Check for updates

A hybrid CNN-LSTM model for predicting PM$_{2.5}$ in Beijing based on spatiotemporal correlation

Chen Ding[1] · Guizhi Wang[1] · Xinyue Zhang[2] · Qi Liu[3] · Xiaodong Liu[4]

Abstract

Long-term exposure to air environments full of suspended particles, especially PM$_{2.5}$, would seriously damage people's health and life (i.e., respiratory diseases and lung cancers). Therefore, accurate PM$_{2.5}$ prediction is important for the government authorities to take preventive measures. In this paper, the advantages of convolutional neural networks (CNN) and long short-term memory networks (LSTM) models are combined. Then a hybrid CNN-LSTM model is proposed to predict the daily PM$_{2.5}$ concentration in Beijing based on spatiotemporal correlation. Specifically, a Pearson's correlation coefficient is adopted to measure the relationship between PM$_{2.5}$ in Beijing and air pollutants in its surrounding cities. In the hybrid CNN-LSTM model, the CNN model is used to learn spatial features, while the LSTM model is used to extract the temporal information. In order to evaluate the proposed model, three evaluation indexes are introduced, including root mean square error, mean absolute percent error, and R-squared. As a result, the hybrid CNN-LSTM model achieves the best performance compared with the Multilayer perceptron model (MLP) and LSTM. Moreover, the prediction accuracy of the proposed model considering spatiotemporal correlation outperforms the same model without spatiotemporal correlation. Therefore, the hybrid CNN-LSTM model can be adopted for PM$_{2.5}$ concentration prediction.

Keywords Convolutional neural networks · Deep learning · Long short-term memory networks · PM$_{2.5}$ prediction · Spatiotemporal correlation

1 Introduction

Air pollution has always attracted substantial attention in environmental sciences (Sun et al. 2017). Long-term exposure to haze has caused various diseases such as lung cancers, heart attacks, and respiratory diseases (Yu and Stuart 2017). Especially, severe haze episodes have erupted in Beijing since January 2013, resulting in excess deaths due to respiratory and circulatory diseases (Gao et al. 2017; Chen et al 2013; David et al. 2014). $PM_{2.5}$ is the most harmful suspended particle to human health. Thus, an accurate prediction approach is essential and positive for decision makers to formulate the prevention measures.

In recent years, the $PM_{2.5}$ concentration prediction approaches have been enriched. Generally, the existing methods designed for $PM_{2.5}$ concentration prediction can be concluded as deterministic methods and statistical methods. Deterministic methods tend to focus on their temporal and spatial evolution process. Specifically, the evolution process consists of emission, dispersion, transformation and diffusion of air pollutants based on meteorological factors and chemical reaction (Bray et al. 2017; Zhou et al. 2017; Woody et al. 2016). In addition, statistical methods widely applied in air pollutant prediction consist of multiple linear regression (MLR) (Donnelly et al. 2015), auto-regression integrated moving average model (ARIMA) (Jian et al. 2012), support vector regression (SVR) (Yang et al. 2018). Nevertheless, the regression models and time series models fail to handle stochastic uncertainty. Thus, the proposed methods have poor performance in extreme points.

In order to handle the shortcomings of linear models, artificial neural networks (ANN) have been employed to predict air pollutants with satisfying performance in recent years. Gennaro et al. (2013) predicted the PM_{10} concentration in two contrasted sites by ANN, respectively. The results proved its availability in air pollutant prediction. To predict the air quality index in Ahvaz, Iran by ANN, Maleki et al. (2019) proved its applicability through the comparison tests. However, the data volume and dimension for model training have been grown rapidly in recent years. A deep learning method as a new artificial intelligence technology has been exploited in different fields such as computer vision (Chan et al. 2015), text processing (Liu et al. 2019) and time series prediction (Wang et al. 2019a, b), etc. Likewise, the deep neural network was applied in air pollutant prediction with excellent performance (Ong et al. 2016; Soh et al. 2018; Li et al. 2016). Previous scholars used LSTM models to conduct the air pollutant prediction (Wen et al. 2019; Wu and Lin 2019). The LSTM model can deal with air pollutant prediction excellently due to its excellent performance in time series problems. Nevertheless, the single LSTM model fails to learn spatial information. Specifically, the air pollutant concentration would change with its emission, diffusion, and reaction with other suspended particles, which indicates the air pollutant is also related to space dimension. A convolutional neural network (CNN) (LeCun et al. 1998) has been proven its strong processing ability in the spatial dimension, which was widely applied in image recognition (Ren et al. 2015). Moreover, the monitoring data in this paper are also spatially relevant. Air pollutants in different areas will affect each other. Thus, the CNN model is a reasonable approach to solve spatial correlation in air pollutant prediction.

Given the limitations of the above methods, a hybrid CNN-LSTM model is proposed, which could handle the air pollutants' complexity and variability. The CNN model can extract spatial features of air pollutants in different cities around Beijing. In this way, it can reflect the spatial effect of different cities when air pollutants diffuse and spread. Then, the output of the CNN model can be used as the input of the LSTM model. Meanwhile, LSTM is used to deal with time series prediction widely. LSTM will achieve better prediction performance due to its strong ability to handle gradient explosion and vanishing problems (Zhang et al. 2018a, b; Zhao et al. 2017). Therefore, the LSTM model is employed to predict the daily average $PM_{2.5}$ concentration by extracting the features of the time dimension.

The remaining part of the article is organized as follows. The relevant literature on the methods of air pollutant prediction is introduced in Sect. 2. Section 3 gives the data description and a specific modeling approach of CNN-LSTM. In Sect. 4, a detailed analysis of the experimental result is given. Finally, Sect. 5 makes a conclusion briefly.

2 Related works

Deep learning methods have been widely applied in the $PM_{2.5}$ prediction instead of conventional prediction models (Ong et al. 2016; Soh et al. 2018; Li et al. 2016). Conventional prediction models consist of deterministic methods and statistical methods. Deterministic methods focus on the emission and diffusion process of air pollutants based on historical data. However, factors such as the lack of prior knowledge and incomplete data may add air pollutant prediction difficulty. Thus, the deterministic methods suffer from low precision and instability. Statistical methods focus on mathematical principles and probability models with flexibility and simplicity. Zhang et al. (2018a, b) utilized the ARIMA approach to predict $PM_{2.5}$ in Fuzhou, China, which indicated that $PM_{2.5}$ concentration experienced seasonal fluctuations. Metia et al. (2016) proposed a hybrid model to overcome the uncertainties related to emission inventory data by integrating a chemical transport model and the Kalman Filter approach.

