LETTER • OPEN ACCESS

An LSTM-based neural network method of particulate pollution forecast in China

View the article online for updates and enhancements.

ENVIRONMENTAL RESEARCH
LETTERS

CrossMark

LETTER

An LSTM-based neural network method of particulate pollution forecast in China

Yarong Chen[1], **Shuhang Cui**[1], **Panyi Chen**[1], **Qiangqiang Yuan**[2], **Ping Kang**[3] and **Liye Zhu**[1,4,5]

[1] School of Atmospheric Sciences, Sun Yat-Sen University, Zhuhai 519082, People's Republic of China
[2] School of Geodesy and Geomatics, Wuhan University, Wuhan, Hubei 430079, People's Republic of China
[3] School of the Atmospheric Science, Plateau Atmosphere and Environment Key Laboratory of Sichuan Province, Chengdu University of Information and Technology, Chengdu 610225, People's Republic of China
[4] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, People's Republic of China
[5] Key Laboratory of Tropical Atmosphere-Ocean System, Ministry of Education, Zhuhai 519082, People's Republic of China

E-mail: zhuly37@mail.sysu.edu.cn

Abstract

Particulate pollution has become more than an environmental problem in rapidly developing economies. Large-scale, long-term and high concentration of particulate pollution occurs much more frequently, which not only affects human health but also economic production. As PM_{10} is one of the main pollutants, the prediction of its concentration is of great significance. In this study, we present a PM_{10} forecast model based on the long short-term memory (LSTM) neural network method and evaluate its performance of predicting PM_{10} daily concentrations at five representative cities (Beijing, Taiyuan, Shanghai, Nanjing and Guangzhou) in China. Our model shows excellent adaptability for various regions in China. The predicted PM_{10} concentrations have good correlations with observations ($R = 0.81$–0.91). We also achieve great predication accuracy (70%–80%) on predicting the next-day changing trend and the model has the best performance for heavy pollution situation ($PM_{10} > 100$ μg m^{-3}). In addition, the comparison of LSTM-based method and other statistical/machine learning methods indicates that our model is not only robust to different pollution intensities and geographic locations, but also with great potential on pollution forecast with temporal-correlated feature.

1. Introduction

The epidemiological investigation, animal toxicology test and human clinical observation of PM_{10} show that PM_{10} has obvious and direct toxic effects on human health, and can cause extensive damage to respiratory system (Leng *et al* 2017), heart and blood system, immune system and endocrine system (Li *et al* 2002). For every 50 μg m^{-3} increase of PM_{10} daily average concentration, the mortality could increase by 4%–5% on average in the research of Utah River Valley (Maynard 1997). In the meantime, the exposure level of PM_{10} concentration is also important. When the mass concentration of PM_{10} is greater than 100 μg m^{-3}, the mortality rate is 11% higher than that when the mass concentration of PM_{10} is less than 50 μg m^{-3} (Maynard 1997). In addition, as a major component of air pollution, aerosol load trends also have an important impact on the prediction of

climate change (Rotstayn *et al* 2015, Westervelt *et al* 2015, Wang *et al* 2016, Yang *et al* 2017). Therefore, accurate and timely prediction of PM_{10} is of great significance in terms of both climatology and socio-economic development.

In recent years, statistical methods have been used in air pollution predictions and gradually formed a research trend for its advantages of high efficiency, convenience and low cost. In addition to a single numerical prediction model (Li *et al* 2016, Zhao *et al* 2016, Yang *et al* 2017), machine learning methods have also been gradually applied to air pollution prediction (Mallet *et al* 2009, Papaleonidas and Iliadis 2013, Nieto *et al* 2018). Deep learning, as one of the most highly sought after class of machine learning algorithms, shows great potential in fitting nonlinear complex relationships between the influencing factors and the pollution concentrations (Hinton *et al* 2006, Li *et al* 2018), such as the recurrent neural

network (RNN) method, which usually refers to considering the influence of the current pollution concentration of the geographical adjacent area of the interested area (Tong *et al* 2019).

Long short-term memory (LSTM) is a RNN with long-term and short-term memory. In recent years, it has become an effective and scalable model to solve some learning problems related to sequential data. The prediction ability of LSTM has been revealed in atmospheric science recently (Asanjan *et al* 2018, Alhirmizy and Qader 2019, Mohapatra *et al* 2020). These studies show that LSTM has better statistical properties (such as correlation coefficient and mean squared error) than traditional RNN and also has better performance than the single environmental meteorological model (Dai *et al* 2019). Therefore, LSTM method brings us another opportunity of better predicting the PM_{10} changing trend in China.

In recent years, the air particulate pollution in China shows an obviously region-based feature. For example, the sources of particulate pollution in Jing–Jin–Ji Area are similar, and the seasonal distribution shows specific characteristics of regions (Dao 2015). Li *et al* (2017) found that aerosol extinction coefficient in Northeast and Northwest China continued to decline. Around 2006, the trend of the Pearl River Delta and southwest China has indeed changed from increasing to decreasing, while the trend of the North China Plain (NCP) and the Yangtze River Delta (YRD) is still increasing. These results indicate the diversity of PM_{10} in different regions and the complexity of its prediction. There is a significant correlation in PM_{10} among the cities in the Yangtze River Delta region and PM_{10} concentration is also correlated with other pollutants (e.g. NO_2 and SO_2) (Shi *et al* 2008). Similar relationships also appears in Pearl River Delta region (Hu *et al* 2011) and Northwest China (Qiu 2010), which are light and heavy polluted PM_{10} regions, respectively.

