

# Novel Convolution and LSTM Model for Forecasting PM2.5 Concentration

Wenfang Zhao<sup>a,b</sup>, Yong Zhou<sup>c,\*</sup>, and Wei Tang<sup>c</sup>

<sup>a</sup>*Institute of Urban Meteorology, China Meteorological Administration, Beijing, 100089, China*

<sup>b</sup>*Beijing Meteorological Information Center, Beijing Meteorological Bureau, Beijing, 100089, China*

<sup>c</sup>*Development and Research Center, China Meteorological Administration, Beijing, 100081, China*

---

## Abstract

Higher levels of PM2.5 concentration are becoming the leading cause of hazy days in China. However, studies have shown that the variations of PM2.5 involve complicated physical and chemical processes, which make their accurate predictions challenging. Meanwhile, the forecast results from numerical models frequently deviate from observation values. The deep learning method is a good substitute for the prediction of mass time series data in the field of meteorology. In the present study, a framework for PM2.5 concentration prediction is presented based on a three-dimensional convolutional neural network (3DCNN) and long short term memory neural network (LSTM). Using preprocessing, correlation analysis, feature extraction, and transformation, spatiotemporal sequence data was generated. In the spatiotemporal feature extraction phase, 3DCNN was used to extract high-level spatial features, and LSTM was used to extract temporal features. In the prediction phase, full connect (FC) was used to combine spatial and temporal features. To examine the efficacy of the proposed model, the PM2.5 concentration data, meteorological observation data, and grid dataset collected at ten observation stations in the Beijing Meteorological Bureau (BMB) were used. After the performance evaluation was compared with several methods including this proposed model, support vector machine (SVM), and the existing PM2.5 forecast system in BMB, root mean square errors (RMSE) and mean absolute errors (MAE) were chosen as evaluation indicators. The experimental results showed that the proposed model performed the best, the minimum MAE value was 3.24 $\mu\text{g}/\text{m}^3$ , and the minimum RMSE value was 13.56 $\mu\text{g}/\text{m}^3$  over the ten stations. In addition, the proposed model overcame the underestimation produced by the existing PM2.5 forecast system in BMB and demonstrated superior performance for different time lengths over a 24-hour period. The results also confirmed the effectiveness of the deep learning method in the prediction of PM2.5 concentration.

**Keywords:** PM2.5 concentration prediction; deep learning; 3DCNN; LSTM

(Submitted on March 20, 2019; Revised on April 14, 2019; Accepted on June 13, 2019)

© 2019 Totem Publisher, Inc. All rights reserved.

---

## 1. Introduction

PM2.5 (particulate matter with an aerodynamic diameter of  $\leq 2.5\mu\text{m}$  suspended in the atmosphere) has been proven to be not only the main criteria pollutant for air pollution, but also the chief factor leading to haze weather [1-3]. A higher concentration of PM2.5 can cause disease to humans, destruction to the throat, lung, and even the brain, and damage to the environment. In recent years, air pollution caused by PM2.5 has received increasing attention worldwide. Jin et al. pointed out that the growth rate of PM2.5 emissions in some provinces of China exceeded 200% from 2005 to 2014 [4]. Donkelaar et al. released a report stating that the concentration of PM2.5 in urban areas of China ranked the top in the world in 2015 [5]. According to the monthly report of 74 cities in China in December 2015, the percentage of days that fell below the national healthy air quality standard accounted for 44.4%, and 39 cities were infected for more than 50% of that year [6]. According to the results of the 2012 Global Burden of Disease Assessment Project, the number of early deaths related to PM2.5 pollution in China was about 1.2 million in 2010, accounting for 1/9 of the total annual deaths of Chinese residents [7]. In [8], studies showed that PM2.5 has a significant positive correlation with the visit rate, hospitalization rate, and mortality of some diseases. Therefore, the accurate prediction of PM2.5 concentration has great significance for strengthening air pollution prevention, making government decisions, and protecting people's daily health.

\* Corresponding author.

E-mail address: [zhouy@cma.gov.cn](mailto:zhouy@cma.gov.cn)

The current forecast methods of PM2.5 concentration are mainly classified into two categories: numerical model methods and statistical forecast methods. The numerical model prediction method is mainly based on aerodynamic theory and physical and chemical processes and is widely applied in the meteorological domain. The Community Multi-scale Air Quality model (CMAQ) developed by the US Environmental Protection Agency is widely used to simulate PM2.5 formation and air pollution research over many countries [9-11]. WRF-Chem is a meteorological-chemical online fully coupled regional air quality model. Zhou et al. developed a numerical forecasting business system based on the WRF-Chem model for the atmospheric environment in East China, and it has many applications in Shanghai [12-13]. The Chinese Unified Atmospheric Chemistry Environment (CUACE/haze-fog), a forecast model developed by the China Meteorological Administration, has a good effect on the prediction of major haze processes, but there is a certain deviation in the peak concentration and visibility of pollutants in the process [14]. The Beijing Regional Environmental Meteorology Prediction System (BREMPS) had a good forecast effect on PM2.5 concentration in Beijing and surrounding areas in 2014. The correlation of most observation stations was above 0.6, but the forecast effect decreased sharply after 24 hours [15]. In general, these models consider physical and chemical processes comprehensively, but due to the large uncertainties in the parameters used in the model, the forecast results are also uncertain [16].

In order to alleviate the error caused by uncertainty in the numerical model prediction method, the statistical forecast method has been presented as an alternative exploration for prediction making, pattern recognition, classification, and regression analysis. SVM, multiple linear regression, artificial neural networks (ANN), genetic algorithms, and wavelet analysis are commonly used methods. In the field of atmosphere science, SVM and ANN have been used for the mass concentration prediction of atmospheric particles, such as PM2.5 and PM10, in which effective algorithms and results have been proposed [17-21]. The advantages of the statistical forecast method are data driven to void sophisticated theoretical models. However, because of their mapping from nonlinear to linear representations, the predictions are of variable accuracy, and limited parameters and variants cannot extract better feature information from actual complex systems.

