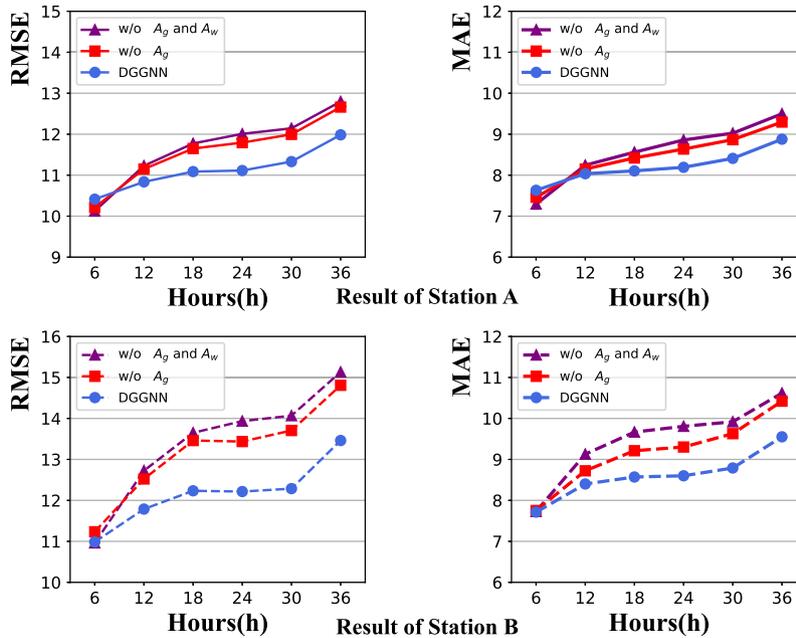


Fig. 11. Ablation experiment result of different stations.

performance on PM<sub>2.5</sub> forecasting using data from East China, its generalizability to other regions with differing geographical and meteorological dynamics remains untested. This raises the need for future studies to validate the model's applicability across diverse environments.

There are some aspects to optimize our model: (i) incorporating real-world geographic information into the model, such as detailed topographical data or urban features, to improve spatial understanding, (ii) achieving a balance between the complexity of dynamic graphs and computational efficiency, (iii) employing an improved structure to better integrate spatial and temporal convolution modules.

In the future, we will extend the method to more regions and pollutants. Considering that inherent relationships between nodes always exist, the method is also worthy of validating the dynamic graph in some other domains.

### CRedit authorship contribution statement

**Qin Zhao:** Writing – review & editing, Methodology, Formal analysis, Conceptualization. **Jiajun Liu:** Writing – original draft, Software, Investigation, Formal analysis. **Xinwen Yang:** Visualization, Data curation. **Hongda Qi:** Supervision, Resources, Formal analysis. **Jie Lian:** Supervision, Project administration, Formal analysis.

### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Qin Zhao reports financial support was provided by National Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

This work was funded in part by the National Key Research and Development Program of China under Grant No. 2022YFB4501704, and in part by the National Natural Science Foundation of China under Grant Nos. 61702333, 62372300, and 62302308.

### Data availability

No data was used for the research described in the article.

### References

Aman, N., Manomaiphiboon, K., Xian, D., Gao, L., Tian, L., Pala-En, N., Wang, Y., Wangyao, K., 2024. Spatiotemporal estimation of hourly PM<sub>2.5</sub> using AOD derived from geostationary satellite fengyun-4A and machine learning models for greater bangkok. *Air Qual. Atmos. Heal.* 1–16.

Ben Bouallègue, Z., Clare, M.C., Magnusson, L., Gascon, E., Maier-Gerber, M., Janoušek, M., Rodwell, M., Pinaut, F., Dramsch, J.S., Lang, S.T., et al., 2024. The rise of data-driven weather forecasting: A first statistical assessment of machine learning-based weather forecasts in an operational-like context. *Bull. Am. Meteorol. Soc.*

Bergen, K.J., Johnson, P.A., de Hoop, M.V., Beroza, G.C., 2019. Machine learning for data-driven discovery in solid earth geoscience. *Sci.* 363 (6433), eaau0323.