This paper aims to explore the potential of a LSTM-based method prediction ability of daily PM_{10} concentration and develop the general prediction model that can be applied to different regions of China. We evaluate the model adaptability in five representative regions of China in section 4. This study is an important progress of neural network method that is applied to air pollution forecast. Although this study focuses on forecasting PM_{10} concentrations, it can play a significant role in generalization to a broader area to forecast other air pollutants.

2. Data

2.1. Data source
We obtain the national urban pollution data from the national urban air quality real-time release platform (http://106.37.208.233:20035/, last access: 22 November 2020). The meteorological data is from the China surface climatic data daily data set V3.0,

which is from the China Meteorological Data Network (https://data.cma.cn, last access: 22 November 2020). We select the national urban pollution data and meteorological data of 359 cities from period 2015 to 2017. We have 1095 data samples for each city in our data set and our LSTM model is based on the time dimension. The details of the data are summarized in table S1 (available online at stacks.iop.org/ERL/16/044006/mmedia). The description of data preprocessing and the temporal feature are in supplemental text S1. The combined data sets for our research are saved at figshare (Zhu 2021).
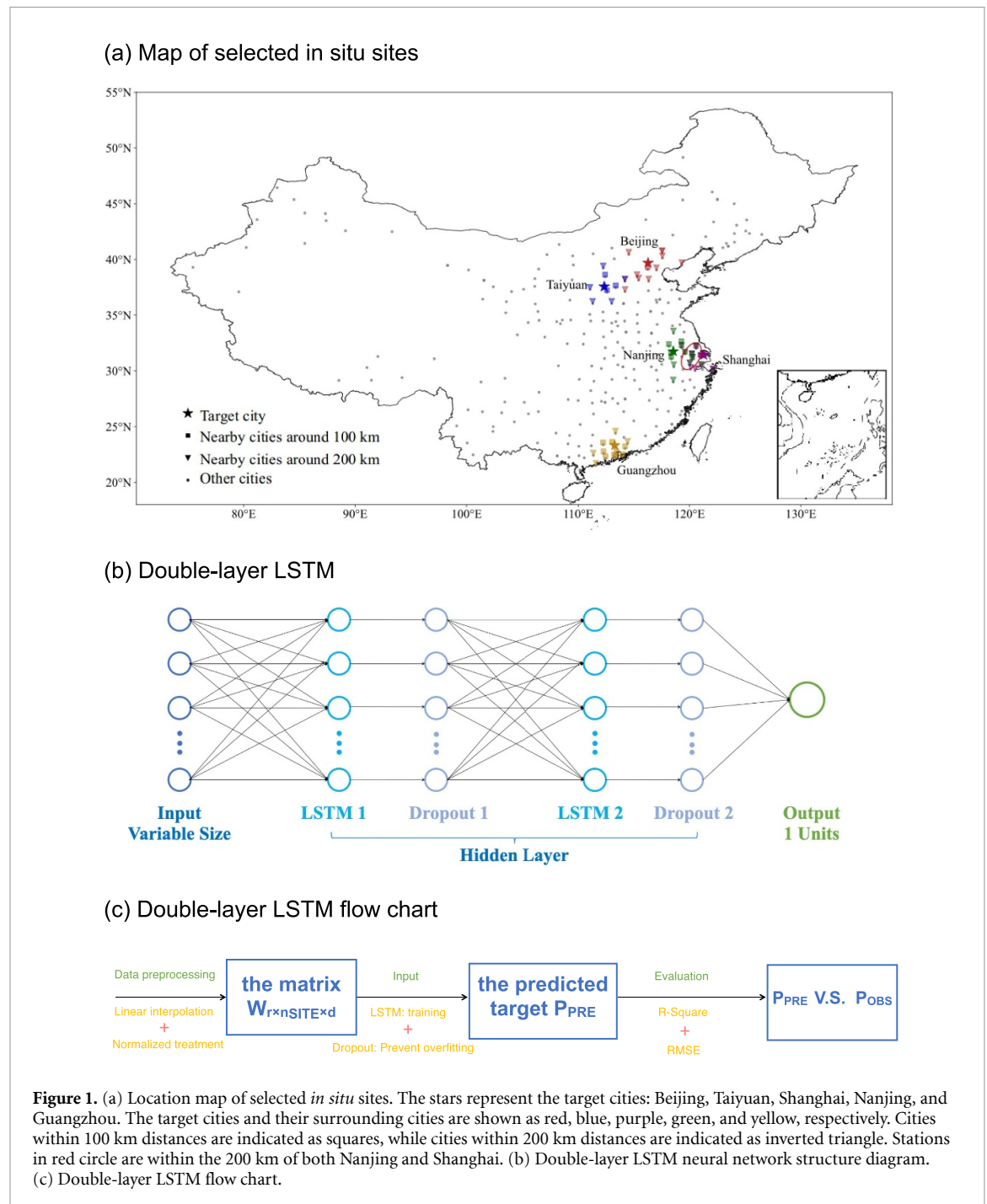
2.2. Spatial patterns of different city metropolitans
Data collection sites are mainly distributed in the eastern region of China, especially the Yangtze River Delta, the Pearl River Delta and the Beijing–Tianjin–Hebei region, with relatively few sites in the western region. Therefore, in this paper, we select a few representative cities (Beijing, Taiyuan, Shanghai, Nanjing and Guangzhou) from different geographic regions (Jing–Jin–Ji, Northwest China, Yangtze River Delta and Pearl River Delta) of China and cities surrounding them within a certain distance (e.g. 100 km, 200 km) for this study (figure 1(a)). Due to the geographical proximity between Beijing and Taiyuan as well as Nanjing and Shanghai, some of the selected cities within 200 km are with overlapped areas (the red circle in figure 1(a)).

