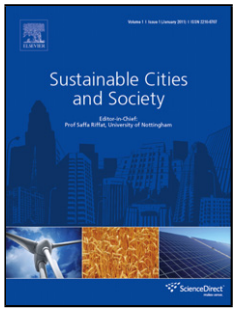
Journal Pre-proof



Modeling air quality prediction using a deep learning approach: Method optimization and evaluation

Wenjing Mao, Weilin Wang, Limin Jiao, Suli Zhao, Anbao Liu

PII: S2210-6707(20)30785-X

DOI: https://doi.org/10.1016/j.scs.2020.102567

Reference: SCS 102567

To appear in: *Sustainable Cities and Society*

Received Date: 18 July 2020

Revised Date: 14 October 2020

Accepted Date: 15 October 2020

Please cite this article as: { doi: https://doi.org/

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

Journal Pre-proof

Modeling Air Quality Prediction using a Deep Learning approach: Method Optimization and Evaluation

Wenjing Mao[a, b,1], Weilin Wang[a, b,1], Limin Jiao[a, b,*], Suli Zhao[a], Anbao Liu[c]

[a] School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China

[b] Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, Wuhan, 430079, China

[c] Fujian Jingwei Surveying and Mapping Information Co., Ltd.

* *Corresponding author*

E-mail addresses: Wenjing Mao wenjingmao@whu.edu.cn; Weilin Wang wangweilin@whu.edu.cn; Limin Jiao lmjiao@whu.edu.cn; Suli Zhao zhaosuli@csepdi.com; Anbao Liu lab115@sina.cn

Competing Interests

The authors declare no competing interests

[1] These authors contributed equally to this work and should be considered as co-first authors.

Journal Pre-proof

Highlights

- A temporal sliding model is proposed for $PM_{2.5}$ concentration long-term predictions.
- Integrating the optimal time lag of the spatiotemporal

Journal Pre-proof

correlations can improve model performance.

- The proposed model achieves higher prediction in the longer time series.
- The proposed model has strong practicability in making atmospheric management decisions.

Abstract: Air pollution is one of the hot issues that attracted widespread attention from urban and society management. Air quality prediction is to issue an alarm when severe pollution occurs, or pollution concentration exceeds a specific limit, contributing to the measure-taking of relevant departments, guiding urban socio-economic activities to promote sustainable urban development. However, existing methods have failed to make full use of the temporal features from spatiotemporal correlations of air quality monitoring stations, and achieved poor performances in long-term predictions (up to or above 24h-predictions). In this study, we proposed a deep learning framework to predict air quality in the following 24 hours: a neural network with a temporal sliding long short-term memory extended model (TS-LSTME). The model integrated the optimal time lag to realize sliding prediction through multi-layer bidirectional long short-term memory (LSTM), involving the hourly historical $PM_{2.5}$ concentration, meteorological data, and temporal data. We applied the proposed model to predict the next 24-hour average $PM_{2.5}$ concentration in Jing-Jin-Ji region, with the most severe air pollution in China. The proposed model had better stability and performances with high correlation coefficient R^2 (0.87), compared to the multiple linear regression (MLR), the support vector regression (SVR), the traditional LSTM, and the long short-term memory extended (LSTME) models. Moreover, the proposed model can achieve $PM_{2.5}$ concentration predictions with high accuracy in long-term series (48 h and 72 h). We also tested the model to predict O_3 concentration. The proposed model could be applied for other air pollutants. The proposed methods can significantly improve air quality prediction information services for the public and provide support for early warning and management of regional pollutants.

Keywords: air pollutant; air quality prediction; deep learning; long-term prediction; temporal sliding

1. Introduction

With the rapid development of the world economy and the acceleration of industrialization and urbanization, many cities around the world are suffering from air pollution. Air pollution problems are becoming more apparent, threatening human

Journal Pre-proof

production, life, and sustainable social development seriously (Li et al., 2019). Wherein, $PM_{2.5}$ (fine particles with a diameter of 2.5 microns or less in the atmosphere) is the main factor of air pollution, and the increasing $PM_{2.5}$ concentration will also directly affect the human health (Kampa and Castanas, 2008; Martins and Carrilho da Graça, 2018; Kumar et al., 2020; Barzeghar et al., 2020; Pak et al., 2020). Therefore, real-time $PM_{2.5}$ concentration predictions in advance have significant practical and social values, which play a significant role in making atmospheric management decisions to control air pollution and are important information needed to facilitate sustainable development.

Previous models for air pollutant predictions can be divided into two major categories: physical prediction models and statistical prediction models. The physical prediction models are based on aerodynamics, atmospheric physics, and chemistry to study pollutant diffusion mechanism (Geng et al., 2015), and apply mathematical methods to calculate the spatiotemporal distributions of pollutants (Jeong et al., 2011; Kim et al., 2010), such as CMAQ models (Jiang and Yoo, 2018; Woody et al., 2016; Yang et al., 2019), WRF-Chem models (Geng et al., 2015; Liu et al., 2018; Ma et al., 2018) and GEO-Chem models (Jeong et al., 2011; Lee et al., 2017; Wang et al., 2014). However, due to the complex pollutant diffusion mechanism, some problems produced by these models result in the usage limitations, such as colossal calculation workload, the complexity of the process, uncertainty of parameters (Pan et al., 2011; Stern et al., 2008; Vautard et al., 2007; J. Wang et al., 2019). The statistical prediction models are based on statistics and use historical time series data to predict future air quality. Compared with the physical prediction models, the statistical prediction models not only avoid the complicated mechanism so that the calculation is faster and the cost of each prediction can be reduced, but also can achieve an almost equivalent level of $PM_{2.5}$ concentration prediction accuracy of the physical prediction models (Engel-Cox et al., 2013). Due to those advantages, the application of the statistical prediction models is more extensive. Commonly used statistical prediction models are included machine learning (ML) and artificial neural network (ANN) models (Perez and Reyes, 2006). Machine learning models include multiple linear regression (MLR) models (Han et al., 2020; Stadlober et al., 2008), support vector regression (SVR) models (Zhu et al., 2018), random forest models (Gariazzo et al., 2020), etc. Machine learning models have clear mathematical logic, in which the relationship between input and output is relatively definite, and the structures are simple. However, these models ignore the influences of external variables and spatial factors (Leng et al., 2017), and are prone to overfitting due to improper index selection (Ausati and Amanollahi, 2016).

With the development of artificial intelligence, artificial neural network models, especially deep learning models, which take the nonlinear relationship between the prediction target and external variables into account, have excellent performance in prediction tasks because of their adaptiveness and robustness. Moreover, many studies showed that these models are suitable for large sample tasks (Franceschi et al., 2018; Kolehmainen et al., 2001; Voukantsis et al., 2011). Commonly used models include multilayer perceptron (MLP) models (Li et al., 2011), back propagation neural network (BPNN) models (W. Wang et al., 2019), general regression neural network (GRNN) models (Antanasijević et al., 2013), recurrent neural network (RNN) models (Feng et

Journal Pre-proof

al., 2011), long short-term memory (LSTM) models (Sak et al., 2016), long short-term memory extended (LSTME) models (Li et al., 2017), lag layer-LSTM-Fully connected network (Lag-FLSTM) models (Ma et al., 2020), etc. Compared with the MLP, BPNN and GRNN models, the RNN models are more excellent in processing time-series information (Ma et al., 2015), but the problems such as suffering from gradient explosion or disappearance limit the long-term dependence of capturing time series. LSTM models can effectively overcome these problems of RNN models (Hochreiter and Schmidhuber, 1997). However, when establishing the LSTM models, some previous studies did not make full use of the spatiotemporal correlations of $PM_{2.5}$ monitoring stations, and some of them ignored the influences of meteorological factors. Studies have shown that meteorological characteristics and temporal characteristics will affect prediction results of $PM_{2.5}$ concentration (Bai et al., 2016; Krishan et al., 2019; Zhang et al., 2018; Zhao et al., 2020), and there exist strong spatial correlations of $PM_{2.5}$ monitoring stations within a certain range (Wen et al., 2019). The LSTME models are based on the LSTM, considering the meteorological factors and the spatiotemporal correlations with the monitoring stations, and they performed well in short-term prediction tasks (Li et al., 2017). In recent years, some researchers combined convolutional neural networks (CNN) and LSTM models to simulate the changing process of air pollutants and achieved good performances of prediction, such as C-LSTME (Wen et al., 2019), graph GC-LSTM (Qi et al., 2019), multi-output deep learning LSTM (DM-LSTM) (Zhou et al., 2019), and Deep CNN-LSTM (Huang and Kuo, 2018) models. However, they have failed to capture the time dependence features effectively, which resulted in exhibiting relatively low precision for long-term prediction tasks (up to or above 24h-predictions) of $PM_{2.5}$ concentration.

Studies have proven that spatiotemporal correlations of $PM_{2.5}$ monitoring stations play a non-substitutable role, which can significantly affect the changing trend of $PM_{2.5}$ concentration (Li et al., 2017; W. Wang et al., 2019; Wen et al., 2019). In our study, we proposed a temporal sliding long short-term extended (TS-LSTME) model, taking the spatiotemporal correlations of $PM_{2.5}$ concentration at the monitoring stations into consideration. This model realized temporal sliding predictions of $PM_{2.5}$ concentration by using the optimal time lag of $PM_{2.5}$ concentration temporal correlations. At the same time, the auxiliary data, including meteorological data and temporal data, were integrated to improve the prediction performance of the model. The model was applied to the long-term predictions of $PM_{2.5}$ concentration in Jing-Jin-Ji (Beijing, Tianjin, and Hebei) area, China. The results showed that the proposed model with high accuracy could be used for long-term predictions of $PM_{2.5}$ concentration. Compared with the MLR model, SVR model, LSTME model, and conventional LSTM model, the TS-LSTME model achieved the best performance, especially in the predictions of long-term series (48 hours and 72 hours). According to the seasonal characteristics of $PM_{2.5}$ concentration, we selected representative time points (January 1st, April 1st, July 1st, and October 1st) to spatially interpolate the model predicted results and observed values of monitoring stations and compared spatiotemporal distributions. The results showed that the proposed TS-LSTME model has spatiotemporal prediction capability, which can be applied for early warning and management of regional pollutants. We also apply

Journal Pre-proof

the proposed model to predict O_3 concentration, and the results indicated that the model is suitable for other air pollutants prediction.