Dai, H., Huang, G., Zeng, H., Zhou, F., 2022. PM<sub>2.5</sub> volatility prediction by XGBoost-MLP based on GARCH models. *J. Clean. Prod.* 356, 131898.

Faraji, M., Nadi, S., Ghaffarpasand, O., Homayoni, S., Downey, K., 2022. An integrated 3D CNN-GRU deep learning method for short-term prediction of PM<sub>2.5</sub> concentration in urban environment. *Sci. Total Environ.* 834, 155324.

Foley, K., Roselle, S., Appel, K., Bhave, P., Pleim, J., Otte, T., Mathur, R., Sarwar, G., Young, J., Gilliam, R., et al., 2010. Incremental testing of the community multiscale air quality (CMAQ) modeling system version 4.7. *Geosci. Model Dev.* 3 (1), 205–226.

Goldberg, D.L., Gupta, P., Wang, K., Jena, C., Zhang, Y., Lu, Z., Streets, D.G., 2019. Using gap-filled MAIAC AOD and WRF-chem to estimate daily PM<sub>2.5</sub> concentrations at 1 km resolution in the eastern United States. *Atmos. Environ.* 199, 443–452.

Han, S., Dong, H., Teng, X., Li, X., Wang, X., 2021. Correlational graph attention-based long short-term memory network for multivariate time series prediction. *Appl. Soft Comput.* 106, 107377.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al., 2020. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* 146 (730), 1999–2049.

Himanen, L., Geurts, A., Foster, A.S., Rinke, P., 2019. Data-driven materials science: status, challenges, and perspectives. *Adv. Sci.* 6 (21), 1900808.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780.

Hong, J., Mao, F., Min, Q., Pan, Z., Wang, W., Zhang, T., Gong, W., 2020. Improved PM<sub>2.5</sub> predictions of WRF-chem via the integration of Himawari-8 satellite data and ground observations. *Environ. Pollut.* 263, 114451. <http://dx.doi.org/10.1016/j.envpol.2020.114451>, URL: <https://www.sciencedirect.com/science/article/pii/S0269749119334955>.

Karimian, H., Li, Q., Wu, C., Qi, Y., Mo, Y., Chen, G., Zhang, X., Sachdeva, S., et al., 2019. Evaluation of different machine learning approaches to forecasting PM<sub>2.5</sub> mass concentrations. *Aerosol Air Qual. Res.* 19 (6), 1400–1410.

Keisler, R., 2022. Forecasting global weather with graph neural networks. arXiv preprint arXiv:2202.07575.

Kim, D.-Y., Jin, D.-Y., Suk, H.-I., 2023. Spatiotemporal graph neural networks for predicting mid-to-long-term PM<sub>2.5</sub> concentrations. *J. Clean. Prod.* 425, 138880.

Kipf, T.N., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., et al., 2022. GraphCast: Learning skillful medium-range global weather forecasting. arXiv preprint arXiv:2212.12794.

Li, L., Gong, J., Zhou, J., 2014. Spatial interpolation of fine particulate matter concentrations using the shortest wind-field path distance. *PLoS One* 9 (5), e96111.

Li, T., Hua, M., Wu, X., 2020. A hybrid CNN-LSTM model for forecasting particulate matter (PM<sub>2.5</sub>). *IEEE Access* 8, 26933–26940.

Li, X., Sun, L., Ling, M., Peng, Y., 2023. A survey of graph neural network based recommendation in social networks. *Neurocomputing* 549, 126441.

Li, J., Yu, S., Chen, X., Zhang, Y., Li, M., Li, Z., Song, Z., Liu, W., Li, P., Xie, M., et al., 2022. Evaluation of the WRF-CMAQ model performances on air quality in China with the impacts of the observation nudging on meteorology. *Aerosol Air Qual. Res.* 22 (4), 220023.

Ma, M., Xie, P., Teng, F., Wang, B., Ji, S., Zhang, J., Li, T., 2023. HiSTGNN: Hierarchical spatio-temporal graph neural network for weather forecasting. *Inform. Sci.* 648, 119580.