3. Methods

3.1. The principle of LSTM
Long short-term memory (LSTM) neural network is a widely used RNN architecture in the field of deep learning. It has a tree-like hierarchical structure, consisting of input layer, hidden layer and output layer (figure 1(b)), while its network nodes are recursive with the input information connected to each other in sequential order. Each hidden layer contains LSTM structure and dropout layer. LSTM structure is the core part of the entire neural network. Each LSTM neuron contains three control gates: input gate, forgetting gate and output gate. The input gate learns to decide when to let the activation signal into the cell, the output gate learns when to let the activation signal out of the cell, and the forgetting gate learns when to let the cell of the last moment into the cell of the next moment. In the forgetting gate, only the information conforming to algorithm authentication will remain, while the information not conforming will be forgotten, thus it effectively solves the problem of long-term dependence (Greff *et al* 2017). In the dropout layer, neural network cells are temporarily discarded from the network with a certain probability during the training process, which is a means to prevent overfitting and has good fault tolerance.

Figure 1. (a) Location map of selected *in situ* sites. The stars represent the target cities: Beijing, Taiyuan, Shanghai, Nanjing, and Guangzhou. The target cities and their surrounding cities are shown as red, blue, purple, green, and yellow, respectively. Cities within 100 km distances are indicated as squares, while cities within 200 km distances are indicated as inverted triangle. Stations in red circle are within the 200 km of both Nanjing and Shanghai. (b) Double-layer LSTM neural network structure diagram. (c) Double-layer LSTM flow chart.

3.2. Model design

In our model, the prediction object is the target pollutant concentration of a central city, and the prediction factors are the historical daily data of meteorological elements and pollution elements of the central city and its surrounding n_{SITE} stations in the past r days. Specifically, the target pollutant is daily PM_{10} in this study. The input factors include d factors such as daily average temperature, daily average humidity, maximum wind speed, maximum wind direction, SO_2 concentration, NO_2 concentration, CO concentration, O_3 concentration for 8 h, $PM_{2.5}$ concentration and PM_{10} concentration (table 1).

We constructed the input factor matrix W, and its dimension is $r \times n_{SITE} \times d$ (figure 1(c)). One or more LSTM cells constitute the middle layer of the prediction model, and the number of layers is denoted as n_{LAYER}. Each LSTM layer contains n_{NEU} neurons and is followed by a dropout layer. The output layer is where we find the predicted target p_{PRE}. The forecaster is denoted as p_{PRE} while the corresponding observed value is denoted as p_{OBS}. After e times of prediction, the prediction sequence p_{PRE} would be achieved. The R-square and root mean square error (RMSE) of p_{PRE} and p_{OBS} are calculated to evaluate the prediction performance of different model configurations.

Table 1. Input factors.

Variable	Unit	Notes
PM_{10}	$\mu g\, m^{-3}$	Target variable to predict with nearby cities
$PM_{2.5}$	$\mu g\, m^{-3}$	
SO_2	$\mu g\, m^{-3}$	
NO_2	$\mu g\, m^{-3}$	
CO	$\mu g\, m^{-3}$	
O_3	$\mu g\, m^{-3}$	
Mean temperature	$0.1\,^{\circ}C$	Unit is 0.1 times of $1\,^{\circ}C$
Mean relative humidity	%	
Great wind speed	$0.1\, m\, s^{-1}$	Unit is 0.1 times of $1\, m\, s^{-1}$
Great wind direction		Numbers are used to represent 16 directions and then be converted to degrees (wind from North is 0°)

During the training process, the optimization function used in the LSTM model is the 'Adam' algorithm (Zhang *et al* 2019) and the loss function is 'Mean_Squared_Error'.

In general, the prediction effect of the model is related to the parameters of *r, n$_{SITE}$, n$_{LAYER}$* and ***n$_{NEU}$***. In this study, we used Python and its deep learning libraries Tensorflow and Keras to build and implement the above neural network algorithms. We set up a control experiment of multiple parameter configurations, and then discuss the capabilities of our LSTM model in different regions of China.

We divided the data set with time length of ***L*** days into two sets, with 70% of the available data allocated as training set and the remaining 30% data as test set. A summary of experiment settings is shown in table S3. We adopted the control variables for the selected representative cities (Beijing, Taiyuan, Shanghai, Nanjing, and Guangzhou) with only one changeable variable in each group of experiments. Finally, we used R^2 and RMSE to evaluate the experimental results to explore the influence of parameters such as the number of layers in LSTM (supplemental text S2), the scope of surrounding cities (supplemental text S3), the days of input data as well as different regions that the model applied to (supplemental text S4). Based on the experimental results, we determined the best model is double-layer LSTM with previous one observed data as input within a certain geographical range (DLP1–LSTM model; DLP1, double-layer with previous ONE observed data).