2. Data and methods

2.1 Study area and data description

Jing-Jin-Ji (113° E–120° E and 36° N–43° N) area was taken as our study area, which includes 13 cities in Beijing, Tianjin and Hebei province. The economy is developing rapidly, the population and heavy industries are concentrated in this area. It is a region where air pollution problems are relatively serious.

Our study collected $PM_{2.5}$ observed data from 65 air quality monitoring stations and meteorological data over Jing-Jin-Ji area from January 1, 2014 to December 31, 2019, and regarded each city as a basic unit to establish a model respectively, which predicted the average $PM_{2.5}$ concentration of each station in the next 24 hours. The hourly scale $PM_{2.5}$ monitoring data came from the national city air quality real-time publishing platform (http://106.37.208.233:20035/), and the day scale meteorological data came from the China Meteorological Data Service Center (CMDC: http://data.cma.cn/en). The data of missing time for each kind were compensated by the collected data of adjacent time. Figure 1 shows the distribution of air quality monitoring stations (red circles) and meteorological monitoring stations (blue triangles) in the Jing-Jin-Ji area. The statistics of air quality monitoring stations in 13 cities are shown in Table 1. The measured $PM_{2.5}$ data at the historical moment, meteorological data, and temporal data were entered into the model. Meteorological characteristics included precipitation (PRE), air pressure (PRS), relative humidity (RHU), sunshine (SSD), temperature (TEM), and wind direction and speed (WIN). Table 2 lists the statistical characteristics of the selected data.

Journal Pre-proof

Journal Pre-proof

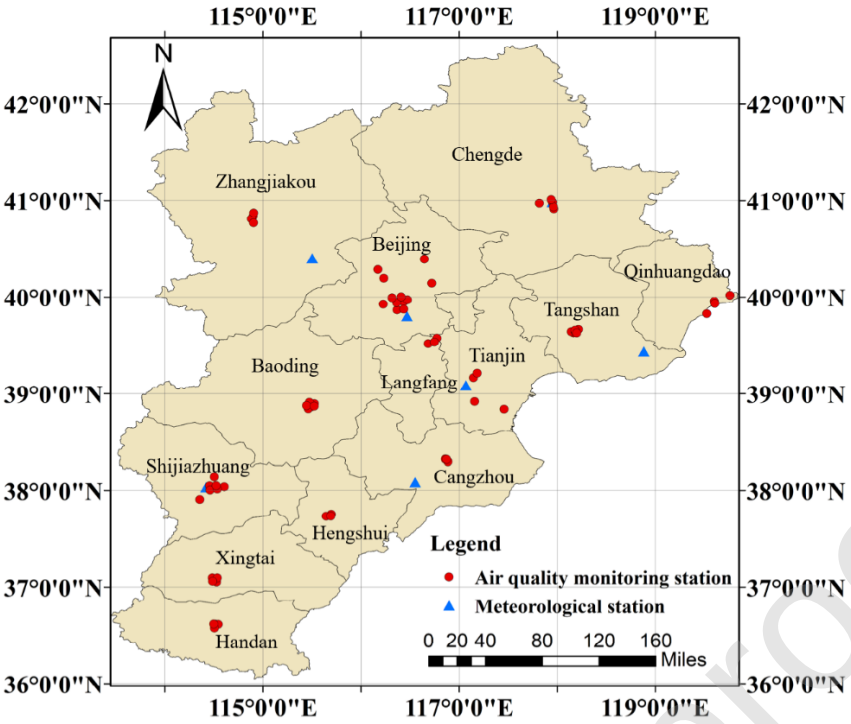


Figure 1. Distribution of air quality monitoring stations and meteorological stations over Jing-Jin-Ji area.

Table 1. Statistics of air quality monitoring stations in various cities.

City	Station code	Count of stations
Beijing	1001A-1012A	12
Tianjin	1015A,1019A,1024A,1027A	4
Shijiazhuang	1029A-1035A	7
Tangshan	1036A-1041A	6
Qinhuangdao	1042A-1045A	4
Handan	1047A-1050A	4
Baoding	1051A-1056A	6
Zhangjiakou	1057A,1059A-1061A	4
Chengde	1062A-1066A	5
Langfang	1067A-1069A	3
Cangzhou	1071A-1073A	3
Hengshui	1074A-1076A	3
Xingtai	1077A-1080A	4

Table 2. Statistical characteristics of experimental data.

Variable	Unit	Range	Variable Mean	St.dev
$PM_{2.5}$	μg /m^3	[1,500]	62.14	62.69
PRE	mm	[0,175.04]	1.51	5.85
PRS	hPa	[893.50,1043.28]	967.06	33.40
RHU	%	[3.04,97.85]	47.67	21.73
SSD	MJ/m^2	[0,13.82]	6.40	3.89

Journal Pre-proof

Journal Pre-proof

TEM	℃	[-22.66,35.13]	12.87	11.22
WIN	m/s	[0.05, 28.77]	2.21	3.02
Season	NA	[1,4]	-	-
Month	NA	[1,12]	-	-

2.2 Spatiotemporal correlation

Pearson correlation coefficient is usually used to reflect the statistics of the linear correlation between two variables (Box and Jen, n.d.). Pearson correlation coefficient formula is as follows:

$$\rho_{xy} = \frac{Cov(X \cdot Y)}{\sigma_x \cdot \sigma_y} \quad (1)$$

Where *X* and *Y* represent two variables, $Cov(X \cdot Y)$ represents the covariance of two variables, $\sigma_x \sigma_y$ respectively represent the standard deviation of two variables *X*, *Y*.

To evaluate the degree of the temporal correlation, Pearson correlation coefficient formula was used to calculate the temporal correlations. The variables *X* and *Y* are the values of $PM_{2.5}$ concentration before and after the time interval. Figure 2 shows the changes in the Pearson correlation coefficients of the 65 air quality monitoring stations over our Jing-Jin-Ji area for 0-50 hours intervals. It can be seen from Figure 2 that the Pearson correlation coefficients decrease with the increase of the time interval. When the time interval is less than 15h, the Pearson correlation coefficients of the front and back time intervals are greater than 0.4. The smaller the time interval, the higher the degree of correlation.

Journal Pre-proof

Journal Pre-proof

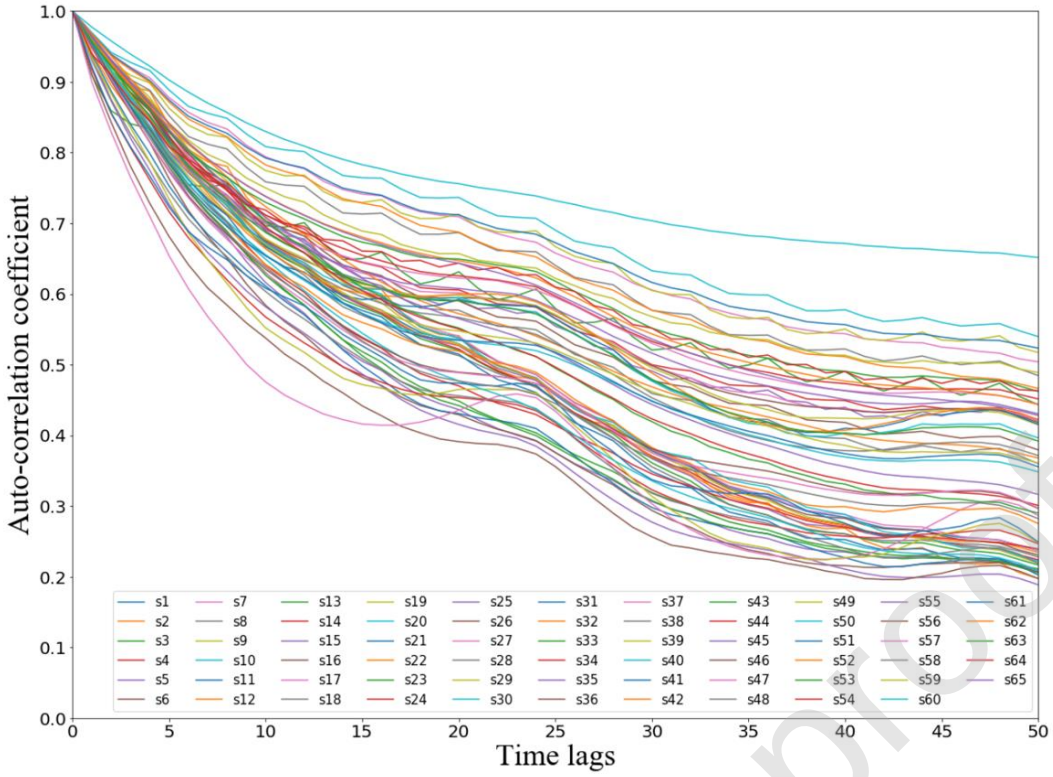


Figure 2. Variations among the time autocorrelation coefficients with respect to different time lags for 65 selected stations (s1-s65).