Deep learning is a novel machine learning method in the field of artificial intelligence. It can effectively recognize the representative characteristics from a large amount of input data and also provides new research ideas and methods for the prediction of mass time series data in the field of meteorology. The main neural network models for deep learning are CNN, recurrent neural network (RNN), LSTM, and anti-neural networks. In recent years, some scholars have tried to apply these network models to challenging air pollution forecasting problems, which so far have not benefited from sophisticated machine learning techniques. However, few applications have considered visibility and gridded meteorological elements observation data. In fact, many researchers have revealed that there is a strong relationship between PM2.5 concentration and visibility, and PM2.5 concentration shows strong correlations in both time and space. The gridded data can be used to extract high-level representative features better than discrete point data from stations.

The present study proposed a novel prediction model based on 3D-CNN and the LSTM neural network. In order to take into account the spatiotemporal correlations between the PM2.5 concentration and meteorological factors, the historical concentration and meteorological observation data of the present station and gridded meteorological observation data centered at this station were entered into the model. In order to keep all data consistent, each kind of data was hourly, which was important for accurate feature extraction. The gridded meteorological element data at each observation time was regarded as an image. All the data from the past time period was regarded as serialized images, and then the spatial features were extracted using 3D-CNN. The LSTM neural network was used to extract time features from the sequence data of the present station. Finally, all the features were put into the fully connected layer to obtain the final prediction result.

The main contributions of this paper are as follows: (1) the surrounding meteorological environment of each station was fully considered, which meant that gridded neighboring stations with high correlation were selected for each station. (2) The visibility data was taken into consideration as feature vectors. (3) The meteorological gridded data and station data were integrated processed. To a certain extent, gridded data surrounding the station can not only provide more relative feature information than k-nearest neighboring stations, but also reflect surrounding weather conditions, which can help the model better track and predict sudden changes in PM2.5 concentration. (4) The proposed method can efficiently extract better spatiotemporal correlation features and achieve high accuracy for PM2.5 concentration prediction of different temporal scales at different locations.

## 2. Related Work

Deep learning methods can be powerful in extracting representative features without much professional knowledge, and many studies have focused on forecasting of air pollution and predicting PM2.5 concentration using applied deep learning methods. Shi et al. proposed a convolution-LSTM (Conv-LSTM) model in 2015, which successfully transformed the precipitation now-casting problem into a spatiotemporal sequence prediction problem. The experimental results show that

Conv-LSTM captures better spatiotemporal correlation and is better than full connectivity-LSTM (FC-LSTM) and the real-time optical flow by variation methods for echoes of radar (ROVER) algorithm [22]. Zhao et al. used historical air quality data and meteorological data to propose a FC-LSTM to predict PM2.5 pollution of specific air quality [23]. Chaudhary et al. presented a multi-layer LSTM to predict the concentration of future air pollutants [24]. Yan et al. constructed two prediction models of multi-level LSTM and encoder-decoder (Encoder-Decoder). The experimental results show that the LSTM model is better than the Encoder-Decoder framework [25]. Huang et al. developed a deep neural network model with historical hourly precipitation, wind speed, direction, and PM2.5 concentration data as the input. First, multiple one-dimensional convolution processing was performed, and the result was input into LSTM to predict PM2.5 concentration [26]. In addition, some scholars have adopted the LSTM model with feature space correlation to apply to PM2.5 concentration prediction. For example, Wen et al. proposed a new space-time convolution long-term memory neural network, and the PM2.5 concentration data from the current station and the nearest neighbor stations was input into the model after 1D convolution operation [27]. Ping et al. developed an LSTM prediction model by using PM2.5 concentration data, meteorological data, and terrain data of the current station and neighboring stations, and terrain data was also used [28]. However, the number of air quality monitoring stations in large cities is very limited, and the distribution may be uneven. The two nearest neighbors may be distributed in different areas. Adding such adjacent stations to the model may lead to inaccurate prediction [29-31].

The gridded meteorological elements observation data has taken into account the complex terrain and dynamic downscaling factors, which could reflect the spatial characteristics more objectively and accurately than the site data. There was a strong correlation between the meteorological elements of the past time period and the weather of the future moments. To sum up, CNN is suitable for solving the problem of strong correlation between input images. It can learn image features, change rules from many historical images, and predict the output image. The 3DCNN model extracts features from spatial and temporal dimensions and then performs 3D convolution to obtain more available information.

### 3. Data and Method

#### 3.1. Data Description Structure

This study was carried out in the city of Beijing, which has a typical warm temperate semi-humid continental monsoon climate often characterized by short spring and autumn. Summers are typically hot and rainy, and winters are usually cold and dry. The smog weather often occurs in the autumn and winter. The data used in this paper was collected from ten stations of meteorological observation network, situated in both urban and suburban area in Beijing. Both weather conditions and PM2.5 concentration are measured at all stations, which is one of the most advantages of meteorological observation networks. In contrast, in many other cities of China, weather parameters are usually not measured at the PM2.5 monitoring sites; we can only use the data of corresponding automatic weather stations located near the PM2.5 monitoring sites. Data was recorded from 2015 to 2017, including the meteorological dataset and PM2.5 concentration dataset. Each PM2.5 record contains the hourly air quality data in Beijing, while each meteorological record contains the values of hourly weather conditions including wind speed, temperature, precipitation, humidity, pressure, and visibility. Gridded meteorological elements observation data is updated hourly with a spatial resolution of 1 km.

#### 3.2. Time Series Analysis of PM2.5 Concentration in Beijing

Based on the hourly weather condition data of ten national meteorological stations in Beijing from 2008 to 2018, the yearly variability and trends of average number of haze days that occur every year is studied and shown in Figure 1. It is obvious that the number of haze days at the stations located in mountainous areas is lower than the average of all stations, and the number of haze days at the stations in plains area are higher than the average level. Moreover, the pattern shows that the peak of the number of haze days appeared in 2014 at all stations except Haidian, and the highest value reached 182 days at the Yanqing station, which is unexpectedly located in a mountainous area. The peak day at the Haidian station occurred in 2016.

