International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Predicting short-term $PM_{2.5}$ concentrations at fine temporal resolutions using a multi-branch temporal graph convolutional neural network

Qingfeng Guan, Jingyi Wang, Shuliang Ren, Huan Gao, Zhewei Liang, Junyi Wang & Yao Yao

To cite this article: Qingfeng Guan, Jingyi Wang, Shuliang Ren, Huan Gao, Zhewei Liang, Junyi Wang & Yao Yao (02 Feb 2024): Predicting short-term $PM_{2.5}$ concentrations at fine temporal resolutions using a multi-branch temporal graph convolutional neural network, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2024.2310737

To link to this article: https://doi.org/10.1080/13658816.2024.2310737

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2024.2310737

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Predicting short-term $PM_{2.5}$ concentrations at fine temporal resolutions using a multi-branch temporal graph convolutional neural network

Qingfeng Guan[a,b], Jingyi Wang[c], Shuliang Ren[d], Huan Gao[a,b], Zhewei Liang[a,b], Junyi Wang[e] and Yao Yao[a,b,f]

[a]School of Geography and Information Engineering, China University of Geoscience, Wuhan, PR China; [b]National Engineering Research Center of GIS, China University of Geoscience, Wuhan, PR China; [c]ChangJiang Waterway Bureau Survey Center, Wuhan, PR China; [d]School of Earth and Space Sciences, Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing, PR China; [e]Tencent Technology (Shenzhen) Co., Ltd., Shenzheng, PR China; [f]Center for Spatial Information Science, The University of Tokyo, Chiba, Japan

ABSTRACT

Predicting $PM_{2.5}$ concentrations at an hourly temporal resolution in urban areas can provide key information for public health protection. The spatiotemporal dependency among monitoring stations and the spatiotemporal correlations between $PM_{2.5}$ and relevant factors (e.g. meteorology and emissions) are both essential for such predictions. This study proposes a multi-branch temporal graph convolutional neural network (MB-TGCN) for short-term predictions of $PM_{2.5}$ concentrations at city monitoring stations. Composed of a set of graph convolutional networks (GCNs) for spatial dependency modeling, a set of gated recurrent units (GRUs) for temporal dependency modeling, and a multi-branch structure for integrating $PM_{2.5}$ and relevant factors, MB-TGCN aims to accurately predict $PM_{2.5}$ concentrations by capturing both spatial and temporal relationships through a graph modeling approach. Experiments with an air quality dataset from 35 stations in Beijing showed that MB-TGCN achieved higher accuracy than several deep learning models for various prediction durations ranging from 1 to 12 h. The method described in this study can help enhance the prediction capability of $PM_{2.5}$ and provide decision support for environment-aware activity planning.

ARTICLE HISTORY
Received 10 May 2022
Accepted 22 January 2024

KEYWORDS
$PM_{2.5}$ prediction; spatiotemporal dependency; graph convolution network; meteorological factors; air pollution

1. Introduction

With rapid urbanization and industrialization, air pollution has become one of the most serious environmental problems (Kurt and Oktay 2010, Aliyu and Botai 2018). Atmospheric pollutants include gaseous pollutants (e.g. CO, SO_2 and NO_x) and particulate matter (PM, e.g. $PM_{2.5}$ and PM_{10}) (Kampa and Castanas 2008, Liu *et al.* 2018). Among air pollutants, fine PM, i.e. particles that are 2.5 microns or less in diameter

CONTACT Yao Yao yaoy@cug.edu.cn; yaoy@csis.u-tokyo.ac.jp

($PM_{2.5}$), can cause serious damage to human health. Long-term exposure to $PM_{2.5}$ pollution can increase the risk of cardiovascular and respiratory diseases and even shorten life expectancy (Brook *et al.* 2010, Chen *et al.* 2018a, 2018b). Hill *et al.* (2023) found a causal relationship between PM air pollutants and lung cancer. It is therefore important to accurately predict $PM_{2.5}$ concentrations.

$PM_{2.5}$ concentration predictions usually include $PM_{2.5}$ concentration estimations at unsampled locations and future $PM_{2.5}$ concentration predictions. The former focuses on estimating the $PM_{2.5}$ concentrations of the current moment at locations where measurements are not available (Hsieh *et al.* 2015), whereas the latter predicts $PM_{2.5}$ concentrations of future moments at locations where historical measurements are available (Shi *et al.* 2022). Both can provide key information for environmental protection, public health care and environment-aware decision-making.

The existing methods for predicting $PM_{2.5}$ can be divided into mechanism-based and data-driven methods. Typical mechanism-based methods include the Community Multiscale Air Quality (CMAQ) model (Tesche *et al.* 2006), the Urban Airshed Model (UAM) (Scheffe and Morris 1993) and the WRF-Chem model (Saide *et al.* 2011). The predictions generated by mechanism-based methods can be corrected objectively, thereby significantly improving the forecasting accuracy (Jia *et al.* 2019). However, such methods often include several parameters for which calibration requires extensive expertise or experience, as well as a large variety of input datasets and extensive computing resources that may not be easily and immediately available (Suleiman *et al.* 2019).

Data-driven methods include statistical models and machine learning models. Statistical models are widely used to capture the relationships between $PM_{2.5}$ concentrations and relevant factors and are easy to implement with acceptable accuracy (Yang *et al.* 2018a, 2018b). Many statistical models, such as the multiple linear regression (MLR) model (Huang and Chun 2017), land use regression (LUR) model (Johnson *et al.* 2010), geographically weighted regression (GWR) (Jiang *et al.* 2017), geographically and temporally weighted regression (GTWR) (Mirzaei *et al.* 2019) and multiscale GTWR (MGTWR) (Liu *et al.* 2021) have been adopted to predict air pollutant concentrations.

Nevertheless, statistical methods have difficulty with large amounts of multidimensional nonlinear data, and analyzing the correlations between air pollutant concentrations and other predictor variables is extremely challenging (Zhang *et al.* 2020a, 2020b, 2020c). Advancements in computational technologies and the availability of larger datasets have led to machine learning techniques being increasingly adopted for air quality prediction (Krishan *et al.* 2019). Typical examples include the support vector machine (SVM) (García Nieto *et al.* 2013) and random forest (Li *et al.* 2021) models. Air quality monitoring data have notable spatiotemporal and multidimensional characteristics. In other words, the $PM_{2.5}$ concentration at a specific location and a specific moment is greatly influenced by the $PM_{2.5}$ concentrations at nearby locations and previous moments. However, conventional machine learning techniques largely ignore these spatiotemporal processes.

For the complicated temporal dependency of monitoring stations, a number of deep learning models have been adopted. Bui *et al.* (2018) predicted air pollutant

concentrations using a recurrent neural network (RNN) and a long short-term memory (LSTM) network. Athira *et al.* (2018) used the gated recurrent unit (GRU) in predicting PM_{10} concentrations. These models can capture the temporal patterns of air pollutant concentrations. Meanwhile, a number of researchers have attempted to explore the spatial characteristics of air pollutants. For example, Huang and Kuo (2018) used a convolutional neural network (CNN) to extract the spatial patterns of $PM_{2.5}$ concentrations. They found that the prediction performance was better than that of traditional machine learning models. Zhang *et al.* (2020a, 2020b, 2020c) used an auto-encoder (AE) to capture the spatial features from air quality data, and greatly improved the training efficiency. These methods can detect hidden spatial features through deep neural networks and are most suited for identifying spatial relationships in Euclidean spaces (e.g. images and regular grids). However, monitoring stations are often located irregularly across an area. The existing methods for extracting spatial features lead to the destruction of the original spatial information, and it can be difficult to characterize their spatial dependency in the model.

Graphs are an effective approach to represent irregularly structured data in non-Euclidean spaces (Scarselli *et al.* 2009). With graph convolutional networks (GCNs) (Bruna *et al.* 2013, Kipf and Welling 2016, Defferrard *et al.* 2016, Veličković *et al.* 2017), the convolution operator can be extended to general graph structure data. In different scenarios, such as traffic forecasting (Zhang *et al.* 2020a, 2020b, 2020c), community detection (Luo and Du 2020) and resource allocation (Zhao *et al.* 2020a, 2020b), the spatial features of non-Euclidean distributions can be effectively captured using GCNs. Therefore, air pollution data from air quality monitoring stations can be reconstructed into graphs, in which the spatial relationships among stations can be easily hard-coded. For example, Qi *et al.* (2019) used a GCN and an LSTM network to capture the monitoring stations' spatial and temporal dependencies, respectively. However, $PM_{2.5}$ concentrations tend to change in space and time simultaneously, and capturing the spatial and temporal characteristics separately may lead to a lower prediction accuracy.

In addition, some relevant factors, such as meteorological variables (MVs), play an important role in the generation, decomposition and dispersion of $PM_{2.5}$ (Yang *et al.* 2015, Mao *et al.* 2016, Gautam *et al.* 2019). Several studies have used MVs to predict $PM_{2.5}$ concentrations (Krishan *et al.* 2019, Liu *et al.* 2019, Zhang *et al.* 2020a, 2020b, 2020c), where they focused on the relationships between MVs and physicochemical processes of air pollution. However, while MVs can also interact in temporal and spatial dimensions (Chen *et al.* 2020), few studies have focused on the spatiotemporal variations of MVs. Furthermore, the emissions of air pollutants are the key factors influencing the spatiotemporal variations of air pollutants (Yao *et al.* 2019). Dong *et al.* (2019) quantified and discussed the influence of CO_2 emissions reduction on $PM_{2.5}$. Duan *et al.* (2021) evaluated the effects of meteorological factors and anthropogenic emissions on $PM_{2.5}$ and O_3 concentrations.

The methods above can be used for both coarse-resolution long-term (e.g. daily, weekly, monthly and annually) and fine-resolution short-term (e.g. hourly) predictions of $PM_{2.5}$ concentrations. Long-term predictions are beneficial for environmental planning to ensure effective control of air pollution in advance (Sharma *et al.* 2018).

Short-term predictions can provide accurate air quality information for the immediate future. The government and environmental protection agencies can use such short-term predictions to implement fast PM reduction measures, such as vacuum sweeping, water flushing and wet suppression (Tucker 2000), at those locations where high concentrations of $PM_{2.5}$ are predicted to occur. Citizens can also use such short-term predictions for trip/activity planning to avoid exposure to air pollution (Liu and Sun 2019). This study was aimed at devising an approach for the short-term (1–12 h) prediction of $PM_{2.5}$ concentrations.

