



Dynamic graph convolution neural network based on spatial-temporal correlation for air quality prediction

Ao Dun, Yuning Yang, Fei Lei*

Beijing University of Technology, Faculty of Information Technology, Beijing 100124, China



ARTICLE INFO

Keywords:

Air quality prediction
Graph convolutional network
Temporal convolutional network
Spatiotemporal correlation

ABSTRACT

Air pollution is a serious threat to both the ecological environment and the physical health of individuals. Therefore, accurate air quality prediction is urgent and necessary for pollution mitigation and residents' travel. However, few existing models are established based on the dynamic spatiotemporal correlation of air pollutants to predict air quality. In this paper, a novel deep learning model combining the dynamic graph convolutional network and the multi-channel temporal convolutional network (DGC-MTCN) is proposed for air quality prediction. To efficiently represent the time-varying spatial dependencies, a new spatiotemporal dynamic correlation calculation method based on gray relation analysis is proposed to construct dynamic adjacency matrices. Then, the spatiotemporal features are sufficiently extracted by the graph convolutional network and the multi-channel temporal convolutional network. Two real-world air quality datasets collected from Beijing and Fushun are applied to verify the performance of our proposed model. The experimental results show that compared with other baselines, the DGC-MTCN model has excellent prediction accuracy. Especially for the prediction of multi-step and different stations, our model performs better temporal stability and generalization ability.

1. Introduction

With the rapid development of economy and urbanization, air pollution attracts increasing attention, especially in some developing countries (Li et al., 2017). Numerous studies show that air pollution has brought great harm to human health such as increasing the risk of respiratory diseases, cancer, heart diseases and so on (Loaiza-Ceballos et al., 2022; Sokoty et al., 2021; Hou et al., 2021). PM_{2.5} is one of six air pollutants (PM_{2.5}, PM₁₀, SO₂, CO, NO₂ and O₃) which has both acute and chronic effects on human health. Therefore, PM_{2.5} concentration prediction has attracted worldwide attention, which can help people to give early warning and take timely measures. Although there are many methods to predict PM_{2.5} concentrations, it is still a very challenging task due to the high spatiotemporal correlation among stations and multiple influential factors such as humidity and wind field.

Existing air quality prediction methods can be divided into three categories: statistical methods, machine learning methods and deep learning methods. Statistical methods such as autoregressive moving average (ARMA) (Hajmohammadi and Heydecker, 2021), autoregressive integrated moving average (ARIMA) (Alada, 2021) and multiple linear regression (MLR) (Napi et al., 2020) construct mathematical

models to fit historical time series curves. Statistical methods are simple and have low computational costs, while a significant shortcoming is that the forecasting accuracy depends on making strong assumptions about stationary processes (Song et al., 2015). Compared with statistical methods, the machine learning methods do not require stationary assumptions, and they can excavate the non-linearity of data. Common machine learning methods include random forest (RF), support vector machine (SVM) and artificial neural networks (ANNs) (Lepioufle et al., 2021; Sun and Xu, 2021; Sakhrieh et al., 2021). Among the machine learning methods, ANNs such as back propagation (BP) (Xia, 2021), multi-layer perceptron (MLP) (Kolasa-Wicek et al., 2021) and radial basis function (RBF) (Silva et al., 2020) are widely applied to predict air quality. Although these methods can mine the non-linearity of data, they are difficult to satisfy the demand of high prediction accuracy.

In recent years, deep learning methods develop rapidly and are widely applied in computer vision (CV) (Xie et al., 2021), natural language processing (NLP) (Akhter et al., 2021) and traffic flow prediction (Yang, 2021). Deep learning methods have more powerful feature mining capabilities than machine learning methods, so they provide new prospects for air quality prediction task. A recurrent neural network (RNN) with back propagation through time for air quality prediction is

* Corresponding author.

E-mail addresses: aodun@bjut.edu.cn (A. Dun), yangyuning@emails.bjut.edu.cn (Y. Yang), leifei@bjut.edu.cn (F. Lei).

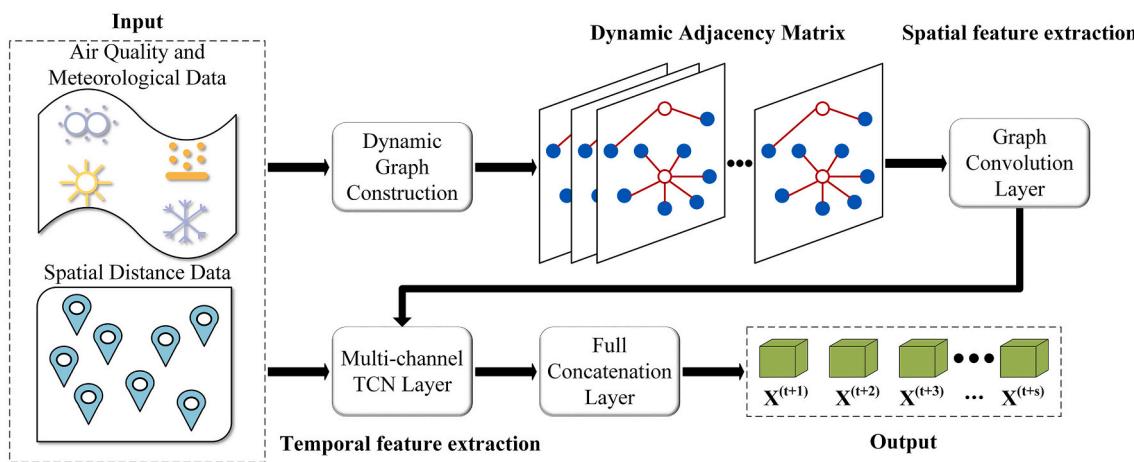


Fig. 1. The framework of the DGC-MTCN model.

proposed to predict the concentrations of air pollutants (Septiawan and Endah, 2018). A multi-channel temporal convolutional network (TCN) is proposed to process time series prediction problems and it is verified that the TCN is better than long short-term memory (LSTM) and gate recurrent unit (GRU) in handling with the problem of long data memory (Liu et al., 2019). An air quality prediction method LSTM is introduced to solve the problem of multiple input time variables according to the characteristics of time series (Du and Lin, 2020). An hourly prediction method of PM_{2.5} concentration in Beijing based on the bidirectional LSTM (Bi-LSTM) neural network is proposed and experimental results show that the performance of proposed model is better than the LSTM and GRU (Zhang et al., 2021b).

The above methods of air quality prediction only consider the data of the single station. However, integrating spatiotemporal correlation between different stations is important for air quality prediction on account of the emission and diffusion of air pollutants (Li et al., 2016). In recent years, the convolutional neural network (CNN) has been usually combined with other neural networks to extract spatial-temporal features from air quality data due to the strong capabilities in mining spatial local dependencies of grid-like data (Pak et al., 2018). A spatiotemporal convolutional long short-term memory neural network extended (C-LSTM) model is proposed to predict air quality concentration which considers the adaptive k-nearest neighboring stations into the model (Cwa et al., 2019). A spatiotemporal convolutional recursive long short-term memory (CR-LSTM) neural network model is proposed to predict the PM_{2.5} concentrations in long-term prediction tasks by combining a convolutional long short-term memory (ConvLSTM) neural network and a recursive strategy to allow for spatiotemporal correlations (Wang et al., 2021). A spatiotemporal causal convolutional neural network (ST-CausalConvNet) is proposed for short-term PM_{2.5} prediction and the model achieves better prediction performance than the other three comparative models (ANNs, GRU and LSTM) (Zhang et al., 2021a).