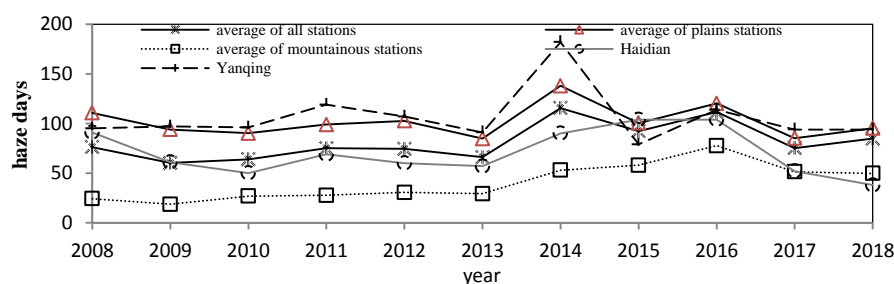


Figure 1. Comparison of haze days in different years over all stations

A haze weather course occurred from December 2 to December 3, 2018. The two nearest stations in the Haidian District were selected, and their trend of PM2.5 concentration over time is shown in Figure 2. The peaks of PM2.5 concentration at these two stations are very different, the trends are not nearly the same, and the correlation between time and space is not significant. It can be seen that there is a significant spatial difference in the concentration of PM2.5. For large cities with sparsely polluted air pollutant monitoring sites and uneven distribution, the spatial feature correlation acquired from gridded data is more accurate and reliable than from discrete sites.

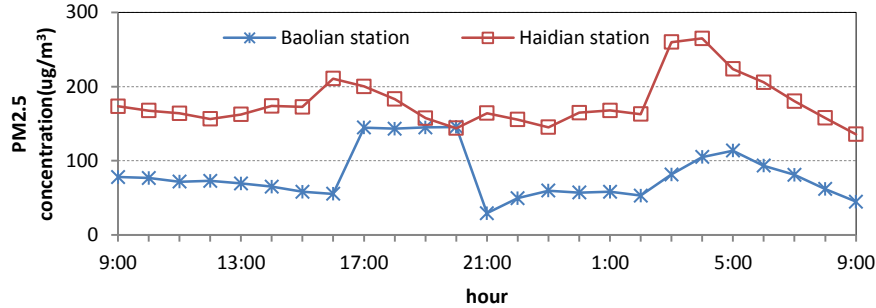


Figure 2. Comparison of PM2.5 concentration between two stations during one haze weather course

### 3.3. Correlation Analysis of PM2.5 Concentration and Meteorological Factors

The impact of meteorological factors on PM2.5 concentration is very complicated, often as a result of the interaction of different meteorological factors. This means that the trend of a single factor of weather condition may not directly lead the trend of PM2.5 concentration. If each factor is considered separately, the coupling effect caused by the interaction of many factors on the PM2.5 concentration cannot be reflected accurately, which will affect the accuracy of the prediction model. Therefore, to quantify the dependence between PM2.5 concentration and meteorological factors as best as possible, a variety of meteorological factors are used, including hourly maximum temperature (T-max), hourly minimum temperature (T-min), hourly temperature (T-mean), hourly wind speed at a height of about two meters from the ground (WS-2m), hourly wind speed at a height of about ten meters from the ground (WS-10m), maximum hourly wind speed (WS-max), hourly precipitation (Pr), hourly relative humidity (RH), maximum hourly relative humidity (RH-max), and hourly visibility (VIS). Pearson correlation analysis is conducted with PM2.5 concentration and meteorological factor variables to determine which factors have a significant impact on hourly PM2.5 concentration, and most correlated factors will be selected as input data for the prediction model.

Correlation coefficients of different meteorological factors of PM2.5 concentration are presented in Figure 3. Through this figure, we can intuitively find that the T-max, T-min, T-mean, HR, and RH-max have high positive correlation coefficients for most of the time. Other meteorological factors have negative correlation coefficients, which is consistent with the conclusions stated in [32]. The concentration of PM10 is strongly correlated with the concentration of PM2.5, while the concentrations of O<sub>3</sub> and SO<sub>2</sub> are almost irrelevant to the PM2.5 concentration.

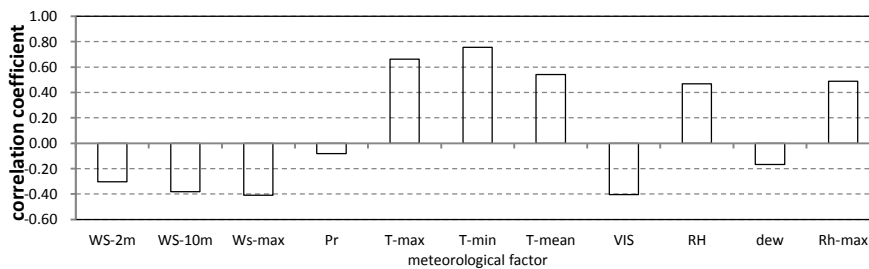


Figure 3. Correlation coefficients between PM2.5 concentration and meteorological factors

### 3.4. Predict Model for PM2.5 Concentration

#### 3.4.1. Formulation of PM2.5 Concentration Predict Problem

The general goal of the PM2.5 concentration forecast is to use the previously observed sequence data with fixed durations (e.g., 24 hours, 48 hours, 72 hours) to forecast a fixed length of the future PM2.5 concentration in a local station or site. All the observed data is updated hourly. Supposing we gain N different observation stations, the input dataset can be denoted as

