



A hybrid deep learning model for regional O₃ and NO₂ concentrations prediction based on spatiotemporal dependencies in air quality monitoring network[☆]

Cui-lin Wu ^a, Hong-di He ^{a,*}, Rui-feng Song ^b, Xing-hang Zhu ^a, Zhong-ren Peng ^c, Qing-yan Fu ^d, Jun Pan ^d

^a Center for ITS and UAV Applications Research, State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

^b Department of Civil and Environmental Engineering, Northeastern University, 360 Huntington Ave., Boston, MA, 02115, USA

^c International Center for Adaptation Planning and Design, College of Design, Construction and Planning, University of Florida, PO Box 115706, Gainesville, FL, 32611-5706, USA

^d Shanghai Environmental Monitoring Center, Shanghai, 200235, China



ARTICLE INFO

Keywords:

Regional prediction
Air quality monitoring network
Spatiotemporal dependencies
Deep learning

ABSTRACT

Short-term prediction of urban air quality is critical to pollution management and public health. However, existing studies have failed to make full use of the spatiotemporal correlations or topological relationships among air quality monitoring networks (AQMN), and hence exhibit low precision in regional prediction tasks. With this consideration, we proposed a novel deep learning-based hybrid model of Res-GCN-BiLSTM combining the residual neural network (ResNet), graph convolutional network (GCN), and bidirectional long short-term memory (BiLSTM), for predicting short-term regional NO₂ and O₃ concentrations. Auto-correlation analysis and cluster analysis were first utilized to reveal the inherent temporal and spatial properties respectively. They demonstrated that there existed temporal daily periodicity and spatial similarity in AQMN. Then the identified spatiotemporal properties were sufficiently leveraged, and monitoring network topological information, as well as auxiliary pollutants and meteorology were also adaptively integrated into the model. The hourly observed data from 51 air quality monitoring stations and meteorological data in Shanghai were employed to evaluate it. Results show that the Res-GCN-BiLSTM model was better adapted to the pollutant characteristics and improved the prediction accuracy, with nearly 11% and 17% improvements in mean absolute error for NO₂ and O₃, respectively compared to the best performing baseline model. Among the three types of monitoring stations, traffic monitoring stations performed the best for O₃, but the worst for NO₂, mainly due to the impacts of intensive traffic emissions and the titration reaction. These findings illustrate that the hybrid architecture is more suitable for regional pollutant concentration.

1. Introduction

Rapid urbanization, industrialization, and motorization have dramatically elevated atmospheric pollution, which has arisen widespread public concern (Gao et al., 2017; Zhu et al., 2022). Especially, surface O₃ is of particular interest due to its acute detrimental effects on human health, ecosystems, and the environment (Monks et al., 2015). According to the report of epidemiological studies, short- and long-term exposure to higher concentrations of O₃ can increase the risk of

respiratory and cardiovascular morbidities and mortalities (Al-Hemoud et al., 2021). NO₂ is another air pollutant and one of the most important O₃ precursors. NO₂ can participate in photochemical reactions and contribute to the direct production of O₃, which is inextricably associated with O₃ levels (He & Lu, 2012). Although considerable efforts have been undertaken to reduce the emission of air pollutants, extreme air pollution episodes still frequently occurred in urban areas of China (Su et al., 2017). Among which Shanghai is the megacity with the most developed economy and the largest energy demand, which inevitably

[☆] This paper has been recommended for acceptance by Pavlos Kassomenos.

* Corresponding author.

E-mail address: hongdihe@sjtu.edu.cn (H.-d. He).

suffers from serious air pollution. Accurate prediction of NO₂ and O₃ can provide real-time air quality information to help the public make rational travel routes and help the government design and implement relevant pollution prevention policies (Wang et al., 2020a). Different from the prediction of air pollutant concentration at individual monitoring station, regional NO₂ and O₃ concentrations prediction implies higher spatial resolution, which is of great significance for the coordinated control and the refined management of urban air pollution (Wang et al., 2020b). Additionally, the high-resolution environmental data is helpful to enhance the digital governance capacity of ecological environment system and accelerate the realization of the urban digital development goal. Therefore, considering the citizens' yearning and demand for an ecological and healthy city, it is extremely necessary to conduct the regional NO₂ and O₃ concentrations prediction.

Nowadays, many regions have established real-time air quality monitoring networks (AQMN) and accumulated a large amount of rich monitoring data, which provides basic information for the accurate estimation of air pollutants. However, the prediction of pollutant concentration in a city remains challenging and strongly depends on the following three complex factors: (1) Temporal dependencies. The pollution levels at the current time point are affected by historical time intervals, both near and far. For example, the concentration of pollutant at 7:00 a.m. will affect that of 8:00 a.m. Additionally, since anthropogenic influences such as traffic emissions and meteorology exhibit cyclic patterns, pollutants' time series are also generally recorded in a daily periodicity, repeating every 24 h (2) Spatial dependencies. The concentrations of pollutant at some monitoring station are correlated with those of nearby stations as well as distant stations, according to the first law of geography (Zhu et al., 2018). (3) External factors. Some external factors, such as other pollutants, meteorological conditions, topographic features, and the height of buildings may affect the production, decomposition and dispersion of pollutants by physical processes and multiphase reactions (Song et al., 2021). Specially, some factors tend to interact with each other, which increases the difficulty of pollutant concentration prediction (Zhang et al., 2022). Hence, an efficient model is required to fully extract and learn time and space dependencies as well as external factors to provide an accurate estimation of regional air quality.

Generally, the approaches to model air quality can be classified into two types: knowledge-driven models and data-driven models. Knowledge-driven models incorporate physical and chemical mechanisms to simulate pollutant emission, transformation, diffusion, and transport processes, and are capable of effectively estimating air pollutant concentration (Zhao et al., 2019), such as Weather Research and Forecasting (WRF) models (Cai et al., 2016) and Nested Air Quality Prediction Modeling System (NAQPMS) (Wang et al., 2017). However, such models require prior knowledge and heavily rely on parameter settings that are difficult to obtain, and thereby are unfavorable for the analysis of air quality problems at microscales, such as in urban areas (Wang et al., 2020b). Consequently, data-driven models offer an alternative approach to realize the prediction analysis of the future concentration of pollution. Compared with knowledge-driven models, such models are purely based on the correlations between pollutant concentration data and dependent variables with no prior knowledge required. Data-driven models can be divided into three sub-categories: statistical models, shallow machine learning (SML) models, and deep learning (DL) models. Statistical models use the historical observed data to predict future values through time-series analysis. As air pollutants prediction is a kind of time series task, it is feasible to exploit statistical models to forecast the levels of air pollutant (Bhatti et al., 2021). Kumar and Jain utilized a model based on autoregressive integrated moving average (ARIMA) to forecast daily mean ambient air pollutants (O₃, CO, NO and NO₂) concentrations at an urban traffic site of Delhi, India and found that the model achieved satisfactory performance in the next step prediction (Kumar & Jain, 2010). Statistical model can well handle the linear features but are incapable of capturing the complicated non-linear

features. Additionally, it is difficult to sufficiently incorporate other external factors in the process of model construction. SMI models resolve the drawbacks of statistical models and begin to flourish in the field of air quality prediction, such as support vector machine (SVM) (Cabaneros et al., 2019) and random forests (RF) (Song et al., 2021). Song et al. applied RF model to predict PM_{2.5} and NO₂ concentrations in Shanghai considering transportation, meteorology, geographical conditions, land use, etc., which obtained better results than traditional land use regression model (Song et al., 2021). SMI models could cover different kinds of data as input but ignore neighboring factors in pollutant monitoring network (Wang et al., 2022). Additionally, their performance greatly depends on manually constructed features, which is challenging to process the complex and huge spatiotemporal pollutant data.

Compared with SML models, DL models can automatically capture crucial features and allow raw data as input to make end-to-end prediction. As a form of machine learning, DL architectures have presented state-of-the-art performances in general environmental prediction issues owing to flexible model structure, powerful non-linear mapping abilities, and strong generalization (Kim et al., 2021). Recurrent neural networks (RNNs) have great advantages in dealing with sequence learning tasks. Hence, RNNs (Septiawan & Endah, 2018) and their variants such as long short-term memory networks (LSTM) (Zhou et al., 2019) and bidirectional long short-term memory (BiLSTM) networks (Ma et al., 2020) are introduced to capture temporal dynamics in the pollutant sequence. However, RNN-based models fail to make full use of the spatial relationships in the monitoring network, which may limit the performance of the models coping with spatiotemporal data. Inspired by the potentials of convolutional neural networks (CNNs) in extracting spatial features, it has become prevailing to integrate CNNs and RNNs to perform air quality prediction (Le et al., 2020; Li et al., 2020; Wen et al., 2019; Yan et al., 2021). For example, Yan et al. combined a one-dimensional CNN and LSTM to propose a hybrid model, CNN-LSTM, which yielded better results than the CNN and the back propagation neural network (BPNN) (Yan et al., 2021). Le et al. proposed the convolutional long short-term memory (ConvLSTM) model, a combination of a two-dimensional CNN and LSTM to simultaneously manipulate temporal and spatial dependencies of the data in Seoul city of Korea (Le et al., 2020). The residual neural network (ResNet) is the latest successor of CNNs and allows longer structures to learn deep abstract spatial dependencies, and hence it has been attempted to model air pollutant concentration. Cheng et al. exploited LSTM in combination with the ResNet for PM_{2.5} forecasting in Beijing, which outperformed other models such as boosting algorithms or general recurrent neural networks (Cheng et al., 2022). Zhang et al. proposed a deep learning network model comprising the ResNet and the ConvLSTM, to deeply extract the temporal and spatial distribution features of pollutant concentration and meteorological data from multiple cities (Zhang et al., 2022). However, CNNs are designed for Euclidean data, which means that all historical data in AQMN must be preprocessed into regular data before applying such models (Zhang et al., 2020). This may result in the loss of irregular topological information in pollutant monitoring network. Thus, graph convolutional network (GCN) is proposed to address the drawbacks of CNNs, and processes graph-structure data to capture topological relationships among AQMN (Gao & Li, 2021; Wang et al., 2022). In light of this, we attempt to construct a hybrid deep learning model to take advantage of both the spatiotemporal features and topological information in pollutant monitoring network for further optimization of urban atmospheric forecasting.