Although existing forecasting models demonstrated prospective performance, two main limitations in extracting the spatiotemporal dependencies exist: (1) previous studies have extracted spatial and temporal patterns separately, which might not completely represent and capture the inter-correlated spatiotemporal variation patterns of $PM_{2.5}$ concentrations; and (2) the spatiotemporal characteristics of relevant factors (e.g. meteorology and emissions) were often less sufficiently considered for predicting $PM_{2.5}$.

To solve these problems, this study proposes a multi-branch temporal graph convolutional neural network (MB-TGCN). This spatiotemporal fusion model aims to predict short-term $PM_{2.5}$ concentrations of monitoring stations over the next 1–12 h. The TGCN is constructed by combining the GCN and GRU models to obtain the dependencies among stations in both space and time. Thus, the spatiotemporal dependencies among $PM_{2.5}$ or relevant factors (e.g. MVs and emission indicators [EIs]) at multiple monitoring stations are extracted using a set of TGCNs in parallel. Finally, through a fully connected neural network (FCN), the prediction results of multiple TGCNs are merged to integrate the spatiotemporal characteristics and, hence, continuously predict future $PM_{2.5}$ concentrations. The performance of the proposed method is validated by using an air quality dataset of Beijing city to predict $PM_{2.5}$ concentrations over different temporal durations.

2. Method

The architecture of the MB-TGCN model proposed in this study is illustrated in Figure 1, and includes the following steps: (1) the datasets of $PM_{2.5}$ concentrations and relevant factors (e.g. MVs and EIs) at monitoring stations are preprocessed to create the feature matrices with complete time sequences (PM_t, $M1_t$, … …, Mm_t); (2) based on the feature matrices and the adjacency matrix (A, representing the graph of monitoring stations), the multivariable time series prediction values (PM'_t, $M1'_t$, … …, Mm'_t) are obtained through a multi-branch structure composed of a set of TGCNs for the spatiotemporal dependency modeling of $PM_{2.5}$ and other factors; and (3) the multidimensional predictions are fused through an FCN to predict the $PM_{2.5}$ concentrations (PM_{t+1}, … …, $PM_{t+T'-1}$, $PM_{t+T'}$) for the future time steps at monitoring stations.

It is worth noting that MB-TGCN is a flexible structure, which can take into account any number of relevant factors (i.e. $m \geq 0$ in Figure 1). Technically, any factor can be included in MB-TGCN to provide auxiliary information for the prediction of $PM_{2.5}$ concentration, as long as it has been theoretically and/or empirically proven to affect the spatiotemporal dynamics of $PM_{2.5}$ concentration. Also, the availability of data greatly

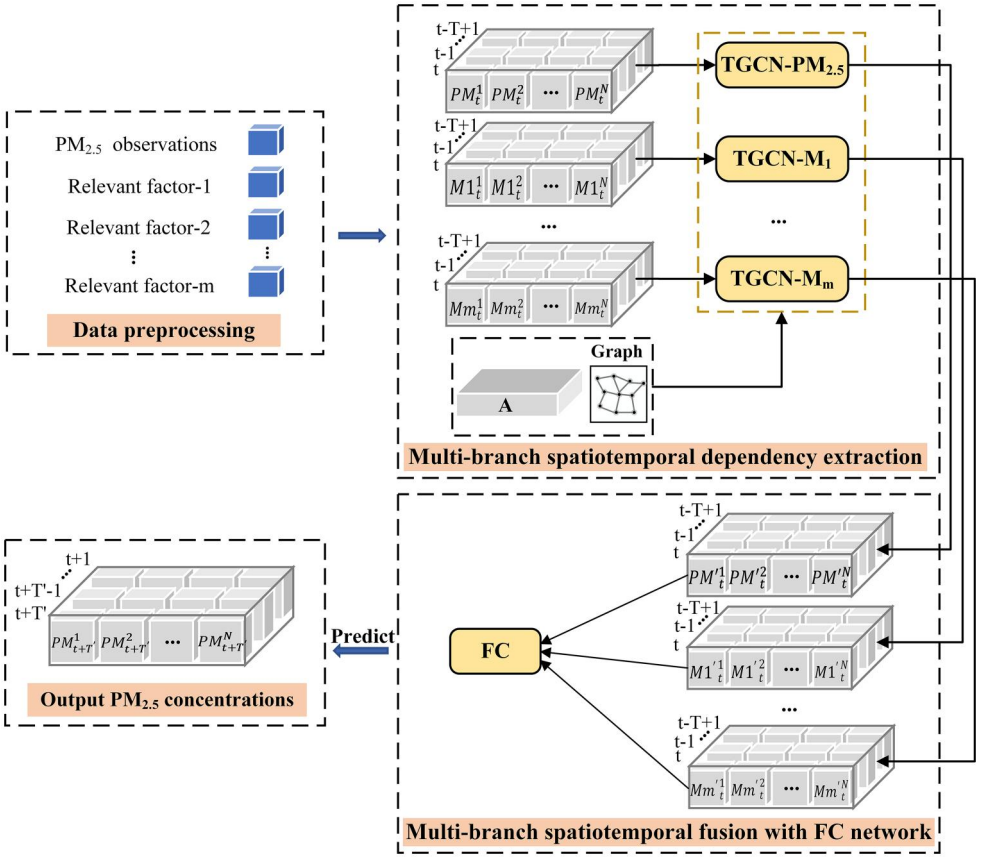

Figure 1. Structure of the proposed MB-TGCN for $PM_{2.5}$ concentration prediction.

determines the inclusion of a specific factor. For example, even though it is well known that emissions are one of the key factors influencing the variation of $PM_{2.5}$ concentration, emission datasets with high spatial and temporal resolutions and high precision are typically unavailable for most regions. In the experiments in this study, several datasets were used as the proxies of emissions, including taxi trajectories, road networks and Points of Interest (POIs), as they may represent the spatiotemporal patterns of emissions to some degree. More details and discussion can be found in Sections 3 and 4.

2.1. Spatial dependency modeling

This study defines an undirected graph $G = (V, E, A)$ (Figure 2) to represent the spatial relationships among stations. $V = \{v_1, v_2, \ldots, v_i, \ldots, v_N\}$ is a set of stations, where v_i is one station and N is the number of stations. (v_i, v_j) represents an edge belonging to E, which is a set of edges. $A \in R^{N \times N}$ is the adjacency matrix, and the spatial dependency between station v_i and station v_j is represented by each element $(A_{i,j})$. Spatial dependency modeling on graphs relies on the topological structure features captured

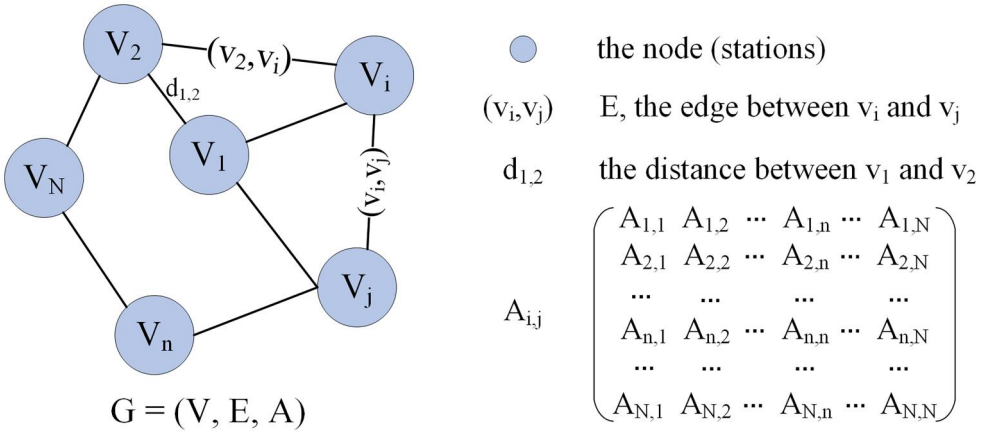

Figure 2. The undirected graph G = (V, E, A).

by the adjacency matrix. An adjacency matrix A is constructed according to the spatial distances between urban air quality monitoring stations:

$$A_{i,j} = \begin{cases} \dfrac{1}{d_{i,j}}, i \neq j \;\; d_{i,j} < D \\ 0, \qquad else \end{cases} \tag{1}$$

where $d_{i,j}$ represents the spatial distance between stations v_i and v_j and D is the distance threshold. A pair of stations is connected when the distance between them is less than D. D is determined by the prediction accuracy, and examples can be found in Section 3.3.

The $PM_{2.5}$ concentration time series at the stations are regarded as the attribute information of the nodes in the graph, expressed as the feature matrix $X \in R^{N \times P}$, where N is the number of air quality monitoring stations and P is the number of station attribute features (i.e. the length of historical sequence of $PM_{2.5}$ concentrations or relevant factors).

Then, through the stacking of several convolutional layers, the GCN model is used to extract the spatial information from the feature matrix. The convolutional layer is defined as:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{l}\theta^{l}\right) \tag{2}$$

where $\tilde{D}$ is the degree matrix such that $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$; $\tilde{A} = A + I$, where I is the identity matrix; H^l is the output of l layer; and $H^0 = X$.

Then, two layers of the GCN model are stacked (Kipf and Welling 2016) to capture spatial dependency:

$$f(X, A) = \sigma\left(\hat{A}ReLU\left(\hat{A}XW_0\right)W_1\right) \tag{3}$$

where $\hat{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$; $W_0 \in R^{P \times H}$ and $W_1 \in R^{H \times T}$ are the weight matrices in the neural network; H is the number of hidden units; $f(X, A) \in R^{N \times T}$ denotes the output with the prediction duration T after the operation of two-layer convolution; and $ReLU()$ is the activation function of the rectified linear units:

$$ReLU(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases} \tag{4}$$

When the input is positive, the output remains the same as the input, which can alleviate the gradient disappearance.

2.2. Temporal dependency modeling

An RNN can learn the features from time series data and generate predictions for future time steps. However, owing to gradient disappearance and gradient explosion, the traditional RNN has limitations in capturing long-term dependencies (Lipton *et al.* 2015). Its variants, such as the LSTM (Hochreiter and Schmidhuber 1997) and GRU (Cho *et al.* 2014), have been proposed to overcome this problem. The GRU model with a simple structure allows each recurrent unit to capture dependencies adaptively at different time scales (Bai *et al.* 2021). Therefore, the GRU is utilized to obtain temporal dependency from the time series of the $PM_{2.5}$ concentration data and relevant factors in our study.