Besides, there exist strong spatial correlations of $PM_{2.5}$ monitoring stations. The Pearson correlation coefficient formula was also used to calculate the correlation coefficients among monitoring stations in each city. Table 3 shows the maximum and minimum values of spatial correlation coefficients of $PM_{2.5}$ monitoring stations in different cities. It can be seen from Table 3 that the correlation coefficients of $PM_{2.5}$ monitoring stations in each city are higher than 0.70, and the most of them are higer than 0.90, indicating that there is a high degree of spatial correlations.

Table 3. Maximum and Minimum of spatial correlation coefficients in different cities

City	Maximum of correlation coefficients	Minimum of correlation coefficients
Beijing	0.97	0.81
Tianjin	0.93	0.84
Shijiazhuang	0.95	0.91
Tangshan	0.95	0.92
Qinhuangdao	0.86	0.77
Handan	0.92	0.89
Baoding	0.95	0.90
Zhangjiakou	0.83	0.73
Chengde	0.92	0.80
Langfang	0.95	0.92
Cangzhou	0.94	0.93
Hengshui	0.94	0.84

Journal Pre-proof

Xingtai	0.94	0.90

2.3 TS-LSTME model

2.3.1 LSTM model

The long short-term memory neural network (LSTM NN) is a variant of the recurrent neural network (RNN). The LSTM NN can make full use of long time-series data, and effectively solve the gradient explosion and disappearance problems generated by the RNN. Figure 3 shows the structure of the LSTM. Compared with the RNN, the LSTM NN introduces three gating mechanisms inside, which can control the path of information transmission through the forget gate F_t, input gate I_t, and output gate O_t. Where C_{t-1} and h_{t-1} represent the internal and external state of the last moment, X_t, C_t, and h_t represent the input, internal and external state of the current moment. Using the two activation functions of sigmoid and tanh, the forget gate F_t, input gate I_t, output gate O_t, and candidate state c_t are calculated through h_{t-1} and X_t. Then combine the forget gate F_t, the input gate I_t, and the candidate state c_t to update the internal state C_t at the current moment. Finally combine the internal state C_t at the current moment and the output gate O_t to obtain the external state h_t at the current moment. The forward propagation formulas of the LSTM NN are as follows:

$$F_t = \sigma\left(W_f X_t + U_f h_{t-1} + b_f\right) \tag{2}$$

$$I_t = \sigma\left(W_i X_t + U_i h_{t-1} + b_i\right) \tag{3}$$

$$O_t = \sigma\left(W_o X_t + U_o h_{t-1} + b_o\right) \tag{4}$$

$$c_t = \tanh\left(W_c X_t + U_c h_{t-1} + b_c\right) \tag{5}$$

$$C_t = F_t \odot C_{t-1} + I_t \odot c_t \tag{6}$$

$$h_t = O_t \odot tanh\left(C_t\right) \tag{7}$$

Where W_f, W_i, W_o and W_c are the weight matrices for input vector X_t at current moment. U_f, U_i, U_o and U_c are the weight matrices for external state h_{t-1} at last moment. b_f, b_i, b_o and b_c are the bias vectors. σ and tanh are activation functions which can bring non-linearity to model. $\odot$ stands for element-wise multiplication of the matrix.

Journal Pre-proof

Journal Pre-proof

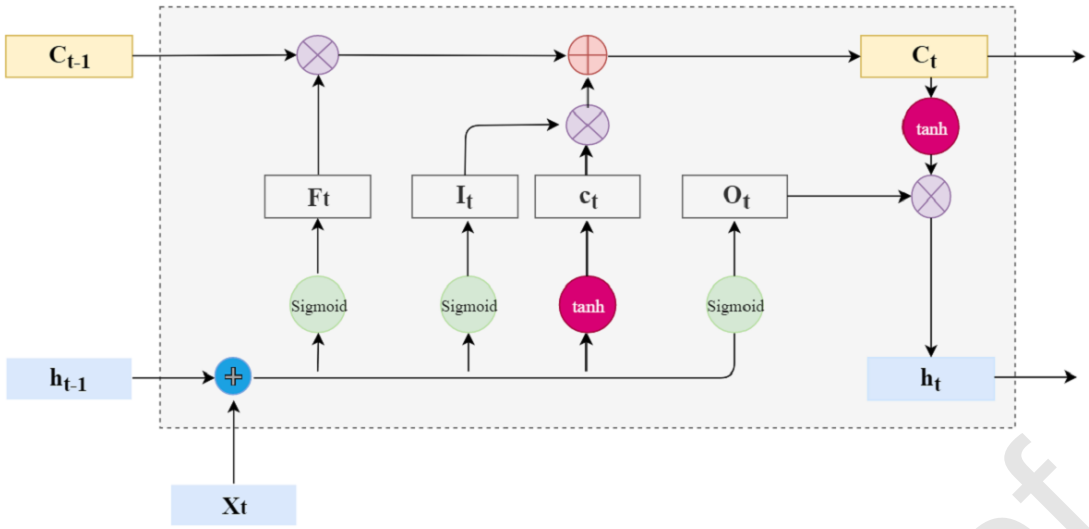


Figure 3. The network structure of the LSTM.

2.3.2 TS-LSTME model

Integrating hourly-scale $PM_{2.5}$ concentration data and day-scale meteorological data, and taking the spatiotemporal correlations of $PM_{2.5}$ monitoring stations into account, our study regarded a city as a basic unit to establish a TS-LSTME model. The model can perform collaborative training on the data of multiple base stations in the unit city to predict the average $PM_{2.5}$ concentration of multiple monitoring stations in the next 24 hours. Figure 4 shows the framework of the TS-LSTME model. The feed-forward process of the TS-LSTME model is as follows: (1) Considering the significant spatiotemporal correlations of $PM_{2.5}$ monitoring stations in each city, the historical data of $PM_{2.5}$ concentration at monitoring stations in each city are used as the inputs of the model; (2) The time-related characteristics of $PM_{2.5}$ concentration data of monitoring stations in each city are extracted through the bidirectional LSTM layer and the fully connected layer, and the results will be the $PM_{2.5}$ concentration for 0-*r* hours (*output1*) in the future. The previous step outputs are used as the inputs for the next prediction to predict the $PM_{2.5}$ concentration for *r*-2 * *r* hours (*output2*) in the future. Through the combination of a bidirectional LSTM layer and a fully connected layer, the above process performs temporal sliding prediction to form the temporal sliding block until the $PM_{2.5}$ concentration in the next 24 hours is predicted, where the value of *r* is determined by the optimal time lag selected. Then, the predicted results (*output1*, *output2*, ..., *outputn*) are concatenated, and the training results are passed through the bidirectional LSTM layer, which is used as the inputs of the next process. This process is forming the TS-LSTME unit; (3) The outputs of the previous process and auxiliary data are integrated into the fully connected layer, which could obtain the results of the $PM_{2.5}$ concentration of each monitoring station in the next 24 hours. The auxiliary data are composed of normalized meteorological data (precipitation, air pressure, humidity, sunshine, temperature, wind direction, and speed), and temporal data (one-hot codes formed by seasons and months).

In the TS-LSTME model, the rectified linear unit (ReLU) function was used as the

Journal Pre-proof

activation function, and dropout and regularization were used respectively in the LSTM layer and the fully connected layer to improve the generalization ability of the model and accelerate the speed of network training. To evaluate the effectiveness of the proposed model, root mean square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2) and normalized root mean square error (NRMSE) were used as metric indices to evaluate model performance. These indicators are defined as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_i - O_i\right)^2} \tag{8}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|P_i - O_i\right| \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(P_i - O_i\right)^2}{\sum_{i=1}^{N}\left(\bar{O}_i - O_i\right)^2} \tag{10}$$

$$NRMSE = \frac{RMSE}{O_{\max} - O_{\min}} \tag{11}$$

where O_i is the observed air pollutant concentration, $\bar{O}_i$ is the average observed air pollutant concentration, $O_{\max}$ is the observed maximum value of air pollutant concentration, $O_{\min}$ is the observed minimum value of air pollutant concentration, P_i is the predicted air pollutant concentration, and N is the number of test samples.