3.3. Other machine learning methods

To evaluate superiority of the LSTM model in PM_{10} prediction in this paper, we compared it with five other methods including traditional statistical methods and machine learning models: ordinary least squares regression (OLSR), Bayesian ridge regression, support vector regression (SVR), multi-layer perceptron (MLP), and Random Forest Regression. We will briefly introduce these methods in the following paragraphs.

OLSR, which seeks the best match for data by minimizing L2 norm errors, is a mathematical optimization technique (Moutinho and Hutcheson 2011). In this algorithm, OLSR is used to fit a linear model using coefficients in Linear Regression to minimize the sum of squares of differences between the actual observed data and the predicted data.
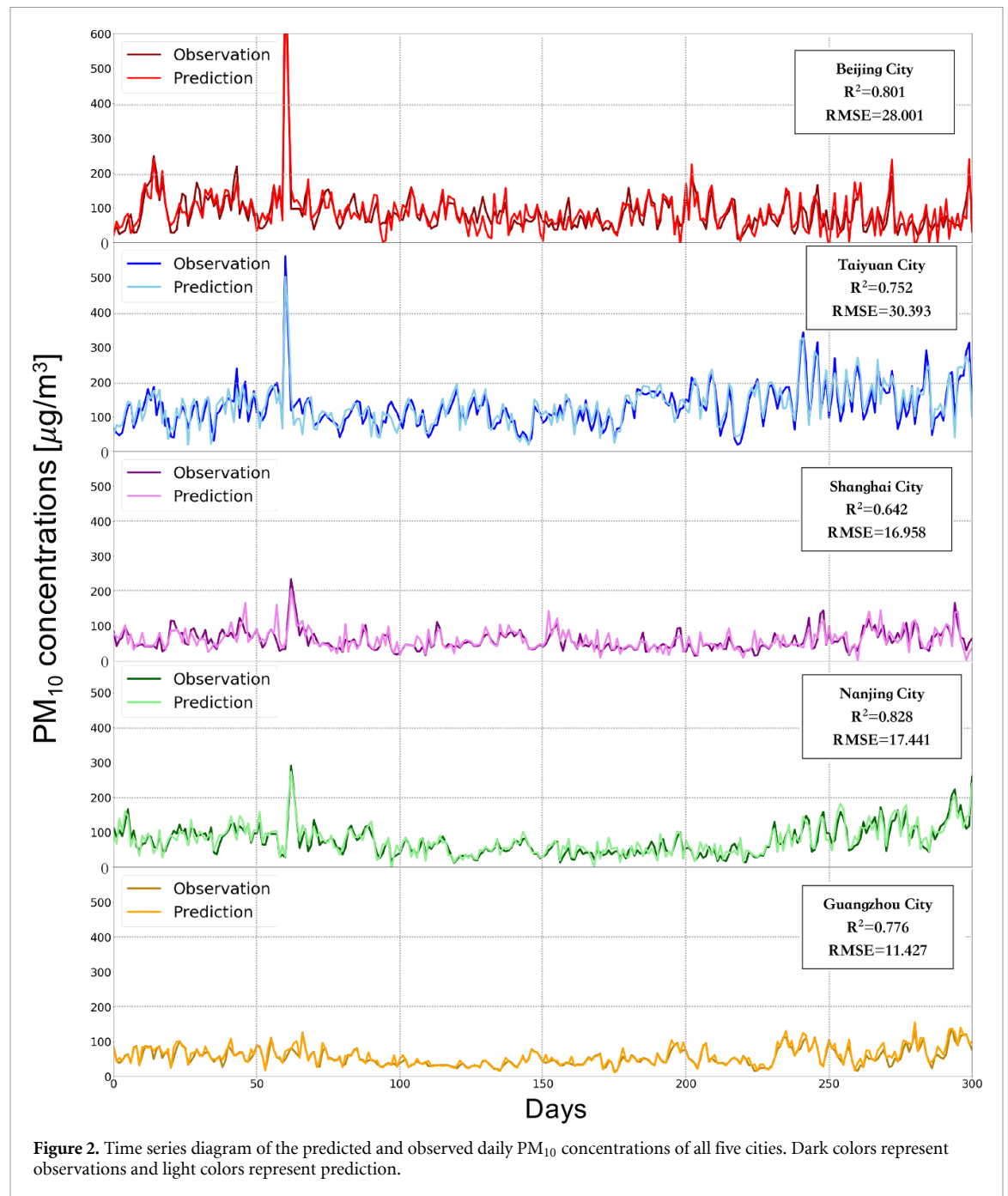
Bayesian Ridge Regression solves these regression problems by imposing a penalty on the Regression coefficients on the basis of OLSR. In other words, a regular term is added to the deviated sum of squares function, which is used to adjust parameters and delete those related terms (Massaoudi *et al* 2020).

SVR is different from traditional regression methods. For SVR, as long as the predication $f(x)$ does not deviate too much from *Y*, the prediction can be regarded as correct and there is no need to calculate the error. SVR is still a convex optimization problem and thus ensures the global solution give it nonlinear prediction ability and generalization performance (Uçak and Günel 2016).