The GRU structure is capable and robust for sequential data modeling. The GRU has two gate structures: r_t is the reset gate and u_t is the update gate (Figure 3(a)). The smaller the r_t is, the more the status information of the last time step will be ignored. u_t controls how much information at the last time step is saved to the status. These two gates can preserve long-term series information, which cannot be removed over time (Cho *et al.* 2014). c_t is the candidate hidden state, and h_t is the hidden

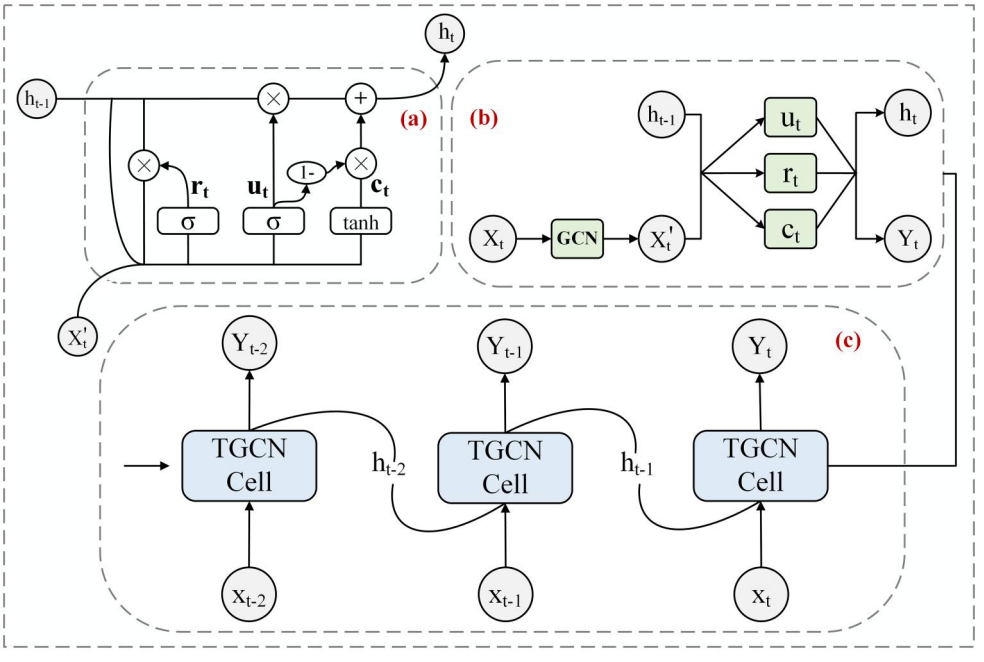

Figure 3. Structure of the temporal graph convolutional network (TGCN) (Zhao et al. 2020a, 2020b). (a) The structure of the GRU model. (b) A cell of TGCN, GCN represents the graph convolution. (c) The overall process of spatiotemporal prediction.

information of the status. In this study, the GRU uses the hidden state h_{t-1} at time $t-1$ and the current $PM_{2.5}$ concentration/relevant factor (x_t) as the inputs to obtain the $PM_{2.5}$ concentration/relevant factor at time $t+1$.

2.3. TGCN for spatiotemporal prediction

To simultaneously capture the spatial and temporal dependencies of $PM_{2.5}$ concentration or relevant factors, this study adopts the basic structure of TGCN proposed by Zhao et al. (2020a, 2020b), in which a GCN model is embedded into a GRU model to form a TGCN cell (Figure 3(b)). The GCN model captures the topological characteristics of the monitoring stations to extract the spatial dependency. The GRU model captures the dynamic variations to extract the temporal patterns. $f(A, X_t)$ corresponds to the graph convolution process. To represent and capture the spatiotemporal variation patterns comprehensively, the TGCN replaces the input feature matrix X_t with the output $f(A, X_t)$ of graph convolution in the GRU structure, as follows:

$$u_t = \sigma(W_u[h_{t-1}, f(A, X_t)] + b_u) \tag{5}$$

$$r_t = \sigma(W_r[h_{t-1}, f(A, X_t)] + b_r) \tag{6}$$

$$c_t = \tanh\left(W_c\left[(r_t * h_{t-1}), f(A, X_t)\right] + b_c\right) \tag{7}$$

$$h_t = u_t * h_{t-1} + (1 - u_t) * c_t \tag{8}$$

where W and b are the weights and biases in the neural network, respectively.

2.4. Multi-branch modeling and fusion

In this study, a multi-branch structure composed of multiple TGCNs is used to extract the spatiotemporal characteristics of $PM_{2.5}$ concentrations and relevant factors (e.g. MVs and EIs) at the monitoring stations. Then, a fully connected network (FCN) is used as a spatiotemporal combinator to fuse the predictions of TGCNs. As illustrated in Figure 1, the MB-TGCN uses the feature matrices of $PM_{2.5}$ and relevant factors as inputs. The second component of MB-TGCN includes multiple TGCNs in parallel to capture the spatiotemporal characteristics and obtain the predicted values of all variables. The predicted values are then combined to form the inputs of a three-layer FCN. Finally, the FCN integrates the predictions of multiple TGCNs to generate the final $PM_{2.5}$ concentrations prediction sequence.

Therefore, given the historical $PM_{2.5}$ concentration time series data, relevant factor data and topology (G) of the monitoring stations, the MB-TGCN aims to predict the $PM_{2.5}$ values for multiple stations over various temporal durations. Thus, the $PM_{2.5}$ concentration prediction problem can be formulated as learning a mapping function, $F(\cdot)$, written as:

$$\left[X^{t+1}, \ldots, X^{t+T'}\right] = F\left(\left[X^{t-T+1}, \ldots, X^{t};\ M^{t-T+1}, \ldots, M^{t}\right]; G(V, E, A)\right) \tag{9}$$

where X^t and M^t are the graph signals of the observed $PM_{2.5}$ concentrations and relevant factors at current time t, respectively. T is the temporal length of the model's input, and T' is the temporal length of the model's output. $G(V, E, A)$ is the undirected graph of air quality monitoring stations.

2.5. Parameter optimization

Time lag is an important parameter in the TGCN and determines the temporal length of the input. In our study, the autocorrelation coefficient (Box *et al.* 1976) is utilized to calculate the temporal correlation between the $PM_{2.5}$ concentrations at the current time and at the previous time for a specific station. This helps identify the best time lag. The more significant the autocorrelation coefficient, the stronger is the temporal correlation. The autocorrelation coefficient is calculated as follows:

$$K_t = \frac{Cov\big(y(t), y(t+k)\big)}{\sigma_{y(t)}\sigma_{y(t+k)}} \tag{10}$$

where K_t denotes the autocorrelation coefficient; $y(t)$ and $y(t+k)$ represent the $PM_{2.5}$ concentrations at time t and time $t+k$, respectively; $Cov()$ is the covariance between variables; and σ is the standard deviation.

In addition, a separate model is trained to predict the values of $PM_{2.5}$ for a specific temporal duration. The number of hidden units in the TGCN can affect the accuracy. For each prediction duration, the most appropriate number of hidden units is selected.

2.6. Performance assessment

To assess the MB-TGCN model's prediction performance, a set of experiments were conducted to compare the proposed model with several commonly used deep learning models, including CNN, GCN and LSTM. The network structure of the CNN includes one convolution layer, one pooling layer and two fully connected layers. The precision curve determines the convolution kernel size and number of neurons. The details of the GCN model are presented in Section 2.1. The LSTM model has two LSTM layers and one fully connected layer. The number of output neurons in the fully connected layer can be adjusted according to different temporal durations of prediction. A persistence model was also included in the experiments to provide a baseline for the comparison. The persistence model simply uses the $PM_{2.5}$ observation at time t as the prediction at the time $t+T'$ $(1 \leq T' \leq 12)$. The persistence model assumes that the $PM_{2.5}$ concentration at a location does not change over a certain period (i.e. T'), which is close to the fact that the $PM_{2.5}$ concentrations at adjacent hours are usually quite similar.

The random forest algorithm (Breiman 2001, Joharestani *et al.* 2019) was used to evaluate the importance of relevant factors that may influence $PM_{2.5}$ concentrations. Then, $PM_{2.5}$ is combined with different relevant factors, in the order of their rankings of importance, to explore the influences of these factors. These factors can be divided into two categories: MVs (including temperature [TP], humidity [HD], wind speed [WS], wind direction [WD] and pressure [PS]) and EIs (including traffic congestion index [TCI], road density and POI density).

Three performance indicators, including the root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2), are used to evaluate the effectiveness of the proposed method. These indicators are formulated as:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{11}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i| \tag{12}$$

$$R^2 = 1 - \frac{\sum_i (\hat{y} - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \tag{13}$$

where m is the number of test samples; y_i is the observed $PM_{2.5}$ concentration; and $\hat{y}_i$ is the predicted $PM_{2.5}$ concentration.

3. Experiments

3.1. Study area and data

Beijing city, China (Figure 4) has adopted a series of policies to mitigate air pollution. The air quality has improved significantly in recent years, but the current $PM_{2.5}$ concentration still exceeds the Chinese National Ambient Air Quality Standards (NAAQS; i.e. an annual average of 35 $\mu g/m^3$) (Huang *et al.* 2021). The hourly $PM_{2.5}$ concentration data at 35 monitoring stations from the Beijing Municipal Ecological and Environmental Monitoring Center (http://www.bjmemc.com.cn/) for the period from 1 January 2017 to 27 March 2018 (a total of 10,806 h) were collected for experiments to assess the performance of the proposed MB-TGCN model.

As mentioned in Section 2, MB-TGCN is a flexible structure that can take into account any number of relevant factors to provide auxiliary spatiotemporal information for the prediction of $PM_{2.5}$ concentrations. Many factors can influence $PM_{2.5}$ concentrations, including meteorology and emissions (Xu *et al.* 2020a, 2020b). For this study area, meteorological data for the same period are publicly available from the American National Climate Data Center (NCDC) (https://www.ncdc.noaa.gov/). Five MVs that were highly correlated with $PM_{2.5}$ were selected for the experiments, namely, HD, PS, TP, WD and WS (Chen *et al.* 2020, Xu *et al.* 2020a, 2020b). The temporal resolution of the meteorological data was the same as that of the $PM_{2.5}$ concentration data. Figure 4 shows the distribution of the two types of monitoring stations, including 35 air quality monitoring stations and 18 meteorological stations.