Journal Pre-proof

Journal Pre-proof

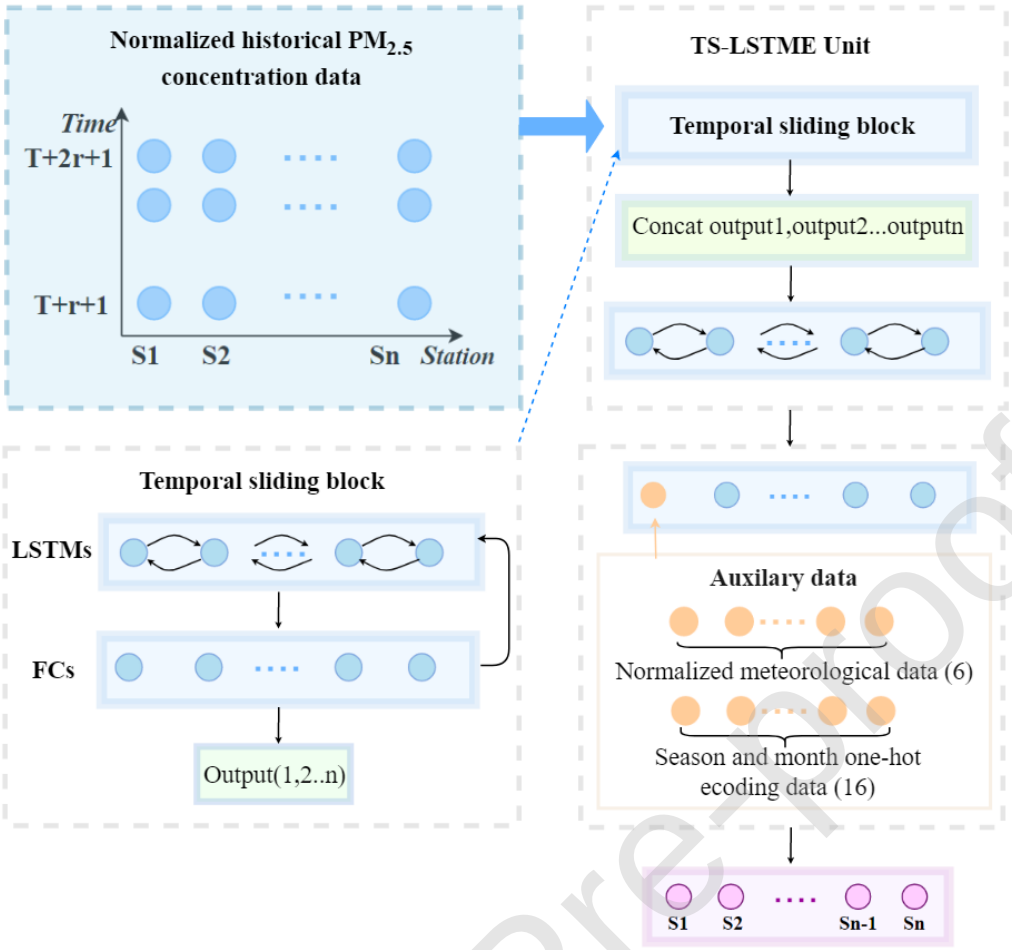


Figure 4. Network framework of the TS-LSTME model for $PM_{2.5}$ concentration predictions. The model inputs consist of historical $PM_{2.5}$ concentration data of all monitoring stations in the unit city in the dark blue dotted box and auxiliary data (meteorological data and temporal data) in the yellow dotted box. *r* represents the time lag, *T* represents the present time, *S* represents the monitoring station, and *n* indicates the number of monitoring stations.

According to the temporal correlations of $PM_{2.5}$ concentration data and the characteristics of the LSTM layer, the inputs of the TS-LSTME network are three-dimensional data: the number of input samples, the time lag for future prediction selection, and the number of monitoring stations in the city. The time lag is also used as the time lag interval of each output in the TS-LSTME unit to determine the value of *r*. As shown in Figure 5, according to the time lag *r*, the next 24 hours are grouped (*T to T + r, T + r + 1 to T + 2r + 1*, etc.) to form the TS-LSTME unit, and it is worth noting that the average concentration is predicted in the last stage (*T + 24-r to T + 24*). Based on the TS-LSTME unit, the model can predict the average $PM_{2.5}$ concentration of multiple monitoring stations in the next 24 hours. According to the previous temporal correlations analysis, the time lag will be selected from {6, 8, and 12}. Taking time lag selection 6 as an example, the TS-LSTME unit is formed by four temporal sliding prediction results (*output1*, *output2*, *output3*, *output4*), which are respectively the predicted $PM_{2.5}$ concentration in the next 1-6 hours, 7-12 hours, 13-18 hours, and the

Journal Pre-proof

predicted average $PM_{2.5}$ concentration in the next 19-24 hours. Experiments with different time lag values were made to choose the best time lag, which is the recursive time interval of the TS-LSTME unit. The experimental results of all cities in our study area have shown that with 12 as the time-lag input, the prediction ability of the model was better. The experimental results of Beijing in Table 4 show that when the time lag is selected as 12, that is, r = 12, the results showed better performance with RMSE 20.00 μg/m^3, MAE 13.54 μg/m^3 and R^2 0.72, compared with the other values of r.

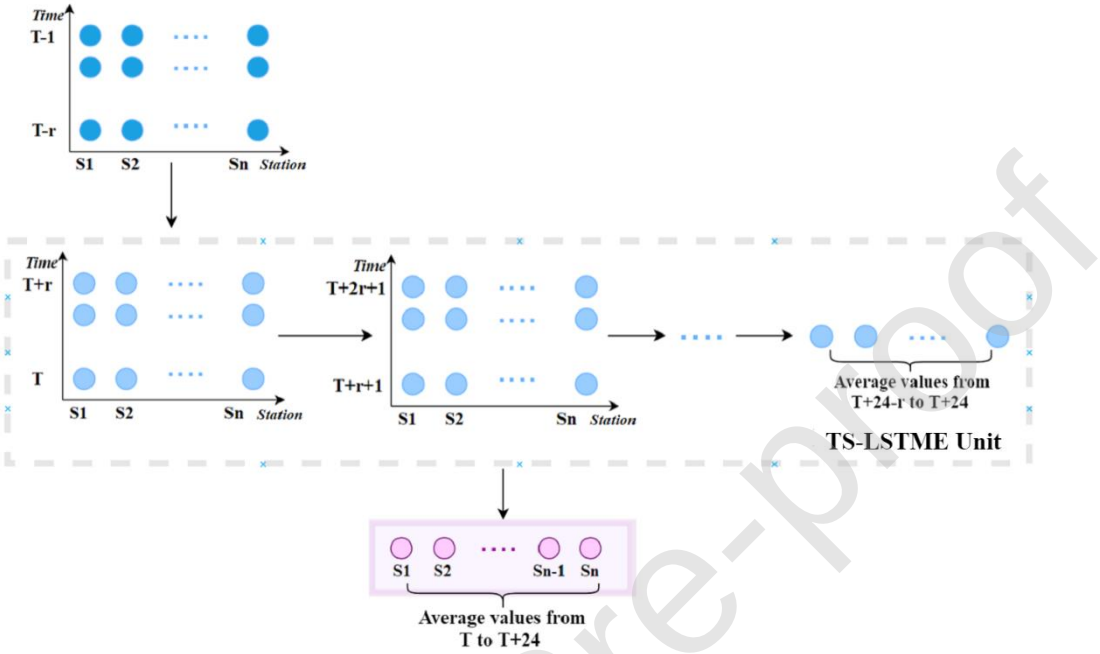


Figure 5. Illustration of TS-LSTME Unit. *r* represents the time lag, *T* represents the present time, *S* represents the monitoring station, and *n* Indicates the number of monitoring stations.

Table 4. Effects of different time lags.

Time lag (value of *r*)	Structure of TS-LSTME Unit	RMSE (μg/m^3)	MAE (μg/m^3)	R^2
6	1-6h→7-12h→13-18h→19-24h average	23.41	16.33	0.64
8	1-8h→9-16h→16-24h average	21.90	15.21	0.69
12	1-13h→14-24h average	20.00	13.54	0.72

3. Results

`3.1 Performance of the model TS-LSTME

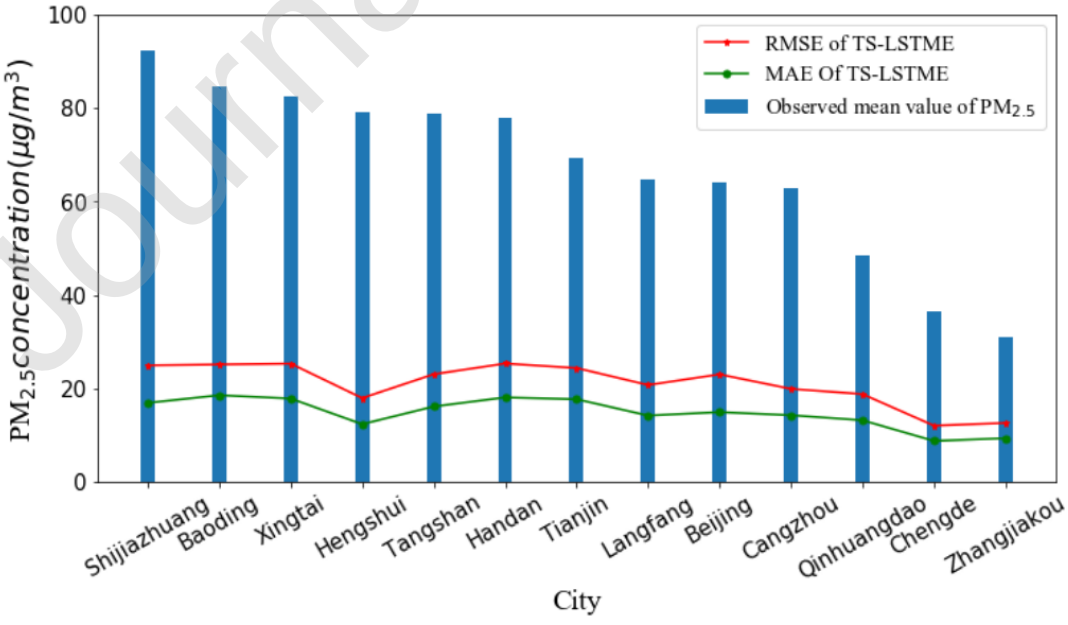
The study obtained the data including a total of 52,584-hour $PM_{2.5}$ concentration data and 2191-day meteorological data (January 1, 2014 to December 31, 2019) in Jing-Jin-Ji monitoring stations, 60% of which were chosen as the training set, 20% as the validation set and 20% as the test set. After determining the best network architecture for the current predicting tasks, models were established in units of cities, and the training set was used to train the current TS-LSTME models. The accuracy indicators of the model prediction results of 13 cities in the study area are shown in Table 5 and

Journal Pre-proof

Figure 6. Table 5 shows that the R^2 values in 13 cities were above 0.70, and the R^2 value of Hengshui city was above 0.80. Moreover, the NRMSE values are between 5.96%-9.42%. Figure 6 shows the average values of the observed $PM_{2.5}$ concentration in each city, the MAE and RMSE values of the corresponding models. The abscissa was sorted from largest to smallest according to the average values of urban $PM_{2.5}$ concentration. We can see that the overall trends of MAE and RMSE were also gradually decreasing, and there existed no apparent fluctuation. In cities with high concentration of air pollution (such as Shijiazhuang, Baoding, and Tianjin city) and cities with low concentration (such as Chengde, Zhangjiakou city), models performed well overall, indicating that the TS-LSTME models have good capabilities of $PM_{2.5}$ concentration predictions.