With the increase of data dimension, the above conventional methods fail to deal with the stochastic uncertainty and have poor performance in predicting the extreme points. Therefore, deep neural network (DNN) as an excellent deep learning method has been adopted widely. A restricted Boltzmann machine was used to predict time series data (Kuremoto et al. 2014). In addition, a deep recurrent neural network (DRNN) was adopted to predict air pollutant concentration with acceptable accuracy.

However, the proposed approaches are usually a single prediction model and ignore air pollutants' spatiotemporal correlation. The prediction performance of a hybrid model outperforms a single model. Based on this viewpoint, a hybrid model called CNN-LSTM is exploited. The CNN model is adopted to extract features, while LSTM can deal with time series prediction well (Huang and Kuo 2018; Qin et al. 2019; Li et al. 2020). Huang et al. (2018) introduced the CNN-LSTM model to predict particulate matter concentration. The proposed model achieved the best

prediction performance compared with other models. However, the above researchers only considered the air pollutant concentration and ignored the impact of air pollutants in different regions. As known to all, the concentration of air pollutants may change with its emission, diffusion, and reaction with other suspended particles. Therefore, it is necessary to consider the spatiotemporal correlation based on this paper's deep neural network.

3 Materials and methods

3.1 Data description

The study area in this paper is Beijing and its surrounding areas, including Tianjin, Hebei, and so on. Figure 1 demonstrates the $PM_{2.5}$ concentration distribution in China in Feb. 2014. It is well known that $PM_{2.5}$ pollution is very concerning in Beijing and its surrounding cities. These areas have experienced industrialization and urbanization over the past years and their geographical location is very close to each other.

In this paper, the historical data from Beijing can be divided into two subsets, including pollutant concentration and meteorological factors. The statistical information of the dataset is shown in Table 1. The dataset contains 1887 samples ranging from Jan. 1st, 2015 to Mar. 1st, 2020. Among them, the pollutant concentration data is collected from the air quality online monitoring platform (https://www.aqist udy.cn/), and the meteorological data is obtained from the weather forecasting website (http://tianqi.2345.com/). Table 1 displays the statistics of different variables. It
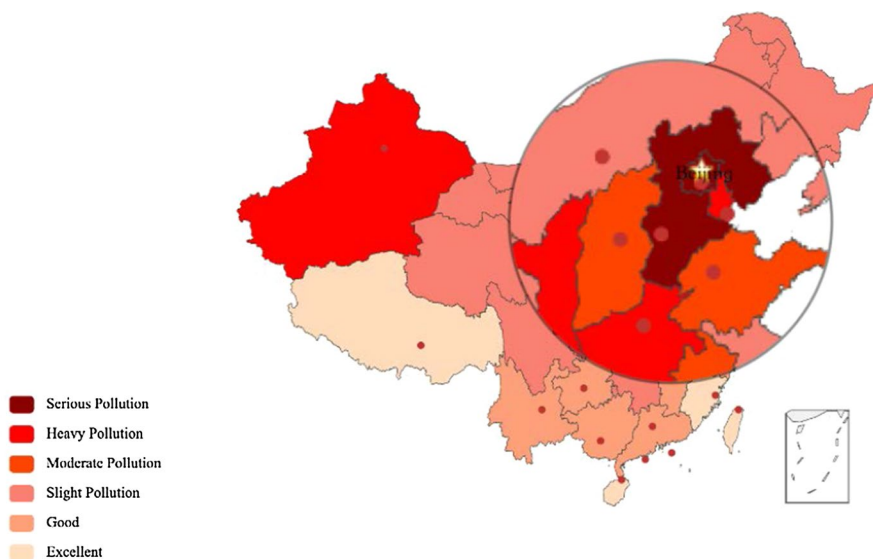

Fig. 1 The $PM_{2.5}$ concentration distribution in China

Table 1 The statistical information of the dataset used in the model

Variable	Data Type	Range	Mean	SD	Unit
$PM_{2.5}$	Numeric	[3, 477]	60.53	56.36	$\mu g/m^3$
PM_{10}	Numeric	[7, 550]	86.97	63.95	$\mu g/m^3$
SO_2	Numeric	[2, 84]	8.03	9.20	$\mu g/m^3$
NO_2	Numeric	[5, 155]	43.76	21.71	$\mu g/m^3$
CO	Numeric	[0.2, 8]	1.00	0.81	$\mu g/m^3$
O_3	Numeric	[2, 311]	97.64	62.71	$\mu g/m^3$
Max Temperature	Numeric	[− 11, 40]	18.58	11.51	°C
Min Temperature	Numeric	[− 16, 27]	8.24	11.12	°C
Wind Direction	Character	–	–	–	–
Wind Grade	Numeric	[0, 6.5]	1.60	1.39	–
Weather	Character	–	–	–	–

is seen that the range of different variables fluctuates wildly. Meanwhile, some character variables need to be converted into numerical variables. Therefore, in order to speed up the model training progress, feature processing techniques are applied as follows:

(1) As shown in Fig. 2, the probability distribution of different continuous variables demonstrates the left-skewed distribution, which is unfavorable for prediction accuracy. Most of the models are based on the assumption of normal distribution. Thus, logarithmic transformation is a good solution of solving data with a biased distribution. The final probability distribution after the logarithmic transformation is shown in Fig. 11.

(2) As for discrete variables, such as wind direction, weather, and wind, an approach called one-hot encoding is utilized to divide into different categories, which is beneficial to modeling.

(3) The present dataset contained 20,757 records for model studying. The dataset is divided into a training set and a test set. We use 80 percent of data as the training set, and the remaining data as the test set to verify the model's effect.