As emission data with high spatiotemporal resolution and high precision are not available for Beijing city, we utilized several datasets as alternatives to emissions. Taxi trajectories and the road network were used as proxies for the impacts of traffic on air pollution, and POIs were used as the proxies for the impacts of socioeconomic activities on air pollution. The taxi data included the trajectories generated by 3500 taxis during May to July. The road data was extracted from OpenStreetMap (https://www.openstreetmap.org/), and includes 66,957 road segments in Beijing. The POI dataset was obtained from the Gaode map (https://maps.gaode.com/), and comprises 1,766,640 POIs.

3.2. Data preprocessing

The locations of the meteorological stations did not match with those of the air quality monitoring stations (Figure 4). Therefore, we used the inverse distance weight (IDW) interpolation to obtain the meteorological values at the $PM_{2.5}$ monitoring

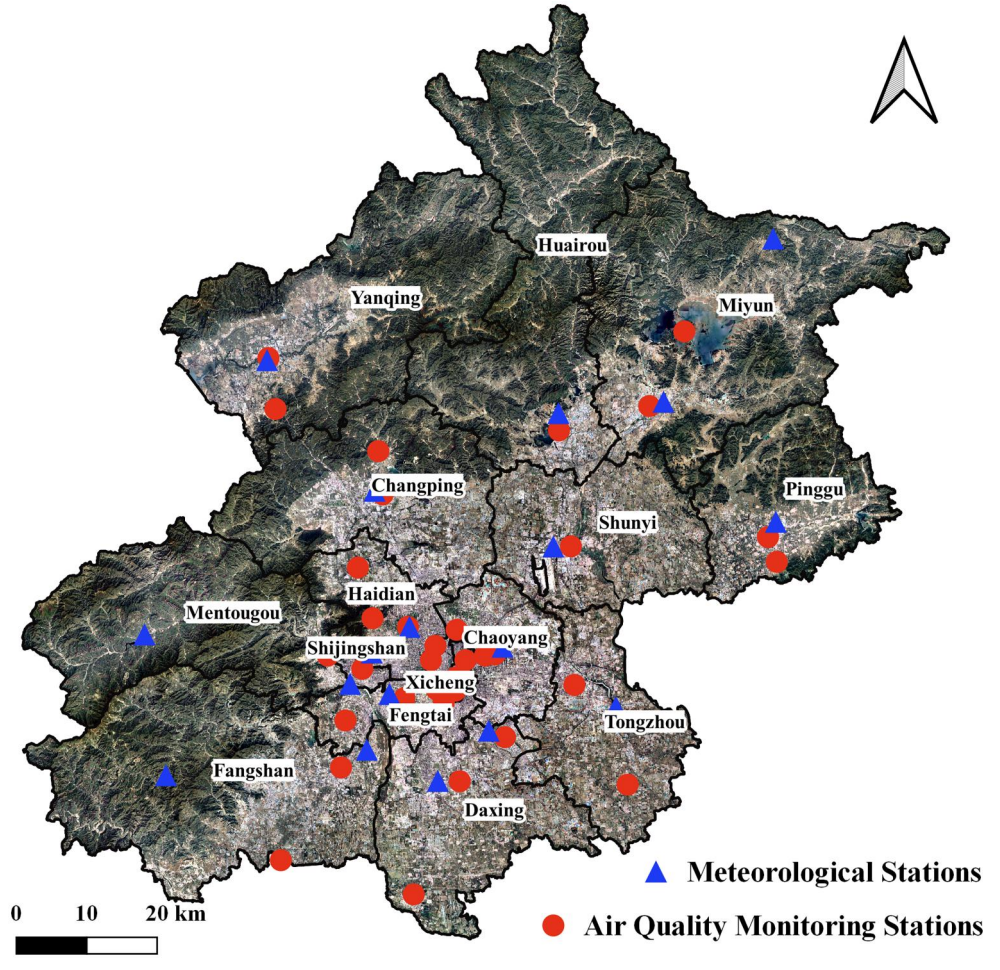

Figure 4. The distribution of air quality monitoring stations and meteorological stations in Beijing.

stations. Thus, a set of feature matrices of MVs for the monitoring stations were constructed, and used as relevant factors for the MB-TGCN model.

Road traffic is one of the key emission sources of $PM_{2.5}$. To represent the traffic conditions of the study period, we used a dataset of taxi trajectories over a week in Beijing, as traffic conditions generally follow a periodic pattern (Zhao *et al.* 2020a, 2020b, Jia *et al.* 2021). Then, the hourly TCI for the road segments within the 5-km buffer zone of each air quality monitoring station was generated using the road network and taxi trajectories, according to the Evaluation Index System of Urban Road Traffic Operation adopted by the Beijing Bureau of Quality and Technical Supervision (BBQTS 2011). The road density and POI density within the 5-km buffer zone of each air quality monitoring station were calculated. The TCI, road density and POI density were regarded as the EIs that represent the spatiotemporal variations of $PM_{2.5}$ emissions to some degree. Thus, a set of feature matrices of EIs for the monitoring stations were constructed and used as relevant factors for the MB-TGCN model. It is worth noting that the TCI values were put into the MB-TGCN model on a one-week cycle. Since the

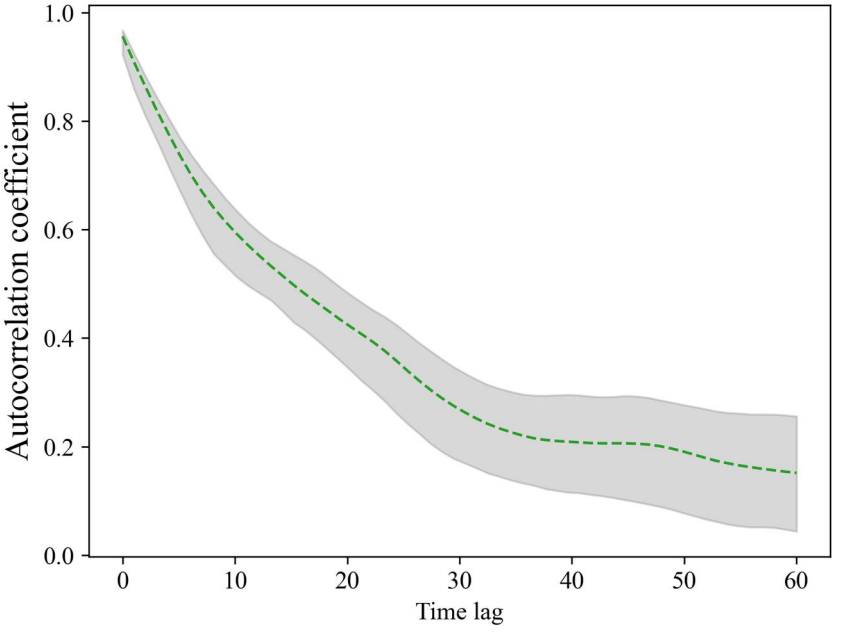

Figure 5. The variation of the autocorrelation coefficients corresponding to different time lags (in hours). The grey area represents the range of autocorrelation coefficients at all monitoring stations. The green dashed line represents the average autocorrelation coefficients.

road density and POI density are relatively static compared with $PM_{2.5}$ and other factors with high-frequency changes, their values stayed the same throughout the time periods of the experiments.

3.3. Hyperparameter settings

In the experiments, the learning rate was set to 0.001 and the batch size was set to 32. The temporal autocorrelation of $PM_{2.5}$ concentration in the dataset decreases as the time lag increases (Figure 5). $PM_{2.5}$ concentration generally exhibits a periodic pattern on a 24-h basis, with higher values during the daytime and lower values during the night time. Therefore, the previous 24-h time series (i.e. historical input length $=$ 24) were empirically utilized to predict the following 12-h $PM_{2.5}$ concentration values (i.e. horizon $=$ 12).

The number of hidden units was determined by analyzing the change in prediction accuracy. We trained models for each prediction duration. Specifically, the number of hidden units was set to 8 for the $+1$ h horizon, 64 for $+2$, $+7$, $+10$ to $+12$ h, 16 for $+3$ h, 100 for $+4$ to $+6$ h, $+9$ h and 128 for $+8$ h. The distance threshold D was set to 60 km according to the accuracy variation when constructing the adjacency matrix. The graph of air quality monitoring stations is shown in Figure 6.

A three-layer FCN was constructed to integrate the outputs of multiple TGCNs. After adjustment, the number of nodes in each hidden layer was set to 3 and the training epochs were selected to 1000. After constructing each variable's feature matrix, 80% of the data were used to train the model, and 20% were used to test the

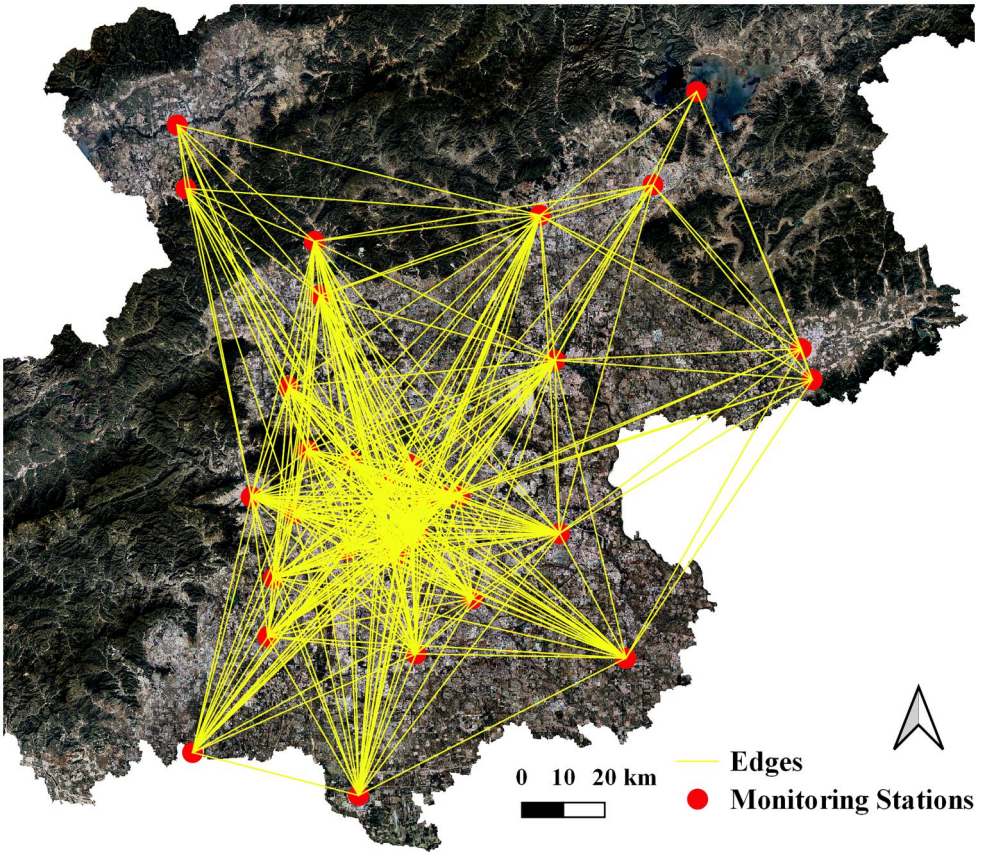

Figure 6. Undirected graph of air quality monitoring stations in Beijing city.

accuracy of the model. The mean squared error (MSE) was set as the loss function of the training process. The model was trained using the Adam optimizer.

