

# PM<sub>2.5</sub> Concentration Prediction Method Based on Temporal Attention Mechanism and CNN-LSTM

Zhuohui Zhou<sup>1</sup>, Xiaofang Liu<sup>1,\*</sup>, Huan Yang<sup>1</sup>

<sup>1</sup>School of Computer Science and Engineering, Sichuan University of Science & Engineering, Sichuan 643000, China

\* Corresponding author: Liu Xiaofang (Email: lxf1969@163.com)

**Abstract:** Accurately predicting PM<sub>2.5</sub> concentration can effectively avoid the harm caused by heavy pollution weather to human health. In view of the non-linearity, time series characteristics, and the problem of large multi-step prediction errors in PM<sub>2.5</sub> concentration data, a method combining Long Short-term Memory Network and Convolutional Neural Network with Time Pattern Attention mechanism (TPA-CNN-LSTM) is proposed. The method uses historical PM<sub>2.5</sub> concentration data, historical meteorological data, and surrounding station data to predict the future 6-hour PM<sub>2.5</sub> concentration of air quality monitoring stations. Firstly, CNN is used to obtain the spatial characteristics between multiple stations, secondly, LSTM is added after CNN to extract the temporal changes of non-linear data, and finally, to capture the key features of temporal information, Temporal Pattern Attention mechanism (TPA) is added. TPA can automatically adjust weights based on the input of each time step, and select the most relevant time step for prediction, thereby improving the accuracy of the model. An example analysis is conducted on the measured data of Beijing's air quality stations in 2018, and compared with other mainstream algorithms. The results show that the proposed model has higher prediction accuracy and performance.

**Keywords:** PM<sub>2.5</sub> concentration prediction, Separated by commas, Temporal pattern attention mechanism, CNN-LSTM.

## 1. Introduction

The introduction should provide background information (including relevant references) and should indicate the purpose of the manuscript. Cite relevant work by others, including research outside your company. Place your work in perspective by referring to other research papers. Inclusion of statements at the end of the introduction regarding the organization of the manuscript can be helpful to the reader.

PM<sub>2.5</sub> refers to particulate matter with a diameter of equal to or less than 2.5 micrometers in ambient air. Due to their small size, these particles can penetrate deep into the respiratory system and pose a significant threat to human health. PM<sub>2.5</sub> particles often adsorb carcinogenic substances, toxic metals, persistent organic pollutants, and other harmful materials, which can directly affect the lungs when they enter the human body, leading to heavy metal poisoning, increased cancer risk, reproductive harm, and other problems [1].

The fluctuations in PM<sub>2.5</sub> concentration have varying degrees of impact on the respiratory system, crop growth, and tourism. Predicting future PM<sub>2.5</sub> concentrations in advance can provide valuable health information for travelers and serve as a warning for relevant departments to take measures to improve air quality [2]. This research topic has positive significance for the long-term development of public health, the economy, and the ecological environment. At the same time, it presents significant challenges to the accuracy and stability of the models, making it a profound and meaningful research topic.

Against the background of rapid development in machine learning, researchers have conducted many studies on air pollutant concentration prediction based on different algorithms and models. For example, Zhang et al. [3] used the PCA method for data feature extraction, which improved the prediction accuracy and reduced the model's time complexity. They input the extracted data into a BP neural network for

prediction. Samal et al. [4] proposed the Multi-Time Convolutional Artificial Neural Network (MTCAN) model, which can perform feature learning and sequence modeling simultaneously, and use a large amount of past observation data for long-term prediction to minimize memory requirements and operating time. Liu et al. [5] optimized the BP neural network using a genetic algorithm and constructed a feature-based PM<sub>2.5</sub> concentration prediction model. Li [6] proposed an AC-LSTM model composed of a one-dimensional convolutional neural network, long short-term memory network, and attention mechanism. This model not only uses air pollutant concentration but also adds PM<sub>2.5</sub> concentration from neighboring air quality monitoring stations as prediction data. They used CNN and LSTM to extract the spatio-temporal correlation and interdependence of multi-variable time series data, and used attention mechanism to capture the importance of different feature states at different time steps in affecting future PM<sub>2.5</sub> concentrations. Liu et al. [7] used historical air pollutant and meteorological data for a region to construct an LSTM prediction model and accurately predict the PM<sub>2.5</sub> concentration for 1, 4, 8, and 12 hours in the future.