$N$  time series,  $ST = \{st_1, st_2, \dots, st_n\}$ , where  $st_n = \{x_{t-L}, x_{t-L+1}, \dots, x_t\}$ , (total  $L$  time steps), and each observation  $x$  in the sequences is a  $d$ -dimensional vector consisting of the concentration of pollutants and observed meteorological elements. The gridded meteorological elements observation data divides the whole Beijing area into an  $M \times N$  grid, which consists of  $M$  rows and  $N$  columns. Inside each cell in the grid, there are  $p$  measurements that vary every hour. For a given time  $t$ , the observation of the whole Beijing area can be represented by a tensor  $R^{P \times M \times N}$ , where  $R$  is denoted as the domain of the observed features. The gridded meteorological elements observation dataset of the previous  $L$  hours can be denoted as sequence tensors  $RT = \{\mathcal{R}_{t-L}^{P \times M \times N}, \mathcal{R}_t^{P \times M \times N}\}$ . Therefore, for the target station  $n$ , its optimal spatial feature can be found from  $RT$ , its temporal feature can be extract from  $ST$ , and a fixed time length  $(t+1, t+2, \dots, t+K)$  of the future PM2.5 concentration can be predicted based on these spatiotemporal features. From the perspective of machine learning, this problem can be seen as a space-time sequence prediction problem. In our proposed 3D-CNN and LSTM methodology, the goal is to forecast PM2.5 24 hours in advance for the given station.

### 3.4.2. Predict Model for PM2.5 Concentration

The PM2.5 concentration prediction model consists of two main modules: 3D-CNN for acquiring spatial features and LSTM for acquiring temporal features. The data used includes hourly air pollutant concentration data and meteorological observed data from ten stations of the Beijing Meteorological Bureau, as well as gridded meteorological elements observation data from Rapid-refresh Multi-scale Analysis and Prediction System-Integration (RMAPS-IN), which was developed by the Institute of Urban Meteorology. Although the gridded meteorological elements observation data includes many different elements, according to the above correlation coefficients of different meteorological factors of PM2.5 concentration, we only chose temperature, relative humidity, and wind speed elements from the gridded meteorological elements observation data. Moreover, in order to obtain more accurate spatial features, a  $112 \times 112$  grid centered at the station location was cropped from the entire gridded meteorological elements observation data. Then, 3D-CNN was used to obtain spatial features of temperature, relative humidity, and wind. The LSTM model for acquiring time features consists of a two-layer LSTM structure with 128 nodes per layer. Hourly air pollutant concentration data and meteorological observed data from stations were used as the feature vector input into the LSTM mode. A full connected layer was used to integrate the spatiotemporal eigenvectors. After one or more fully connected layers, the PM2.5 value of the target station at time  $T$  was obtained. The network framework of the entire model is shown in Figure 4.

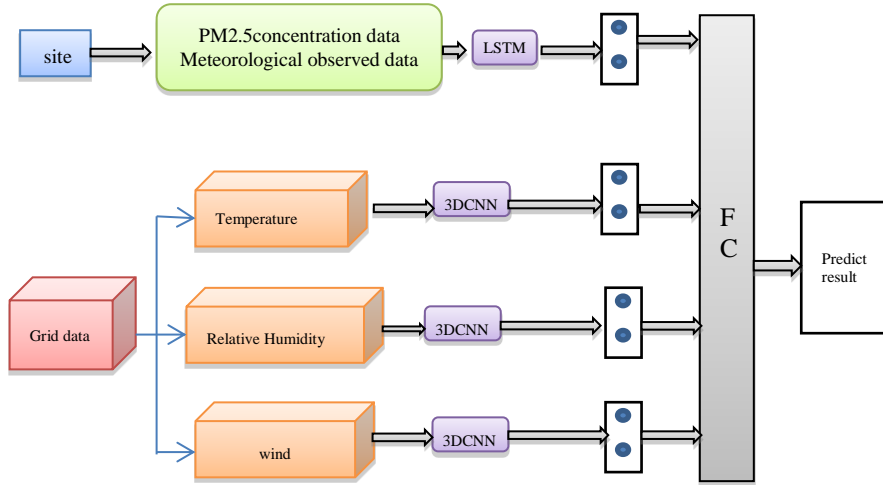


Figure 4. Network framework of predict model for PM2.5 concentration

In order to evaluate the performance of the model, root mean square error (RMSE), mean absolute error (MAE), and mean absolute percent error (MAPE) were used as evaluation indicators. RMSE and MAE are used to measure the overall closeness of the predict result with the target values and evaluate absolute error, while MAPE is used to measure relative error. RMSE and MAE can reflect predicted extreme value effects and error ranges, and MAPE can reflect the specific prediction of the mean value [32].

### 3.4.3. 3DCNN Component

After convolution, sampling, activation function, and flattening, a one-dimensional column vector is obtained as the output

of the spatial features for each meteorological element. 3DCNN only needs to learn the short-term space-time features that are highly correlated with the site, so it does not need to be very deep. Similar to the design in literatures [33], only four convolutional layers and three pooling layers are therefore constructed, as displayed in Figure 5. The kernel size of each convolutional layer is  $3 \times 3 \times 3$  with stride  $1 \times 1 \times 1$ . The kernel size of the first layer of the pooling layer is  $1 \times 2 \times 2$  with stride  $1 \times 2 \times 2$ , and the kernel size of the remaining pooling layer is  $2 \times 2 \times 2$  with stride  $1 \times 2 \times 2$ . The convolution kernels of all convolutional layers are 64, 128, 256, and 256, respectively. The input image size is  $112 \times 112$ , and the output image size is  $28 \times 28$ . After the fully connected layer, the feature vector of  $4096 \times 1$  is obtained.