Table 5. Effects of the TS-LSTME models in different cities, using three indicators: the root mean square error (RMSE), the mean absolute error (MAE), coefficient of determination (R^2) and the normalized root mean square error (NRMSE).

City	RMSE (μg/m^3)	MAE (μg/m^3)	R^2	NRMSE (%)
Beijing	20.00	13.54	0.72	7.52
Tianjin	24.37	17.69	0.72	8.04
Shijiazhuang	24.94	16.91	0.78	7.16
Tangshan	23.04	16.13	0.71	8.36
Qinhuangdao	18.79	13.19	0.74	6.40
Handan	25.33	18.08	0.74	9.12
Baoding	28.44	20.26	0.72	5.96
Zhangjiakou	12.61	9.34	0.72	9.42
Chengde	12.04	8.77	0.73	9.02
Langfang	20.74	14.20	0.72	8.30
Cangzhou	19.91	14.27	0.74	9.16
Hengshui	17.94	12.35	0.81	9.31



Journal Pre-proof

Journal Pre-proof

Figure 6. Distribution of average values of $PM_{2.5}$ concentration at each city in study area (histogram) and distribution of MAE values and RMSE values in TS-LSTME models of each cities.

To explore the prediction performances of the TS-LSTME models and the temporal correlations of $PM_{2.5}$ concentration, Figure 7 shows the changes in the daily average observed $PM_{2.5}$ concentration and predicted $PM_{2.5}$ concentration by our models in our study area from January 1, 2019 to December 31, 2019 (the regression curve corresponds to Figure 8 (b)). As we can see, the predicted trend curve can simulate the trend change of the actual observed values, indicating the proposed model do well in capturing temporal correlations of $PM_{2.5}$ concentration by virtue of the way of temporal sliding predictions. Figure 8 shows the regression curves of predicted and observed $PM_{2.5}$ concentration for all test samples (a) and daily samples (b). The values of R^2 are 0.86 and 0.87, which suggests that the proposed model can achieve predictions with high accuracy. The model can accurately predict the changing trend of $PM_{2.5}$ concentration, but the forecast results appeared "high values underestimation, low values overestimation" phenomenon, and the intervals of the prediction results were smaller than the intervals of the observation values, which also reflects the complexity of $PM_{2.5}$ concentration predictions.

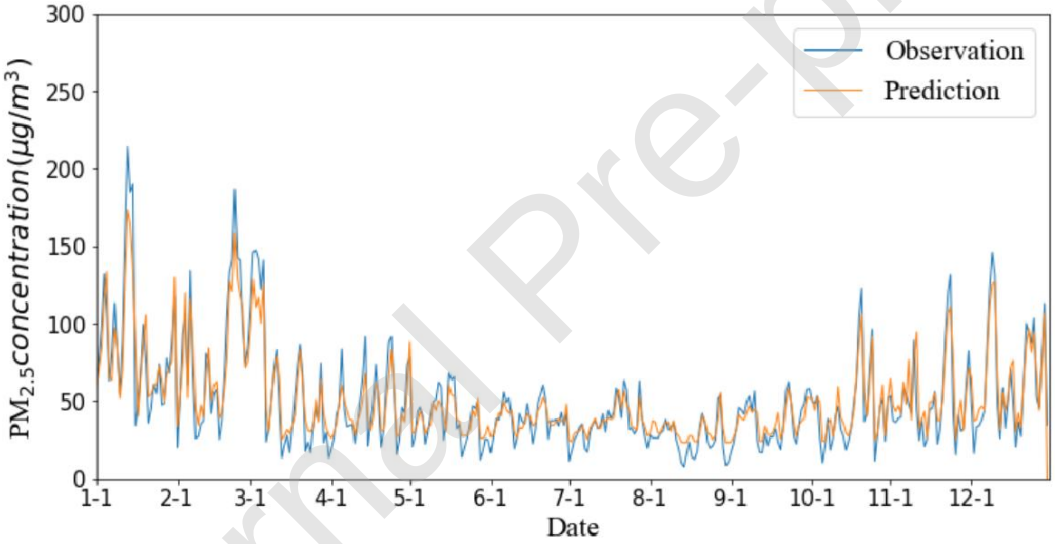


Figure 7. Distribution of daily averaged observed $PM_{2.5}$ concentration of all stations and the corresponding predictions by TS-LSTME models for Jing-Jin-Ji area from January 1, 2019 to December 31, 2019.

Journal Pre-proof

Journal Pre-proof

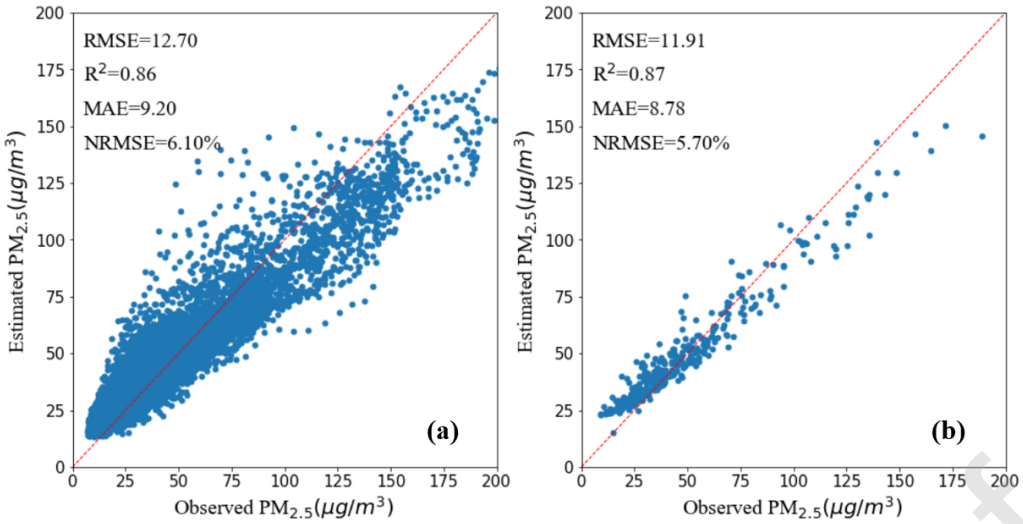


Figure 8. Correlation between the predicted and observed average values by TS-LSTME models for Jing-Jin-Ji area from January 1, 2019 to December 31, 2019: (a) all test samples (b) daily samples. The red dashed line is y = x reference line.

To further explore the spatiotemporal prediction capabilities of the TS-LSTME models and the spatial correlations of $PM_{2.5}$ monitoring stations, we compared predicted spatial interpolation results of $PM_{2.5}$ concentration by TS-LSTME models with observed results at monitoring stations, and selected representative comparison results on four days (January 1st, April 1st, July 1st, and October 1st) according to the seasonal characteristics of $PM_{2.5}$ concentration. From January to March and October to December each year, the $PM_{2.5}$ concentration is generally high, while from April to September, $PM_{2.5}$ concentration is generally low. Figure 9 shows the spatiotemporal distributions of the observed data and predicted data on January 1st, April 1st, July 1st, October 1st, 2019. From Figure 9, it can be found that the heavily polluted areas in the Jing-Jin-Ji area are mainly concentrated in cities in the southern and eastern regions. By comparison, the predicted results on January 1st in the southern area were lower than the observed values, but the predicted spatial distributions on January 1st, April 1st, July 1st, and October 1st were generally consistent with the observed spatial distributions of $PM_{2.5}$ concentration. The predicted spatial distributions were accurate, which indicated that the TS-LSTME models have strong capabilities for extracting spatiotemporal features to realize high accuracy predictions.

Journal Pre-proof

Journal Pre-proof

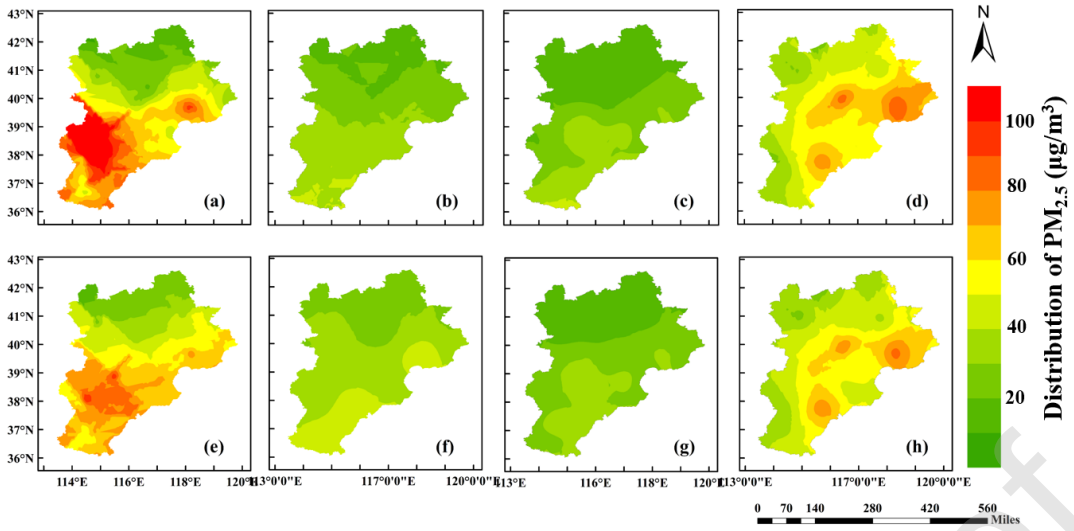


Figure 9. Spatial interpolation results of observed values and predicted values by TS-LSTME models on Jan 1st, 2019 (a, e), on April 1st, 2019 (b, f), on July 1st, 2019 (c, g), and on Oct 1st, 2019 (d, h).