3.4. Prediction performance

The prediction accuracies of the MB-TGCN model and other models under different prediction horizons (i.e. prediction durations) are shown in Figure 7. Among the compared models, CNN exhibited the lowest accuracies, while models with temporal dependencies, i.e. MB-TGCN, LSTM and persistence model, exhibited better prediction accuracies. The GCN model performed better than CNN. The result demonstrates that the GCN is more suitable for predicting $PM_{2.5}$ concentrations at such irregularly distributed stations. We further found that the GCN model predicted better than LSTM for the $+1$–$+5$ h predictions, but the accuracy of LSTM increased after $+5$ h horizons.

Figure 7 shows that the proposed MB-TGCN model achieved the highest accuracies in both 1-h and 12-h predictions compared to all other models with respect to all three indicators (RMSE, MAE and R^2). Table 1 presents detailed results for two prediction durations. For the 1-h prediction task, the RMSE of MB-TGCN was 8%, 16%, 21% and 51% lower, and R^2 was 1%, 4%, 4% and 30% higher than those of the persistence

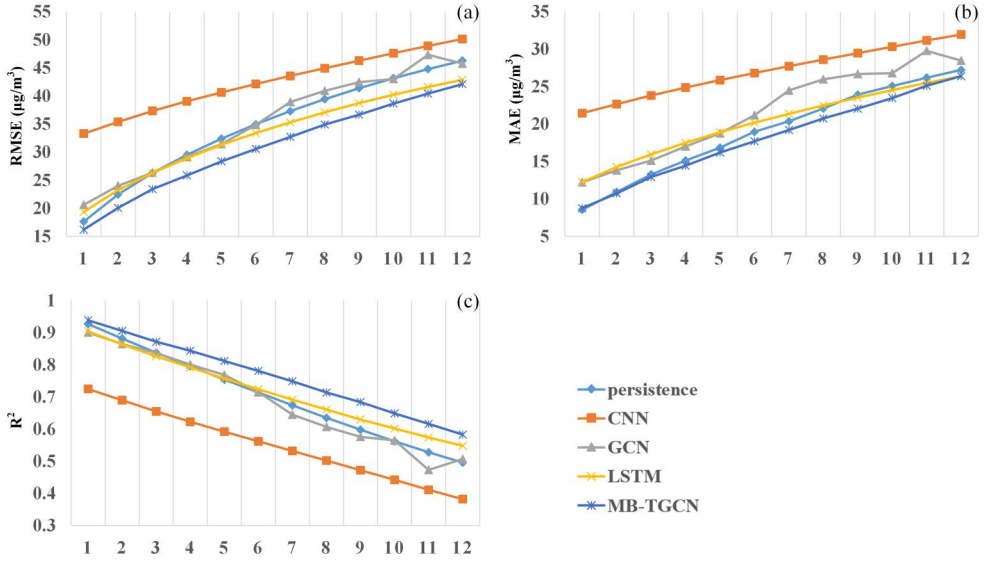

Figure 7. Prediction accuracies of the MB-TGCN model and other methods for different prediction horizons (i.e. prediction temporal durations), including (a) RMSE, (b) MAE, (c) R^2.

Table 1. Accuracies for the 1-h and 12-h prediction tasks.

Prediction duration	indicator	Persistence	CNN	GCN	LSTM	MB-TGCN
1 h	RMSE	17.641	33.306	20.621	19.365	**16.209**
	MAE	**8.576**	21.450	12.229	12.245	8.768
	R^2	0.927	0.725	0.901	0.905	**0.939**
12 h	RMSE	46.281	50.132	45.754	42.880	**42.108**
	MAE	27.221	31.952	28.470	26.384	**26.380**
	R^2	0.496	0.382	0.507	0.548	**0.583**

The bold values in the table represent the best performance.

model, LSTM, GCN and CNN, respectively. In this case, because the $PM_{2.5}$ concentration to be predicted was closely similar to the concentration in the previous hour, the persistence model achieved the second-highest accuracy among all models, which is understandable and in line with the expectation. MB-TGCN's R^2 was 1% higher and its RMSE was 8% lower than those of the persistence model, indicating its effective improvement over the persistence model.

For the 12-h prediction task, the RMSE of MB-TGCN was 9%, 2%, 8% and 16% lower, and R^2 was 18%, 6%, 15% and 53% higher than those of the persistence model, LSTM, GCN and CNN, respectively. In this case, the $PM_{2.5}$ concentrations to be predicted were different from the values in the previous 24 h, and, therefore, more complex models exhibited better prediction performance. LSTM, with its capability of extracting complex temporal patterns, achieved the second-highest accuracy. MB-TGCN's RMSE was 2% lower and its R^2 was 6% higher than those of LSTM, demonstrating a consistent improvement over LSTM. In addition, for a longer prediction duration, MB-TGCN also achieved more improvements in R^2 over other models.

Such results demonstrated that capturing both the spatial relationships and temporal patterns did help improve the prediction accuracy.

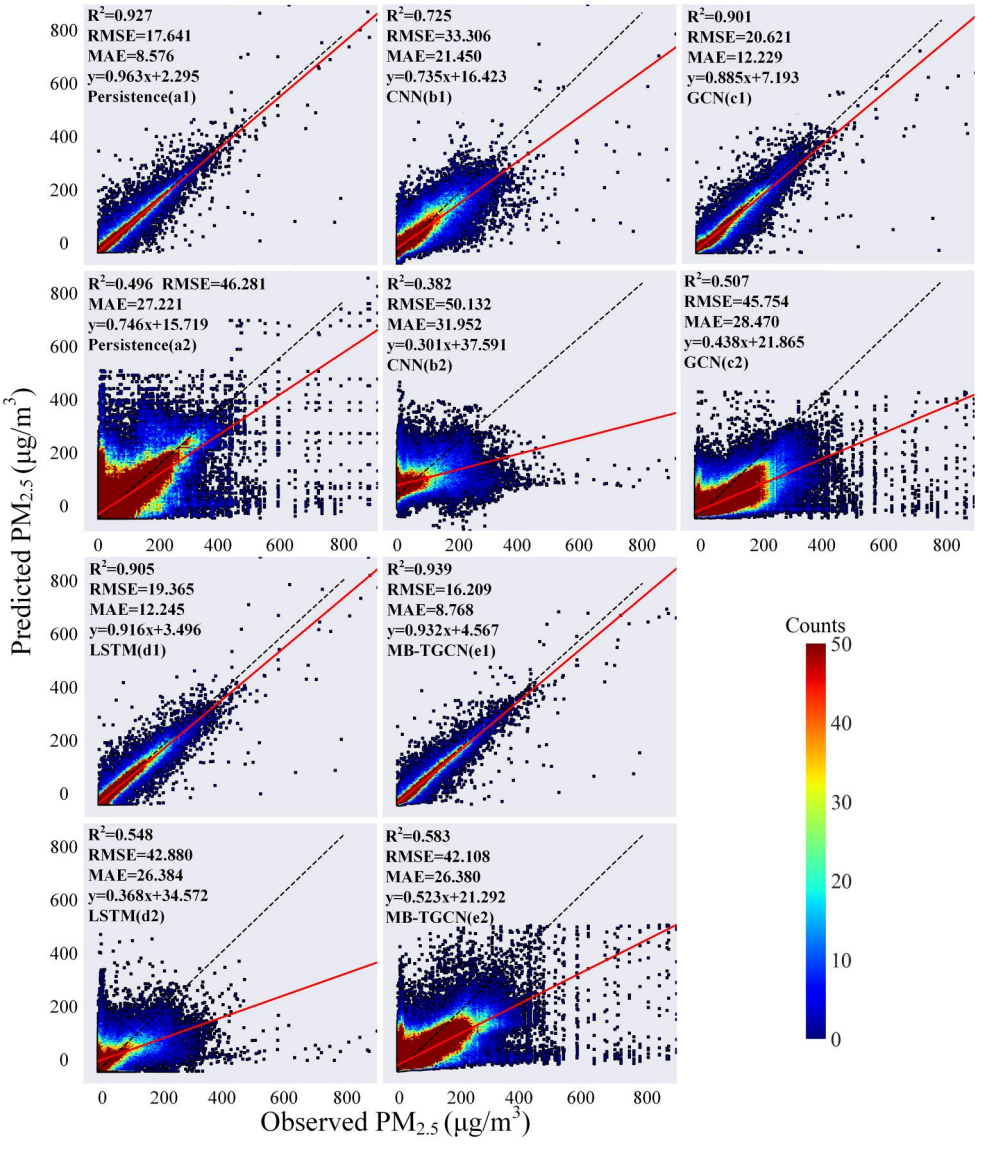

Figure 8. The results of the proposed MB-TGCN model and other models: (a1)∼(e1) 1-h prediction tasks, and (a2)∼(e2) 12-h prediction tasks. The solid lines are the regression lines, and the dashed lines are the y = x reference lines.