Figure 5. The 3DCNN component

## 4. Experiments

### 4.1. General Description of the Experiments

The experimental data used in this paper is from the Beijing Meteorological Bureau and is divided into three categories: 1) hourly concentration observations from ten stations, including PM2.5 concentration, PM10 concentration, O<sub>3</sub> concentration, SO<sub>2</sub> concentration, NO concentration, CO concentration, and meteorological observation data from the target station. All the data is updated hourly. To sum up, for the future 24 hour forecast in a single station, all the selected features of this station are injected; 2) meteorological gridded observation data with  $112 \times 112$  grid, updated hourly, including temperature, relative humidity, and wind speed; 3) station meter information, including site geography location, ID, type, etc.

Our implementations of the models were in Python with the help of the Scikit-learn machine learning library, Tensorflow, and Keras. The Scikit-learn machine learning library was applied to complete the data preprocessing and build a PM2.5 concentration prediction model based on SVM, which was compared with the deep neural network model. The LSTM and 3DCNN deep neural network structures were built and trained by Tensorflow and Keras.

### 4.2. Fine-Tuned Model

For any model based on neural networks, hyperparameters are required to be configured ahead, including those that specify the structure of the model itself and those that determine how the model is trained. We call them training parameters and model hyperparameters here. 3D-CNN and LSTM are the two parts of the proposed model, but the hyperparameters of LSTM should be determined. The input features have already been determined by what data is taken as input, so for the LSTM model, the number of hidden layers and number of hidden units in each hidden layer will be determined by hyperparameter exploration. The multi-layer stacked LSTM structure can increase the learning ability of the model, but the network parameters will also dramatically increase, which has a negative impact on the generalization ability of the model and the training time.

In order to find the optimal parameters of the model, a large number of tests were taken. Five different structural LSTM models were tested by using a validation dataset. The number of layers, hidden units in each hidden layer, and MAPE, RMSE, and MAE errors for each structure are shown in Table 1. It can be seen that the errors of the model will not decrease by simply increasing the number of layers or the number of nodes, the error increases significantly when the number of hidden units in each hidden layer is more than 400, and the RMSE error increases by 3.16 times when the stacked LSTM exceeds four layers. The LSTM model consisting of a two-layer LSTM structure with 128 hidden units of each hidden layer has the best performance; therefore, it is chosen as the fixed structure of LSTM model for subsequent experiments.

Table 1. Error comparison among five LSTM structure

LSTM structure	Number of layers	Hidden units	MAPE	RMSE	MAE
LSTM-1	1	128	27.14	17.85	9.91
<b>LSTM-2</b>	<b>2</b>	<b>256</b>	<b>23.8</b>	<b>16.32</b>	<b>8.62</b>
LSTM-3	3	384	25.85	15.57	9.72
LSTM-4	4	512	46.53	64.79	9.29
LSTM-5	5	640	48.62	64.06	27.45

The hyperparameters involved in the Adam algorithm include the learning rate, loss function, mini-batch size, and number of training iterations. In this paper, the mini-batch size, learning rate, and loss function are fixed according to the experience of previous study. Only the number of training iterations, also called epochs, needs to be determined. Excessive

training iterations can cause the model to overfit the training data and consume more time. Figure 6 shows the MAPE variation of the training dataset and validation dataset over the training epochs. The MAPE error is gradually reduced as the number of iterations increases. When the number of iterations exceeds 3000, the model seems to be over-adapted; not only does the generalization ability not improve, but it also has weak fluctuations. Therefore, the proper iteration step size should be 3000.

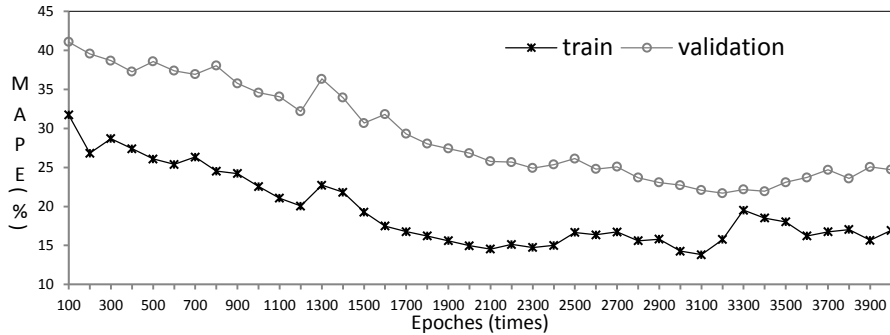


Figure 6. The trend of MAPE over epochs

4.3. Results and Discussion

According to the results of the above tests, the LSTM model with two layers and 128 hidden neurons per layer was selected and combined with 3D-CNN to predict PM2.5 con-centration. The total training step size was set to 3000, the learning rate was set to 10-3, the decay rate was set to 0.9995, the initialization range of parameter was set to [-0.08, 0.08], and the adaptive moment estimation (Adam) algorithm was chosen to optimize the model. The Adam method is efficient in terms of computation, requires little memory, and is suitable for problems with numerous data or parameters. The previous sequence observed data with a fixed duration of 72 hours to predict the PM2.5 concentration for each station over a 24-hour period, each sample contained 73 time series (input 72, output 1) data, and each time series data consisted of 19 feature vectors. Batch training input 64 samples at a time. The training dataset spanned from 2015 to 2017, with 80% of the data used to train the model and 20% of the data used to validate the model. The time span of the test set was 10-11 in 2018.

4.3.1. Prediction Results

The trained model used the test dataset for prediction. Considering the large difference in the spatial distribution of PM2.5 concentration, the performance of the RMSE, MAE, and MAPE of ten stations was calculated respectively. The specific results are shown in Table 2. The RMSE varied from 12.6 to 15.32 $\mu\text{g}/\text{m}^3$ , the MAE varied from 3.93 to 6.67 $\mu\text{g}/\text{m}^3$ , and the MAPE varied from 14.23 to 20.52%. The ID3 site had a minimum MAE value of 3.24 $\mu\text{g}/\text{m}^3$ , a MAEP value of 19.21%, and an RMSE value of 13.56 $\mu\text{g}/\text{m}^3$ , which was the best predictive effect at all stations. Overall, the prediction results were consistent with the observed data trend. The RMSE and MAE of the ten sites were not too different, and the absolute errors were within the acceptable range. Four of the ten sites were located in the suburbs, one of which was located in a mountainous area, and the other six were in an urban area. The geographical locations of the sites were quite different, but the prediction results of the model do not show a particularly large deviation from the observation, which indicates the learning effect on spatial features of the proposed model is better.