Compared with CNNs, the graph convolutional network (GCN) has successfully solved spatial dependency problems in a non-Euclidean space and has been widely applied in skeleton-based action recognition (Yan et al., 2018), urban traffic flow forecasting (Liao et al., 2022) and graph classification (Bail et al., 2022). A hybrid model (GC-LSTM) integrating GCN and LSTM is proposed to model and forecast the spatiotemporal variation of PM_{2.5} concentrations (Qi et al., 2019). A synthesis prediction method combining air quality spatial-temporal network and GCN is proposed to simultaneously excavate the time series changing law and the spatial propagation effect (Zhao et al., 2021). A multi-scale spatiotemporal graph convolutional network (MST-GCN) which combines the grouped features and the constructed graphs is proposed, and each spatial-temporal block containing a graph

convolution layer and a temporal convolutional layer is applied to model the spatial correlations and long-term temporal dependencies (Ge et al., 2021).

The above models only create a static graph based on the distance of stations without considering the dynamics of the relationships of nodes. However, constructing dynamic graphs is significant for effectively mining spatial dependencies that change over time (Li et al., 2020). Recently, the dynamic graph convolutional network has attracted wide attention in traffic flow prediction. An attention-based graph neural network predictor (Jiang et al., 2021) and a long-term traffic flow prediction method based on dynamic traffic flow probability graphs (Peng et al., 2021) are proposed to predict traffic flow. While for air quality prediction, few researches establish dynamic graphs based on the spatiotemporal correlation between air pollutants. Therefore, this paper constructs dynamic graphs by a new correlation calculation method and proposes a novel deep learning model combining the dynamic graph convolutional network and the multi-channel temporal convolutional network (DGC-MTCN) for PM_{2.5} concentration prediction. The main contributions are listed as below:

(1) A new calculation method namely dynamic distance gray relation analysis (DD-GRA) is proposed to construct dynamic adjacency matrices. The DD-GRA method can calculate the time-varying spatial correlation between different stations by PM_{2.5} concentrations and the spatial location of stations.

(2) A hybrid deep learning model which combines the dynamic GCN and the multi-channel TCN (DGC-MTCN) is proposed to extract spatiotemporal features. The GCN is applied to extract spatial information from dynamic adjacency matrices and the multi-channel TCN is used to sufficiently extract temporal features which can effectively avoid information leakage.

(3) Two real-world air quality datasets (Beijing PM_{2.5} Dataset and Fushun PM_{2.5} Dataset) are selected to evaluate the performance of DGC-MTCN model. Compared with other baselines, experimental results show that the DGC-MTCN model performs the best effect especially for PM_{2.5} concentration prediction of multi-step and different stations.

The remaining part of this article is organized as follows. Section 2 provides the methodology including the problem definition and the feature extraction modeling. Section 3 describes the experimental settings and data. Section 4 displays the experimental results and the detailed analysis. Section 5 summarizes the research results and puts forward the limitations and prospects.

2. Methodology

In this section, we give the definition of PM_{2.5} concentration prediction problem and describe the modeling process of the proposed

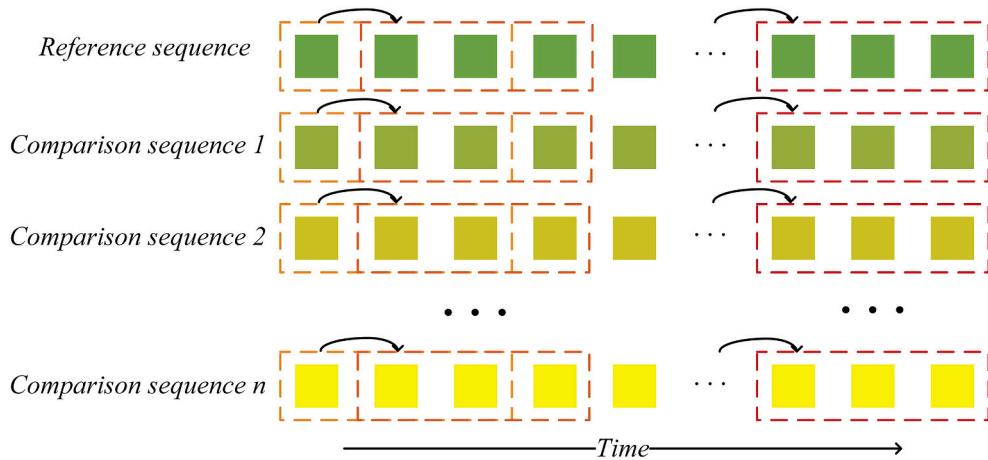


Fig. 2. The diagram of adding sliding windows to calculate dynamic correlation.

DGC-MTCN model in detail. The framework of DGC-MTCN model is presented in Fig. 1. The main structure of model includes input data (air quality data, meteorological data and spatial location information of the stations), dynamic graph construction, spatial feature extraction modeling, temporal feature extraction modeling and the full concatenation layer.

2.1. Preliminaries

2.1.1. Problem definition

The historical T timestamp data of N monitoring stations is used to predict PM_{2.5} concentrations of the next S timestamp in one of the monitoring stations. The function $f(\cdot)$ which combines the dynamic graph convolutional network and the multi-channel temporal convolutional network is applied to extract spatiotemporal features and the final predicted values are obtained by the full concatenation layer.

$$[X_{t+1}, \dots, X_{t+S}] = f([X_{t-T}, \dots, X_t]) \quad (1)$$

2.1.2. Graph definition

Given the graph signal dataset $X \in \mathbb{R}^{N \times T}$, which contains the historical PM_{2.5} concentration data, where N denotes the number of nodes, that is, the number of monitoring stations, and T represents the quantitative characteristics of node attributes in the graph, that is, the size of the input historical time window of the corresponding model. $X^{(t)} \in X$ denotes the graph signal at time t . Given the auxiliary factor dataset $X_A \in \mathbb{R}^{P \times T}$, where P denotes the quantitative characteristics of auxiliary factors. The auxiliary factors include five air pollutants except PM_{2.5} and meteorological data from the target prediction station. $X_A^{(t)} \in X_A$ denotes the auxiliary factor at time t . $X_I = [X; X_A]$ is used as the inputs of the model. $G = (V, A)$ represents an undirected graph, where V denotes the set of nodes, $|V| = N$. $A \in \mathbb{R}^{N \times N}$ is the adjacency matrix of a graph, which represents proximity between the nodes of graph, and $A^t[i, j]$ in A denotes the similarity between v_i and v_j in V at t th time interval.

$$\{[(X^{(t-T):(t)}; G) \oplus X_A^{(t-T):(t)}]; TC\} \rightarrow [X^{(t+1):(t+S)}] \quad (2)$$

where \oplus represents that the new node representations of $X^{(t-T):(t)}$ obtained by the spatial feature extraction G are concatenated with $X_A^{(t-T):(t)}$. Then, they will be passed through temporal feature extraction TC .

2.2. Spatial feature extraction modeling

Constructing a graph structure is usually preferred to represent the spatial relationships of geospatial data, thereby it is also suitable for solving irregularly distributed air quality monitoring stations studied in

this paper. At present, when constructing the adjacency matrix for air quality prediction, most methods use spatial distance to construct a static graph. However, the interaction between different stations not only is related to the distance, but also changes over time. Consequently, we construct dynamic adjacency matrices based on spatiotemporal correlation calculated by PM_{2.5} concentrations and spatial location of stations. The spatiotemporal correlation represents the time-varying spatial dependencies between different stations. To calculate the spatiotemporal correlation, we propose a new calculation method namely dynamic distance gray relation analysis (DD-GRA) on the basis of gray relation analysis (GRA). The GRA is a new analysis method based on grey system theory, which measures the degree of correlation according to the similarity or difference degree of development situation among factors. It has been demonstrated that the GRA is very suitable for dynamic and fast correlation analysis (Bai et al., 2020).