3.2 Spatiotemporal correlation analysis

Due to severe pollution in Beijing and its close geographical location, we consider the spatial correlation of $PM_{2.5}$ concentration from different cities. Pearson's correlation coefficient is a common approach used in measuring the correlation between different variables. The model features can be filtered according to their correlation coefficients. Figure 3 shows the calculation results of variables from different cities. The correlation coefficient values range from − 0.289 to 0.761. It is observed that the further the distance is away from Beijing, such as Henan and Shandong, the smaller the correlation coefficient is. Besides, the correlation coefficient's threshold value is selected as 0.5 for feature selection in this paper. The coefficient is more than 0.5, indicating a significant correlation between variables
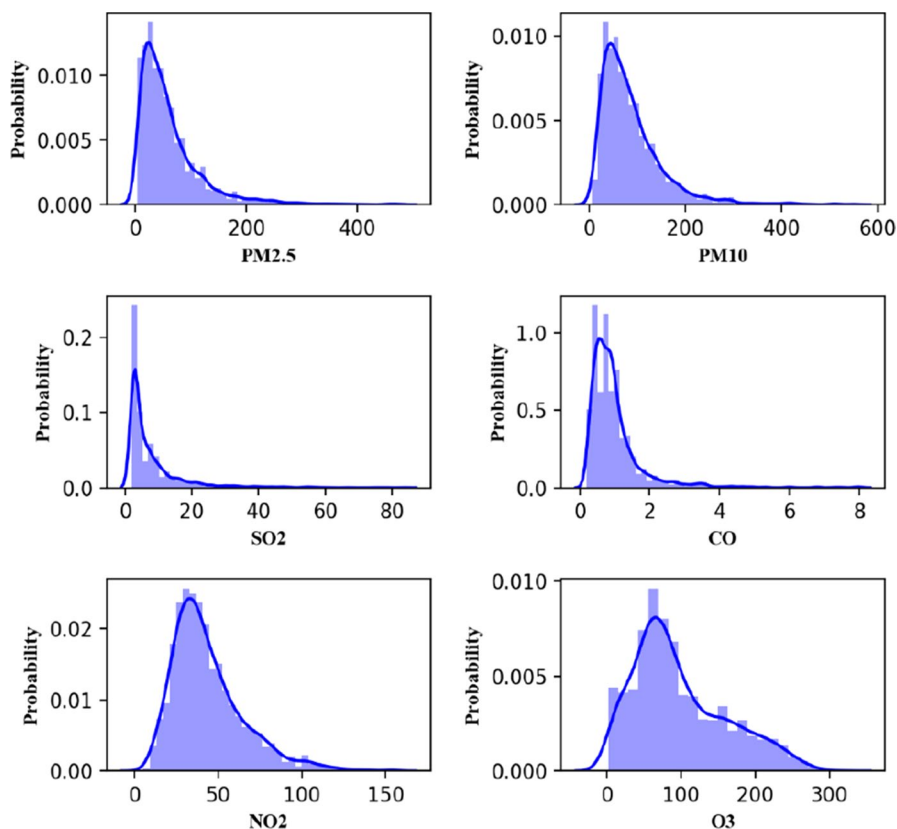
Fig. 2 The probability distribution of different continuous variables

(Li et al. 2017). Apparently, the CO, $PM_{2.5}$ and PM_{10} from Tianjin and Hebei strongly correlate with $PM_{2.5}$ in Beijing. Thus, the spatial correlation provides powerful support for improving the prediction performance instead of establishing a separate model for each city.

Then, we analyze the temporal correlations according to autocorrelation functions. The formula can be written as follows:

$$\rho_k = \frac{Cov(y(t), y(t+i))}{\sigma_{y(t)}\sigma_{y(t+i)}}, i = 1, 2, 3..., \tag{1}$$

where $Cov(\cdot)$ represents the covariance, $\sigma(\cdot)$ denotes the standard deviation, $y(t)$ and $y(t+i)$ represent the target time series at time t and the delayed time series with a time delay i, respectively.

Figure 4 demonstrates the autocorrelation coefficients of $PM_{2.5}$ from different cities. It is obvious that the curve shows a descending trend with the lag time. The trend reflects that the longer the time, the less impact the $PM_{2.5}$ concentration data has on the current state. In addition, the rate of decline is also gradually slowed
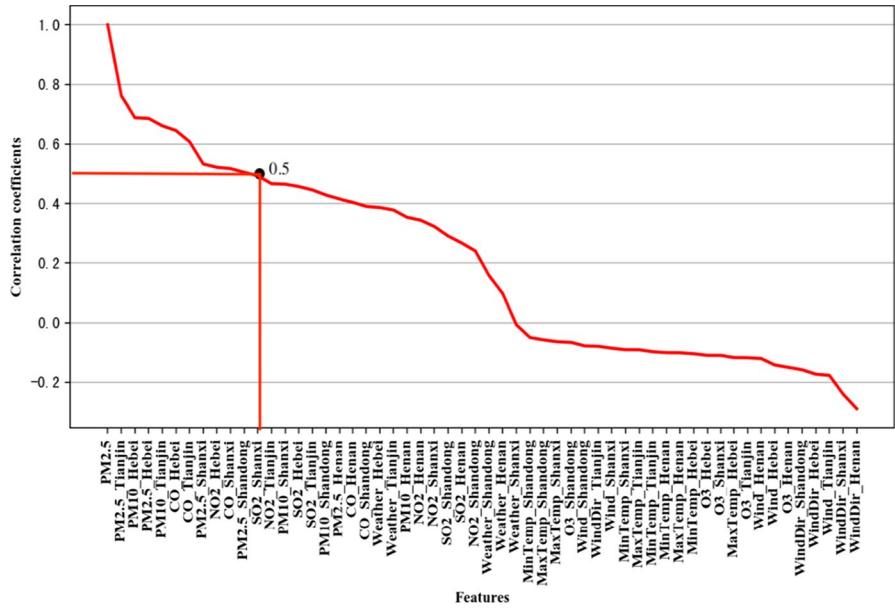
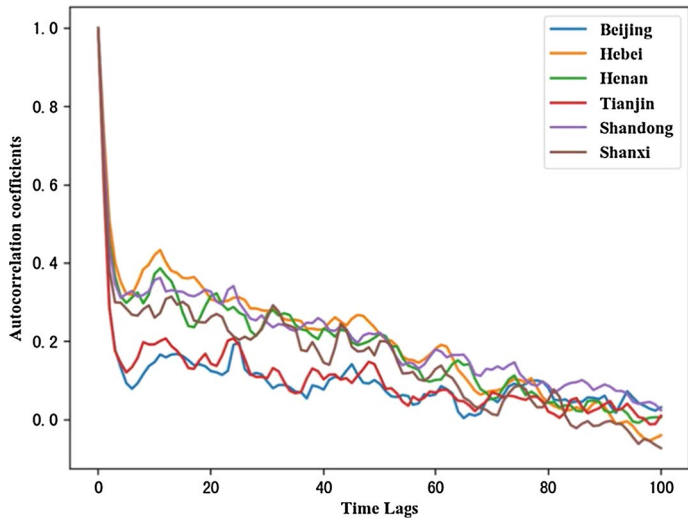**Fig. 3** The correlation coefficients of different features from surrounding cities

Fig. 4 The autocorrelation coefficients of different time lags

down with the increase of the lag time, and the descent speed at the beginning is the largest.

Based on the above research, it is readily observed that PM$_{2.5}$ in Beijing has a significant spatiotemporal correlation with surrounding cities, which is beneficial to prediction accuracy.

