

## Research Article

# PM2.5 Concentration Forecasting in Industrial Parks Based on Attention Mechanism Spatiotemporal Graph Convolutional Networks

Qingtian Zeng <sup>1</sup>, Chao Wang <sup>1</sup>, Geng Chen <sup>1</sup> and Hua Duan <sup>2</sup>

<sup>1</sup>College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China

<sup>2</sup>College of Mathematics and System Science, Shandong University of Science and Technology, Qingdao 266590, China

Correspondence should be addressed to Geng Chen; gengchen@sdust.edu.cn

Received 3 August 2021; Accepted 15 October 2021; Published 3 November 2021

Academic Editor: Chunguo Li

Copyright © 2021 Qingtian Zeng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Industrial parks are one of the main sources of air pollution; the ability to forecast PM2.5, the main pollutant in the industrial park, is of great significance to the health of the workers in the industrial park and environmental governance, which can improve the decision-making ability of environmental management. Most of the existing PM2.5 concentration forecast methods lack the ability to model the dynamic temporal and spatial correlations of PM2.5 concentration. In an industrial park environment, in order to improve the accuracy of PM2.5 concentration forecast, based on deep learning technology, this paper proposes a spatiotemporal graph convolutional network based on the attention mechanism (STAM-STGCN) to solve the PM2.5 concentration forecast problem. When constructing the adjacency matrix, we not only use the Euclidean distance between sites but also consider the impact of wind fields and the impact of pollution sources near the nodes. In the process of model construction, we first use the spatiotemporal attention mechanism to capture the dynamic spatiotemporal correlations in PM2.5 data. In the spatiotemporal convolution module, we use graph convolutional neural networks to capture spatial features and standard convolution to describe temporal features. Finally, the output module adjusts the output shape of the data to produce the final forecast result. In this paper, the mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) are used as the performance evaluation metrics of the model, and the Dongmingnan Industrial Park atmospheric dataset is used to verify the effectiveness of the proposed algorithm. The experimental results show that our STAM-STGCN model can more fully capture the spatial-temporal characteristics of PM2.5 concentration data; compared with the most advanced model in the comparison model, the RMSE can be improved about 24.2%, the MAE is improved about 35.8%, and the MAPE is improved about 34.6%.

## 1. Introduction

PM2.5 refers to the particulate matter with a diameter less than or equal to 2.5 microns in the atmosphere, also known as fine particulate matter or particulate matter that can enter the lung [1]. Scientists use PM2.5 concentration to indicate the content of such particles per cubic meter of air; the higher the value, the more serious the air pollution. The main sources are industrial fuels, dust, motor vehicle exhaust, photochemical smog, and other pollutants [2]. Although fine particulate matter is only a small component of the earth's atmosphere, it has an important impact on

air quality and visibility. Compared with coarser atmospheric particulate matter, fine particulate matter has a small particle size and is rich in a large amount of toxic and harmful substances stay in the atmosphere for a long time, so they have a greater impact on human health and the quality of the atmospheric environment. There is a high correlation between the time of exposure to a high concentration of PM2.5 environment and mortality [3]. With the rapid development of our country's economy, the process of industrialization and urbanization has accelerated, and air pollution problems caused by air pollutants mainly PM2.5 have become more and more prominent [4], which has caused

more and more serious impacts on people's production and life. Therefore, forecasting the PM2.5 concentration of industrial parks as the main source of air pollution is of great significance to human health and environmental governance.

PM2.5 concentration has nonlinear characteristics in time and space [5] and has a very complicated formation mechanism and process [6]. Due to this complexity, it is difficult to forecast PM2.5 concentration. Therefore, it has become a hot topic of PM2.5 forecast to propose a model that can fully exploit this complex feature. Atmospheric data in industrial parks is closely related to time and is a typical time series data. Therefore, forecasting the PM2.5 concentration in industrial parks is essentially a forecast of time series data. At this stage, there are many mature time series forecasting methods, including support vector regression (SVR) [7], autoregressive moving average model (ARMA) [8], and BP neural network [9]. However, with the exponential increase in the amount of time series data and the increase in complexity, these traditional forecast methods can only extract relatively simple linear features, and it is difficult to effectively extract more complex nonlinear features and the training time is too long, and the forecast accuracy is limited and cannot meet the actual demand.

In recent years, with the rise of deep learning applications in various fields, many excellent algorithms have been developed, including recurrent neural network (RNN) [10] and long short-term memory (LSTM) networks [11]. These algorithms focus on modelling the temporal correlation of data, while ignoring the modelling of spatial correlation of data. In PM2.5 concentration forecast, in order to consider the spatial dependence between each monitoring station, people began to combine spatial learning and time series learning [12], which can make full use of the spatial and temporal dependence of data [13]. Literature [14] proposed a hybrid CNN-LSTM model to forecast PM2.5 concentration; compared with the LSTM model, the experimental results show that the accuracy of the model exceeds the LSTM model, which proves the effectiveness of the hybrid CNN-LSTM model. Literature [15] combined the convolutional neural network and bidirectional gated loop unit to achieve multistep wind speed forecast. Literature [16] constructed a CNN-LSTM hybrid model to forecast the ozone concentration in Beijing. Literature [17] combined attention-based bidirectional gated recurrent neural network and convolutional neural network for emotion classification and conducted experiments on four public datasets; the results show that the proposed model is better than the existing model. However, these regular convolution operations based on the CNN model [18] are only suitable for processing grid-structured data, in order to solve this problem; [19] constructed a site topology map based on the spatial distribution of the site to describe the spatial relationship; through graph convolutional neural network, the spatial relationship between monitoring stations is extracted. [20] uses graph convolution as a spatial convolution operation and use one-dimensional convolution in time to form a spatiotemporal convolution block; the experimental results show that the model can effectively capture the spatiotemporal correla-