Firstly, a sequence matrix is established based on PM_{2.5} data of L hours from N monitoring stations, as shown in Eq. (3).

$$\begin{pmatrix} X_1(1) & X_2(1) & \dots & X_N(1) \\ X_1(2) & X_2(2) & \dots & X_N(2) \\ \dots & \dots & \dots & \dots \\ X_1(L) & X_2(L) & \dots & X_N(L) \end{pmatrix} \quad (3)$$

Where each column represents the PM_{2.5} concentrations of a station during L hours.

To calculate dynamic correlation, we add the sliding windows. The size of the windows is T and the stride is 1. Therefore, we can obtain $L - T + 1$ sequence matrices by sliding windows. Fig. 2 shows the principle of sliding windows.

For all the new sequential matrices, we perform the same calculation operation. In each sequential matrix, each column is used as a reference sequence and the rest as comparison sequences. The sequence difference between the reference sequence and the comparison sequence is calculated in Eq. (4). And the correlation coefficient is given in Eq. (5).

$$\Delta_{i,j}(t) = X_i(t) - X_j(t) \quad (4)$$

where $\Delta_{i,j}(t)$ denotes the difference between station i and station j at time t , $X_i(t)$ and $X_j(t)$ respectively represent the PM_{2.5} concentration values of station i and station j at time t . The range of i and j is from 1 to N and the range of t is 1 to T .

$$\Phi_{i,j}(t) = \frac{\Delta_{\min} + \rho \Delta_{\max}}{\Delta_{i,j}(t) + \rho \Delta_{\max}} \quad (5)$$

where $\Phi_{i,j}(t)$ denotes the correlation between the reference sequence X_i and the comparison sequence X_j at time t . Δ_{\min} is minimum difference of

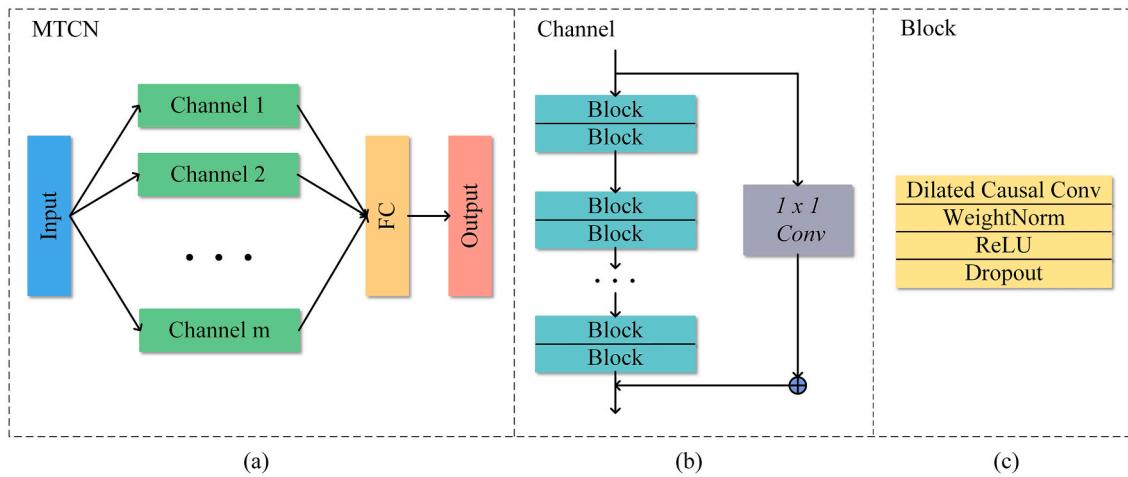


Fig. 3. The structure of multi-channel TCN.

two levels and Δ_{max} is maximum difference of two levels. ρ represents distinguish coefficient, usually takes 0.5.

Then, the average value of the correlation coefficients between the reference sequence X_i and the comparison sequence X_j in each time step is taken as the correlation degree, which is shown as Eq. (6).

$$r_{ij} = \frac{1}{T} \sum_{t=1}^T \phi_{ij}(t) \quad (6)$$

In order to fully consider the influence of spatial distance between stations, we add the inverse distance weight to the calculation formula of correlation coefficients. Now, the distance is calculated according to the longitude and latitude information of each monitoring station. The calculation equation is shown in Eq. (7).

$$S_{ij} = 2\arcsin \sqrt{\sin^2 A + \cos(\text{lat}_i) \times \cos(\text{lat}_j) \times \sin^2 B \times R} \quad (7)$$

where S_{ij} donates the distance between station i and station j . $A = \frac{|\text{lat}_i - \text{lat}_j|}{2}$, $B = \frac{|\text{lon}_i - \text{lon}_j|}{2}$, lat_i , lat_j are the latitude of station i and j , lon_i , lon_j are the longitude of station i and j , R is the radius of the earth, which is 6378 km.

According to the results obtained by Eq. (8), the inverse distance weight of station i and station j is defined by Eq. (8).

$$w_{ij} = 1 - \frac{S_{ij}}{\sum_{i=1, j=1}^{i=N, j=N} S_{ij}/N} \quad (8)$$

The final dynamic adjacency matrix constructed by the DD-GRA method is shown in Eq. (9).

$$A = \begin{pmatrix} 1 & w_{2,1} & \cdots & w_{N,1} \\ w_{1,2} & 1 & \cdots & w_{N,1} \\ \cdots & \cdots & \cdots & \cdots \\ w_{1,N} & w_{2,N} & \cdots & 1 \end{pmatrix} * \begin{pmatrix} 1 & r_{2,1} & \cdots & r_{N,1} \\ r_{1,2} & 1 & \cdots & r_{N,1} \\ \cdots & \cdots & \cdots & \cdots \\ r_{1,N} & r_{2,N} & \cdots & 1 \end{pmatrix} \quad (9)$$

GCNs are widely used in the field of spatial dependency modeling since its powerful spatial information aggregation ability. After constructing graphs, we apply the convolution in the spatial domain to extract spatial features. $\tilde{A} = D - A$, where D is a diagonal degree matrix of the graph, and $D_{ii} = \sum_{j=1}^N A_{ij}^{(t)}$. The graph convolutional layer over a graph signal is defined as:

$$\tilde{X}_G = D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}} X W \quad (10)$$

where \tilde{X}_G is the new node representation, G represents the graph convolution operation and W is the trainable weight matrix.

2.3. Temporal feature extraction modeling

After obtaining the new node representation, we combine it with auxiliary factors of the target prediction station. Then, the multi-channel TCN is proposed to extract temporal features, which is more efficiently and conveniently than RNN. Fig. 3 shows the detailed structure of multi-channel TCN. Fig. 3(a) shows the overall structure of the multi-channel TCN. The composition of each channel is shown in Fig. 3(b) and (c) is the detailed algorithm of each block in the channels. The main compositions of TCN include causal convolution, dilated convolution and residual structure.

2.3.1. Causal convolution

The function of causal convolution is to avoid information leakage. For air quality forecasting, the forecast value is only influenced by the current and previous moments and is not influenced by future information by causal convolution. For example, for a three-layer network, only 3 units X_{t-2} , X_{t-1} , X_t participate in the convolution to obtain the prediction values.