3.3 The introduction of the Artificial Neural Network

Artificial Neural Network is an effective mathematical model in the early stages due to its strong capacity of handling nonlinear problems, which simulates the structure of brain neurons. Among them, Multilayer Perceptron (MLP) as a typical neural network structure has been widely applied over the past years. MLP contains the input layer, output layer, and hidden layer. As shown in Fig. 5, the simplest neural structure of MLP consists of one hidden layer. However, with the increase of data volume and feature dimension, the traditional MLP model with a three-layer neural structure cannot achieve good performance. Therefore, popular neural networks such as CNN (Chu and Thuerey 2017) and LSTM (Song et al. 2019) are put forward by increasing network structure complexity. In this study, the CNN and LSTM models are combined to deal with the time series prediction problem.

3.3.1 Convolutional neural network model

Convolutional Neural Network (CNN) comes from the lenet-5 neural network proposed by Lecun in 1998 (Lecun et al. 1998). The proposed network has achieved remarkable recognition performance in the research of handwritten font recognition, which has aroused scholars' close attention. The network structure of the convolutional neural network is shown in Fig. 6.

Different from the traditional neural network model (NN), CNN has multiple feature maps in every layer, and every feature map contains multiple neurons. The
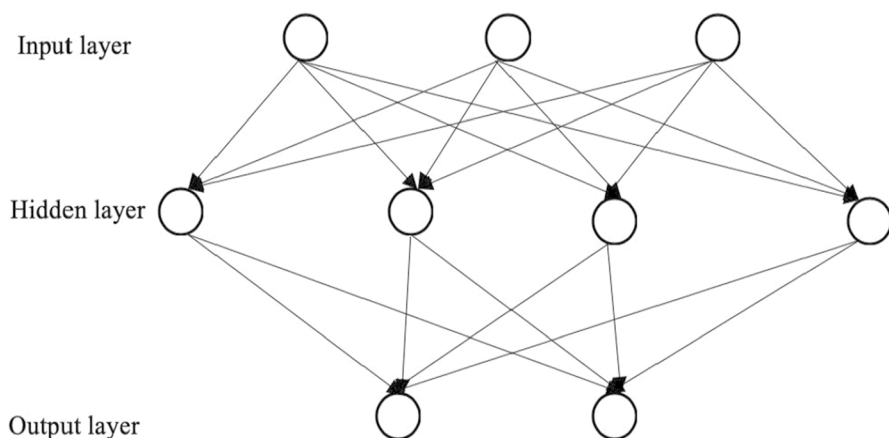

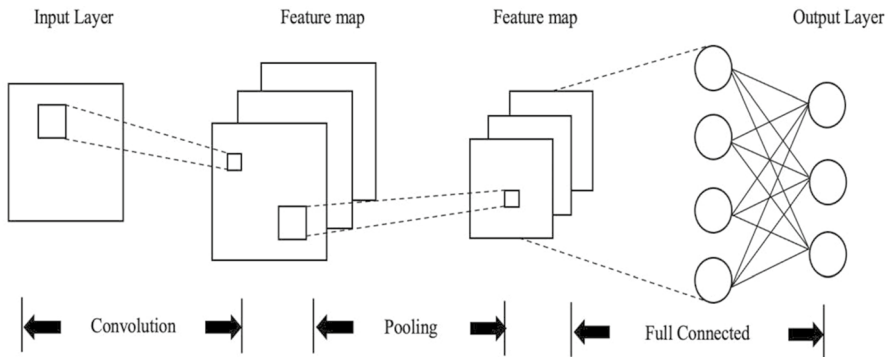**Fig. 5** The specific structure of three-layer perceptron network

Fig. 6 The structure of a simple convolutional neural network

current neuron is convoluted by the output of the upper layer neuron and a convolutional kernel. The convolutional kernel is essentially a defined weight matrix, which is used to extract the features of the local sensing domain.

The structure of a convolutional neural network mainly includes a convolutional layer, pooling layer and fully connected layer. The convolutional layer and pooling layer in the hidden layer are the essential modules of CNN. The convolutional layer is responsible for extracting local features of data while the pooling layer is employed to extract further features based on the down-sampling approach.

Convolutional Neural networks (CNN) can automatically learn features from sequence data, such as text and image data. Its standard network structure contains 1D, 2D and 3D CNN. Given that $PM_{2.5}$ data is one-dimension data, 1D CNN was utilized for feature learning in this study. The specific process of 1D CNN is demonstrated in Fig. 7. The blue part indicates a filter, which represents a sliding window that convolves
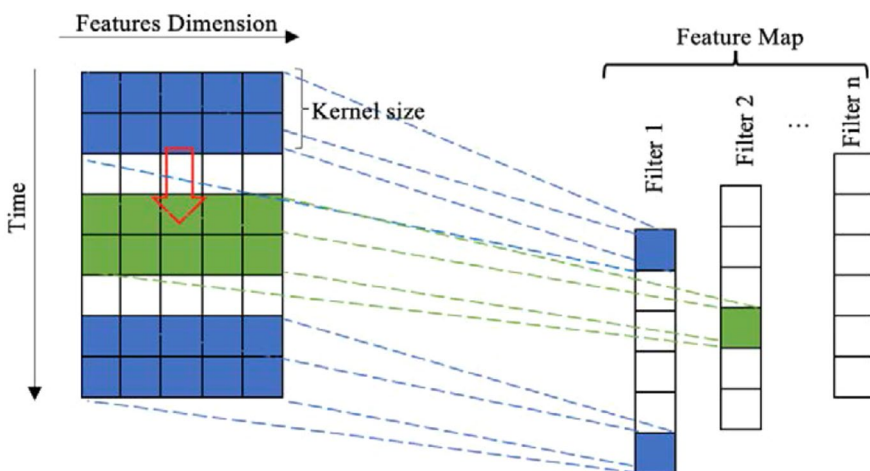

Fig. 7 The learning mechanism of 1D CNN

across the data. The input data and the extracted feature after a sliding window have the same dimension. The green part denotes another filter, and its sliding process is the same as before. Suppose the dimension of input data is M and the number of filters is N, then the total number of the extracted features is M*N (Huang and Kuo 2018).

3.3.2 Long Short-term memory model

Another important neural network widely applied in sequential data is the Recurrent Neural Network (RNN). Unlike other neural networks, RNN tends to focus on the relationship between input data and output data. The basic structure of RNN is shown in Fig. 8.

As shown in Fig. 8, x denotes input data, o denotes output data, U represents weight matrix from input layer to hidden layer, V represents weight matrix from hidden layer to output layer, W represents weight matrix from hidden layer to the hidden layer, s is state value of hidden layer.

However, gradient vanishing problem often occurs in the training process of RNN. Then the training parameters are reduced to zero. Therefore, Long Short-Term Memory Model (LSTM) was introduced to solve the problem of gradient vanishing. LSTM model was first proposed in 1997 and it is a special RNN model (Hochreiter and Schmidhuber 1997). Figure 9 displays the specific network structure of the LSTM model.