tions of data and has achieved good results in traffic forecast. Literature [21] proposed a novel attention-based spatial-temporal graph convolutional network model to solve traffic flow forecasting problem; the results show that the spatio-temporal attention mechanism can effectively capture the dynamic spatial-temporal correlations in traffic data. In order to more fully capture the spatial correlation of data, [13] proposed a dynamic directed spatiotemporal graph convolutional network, which uses directed graph time series to describe the topological relationship between vertices and vertices; the wind field diffusion distance is used to replace the traditional Euclidean distance to describe the proximity relationship between the vertices; the experimental results show that the forecast ability of this model is better than that of the comparison model. It can be seen from the above that the graph convolutional neural network can model irregular graph structure data, and the ability to extract spatial correlation is ideal.

The PM2.5 concentration of each monitoring station in the industrial park has more complex nonlinear characteristics under the influence of various pollution sources, and the influence of external factors on the PM2.5 concentration is difficult to express. In conclusion, the challenges and problems of PM2.5 concentration forecast in industrial parks mainly include the following three points: First, how to make full use of the time and space characteristics of the PM2.5 concentration data in the industrial park and dynamically capture this characteristics? Secondly, how to make full use of other data to make effective forecast, for example, wind direction and wind speed, etc.? Finally, which network structure should be adopted to meet the above two requirements? Existing forecast methods, for example, [14, 20] cannot simultaneously model the temporal and spatial characteristics and dynamic correlations of PM2.5 concentration data. In order to solve this problem, we propose a spatiotemporal graph convolutional network based on the attention mechanism (STAM-STGCN), which is used to centrally forecast the PM2.5 concentration of each monitoring station in the industrial park; this model can capture the dynamic spatial-temporal characteristics of data more effectively. The main contributions of this paper are summarized as follows:

- (1) Use the number of pollution sources and the dynamic wind field information to jointly construct an adjacency matrix to define the spatial correlation between the stations, so that the spatial relationship can be better described
- (2) The dynamic spatiotemporal correlations of PM2.5 data is modelled through the spatiotemporal attention mechanism. Spatial attention is used to model the dynamic spatial correlation of monitoring stations in different locations in the industrial park, and temporal attention is used to capture the dynamic temporal correlation between different times
- (3) Combining graph convolution and temporal convolution to construct a spatiotemporal convolution module for extracting the spatiotemporal correlations of data

- (4) Experiments were carried out on the atmospheric dataset of Dongmingnan Industrial Park. Compared with the comparison model, our model achieved the best forecast results

This paper is divided into five sections, and the rest of the structure is as follows. In Section 2, we introduce the source of the dataset, complete data preprocessing, and data analysis. In Section 3, the construction of graph data and the network structure of the forecast model are introduced in detail. In Section 4, the experimental results are displayed and analysed. Finally, we summarize and forecast the paper in Section 5.

## 2. Dataset

**2.1. Data Sources.** The dataset used in this paper comes from the real atmospheric data of Dongmingnan Industrial Park; the panoramic view of Dongmingnan Industrial Park is shown in Figure 1. The equipment that collects these data is mainly IoT sensing equipment that monitors the smoke and toxic and harmful gases emitted by the industrial park, as shown in Figure 2. These sensing equipment for atmospheric monitoring are distributed on the boundary of the park, the boundary of the enterprise, the inside of the enterprise, the sensitive area, and the mobile monitoring point according to the layout principle of points, lines, and surfaces. Through the atmospheric monitoring gateway device, the data collected by the monitoring atmospheric sensing device is uploaded to the database using 4G or wired network, and the upload frequency is 30 seconds. We selected 9 monitoring stations with relatively complete data, and the time span is from 0:00 on August 25, 2020 to 0:00 on February 2, 2021. The information of the dataset is shown in Table 1.

**2.2. Data Preprocessing.** The data preprocessing in this paper mainly has the following three steps; the flow chart is shown in Figure 3.

- (1) For the missing values of the data we obtained, the value of the monitoring station with the largest correlation coefficient is used to fill in
- (2) The frequency of data collected by the atmospheric monitoring station is 30 seconds, but due to network delays and other reasons, the data interval actually stored in the database is not strictly 30 seconds, resulting in messy data. So, we resample the data and adjust the time interval to 10 minutes to ensure the regularity of the dataset
- (3) Standardize the data through the z-score method to speed up the training process

**2.3. Data Analysis.** We visualize part of the PM<sub>2.5</sub> concentration data, as shown in Figure 4. It can be seen from the figure that the time of the highest PM<sub>2.5</sub> concentration occurs in the morning; the concentration value gradually drops to the bottom in the afternoon and then gradually rises at night, until the next morning, basically cyclical

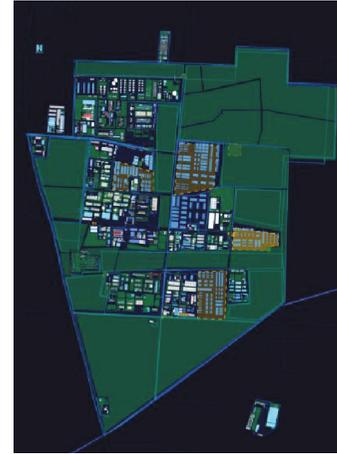


FIGURE 1: The panorama of Dongmingnan Industrial Park.

changes; and there is a strong correlation between different monitoring stations, and there are differences in numerical values. For example, the 7# monitoring station has the lowest PM<sub>2.5</sub> concentration, because the 7# monitoring station is located at the edge of Dongmingnan Industrial Park and there is no pollution source nearby.