Table 2. RMSE, MAE, and MAPE among different stations

Station-ID	MAPE	RMSE	MAE
1	14.91	5.41	19.91
2	15.32	3.93	19.21
3	<b>13.56</b>	<b>3.24</b>	<b>14.23</b>
4	17.91	3.41	20.52
5	14.10	4.18	19.60
6	13.23	4.73	19.50
7	12.6	4.30	15.87
8	13.6	6.67	17.2
9	14.4	4.11	18.58
10	13.2	4.52	19.91

The input sequence and target values of the samples were modified so that the models predicted PM2.5 concentrations for the next hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 12 hours, and 24 hours. MAE was selected to evaluate the



prediction accuracy of the model on different time scales over the 24-hour period. Comparative experiments were performed among the proposed model, RMAPS-IN system, and SVM model, and the results are shown in Table 3.

Table 3. Comparison of the MAE of different models for different time scales over the next 24 hours

Model	1 <sup>st</sup> hour	2 <sup>nd</sup> hour	3 <sup>rd</sup> hour	4 <sup>th</sup> hour	5 <sup>th</sup> hour	6 <sup>th</sup> hour	12 <sup>th</sup> hour	24 <sup>th</sup> hour
3DCNN-LSTM	2.31	3.40	5.06	7.07	7.96	8.66	10.02	12.28
RMAPS-IN	7.34	9.68	11.08	15.51	16.33	19.04	23.85	28.46
SVM	16.59	21.18	25.37	27.35	30.71	33.80	35.22	37.67

It can be seen that the MAE of several models has an increasing trend over time. The MAE of the 3DCNN-LSTT model changed from 2.31 to 12.28, and the increasing rate with time was the smallest, which is a relatively steady trend. The MAE of the RMAPS-IN system varied smoothly over the next 1-3 hours, while the MAE of the SVM model had a sharp increase over time. The MAE of the 3DCNN-LSTM model was much lower than that of RMAPS-IN. The MAE for the next hour was 68.5% lower than that of RMAP-IN. The MAE for the next 24 hours was 56.8% lower than that of MAPS-IN. The proposed model's accuracy was higher than that of the RAMPS-IN system, which is the current operational forecast system in Beijing Meteorological Bureau.

#### 4.3.2. The Comparison of Evaluation Indicators

The RMSE, MAE, and MAPE were selected as the accuracy evaluation indicators for the comparison among the 3DCNN-LSTM deep learning network model, SVR model, and RMAPS-IN system, and the error distribution at ten stations is displayed in Figures 7(a)-(c). 3DCNN-LSTM had the lowest RMSE over ten stations with a range of [10,15]; its MAE was also significantly lower than the other two models, with a peak of 6.67 $\mu\text{g}/\text{m}^3$ , and its MAPE had the same situation. In the RMAPS-IN model, the RMSE and MAPE of the four stations were significantly higher than those of the other models. The SVR model had much worse performance compared with the other two models.

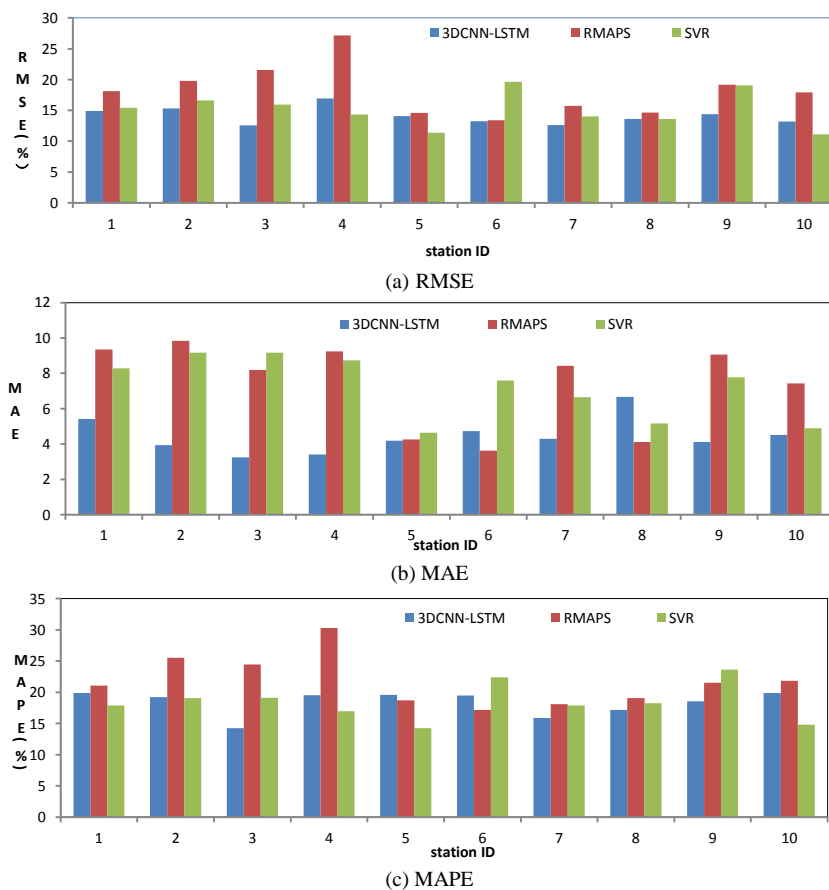


Figure 7. Comparison among different models

To sum up, the 3DCNN-LSTM model presents a more accurate PM2.5 concentration prediction than RMAPS-IN and