As shown in Fig. 9, σ and $tanh$ represent the activation function, where σ is designed to map the value between 0 and 1, while $tanh$ is adopted to map the output between -1 and 1. The formulas of activation functions are written in Eq. (2) and (3).

$$\sigma = \frac{1}{1 + e^{-x}}, \tag{2}$$

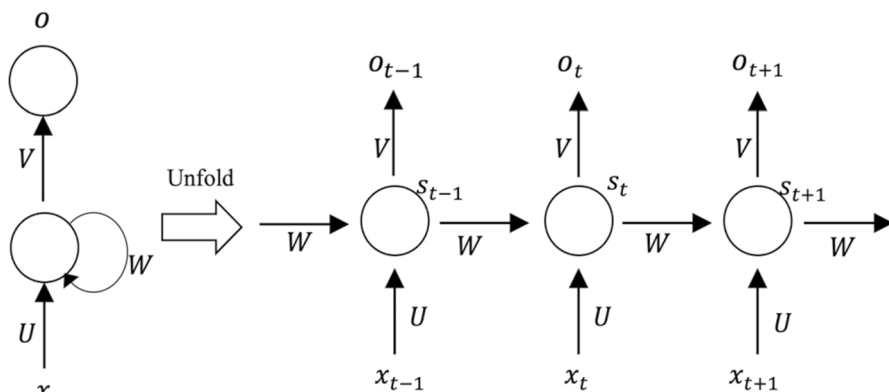$$tanh = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \tag{3}$$

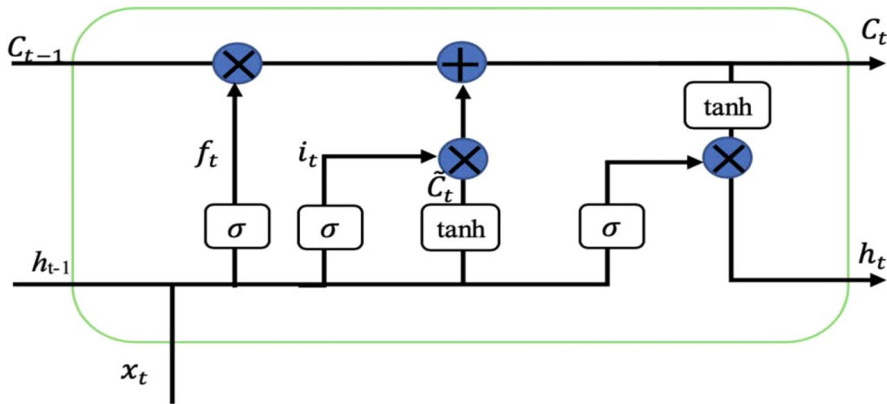**Fig. 8** The structure of a simple recurrent neural network

Fig. 9 The specific network structure of LSTM model

Unlike the internal structure of RNN, the state of LSTM is controlled by an input gate i_t, a forget gate f_t and an output gate o_t. Among them, the forget gate is designed to discard information of the memory cell. The forget gate mechanism receives the output value h_{t-1} of the upper layer and the input value x_t of the current time. Then a probability value C_{t-1} is calculated through the sigma function, which is used to determine the retention of the unit state at the previous time. Also, the input gate is responsible for updating new information to the cell state. Specifically, the probability of state update is controlled according to the output value of σ function, and then a new input value C_t is generated through *tanh* function. The output gate determines to control the output of the external state h_t according to the internal state C_t at the current time. The specific process can be described as Eqs. (4)–(9).

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right), \tag{4}$$

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right), \tag{5}$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right), \tag{6}$$

$$\tilde{C}_t = \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right), \tag{7}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \tag{8}$$

$$h_t = o_t \odot \tanh\left(C_t\right), \tag{9}$$

where W_f, W_i, W_o and W_c represent the weight matrices for input vector x_t. U_f, U_i, U_o and U_c denote the weight matrices from the previous state to hidden state. b_f, b_i, b_o and b_c are bias weights. $\odot$ represents the multiplication of the matrix. x_t is input vector at time t. h_t denotes output vector at time t. C_t represents the cell status at time t.

3.3.3 The hybrid CNN-LSTM model

The hybrid CNN-LSTM model was applied in computer vision and text processing at an early stage. CNN was used as a feature extractor on image and text data, and then input to LSTM for further processing. Likewise, CNN is adopted to extract features of time series data, while LSTM is designed for prediction according to the output from the CNN model in this study.

Figure 10 demonstrates the specific structure of the CNN-LSTM model. A one-dimensional convolutional layer and a pooling layer are designed as the base layer of the hybrid model due to the particularity of time series. In order to input the output of CNN into LSTM, a flatten layer is constructed between CNN layer and LSTM layer. Also, the fully connected layer is constructed to decode the LSTM output. Finally, the prediction results can be obtained from the proposed model.

Aimed at improving the robustness of the model, we use 336 samples as validation set to adjust model parameters and the remaining 28 samples to predict. The parameter selection method is determined by grid search. The specific parameters of CNN-LSTM in this paper are shown in Table 2. Among them, we adopt the *relu* function as an activation function instead of other common activation functions. The *relu* function can solve the problem of gradient disappearance in neural networks due to its special structure. In addition, an efficient parameter optimizer called Adam is utilized in this study instead of the gradient descent approach. In Adam's parameter optimizer, the learning rate of parameters can be dynamically updated. Thus, the parameter has more opportunities to jump out of the local optimum.

The popular performance indices are employed to evaluate the model accuracy, which are expressed as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2}, \tag{10}$$

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{y_t - \hat{y}_t}{y_t} \right|, \tag{11}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{N} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{N} (y_t - \bar{y})^2}, \tag{12}$$