The pollution sources around the monitoring station greatly affect the concentration of PM<sub>2.5</sub>, but this effect is not static; the influence of pollution sources on PM<sub>2.5</sub> concentration is dynamic, the uneven distribution of pollution sources in space affects the spatial correlation between sites, and this spatial correlation is also dynamic. Therefore, when designing the forecast framework, an effective method is needed to simultaneously capture the dynamic temporal and spatial dependence of PM<sub>2.5</sub> concentration.

## 3. Graph Data Construction and Forecast Model

In this section, according to the characteristics of the industrial park, we first introduce the construction idea of the graph of the industrial park monitoring station and the construction method of graph data and then introduce our deep learning forecast model.

**3.1. Graph Data Construction.** In the industrial park scenario, it is impractical to construct grid data, so we construct the distribution map of the monitoring stations in the industrial park into a graph to characterize the spatial correlation between the monitoring stations.

We abstract the spatial distribution of  $N$  monitoring stations at a certain time as a graph  $G = (V, E, A)$ ; under the influence of the wind field, this graph is a directed graph, where  $V$  is a limited set of monitoring stations;  $E$  is the edge set;  $A \in \mathbb{R}^{N \times N}$  is the weighted adjacency matrix of the graph. Each monitoring station on the graph detects the PM<sub>2.5</sub> concentration value at the same sampling frequency, thereby composing the graph sequence data, as shown in Figure 5.

The PM<sub>2.5</sub> concentration forecast problem is essentially a time series forecast problem, that is, using historical PM<sub>2.5</sub> concentration data of  $m$  continuous time steps to forecast



FIGURE 2: Atmosphere monitoring miniature sensing equipment.

TABLE 1: Dataset information.

Parameters	Value
Location	Dongmingnan Industrial Park
Span of time	2020/08/25~2021/02/02
Time interval	10 minutes
Type of data	PM2.5 and wind direction

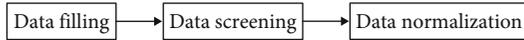


FIGURE 3: Data preprocessing.

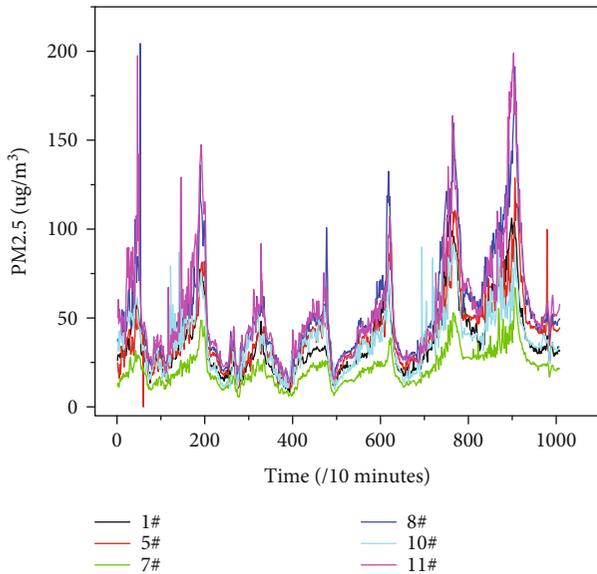


FIGURE 4: Line graph of PM2.5 concentration in some monitoring stations.

the PM2.5 concentration of  $n$  continuous time steps in the future. We use the sliding window method to divide the data, as shown in Figure 6, where  $X_t$  represents the graph data composed of all monitoring stations at time  $t$ .

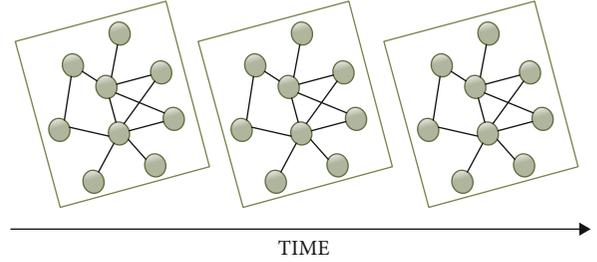


FIGURE 5: The spatiotemporal structure of PM2.5 concentration data.

The graph can represent the spatial relationship between geospatial data. When we forecast the PM2.5 concentration, we need to consider the spatial relationship between monitoring stations. In general, we use the Euclidean distance between sites to represent the spatial association between sites; this value can be understood as the difficulty of interaction between sites. PM2.5 can diffuse completely freely and will be affected by wind. Therefore, in the industrial park scenario, we need to consider the impact of wind [13]. The number of pollution sources around the monitoring station largely affects the PM2.5 concentration; the uneven distribution of pollution sources in space affects the correlation between the stations, so we also need to consider the pollution sources around the stations.

In order to obtain the influence of wind on the monitoring stations, we introduced the Gaussian diffusion model. The Gaussian diffusion model is a standard model to solve the problem of wind field diffusion; the basic formula of the model is as follows:

$$C_0(x, y, z, u) = \frac{Q}{\pi u \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2} - \frac{z^2}{2\sigma_z^2}\right), \quad (1)$$

where  $C_0$  represents the concentration of air pollutants.  $x$  and  $y$  represent the downwind distance and the horizontal distance between the point of interest and the centerline, respectively.  $z$  represents the height of the pollution source.  $u$  is the horizontal wind speed.  $\sigma_y$  and  $\sigma_z$  represent the standard deviation of horizontal dispersion and the standard deviation of vertical dispersion, respectively.  $Q$  is the source strength, which means the amount of pollutants discharged per unit time. When only horizontal diffusion is considered, the formula can be simplified to Equation (2) [22].

$$\text{cost}(E_{AB}) = [F(D_A, D_M) + F(D_B, D_M)] \times L_{AB}, \quad (2)$$

where  $A$  and  $B$  are the starting point and the end point;  $\text{cost}(E_{AB})$  is used to describe the difficulty of air pollutant diffusion from point  $A$  to point  $B$ ;  $E_{AB}$  is the edge between two points;  $D_A$  and  $D_B$  represent the wind direction azimuth of these two points;  $D_M$  is the azimuth of  $E_{AB}$ ;  $L_{AB}$  is the length of  $E_{AB}$ , that is, the distance between point  $A$  and point  $B$ , the unit is kilometres;  $F$  is a function to calculate the absolute value of azimuth difference.