In this study, we propose a hybrid deep learning-based model called Res-GCN-BiLSTM for the regional prediction of NO₂ and O₃ concentrations by integrating ResNet, GCN, and BiLSTM. Specially, ResNet is employed to adaptively learn deep spatial dependencies between monitoring stations, considering the temporal close and daily patterns of pollutants. GCN is applied to well capture the topological information of the entire monitoring network, while BiLSTM is utilized to effectively

extract temporal dependencies of external factors related to target pollutants including auxiliary pollutants and meteorological conditions. The hourly six air pollutants (i.e., NO₂, CO, PM_{2.5}, PM₁₀, SO₂, and O₃), geographical information, and meteorological data collected in Shanghai, China is used for training, validating and testing the proposed model. To do so, the auto-correlation analysis is first performed to pinpoint the periodical patterns of pollutants. Then the hierarchical clustering algorithm is applied to examine spatial similar behaviors in the pollutant monitoring network. Finally, the Res-GCN-BiLSTM model is constructed to predict the pollutants concentration at the next hour on a network scale based on the identified spatiotemporal characteristics. The refined forecast information is expected to help urban managers improve air quality and help citizens plan travel routes.

2. Materials and methods

2.1. Study area and data description

The data in this study mainly were supported by Shanghai Environmental Monitoring Center, and hourly air quality data (i.e., NO₂, CO, PM_{2.5}, PM₁₀, SO₂, and O₃) from 72 air quality monitoring stations in Shanghai from January 1, 2019 to December 30, 2019 were selected for research. Shanghai is a megacity located in east China, consisting of 17 districts. In this study, the research area of interest was composed of 16 districts, except Chongming District. This is because Chongming District is an independent isolated island in the estuary of the Yangtze River, with a great difference compared with the intro-urban areas. Hence, it was excluded from the scope of this study. Additionally, geographical information of each station and meteorological data (i.e., temperature, relative humidity, wind speed, and wind direction) during the same period were also introduced as auxiliary parameters as they have strong relations with pollutant levels (Chen et al., 2020). It is worth mentioning that most air quality monitoring stations are not equipped with meteorological stations nearby because air quality monitoring stations are located close to each other, resulting in similarity of hourly meteorological conditions. Thus, the data of the meteorological station closest to the Hongqiao Airport in Shanghai was selected to represent the meteorological conditions of the whole city. The specific information of the variables is shown in Appendix Table 1s.

To further understand the spatial distribution characteristics of air pollutants, monitoring stations across the city were divided into three categories according to geographical locations, namely suburban, traffic, and urban monitoring stations. The stations labelled by 0–24, 25 to 26, and 27 to 50 refer to suburban, traffic, and urban monitoring stations, respectively. Wherein traffic monitoring stations are located on the roadside to monitor the direct impact of road traffic sources on air pollution, while suburban and urban monitoring stations are set to monitor the background concentrations of pollutants in suburban and urban environments, respectively. The geographical location is shown in Fig. 1.

In the data process preprocessing, the monitoring stations located in Chongming District were eliminated due to the precondition of this study. Additionally, the missing data increases the uncertainty of the data, which makes it difficult to capture the temporal characteristics in the pollution data. Therefore, it is necessary to process it before modeling. The missing rate of data at some monitoring stations was as high as 35%, which could not reflect the local pollution levels. Hence, they are firstly discarded. Then, we used the linear interpolation imputation algorithm to fill the missing data as it can restore real data to a certain extent, and improve the model effect. To eliminate the dimensional impact of features and improve the run efficiency of the model, we used to max-min normalization method to make the data map within the same range (0,1). Finally, there were 446,760 records of hourly averages from 51 stations.

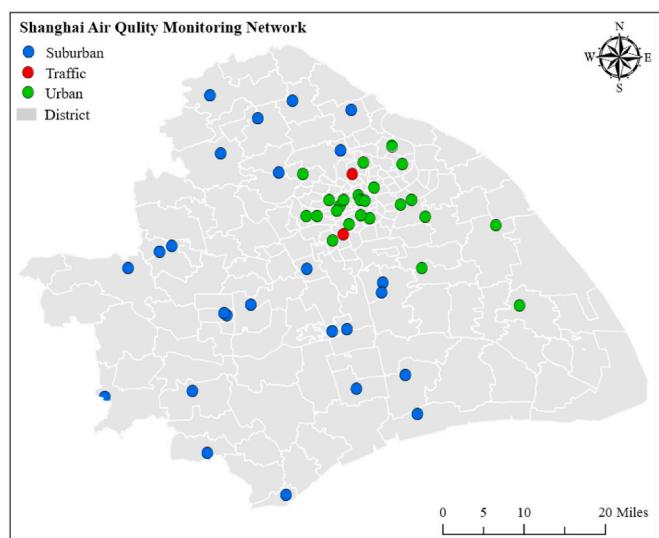


Fig. 1. Geographical locations of the air quality monitoring network in Shanghai (except Chongming District).

2.2. Res-GCN-BiLSTM

Regional pollutant concentration prediction is a typical spatiotemporal modeling problem. Hence, we propose a hybrid deep learning-based model of Res-GCN-BiLSTM, combining the ResNet, GCN, and BiLSTM, comprehensively considering the spatiotemporal, topological, and external influence. The Res-GCN-BiLSTM model comprises three components. Wherein, **Branch 1** utilized the ResNet model to deeply extract nearby and distant spatial dependencies between stations in the close and daily patterns, **Branch 2** applied the GCN model to well capture the topological information of the entire monitoring network, and **Branch 3** employed the BiLSTM model to adaptably learn the temporal correlations of external factors, such as auxiliary pollutants and meteorology, to further improve the accuracy of the model. Furthermore, the outputs of **Branches 1–3** were integrated and then input into a BiLSTM layer to extract the high-level features. Following the BiLSTM layer, the output was then flattened and input into a fully connected layer to obtain the output data. The model architecture is presented in Fig. 2.

2.2.1. ResNet

The ResNet model, a successful variant of CNN, is a powerful feature extraction tool without manual feature extraction, which can effectively learn the spatial dependency of Euclidean data. Compared with the original CNN, the ResNet model allows CNN to have a super deep structure to extract more enriched features and avoid the problem of vanishing gradients or exploding gradients (He et al., 2016; Zhang et al., 2017). Thus, we utilized the ResNet model to deal with historical pollutant concentration data. The pollutant data in monitoring network were treated as an image-like matrix to allow the execution convolution operation. Two temporal properties, namely close and daily patterns, were taken into account to extract temporal dependencies. Wherein the close pattern refers to the concentration of pollutant in the closest time intervals and the daily pattern represents the corresponding information in the same time intervals but in daily periodicities (Jinlei Zhang et al., 2020). If the target time interval is t, the close pattern refers to the time close to t, such as t-1, t-2, t-3, etc. The daily pattern refers to t every day for the previous days. The close and daily patterns can be described as follows.