The above studies have made certain improvements in predicting PM<sub>2.5</sub> concentration, but these studies often only consider the relevant features of a single station itself or ignore the correlation between different time steps and the influence of meteorological factors on PM<sub>2.5</sub> concentration. In fact, PM<sub>2.5</sub> concentration is not only affected by local historical conditions but also by the transport of pollutants from surrounding areas and meteorological factors. Therefore, this paper proposes a new type of TPA-CNN-LSTM network, which uses historical pollutant concentrations, surrounding PM<sub>2.5</sub> concentrations, and meteorological factors as prediction features. LSTM overcomes the problems of gradient explosion and vanishing in RNN and can extract the time features of long-term prediction. CNN can extract the spatiotemporal features in the input data and the relevant

features of  $PM_{2.5}$  concentration between different monitoring stations. The time pattern attention mechanism can capture periodic features in time series data and handle multiple time series inputs, improving the accuracy, adaptability, and practicality of the model.

## 2. Data and Method

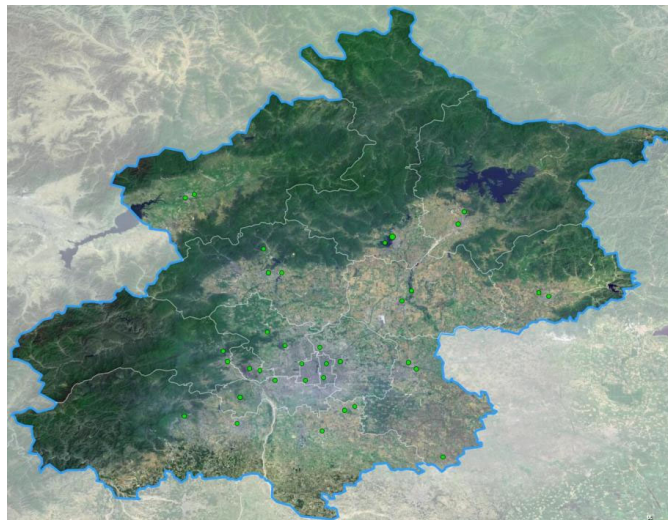
### 2.1. Data Description

The temporal and spatial variations of atmospheric particulate matter (PM) are influenced by multiple factors, including pollution sources, meteorological conditions, and others [14,15]. The variation of  $PM_{2.5}$  concentration is not only affected by the previous atmospheric state and  $PM_{2.5}$  concentration but also related to the  $PM_{2.5}$  concentration in adjacent areas [16,17]. Therefore, historical air pollution data and meteorological data need to be considered in the

prediction model. In this study, we selected 35 environmental monitoring stations in Beijing and collected hourly historical air quality concentration and meteorological data from January 1, 2018, to December 31, 2018, with a total of 8674 sets of data. The statistical information of the specific data is shown in Table 1. The locations of the air quality monitoring stations are shown in Figure 1. The experimental data includes concentrations of  $PM_{10}$ ,  $PM_{2.5}$ ,  $SO_2$ ,  $NO_2$ ,  $O_3$ , and CO. The detailed characteristics of the air quality monitoring stations are also listed in Table 1. The meteorological data includes indicators such as atmospheric pressure, temperature, wind speed, humidity, and weather. Data sources include the Beijing Environmental Protection Monitoring Center and the European Center for Medium-Range Weather Forecasting reanalysis data. For individual missing data in the used data, linear interpolation, pre-filling, or post-filling methods were used for filling, and then normalization was carried out.