where N is the sample size of test set, $\hat{y}_t$ represents the predicted value of $PM_{2.5}$ at time t, $\bar{y}$ is the mean value of $PM_{2.5}$, while y_t denotes the observed value of $PM_{2.5}$ at time t.
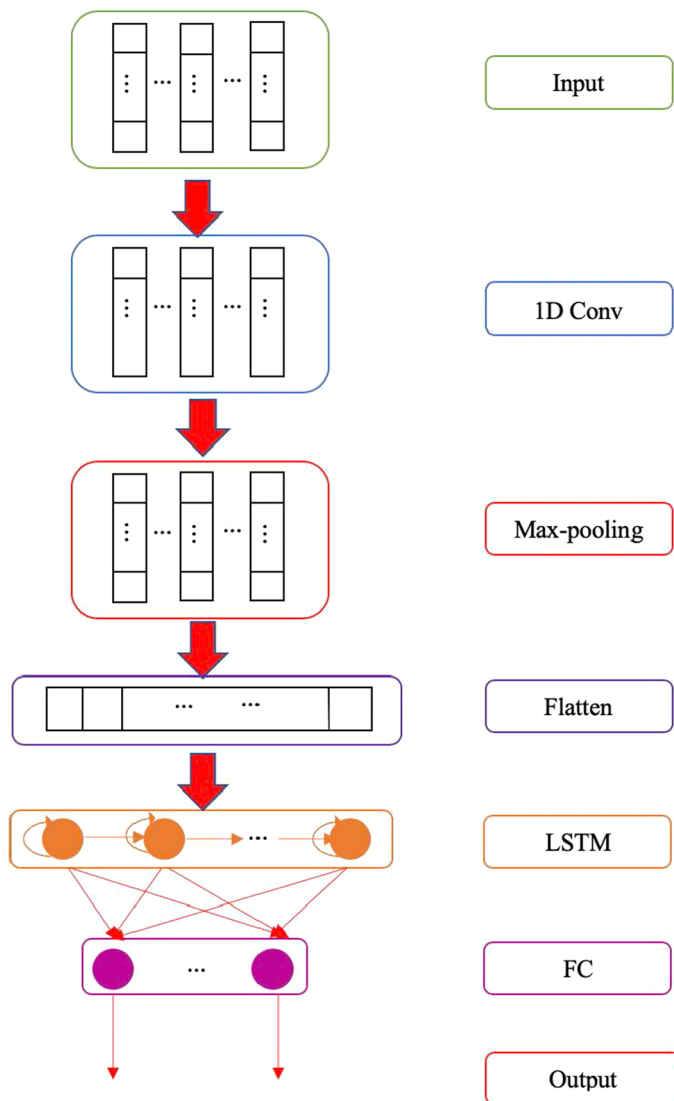
**Fig. 10** The specific network structure of CNN-LSTM

4 Results and discussion

4.1 Prediction performance

The hybrid CNN-LSTM model based on spatiotemporal correlation is conducted to predict the daily average PM$_{2.5}$ concentration from February 2020 to March 2020. Figure 11 displays the prediction performance. It is obviously seen that the predicted

**Table 2** The specific parameters of the hybrid CNN-LSTM model

Parameter	Value
Number of convolutional filters	64
Kernel size of convolutional layer	1
Kernel size of pooling layer	1
Number of pooling layers	1
Number of LSTM neurons	200
Number of epochs	100
Batch size	8
Activation function	*relu*
Training approach for parameters	Adam
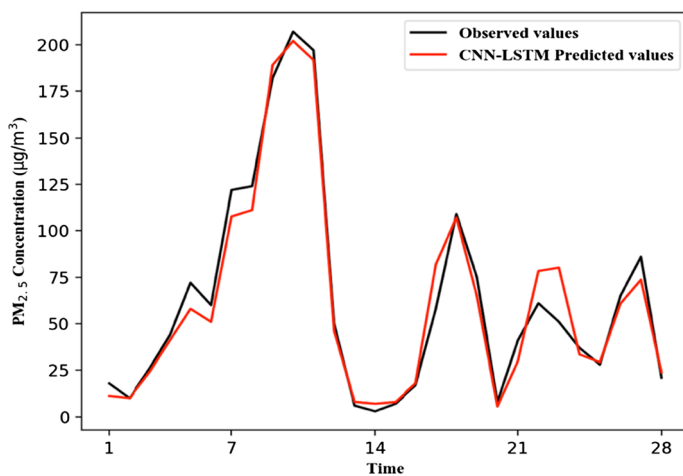
Fig. 11 The daily average PM$_{2.5}$ concentration prediction result

values are close to the observed values in the whole prediction range. The proposed model demonstrates an accurate performance, especially at local high points. This phenomenon indicates that the hybrid CNN-LSTM based on spatiotemporal correlation can deal with nonlinear characteristics and the sudden changes of time series excellently. More specifically, the performance indexes RMSE, MAPE and R^2 of train set are 11.56, 41.91%, 94.72%, while the RMSE, MAPE and R^2 of test set are 10.60, 39.58%, 96.47%, respectively. The above performance indexes indicate that the proposed model obtains high prediction accuracy and avoid the model over-fitting issue. It is strongly proved that CNN can extract the inherent features efficiently and then improve the prediction accuracy of LSTM.

Table 3 The comparisons of different evaluation indexes in prediction performance	Model	RMSE	MAPE (%)	$R^2(\%)$
	MLP	22.16	87.43	84.56
	LSTM	14.84	52.53	93.08
	CNN-LSTM	10.60	39.58	96.47
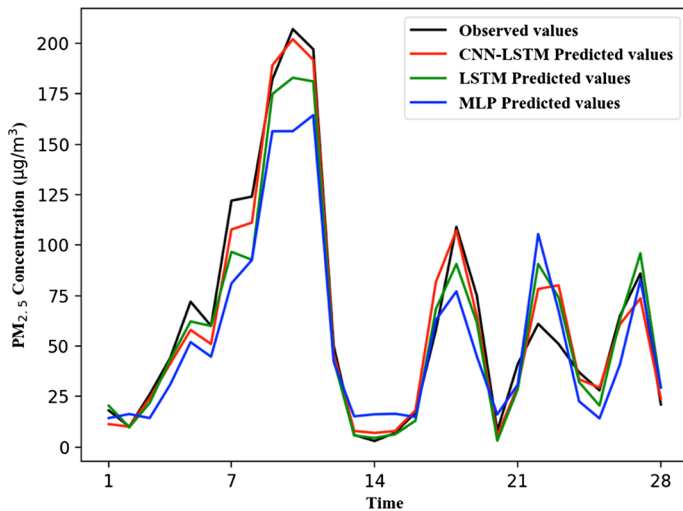
Fig. 12 Comparison of prediction results of different models

4.2 Comparison with other neural network models

To compare different models' performance, we select two commonly used neural networks, including Multilayer perceptron (MLP) and Long Short-Term Memory (LSTM). Among them, MLP was widely used to predict air pollution with excellent performance at early stages. Table 3 shows the prediction performance of different evaluation indexes, while Fig. 12 demonstrates the specific prediction results. It is readily observed that the prediction performance of the hybrid model outperforms the single model. Especially, the forecasting values by CNN-LSTM are consistent with the observed values. In Table 3, the CNN-LSTM model achieves the lowest RMSE and MAPE values, while the highest R^2 value in daily air pollutant prediction. More specifically, the performance indexes of LSTM are RMSE 14.84, MAPE 52.53% and R^2 93.08%, while the RMSE, MAPE and R^2 of MLP are 22.16, 87.43% 84.56%, respectively. It is observed that the prediction accuracy of deep neural network model including LSTM and CNN-LSTM are superior to MLP. Moreover, the prediction performance of CNN-LSTM outweighs LSTM. In general, the above two single models' prediction accuracy is less than that of CNN-LSTM according to the experimental results. In contrast, the CNN-LSTM model makes full use of both models' advantages to well account for the spatiotemporal correlation and reduce prediction error. Therefore,

the hybrid CNN-LSTM model achieves much better prediction accuracy than the proposed neural networks.