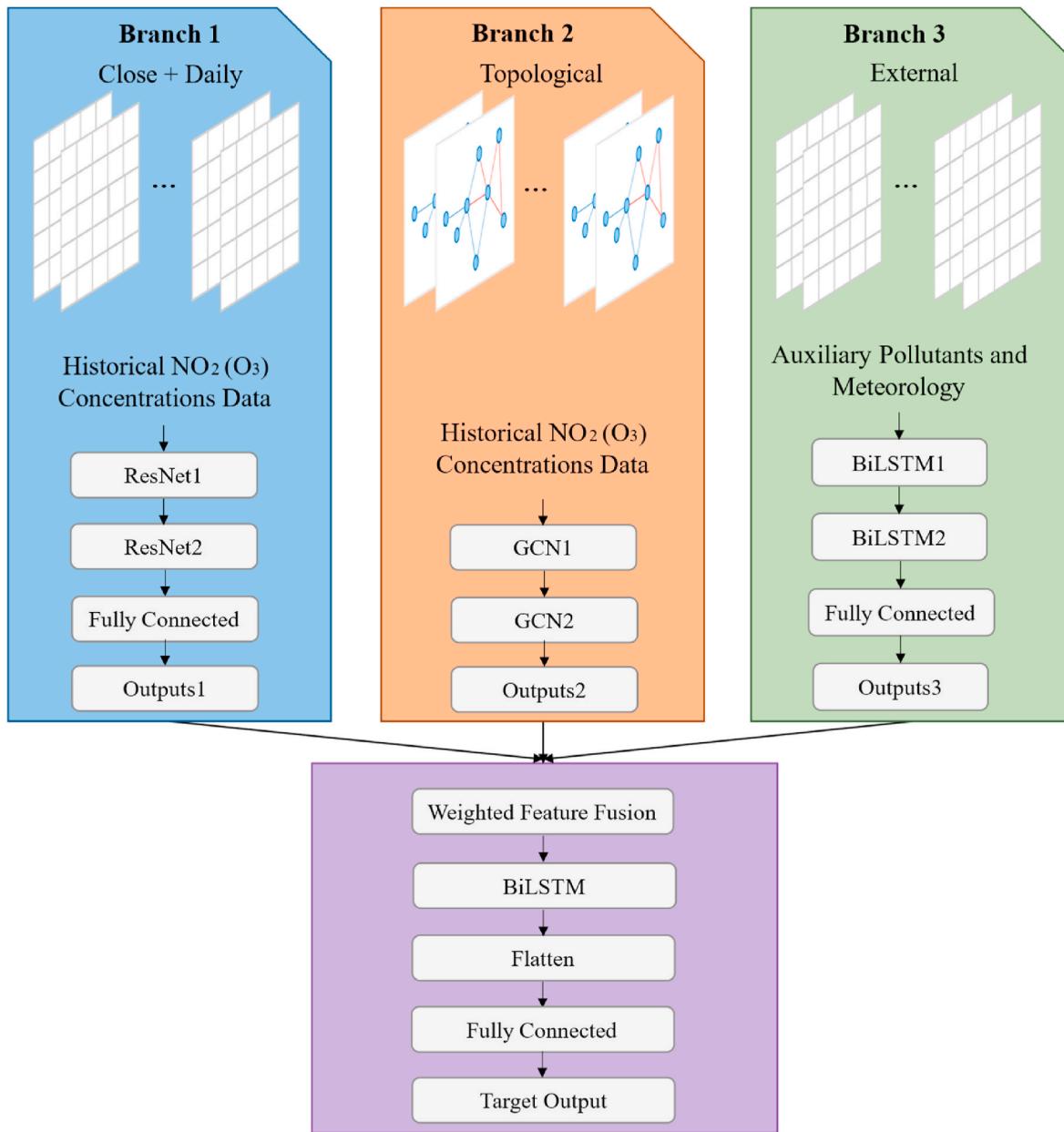


Fig. 2. The architecture of the hybrid Res-GCN-BiLSTM model.

$$X_{s,t}^C = \begin{pmatrix} x_{1,t-n} & x_{1,t-(n-1)} & \cdots & x_{1,t-1} \\ x_{2,t-n} & x_{2,t-(n-1)} & \cdots & x_{2,t-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{s,t-n} & x_{s,t-(n-1)} & \cdots & x_{s,t-1} \end{pmatrix} \quad (1)$$

$$X_{s,t}^D = \begin{pmatrix} x_{1,t-n+24} & x_{1,t-(n-1)*24} & \cdots & x_{1,t-24} \\ x_{2,t-n+24} & x_{2,t-(n-1)*24} & \cdots & x_{2,t-24} \\ \vdots & \vdots & \ddots & \vdots \\ x_{s,t-n+24} & x_{s,t-(n-1)*24} & \cdots & x_{s,t-24} \end{pmatrix} \quad (2)$$

where s is the number of monitoring stations, t is the target time, and n represents the historical time lags for each station. $X_{s,t}^C$ and $X_{s,t}^D$ represent pollutant time series corresponding to the same day or previous day, respectively.

In **Branch 1**, the two temporal patterns were integrated and then fed into a two stacked ResNet layers. Considering one ResNet layer that only

accounts for spatial near dependencies, and thus we need to construct network with many layers to capture the spatial dependencies of any station. However, deeper ResNets are not always better because of vanishing or exploding gradients. Hence, we finally determined the two-layer ResNet to deeply extract spatial dependency between nearby and distant stations. Following the ResNet layer, the output was input into one fully connected layer. Herein, the first ResNet layer had 32 filters; the second layer had 64 filters; the kernel size was 3*3; the fully connected layer consisted of 256 neurons. Through **Branch 1**, the high-level spatiotemporal information could be well captured.

2.2.2. GCN

GCN is one of Graph Neural Networks (GNN), which is suitable to process graph-structure data. GCN has achieved brilliant performance in air quality prediction tasks because of its powerful ability to capture the topological characters of non-Euclidean data (Ge et al., 2021; Qi et al., 2019). Thus, we applied GCN to predict pollutant concentration at a city level. To reduce the complex computation and mitigate gradient

vanishing issues during backpropagation (Zhang et al., 2021), a two-layer of GCN was employed to capture the topological relationships in pollutant monitoring network. The pollutant monitoring network at a certain time was abstracted into an undirected topology graph $G = (V, E, A)$, where V is the node, denoting monitoring stations, E is the set of edges representing the connectivity between stations, and A is the adjacent matrix representing the spatial topological relationships between stations. The equation of GCN can be formulated as follows.

$$X^{l+1} = \delta\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}X^lW^l + b\right) \quad (3)$$

$$\tilde{A} = A + I \quad (4)$$

where $A \in R^{n \times n}$ is an adjacent matrix based on the geographical distance between stations, $I \in R^{n \times n}$ is the identity matrix, $\tilde{D} \in R^{n \times n}$ is the diagonal mode degree matrix of \tilde{A} , $X^l \in R^{n \times m}$ is the feature matrix of the l th layer in which m represents the time steps used to predict the pollutants in the next time step, $W \in R^{m \times k}$ is the weight matrix in which k is the kernel number, $b \in R^{k \times 1}$ is the bias vector, and δ is the activation function.

Note that the adjacent matrix was constructed based on the geographical distance between two monitoring stations, which is defined as follows (Qi et al., 2019).

$$d_{ij} = d_{geo}((x_i, y_i), (x_j, y_j)) \quad (5)$$

$$A_{ij} = \begin{cases} \frac{1}{d_{ij}}, & i \neq j \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where (x_i, y_i) represents the latitude and longitude coordinates of station v .

In Branch 2, the historical pollutant concentration data were fed into a two stacked GCN layers and then input into one fully connected layer with 256 neurons. We designed the GCN with two layers because stacking multiple GCN layers not only leads to higher complexity during backpropagation but also to gradient vanishing, thereby degrading the performance of deeper GCNs. In this way, the topological relationships among the entire monitoring network could be accurately obtained.

2.2.3. BiLSTM

LSTM has been reported to achieve state-of-the-art results on the sequence learning tasks due to effectively extract the long-term temporal dependency of the time series. However, such a network only considers the previous information. To address the issue, BiLSTM was proposed to increase the amount of input information to the network by containing two LSTMs for both forward pass and backward pass (Mao et al., 2021). Thus, BiLSTM was employed to deal with the historical external factors, including auxiliary pollutants (i.e., NO₂/O₃, CO, PM_{2.5}, PM₁₀, and SO₂) and meteorology (i.e., temperature, humidity, wind direction, and wind speed), to obtain the forward and backward characteristics of external factors and further improve the prediction accuracy. The equation of BiLSTM can be formulated as follows (Wang et al., 2020).

$$\vec{h}_t = \delta\left(W_{x \rightarrow h}x_t + W_{\vec{h} \rightarrow \vec{h}}\vec{h}_{t-1} + b_{\vec{h}}\right) \quad (7)$$

$$\bar{h}_t = \delta\left(W_{\bar{x} \rightarrow \bar{h}}\bar{x}_t + W_{\bar{h} \rightarrow \bar{h}}\bar{h}_{t+1} + b_{\bar{h}}\right) \quad (8)$$

$$y_t = W_{\bar{h} \rightarrow y}\bar{h}_t + W_{\bar{h} \rightarrow \bar{h}}\bar{h}_t + b_y \quad (9)$$

where, x_t , \vec{h}_t , \bar{h}_t , and y_t respectively represent input vector, forward hidden states, backward hidden states, and predictions at time t . $W_{x \rightarrow \vec{h}}$, $W_{\bar{x} \rightarrow \bar{h}}$, $W_{\vec{h} \rightarrow \vec{h}}$, $W_{\bar{h} \rightarrow \bar{h}}$, and $W_{\bar{h} \rightarrow y}$ are the weight matrix, $b_{\vec{h}}$, $b_{\bar{h}}$, and b_y

are the bias vectors.

In Branch 3, we obtained the preprocessed input data as follows:

$$X_{w,t} = \begin{pmatrix} x_{1,t-n} & x_{1,t-(n-1)} & \cdots & x_{1,t-1} \\ x_{2,t-n} & x_{2,t-(n-1)} & \cdots & x_{2,t-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{w,t-n} & x_{w,t-(n-1)} & \cdots & x_{w,t-1} \end{pmatrix} \quad (10)$$

where w represents the 9 indicators used for auxiliary pollutants and meteorology.

Then, preprocessed input data were flattened and subsequently sent into the a two stacked BiLSTM layers. In this way, their impacts influencing the target pollutants could be analyzed. The number of hidden layers and neurons represent the complexity of the model, which are correlated with the learning ability of the model. Herein, two stacked BiLSTM layers consisted of 128 and 256 neurons, respectively; the fully connected layer had 256 neurons.

2.2.4. Feature Fusion

The parametric matrix-based method was applied to merge the above three outputs (Han et al., 2021). A BiLSTM layer with 64 neurons was conducted after feature fusion. Subsequently, the BiLSTM output was flattened and input into a fully connected layers with 256 neurons to generate the final outputs. The equation of fusion can be formulated as follows.

$$Fusion = \delta(W_1 \odot O_1 + W_2 \odot O_2 + W_3 \odot O_3 + b) \quad (11)$$

where $Fusion$ is the prediction target after fusion, \odot represents the Hadamard product, O_1 , O_2 , and O_3 are the outputs from the three parts, and W_1 , W_2 , and W_3 are the trainable weights matrix, b is the bias, δ is the activation function.

2.3. Experiments settings

The experiments were conducted on a server equipped with Ubuntu 16.04 Linux System with 128 GB RAM and one NVIDIA Titan RTX Graphics Cards (24 GB GDDR5 VRAM). Python (version 3.6) with Numpy, Pandas, GeoPandas, Tensorflow, and Keras packages were used for data preprocessing and modeling. In the experiments, the dataset was sequentially partitioned into three sets according to the time series, of which 70% for training, 10% for validating, and the remaining 20% for testing. The training set was used to model fitting; the validation set was used to adjust the hyper-parameters; the test set was used to assess the ultimate universality of the model. Mean Absolute Error (MAE) was adopted as the loss function and the Adam optimization algorithm was used as an optimizer to update the parameters. The EarlyStopping mechanism was implemented to save the computational costs and mitigate overfitting. In our experiments, the number of time intervals for close and daily patterns were set as $n = 6$; the initial learning rate was set to 0.001; the weight decay rate was set to 5e-4; the epoch was 200; the patience of the early stopping mechanism was set to 50; the batch size was 64; the activation function was Rectified Linear Unit (ReLU). The details of the experimental setup are shown in Appendix Table 2s.