The creation of feedforward neural network of MLP starts from the most basic form of a single perceptron. A sensor has one or more inputs, biases, activation functions, and a single output. The sensor takes the inputs, multiplies them by certain weights, and passes them to the activation function to produce the output. The input layer of an MLP consists of all the variables in the individual neurons, and the output layer consists of response variables. The input layer and the hidden layer include a constant neuron associated with the intercepting synapse or deviation (that is, a synapse not directly affected by any covariable called interception) (Egbo 2018).

Random Forest Regressor is an ensemble method by creating forest with multiple decision trees during training. Each tree contains an arrangement of decision nodes, based on which the tree is divided into different branches until the termination point (leaf) is reached. Each decision node depends on whether the value of the input feature exceeds a certain threshold. The final prediction is obtained by averaging each tree prediction that provided by training it independently (Pillai *et al* 2020).

In this paper, the modules in the Scikit-earn library are used to implement the models for comparison. To ensure reliability of the comparison results, the data preprocessing and the input factors in other machine learning methods are kept consistent with our LSTM model.

Figure 2. Time series diagram of the predicted and observed daily PM$_{10}$ concentrations of all five cities. Dark colors represent observations and light colors represent prediction.

4. Results

4.1. DLP1–LSTM model performance between representative cities

In order to study the applicability of double-layer LSTM model to PM$_{10}$ prediction in different regions of China, we applied our DLP1–LSTM model to several typical metropolitans that represent different features of pollution in China. They are Beijing, Taiyuan, Shanghai, Nanjing and Guangzhou. We set up the configuration of using observation data from cities within 100 km of circle around of the target city and 1 d before the prediction day. The training convergence curves (figure S2) show that the DLP1–LSTM prediction model gradually converges for all

five target cities. This further reflects the good applicability of the prediction model in China.

Figure 2 is the time series diagram of the predictions and observations of all five cities. In general, the predictions match the trends of observations of all five cities. The predictions could match most of the peaks and valleys. We also know from figure 2 that the observed PM$_{10}$ of Guangzhou is relatively lower than that of other cities and has fewer sharp changes. Figure S3 shows that in the representative cities, the predicted values of the DLP1–LSTM is roughly positively correlated with the observations, and the slope of the fitting curve is close to 1. In addition, Nanjing has the highest R^2 (0.828), while Shanghai has the lowest R^2 (0.642) among these five cities. The predictions of

PM_{10} concentrations in these cities all have good correlations ($R = 0.81–0.91$) with observations. Guangzhou has the smallest RMSE, while Taiyuan has the largest RMSE.

In summary, our configuration of double LSTM model is able to make accurate predictions of daily PM_{10} concentrations for various city metropolitans within a wide range of pollution levels in China.

4.2. DLP1–LSTM model performance of next-day trend prediction at different pollution levels

To further explore the predictive power of the DLP1–LSTM model, we analyzed the sensitivities of the model to different levels of pollution and also its performance in the five representative cities of different regions. We divided the pollution into three levels: mild (0–50 μg m^{-3}), moderate (50–100 μg m^{-3}) and polluted (>100 μg m^{-3}), and then tested the prediction accuracy of our model on next-day trend (increase or decrease from the previous day) for different pollution levels. Overall, the model has the best trend prediction capability for PM_{10} concentration in heavy pollution situation (table S7).

Figure S4 compares the next-day changes of predictions to observations. Positive value means the PM_{10} concentration increases when compared to the previous day, while negative value means it decreases when compared to the previous day. Beijing has more days with moderate and heavy pollution, Nanjing has more days with moderate and mild pollution, and Taiyuan has more days with heavy pollution. The prediction accuracy of the next-day trend is much better for polluted cases of these three cities. However, Guangzhou has the most days of mild pollution, and the corresponding correlation of prediction of mild pollution is much better than other cities (figure S3 and table 2). Shanghai has the lowest accuracy of the next-day trend forecast, at about 70%. The prediction ability of the model is weak for mild ($R^2 = 0.12$) and moderate ($R^2 = 0.17$) pollution days.

It can be seen from figure S3 and table 2 that the prediction ability of next-day trend in Shanghai is weaker than that in the other cities. It is due to the large coastal area of Shanghai, and its landform is mostly plain, which is conducive to the diffusion of pollutants. In addition, there is a lot of rain in summer, especially in the Meiyu season. The relative humidity is high and is negatively correlated with the concentration of PM_{10} (Jian *et al* 2019), that is, it is conducive to the settlement of PM_{10}. Moreover, Shanghai has less pollution in general and fewer days of heavy pollution, which reduces the ability of the model to capture its characteristics. Guangzhou, which also faces the sea and has a relatively lower degree of pollution, is special with more hills and high in the northeast and low in the southwest. As a result, the terrain of Guangzhou is not conducive to the diffusion of pollutants and the DLP1–LSTM model is able to capture the data features well.

In general, the prediction accuracy of increasing or decreasing of the next day is about 70%–80% when the PM_{10} concentration of the current day is known, and it varies slightly among different representative cities. The model is more suitable for predicting heavy pollution events (PM_{10} concentration >100 μg m^{-3}) with some regional differences.