3.2 Comparison of experiments

In order to verify the performances of the TS-LSTME models, our study also established deep learning models (LSTM and LSTME) based on cities and established MLR models and SVR models based on stations. The inputs of the models based on deep learning, the time lag, the number of nodes in each neuron layer structure, and the output structures remained unchanged, but the LSTM models did not consider the influences of meteorological factors. The MLR and SVR models are only time series prediction models, without considering the spatial correlation characteristics of $PM_{2.5}$ monitoring stations, and using the stations as the basic unit to perform prediction experiments separately. Table 6 shows the evaluation indices of experimental prediction results obtained by using three deep learning models TS-LSTME, LSTME and LSTM in 13 cities. The MAE values predicted by the TS-LSTME, LSTME, LSTM, MLR, and SVR models at each monitoring station were shown in Figure 10. It can be seen from Figure 10 that the MAE values of each monitoring station of the TS-LSTME, LSTME, LSTM models are less than the MAE values of the MLR, SVR models, indicating that the models based on deep learning which take into account the spatiotemporal correlations of $PM_{2.5}$ monitoring stations and possess the complexity of the networks, are more suitable for complex $PM_{2.5}$ concentration prediction tasks and can obtain prediction results of multiple monitoring stations at the same time. As can be seen from Table 6, the R^2 values of the LSTME models in each city are between 0.51-0.66, the R^2 values of the LSTM models are between 0.45-0.63, and R^2 values of the TS-LSTME models are above 0.70, of which evaluation indicators are better than other models. It is worth noting that has the R^2 value of the TS-LSTME model in Hengshui city is 0.81, and the MAE value at 1076A monitoring station is 5.98 $\mu g/m^3$, which is better than the other two models. The results show that compared with the directly predicted LSTM models and LSTME models, the TS-LSTME models use a highly time-correlated delay

Journal Pre-proof

for temporal sliding prediction, which can achieve higher accuracy and stability in long-term prediction, and add auxiliary data (meteorological data and temporal data) can improve the prediction performances of the models.

Table 6. Effects of the different models (TS-LSTME, LSTME, LSTM), using three indicators: the root mean square error (RMSE), the mean absolute error (MAE), and R^2.

City	TS-LSTME			LSTME			LSTM		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
Beijing	20.00	13.54	0.72	24.76	17.54	0.52	24.84	17.76	0.52
Tianjin	24.37	17.69	0.72	28.51	20.30	0.52	30.45	23.24	0.45
Shijiazhuang	24.94	16.91	0.78	29.76	20.79	0.66	30.89	21.80	0.63
Tangshan	23.04	16.13	0.71	25.87	18.02	0.52	27.15	19.89	0.47
Qinhuangdao	18.79	13.19	0.74	22.14	15.01	0.51	26.21	17.13	0.46
Handan	25.33	18.08	0.74	31.14	22.48	0.58	36.65	25.61	0.51
Baoding	28.44	20.26	0.72	29.34	22.12	0.66	30.67	22.49	0.62
Zhangjiakou	12.61	9.34	0.72	13.65	10.44	0.57	13.93	11.68	0.52
Chengde	12.04	8.77	0.73	14.20	10.00	0.51	14.85	10.10	0.47
Langfang	20.74	14.20	0.72	23.54	16.29	0.59	24.46	17.12	0.56
Cangzhou	19.91	14.27	0.74	22.83	16.75	0.62	23.60	16.86	0.59
Hengshui	17.94	12.35	0.81	32.71	27.09	0.41	33.13	27.21	0.39
Xingtai	25.31	17.84	0.76	27.73	19.38	0.68	29.68	21.74	0.65

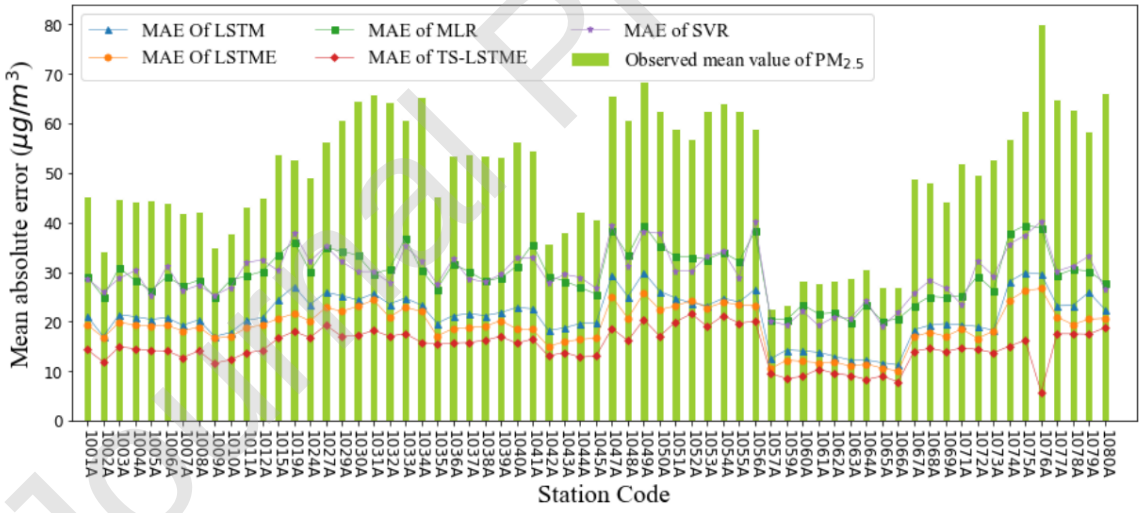


Figure 10. Comparison of the results (MAE) of different models (TS-LSTME, LSTME, LSTM, MLR, SVR) at each station in study area.

3.3 Long term predictions

To validate the broad applicability of the TS-LSTME model predictions for more extended time series, our study used Beijing as an experimental object to establish the TS-LSTME models to predict the average $PM_{2.5}$ concentration in the next 24h, 48h, and 72h. The most suitable hyperparameters were used in each model. The optimal time lag

Journal Pre-proof

(the value of r in the TS-LSTME Unit) and the auxiliary data increases as the prediction time extends. The model for the prediction task in the next 24h involved one-day meteorological data, while the models in the next 48h and 72h respectively involved two-day and three-day meteorological data. It can be seen from Table 7 that as the prediction time extends, the indicators RMSR, MAE values also increase, but the proposed model can also show satisfactory performances in the prediction tasks of 48h and 72h, which the MAE values are 17.86 μg/m^3 and 20.63 μg/m^3 respectively, comparing the MAE values of the LSTME model and the LSTM model in the 24-hour prediction of 17.54 μg/m^3 and 17.76 μg/m^3. The results showed that TS-LSTME models for the 48-hour prediction task (MAE 17.86 μg/m^3 and RMSE 24.96 μg/m^3) can achieve an almost equivalent level of prediction accuracy of LSTME models (MAE 17.54 μg/m^3, RMSE 24.76μg/m^3) and LSTM models (MAE 17.76 μg/m^3, RMSE 24.84 μg/m^3) for the 24-hour prediction task, demonstrating the TS-LSTME models which are based on the idea of temporal sliding prediction, can yield high accuracy in the long-term prediction tasks.

Table 7. Structure and prediction accuracy of multiscale air pollutant concentration predictions.

Task	Time lag (value of r)	Structure of TS-LSTME Unit	RMSE (μg/m^3)	MAE (μg/m ^{3})
24h	12	1-13h→14-24h average	20.00	13.54
48h	24	1-24h→25-48h average	24.96	17.86
72h	36	1-36h→37h-72h average	28.75	20.63

4. Discussion

4.1 Application of predicting O_3 concentration

In order to further reflect the performance of the present model for other air quality predictions, we applied the proposed model to predict O_3 concentration over Jing-Jin-Ji area. According to the correlation analysis (Table 8), it can be known that O_3 is less relevant to other pollutants, and O_3 is a critical atmospheric pollutant in summer, which has become a limiting factor affecting the number of days to achieve the standard. O_3 also has significant spatial and temporal characteristics, seasonal characteristics, and will be affected by meteorological conditions, which we could apply the proposed model to predict. Data from January 1, 2014 to December 31, 2016, over Jing-jin-ji area were input into the model to train and adjust the model's structure to achieve the best results. Figure 11 shows the correlation between the predicted and observed daily average O_3 values by TS-LSTME models on train set (2014-2015) and test set (2016), respectively. The proposed model obtained high R^2 values (0.96 and 0.86), indicating strong consistency between predicted and observed values. The results show that the proposed model is suitable for predicting multiple air pollutants.