SVR. All the results indicate that the accuracy of the space-time model is higher than only the time series model, especially for PM<sub>2.5</sub> concentration prediction. The deep neural network method can autonomously learn more information than unsupervised machine learning, thereby improving prediction performance. In addition, 3DCNN-LSTM is superior to the RMAPS-IN system. There are two main reasons for this. First, although the RMAPS-IN system uses fusion technology, complex terrain mode correction techniques, and dynamic downscaling techniques, it is still impossible to remove the limitation of the uncertainties parameters used in the model. The 3DCNN-LSTM deep neural network can learn the complex spatiotemporal features of the dataset through nonlinear and convolution, which can reduce the error caused by uncertainty. The second reason is that in real life, PM<sub>2.5</sub> is characterized by many sudden occurrences. If the 3DCNN-LSTM network ever sees a similar pattern during training, it could find this type of sudden change and give a reasonable prediction in the prediction.

## 5. Conclusions

Meteorological factors have always been the main factors affecting the prediction of PM<sub>2.5</sub>, and the coupling effect of multiple meteorological factors has a very complicated impact on it. Based on the analysis of the correlation between meteorological elements and PM<sub>2.5</sub>, this paper proposes a 3DCNN-LSTM model to predict PM<sub>2.5</sub> concentration that uses meteorological grid data and station data. The comparison between the prediction results and the errors of several models shows that the 3DCNN-LSTM prediction model can effectively acquire the spatiotemporal features, which is better than the current operational RMAPS-IN system, and is suitable for solving the prediction problem of spatiotemporal sequence data. However, the comparative experiments were not conducted with this model and other popular deep learning models. In the subsequent research, the efficiency and accuracy of the model will be further improved through comparative analysis of various models.

## Acknowledgments

This work was partly financially supported through a grant from the Special Science Research Project of China Meteorological Administration (No. 2019ZDIANXM19). The authors thank the three anonymous reviewers for their helpful suggestions.

## References

1. Y. J. Kaufman, T. Didier, and B. Olivier, "A Satellite View of Aerosols in the Climate System," *Nature*, Vol. 419, pp. 215-233, 2002
2. C. K. Chan and X. H. Yao, "Air Pollution in Mega Cities in China," *Atmospheric Environment*, Vol. 42, No. 11, pp. 1-42, 2008
3. X. Y. Zhang and H. B. Hu, "Risk Assessment of Exposure to PM<sub>2.5</sub> in Beijing using Multi-Source Data," *Acta Scientiarum Naturalium Universitatis Pekinensis*, Vol. 54, No. 1, pp. 1103-1113, 2018
4. Q. Jin, X. Fang, B. Wen, and A. Shan, "Spatiotemporal Variations of PM<sub>2.5</sub> Emission in China from 2005 to 2014," *Chemosphere*, Vol. 183, pp. 429-436, 2014
5. A. V. Donkelaar, R. V. Martin, M. Brauer, and B. L. Boys, "Use of Satellite Observations for Long-Term Exposure Assessment of Local Concentrations of Fine Particulate," *Environmental Health Perspectives*, Vol. 123, No. 2, pp. 135-143, 2015
6. C. L. Fang, Z. B. Wang, and G. Xu, "Spatial-Temporal Characteristics of PM<sub>2.5</sub> in China: A City Level Perspective Analysis," *Journal of Geographical Sciences*, Vol. 26, pp. 1519-1532, 2016
7. X. Su, W. Gough, and Q. Shen, "Correlation of PM<sub>2.5</sub> and Meteorological Variables in Ontario Cities: Statistical Downscaling Method Coupled with Artificial Neural Network," in *Proceedings of the 24<sup>th</sup> International Conference on Modeling, Monitoring and Management of Air Pollution*, pp. 215-226, 2016
8. R. Chen, X. Wang, X. Meng, J. Hua, Z. J. Zhou, B. H. Chen, et al., "Communicating Air Pollution-Related Health Risks to the Public: An Application of the Air Quality Health Index in Shanghai, China," *Environment International*, Vol. 1, No. 5, pp. 168-173, 2013
9. J. Chen, J. Lu, J. C. Avise, J. A. DaMass, M. J. Kleeman, and A. P. Kaduwela, "Seasonal Modeling of PM<sub>2.5</sub> in California's San Joaquin Valley," *Atmospheric Environment*, Vol. 92, pp. 182-190, 2014
10. Q. Z. Wu, W. S. Xu, A. Shi, Y. Li, X. J. Zhao, Z. F. Wang, et al., "Air Quality Forecast of PM<sub>10</sub> in Beijing with Community Multi-Scale Air Quality Modeling (CMAQ) System: Emission and Improvement," *Geoscience Model*, Vol. 12, No. 7, pp. 2243-2259, 2014
11. L. Chen, D. M. Wu, and Q. Chen, "Prediction of Air Pollution based on Wavelet Analysis and Support Vector Machine," *Journal of XTAN University of Science and Technology*, Vol. 30, No. 6, pp. 726-730, 2010
12. G. -Q. Zhou, Y. Xie, J. -B. Wu, Z. -Q. Yu, L. -Y. Chang, and W. Gao, "WRF-Chem based PM<sub>2.5</sub> Forecast and Bias Analysis over the East China Region," *China Environment Science*, Vol. 36, No. 8, pp. 2251-2259, 2016
13. G. Yi and M. G. Zhang, "Numerical Simulation of a Heavy Fog-Haze Episode over the North China Plain in January 2013," *Climatic and Environmental Research*, Vol. 19, No. 2, pp. 40-152, 2014
14. H. Z. De, H. X. Yun, and H. X. Yong, "Haze Forecast based on Time Series Analysis and Kalman Filtering," *Journal of Computer Applications*, Vol. 37, No. 11, pp. 3311-3316, 2017