2.3.2. Dilated Convolution

The dilated convolution is originally proposed to solve the problem of image semantic segmentation. Compared with the original standard convolution, under the same convolution kernel size, the dilated convolution can have a larger receptive field. The dilated convolution has one more hyper-parameter called the dilated rate, which is the number of intervals before each point of the convolution kernel. For example, the original 3×3 convolution kernel can have 5×5 (dilated rate = 2) receptive field.

2.3.3. Residual structure

The residual structure is proposed to alleviate the gradient disappearance problem caused by increasing depth in deep neural networks. For the TCN, the receptive field can be climbed by increasing the number of layers, kernel size, and dilated rate. When a prediction task needs a receptive field to be 2^{10} , the TCN may require 10 layers. While, as the depth of the network increases, the degradation problem will occur. Introducing the residual structure can effectively solve the problem of deep network degradation.

Algorithm 1. (DGC-MTCN model)

Data: The $PM_{2.5}$ data of N stations;
The location information of N stations;
The auxiliary factors from the target prediction station.

Result: The prediction results of $PM_{2.5}$ in the next few hours;
The evaluation metrics, including mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R^2).

```

1 for each sample do
2   Constructing dynamic adjacency matrices by Eq. 3 to Eq. 9.
3   The new node representation is obtained by Eq. 10.
4   The auxiliary factors are connected with the new node representation.
5   Applying a MTCN layer to extract temporal features.
6   Using full concatenation layer to obtain the final prediction values.
7 end
8 return prediction values, RMSE, MAE and  $R^2$ 
```

3. Experiments

3.1. Data source

To verify the effects of models, two real-world datasets are selected for the experiments: Beijing PM_{2.5} Dataset and Fushun PM_{2.5} Dataset. Each dataset includes hourly air pollutant data and meteorological data from multiple stations.

Beijing PM_{2.5} Dataset It is a public dataset (<https://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data>) which collects hourly air pollutant concentrations and meteorological data from 12 monitoring stations. The detailed description of the dataset is shown in Table 1. We select a total of 35064 pieces of data from January 1, 2013 to February 28, 2017.

Fushun PM_{2.5} Dataset It is collected from 34 monitoring stations deployed by the partner company in Fushun City, Liaoning Province, China. The data is sampled continuously every hour, from 0 o'clock to 23 o'clock every day, once an hour. The detailed description of the dataset is shown in Table 2. A total of 8784 pieces of data from January 1, 2020 to December 31, 2020 is selected as the dataset.

The prior data cleaning is vital for accurate air quality prediction. However, there are often missing data due to equipment failure, sensor malfunctions and other reasons. If the abnormal data is not processed in advance, the training effect of the models will deteriorate, and the predicted values will have a large deviation from the observed values. In this paper, linear interpolation is selected to fill in the missing values. Besides, the normalization method of unifying variables in a range is necessary for the fast convergence of the models. From Table 1 and 2, it is obviously observed that the units and ranges of variables are different. Therefore, maximum-minimum normalization method is selected to change the range of variables to [0, 1]. Among the processed data, 80% is used as the training set, and 20% is used as the test set.

Table 1
The detailed description of Beijing PM_{2.5} Dataset.

Variables	Range	Mean	Unit
PM _{2.5}	[3.0, 999.0]	52.0	$\mu\text{g}/\text{m}^3$
PM ₁₀	[2.0, 999.0]	104.5	$\mu\text{g}/\text{m}^3$
NO ₂	[0.3, 500.0]	15.8	$\mu\text{g}/\text{m}^3$
SO ₂	[1.0, 290.0]	50.6	$\mu\text{g}/\text{m}^3$
O ₃	[100, 10000]	1230.6	$\mu\text{g}/\text{m}^3$
CO	[0.2, 500]	57.3	$\mu\text{g}/\text{m}^3$
TEMP	[41.4, -19.9]	13.5	°C
DEWP	[982.4, 1042.8]	1010.7	hPa
Wind speed (w-sp)	[0.0, 11.2]	1.7	m/s
Wind direction (w-dr)	-	-	-

3.2. Baselines

In order to verify the superiority of our proposed model, three single models (LSTM, GRU and TCN) and four hybrid models (CNN-LSTM, CNN-GRU, CNN-TCN and GCN-TCN) are selected as baselines. **LSTM** ([Hochreiter and Schmidhuber, 1997](#)): The long short term memory (LSTM) is a variant of Recurrent neural network (RNN), which can efficiently extract temporal features from time series. **GRU** ([Cho et al., 2014](#)): The gate recurrent unit (GRU) is also a variant of RNN, whose structure is simpler than LSTM. **TCN** ([Bai et al., 2018](#)): The temporal convolutional network (TCN) is a kind of network which uses convolution operation to extract temporal features. It has the advantage of efficient parallel data processing. **CNN-LSTM** ([Ding et al., 2021](#)): This is a hybrid model which combines the convolutional neural network (CNN) and LSTM. This model can efficiently extract spatiotemporal characteristics. **CNN-GRU**: Similar to the above model, this hybrid model combines the CNN and GRU. **CNN-TCN** ([Zhang et al., 2021a](#)): Similar to the above model, this hybrid model combines the CNN and TCN. **GCN-TCN**: The spatial distance of the stations is used to construct a static adjacency matrix, and the prediction model combines the GCN and TCN.

3.3. Experimental settings

It is important to choose the appropriate hyper-parameters for the models. In the experiments, the number of channels in the TCN is 3 and each channel has 6 blocks which can ensure sufficient temporal feature extraction without causing information leakage as much as possible. The kernel size in the TCN is set as 3. Stochastic gradient descent (SGD) is selected as the optimizer of parameters. We set the learning rate and momentum as 0.01 and 0.9, respectively. And we also introduce the weight decay and dropout which are respectively set as 5×10^{-4} and 0.5 to avoid over-fitting. Besides, the batch size and epoch are respectively set as 256 and 200 to ensure the fully training of model.

For the experimental setting of baselines, we divide the seven baselines into two groups. For the LSTM, GRU and TCN, we consider only air pollutants and meteorological data from the predicted stations. The hidden size of the LSTM and GRU is set as 10. A single channel TCN is modelled and the number of blocks is 8. For the CNN-LSTM, CNN-GRU and CNN-TCN, Pearson correlation coefficient is used to calculate the correlation among stations. When the correlation coefficient of the two stations is higher than 0.7, we choose this station as the spatial features of the models and the CNN is applied to extract spatial features. To be fair, the kernel size, batch size, epoch, learning rate and the method of optimizer are same as the DGC-MTCN model. Moreover, we set the size of time window as 24, that is, the past 24-hour historical data is used to predict future $t + 1, t + 2, t + 3$ and $t + 4$ hour PM_{2.5} concentrations.

Table 2
The detailed description of Fushun PM_{2.5} Dataset.

Variables	Range	Mean	Unit
PM _{2.5}	[1.0, 500.0]	51.0	µg/m ³
PM ₁₀	[1.0, 569.0]	75.2	µg/m ³
NO ₂	[1.0, 189.0]	66.6	µg/m ³
SO ₂	[1.0, 127.0]	20.6	µg/m ³
O ₃	[3.0, 804.0]	67.1	µg/m ³
CO	[0.0, 7.3]	1.5	mg/m ³
Humidity(hum)	[15.0, 100.0]	58.5	%
Air pressure(pres)	[22.0, 102.0]	98.9	kPa
Wind speed(w-sp)	[0.1, 4.2]	0.48	m/s
Wind direction(w-dr)	[4.5, 315.0]	106.7	°

Table 3
The comparison of different models for single-step PM_{2.5} prediction.