Table 8. Pearson's coefficients between multiple air pollutants

R	$PM_{2.5}$	SO_2	NO_2	CO	O_3	PM_{10}
$PM_{2.5}$	1.00	0.66	0.65	0.81	-0.25	0.76
SO_2	-	1.00	0.56	0.80	-0.31	0.54
NO_2	-	-	1.00	0.70	-0.54	0.54

Journal Pre-proof

CO	-	-	-	1.00	-0.45	0.62
O_3	-	-	-	-	1.00	-0.19
PM_{10}	-	-	-	-	-	1.00

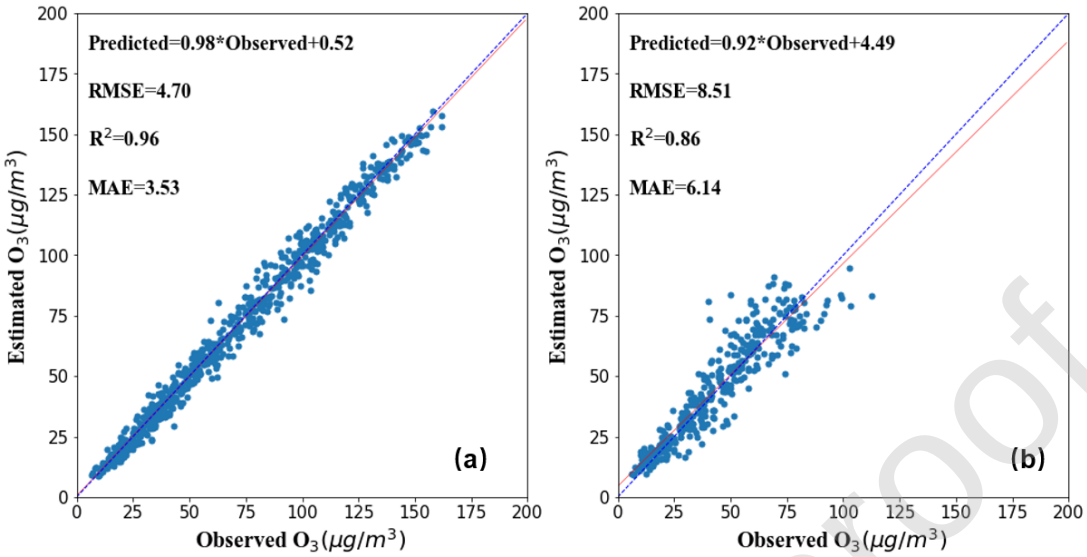


Figure 11. Correlation between the predicted and observed daily average O_3 values by TS-LSTME models on different datasets: (a) on train dataset (2014-2015), (b) on test dataset (2016). The solid line and dashed line are the regression line and y = x reference line, respectively.

4.2 Limitations and future work

In methodology, this study constructed a deep learning-based, nonlinear and adaptive $PM_{2.5}$ spatiotemporal temporal sliding prediction model by the analysis of the temporal and spatial correlations and influencing factors of atmospheric pollutants, based on the historical data of atmospheric pollutants from monitoring stations. The proposed model can more effectively capture the long-term change law of the target variable in the way of time sliding and thus performs well in long-term prediction tasks. It is suitable for predicting multiple air pollutants such as O_3, PM_{10}, etc. Moreover, the proposed model structure can be adjusted to predict other target variables by choosing suitable influencing factors and time lag appropriately according to the actual situation. However, there exist some limitations of the proposed model. First, the intervals of the prediction results are often smaller than the ranges of the actual values. There exist cases of underestimation and overestimation, which also reflects the complexity of $PM_{2.5}$ concentration predictions. In the future study, we will propose a global-local neural network model. A global model based on TS-LSTME is established to learn the overall change trend of air pollutants in the area, and a local model is set to learn the changes of air pollutant concentrations with higher or lower values, and fine-tune predictions of the global model to obtain the final forecast results. Second, we established the model in units of cities, without fully considering the regional effects of pollutants in a large area. In the future study, we will combine CNN and TS-LSTME model to extract more essential spatiotemporal correlation features, and apply it to a country.

Journal Pre-proof

5. Conclusions

Most current studies directly predict air pollutants by inputting data for a fixed period of time in the future. The prediction results performed well in the short-term prediction such as 1h-predictions, but in the long-term prediction (up to or above 24h-predictions), the time interval between forecasting and training will become very long, which results in lowering the temporal correlations between them. Thus the accuracy of the long-term prediction is generally low. In this paper, the TS-LSTME model based on deep artificial neural networks was proposed for air quality in long-term prediction. The proposed model involved hourly historical $PM_{2.5}$ concentration data and auxiliary data (meteorological data and temporal data). The model, which integrated the optimal time lag by the spatiotemporal correlation analysis, is capable of capturing the long-term change in the way of temporal sliding prediction through multi-layer bidirectional LSTM, which can maintain the high temporal correlations. The results of the performance evaluation and comparison with other models demonstrated that the proposed model based on the method of temporal sliding prediction could be more effective in extracting temporal features from spatiotemporal correlations of $PM_{2.5}$ monitoring stations for predicting $PM_{2.5}$ concentration in long-term series. The proposed model is suitable for long-term forecasting tasks and has strong practicability. The proposed model can be applied in other regions for multiple air pollutant predictions such as O_3, PM_{10}, etc. By spatializing predicted results of the air quality monitoring stations by the TS-LSTME model, the spatial distribution characteristics of air pollutants in the region can be evaluated in real-time.

Based on the forecast results, the relevant departments can predict where will appear more severe pollutant pollution and adjust some measures in time to effectively control pollution emissions according to the specific conditions of different cities, such as warn the public to go out as little as possible to reduce exhaust emissions caused by cars driving, supervise and control the construction site to reduce the dust caused, require some industrial enterprises to reduce production to ensure the stable discharge and meet the standards, strengthen the comprehensive management of catering industry fume and so on. The conceptual and methodological advancement of this study could assist in joint prevention and control of regional pollutants which could promote sustainable urban development.

Conflict of Interest

We would like to submit the revised manuscript entitled “Modeling air quality prediction using a deep learning approach: method optimization and evaluation”, which we wish to be considered for publication in “Sustainable Cities and Society”. No conflict of interest exits in the submission of this manuscript, and manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published

Journal Pre-proof

previously, and not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

References:

Antanasijević, D.Z., Pocajt, V.V., Povrenović, D.S., Ristić, M.Đ., Perić-Grujić, A.A., 2013. PM10 emission forecasting using artificial neural networks and genetic algorithm input variable optimization. Sci. Total Environ. 443, 511–519. https://doi.org/10.1016/j.scitotenv.2012.10.110

Ausati, S., Amanollahi, J., 2016. Assessing the accuracy of ANFIS, EEMD-GRNN, PCR, and MLR models in predicting $PM_{2.5}$. Atmos. Environ. 142, 465–474. https://doi.org/10.1016/j.atmosenv.2016.08.007

Bai, Y., Li, Y., Wang, X., Xie, J., Li, C., 2016. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. Atmospheric Pollut. Res. 7, 557–566. https://doi.org/10.1016/j.apr.2016.01.004

Barzeghar, V., Sarbakhsh, P., Hassanvand, M.S., Faridi, S., Gholampour, A., 2020. Long-term trend of ambient air PM_{10}, $PM_{2.5}$, and O_3 and their health effects in Tabriz city, Iran, during 2006–2017. Sustain. Cities Soc. 54, 101988. https://doi.org/10.1016/j.scs.2019.101988

Box, E.P., Jen, M., n.d. TIME SERIES ANALYSIS: FORECASTING AND CONTROL. (Revised Edition) by 1.

Engel-Cox, J., Kim Oanh, N.T., van Donkelaar, A., Martin, R.V., Zell, E., 2013. Toward the next generation of air quality monitoring: Particulate Matter. Atmos. Environ. 80, 584–590. https://doi.org/10.1016/j.atmosenv.2013.08.016

Feng, Y., Zhang, W., Sun, D., Zhang, L., 2011. Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification. Atmos. Environ. 45, 1979–1985. https://doi.org/10.1016/j.atmosenv.2011.01.022

Franceschi, F., Cobo, M., Figueredo, M., 2018. Discovering relationships and forecasting PM_{10} and $PM_{2.5}$ concentrations in Bogotá, Colombia, using Artificial Neural Networks, Principal Component Analysis, and k-means clustering. Atmospheric Pollut. Res. 9, 912–922. https://doi.org/10.1016/j.apr.2018.02.006

Gariazzo, C., Carlino, G., Silibello, C., Renzi, M., Finardi, S., Pepe, N., Radice, P., Forastiere, F., Michelozzi, P., Viegi, G., Stafoggia, M., 2020. A multi-city air pollution population exposure study: Combined use of chemical-transport and random-Forest models with dynamic population data. Sci. Total Environ. 724, 138102. https://doi.org/10.1016/j.scitotenv.2020.138102

Geng, G., Zhang, Q., Martin, R.V., van Donkelaar, A., Huo, H., Che, H., Lin, J., He, K., 2015. Estimating long-term $PM_{2.5}$ concentrations in China using satellite-based aerosol optical depth and a chemical transport model. Remote Sens. Environ. 166, 262–270. https://doi.org/10.1016/j.rse.2015.05.016

Han, L., Zhao, J., Gao, Y., Gu, Z., Xin, K., Zhang, J., 2020. Spatial distribution characteristics of $PM_{2.5}$ and PM_{10} in Xi'an City predicted by land use regression models. Sustain. Cities Soc.