15. J. Z. Xiu, X. Jing, and Z. Z. Yin, "Beijing Regional Environmental Meteorology Prediction System and its Performance Test of PM2.5 Concentration," *Journal of Application Meteorological Science*, Vol. 27, No. 2, pp. 160-172, 2016
16. T. J. Wang, F. Jiang, J. J. Deng, Y. Shen, Q. Y. Fu, Q. Wang, et al., "Urban Air Quality and Regional Haze Weather Forecast for Yangtze River Delta," *Atmospheric Environment*, Vol. 58, No. 15, pp. 70-83, 2012
17. J. Liu, P. Yang, W. S. Lv, A. Liu, and J. X. Liu, "Prediction Model of PM2.5 Mass Concentrations based on Fuzzy Time Series and Support Vector Machine," *Journal of University of Science and Technology Beijing*, Vol. 36, No. 12, pp. 1694-1703, 2014
18. L. Li, L. Ma, J. F. He, D. G. Shao, S. L. Yi, Y. Xiang, et al., "PM2.5 Concentration Prediction Model of Least Squares Support Vector Machine based on Feature Vector," *Journal of Computer Applications*, Vol. 34, No. 8, pp. 2212-2216, 2014
19. J. Pan, H. Z. Wang, H. Gao, W. X. Zhao, H. X. Huo, and H. R. Dong, "Paradise Pointer: A Sightseeing Scenes Images Search Engine based on Big Data Processing," in *Proceedings of the International Conference of Young Computer Scientists*, pp. 448-452, Harbin, China, January 2015
20. D. Mishra, P. Goyal, and A. Upsfhyssy, "Artificial Intelligence based Approach to Forecast PM2.5 During Haze Episodes: A Case Study of Delhi, India," *Atmospheric Environment*, Vol. 120, pp. 239-248, 2015
21. G. O. Philip, L. Gunnar, and D. Ottfried, "Relationship Between Rice Yield and Climate Variables in Southwest Nigeria using Multiple Linear Regression and Support Vector Machine Analysis," *International Journal of Biometeorology*, Vol. 62, No. 3, pp. 459-469, 2018
22. X. J. Shi, Z. R. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," in *Proceedings of the International Conference of Language Process*, pp. 1-12, Los Angeles, USA, August 2015
23. J. C. Zhao, F. Deng, Y. Y. Cai, and J. Chen, "Long Short-Term Memory-Fully Connected (LSTM-FC) Neural Network for PM2.5 Concentration Prediction," *Chemosphere Volume*, Vol. 220, pp. 486-492, 2019
24. C. Vidushi, D. Anand, K. Vijayan, et al., "Time Series based LSTM Model to Predict Air Pollutant's Concentration for Prominent Cities in India," in *Proceedings of 1st International Workshop on Utility-Driven Mining*, pp. 1-9, London, United Kingdom, August 2018
25. S. A. Weber, T. Z. Insaf, E. S. Hall, T. O. Talbot, and A. K. Huff, "Assessing the Impact of Fine Particulate Matter (PM2.5) on Respiratory-Ardiovascular Chronic Diseases in the New York City Metropolitan Area using Hierarchical Bayesian Model Estimates," *Environment Research*, Vol. 151, pp. 399-409, 2016
26. J. H. Chou and H. K. Ping, "A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities," *Sensors*, Vol. 18, No. 2, pp. 2220-2241, 2018
27. C. C. Wen, S. F. Liu, X. J. Yao, L. Peng, X. Li, Y. Hu, et al., "A Novel Spatiotemporal Convolutional Long Short-Term Neural Network for Air Pollution Prediction," *Science of the Total Environment*, Vol. 654, pp. 1091-1099, 2019
28. W. S. Ping, W. C. Jia, and W. H. Jin, "Adaptive Deep Learning-based Air Quality Prediction Model using the Most Relevant Spatial-Temporal Relations," *IEEE Access*, Vol. 6, No. 38, pp. 186-382, 2018
29. D. Chu, Y. J. Kaufman, G. Zibordi, J. Chern, J. T. Mao, C. C. Li, et al., "Global Monitoring of Air Pollution Over Land from the Earth Observing System-Terra Moderate Resolution Imaging Spectro-Radiometer (MODIS)," *Journal of Geophysical Research*, Vol. 21, No. 108, pp. 4661-4667, 2003
30. P. E. Saide, G. R. Carmichael, S. N. Spak, L. Gallardo, A. E. Osses, M. A. Mena-Carrasco, et al., "Forecasting Urban PM10 and PM2.5 Pollution Episodes in Very Stable Nocturnal Conditions and Complex Terrain using WRF-Chem CO Tracer Model," *Atmospheric Environment*, Vol. 45, No. 16, pp. 2769-2780, 2011
31. T. Du, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning Spatiotemporal Features with 3D Convolutional Networks," in *Proceedings of the International Conference of Computer Vision ICCV*, pp. 4489-4497, Santiago, Chile, Dec. 2015
32. Y. Chen, R. Shi, S. Shu, and W. Gao, "Ensemble and Enhanced PM10 Concentration Forecast Model based on Stepwise Regression and Wavelet Analysis," *Atmospheric Environment*, Vol. 74, pp. 346-359, 2013
33. Z. Liang, G. Z. Guan, and Y. S. Pei, "Learning Spatiotemporal Features using 3DCNN and Convolutional LSTM for Gesture Recognition," in *Proceedings of the International Conference of Computer Vision*, pp. 1-9, Venice, Italy, October 2017

**Wenfang Zhao** is a senior engineer at the Beijing Meteorological Information Center, Beijing Meteorological Bureau, Beijing, China. Her research interests include big data, deep learning, machine learning, and big data analysis.

**Yong Zhou** is a senior engineer at the Development and Research Center, China Meteorological Administration, Beijing, China. His main research interests include atmospheric sounding, instrument development, and atmospheric environment.

**Wei Tang** is a senior engineer at the Development and Research Center, China Meteorological Administration, Beijing, China. Her current research interests include data mining and machine learning.