	Models	RMSE	MAE	R ²
Beijing	LSTM	13.670	8.766	0.899
	GRU	14.539	8.887	0.887
	TCN	13.719	7.637	0.897
	CNN-LSTM	12.298	7.237	0.920
	CNN-GRU	13.015	7.669	0.907
	CNN-TCN	13.118	8.481	0.902
	GCN-TCN	11.079	5.915	0.937
	DGC-MTCN	9.779	5.540	0.950
Fushun	LSTM	20.115	11.420	0.776
	GRU	20.101	11.071	0.778
	TCN	22.821	12.631	0.737
	CNN-LSTM	19.114	10.572	0.781
	CNN-GRU	19.110	10.509	0.788
	CNN-TCN	18.039	9.451	0.842
	GCN-TCN	17.863	9.212	0.859
	DGC-MTCN	12.961	8.391	0.912

3.4. Evaluation metrics

We evaluate the experimental results by Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and coefficient of determination (R^2), which are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (11)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N \| y_i - \hat{y}_i \| \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (13)$$

where y_i is the real value of output, \hat{y}_i is the prediction of output, N is the sample size and \bar{y}_i denotes the mean of all real values.

4. Results and discussions

In this section, two target stations (Beijing-Aotizhongxin and Fushun-station 01) are selected to evaluate the single-step forecasting performance (Section 4.1) and multi-step forecasting performance (Section 4.2). Besides, Fushun-station 17 and Fushun-station 22 are used to evaluate the performance of our DGC-MTCN model across different stations (Section 4.3).

4.1. Single-step forecasting results comparison

Table 3 displays the results of two datasets for the single-step PM_{2.5} prediction. The comparison of other baselines reveals that our proposed model achieves the best performance for both datasets. This highlights the importance of considering the dynamic spatiotemporal correlation in forecasting PM_{2.5} concentrations. In order to further show the prediction capabilities of our model, Figs. 4 and 5 show the comparison of the predicted values and the observed values at Beijing-Aotizhongxin and Fushun-station 01 during 200 hours of test set. We can observe that our DGC-MTCN model shows great prediction accuracy in two datasets, and the predicted values can well fit the observed values to a large extent.

Compared with the LSTM, GRU and TCN which only consider temporal features, the models that consider the spatial-temporal correlation including the CNN-LSTM, CNN-GRU, CNN-TCN and DGC-TCN exhibit better results. The poor results obtained from the LSTM, GRU and TCN are due to that they do not consider spatiotemporal correlation when facing the inputs of multiple monitoring stations. For the CNN-TCN model, the RMSE, MAE and R^2 values are 13.118, 8.481 and 0.902 respectively at Beijing-Aotizhongxin; 18.039, 9.451 and 0.842 respectively at Fushun-station 01. It turns out that there is a significant increase in accuracy after accounting for spatial correlation. However, there is still a large deviation between the predicted values and the observed values. This is because the CNN can only be applied for Euclidean data like images while the air quality monitoring station are distributed in a non-Euclidean space.

Compared to the CNN-TCN model, the GCN-TCN model shows not bad prediction ability in PM_{2.5} prediction task. The RMSE and MAE

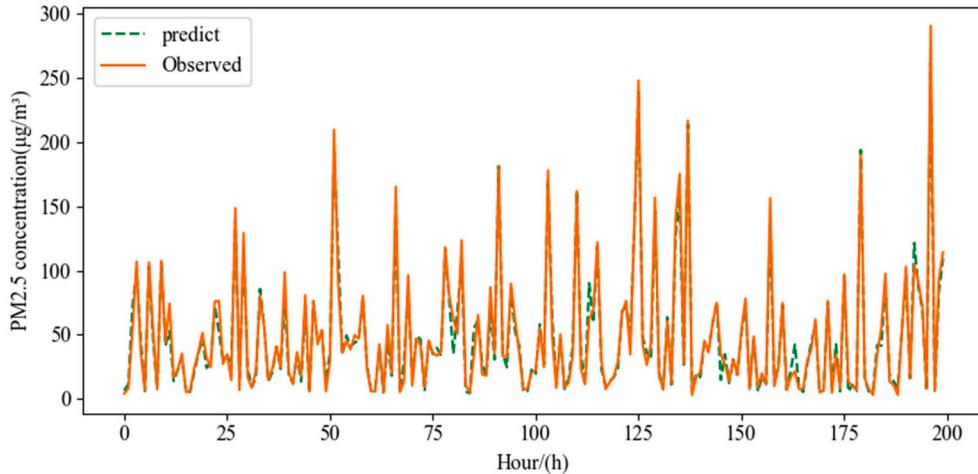


Fig. 4. The comparison of the predicted values and the observed values for single-step PM_{2.5} prediction at Beijing-Aotizhongxin.

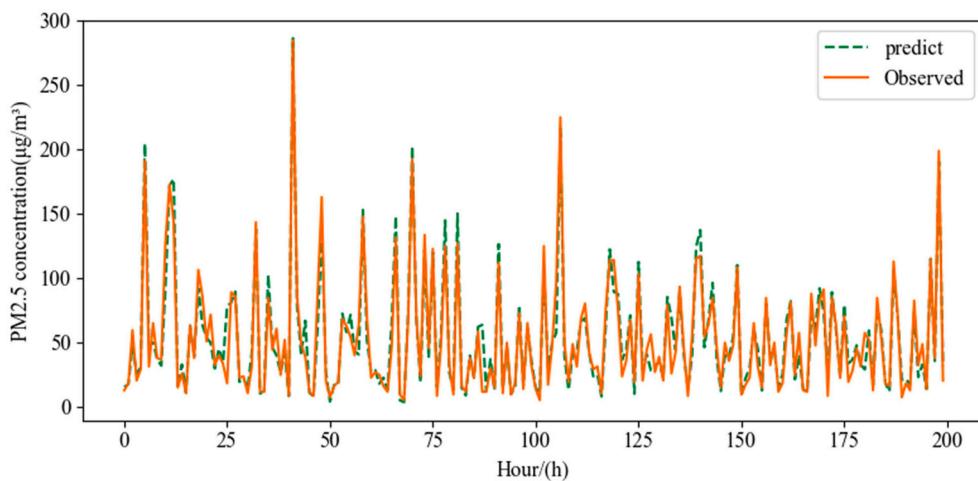


Fig. 5. The comparison of the predicted values and the observed values for single-step PM_{2.5} prediction at Fushun-station 01.

Table 4
The comparison of different models for multi-step PM_{2.5} prediction.