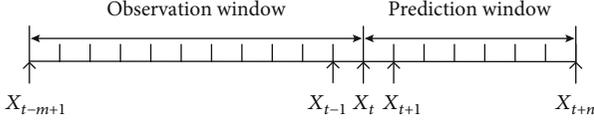


FIGURE 6: Sliding window method.

Due to the limited geographical space of the industrial park, the wind direction of each monitoring station at the same time can be regarded as the same, so the formula can be simplified as Equation (3), and the constant term can be omitted.

$$\text{cost}(E_{AB}) = [2 \times F(D_A, D_M)] \times L_{AB}. \quad (3)$$

In order to measure the impact of pollution sources on monitoring stations, we set the distance  $S$ , take a certain monitoring station as the center and the number of pollution sources within a radius of  $S$  as the number of pollution sources affecting the monitoring station and  $S$  defaults to 0.2 km. Since we cannot know the types of pollution sources around the monitoring station, we can only know the approximate number of pollution sources, in order to reduce the error caused by uncertain pollution source types, we do not directly use the quantity data of pollution sources, but by setting the scope, we obtain the influence coefficient  $c$  of the quantity of pollution sources on the monitoring station. Table 2 shows our division of impact coefficients, where  $r$  is a variable, representing the span of the number of pollution sources, and the default is 3. For example, if there are six pollution sources around monitoring station 1, then its influence coefficient is 3, that is,  $c_1$  is equal to 3.

Based on the above analysis, we express the weighted adjacency matrix as

$$A_{ij} = \begin{cases} \frac{1}{\text{cost}(E_{ij}) \times c_i}, & i \neq j, \\ 0, & i = j, \end{cases} \quad (4)$$

where  $c_i$  represents the influence coefficient of the number of pollution sources on the monitoring station  $i$ ,  $\text{cost}(E_{ij})$  means the difficulty of air pollutant diffusion from site  $i$  to station  $j$ , and the value of  $\text{cost}(E_{ij})$  will be relatively large; we normalized it. Here, we use the min-max normalization method for normalization, so that the data is distributed in a smaller range, and the data distribution is more reasonable.

**3.2. Proposed Forecasting Model.** We use  $x_t^i$  to represent PM2.5 concentration of node  $i$  at time  $t$ ,  $X_t = (x_t^1, x_t^2, \dots, x_t^N)^T \in \mathbb{R}^{N \times 1}$  represents PM2.5 concentration of all nodes at time  $t$ ,  $\chi = (X_1, X_2, \dots, X_m)^T \in \mathbb{R}^{N \times 1 \times m}$  denotes PM2.5 concentration of all nodes over  $m$  time slices,  $y^\wedge^i = (\hat{y}_{m+1}^i, \hat{y}_{m+2}^i, \dots, \hat{y}_{m+n}^i)$  represents the forecasted value of node  $i$  in the next  $n$  time slices,  $\hat{Y} = (y^\wedge^1, y^\wedge^2, \dots, y^\wedge^N)^T \in \mathbb{R}^{N \times n}$  represents the forecasted value of all nodes in the next  $n$  time slices, and  $Y = (y^1, y^2, \dots, y^N)^T \in \mathbb{R}^{N \times n}$  represents the true

TABLE 2: Division of influence coefficient.

Influence coefficient	0- $r$	$r-2r$	$\geq 2r$
$c$	1	2	3

value of all nodes in the next  $n$  time slices. PM2.5 concentration forecast can be expressed as follows:

$$\hat{Y} = f(\chi, A), \quad (5)$$

where  $f$  indicates the forecast method and  $A$  is the adjacency matrix. That is to say, the PM2.5 concentration data of all monitoring stations in the past  $m$  moments are used to forecast the PM2.5 concentration of all monitoring stations in the future  $n$  moments.

Figure 7 shows the overall framework of the proposed STAM-STGCN model, which is composed of the spatiotemporal attention block, spatiotemporal convolution block, and output block. The details of each part are as follows.

**3.2.1. Spatiotemporal Attention Block.** This block contains two kinds of attention, namely, spatial attention and temporal attention, which are used to capture the dynamic spatiotemporal correlation of data.

In the time dimension, PM2.5 concentrations in different time periods are correlated, but the correlation is different in different situations. We use the temporal attention mechanism to adaptively assign different weights to the data to capture this dynamic correlation.

$$E = V_e \cdot \sigma \left( \left( (\chi)^T U_1 \right) U_2 (U_3 \chi) + b_e \right), \quad (6)$$

where  $V_e$ ,  $b_e$ ,  $U_1$ ,  $U_2$ , and  $U_3$  are learnable parameters,  $\chi = (X_1, X_2, \dots, X_m)^T \in \mathbb{R}^{N \times 1 \times m}$  is the input of the model,  $\sigma$  is the activation function, and  $E$  is the time attention matrix; we normalize  $E$  by softmax function. We perform the Hadamard product of the normalized time attention matrix and the input  $\chi$  to obtain the dynamically adjusted input  $\chi'$ .