Furthermore, we visualized the predicted $PM_{2.5}$ concentrations for 1-h and 12-h by MB-TGCN and the other four models (Figure 8). The predicted $PM_{2.5}$ values of GCN and MB-TGCN are generally smaller than the observed values. In general, MB-TGCN has a superior predictive capability compared with other models.

3.5. Effects of relevant factors on $PM_{2.5}$ prediction

A series of experiments were conducted using MB-TGCN with different combinations of relevant factors to explore the contributions of relevant factors to the predictions.

Table 2. The accuracies of MB-TGCN with different combinations of relevant factors.

	$PM_{2.5}$			$PM_{2.5}$+ MVs + EIs			$PM_{2.5}$+ MVs		
T	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
+1 h	16.395	**8.677**	0.937	16.423	8.812	0.937	**16.209**	8.768	**0.939**
+4 h	27.108	15.090	0.828	26.003	14.737	0.842	**25.857**	**14.435**	**0.844**
+8 h	36.666	21.857	0.685	35.348	21.226	0.707	**34.915**	**20.738**	**0.714**
+12 h	43.092	27.078	0.563	42.767	26.748	0.570	**42.108**	**26.380**	**0.583**

As shown in Table 2, the models with relevant factors outperformed the model without any relevant factor, indicating the multi-branch structure of MB-TGCN that takes into account the spatiotemporal dependencies of relevant factors effectively improved the model's prediction capability. However, the prediction accuracy of MB-TGCN with both MVs and EIs was slightly lower than that of the model with only MVs. This is most likely because the EIs used in the experiments are insufficient in representing the spatiotemporal variations of actual emissions. More discussion can be found in Section 4.1.

To explore the effects of MVs on $PM_{2.5}$ prediction, $PM_{2.5}$ was combined with different MVs in the order of importance (HD > WS > TP > WD > PS) in the experiments. As shown in Figure 9, the accuracy of the model with MVs was higher than that of the model with only $PM_{2.5}$. For the 12-h forecasting task, MB-TGCN with MVs improved R^2 by 4% compared to the single-input model.

Variations were observed among different combinations of MVs for different prediction horizons. The minimum RMSE (16.161 $\mu g/m^3$) was achieved by integrating the first four MVs for the 1-h forecasting task. In general, the combinations that include HD, WS, TP, WD and PS can produce better results than $PM_{2.5}$ alone. In our proposed MB-TGCN model, using $PM_{2.5}$ with HD, WS, TP, WD and PS performed the best when the prediction duration was less than 8 h. However, when predicting for $9 \sim 11$ h, the errors of $PM_{2.5}$ with HD, WS, TP and WD were more minor than other combinations. When the prediction duration was 12 h, the prediction error using $PM_{2.5}$ with HD and WS (shown in green line) was relatively more minor compared with other combinations, indicating that HD and WS are the most relevant MVs.

4. Discussion

4.1. Interpretation of the findings in the context of available evidence

The experiment results showed that MB-TGCN achieved higher accuracy in short-term prediction of $PM_{2.5}$ concentrations than the persistence model and other deep learning methods (e.g. CNN, GCN and LSTM). The proposed model is able to effectively extract and fuse both the spatial and temporal dependencies of $PM_{2.5}$ and relevant factors, while CNN and GCN only capture the spatial dependency and LSTM only captures the temporal dependency of $PM_{2.5}$.

These findings proved our initial hypothesis that the method with spatiotemporal dependencies has higher accuracy than the methods with a single aspect. $PM_{2.5}$ concentrations exhibit heterogeneities in both temporal and spatial distributions, thus the spatiotemporal correlations among air quality stations are beneficial for future

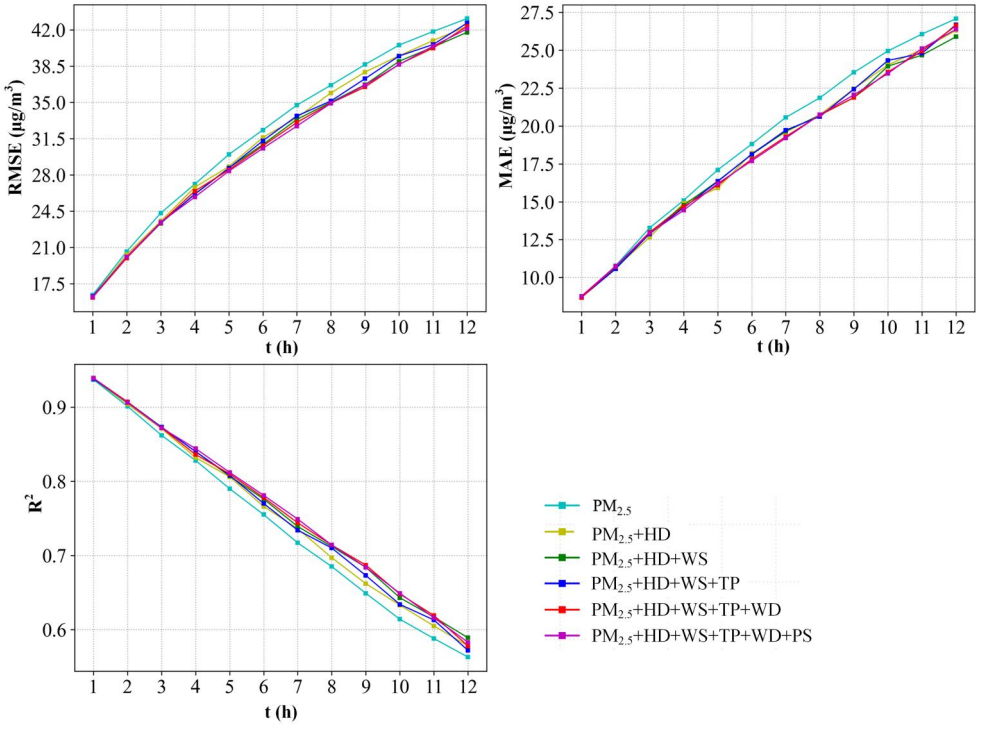

Figure 9. Prediction accuracies of different combinations of meteorological variables.

prediction, owing to the interaction between $PM_{2.5}$ concentrations at different locations (Wen *et al.* 2019, Zhao *et al.* 2019).

Our hypothesis that the spatial dependency among stations can enhance the prediction performance is quantitatively confirmed. The results showed GCN performed better than CNN in extracting the spatial features of irregularly distributed stations and achieved a noticeable forecast accuracy improvement. This reflects the necessity of incorporating the spatial interactions among stations into predictions, which has not been considered in previous studies of spatiotemporal feature extraction (Huang and Kuo 2018, Zhao *et al.* 2019). The GCN model can be operated on the graph's adjacency matrix to obtain the topological characteristics of the graph and capture the spatial dependency among stations (Zhang *et al.* 2020a, 2020b, 2020c).

Besides the capability of capturing the spatiotemporal dependencies of $PM_{2.5}$, the multi-branch structure of MB-TGCN provides a flexible approach to integrate the spatiotemporal dependencies of other factors that may influence the variation of $PM_{2.5}$ concentration, hence to give extra information to the prediction. Our experiments demonstrated that the models with relevant factors achieved higher accuracies than the model without other factors for all prediction durations, confirming the effectiveness of the proposed model. The model with MVs performed the best in the experiments. Such results are consistent with other studies' findings that the variation of $PM_{2.5}$ concentration is greatly influenced by the variations of some particular MVs (e.g. HD and WS) (Li *et al.* 2019, Zhai *et al.* 2019). In many studies on $PM_{2.5}$ modeling and prediction, MVs were used as the inputs to the models, but their spatial and temporal

dependencies were usually insufficiently considered (Bai *et al.* 2016, Li *et al.* 2017). In MB-TGCN, the multi-branch structure composed of multiple TGCNs enables extracting the spatiotemporal features of MVs in parallel with the spatiotemporal features of $PM_{2.5}$ concentration, helping improve the prediction capability. We also found that each MV's contribution was not consistent over prediction durations. The prediction results of $PM_{2.5}$ with HD, WS, TP, WD and PS were better in 8 h. While in 9–11 h, PS may partly weaken the prediction accuracy. Especially, to predict 12-h $PM_{2.5}$ concentrations, we found that using $PM_{2.5}$ with HD and WS was the best combination, and other MVs became interference factors. Previous studies also indicated that the effects of MVs change over time, such as TP and WS (Zhang *et al.* 2020a, 2020b, 2020c). With the increase in prediction duration, the MVs may interact with each other to increase the uncertainty of the prediction.

In addition, the model with both MVs and EIs performed slightly less well than the model with only MVs in our experiments. The feature matrices of MVs at the air quality stations included the time series of MVs with the same length and temporal resolution as the time series of $PM_{2.5}$ concentration, containing sufficient information of the spatiotemporal dependencies of MVs, and therefore effectively improved the prediction accuracies. In contrast, the EI datasets used in our experiments (i.e. taxi trajectories, road network and POIs) are merely proxies of emissions. Also, road network and POIs are relatively static compared with the highly dynamic $PM_{2.5}$ concentrations. Therefore, their contributions to the high-frequency spatiotemporal variations of $PM_{2.5}$ concentration are minimal. Thus, emission data with high spatiotemporal resolutions and high precisions are needed to improve the prediction accuracy further.

Our study provides a deep learning method to predict hourly short-term $PM_{2.5}$ concentrations, while some studies predicted the air pollutant concentrations for longer terms and coarser temporal resolutions (Wang *et al.* 2020, 2021, Xiao *et al.* 2021). Both are predictions of future air pollutant concentrations. Another type of prediction is the real-time estimation of air quality, which uses real-time monitoring data and multi-source urban information to speculate the air quality index at any location of the city at the current time (Tian *et al.* 2021, Jiang *et al.* 2021, Zhang *et al.* 2021). Future forecasting and real-time estimation serve different purposes. Real-time estimation is to predict the values at spatial locations where measured values are unavailable, providing the current air quality information for environment evaluation. Future forecasting is to predict the values at future moments, providing temporally advanced air quality information for environmental risk assessment and decision-making of government management and residents' activity planning.

4.2. Strengths and limitations

Some strengths of this study need to be emphasized. First, we make a new attempt to utilize graph deep learning models to extract stations' spatiotemporal dependencies for $PM_{2.5}$ prediction, as an alternative to the widely used CNN models. Both methods can extract features from the spatial perspective. However, the undirected graph is more suitable for irregularly located stations. Second, the multi-branch structure of our

model enables the extraction of spatiotemporal dependencies of multiple auxiliary factors that influence the spatiotemporal variation of $PM_{2.5,}$ and the integration of such information to enhance the prediction capability. With these strengths, the proposed MB-TGCN is able to achieve high-precision short-term forecasts, and the results can provide references and suggestions for governments' measures of air pollution mitigation and residents' activity planning.