		T+2		T+3		T+4	
		RMSE	MAE	RMSE	MAE	RMSE	MAE
Beijing	LSTM	16.522	9.579	18.773	10.900	19.826	12.358
	GRU	16.761	9.807	20.016	11.733	20.976	12.518
	TCN	15.707	9.171	19.162	10.677	20.210	12.345
	CNN-LSTM	14.367	8.502	16.899	9.495	17.968	10.393
	CNN-GRU	14.856	8.204	17.956	10.613	18.291	10.523
	CNN-TCN	15.261	8.744	17.324	10.493	19.858	11.518
	GCN-TCN	14.072	8.773	16.627	9.449	18.603	11.007
Fushun	DGC-MTCN	13.455	8.098	15.825	9.208	17.692	10.113
	LSTM	25.814	16.892	27.421	18.196	34.102	20.215
	GRU	25.797	15.810	27.279	17.975	33.291	19.466
	TCN	23.053	14.035	27.090	17.900	32.279	18.975
	CNN-LSTM	22.204	13.378	27.002	16.356	30.200	17.856
	CNN-GRU	22.367	13.671	27.030	16.751	30.064	18.387
	CNN-TCN	22.191	13.305	26.941	16.529	31.385	18.778
	GCN-TCN	22.377	13.779	25.398	15.626	31.120	17.117
	DGC-MTCN	21.063	12.152	22.916	14.775	23.635	15.921

values decrease by 15.5% and 30.3% respectively, and the R^2 value increases by 3.9% at Beijing-Aotizhongxin; the RMSE and MAE values decrease by 1.0% and 2.5% respectively, and the R^2 value increases by 3.9% at Fushun-station 01. However, the GCN-TCN model is still hard to achieve a satisfactory prediction performance due to that it ignores the dynamic correlation between different stations. The proposed DGC-MTCN model takes the complex dynamic correlation into consideration, and is significantly superior to other baselines of single-step prediction performance on PM_{2.5} prediction. The RMSE, MAE and R^2 values are 9.779, 5.540 and 0.950 respectively at Beijing-Aotizhongxin; 12.961, 8.391 and 0.912 respectively at Fushun-station 01. The results reveal that our DGC-MTCN model can well represent the dynamic spatial-temporal correlation between stations and achieve high prediction accuracy.

4.2. Multi-step forecasting results comparison

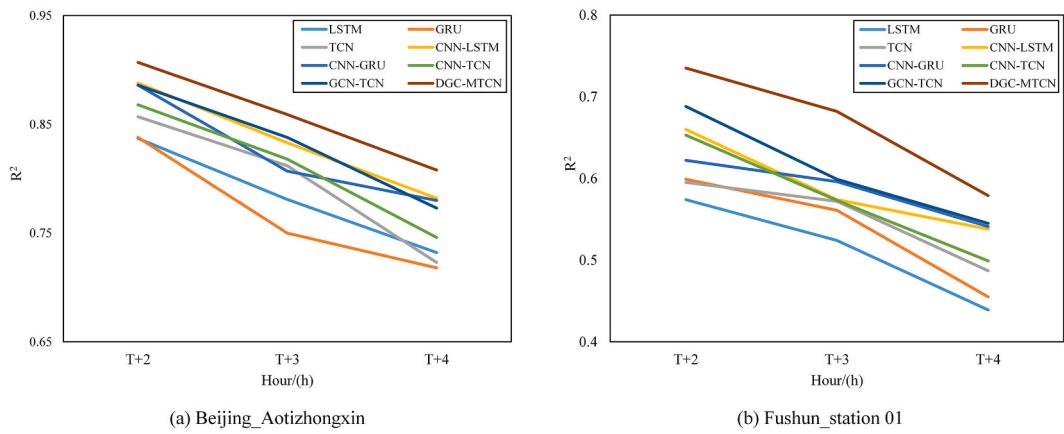
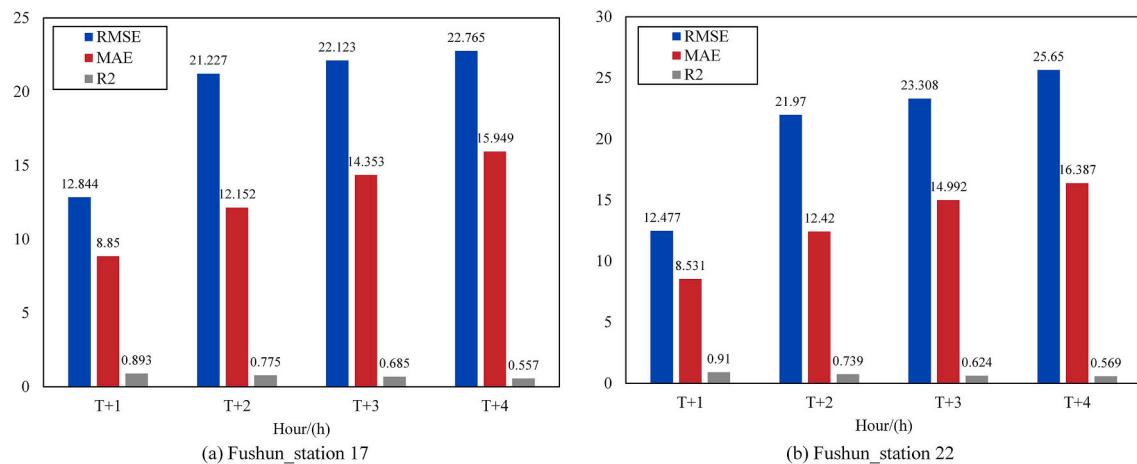
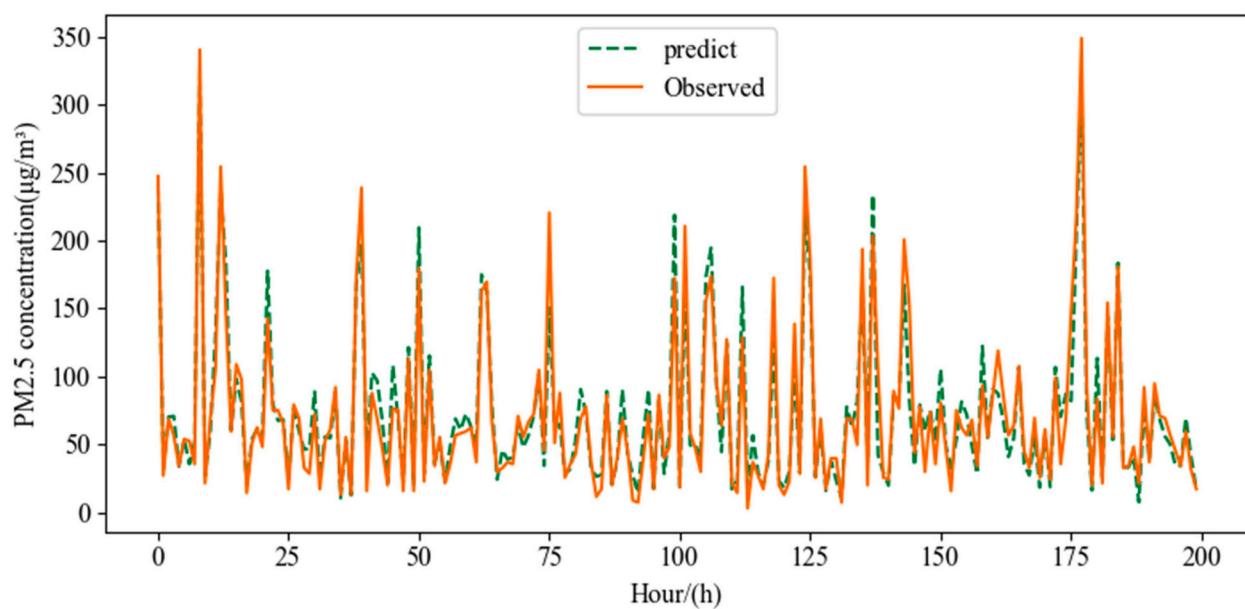
The results of multi-step PM_{2.5} prediction are summarized in Table 4, which lists the RMSE and MAE values for different baselines and the proposed model. From the results, it can be observed that with the increase of prediction steps, the performance of all models gradually deteriorates. We can infer that the increase of prediction steps leads to the gradual weakening of data correlation. While our DGC-MTCN model exhibits the lowest RMSE and MAE values for different multi-step PM_{2.5} prediction task and has a significant performance increasing compared

with other models.