5. Comparison of LSTM with other methods

In this section, we compared the prediction performances of LSTM with several traditional statistical prediction methods (OLSR, Bayesian ridge regression, SVR) as well as two machine learning methods (MLP, Random Forest Regression). Here we show the comparison results in Guangzhou as an example. We set up the same input factors for DLP1–LSTM and the other five methods, which are the meteorological factors and pollution factors recorded in the previous day of Guangzhou as well as partial stations (Zhongshan city, Dongguan city, Huizhou city and Foshan city) within 100 km around it (table 1). Similarly, the first 70% of the data set is the training set while the remaining 30% is the test set.
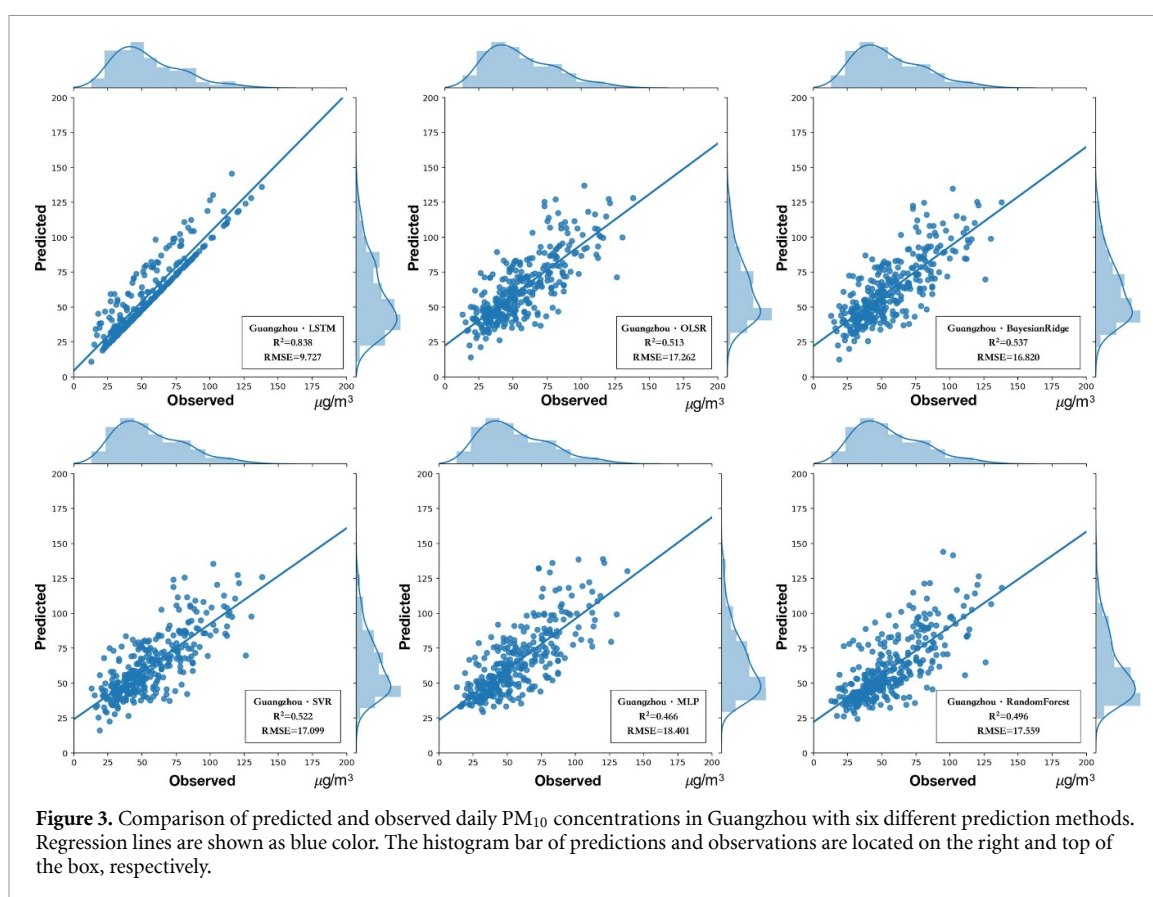
In the case of Guangzhou, the DLP1–LSTM method does a much better job than the other five methods on predicting the daily PM_{10} concentrations with R^2 equals to 0.838 while other methods only have R^2 in the range of 0.466–0.537 (figure 3). From the time series plots (figure S5), it is also obvious that the performance of DLP1–LSTM model is much better than the other five models which show large discrepancies with observations and always miss the peaks, either behind or ahead.

Besides the methods we applied in this section, there are many combined methods applied by other studies in order to better predict the air quality. Guo *et al* (2020) explored the potential of including wavelet method in the artificial neural networks (ANNs) and evaluated the performance of several combined algorithms of wavelet and ANNs. The R^2 of the best prediction case of Air Pollution Index is 0.78 and 0.79 at Xi'an and Lanzhou in China, respectively. Cortina–Januchs *et al* (2015) used a combination of Multilayer Perceptron Neural Network and clustering algorithm to predict the daily PM_{10} concentration at three monitoring stations of Salamanca city in Mexico with R^2 between 0.49 and 0.59. It specified that the combined method of ANNs with clustering algorithms had better generalization capacities than those based on a simple ANN method.

In our study, our LSTM-based method shows obvious superiority in PM_{10} prediction and the good adaptability to different regions of China. In the future, we would like to explore the opportunity of combined methods of LSTM and data analyzing algorithms as the studies we mentioned above.

Table 2. Prediction accuracy of next-day trend on increasing or decreasing for five representative cities at three pollution levels: 0–50 μg m^{-3} (mild), 50–100 μg m^{-3} (moderate), >100 μg m^{-3} (polluted).