Ma, J., Yu, Z., Qu, Y., Xu, J., Cao, Y., et al., 2020. Application of the xgboost machine learning method in PM<sub>2.5</sub> prediction: A case study of shanghai. *Aerosol Air Qual. Res.* 20 (1), 128–138.

Muthukumar, P., Cocom, E., Nagrecha, K., Comer, D., Burga, I., Taub, J., Calvert, C.F., Holm, J., Pourhomayoun, M., 2021. Predicting PM<sub>2.5</sub> atmospheric air pollution using deep learning with meteorological data and ground-based observations and remote-sensing satellite big data. *Air Qual. Atmos. Heal.* 1–14.

Ouyang, X., Yang, Y., Zhang, Y., Zhou, W., Guo, D., 2023. Dual-channel spatial-temporal difference graph neural network for PM<sub>2.5</sub> forecasting. *Neural Comput. Appl.* 35 (10), 7475–7494.

Park, S., Lee, J., Im, J., Song, C.-K., Choi, M., Kim, J., Lee, S., Park, R., Kim, S.-M., Yoon, J., et al., 2020. Estimation of spatially continuous daytime particulate matter concentrations under all sky conditions through the synergistic use of satellite-based AOD and numerical models. *Sci. Total Environ.* 713, 136516.

Qi, Y., Li, Q., Karimian, H., Liu, D., 2019. A hybrid model for spatiotemporal forecasting of PM<sub>2.5</sub> based on graph convolutional neural network and long short-term memory. *Sci. Total Environ.* 664, 1–10.

Qin, D., Yu, J., Zou, G., Yong, R., Zhao, Q., Zhang, B., 2019. A novel combined prediction scheme based on CNN and LSTM for urban PM<sub>2.5</sub> concentration. *IEEE Access* 7, 20050–20059.

Saïde, P.E., Carmichael, G.R., Spak, S.N., Gallardo, L., Osses, A.E., Mena-Carrasco, M.A., Pagowski, M., 2011. Forecasting urban PM<sub>10</sub> and PM<sub>2.5</sub> pollution episodes in very stable nocturnal conditions and complex terrain using WRF-chem CO tracer model. *Atmos. Environ.* 45 (16), 2769–2780.

Tan, J., Liu, H., Li, Y., Yin, S., Yu, C., 2022. A new ensemble spatio-temporal PM<sub>2.5</sub> prediction method based on graph attention recursive networks and reinforcement learning. *Chaos Solitons Fractals* 162, 112405.

Teng, M., Li, S., Xing, J., Fan, C., Yang, J., Wang, S., Song, G., Ding, Y., Dong, J., Wang, S., 2023. 72-Hour real-time forecasting of ambient PM<sub>2.5</sub> by hybrid graph deep neural network with aggregated neighborhood spatiotemporal information. *Environ. Int.* 176, 107971.

Tsai, Y.-T., Zeng, Y.-R., Chang, Y.-S., 2018. Air pollution forecasting using RNN with LSTM. In: 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). IEEE, pp. 1074–1079.

Udristiou, M.T., Mghouchi, Y.E., Yildizhan, H., 2023. Prediction, modelling, and forecasting of PM and AQI using hybrid machine learning. *J. Clean. Prod.* 421, 138496.