From Table 4, the baselines without considering spatiotemporal correlation show the poor results due to that spatial-temporal dependency is vital for multi-step PM_{2.5} prediction. In most cases, the GCN-TCN model performs better than the models applying CNN to extract spatial features. However, the poor temporal stability of these models indicates that dynamic spatiotemporal correlation between different monitoring stations is essential for multi-step PM_{2.5} prediction task. It is worth noting that the errors of the baseline methods increase slightly when the prediction time step increases for the Beijing-Aotizhongxin, whereas the errors increase significantly for the Fushun-station 01. Compared with the baselines, the errors of our DGC-MTCN model also increase as the time step increase, but it still yields great performance for the multi-step PM_{2.5} forecasting of both datasets. Fig. 6 shows the R^2 values of different models for multi-step PM_{2.5} prediction. It is readily observed that the R^2 values of our DGC-MTCN model can reach above 0.8 for multi-step prediction at Beijing-Aotizhongxin, and above 0.55 at Fushun-station 01. This experiment validates that our model is stable and excellent in multi-step predictions across different datasets.

4.3. Performance comparison across different stations

Fushun-station 17 and Fushun-station 22 in different districts are taken as examples to analyze the performance of our DGC-MTCN model for different stations. Station 17 is located in the motor vehicle

**Fig. 6.** The R^2 values of different models for multi-step PM_{2.5} prediction.**Fig. 7.** The prediction results of Fushun-station 17 and Fushun-station 22 by DGC-MTCN model.**Fig. 8.** The comparison of the predicted values and the observed values for single-step PM_{2.5} prediction at Fushun-station 17.

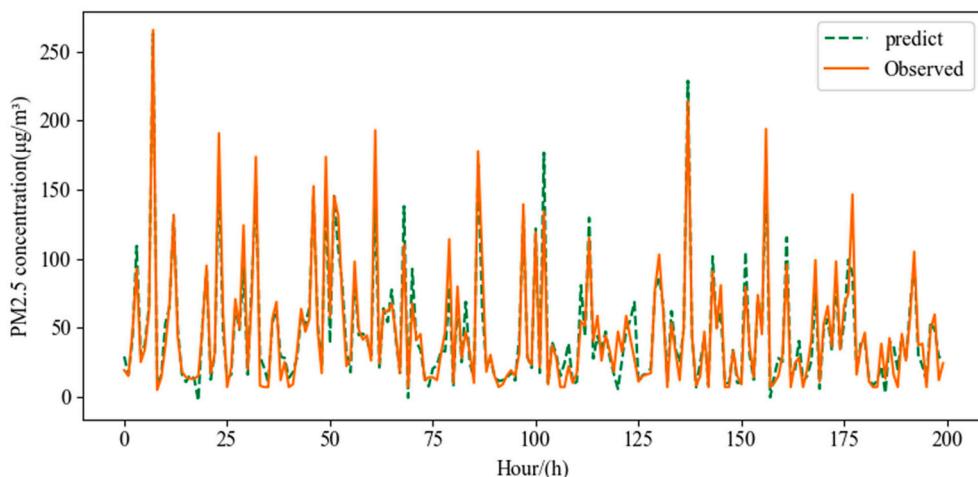


Fig. 9. The comparison of the predicted values and the observed values for single-step PM_{2.5} prediction at Fushun-station 22.

inspection center of Dongzhou district and station 12 is located in a park of Shuncheng district, Fushun City. The changes of air pollutants at different stations are different. The concentrations of PM_{2.5} at station 17 are relatively high due to the large amount of automobile exhaust emissions and fluctuate greatly because of the large flow of vehicles. While the concentrations of PM_{2.5} at station 22 are relatively low and fluctuate slightly due to few air pollution incidents of this area.

For single-step PM_{2.5} prediction, Fig. 7 shows that our DGC-MTCN model achieves good prediction results. The RMSE, MAE and R^2 values of Fushun-station 17 are 12.844, 8.85 and 0.893, respectively. The RMSE, MAE and R^2 values of Fushun-station 22 are 12.477, 8.531 and 0.91, respectively. To show the prediction effect of the model, the comparison of the predicted values and the observed values is given in Figs. 8 and 9. It is worthy to notice that our model can well fit the observed values when the concentrations of PM_{2.5} are very high. Besides, it can be observed that the prediction effect of station 22 is slightly better than station 17 due to that the model has better training effect on stable data.

For multi-step PM_{2.5} prediction, we can observe that the prediction performance of the model gradually deteriorates as the prediction step increases. However, the R^2 values of our model are still all above 0.55 at two stations which show that our model has a great fitting effect for multi-step PM_{2.5} prediction task. Overall, this experiment indicates that our DGC-MTCN model has good prediction performance in both single-step and multi-step PM_{2.5} prediction task, regardless of the stations with large or slight fluctuations in PM_{2.5} concentrations. Moreover, it turns out that our DGC-MTCN model has remarkably better temporal stability and generalization ability.

5. Conclusions

In this paper, we propose a novel deep learning model DGC-MTCN for PM_{2.5} concentration hourly prediction. To begin with, dynamic adjacency matrices are constructed to represent the spatial correlation over time between multiple stations. A new spatiotemporal dynamic correlation calculation method DD-GRA is proposed to calculate the spatiotemporal correlation combining PM_{2.5} concentrations and spatial distance between stations. In order to improve prediction accuracy, auxiliary factors including five air pollutants and meteorological data from the target prediction station are combined with the new node representation obtained through the GCN layer. Moreover, the multi-channel TCN layer is proposed to sufficiently extract temporal features of inputs which is beneficial to avoid information leakage.

Two real-world air quality datasets collected from Beijing and Fushun are selected to evaluate the performance of the proposed DGC-

MTCN model. As a result of evaluation, our DGC-MTCN model shows great prediction performance in single-step prediction and multi-step prediction compared with other baselines. The R^2 values of Beijing dataset and Fushun dataset are 0.95 and 0.912 respectively for single-step PM_{2.5} prediction which prove that the model has a good fitting effect. Moreover, the R^2 values of two datasets are above 0.8 and 0.55 for multi-step prediction which show that our model has great temporal stability in multi-step prediction. Besides, the DGC-MTCN model has good prediction performance in two stations with different pollution conditions and it is proved that the method has good stability and generalization ability.

Although the proposed model can improve the prediction performance compared with other baselines, this model still has some limitations. This method does not fully consider the factors affecting the diffusion of PM_{2.5} such as rainfall, vehicle exhaust emissions and human activities. For the future work, we will explore more methods to construct dynamic adjacency matrices and introduce the attention mechanism to further improve the prediction accuracy.