Our study is limited in several ways, however. First, the two main factors influencing PM concentration are meteorological conditions and emissions (He *et al.* 2017, Yang *et al.* 2018a, 2018b). While meteorological data are widely available in the public domain, high-quality emission data (including both emission sources and amounts) with satisfactory spatiotemporal resolutions are not often available. Therefore, in the experiments of this study, only the taxi trajectories, road density and POIs were used as the proxies of emissions. Nevertheless, these types of data lack specific emission sources and precise emission amounts, providing incomplete information on pollutant emissions. As a result, the inclusion of these 'emission indicators' in the model did not enhance the prediction accuracies.

In recent years, efforts have been made to produce high-quality emission datasets. For example, China is establishing the emission inventory and datasets (Tong *et al.* 2020, Shan *et al.* 2020). The availability of such emission datasets will certainly improve the prediction accuracy of air pollution. In terms of modeling capability, the experiments demonstrated that MB-TGCN was able to generate accurate predictions of $PM_{2.5}$ concentrations without emission data. On the one hand, emissions information and patterns are implicitly embedded in the $PM_{2.5}$ data and can be extracted by MB-TGCN. On the other hand, MB-TGCN is a flexible multi-branch structure, which provides the capability of utilizing multiple datasets of auxiliary factors, including emissions if their datasets are available. To utilize an emission dataset for $PM_{2.5}$ prediction, a new TGCN can be added to the multi-branch structure of MB-TGCN, and is used to capture the spatiotemporal characteristics of emissions and predict future emissions. The FCN then integrates the predicted values from all TGCNs to finally predict $PM_{2.5}$ concentrations.

Second, under the assumption of a constant spatial dependency over time, MB-TGCN uses a pre-determined graph structure. However, the correlations among the $PM_{2.5}$ concentrations at multiple stations may have a temporal heterogeneity. Therefore, an adaptive mechanism is needed to capture the temporally dynamic spatial dependency among monitoring stations to improve the prediction capability further.

Third, the impacts of MVs on $PM_{2.5}$ concentrations vary with season and location (Chen *et al.* 2020). It remains unclear how these MVs affected $PM_{2.5}$ concentrations. Therefore, we will try to combine data-driven models and Chemical Transport Models to explore the interaction between $PM_{2.5}$ and MVs in future studies, including the establishment of causality (Chen *et al.* 2018a, 2018b, He *et al.* 2021).

In addition, improving the performance of predicting sudden significant changes in $PM_{2.5}$ concentration is one of the future directions of our work. Emergency predictions are needed for government agencies to devise measures for environment enhancement and forecast serious deterioration of air quality.

5. Conclusion

In this study, a MB-TGCN model was proposed to predict short-term $PM_{2.5}$ concentrations at fine temporal resolutions. Using a multi-branch structure of TGCNs, the model was able to extract both the spatial and temporal dependencies of $PM_{2.5}$ and relevant factors. A FCN was then used to fuse the predictions of all TGCNs and to predict $PM_{2.5}$ concentrations for future moments.

Experiments using the datasets of $PM_{2.5}$ and relevant factors in Beijing showed that the proposed model achieved the best prediction accuracies for 1–12-h durations, in comparison with several deep learning models (i.e. CNN, GCN and LSTM). In the experiments, we also found that using $PM_{2.5}$ with HD and WS was the best combination when predicting 12-h concentrations. Furthermore, emission data with high spatiotemporal resolutions and high precisions are greatly needed to improve the prediction accuracy further.

Our findings highlight that the method with spatiotemporal dependencies outperforms methods based on a single aspect (e.g. GCN with spatial dependency, and LSTM with temporal dependency), and can produce higher prediction accuracies for various prediction durations. Although some challenges still need to be overcome, graph deep learning models provide a new perspective for $PM_{2.5}$ concentration prediction. Future studies will be concentrated on building the emissions inventory and extending potential driving factors datasets.

Acknowledgment

We are deeply grateful to Professor May Yuan, Professor Shawn Laffan, and the anonymous reviewers for their meticulous suggestions and diligent works.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Key Research and Development Program of China (Grant No. 2023YFB3906803).

Data and codes availability statement

The data and codes that support the findings of this study are available on Figshare at https://doi.org/10.6084/m9.figshare.19729480.

Notes on contributors

Dr. Qingfeng Guan is a Professor in the School of Geography and Information Engineering at China University of Geosciences (Wuhan). His research focuses on high-performance spatial intelligent computing and urban modeling. Prof. Qingfeng Guan contributed to the conception of the research idea, the supervision of the experiments, and the writing of the manuscript.

Ms. Jingyi Wang received the M.S. degree in cartography and geographical information engineering from China University of Geosciences (Wuhan) in 2022. Her research focuses on data mining and spatiotemporal modeling. Ms. Jingyi Wang contributed to the formulation of the model, the experiments, and the writing of the manuscript.

Mr. Shuliang Ren is currently a Ph.D. student in GIScience at the Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing. He received the M.S. degree in cartography and geographical information engineering from China University of Geosciences (Wuhan) in 2022. His research interests are GeoAI, spatial analysis, and urban data mining. Mr. Shuliang Ren contributed to the research design and methodology.

Mr. Huan Gao received the B.E. degree from the China University of Geosciences (Wuhan) in 2018, where he is pursuing the Ph.D. degree in surveying and mapping. His research focuses on high-performance geospatial computing. Mr. Huan Gao contributed to the experiments and the writing of the manuscript.

Mr. Zhewei Liang is a master candidate in the School of Geography and Information Engineering at China University of Geosciences (Wuhan). His research focuses on data mining and parallel computing. Mr. Zhewei Liang contributed to the writing of the manuscript.

Mr. Junyi Wang received the M.S. degree in cartography and geographical information engineering from the China University of Geosciences (Wuhan) in 2019. His research focuses on data mining and spatiotemporal modeling. Mr. Junyi Wang contributed to the experiments.

Dr. Yao Yao (Corresponding author) is currently a Professor in the School of Geography and Information Engineering at China University of Geosciences (Wuhan) and a researcher of Center for Spatial Information Science at the University of Tokyo. His research focuses on geospatial big data mining and computational urban sciences. Prof. Yao Yao contributed to the research design, the supervision of the experiments, and the writing of the manuscript.

ORCID

Huan Gao http://orcid.org/0000-0003-0401-5376

References

Aliyu, Y.A. and Botai, J.O., 2018. Reviewing the local and global implications of air pollution trends in Zaria, northern Nigeria. *Urban Climate*, 26, 51–59.

Athira, A., *et al.*, 2018. DeepAirNet: applying recurrent networks for air quality prediction. *Procedia Computer Science*, 132, 1394–1403.

Bai, Y., *et al.*, 2016. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmospheric Pollution Research*, 7 (3), 557–566.

Bai, Y., *et al.*, 2021. Regression modeling for enterprise electricity consumption: A comparison of recurrent neural network and its variants. *International Journal of Electrical Power & Energy Systems*, 126, 106612.

BBQTS. 2011. *The evaluation index system of urban road traffic operation.* Beijing, China: BBQTS.

Box, G.E.P., *et al.*, 1976. Time series analysis: forecasting and control. *Journal of Time*, 31 (2), 238–242.

Breiman, L., 2001. Random forests. *Machine Learning*, 45 (1), 5–32.

Brook, R.D., *et al.*, 2010. Particulate matter air pollution and cardiovascular disease an update to the scientific statement from the American Heart Association. *Circulation*, 121 (21), 2331–2378.

Bruna, J., *et al.*, 2013. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*.

Bui, T., Le, V., and Cha, S., 2018. A deep learning approach for forecasting air pollution in South Korea using LSTM. *arXiv preprint arXiv:1804.07891*.

Chen, C., *et al.*, 2018a. Short-term exposures to $PM_{2.5}$ and cause-specific mortality of cardiovascular health in China. *Environmental Research*, 161, 188–194.

Chen, Z., *et al.*, 2018b. Understanding meteorological influences on $PM_{2.5}$ concentrations across China: a temporal and spatial perspective. *Atmospheric Chemistry and Physics*, 18 (8), 5343–5358.

Chen, Z., *et al.*, 2020. Influence of meteorological conditions on $PM_{2.5}$ concentrations across China: a review of methodology and mechanism. *Environment International*, 139, 105558.

Cho, K., *et al.*, 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Defferrard, M., Bresson, X., and Vandergheynst, P., 2016. Convolutional neural networks on graphs with fast localized spectral filtering. *30th Conference on neural information processing systems (NIPS)*. Barcelona, Spain, 3844–3852.

Dong, F., Yu, B., and Pan, Y., 2019. Examining the synergistic effect of CO_2 emissions on $PM_{2.5}$ emissions reduction: evidence from China. *Journal of Cleaner Production*, 223, 759–771.

Duan, W., *et al.*, 2021. Influencing factors of $PM_{2.5}$ and O_3 from 2016 to 2020 based on DLNM and WRF-CMAQ. *Environmental Pollution*, 285, 117512.

García Nieto, P.J., *et al.*, 2013. A SVM-based regression model to study the air quality at local scale in Oviedo urban area (Northern Spain): a case study. *Applied Mathematics and Computation*, 219 (17), 8923–8937.

Gautam, S., Patra, A.K., and Kumar, P., 2019. Status and chemical characteristics of ambient $PM_{2.5}$ pollution in China: a review. *Environment, Development and Sustainability*, 21 (4), 1649–1674.

He, J., *et al.*, 2017. Air pollution characteristics and their relation to meteorological conditions during 2014–2015 in major Chinese cities. *Environmental Pollution (Barking, Essex: 1987)*, 223, 484–496.

He, S., *et al.*, 2021. Dynamic relationship between meteorological conditions and air pollutants based on a mixed Copula model. *International Journal of Climatology*, 41 (4), 2611–2624.

Hill, W., *et al.*, 2023. Lung adenocarcinoma promotion by air pollutants. *Nature*, 616 (7955), 159–167.