MAE, Root Mean Square Error (RMSE), Pearson Correlation Coefficient (r), and Coefficient of Determination (R^2) were utilized to evaluate the air pollutants forecasting performance. Higher r , R^2 and lower MAE, RMSE values are associated with better performance of the model. Four performance indexes are defined as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |O_i - P_i| \quad (12)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - P_i)^2} \quad (13)$$

$$r = \frac{\sum_{i=1}^N (O_i - \bar{O}_i)(P_i - \bar{P}_i)}{\sqrt{\sum_{i=1}^N (O_i - \bar{O}_i)^2} \sqrt{\sum_{i=1}^N (P_i - \bar{P}_i)^2}} \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2} \quad (15)$$

where N denotes the number of samples, O_i and P_i denote the observed value, and the predicted value of the i th sample, separately.

3. Results and discussions

3.1. Identification of the temporal periodicity in NO_2 and O_3 monitoring networks

To verify whether the temporal variations of NO_2 and O_3 concentrations exhibit the daily cycle patterns, the analysis of the diurnal variations was first utilized to preliminarily observe and the results are shown in Fig. 3a and b, using the hourly averaged concentrations over the same period in 2019 at the different stations. Generally, similar patterns of the diurnal cycle were observed in three categories of monitoring stations for each gas species, while the magnitudes of variations were different. NO_2 concentrations exhibited two distinct peaks in the diurnal cycle, and the two daily peaks separately occurred around 8:00 and 20:00, coinciding with peak-traffic hours reported in Shanghai. In contrast, the daily variation of O_3 was opposite to that of NO_2 , indicating strong photochemical conversions from NO_2 to O_3 ($\text{NO}_2 + \text{O}_2 + h\nu \rightarrow \text{NO} + \text{O}_3$). The trend of diurnal cycle revealed in this study is also

consistent with previous studies (Wu et al., 2021; Wu et al., 2022). Regarding the spatial variations, NO_2 levels at traffic monitoring stations ($45\text{--}65 \mu\text{g}/\text{m}^3$) were notably higher than those of the other two monitoring stations ($25\text{--}38 \mu\text{g}/\text{m}^3$), which is expected due to the impacts of automobile exhaust emissions on NO_2 levels. O_3 concentrations were obviously higher at suburban and urban monitoring stations ($38\text{--}65 \mu\text{g}/\text{m}^3$) than those of traffic monitoring stations ($22\text{--}42 \mu\text{g}/\text{m}^3$), while the diurnal O_3 peak was delayed by about 1 h later at suburban and urban monitoring stations. When NO pollutants are sufficient in the atmosphere, O_3 can be depleted through the titration reaction ($\text{NO} + \text{O}_3 \rightarrow \text{NO}_2 + \text{O}_2$). Consequently, the high concentrations of NO at traffic monitoring stations resulted in the low concentrations of O_3 in that region.

Then, the auto-correlation analysis (ACF), a fine-grained periodicity detector, was performed to pinpoint the periodical pattern in air pollutants and the results are depicted in Fig. 3c and d. The autocorrelation coefficients of the two pollutants reached the peak when the historical time lag was 24 h, which demonstrated that there exists a clear 24-h periodicity in both NO_2 and O_3 time series at each monitoring station. This finding suggests that the urban atmospheric forecasting models would be expected to be further optimized if such daily periodical characteristic was fully taken into account. Additionally, it is worth noting that the daily periodicities of NO_2 and O_3 at suburban and urban monitoring stations were more significant than that at traffic monitoring stations, with the corresponding autocorrelation coefficients of 0.61 and 0.71 at suburban monitoring stations, 0.57 and 0.65 at urban monitoring stations, and 0.58 and 0.59 at traffic monitoring stations respectively. One possible explanation is that pollutant participates in complex

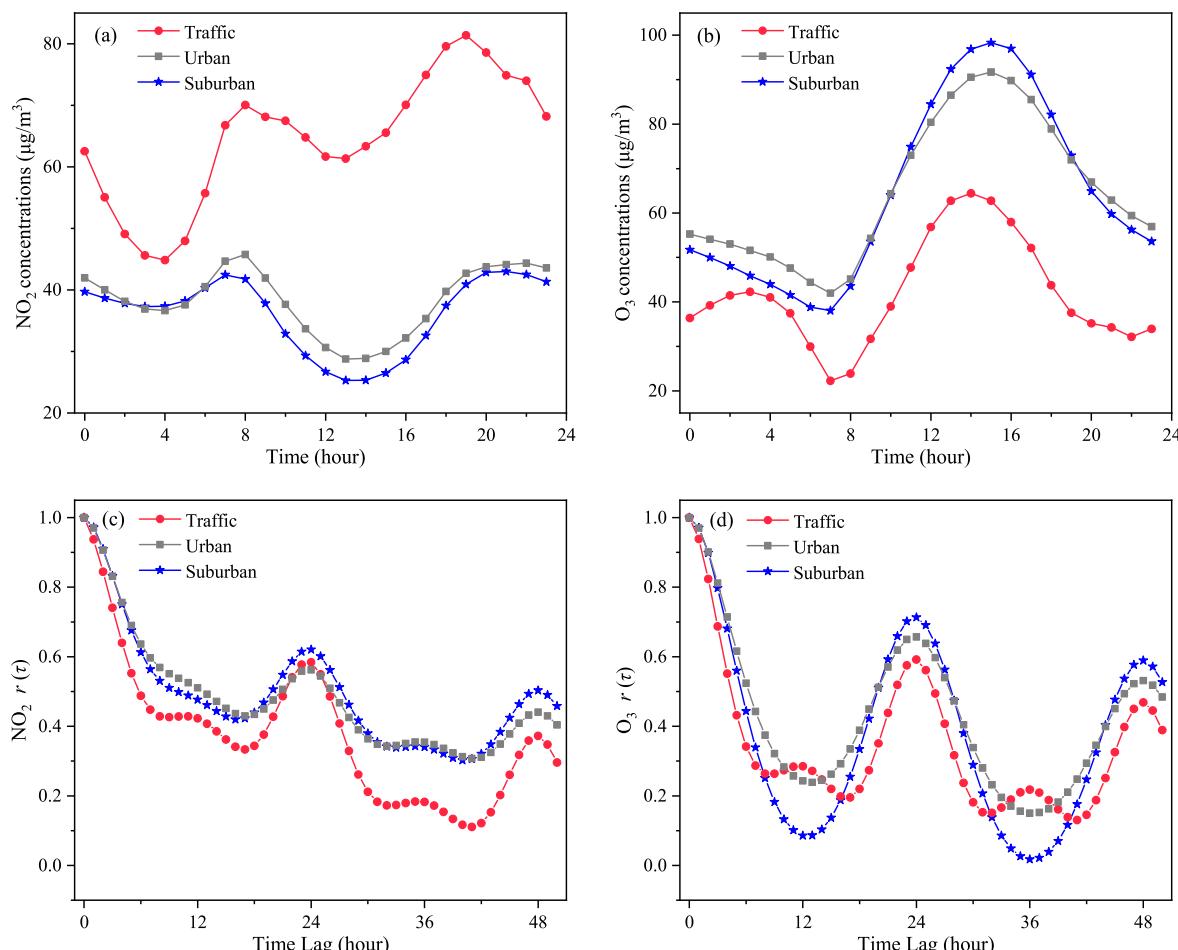


Fig. 3. Temporal variations analysis of pollutants in three categories of monitoring stations. (a)–(b) Diurnal variations of NO_2 and O_3 . (c)–(d) Autocorrelation function (with 95% confidence intervals) of NO_2 and O_3 .

chemical reactions at traffic monitoring stations, which makes the linear correlation between pollutant and its historical time series weak, and thus it exhibits less distinct daily periodic characteristic in this area.

3.2. Identification of the spatial similarity in NO_2 and O_3 monitoring networks

Apart from temporal periodical properties, previous studies also found there exists large spatial similarity between monitoring stations (He et al., 2018). To reveal potential spatial correlations in NO_2 and O_3 monitoring networks, the hierarchical clustering based on Ward's minimum deviation square sum method was conducted and the results are displayed in Fig. 4. For a better illustration, the NO_2 result was selected as an example for analysis. According to the hierarchical clustering distance value, the 51 monitoring stations were classified into three main categories, i.e., A, B, and C (Fig. 4a). The lower the hierarchical clustering distance value within the category, the more similar behaviors the category presents. Wherein the stations in categories A are mainly situated in remote suburban areas with less traffic, while the stations in category B and C are mainly located in urban areas with congested traffic and presented common behavior. It can be concluded that these grouped categories are evidently depended on the station characteristics. Similar classification can also be observed in the O_3 monitoring network (Fig. 4b). The spatial similarity patterns revealed in NO_2 and O_3 monitoring networks give a rise to estimating the pollutants concentrations using the surrounding stations instead of measurements in the future, and provide a scientific basis for regional joint prevention and control of refined urban atmosphere control.