In the spatial dimension, we use the spatial attention mechanism to adaptively capture the dynamic correlation between nodes.

$$S = V_s \cdot \sigma \left( \left( \chi' W_1 \right) W_2 \left( W_3 \chi' \right)^T + b_s \right), \quad (7)$$

where  $V_s$ ,  $b_s$ ,  $W_1$ ,  $W_2$ , and  $W_3$  are learnable parameters and  $\sigma$  is the activation function. The spatial attention matrix  $S \in \mathbb{R}^{N \times N}$  is dynamically calculated based on the input and  $S_{ij}$  represents the correlation strength between node  $i$  and node  $j$ . We use the softmax function to ensure that the sum of the attention weights between nodes is 1; then, we accompany the adjacency matrix  $A$  with the spatial attention matrix  $S$  to dynamic adjust the weights between nodes and get the dynamically adjusted adjacency matrix  $A'$  and  $\chi''$ .  $\chi''$  is the output of the spatial attention module, which is

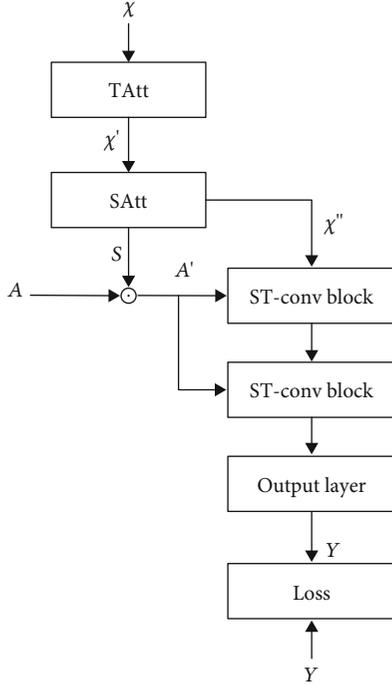


FIGURE 7: The framework of STAM-STGCN. TAtt is temporal attention. SAtt is spatial attention. ST-Conv Block is a spatiotemporal convolutional block.  $S$  is the spatial attention matrix.  $A$  is the adjacency matrix.  $\odot$  is the Hadamard product.

TABLE 3: Parameters of training model.

Parameters	Value
Observation window	144
Forecast window	1
Epoch	100
Batch size	128
The size of the graph convolution kernel	3
The size of the temporal convolution kernel	3
Learning rate	0.001
Optimization technology	Adam
Loss function	MSE
Training set: validation set: test set	8 : 1 : 1

the product of the output of the temporal attention module and the spatial attention matrix  $S$ .

**3.2.2. Spatiotemporal Convolutional Block.** This block uses the method of Yu et al. and the methods description partly reproduces their wording [20]. Spatiotemporal convolutional block is composed of two temporal convolution layers and one spatial graph convolution layer. In this module, the input data is first convolved in the time dimension, and then, the output result is convolved in the graph. The output result of the graph convolution is subjected to an activation function, and then, a time-dimension convolution is per-

TABLE 4: RMSE, MAE, and MAPE of each model.

Model	RMSE	MAE	MAPE (%)
HA	29.078	20.764	32.984
LSTM	22.437	16.794	28.796
GRU	20.583	14.530	24.211
STGCN	17.274	12.994	23.710
STAM-STGCN	13.085	8.334	15.506

formed. The calculation process is as follows:

$$O = \Gamma_1 *_{\tau} \text{ReLU}(\theta *_{\text{g}} (\Gamma_0 *_{\tau} \chi')), \quad (8)$$

where  $\Gamma_0$  and  $\Gamma_1$  are the upper and lower temporal kernel within block, respectively;  $\theta$  is the convolution kernel of graph convolution;  $\text{ReLU}$  denotes the rectified linear units function;  $*_{\tau}$  denotes time dimension convolution;  $*_{\text{g}}$  denotes the graph convolution; and  $\chi'$  is the input; and  $O$  is the output.

**3.2.3. Output Block.** The output block consists of a temporal convolution layer and a fully connected layer. First, the temporal convolution layer merges the temporal dimensions output by the spatiotemporal convolution block to obtain  $O'$  and then gets the final forecast result through the fully connected layer.

$$\hat{Y} = O'W + b, \quad (9)$$

where  $W$  is a weight vector and  $b$  is a bias, and both  $W$  and  $b$  are learnable parameters.

## 4. Experiment and Evaluation

**4.1. Performance Evaluation Method.** In order to measure the effectiveness of our proposed method and compare with the comparison model conveniently, we considered three different metrics, including mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). MAE, RMSE, and MAPE can be calculated by Equations (10)–(12).

$$\text{MAE} = \frac{1}{n} \sum_{i=0}^n |\hat{y}_i - y_i|, \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=0}^n (y \wedge_i - y_i)^2}, \quad (11)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=0}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right|, \quad (12)$$

where  $y_i$  represents the true value,  $\hat{y}_i$  represents the forecasted value, and  $n$  represents the number of samples.