Journal Pre-proof

61, 102329. https://doi.org/10.1016/j.scs.2020.102329

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, C.-J., Kuo, P.-H., 2018. A Deep CNN-LSTM Model for Particulate Matter ($PM_{2.5}$) Forecasting in Smart Cities. Sensors 18, 2220. https://doi.org/10.3390/s18072220

Jeong, J.I., Park, R.J., Woo, J.-H., Han, Y.-J., Yi, S.-M., 2011. Source contributions to carbonaceous aerosol concentrations in Korea. Atmos. Environ. 45, 1116–1125. https://doi.org/10.1016/j.atmosenv.2010.11.031

Jiang, X., Yoo, E., 2018. The importance of spatial resolutions of Community Multiscale Air Quality (CMAQ) models on health impact assessment. Sci. Total Environ. 627, 1528–1543. https://doi.org/10.1016/j.scitotenv.2018.01.228

Kampa, M., Castanas, E., 2008. Human health effects of air pollution. Environ. Pollut. 151, 362–367. https://doi.org/10.1016/j.envpol.2007.06.012

Kim, Y., Fu, J.S., Miller, T.L., 2010. Improving ozone modeling in complex terrain at a fine grid resolution: Part I – examination of analysis nudging and all PBL schemes associated with LSMs in meteorological model. Atmos. Environ. 44, 523–532. https://doi.org/10.1016/j.atmosenv.2009.10.045

Kolehmainen, M., Martikainen, H., Ruuskanen, J., 2001. Neural networks and periodic components used in air quality forecasting. Atmos. Environ. 35, 815–825. https://doi.org/10.1016/S1352-2310(00)00385-X

Krishan, M., Jha, S., Das, J., Singh, A., Goyal, M.K., Sekar, C., 2019. Air quality modelling using long short-term memory (LSTM) over NCT-Delhi, India. Air Qual. Atmosphere Health 12, 899–908. https://doi.org/10.1007/s11869-019-00696-7

Kumar, A., Ambade, B., Sankar, T.K., Sethi, S.S., Kurwadkar, S., 2020. Source identification and health risk assessment of atmospheric $PM_{2.5}$-bound polycyclic aromatic hydrocarbons in Jamshedpur, India. Sustain. Cities Soc. 52, 101801. https://doi.org/10.1016/j.scs.2019.101801

Lee, H.-M., Park, R.J., Henze, D.K., Lee, S., Shim, C., Shin, H.-J., Moon, K.-J., Woo, J.-H., 2017. $PM_{2.5}$ source attribution for Seoul in May from 2009 to 2013 using GEOS-Chem and its adjoint model. Environ. Pollut. 221, 377–384. https://doi.org/10.1016/j.envpol.2016.11.088

Leng, X., Wang, J., Ji, H., Wang, Q., Li, H., Qian, X., Li, F., Yang, M., 2017. Prediction of size-fractionated airborne particle-bound metals using MLR, BP-ANN and SVM analyses. Chemosphere 180, 513–522. https://doi.org/10.1016/j.chemosphere.2017.04.015

Li, C., Hsu, N.C., Tsay, S.-C., 2011. A study on the potential applications of satellite data in air quality monitoring and forecasting. Atmos. Environ. 45, 3663–3675. https://doi.org/10.1016/j.atmosenv.2011.04.032

Li, T., Guo, Y., Liu, Y., Wang, J., Wang, Q., Sun, Z., He, M.Z., Shi, X., 2019. Estimating mortality burden attributable to short-term $PM_{2.5}$ exposure: A national observational study in China. Environ. Int. 125, 245–251. https://doi.org/10.1016/j.envint.2019.01.073

Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., Chi, T., 2017. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. Environ. Pollut. 231, 997–1004. https://doi.org/10.1016/j.envpol.2017.08.114

Liu, S., Hua, S., Wang, K., Qiu, P., Liu, H., Wu, B., Shao, P., Liu, X., Wu, Y., Xue, Y., Hao, Y., Tian,

Journal Pre-proof

H., 2018. Spatial-temporal variation characteristics of air pollution in Henan of China: Localized emission inventory, WRF/Chem simulations and potential source contribution analysis. Sci. Total Environ. 624, 396–406. https://doi.org/10.1016/j.scitotenv.2017.12.102

Ma, J., Ding, Y., Cheng, J.C.P., Jiang, F., Gan, V.J.L., Xu, Z., 2020. A Lag-FLSTM deep learning network based on Bayesian Optimization for multi-sequential-variant $PM_{2.5}$ prediction. Sustain. Cities Soc. 60, 102237. https://doi.org/10.1016/j.scs.2020.102237

Ma, X., Sha, T., Wang, J., Jia, H., Tian, R., 2018. Investigating impact of emission inventories on $PM_{2.5}$ simulations over North China Plain by WRF-Chem. Atmos. Environ. 195, 125–140. https://doi.org/10.1016/j.atmosenv.2018.09.058

Ma, X., Tao, Z., Wang, Yinhai, Yu, H., Wang, Yunpeng, 2015. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transp. Res. Part C Emerg. Technol. 54, 187–197. https://doi.org/10.1016/j.trc.2015.03.014

Martins, N.R., Carrilho da Graça, G., 2018. Impact of $PM_{2.5}$ in indoor urban environments: A review. Sustain. Cities Soc. 42, 259–275. https://doi.org/10.1016/j.scs.2018.07.011

Pak, U., Ma, J., Ryu, U., Ryom, K., Juhyok, U., Pak, K., Pak, C., 2020. Deep learning-based $PM_{2.5}$ prediction considering the spatiotemporal correlations: A case study of Beijing, China. Sci. Total Environ. 699, 133561. https://doi.org/10.1016/j.scitotenv.2019.07.367

Pan, L., Sun, B., Wang, W., 2011. City Air Quality Forecasting and Impact Factors Analysis Based on Grey Model. Procedia Eng. 12, 74–79. https://doi.org/10.1016/j.proeng.2011.05.013

Perez, P., Reyes, J., 2006. An integrated neural network model for PM_{10} forecasting. Atmos. Environ. 40, 2845–2851. https://doi.org/10.1016/j.atmosenv.2006.01.010

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

Sak, H., Yang, G., Li, B., Li, W., 2016. Modeling Dependence Dynamics of Air Pollution: Pollution Risk Simulation and Prediction of $PM_{2.5}$ Levels. ArXiv160205349 Stat.

Stadlober, E., Hörmann, S., Pfeiler, B., 2008. Quality and performance of a PM_{10} daily forecasting model☆. Atmos. Environ. 42, 1098–1109. https://doi.org/10.1016/j.atmosenv.2007.10.073

Stern, R., Builtjes, P., Schaap, M., Timmermans, R., Vautard, R., Hodzic, A., Memmesheimer, M., Feldmann, H., Renner, E., Wolke, R., 2008. A model inter-comparison study focussing on episodes with elevated PM_{10} concentrations. Atmos. Environ. 42, 4567–4588. https://doi.org/10.1016/j.atmosenv.2008.01.068

Vautard, R., Builtjes, P.H.J., Thunis, P., Cuvelier, C., Bedogni, M., Bessagnet, B., Honoré, C., Moussiopoulos, N., Pirovano, G., Schaap, M., 2007. Evaluation and intercomparison of Ozone and PM_{10} simulations by several chemistry transport models over four European cities within the CityDelta project. Atmos. Environ. 41, 173–188. https://doi.org/10.1016/j.atmosenv.2006.07.039

Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A., Kolehmainen, M., 2011. Intercomparison of air quality data using principal component analysis, and forecasting of PM_{10} and $PM_{2.5}$ concentrations using artificial neural networks, in Thessaloniki and Helsinki. Sci. Total Environ. 409, 1266–1276. https://doi.org/10.1016/j.scitotenv.2010.12.039

Wang, J., Bai, L., Wang, S., Wang, C., 2019. Research and application of the hybrid forecasting model based on secondary denoising and multi-objective optimization for air pollution

Journal Pre-proof

early warning system. J. Clean. Prod. 234, 54–70. https://doi.org/10.1016/j.jclepro.2019.06.201

Wang, L., Wang, S., Zhang, L., Wang, Y., Zhang, Y., Nielsen, C., McElroy, M.B., Hao, J., 2014. Source apportionment of atmospheric mercury pollution in China using the GEOS-Chem model. Environ. Pollut. 190, 166–175. https://doi.org/10.1016/j.envpol.2014.03.011

Wang, W., Zhao, S., Jiao, L., Taylor, M., Zhang, B., Xu, G., Hou, H., 2019. Estimation of $PM_{2.5}$ Concentrations in China Using a Spatial Back Propagation Neural Network. Sci. Rep. 9, 13788. https://doi.org/10.1038/s41598-019-50177-1

Wen, C., Liu, S., Yao, X., Peng, L., Li, X., Hu, Y., Chi, T., 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

Woody, M.C., Wong, H.-W., West, J.J., Arunachalam, S., 2016. Multiscale predictions of aviation-attributable $PM_{2.5}$ for U.S. airports modeled using CMAQ with plume-in-grid and an aircraft-specific 1-D emission model. Atmos. Environ. 147, 384–394. https://doi.org/10.1016/j.atmosenv.2016.10.016

Yang, X., Wu, Q., Zhao, R., Cheng, H., He, H., Ma, Q., Wang, L., Luo, H., 2019. New method for evaluating winter air quality: $PM_{2.5}$ assessment using Community Multi-Scale Air Quality Modeling (CMAQ) in Xi'an. Atmos. Environ. 211, 18–28. https://doi.org/10.1016/j.atmosenv.2019.04.019

Zhang, B., Jiao, L., Xu, G., Zhao, S., Tang, X., Zhou, Y., Gong, C., 2018. Influences of wind and precipitation on different-sized particulate matter concentrations ($PM_{2.5}$, PM_{10}, $PM_{2.5–10}$). Meteorol. Atmospheric Phys. 130, 383–392. https://doi.org/10.1007/s00703-017-0526-9

Zhao, R., Zhan, L., Yao, M., Yang, L., 2020. A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of $PM_{2.5}$. Sustain. Cities Soc. 56, 102106. https://doi.org/10.1016/j.scs.2020.102106

Zhou, Y., 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, S., Lian, X., Wei, L., Che, J., Shen, X., Yang, L., Qiu, X., Liu, X., Gao, W., Ren, X., Li, J., 2018. $PM_{2.5}$ forecasting using SVR with PSOGSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors. Atmos. Environ. 183, 20–32. https://doi.org/10.1016/j.atmosenv.2018.04.004

Journal Pre-proof