3.3. Predictions of NO_2 and O_3 based on the spatiotemporal dependencies in monitoring networks

3.3.1. Comparison of models

We employed the Res-GCN-BiLSTM model for the regional NO_2 and O_3 concentrations prediction. To evaluate the effectiveness of the proposed model, several popular deep learning models were selected for comparison, including gate recurrent unit (GRU) (Wang et al., 2019), LSTM, BiLSTM, CNN, ResNet, ConvLSTM, and ResNet-LSTM (Cheng et al., 2022). Wherein GRU, LSTM, and BiLSTM are commonly used to handle sequence prediction tasks. CNN and ResNet are powerful feature extractor for spatial grid structure information. ConvLSTM is a well-known hybrid model based on the integration of two deep learning models, namely a two-dimension CNN and LSTM, to capture both temporal and spatial dependencies. ResNet-LSTM is another hybrid deep-learning-based model that combines a ResNet and LSTM to learn abstract spatial features and the temporal features, respectively. All

these models used the close and daily patterns as inputs to ensure the fairness of the comparison. Similar to the hybrid Res-GCN-BiLSTM, these baselines models were separately trained and tuned hyperparameters to obtain the optimal predictive performance. The MAE, RMSE, and r metrics between observed and predicted values were utilized to provide the accuracy of the forecasting model.

Table 3s (see Appendix) presents the predictive results of the Res-GCN-BiLSTM model and baseline models on the NO_2 and O_3 test datasets. Generally, the Res-GCN-BiLSTM model achieved the best performance for both NO_2 and O_3 , with the lowest prediction errors in MAE and RMSE and a best score of r . Take the NO_2 result as an example: compared with the temporal models, like GRU, LSTM, and BiLSTM, the hybrid Res-GCN-BiLSTM obtained a 47%–54% reduction in MAE and an 18%–22% relative improvement in r ; compared with the spatial models, like CNN and ResNet, the hybrid Res-GCN-BiLSTM exhibited a MAE reduction of 35%–43% and a r improvement of 11%–14%. The reason why the proposed Res-GCN-BiLSTM outperformed these single models (i.e., temporal or spatial models) is that either temporal models or spatial models only capture limited temporal or spatial information. However, the proposed Res-GCN-BiLSTM combined the advantage of the hybrid model for capturing spatiotemporal dependencies simultaneously, and thus exhibited better pollutants estimates in terms of the evaluation metrics. Additionally, one interesting finding is that models concerning spatial dependency performed better than the models concerning temporal dependency. Particularly, compared with the optimal BiLSTM model in the sequence model, ResNet exhibited a reduction of 18% in MAE and a reduction of 16% in RMSE in terms of NO_2 and had a 17% relative reduction in MAE and 10% relative reduction in RMSE for O_3 . This finding indicates that it is much easier to learn spatial information than temporal information for regional air quality prediction.

In comparison to the raster-based hybrid models, such as ConvLSTM and ResNet-LSTM, we found that the hybrid Res-GCN-BiLSTM performed a little better. Specially, compared with the best performing model (ResNet-LSTM), the MAE and RMSE of NO_2 by Res-GCN-BiLSTM decreased by 11% and 7% respectively.; MAE for O_3 decreased by 17% and RMSE by 11%. Such performance is probably due to the GCN component in the hybrid Res-GCN-BiLSTM, which captures the irregular topological information in pollutant monitoring network to compensates for the raster-based models' problem. Conclusively, by integrating ResNet, GCN, and BiLSTM, the hybrid Res-GCN-BiLSTM can efficiently capture the spatiotemporal dependencies better than several existing models, and can be well applied to regional air quality prediction.

To quantify the effects of external factors on the predictive performance, different inputs experiments were conducted for comparison, and the results are displayed in Fig. 5. Among the different inputs, the complete Res-GCN-BiLSTM performed best. Such performance is mainly

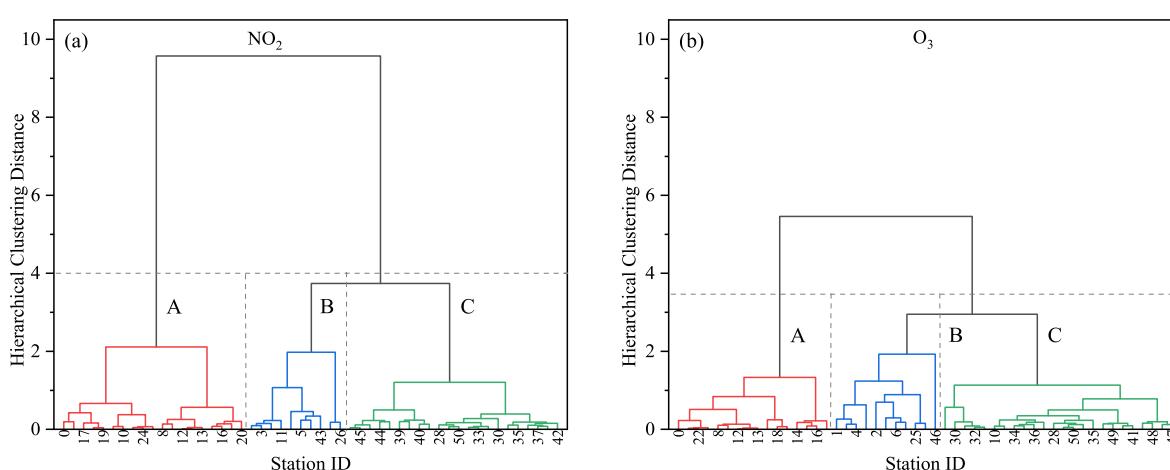


Fig. 4. Dendograms for NO_2 and O_3 .

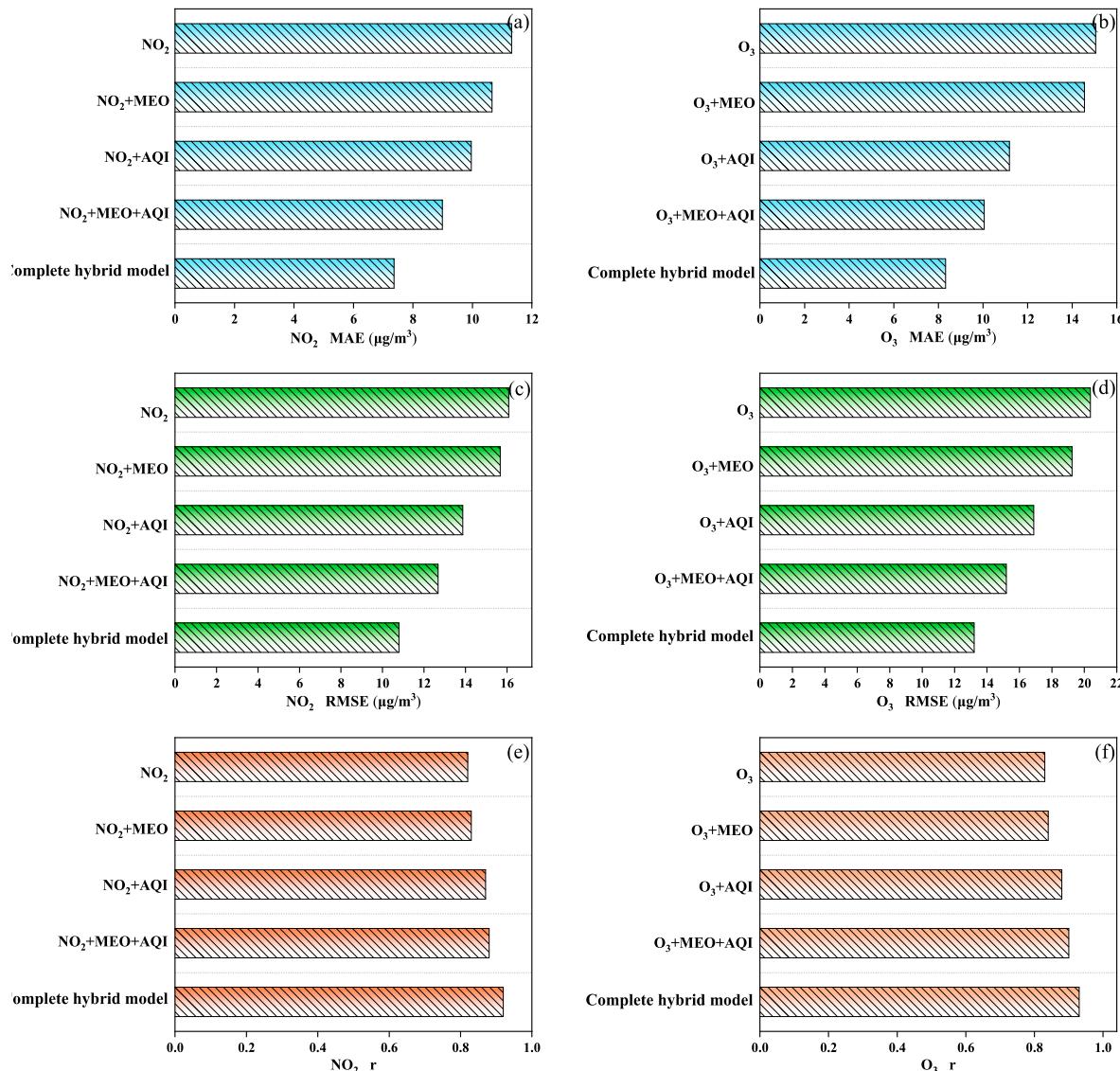


Fig. 5. Inputs variables comparison in the pollutant predictions. (a) NO₂, MAE, (b) O₃, MAE, (c) NO₂, RMSE, (d) O₃, RMSE, (e) NO₂, r, (f) O₃, r. Note: NO₂: input variable groups are only historical target pollutant data. NO₂ + MEO: input variable groups are historical target pollutant data and meteorology. NO₂ + AQI: input variable groups are historical target pollutant data and auxiliary pollutants, such as CO, PM_{2.5}, PM₁₀, SO₂, and O₃. NO₂ + MEO + AQI: input variable groups are historical target pollutant data, meteorology, and auxiliary pollutants. NO₂ + MEO + AQI + Graph: input variable groups are historical target pollutant data, meteorology, auxiliary pollutants, and topology structure.