The city	Beijing			Taiyuan			Nanjing			Guangzhou			Shanghai		
The levels (μg m^{-3})	0–50	50–100	>100	0–50	50–100	>100	0–50	50–100	>100	0–50	50–100	>100	0–50	50–100	>100
Numbers	73	143	85	18	89	194	104	130	67	157	128	16	158	120	23
Percentage of total(%)	24.33	47.67	28.33	29.67	29.67	64.67	34.67	43.33	22.33	52.33	42.67	5.33	52.67	40.00	7.67
RMSE (μg m^{-3})	24.39	29.56	33.47	19.76	24.76	29.18	14.70	16.05	20.48	9.44	11.78	14.10	17.34	23.15	24.97
R^2	0.13	0.17	0.86	0.14	0.23	0.74	0.34	0.41	0.70	0.51	0.58	0.32	0.12	0.17	0.64
Accuracy		77.30%			76.00%			74.33%			79.00%			69.67%	
R^2		0.79			0.81			0.85			0.86			0.56	
RMSE (μg m^{-3})		29.59			27.45			16.71			10.77			20.47	

Figure 3. Comparison of predicted and observed daily PM$_{10}$ concentrations in Guangzhou with six different prediction methods. Regression lines are shown as blue color. The histogram bar of predictions and observations are located on the right and top of the box, respectively.

6. Conclusions

In this paper, we explored the LSTM neural network method with application in atmospheric particulate pollution prediction. Our DLP1–LSTM model shows excellent adaptability for various regions in China with different geographical conditions and PM$_{10}$ characteristics. First, the predictions of PM$_{10}$ concentrations in cities of different regions all have good correlations ($R = 0.81$–0.91) with observations. Second, the DLP1–LSTM model also performances well on predicting the changing trend of next-day's PM$_{10}$ concentration. The prediction accuracy of whether the next day would increase or decrease is 70%–80%. In addition, by dividing the pollution degree into three levels (mild: <50 μg m^{-3}, moderate: 50–100 μg m^{-3} and polluted: >100 μg m^{-3}), the model has the best trend prediction capability for PM$_{10}$ concentration in heavy pollution situation. This shows great potential for our model to contribute to the prediction, protection and regulation of seasonal concentrated pollution in PM$_{10}$ heavily polluted areas, such as northwest China.

Among various prediction methods (LSTM, OLSR, Bayesian ridge, SVR, MLP and Random Forest), the DLP1–LSTM model shows superior performance than the others, and it indicates the great application prospect of LSTM method on pollution forecast with temporal-correlated feature.

The significance of this study is not only the application of LSTM method for PM$_{10}$ daily concentration prediction, but also the great potential of implementing RNN method on better forecasting particulate matter pollution. In the future, our research will focus more on the adjustment of model structure and hyperparameters tuning in order to improve the spatio-temporal scale of model prediction (e.g. hourly prediction on a wider geographical area). In addition, the capability of DLP1–LSTM model on PM$_{2.5}$ prediction as well as other pollutants is worthy to explore.

Data availability statement

Acknowledgments

data sets for this research are available at Figshare (https://doi.org/10.6084/m9.figshare.13627961).

References

Alhirmizy S and Qader B 2019 Multivariate time series forecasting with LSTM for Madrid, Spain pollution *Paper Presented at 2019 Int. Conf. on Computing and Information Science and Technology and Their Applications (ICCISTA)* (https://doi.org/10.1109/ICCISTA.2019.8830667)

Asanjan A A, Yang T, Hsu K, Sorooshian S, Lin J and Peng Q 2018 Short-term precipitation forecast based on the PERSIANN system and LSTM recurrent neural networks *J. Geophys. Res.* **123** 12–563

Cortina–Januchs M G, Quintanilla–Dominguez J, Vega–Corona A and Andina D 2015 Development of a model for forecasting of PM10 concentrations in Salamanca, Mexico *Atmos. Pollut. Res.* **6** 626–34

Dai Y, Lu Z, Zhang H, Zhan T, Lu J and Wang P 2019 A correction method of environmental meteorological model based on long-short-term memory neural network *Earth Space Sci.* **6** 2214–26

Dao X 2015 Characteristics of mass and ionic compounds of atmospheric particles in winter and summer of Beijing–Tianjin–Hebei area, China *Environ. Chem.* **001** 60–9

Egbo N M and Bartholomew D C 2018 Forecasting students' enrollment using neural networks and ordinary least squares regression models *J. Adv. Stat.* **3**

Greff K, Srivastava R K, Koutnik J, Steunebrink B R and Schmidhuber J 2017 LSTM: a search space odyssey *IEEE Trans. Neural Netw.* **28** 2222–32

Guo Q, He Z, Li S, Li X, Meng J, Hou Z, Liu J and Chen Y 2020 Air pollution forecasting using artificial and wavelet neural networks with meteorological conditions *Aerosol Air Qual. Res.* **20** 1429–39

Hinton G E, Osindero S and Teh Y W 2006 A fast learning algorithm for deep belief nets *Neural Comput.* **18** 1527–54

Hu X, Li Y, Li J, Wang X and Zhang Y 2011 Interaction of ambient PM_{10} among the cities over the Pearl River Delta *Acta Sci. Nat. Univ. Pekin.* **47** 519–24 (http://www.cnki.net/kcms/detail/11.2442.N.20110427.1216.004.html)