4.3 Comparison of the spatiotemporal correlation results

In this section, we train the same model with different data in order to evaluate the spatiotemporal correlation on the prediction performance (Russo and Soares 2014; PSoh et al. 2018). For the former, we train the proposed three different models with the air pollutant concentration data and meteorological factors in Beijing. In the latter case, the above input data is integrated with the air pollutant concentration data in other cities around Beijing. Then, the integrated data is put into the same model. The evaluation results are shown in Table 3 and Table 4. For the same model, the latter obtains the lower RMSE and MAPE values.

Meanwhile, the model considering spatiotemporal correlation has a higher R^2. Specifically, the RMSE, MAPE and R^2 of the CNN-LSTM model without considering spatiotemporal correlation are 16.46, 58.45%, 91.49%, respectively. Apparently, the approach has a higher error compared with the CNN-LSTM with spatiotemporal correlation. By comparing the above results, the hybrid CNN-LSTM model combined with spatiotemporal correlation has less error than other neural network models. It is proved that the spatiotemporal correlation plays an important part for higher accuracy.

5 Conclusion

An effective model with high accuracy and stability is essential to protect humans from suffering from the adverse effects of haze. In this study, a hybrid CNN-LSTM model based on spatiotemporal correlation is proposed to predict the daily $PM_{2.5}$ concentration in Beijing. More specifically, we not only focus on the $PM_{2.5}$ in Beijing, but also its surrounding cities with Beijing due to the fluidity of air pollutants. Moreover, meteorological factors could affect the transmission and diffusion of air pollutants. Thus, it is necessary to consider the meteorological data in model training for better prediction accuracy. To explore the spatiotemporal correlation of $PM_{2.5}$ in Beijing, we adopt Pearson's correlation coefficient in this paper and find air pollutants with high correlation in its surrounding cities. It is shown that the model considering spatiotemporal correlation achieves an excellent prediction performance. Thus, the advantage of the proposed hybrid model is that the CNN model can acquire spatial features in input data while the LSTM

Table 4 The comparisons of different evaluation indexes in prediction performance without spatiotemporal correlation	Model	RMSE	MAPE (%)	R^2(%)
	MLP	27.30	96.79	78.98
	LSTM	18.19	67.85	89.60
	CNN-LSTM	16.46	58.45	91.49

model can deal with the time correlation in time series data. Generally, the CNN-LSTM model is verified to be suitable for $PM_{2.5}$ prediction. More attention could be paid on more training data to verify the generalization of the proposed model in future work. Besides, more meteorological factors related to $PM_{2.5}$ concentration need to be taken into account.
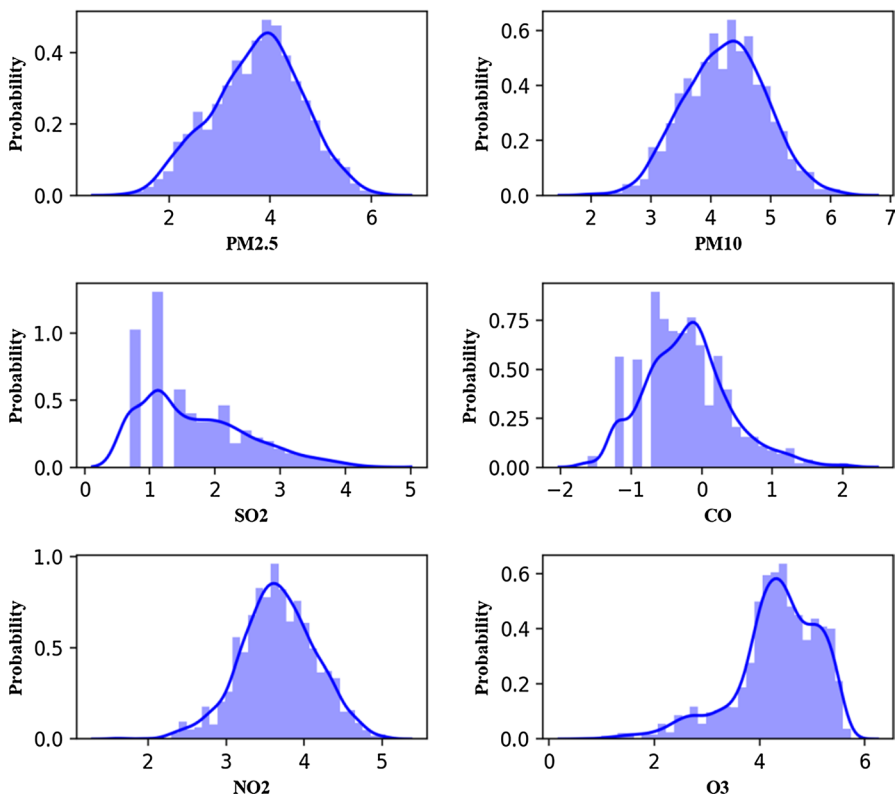
Appendix

See Fig. 13.

Fig. 13 The probability distribution of different continuous variables after logarithmic transformation

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

Bray CD, Battye W, Aneja VP, Tong D, Lee P, Tang Y, Nowak JB (2017) Evaluating ammonia (NH_3) predictions in the NOAA National Air Quality Forecast Capability (NAQFC) using in-situ aircraft and satellite measurements from the CalNex2010 campaign. Atmos Environ 163:65–76

Chan T, Jia K, Gao S, Lu J, Zeng Z, Ma Y (2015) PCANet: a simple deep learning baseline for image classification. IEEE Trans Image Process 24:5017–5032

Chen R, Zhao Z, Kan HD (2013) Heavy smog and hospital visits in Beijing, China. Am J Respir Crit Care Med 188:1170–1171

Chu M, Thuerey N (2017) Data-driven synthesis of smoke flows with CNN-based feature descriptors. ACM Trans Graph 36:69

David YH, Chen SC, Zuo Z (2014) $PM_{2.5}$ in China: Measurements, sources, visibility and health effects and mitigation. Particuology 13:1–26

Donnelly A, Misstear B, Broderick B (2015) Real time air quality forecasting using integrated parametric and non-parametric regression techniques. Atmos Environ 103:53–65