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

Hsieh, H., Lin, S., and Zheng, Y., 2015. Inferring air quality for station location recommendation based on urban big data. *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. Sydney, NSW, Australia: Association for Computing Machinery, 437–446.

Huang, C.J. and Kuo, P.H., 2018. A deep CNN-LSTM model for particulate matter ($PM_{2.5}$) forecasting in smart cities. *Sensors*, 18 (7), 2220.

Huang, R., and Chun, L., 2017. Seasonal variation characteristics and forecasting model of $PM_{2.5}$ in Changsha, Central City in China. *Journal of Environmental & Analytical Toxicology*, 7 (1), 429–435.

Huang, X., *et al.*, 2021. Characteristics of $PM_{2.5}$ pollution in Beijing after the improvement of air quality. *Journal of Environmental Sciences (China)*, 100, 1–10.

Jia, M., *et al.*, 2019. Regional air quality forecast using a machine learning method and the WRF model over the Yangtze River Delta, East China. *Aerosol and Air Quality Research*, 19 (7), 1602–1613.

Jia, X., *et al.*, 2021. Missing data imputation for traffic congestion data based on joint matrix factorization. *Knowledge-Based Systems*, 225, 107114.

Jiang, M., *et al.*, 2017. Modelling seasonal GWR of daily $PM_{2.5}$ with proper auxiliary variables for the Yangtze River Delta. *Remote Sensing*, 9 (4), 346.

Jiang, T., *et al.*, 2021. Estimation of hourly full-coverage $PM_{2.5}$ concentrations at 1-km resolution in China using a two-stage random forest model. *Atmospheric Research*, 248, 105146.

Joharestani, M.Z., *et al.*, 2019. $PM_{2.5}$ prediction based on random forest, xgboost, and deep learning using multisource remote sensing data. *Atmosphere*, 10 (7), 373.

Johnson, M., *et al.*, 2010. Evaluation of land-use regression models used to predict air quality concentrations in an urban area. *Atmospheric Environment*, 44 (30), 3660–3668.

Kampa, M. and Castanas, E., 2008. Human health effects of air pollution. *Environmental Pollution (Barking, Essex: 1987)*, 151 (2), 362–367.

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

Krishan, M., *et al.*, 2019. Air quality modelling using long short-term memory (LSTM) over NCT-Delhi, India. *Air Quality, Atmosphere & Health*, 12 (8), 899–908.

Kurt, A. and Oktay, A.B., 2010. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. *Expert Systems with Applications*, 37 (12), 7986–7992.

Li, R., *et al.*, 2019. Air pollution characteristics in China during 2015–2016: spatiotemporal variations and key meteorological factors. *The Science of the Total Environment*, 648, 902–915.

Li, X., *et al.*, 2017. Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation. *Environmental Pollution (Barking, Essex: 1987))*, 231 (Pt 1), 997–1004.

Li, Z., *et al.*, 2021. A practical framework for predicting residential indoor $PM_{2.5}$ concentration using land-use regression and machine learning methods. *Chemosphere*, 265, 129140.

Lipton, Z.C., Berkowitz, J., and Charles, E., 2015. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*.

Liu, D. and Sun, K., 2019. Short-term $PM_{2.5}$ forecasting based on CEEMD-RF in five cities of China. *Environmental Science and Pollution Research International*, 26 (32), 32790–32803.

Liu, N., *et al.*, 2021. Prediction of $PM_{2.5}$ concentrations at unsampled points using multiscale geographically and temporally weighted regression. *Environmental Pollution*, 284, 117116.

Liu, W., *et al.*, 2019. Meteorological pattern analysis assisted daily $PM_{2.5}$ grades prediction using SVM optimized by PSO algorithm. *Atmospheric Pollution Research*, 10 (5), 1482–1491.

Liu, W., Xu, Z., and Yang, T., 2018. Health effects of air pollution in China. *International Journal of Environmental Research and Public Health*, 15 (7), 1471.

Luo, J. and Du, Y., 2020. Detecting community structure and structural hole spanner simultaneously by using graph convolutional network based auto-encoder. *Neurocomputing*, 410, 138–150.

Mao, Y., *et al.*, 2016. Impacts of meteorological parameters and emissions on decadal and interannual variations of black carbon in China for 1980–2010. *Journal of Geophysical Research: Atmospheres*, 121 (4), 1822–1843.

Mirzaei, M., Amanollahi, J., and Tzanis, C.G., 2019. Evaluation of linear, nonlinear, and hybrid models for predicting $PM_{2.5}$ based on a GTWR model and MODIS AOD data. *Air Quality, Atmosphere & Health*, 12 (10), 1215–1224.

Qi, Y., *et al.*, 2019. A hybrid model for spatiotemporal forecasting of $PM_{2.5}$ based on graph convolutional neural network and long short-term memory. *The Science of the Total Environment*, 664, 1–10.

Saide, P.E., *et al.*, 2011. Forecasting urban PM_{10} and $PM_{2.5}$ pollution episodes in very stable nocturnal conditions and complex terrain using WRF-Chem CO tracer model. *Atmospheric Environment*, 45 (16), 2769–2780.

Scarselli, F., *et al.*, 2009. The graph neural network model. *IEEE Transactions on Neural Networks*, 20 (1), 61–80.

Scheffe, R.D. and Morris, R.E., 1993. A review of the development and application of the Urban Airshed model. *Atmospheric Environment Part B. Urban Atmosphere*, 27 (1), 23–39.

Shan, Y., *et al.*, 2020. China CO_2 emission accounts 2016–2017. *Scientific Data*, 7 (1), 54.

Sharma, N., *et al.*, 2018. Forecasting air pollution load in Delhi using data analysis tools. *Procedia Computer Science*, 132, 1077–1085.

Shi, L., *et al.*, 2022. A balanced social LSTM for $PM_{2.5}$ concentration prediction based on local spatiotemporal correlation. *Chemosphere*, 291 (Pt 3), 133124.

Suleiman, A., Tight, M.R., and Quinn, A.D., 2019. Applying machine learning methods in managing urban concentrations of traffic-related particulate matter (PM_{10} and $PM_{2.5}$). *Atmospheric Pollution Research*, 10 (1), 134–144.

Tesche, T.W., *et al.*, 2006. CMAQ/CAMx annual 2002 performance evaluation over the eastern US. *Atmospheric Environment*, 40 (26), 4906–4919.

Tian, H., *et al.*, 2021. Estimating $PM_{2.5}$ from multisource data: a comparison of different machine learning models in the Pearl River Delta of China. *Urban Climate*, 35, 100740.

Tong, D., *et al.*, 2020. Dynamic projection of anthropogenic emissions in China: methodology and 2015-2050 emission pathways under a range of socio-economic, climate policy, and pollution control scenarios. *Atmospheric Chemistry and Physics*, 20 (9), 5729–5757.

Tucker, W.G., 2000. An overview of $PM_{2.5}$ sources and control strategies. *Fuel Processing Technology*, 65-66, 379–392.

Veličković, P., *et al.*, 2017. Graph attention networks. *arXiv preprint arXiv: 1710.10903*.

Wang, Y., Wang, H., and Zhang, S., 2020. Prediction of daily $PM_{2.5}$ concentration in China using data-driven ordinary differential equations. *Applied Mathematics and Computation*, 375, 125088.

Wang, Z., *et al.*, 2021. High-resolution prediction of the spatial distribution of $PM_{2.5}$ concentrations in China using a long short-term memory model. *Journal of Cleaner Production*, 297, 126493.

Wen, C., *et al.*, 2019. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *The Science of the Total Environment*, 654, 1091–1099.

Xiao, Q., *et al.*, 2021. Evaluation of gap-filling approaches in satellite-based daily $PM_{2.5}$ prediction models. *Atmospheric Environment*, 244, 117921.

Xu, G., *et al.*, 2020a. Analysis of the driving factors of $PM_{2.5}$ concentration in the air: a case study of the Yangtze River Delta, China. *Ecological Indicators*, 110, 105889.

Xu, X., *et al.*, 2020b. Impacts of meteorology and emission control on the abnormally low particulate matter concentration observed during the winter of 2017. *Atmospheric Environment*, 225, 117377.

Yang, D., *et al.*, 2018a. Quantifying the influence of natural and socioeconomic factors and their interactive impact on $PM_{2.5}$ pollution in China. *Environmental Pollution (Barking, Essex: 1987))*, 241, 475–483.

Yang, W., *et al.*, 2018b. Prediction of hourly $PM_{2.5}$ using a space-time support vector regression model. *Atmospheric Environment*, 181, 12–19.

Yang, Y., Liao, H., and Lou, S., 2015. Decadal trend and interannual variation of outflow of aerosols from East Asia: roles of variations in meteorological parameters and emissions. *Atmospheric Environment*, 100, 141–153.

Yao, Y., *et al.*, 2019. Properties of particulate matter and gaseous pollutants in Shandong, China: daily fluctuation, influencing factors, and spatiotemporal distribution. *The Science of the Total Environment*, 660, 384–394.

Zhai, S., *et al.*, 2019. Fine particulate matter ($PM_{2.5}$) trends in China, 2013–2018: separating contributions from anthropogenic emissions and meteorology. *Atmospheric Chemistry and Physics*, 19 (16), 11031–11041.

Zhang, B., *et al.*, 2020a. Constructing a $PM_{2.5}$ concentration prediction model by combining auto-encoder with Bi-LSTM neural networks. *Environmental Modelling & Software*, 124, 104600.

Zhang, L., *et al.*, 2020b. Spatiotemporal variations and influencing factors of $PM_{2.5}$ concentrations in Beijing, China. *Environmental Pollution*, 262, 114276.

Zhang, T., *et al.*, 2021. Satellite-based ground $PM_{2.5}$ estimation using a gradient boosting decision tree. *Chemosphere*, 268, 128801.

Zhang, Y., *et al.*, 2020b. A novel residual graph convolution deep learning model for short-term network-based traffic forecasting. *International Journal of Geographical Information Science*, 34 (5), 969–995.

Zhao, D., *et al.*, 2020a. A graph convolutional network-based deep reinforcement learning approach for resource allocation in a cognitive radio network. *Sensors*, 20 (18), 5216.

Zhao, J., *et al.*, 2019. Long short-term memory-fully connected (LSTM-FC) neural network for $PM_{2.5}$ concentration prediction. *Chemosphere*, 220, 486–492.

Zhao, L., *et al.*, 2020b. T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21 (9), 3848–3858.