RMSE is used to measure the deviation between the forecasted value and the true value, it is one of the most common evaluation metrics, and it is numerically equal to the

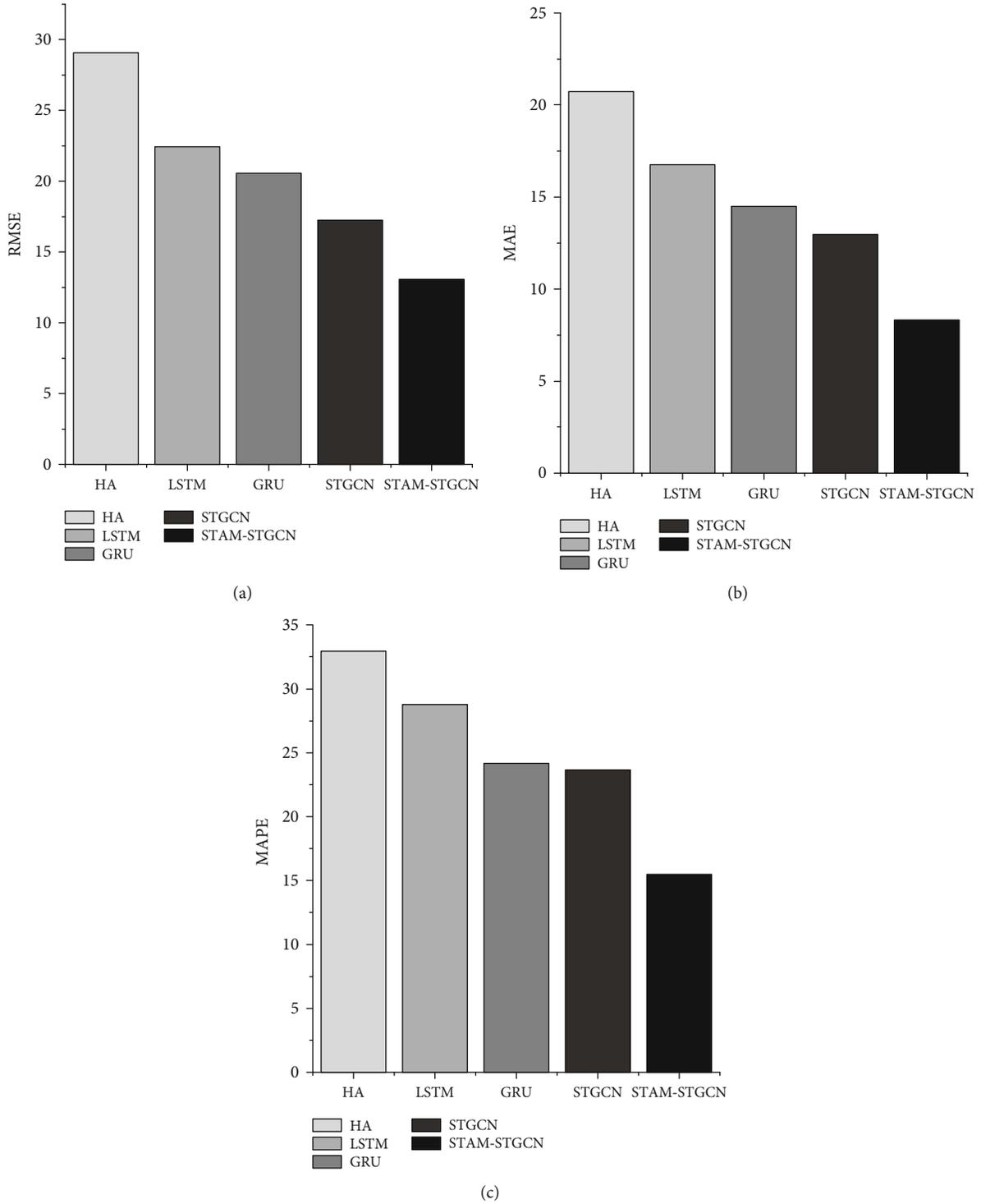


FIGURE 8: Evaluation metrics histogram. (a) RMSE. (b) MAE. (c) MAPE.

square root of the ratio of the square sum of the difference between the real value and the forecasted value and the number of samples  $n$ ; the smaller the value, the better the effect of the model. MAE can better reflect the actual situation of the forecasted value error, the smaller the MAE, the better the effect of the model. If MAPE is 0, it means a perfect model, and if it is greater than 100, it means an inferior model.

**4.2. Training Procedure.** STAM-STGCN uses Adam optimization technology for optimization and trained for 100 epochs with batch size 128; the size of the graph convolution kernel and the temporal convolution kernel are 3, and the learning rate is set to 0.001. ReLU is used as the activation function. We use one day as the historical time window, that is, using 144 data points ( $m = 144$ ) to forecast the PM2.5 concentration for the next 10 minutes ( $n = 1$ ). In the training

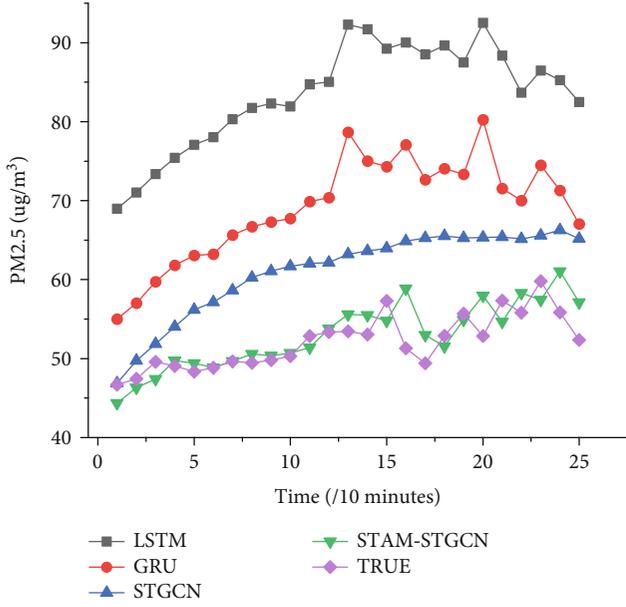


FIGURE 9: Line charts of forecasted and observed value on test data.