Jian W, Xiaobao Q, Jinghua S, Jiafeng L, Chenyan S, Lijun X and Min W 2019 Analysis on variation characteristics of $PM_{2.5}$ and PM_{10} concentration and influence of meteorological factors in Shanghai Chongming Island in 2015 *Meteorol. Environ. Sci.* **42** 1–8

Leng X, Wang J, Ji H, Wang Q, Li H, Qian X, Li F and Yang M 2017 Prediction of size-fractioned airborne particle-bound metals using MLR, BP-ANN and SVM analyses *Chemosphere* **180** 513–22

Li H, Zeng F, Shao L and Shi Z 2002 Current status of study on the human health effects of inhalable particulates *J. Environ. Health* **19** 85–87

Li J, Li C and Zhao C 2017 Different trends between extreme and median surface aerosol extinction coefficients over China inferred from quality controlled visibility data *Atmos. Chem. Phys.* **18** 3289–98

Li T, Shen H, Yuan Q and Zhang L 2018 Deep learning for ground-level $PM_{2.5}$ prediction from satellite remote sensing data *J. Meteor. Environ.* **32** 10–8

Li X, Ma Y, Wang Y, Zhao H and Liu H 2016 Verification and modification to spring air quality forecasted by CUACE system in Shenyang *J. Meteorol. Environ.* **32** 10–8

Mallet V, Stoltz G and Mauricette B 2009 Ozone ensemble forecast with machine learning algorithms *J. Geophys. Res.* **114** D05307

Massaoudi M, Refaat S S, Abu-Rub H, Chihi I and Wesleti F S 2020 A hybrid Bayesian ridge regression-CWT-catboost model for PV power forecasting *Paper Presented at 2020 IEEE Kansas Power and Energy Conf. (KPEC)*

Maynard R L 1997 Smog alert: managing urban air quality *Occup. Environ. Med.* **54** 143–4

Mohapatra M, Nikam B R, Thakur P K, Aggarwal S P and Moothedan A J 2020 Time series gridded precipitation prediction using long short-term memory (lstm) model *Paper Presented at National Seminar on Recent Advances in Geospatial Technology and Application*

Moutinho L and Hutcheson G D 2011 Ordinary least-squares regression *The SAGE Dictionary of Quantitative Management Research* (Thousand Oaks, CA: SAGE) pp 224–8

Nieto P J G, Garciagonzalo E, Sanchez A B and Miranda A A R 2018 Air quality modeling using the PSO-SVM-based approach, MLP neural network, and M5 model tree in the metropolitan area of Oviedo (Northern Spain) *Environ. Model. Assess.* **23** 229–47

Papaleonidas A and Iliadis L S 2013 Neurocomputing techniques to dynamically forecast spatiotemporal air pollution data *Evol. Syst.* **4** 221–33

Pillai M A, Ghosh A, Joy J, Kamal S, Satheesh C C, Balakrishnan A A and Supriya M H 2020 Acoustic source localization using random forest regressor *Paper Presented at 2019 Int. Symp. on Ocean Technology (SYMPOL)*

Qiu Y J 2010 Air pollution characteristics of PM10 in typical cities in Northwest of China *Environ. Monit. China* **26** 65–68

Rotstayn L D, Collier M and Luo J 2015 Effects of declining aerosols on projections of zonally averaged tropical precipitation *Environ. Res. Lett.* **10** 044018

Shi C, Zhai W, Yang J, Shui W and Yao K Y 2008 Characteristics of $PM_{(10)}$ pollution at four provincial cities in Yangtze River Delta District *Plateau Meteorol.* **27** 408–14 (http://en.cnki.com.cn/Article_en/CJFDTOTAL-GYQX200802022.htm)

Tong W, Li L, Zhou X, Hamilton A and Zhang K 2019 Deep learning $PM_{2.5}$ concentrations with bidirectional LSTM RNN, air quality *Atmos. Health* **12** 411–23

Uçak K and Günel G Ö 2016 An adaptive support vector regressor controller for nonlinear systems *Soft Comput.* **20** 2531–56

Wang Z, Lin L, Yang M and Xu Y 2016 The effect of future reduction in aerosol emissions on climate extremes in China *Clim. Dyn.* **47** 2885–99

Westervelt D M, Horowitz L W, Naik V, Golaz J and Mauzerall D L 2015 Radiative forcing and climate response to projected 21st century aerosol decreases *Atmos. Chem. Phys.* **15** 12681–703

Yang G, Deng X, Biwen W U, Huo Y, Zhai J, Caixia Y U and Tian L 2017 Verification of CUACE model in Hefei, Anhui province *J. Meteorol. Environ.* **33** 51–7

Zhang T, Song S, Li S, Ma L, Pan S and Han L 2019 Research on gas concentration prediction models based on LSTM multidimensional time series *Energies* **12** 161

Zhao X, Jing X, Zhang Z, Zhang X, Fan S and Jie S 2016 Beijing regional environmental meteorology prediction system and its performance test of $PM_{(2.5)}$ concentration *J. Appl. Meteorol. Sci.* **27** 160–72

Zhu L 2021 Data_processed (dataset) (https://doi.org/10.6084/m9.figshare.13627961)