due to four factors: (1) Prior knowledge about temporal properties, namely close and daily patterns, were taken into account. (2) The nearby and distant spatial dependencies between stations were fully extracted. (3) The introduction of network topology boosted the predictive performance. (4) The inclusion of auxiliary pollutants and meteorological information improved model accuracy further. For the topological information, the predictive performances decreased obviously after removing the graph structure from the complete model. In particular, the predictive error MAE increased by 1.63 $\mu\text{g}/\text{m}^3$ in NO₂ and 1.72 $\mu\text{g}/\text{m}^3$ in O₃, respectively when the graph structure was excluded into the model input. This result suggests that the topological information has a strong influence on the regional NO₂ and O₃ concentrations forecasting, which need to be taken into account in future air quality forecasts. For auxiliary pollutants (AQI) and meteorological conditions (MEO), compared with the input only considering the historical target pollutant, the introduction of MEO and AQI variables enhanced the predictive accuracy. Especially for the AQI variables, the predictive error MAE decreased by 1.36 $\mu\text{g}/\text{m}^3$ in NO₂ and 3.87 $\mu\text{g}/\text{m}^3$ in O₃, respectively when the AQI variables were added into the model. This result may be

due to the dominant impact of similar sources of emissions, and thus the influence of auxiliary pollutants on target pollutants is greater than that of meteorological variables. In sum, external factors further improved the accuracy of regional air quality prediction.

3.3.2. Comparison of NO₂ and O₃ predictions at different categories of monitoring stations using the Res-GCN-BiLSTM model

To evaluate the spatial predictive capabilities of the Res-GCN-BiLSTM model, suburban, traffic, and urban monitoring stations were selected for comparison. As illustrated in Fig. 6, in general, there was a good fitness (0.78–0.88 in R²) between ground observational data and predicted values for short-term NO₂ and O₃ forecasts in three types of monitoring stations, indicating that the proposed Res-GCN-BiLSTM presents strong robustness. In terms of NO₂, among three types of monitoring stations, suburban monitoring stations presented the lowest prediction error (10.60 in RMSE and 0.85 in R²) (Fig. 6a), followed by urban monitoring stations (10.81 in RMSE, 0.84 in R²) (Fig. 6c), and traffic monitoring stations exhibited the largest error (12.61 in RMSE, 0.78 in R²) (Fig. 6e). Considering NO₂ is a traffic-related pollutant and

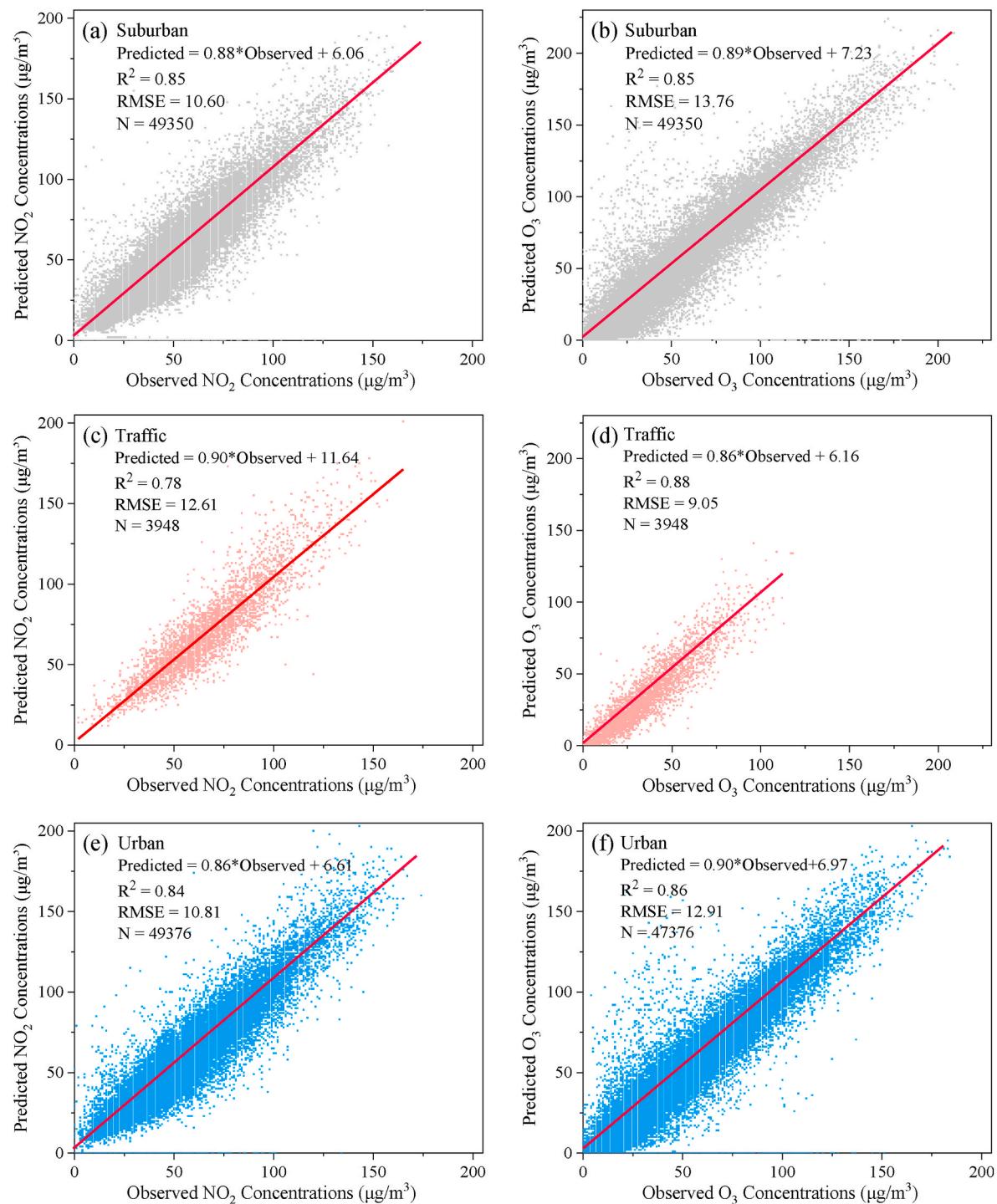


Fig. 6. Predicted and observed NO_2 and O_3 concentrations of the Res-GCN-BiLSTM model in the suburban, traffic, and urban monitoring stations. (a) NO_2 , suburban, (b) O_3 , suburban, (c) NO_2 , traffic, (d) O_3 , traffic, (e) NO_2 , urban, (f) O_3 , urban.

its levels are heavily influenced by local traffic emissions. This result could be attributed to the fact that suburban monitoring stations are located in remote areas of Shanghai, where the diurnal cycle patterns of NO_2 concentrations exhibit relatively small amplitude ($25\text{--}38 \mu\text{g}/\text{m}^3$) due to the impacts of less traffic emissions. This phenomenon is conducive to the DL models to capture the characteristics of NO_2 and to obtain better performance. In contrast, traffic monitoring stations are located in the downtown area of Shanghai, where NO_2 concentrations present stronger diurnal variations ($45\text{--}65 \mu\text{g}/\text{m}^3$) mainly due to the effects of more intensive traffic emissions. Additionally, complex

photochemical reactions also make its diurnal variation patterns complicated. Therefore, the complex temporal patterns at such stations are quite challenging to predict.

In terms of O_3 , the opposite trend to NO_2 occurred. Among the three categories of monitoring stations, traffic monitoring stations performed the best, with the RMSE and R^2 values of 9.05 and 0.88, respectively (Fig. 6d), followed by urban monitoring stations, with the RMSE and R^2 values of 12.91 and 0.86, respectively (Fig. 6f), and suburban monitoring stations performed the worst, with the RMSE and R^2 values of 13.76 and 0.85, respectively (Fig. 6b). This is probably caused by the

fact that O_3 concentrations exhibit much smaller diurnal variations ($22\text{--}42 \mu\text{g}/\text{m}^3$) at traffic monitoring stations, where the key precursors NO prevail and some of which consume O_3 within the titration reaction. Meanwhile, O_3 concentrations at such stations were less than $125 \mu\text{g}/\text{m}^3$ (Fig. 6d), indicating that extreme O_3 pollution episodes rarely occur. This phenomenon benefits the models to learn the pattern of O_3 generating and dispersing processes, and to achieve better performance (Wang et al., 2020a). In comparison to traffic monitoring stations, suburban monitoring stations are located in remote areas, where O_3 concentrations are influenced by both local traffic and regional transport, and thus the diurnal variation patterns of O_3 concentrations presented relatively larger fluctuations ($38\text{--}65 \mu\text{g}/\text{m}^3$). This pattern is difficult for air quality prediction models to capture and thus leads to larger O_3 forecast errors at such stations. These results imply that monitoring stations located in area with heavy traffic pollution tend to achieve worse performance in NO_2 prediction but better performance in O_3 prediction.

3.3.3. Spatial distributions of NO_2 and O_3 predictions using Res-GCN-BiLSTM

For a more intuitive understanding of the spatial differences, the spatial distributions of observations and RMSE in different geographical locations on the test set were compared, using kriging interpolation (Fig. 7). In terms of observations, the spatial distribution of NO_2 concentrations exhibited the higher in the west and lower in the east, and higher in inland areas than in coastal areas, with the highest value in the downtown area (Fig. 7a), demonstrating that NO_2 pollution in Shanghai could be predominantly determined by local traffic emissions. While spatial distribution pattern of O_3 presented the opposite of that of NO_2 , and the highest levels of O_3 were concentrated in the eastern region (Fig. 7b). This phenomenon is probably since O_3 values are significantly

negatively correlated with its key precursor NO_2 . Hence, the high concentrations of NO_2 in a region generally coincide with the low concentrations of O_3 in that region and vice-versa.