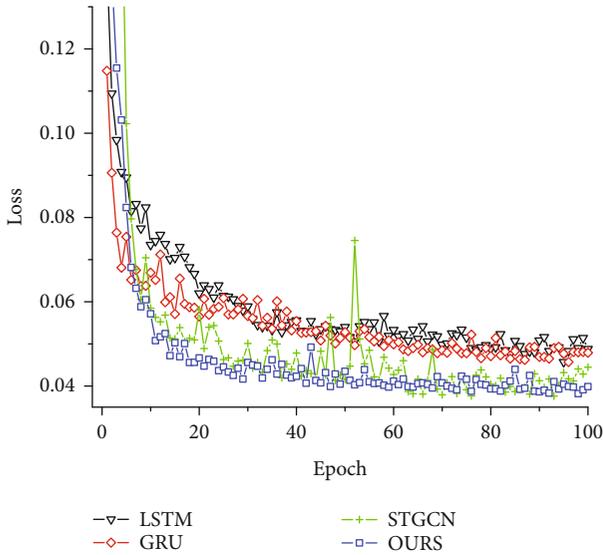


FIGURE 10: The loss of training epochs.

process, the first 80% of the entire dataset is used as the training set, use 10% of the data as a validation set and the last 10% of the data is used as the test set. We implemented the STAM-STGCN in the PyTorch framework [23]. Parameters of the model are also shown in Table 3.

We use mean square error to measure the performance of our model, the process of training the model is the process of continuously optimizing parameters and minimizing the cost function, and the cost function of STAM-STGCN can be expressed as

$$J(\theta) = \frac{1}{2N} \sum_{i=1}^M (y \wedge_i - y_i)^2 = \frac{1}{2N} \sum_{i=1}^M (h_{\theta}(x_i) - y_i)^2, \quad (13)$$

where  $\theta$  is the parameter of STAM-STGCN, which can be continuously updated in training optimization,  $\hat{y}_i$  is the forecasted value of PM2.5 concentration,  $y_i$  is the true value,  $h_{\theta}(x_i)$  is the value forecasted by the parameter  $\theta$  and input  $x$ , and  $M$  is the number of training samples.

**4.3. Results and Analysis.** In order to verify the advantages of the proposed model, we selected four models for time series forecasting for comparison, namely, HA, LSTM, GRU, and STGCN. Each model uses the same dataset as the proposed method; the experimental results are shown in Table 4. In order to show the advantages of the proposed model more intuitively, a histogram of each model evaluation metrics is made, as shown in Figure 8.

It can be seen from the experimental results in Table 4 that compared with HA, LSTM, GRU, and STGCN, the RMSE of our proposed model is reduced by 15.993, 9.352, 7.498, and 4.189, respectively, and the RMSE performance can be improved about 55%, 41.6%, 36.4%, and 24.2%, respectively; the MAE of our proposed model is reduced by 12.43, 8.46, 6.196, and 4.66, respectively, and the performance can be improved about 59.8%, 50.3%, 42.6%, and 35.8%, respectively; the MAPE of our proposed model is reduced by 17.478, 13.29, 8.705, and 8.204, respectively, and the performance can be improved about 53%, 46.1%, 35.9%, and 34.6%, respectively. The results show that this method has better forecast effect than the comparison model. We can also further see that the forecast results of traditional time series analysis methods are not ideal, which proves that the ability of these methods to deal with complex spatiotemporal data is limited. In contrast, methods based on deep learning usually obtain better forecast results than traditional time series analysis methods. In the deep learning method, both the STGCN model and the model we proposed consider spatial-temporal correlations, and the results are better than traditional deep learning models, such as LSTM and GRU; this shows that in the industrial park scenario, considering the spatial-temporal correlations of monitoring stations is useful for forecasting PM2.5. For further analysis of the experimental results, the results of our proposed model are better than the STGCN model, which fully shows that the attention mechanism is effective in capturing the dynamic spatial-temporal correlations of data.

In order to further compare the forecast capabilities of each model, we drew a line graph of the forecasted and true values of each model on the test set, as shown in Figure 9. It can be seen from the figure that our model shows the best fit between the forecasted value and the observed value; among them, the green line in the figure represents the forecast of our model, and the purple line represents the true value.

In order to compare the performance of the model, we draw a graph of the loss during the training process, as shown in Figure 10. It can be seen from the figure that our model can achieve convergence faster and is relatively stable after convergence. However, while the performance of the model is improved, our model has a small increase in complexity compared to the comparison model; this is within our controllable range, which is acceptable relative to the

improvement in performance. In the future, we will work to improve this problem.

## 5. Conclusion and Future Work

In this research, we propose a deep learning model STAM-STGCN to forecast PM<sub>2.5</sub> concentration in industrial parks. We constructed the PM<sub>2.5</sub> concentration data of all monitoring stations in the industrial park into graph time series data, using temporal convolution and graph convolution combined with the spatiotemporal attention mechanism to simultaneously capture the dynamic spatiotemporal characteristics of PM<sub>2.5</sub> concentration in industrial parks. When constructing the adjacency matrix, we also considered the wind field information and the number of pollution sources around the monitoring stations and formulated an information fusion strategy to represent the adjacency matrix. Our model is verified on the real atmospheric dataset of Dongmingnan Industrial Park, experimental results show that compared with the most advanced model in the comparison model, the RMSE of our model is improved by about 24.2%, the MAE performance is improved by 35.8%, and the MAPE performance is improved by 34.6%, the forecast accuracy of the model is better than the comparison model.