Gao J, Woodward A, Vardoulakis S et al (2017) Haze, public health and mitigation measures in China: a review of the current evidence for further policy response. Sci Total Environ 578:148–157

Gennaro GD, Trizio L, Gilio AD, Pey J, PerezN CM, Alastuey A, Querol X (2013) Neural network model for the prediction of PM_{10} daily concentrations in two sites in the Western Mediterranean. Sci Total Environ 463–464:875–883

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

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

Jian L, Zhao Y, Zhu Y, Zhang M, Bertolatti D (2012) An application of ARIMA model to predict submicron particle. concentrations from meteorological factors at a busy roadside in Hangzhou. China Sci Total Environ 426:336–345

Kuremoto T, Kimura S, Kobayashi K, Obayashi M (2014) Time series forecasting using a deep belief network with restricted Boltzmann machines. Neurocomputing 137:47–56

LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86:2278–2324

Li T, Hua M, Wu X (2020) A hybrid CNN-LSTM model for forecasting particulate matter ($PM_{2.5}$). IEEE Access 8:26933–26940

Li X, Peng L, Hu Y, Shao J, Chi T (2016) Deep learning architecture for air quality predictions. Environ Sci Pollut Res Int 23:22408–22417

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

Liu S, Li Z, Li T, Srikumar V, Pascucci V, Bremer PT (2019) NLIZE: A perturbation-driven visual interrogation tool for analyzing and interpreting natural language inference models. IEEE Trans Vis Comput Graph 25:651–660

Maleki H, Sorooshian A, Goudarzi G, Baboli Z, Birgani YT, Rahmati M (2019) Air pollution prediction by using an artificial neural network model. Clean Techn Environ Policy 21:1341–1352

Metia S, Oduro SD, Duc HN (2016) Inverse air-pollutant emission and prediction using extended fractional Kalman filtering. IEEE J Sel Top Appl Earth Observ Remote Sens 9:2051–2063

Ong BT, Sugiura K, Zettsu K (2016) Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting $PM_{2.5}$. Neural Comput Appl 27:1553–1566

Pak U, Ma J, Ryu U, Ryom K, Juhyok U, Pak K, Pak C (2019) Deep learning based PM$_{2.5}$ prediction considering the. spatiotemporal correlations: a case study of Beijing, China. Sci Total Environ 699:133561

PSoh PW, Chang J, Huang J (2018) Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. IEEE Access 6:38186–38199

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

Qin D, Yu J, Zou G, Yong R, Zhao Q, Zhang B (2019) A novel combined prediction scheme based on CNN and LSTM. for urban PM$_{2.5}$ concentration. IEEE Access 99:1–1

Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. Proc Adv Neural Inf Process Syst 39:91–99

Russo A, Soares AO (2014) Hybrid model for urban air pollution forecasting: a stochastic spatio-temporal approach. Math Geosci 46:75–93

Soh P, Chang J, Huang J (2018) Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. IEEE Access 6:38186–38199

Song X, Liu Y, Xue L, Wang J, Zhang J, Wang J, Jiang L, Cheng Z (2019) Time-series well performance prediction based on long short-term memory (LSTM) neural network model. J Pet Sci Eng 186:106682

Sun C, Kahn ME, Zheng S (2017) Self-protection investment exacerbates air pollution exposure inequality in urban China. Ecol Econ 131:468–474

Wang K, Li K, Zhou L, Hu Y, Cheng Z, Liu J, Chen C (2019a) Multiple convolutional neural networks for multivariate time series prediction. Neurocomputing 360:107–119

Wang K, Qi X, Liu H (2019b) Photovoltaic power forecasting based LSTM-convolutional network. Energy 189:116225

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

Woody MC, Wong HW, West JJ, 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

Wu Q, Lin H (2019) A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors. Sci Total Environ 683:801–821

Yang W, Deng M, Xu F, Wang H (2018) Prediction of hourly PM$_{2.5}$ using a space-time support vector regression model. Atmos Environ 181:12–19

Yu H, Stuart AL (2017) Impacts of compact growth and electric vehicles on future air quality and urban exposures may be mixed. Sci Total Environ 576:148–158

Zhang L, Lin J, Qiu R, Hu X, Zhang H, Chen Q, Tan H, Lin D, Wang J (2018a) Trend analysis and forecast of PM$_{2.5}$ in Fuzhou, China using the ARIMA model. Ecol Indic 95:702–710

Zhang J, Zhu Y, Zhang X, Ye M, Yang J (2018b) Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. J Hydrol 561:918–929

Zhao R, Yan R, Wang J, Mao K (2017) Learning to monitor machine health with convolutional Bi-directional LSTM networks. Sensors 17:273

Zhou G, Xu J, Xie Y, Chang L, Gao W, Gu Y, Zhou J (2017) Numerical air quality forecasting over eastern China: an operational application of WRF-Chem. Atmos Environ 153:94–108

Chen Ding is a Master's graduate from the School of Mathematics and Statistics, Nanjing University of Information Science and Technology, China. His research interests include mathematical statistics and its applications.

Guizhi Wang is Professor at the School of Mathematics and Statistics, Nanjing University of Information Science and Technology, China. Her research interests include mathematical statistics, non-parametric statistical theory, and their applications.

Xinyue Zhang is Assistant Professor at the Institute of Environment and Sustainable Development in Agriculture, Chinese Academy of Agricultural Science. Her research work focuses on climate change impact and adaptation.

Qi Liu received the B.S. degree in Computer Science and Technology from Zhuzhou Institute of Technology, China in 2003, and the M.S. and Ph.D. degrees in Data Telecommunications and Networks from the University of Salford, UK in 2006 and 2010. His research interests include context awareness, data communication in MANET and WSN, and smart grid analysis. His recent research work focuses on intelligent agriculture and meteorological observation systems based on WSN.

Xiaodong Liu received his Ph.D. in Computer Science from De Montfort University, and joined Napier in 1999. He is Reader and currently leads the Software Systems research group in the IIDI, Edinburgh Napier University. He was the Director of Centre for Information & Software Systems. He is an active researcher in software engineering, with international reputation and expertise in service evolution, mobile clouds, pervasive computing, software reuse, and green software engineering.

Authors and Affiliations

Chen Ding[1] · Guizhi Wang[1] · Xinyue Zhang[2] · Qi Liu[3] · Xiaodong Liu[4]

[1]   School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China

[2]   Institute of Environment and Sustainable Development in Agriculture, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[3]   Shandong Beiming Medical Technology Ltd, Jinan 250000, China

[4]   School of Computing, Edinburgh Napier University, 10 Colinton Road, Edinburgh EH10 5DT, UK