In terms of RMSE, satisfactory prediction results were obtained for the two target pollutants monitoring networks in the Res-GCN-BiLSTM model, which verifies that the proposed model has a certain extrapolation capacity. However, subtle differences emerged. The higher RMSE values of NO_2 were concentrated in Shanghai downtown (Fig. 7c), which is mainly caused by the intensive emission of motor vehicles in this area, and leads to the large amplitude of NO_2 diurnal variation. By comparison, the larger prediction errors in O_3 occurred in the eastern regions (Fig. 7d), which is probably since the observed O_3 concentrations in such regions were inherently high. It is also acknowledged that there exist cases of underestimation in severe pollution forecasting, which reflects the complexity of NO_2 and O_3 concentrations forecasting in the monitoring network proposed model. Extreme pollution prediction will be one of the most important research directions in the future studies.

4. Conclusions

In this study, a novel hybrid deep learning-based model, called Res-GCN-BiLSTM, was proposed to predict the next hour's NO_2 and O_3 concentrations in air quality monitoring networks. The Res-GCN-BiLSTM model is consist of the ResNet, GCN, and bidirectional LSTM (BiLSTM), comprehensively leveraging spatiotemporal features and monitoring network topological information, as well as auxiliary pollutants and meteorology. Several meaningful findings are summarized as follows.

- (1) The variations of NO_2 and O_3 in the three types of monitoring stations exhibited daily periodical patterns, and the periodicity

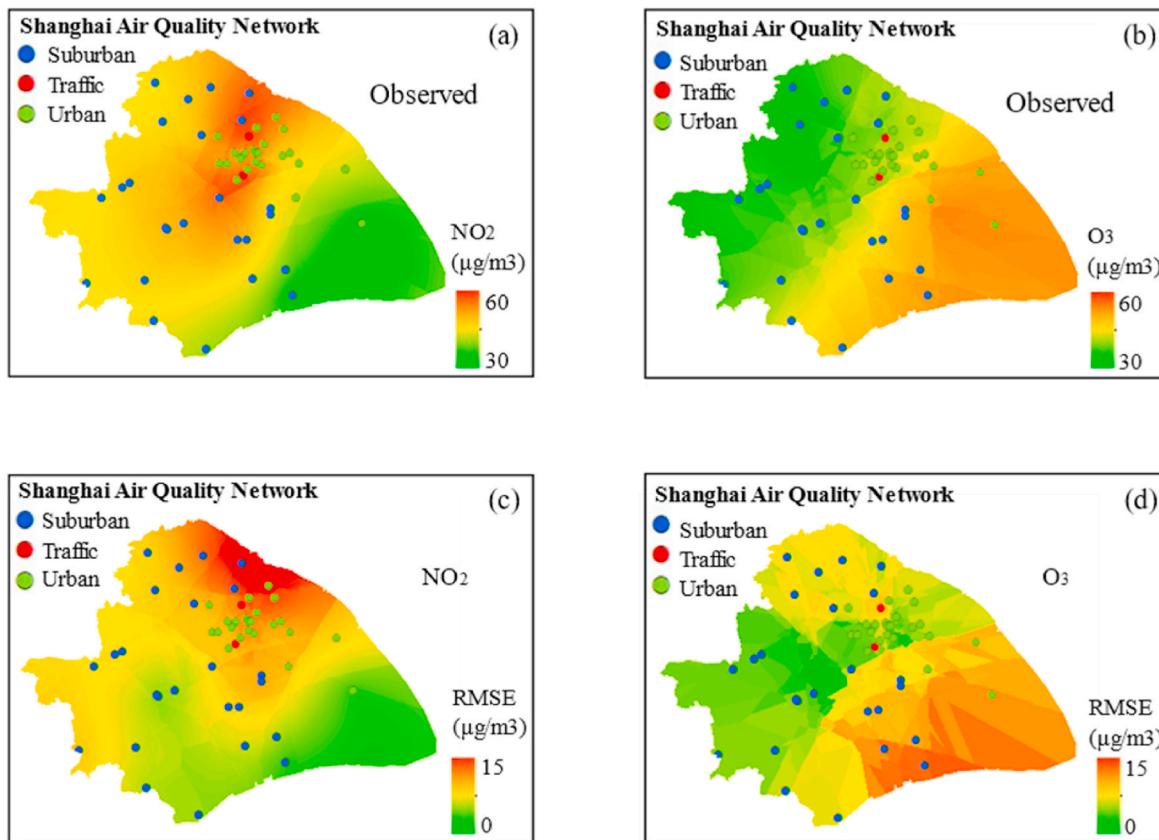


Fig. 7. Spatial distributions of observed mean pollutants concentrations and mean RMSE of next 1-h predictions based on the Res-GCN-BiLSTM model in the monitoring network. (a) NO_2 , observed, (b) O_3 , observed, (c) NO_2 , RMSE, (d) O_3 , RMSE.

- characteristics at suburban and urban monitoring stations were more notable than those at traffic monitoring stations.
- (2) The cluster analysis results confirmed the similar behavior of spatial variations caused by the type of stations in NO₂ and O₃ monitoring networks. It gives a rise to a possible of estimating the pollutants concentrations using the surrounding stations instead of measurements in the future.
 - (3) Res-GCN-BiLSTM model showed a higher forecasting accuracy and robustness of NO₂ and O₃ concentrations in contrast to 7 baseline models (i.e., GRU, LSTM, BiLSTM, CNN, ResNet, ConvLSTM, and ResNet-LSTM). In terms of MAE, the significant improvements were nearly 11% and 17% respectively for NO₂ and O₃ compared with the best performing baseline model (ResNet-LSTM).
 - (4) Monitoring network topological information were proven to have a recognizable influence on the promotion of prediction ability, with approximately 17% reduction after removing such structure from the complete model.
 - (5) Traffic monitoring stations achieved the best performance (9.05 in RMSE and 0.88 in R²), followed by urban monitoring stations (12.91 in RMSE and 0.86 in R²), and suburban monitoring stations performed the worst (13.76 in RMSE and 0.85 in R²) in terms of O₃ prediction, while the trend is opposite for NO₂.

In summary, this study considered the spatiotemporal properties in pollutant monitoring network and proposed a novel deep learning-based predictor to conduct air pollutants forecasts. It provides different insights into short-term spatiotemporal prediction of air quality in the monitoring network. Although the results are only referred to the Shanghai NO₂ and O₃ monitoring networks, there is no doubt the analysis procedure can also be extended to other cities. One limitation of this study is that the Res-GCN-BiLSTM model only considers the short-term forecasts of pollutants. The long-term forecasting capabilities of the proposed model will be tested in the future.

Credit author statement

Cui-Lin Wu: Conceptualization, Methodology, Visualization. **Hong-Di He:** Conceptualization, Writing – review & editing. **Rui-Feng Song, Zhong-Ren Peng:** Formal analysis, Writing-review & editing. **Qing-yan Fu, Jun Pan:** Data collection and analysis. All authors have read and agreed to the published version of the manuscript

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.

Acknowledgments

This study was partially funded by the National Natural Science Foundation of China (No. 12072195).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envpol.2023.121075>.