Authors contributions

All authors contributed to the study conception and design. Method investigation was performed by Ao Dun and Yuning Yang. Software implementation was performed by Yuning Yang and Fei Lei. Data collection and analysis were performed by Fei Lei and Ao Dun. The first draft of the manuscript was written by Ao Dun, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Availability of data and material

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Funding

This work is supported by the National Natural Science Foundation of China under Grant 62173008, 61873007 and 61603009.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- Akhter, M.P., Zheng, B.J.C., Naqvi, I.R., Abdelmajeed, M., Zia, T., 2021. Abusive language detection from social media comments using conventional machine learning and deep learning approaches. *Multimed. Sys.* 1–16.
- Alada, E., 2021. Forecasting of particulate matter with a hybrid arima model based on wavelet transformation and seasonal adjustment. *Urban Clim.* 39 (1), 100930.
- Bai, S., Zico Kolter, J., Koltun, V., 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling.
- Bai, Y.T., Jin, X.B., Wang, X.Y., Wang, X.K., Xu, J.P., 2020. Dynamic correlation analysis method of air pollutants in spatio-temporal analysis. *Int. J. Environ. Res. Pub. Health* 17 (1), 360.
- Bail, L., Cui, L., Jiao, Y., Rossi, L., Hancock, E.R., 2022. Learning backtracking aligned-spatial graph convolutional networks for graph classification. *IEEE Trans. Patt. Anal. Mach. Intel.* 44 (2), 783–798.
- Cho, K., Van Merriënboer, B., Gulcehre, C., BaCHdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *Comput. Sci.* <https://doi.org/10.3115/v1/D14-1179>.
- Cwa, B., Sl, A., Xy, A., Ling, P.A., Xiang, L., Hab Yuan, A., Tc, A., 2019. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *Sci. Total Environ.* 654, 1091–1099.
- Ding, C., Wang, G., Zhang, B.X.C., Liu, Q., Liu, B.X.C., 2021. A hybrid CNN-lstm model for predicting PM_{2.5} in beijing based on spatiotemporal correlation. *Environ. Ecol. Stat.* 28 (3), 503–522.
- Du, Z., Lin, X., 2020. Air quality prediction based on neural network model of long short-term memory. *IOP Conf. Series Earth Environ. Sci.* 508, 012013.
- Ge, L., Wu, K., Zeng, Y., Chang, F., Wang, Y., Li, S., 2021. Multi-scale spatiotemporal graph convolution network for air quality prediction. *Appl. Intel.* 51 (6), 3491–3505.
- Hajmohammadi, H., Heydecker, B., 2021. Multivariate time series modelling for urban air quality. *Urban Clim.* 37, 100834.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780.
- Hou, J., Gu, J., Liu, X., Tu, R., Wang, C., 2021. Long-term exposure to air pollutants enhanced associations of obesity with blood pressure and hypertension. *Clin. Nutr.* 40 (4), 1442–1450.
- Jiang, S., Zhu, M., Li, J., 2021. Traffic Flow Forecasting Using a Spatial-Temporal Attention Graph Convolutional Network Predictor.
- Kolasa-Wicek, A., Suszanicz, D., Pilarska, A.A., Pilarski, K., 2021. Modelling the interaction between air pollutant emissions and their key sources in Poland. *Energies* 14 (21), 6891.
- Lepioulle, J.M., Marsteen, L., Johnsrud, M., 2021. Error prediction of air quality at monitoring stations using random forest in a total error framework. *Sensors* 21 (6), 2160.
- Li, J., Liu, Y., Zou, L., 2020. DynGCN: A Dynamic Graph Convolutional Network Based on Spatial-Temporal Modeling, 12342. Lecture Notes in Computer Science, Springer Cham, pp. 83–95.
- Li, X., Peng, L., Hu, Y., Shao, J., Chi, T., 2016. Deep learning architecture for air quality predictions. *Environ. Sci. Pollut. Res.* 23 (22), 22408–22417.
- Li, X., Zheng, W., Yin, L., Yin, Z., Tian, X., 2017. Influence of social-economic activities on air pollutants in Beijing, China. *Open Geosci.* 9 (1), 314–321.
- Liao, L., Hu, Z., Zheng, Y., Bi, S., Zhang, M., 2022. An improved dynamic chebyshev graph convolution network for traffic flow prediction with spatial-temporal attention. *Appl. Intel.* 1–13.
- Liu, Y., Dong, H., Wang, X., Han, S., 2019. Time series prediction based on temporal convolutional network. In: 2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS), pp. 300–305.
- Loaiza-Ceballos, M.C., Marin-Palma, B.D.C., Zapata, W., Hernandez, B.J.C.C., 2022. Viral respiratory infections and air pollutants. *Air Quality, Atmos. Health* 15 (1), 105–144.
- Napi, N.N.L.M., Mohamed, M.S.N., Abdullah, S., Mansor, A.A., Ahmed, A.N., Ismail, M., 2020. Multiple linear regression (MLR) and principal component regression (PCR) for ozone (O₃) concentrations prediction. *IOP Conf. Series: Earth Environ. Sci.* 616 (1), 012004.
- Peng, H., Du, B., Liu, M., Liu, M., He, L., 2021. Dynamic graph convolutional network for long-term traffic flow prediction with reinforcement learning. *Inf. Sci.* 579, 401–416.
- Qi, Y., Li, Q., Karimian, H., Liu, D., 2019. A hybrid model for spatiotemporal forecasting of PM_{2.5} based on graph convolutional neural network and long short-term memory. *Sci. Total Environ.* 664, 1–10.
- Sakhrieh, A., Hamdan, M.A., Ata, M.F.B., 2021. Air quality assessment and forecasting using neural network model. *J. Ecol. Eng.* 6, 1–11.
- Septiawan, W.M., Endah, S.N., 2018. Suitable recurrent neural network for air quality prediction with backpropagation through time. In: 2nd International Conference on Informatics and Computational Sciences (ICICoS).
- Silva, S., Clemente, F.M., Castelli, M., Popovi, A., Vanneschi, L., 2020. A machine learning approach to predict air quality in California. *Complexity 2020*, 8049504.
- Sokoto, L., Rimaz, S., Hassanlouei, B., Kermani, M., Janani, L., 2021. Short-term effects of air pollutants on hospitalization rate in patients with cardiovascular disease: a case-crossover study. *Environ. Sci. Pollut. Res.* 10, 1–8.
- Song, Y., Qin, S., Qu, J., Liu, F., 2015. The forecasting research of early warning systems for atmospheric pollutants: a case in yangtze river delta region. *Atmos. Environ.* 118, 58–69.
- Sun, W., Xu, Z., 2021. A novel hourly PM_{2.5} concentration prediction model based on feature selection, training set screening, and mode decomposition-reorganization. *Sustain. Cities Soc.* 75, 103348.
- Pak, U., Kim, C., Ryu, U., Sok, K., Pak, S., 2018. A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction. *Air Qual. Atmos. Health* 11 (8), 833–895.
- Wang, W., Mao, W., Tong, X., Xu, G., 2021. A novel recursive model based on a convolutional long short-term memory neural network for air pollution prediction. *Remote Sensing* 13 (7), 1284.
- Xia, X., 2021. Study on the application of bp neural network in air quality prediction based on adaptive chaos fruit fly optimization algorithm. *MATEC Web Conf.* 336 (1), 07002.
- Xie, J., Zheng, Y., Du, R., Xiong, W., Guo, J., 2021. Deep learning-based computer vision for surveillance in its: evaluation of state-of-the-art methods. *IEEE Trans. Vehicular Technol.* 70 (4), 3027–3042.
- Yan, S., Xiong, Y., Lin, D., 2018. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition.
- Yang, C., 2021. Spatial-temporal 3d residual correlation network for urban traffic status prediction. *Symmetry* 14 (1), 33.
- Zhang, L., Na, J., Zhu, J., Shi, Z., Yang, L., 2021a. Spatiotemporal causal convolutional network for forecasting hourly PM_{2.5} concentrations in beijing. *Comput. Geosci.* 155 (11), 104869.
- Zhang, M., Wu, D., Xue, R., 2021b. Hourly prediction of PM_{2.5} concentration in beijing based on bi-lstm neural network. *Mutimed. Tools Appl.* 80 (16), 24455–24468.
- Zhao, G., He, H., Huang, Y., Ren, J., 2021. Near-surface PM_{2.5} prediction combining the complex network characterization and graph convolution neural network. *Neural Comput. Appl.* 33 (24), 17081–17101.