In fact, in industrial parks, the PM<sub>2.5</sub> concentration is also affected by many other factors, such as weather conditions and wind speed. In the future, we will consider these influencing factors to further improve the forecasting accuracy. Since STAM-STGCN is a general spatial-temporal forecasting framework for the graph structure data in industrial park scenarios, it has a certain universality; we can also apply it to the forecast of PM<sub>10</sub> and other air pollutants in industrial parks.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant numbers 61701284, U1931207, and 61702306), the Innovative Research Foundation of Qingdao (grant number 19-6-2-1-cg), the Application Research Project for Postdoctoral Researchers of Qingdao, the Sci. & Tech. Development Fund of Shandong Province of China (grant numbers ZR202102230289, ZR202102250695, and ZR2019LZH001), the Humanities and Social Science Research Project of the Ministry of Education of the People's Republic of China (grant number 18YJAZH017), the Taishan Scholars Program of Shandong Province, the Shandong Chongqing Science and technology cooperation project (grant number cstc2020jcsx-lyj-

sAX0008), the Sci. & Tech. Development Fund of Qingdao (grant number 21-1-5-zlyj-1-zc), SDUST Research Fund (grant number 2015TDJH102), and the Science and Technology Support Plan of Youth Innovation Team of Shandong higher School (grant number 2019KJN024).

## References

- [1] H. Yue, C. He, Q. Huang, D. Yin, and B. A. Bryan, "Stronger policy required to substantially reduce deaths from PM<sub>2.5</sub> pollution in China," *Nature Communications*, vol. 11, no. 1, 2020.
- [2] Y. Tian, X. Liu, R. Huo, Z. Shi, and R. M. Harrison, "Organic compound source profiles of PM<sub>2.5</sub> from traffic emissions, coal combustion, industrial processes and dust," *Chemosphere*, vol. 278, article 130429, 2021.
- [3] M. A. Kioumourtzoglou, J. Schwartz, P. James, F. Dominici, and A. Zanobetti, "PM<sub>2.5</sub> and mortality in 207 US cities: modification by temperature and city characteristics," *Epidemiology*, vol. 27, no. 2, pp. 221–227, 2016.
- [4] S. S. Peng, Y. U. Zhong-Hua, and J. Q. Qin, "Experience of domestic urban air pollution control and the enlightenment to Nanjing," *Pollution Control Technology*, 2019, Chinese Society for Environmental Sciences.
- [5] W. Zhou, X. Wu, S. Ding, X. Ji, and W. Pan, "Predictions and mitigation strategies of PM<sub>2.5</sub> concentration in the Yangtze River Delta of China based on a novel nonlinear seasonal grey model," *Environmental Pollution*, vol. 276, article 116614, 2021.
- [6] H. Liu, H. Tian, K. Zhang et al., "Seasonal variation, formation mechanisms and potential sources of PM<sub>2.5</sub> in two typical cities in the Central Plains urban agglomeration, China," *Science of the Total Environment*, vol. 657, pp. 657–670, 2019.
- [7] Y. Li, J. Wang, X. Sun, Z. Li, M. Liu, and G. Gui, "Smoothing-aided support vector machine based nonstationary video traffic prediction towards 5G networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7493–7502, 2020.
- [8] Y. Wang, D. Wang, and Y. Tang, "Clustered hybrid wind power prediction model based on ARMA, PSO-SVM, and clustering methods," *IEEE Access*, vol. 8, pp. 17071–17079, 2020.
- [9] L. Xu, H. Wang, and T. A. Gulliver, "Outage probability performance analysis and prediction for mobile IoV networks based on ICS-BP neural network," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3524–3533, 2021.
- [10] X. Dai, J. Liu, and Y. Li, "A recurrent neural network using historical data to predict time series indoor PM<sub>2.5</sub> concentrations for residential buildings," *Indoor Air*, vol. 31, no. 4, pp. 1228–1237, 2021.
- [11] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, 2019.
- [12] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, and N. Carvalhais, "Deep learning and process understanding for data-driven earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.
- [13] H. Zhou, F. Zhang, Z. Du, and R. Liu, "Forecasting PM<sub>2.5</sub> using hybrid graph convolution-based model considering dynamic wind-field to offer the benefit of spatial interpretability," *Environmental Pollution*, vol. 273, article 116473, 2021.

- [14] T. Li, M. Hua, and X. Wu, "A hybrid CNN-LSTM model for forecasting particulate matter (PM<sub>2.5</sub>)," *IEEE Access*, vol. 8, pp. 26933–26940, 2020.
- [15] J. Chen, Q. Zhu, D. Shi et al., "A multi-step wind speed prediction model for multiple sites leveraging spatio-temporal correlation," *Proceedings of the CSEE*, vol. 39, no. 7, pp. 2093–2106, 2019.
- [16] P. Unjin, K. Chungsong, R. Unsok, S. Kyongjin, and P. Sungnam, "A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction," *Air Quality Atmosphere & Health*, vol. 11, no. 8, pp. 883–895, 2018.
- [17] F. Liu, J. Zheng, L. Zheng, and C. Chen, "Combining attention-based bidirectional gated recurrent neural network and two-dimensional convolutional neural network for document-level sentiment classification," *Neurocomputing*, vol. 371, pp. 39–50, 2020.
- [18] N. Ketkar, *Convolutional Neural Networks*, Springer International Publishing, 2017.
- [19] Y. Qi, Q. Li, H. Karimian, and D. Liu, "A hybrid model for spatiotemporal forecasting of PM<sub>2.5</sub> based on graph convolutional neural network and long short-term memory," *Science of the Total Environment*, vol. 664, pp. 1–10, 2019.
- [20] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pp. 3634–3640, Stockholm, Sweden, 2017.
- [21] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, pp. 922–929, 2019.
- [22] L. Li, J. Gong, J. Zhou, and Q. Sun, "Spatial interpolation of fine particulate matter concentrations using the shortest wind-field path distance," *Plos One*, vol. 9, no. 5, 2014.
- [23] A. Paszke, S. Gross, F. Massa et al., "PyTorch: an imperative style, high-performance deep learning library," *Advances in Neural Information Processing Systems*, vol. 32, pp. 8026–8037, 2019.