- Wang, Z., Maeda, T., Hayashi, M., Hsiao, L.-F., Liu, K.-Y., 2001. A nested air quality prediction modeling system for urban and regional scales: Application for high-ozone episode in Taiwan. *Water Air Soil Pollut.* 130, 391–396.
- Wang, H., Zhang, L., Wu, R., Cen, Y., 2023. Spatio-temporal fusion of meteorological factors for multi-site PM<sub>2.5</sub> prediction: A deep learning and time-variant graph approach. *Environ. Res.* 239, 117286.
- Wu, S., Li, H., 2022. Prediction of PM<sub>2.5</sub> concentration in urban agglomeration of China by hybrid network model. *J. Clean. Prod.* 374, 133968.
- Wu, F., Min, P., Jin, Y., Zhang, K., Liu, H., Zhao, J., Li, D., 2023. A novel hybrid model for hourly PM<sub>2.5</sub> prediction considering air pollution factors, meteorological parameters and GNSS-ZTD. *Environ. Model. Softw.* 167, 105780.
- Xiao, X., Jin, Z., Wang, S., Xu, J., Peng, Z., Wang, R., Shao, W., Hui, Y., 2022. A dual-path dynamic directed graph convolutional network for air quality prediction. *Sci. Total Environ.* 827, 154298.
- Xu, W., Jin, X., Liu, M., Ma, Z., Zhou, Y., et al., 2021. Analysis of spatiotemporal variation of PM<sub>2.5</sub> and its relationship to land use in China. *Atmos. Pollut. Res.* 12 (9), 101151.
- Ye, Z., Kumar, Y.J., Sing, G.O., Song, F., Wang, J., 2022. A comprehensive survey of graph neural networks for knowledge graphs. *IEEE Access* 10, 75729–75741.
- Yeganeh, B., Hewson, M.G., Clifford, S., Knibbs, L.D., Morawska, L., 2017. A satellite-based model for estimating PM<sub>2.5</sub> concentration in a sparsely populated environment using soft computing techniques. *Environ. Model. Softw.* 88, 84–92.
- Yousefi, R., Shaheen, A., Wang, F., Ge, Q., Wu, R., Lelieveld, J., Wang, J., Su, X., 2023. Fine particulate matter (PM<sub>2.5</sub>) trends from land surface changes and air pollution policies in China during 1980–2020. *J. Environ. Manag.* 326, 116847.
- Yu, M., Masrur, A., Blaszcak-Boxe, C., 2023. Predicting hourly PM<sub>2.5</sub> concentrations in wildfire-prone areas using a SpatioTemporal transformer model. *Sci. Total Environ.* 860, 160446.
- Yu, B., Yin, H., Zhu, Z., 2017. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*.
- Zhan, Y., Luo, Y., Deng, X., Chen, H., Grieneisen, M.L., Shen, X., Zhu, L., Zhang, M., 2017. Spatiotemporal prediction of continuous daily PM<sub>2.5</sub> concentrations across China using a spatially explicit machine learning algorithm. *Atmos. Environ.* 155, 129–139.
- Zhang, B., Chen, W., Li, M.-Z., Guo, X., Zheng, Z., Yang, R., 2024. MGAtt-LSTM: A multi-scale spatial correlation prediction model of PM<sub>2.5</sub> concentration based on multi-graph attention. *Environ. Model. Softw.* 106095.
- Zhang, X.-M., Liang, L., Liu, L., Tang, M.-J., 2021. Graph neural networks and their current applications in bioinformatics. *Front. Genet.* 12, 690049.
- Zhang, Q., Wu, S., Wang, X., Sun, B., Liu, H., 2020a. A PM<sub>2.5</sub> concentration prediction model based on multi-task deep learning for intensive air quality monitoring stations. *J. Clean. Prod.* 275, 122722.
- Zhang, B., Zhang, H., Zhao, G., Lian, J., 2020b. Constructing a PM<sub>2.5</sub> concentration prediction model by combining auto-encoder with Bi-LSTM neural networks. *Environ. Model. Softw.* 124, 104600.
- Zhao, J., Deng, F., Cai, Y., Chen, J., 2019. Long short-term memory-fully connected (LSTM-FC) neural network for PM<sub>2.5</sub> concentration prediction. *Chemosphere* 220, 486–492.
- Zhao, L., Li, Z., Qu, L., 2024. A novel machine learning-based artificial intelligence method for predicting the air pollution index PM<sub>2.5</sub>. *J. Clean. Prod.* 468, 143042.
- Zhou, H., Zhang, F., Du, Z., Liu, R., 2021. Forecasting PM<sub>2.5</sub> using hybrid graph convolution-based model considering dynamic wind-field to offer the benefit of spatial interpretability. *Environ. Pollut.* 273, 116473.
- Zhou, H., Zhang, F., Du, Z., Liu, R., 2022. A theory-guided graph networks based PM<sub>2.5</sub> forecasting method. *Environ. Pollut.* 293, 118569.