References

- Al-Hemoud, A., Gasana, J., Alajeel, A., Alhamoud, E., Al-Shatti, A., Al-Khayat, A., 2021. Ambient exposure of O₃ and NO₂ and associated health risk in Kuwait. Environ. Sci. Pollut. Control Ser. 28, 14917–14926. <https://doi.org/10.1007/s11356-020-11481-w>.
- Bhatti, U.A., Yan, Y., Zhou, M., Ali, S., Hussain, A., Qingsong, H., Yu, Z., Yuan, L., 2021. Time series analysis and forecasting of air pollution particulate matter (PM2.5): an SARIMA and factor analysis approach. IEEE Access. 9, 41019–41031. <https://doi.org/10.1109/ACCESS.2021.3060744>.
- Cabaneros, S.M., Calautit, J.K., Hughes, B.R., 2019. A review of artificial neural network models for ambient air pollution prediction. Environ. Model. Software. 119, 285–304. <https://doi.org/10.1016/j.envsoft.2019.06.014>.
- Cai, C.J., Zhang, X., Wang, K., Zhang, Y., Wang, L.T., Zhang, Q., Duan, F.K., He, K.B., Yu, S.C., 2016. Incorporation of new particle formation and early growth treatments into WRF/Chem: model improvement, evaluation, and impacts of anthropogenic aerosols over East Asia. Atmos. Environ. 124, 262–284. <https://doi.org/10.1016/j.atmosenv.2015.05.046>.
- Chen, Z.Y., Li, R.Y., Chen, D.L., Zhuang, Y., Gao, B.B., Yang, L., Li, M.C., 2020. Understanding the causal influence of major meteorological factors on ground ozone concentrations across China. J. Clean. Prod. 242, 118498. <https://doi.org/10.1016/j.jclepro.2019.118498>.
- Cheng, X.W., Zhang, W.W., Wenzel, A., Chen, J., 2022. Stacked ResNet-LSTM and CORAL model for multi-site air quality prediction. Neural Comput. Appl. 34, 13849–13866. <https://doi.org/10.1007/s00521-022-07175-8>.
- Gao, J.H., Woodward, A., Vardoulakis, S., Kovats, S., Wilkinson, P., Li, L., Xu, L., Li, J., Yang, J., Li, J., Cao, L.N., Liu, X.B., Wu, H.X., Liu, Q.Y., 2017. Haze, public health and mitigation measures in China: a review of the current evidence for further policy response. Sci. Total Environ. 578 (155), 148–157. <https://doi.org/10.1016/j.scitotenv.2016.10.231>.
- Gao, X., Li, W.D., 2021. A graph-based LSTM model for PM2.5 forecasting. Atmos. Pollut. Res. 12 (9), 101150. <https://doi.org/10.1016/j.apr.2021.101150>.
- Ge, L., Wu, K.Y., Zeng, Y., Chang, F., Wang, Y.Q., Li, S.Y., 2021. Multi-scale spatiotemporal graph convolution network for air quality prediction. Appl. Intell. 51 (6), 3491–3505. <https://doi.org/10.1007/s10489-020-02054-y>.
- Han, Y., Peng, T.X., Wang, C., Zhang, Z.H., Chen, G., 2021. A hybrid glm model for predicting citywide spatio-temporal metro passenger flow. ISPRS Int. J. Geo-Inf. 10, 222 <https://doi.org/10.3390/ijgi10040222>.
- He, H.D., Li, M., Wang Wang, Z.Y., W.L., Xue, Y., 2018. Prediction of PM2.5 concentration based on the similarity in air quality monitoring network. Build. Environ. 137, 11–17. <https://doi.org/10.1016/j.buildenv.2018.03.058>.
- He, H.D., Lu, W.Z., 2012. Decomposition of pollution contributors to urban ozone levels concerning regional and local scales. Build. Environ. 49 (1), 97–103. <https://doi.org/10.1016/j.buildenv.2011.09.019>.
- He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J., 2016. Deep residual learning for image recognition. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- Kim, J., Wang, X.L., Kang, C.Y., Yu, J.W., Li, P.H., 2021. Forecasting air pollutant concentration using a novel spatiotemporal deep learning model based on clustering, feature selection and empirical wavelet transform. Sci. Total Environ. 801, 149654 <https://doi.org/10.1016/j.scitotenv.2021.149654>.
- Kumar, U., Jain, V.K., 2010. ARIMA Forecasting of Ambient Air Pollutants (O₃, NO, NO₂ and CO). Stoch Environ Res Risk Assess 24, 751–760. <https://doi.org/10.1007/s00477-009-0361-8>.
- Le, V.D., Bui, T.C., Cha, S.K., 2020. Spatiotemporal deep learning model for citywide air pollution interpolation and prediction. Proceedings - 2020 IEEE International Conference on Big Data and Smart Computing, BigComp 2020. 55–62. doi:10.1109/BigComp48618.2020.9000009.
- Li, T., Hua, M., Wu, X.U., 2020. A hybrid CNN-LSTM model for forecasting Particulate Matter (PM2.5). IEEE Access. 26933–26940.
- Ma, J., Li, Z., Cheng, J.C.P., Ding, Y.X., Lin, C.Q., Xu, Z.R., 2020. Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. Sci. Total Environ. 705, 135771 <https://doi.org/10.1016/j.scitotenv.2019.135771>.
- Mao, W.J., Wang, W.L., Jiao, L.M., Zhao, S.L., Liu, A.B., 2021. Modeling air quality prediction using a deep learning approach: method optimization and evaluation. Sustain. Cities Soc. 6, 102567. <https://doi.org/10.1016/j.scs.2020.102567>.
- Monks, P.S., Archibald, A.T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K.S., Mills, G.E., Stevenson, D.S., Tarasova, O., Thouret, V., Von Schneidemesser, E., Sommariva, R., Wild, O., Williams, M.L., 2015. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. Atmos. Chem. Phys. 15, 8889–8973. <https://doi.org/10.5194/acp-15-8889-2015>.
- 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. <https://doi.org/10.1016/j.scitotenv.2019.01.333>.
- Septiawan, W.M., Endah, S.N., 2018. Suitable recurrent neural network for air quality prediction with backpropagation through time. 2nd International Conference on Informatics and Computational Sciences, ICICoS 2018. 196–201. <https://doi.org/10.1109/ICICoS.2018.8621720>.
- Song, X.Y., Gao, Y., Peng, Y.B., Huang, S., Liu, C., Peng, Z.R., 2021. A machine learning approach to modelling the spatial variations in the daily fine particulate matter (PM2.5) and nitrogen dioxide (NO₂) of Shanghai, China. Environment and Planning B: Urban Analytics and City Science. 48, 467–483. <https://doi.org/10.1177/2399808320975031>.

- Su, W.J., Liu, C., Hu, Q.H., Fan, G.Q., Xie, Z.Q., Huang, X., Zhang, T.S., Chen, Z.Y., Dong, Y.S., Ji, X.G., Liu, H.R., Wang, Z., Liu, J.G., 2017. Characterization of ozone in the lower troposphere during the 2016 G20 conference in Hangzhou. *Sci. Rep.* 7, 1–11. <https://doi.org/10.1038/s41598-017-17646-x>.
- Wang, B.W., Kong, W.W., Guan, H., Xiong, N.N., 2019. Air quality forecasting based on gated recurrent long short term memory model in internet of things. *IEEE Access*. 7, 69524–69534. <https://doi.org/10.1109/ACCESS.2019.2917277>.
- Wang, Dong sheng, Wang, H.W., Li, C., Lu, K.F., Peng, Z.R., Zhao, J., Fu, Q., Pan, J., 2020a. Roadside air quality forecasting in shanghai with a novel sequence-to-sequence model. *Int. J. Environ. Res. Publ. Health* 17 (24), 1–16. <https://doi.org/10.3390/ijerph17249471>.
- Wang, D.S., Wang, H.W., Lu, K.F., Peng, Z.R., Zhao, J., 2022. Regional prediction of ozone and fine particulate matter using diffusion convolutional recurrent neural network. *Int. J. Environ. Res. Publ. Health*. 19, 3988. <https://doi.org/10.3390/ijerph19073988>.
- Wang, H.W., Li, X.B., Wang, D., Zhao, J., He, H. di, Peng, Z.R., 2020b. Regional prediction of ground-level ozone using a hybrid sequence-to-sequence deep learning approach. *J. Clean. Prod.* 253, 119841 <https://doi.org/10.1016/j.jclepro.2019.119841>.
- Wang, H., Chen, H.S., Wu, Q.Z., Lin, J.M., Chen, X.S., Xie, X.W., Wang, R.R., Tang, X., Wang, Z.F., 2017. Accelerating the Global Nested Air Quality Prediction Modeling System (GNAQPMS) Model on Intel Xeon Phi Processors. *Geoscientific Model Development Discussions*. <https://doi.org/10.5194/gmd-2016-307>, 1–18.
- Wen, C.C., Liu, S.F., Yao, X.J., Peng, L., Li, X., Hu, Y., Chi, T.H., 2019. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *Sci. Total Environ.* 654, 1091–1099. <https://doi.org/10.1016/j.scitotenv.2018.11.086>.
- Wu, C.L., He, H.D., Song, R.F., Peng, Z.R., 2022. Prediction of air pollutants on roadside of the elevated roads with combination of pollutants periodicity and deep learning method. *Build. Environ.* 207, 108436 <https://doi.org/10.1016/j.buildenv.2021.108436>.
- Wu, C.L., Wang, H.W., Cai, W.J., He, H.D., Ni, A.N., Peng, Z.R., 2021. Impact of the COVID-19 Lockdown on Roadside Traffic-Related Air Pollution in Shanghai, China. *Building and Environment* 194. <https://doi.org/10.1016/j.buildenv.2021.107718>.
- Yan, R., Liao, J.Q., Yang, J., Sun, W., Nong, M.Y., Li, F.P., 2021. Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Syst. Appl.* 169, 114513. <https://doi.org/10.1016/j.eswa.2020.114513>.
- Zhang, B., Zou, G.J., Qin, D.M., Ni, Q., Mao, H.W., Li, M.Z., 2022. RCL-Learning : ResNet and convolutional long short-term memory-based spatiotemporal air pollutant concentration prediction model. *Expert Syst. Appl.* 207, 118017 <https://doi.org/10.1016/j.eswa.2022.118017>.
- Zhang, J.L., Chen, F., Cui, Z.Y., Guo, Y.N., Zhu, Y.D., 2021. Deep learning architecture for short-term passenger flow forecasting in urban rail transit. *IEEE Trans. Intell. Transport. Syst.* 22, 7004–7014. <https://doi.org/10.1109/TITS.2020.3000761>.
- Zhang, J.L., Chen, F., Guo, Y.N., Li, X.H., 2020. Multi-graph convolutional network for short-term passenger flow forecasting in urban rail transit. *IET Intell. Transp. Syst.* 14, 1210–1217. <https://doi.org/10.1049/iet-its.2019.0873>.
- Zhang, J.B., Zheng, Y., Qi, D.K., 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In: 31st AAAI Conference on Artificial Intelligence, AAAI 2017. 1655–1661.doi:10.48550/arXiv.1610.00081.
- Zhao, G.Y., Huang, G.Y., He, H.D., He, H.T., Ren, J.D., 2019. Regional spatiotemporal collaborative prediction model for air quality. *IEEE Access*. 7, 134903–134919. <https://doi.org/10.1109/ACCESS.2019.2941732>.
- Zhou, Y.L., Chang, F.J., Chang, L.C., Kao, I.F., Wang, Y.S., 2019. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *J. Clean. Prod.* 209, 134–145. <https://doi.org/10.1016/j.jclepro.2018.10.243>.
- Zhu, A.X., Lu, G.N., Liu, J., Qin, C.Z., Zhou, C.H., 2018. Spatial prediction based on third law of geography. *Spatial Sci.* 24, 225–240. <https://doi.org/10.1080/19475683.2018.1534890>.
- Zhu, X.H., He, H.D., Lu, K.F., Peng, Z.R., Gao, H.O., 2022. Characterizing carbon emissions from China V and China VI gasoline vehicles based on portable emission measurement systems. *J. Clean. Prod.* 378, 134458. <https://doi.org/10.1016/j.jclepro.2022.134458>.