**Table 1.** Description of dataset

data type	data name	Variable name	unit	type of variable
Air quality data	The concentration of $PM_{2.5}$	$PM_{2.5}$	$\mu g/m^3$	numerical
	The concentration of $PM_{10}$	$PM_{10}$	$\mu g/m^3$	numerical
	The concentration of $NO_2$	$NO_2$	$\mu g/m^3$	numerical
	The concentration of CO	CO	$\mu g/m^3$	numerical
	The concentration of $O_3$	$O_3$	$\mu g/m^3$	numerical
	The concentration of $SO_2$	$SO_2$	$\mu g/m^3$	numerical
	Climatological data	weather	Weather	-
temperature		Temperature	$^{\circ}C$	numerical
pressure		Pressure	hPa	numerical
relative humidity		Relative humidity	%	numerical
wind speed		wind_speed	m/s	numerical
wind direction		wind_direction	-	Category (None, Southwest, Unstable, Southeast, Southwest, Northwest)



**Figure 1.** distribution of air quality monitoring stations in Beijing

### 2.2. TPA mechanism

The proposed model utilizes one-dimensional CNN to

learn the temporal pattern information of time series data, referred to as TPA (temporal pattern attention), as a local feature learning method in the network. The TPA mechanism

structure is shown in Figure 2. The leftmost arrow in Figure 2 represents the processing of variables, and each row represents a variable's time series data. The time pattern matrix  $H_{i,j}^C$  of the variable within the convolutional kernel range is obtained by convolution calculation. The scoring function calculates the score of the time pattern matrix and normalizes the score using the sigmoid function to obtain the attention weight  $\alpha$ . The context vector  $v_t$  is obtained by combining the time pattern matrix and attention weight. The context vector  $v_t$  from the encoder and the hidden state  $h$  are concatenated and connected with the hidden state  $s$  from the decoder. The output prediction value is calculated using the output layer and softmax function. The following describes in

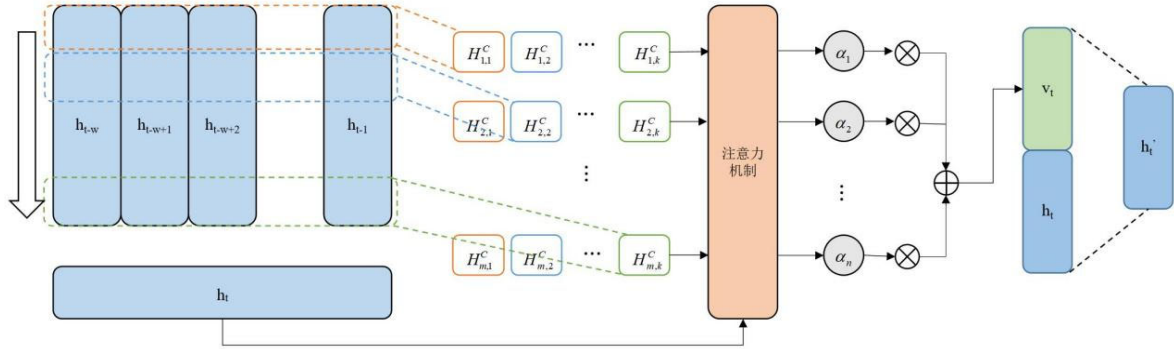


Figure.2 TPA Mechanism Structure

Among them,  $H_{i,j}^C$  represents the temporal pattern matrix of the  $j$ -th filter for the attention mechanism of the  $i$ -th row in the interval  $[t-\omega, t-1]$  through convolution calculation.  $H_{i,(t-\omega-1+l)}$  determines the scope of the convolution calculation. The convolution kernel  $C_{i,(T-\omega+l)}$  performs convolution calculation on the data within this range. The learned temporal patterns are scored using the following formula:

$$f(H_i^C, h_t) = (H_i^C)^T W_\alpha h_t \quad (2)$$

Where  $H_i^C$  represents the  $i$ -th row of the time mode matrix  $H^C$ ,  $h_t$  represents the hidden state of the encoder output, and  $W_\alpha$  represents the weight parameter. The attention weights  $\alpha_i$  are:

$$\alpha_i = \text{sigmoid}\left(f(H_i^C, h_t)\right) \quad (3)$$

Weighted sum each row of HCC to obtain the context vector  $v_t$ :

$$v_t = \sum_{i=1}^m \alpha_i H_i^C \quad (4)$$

Among them,  $n$  is the number of features of the input

detail the role of the TPA mechanism in the proposed model.

Firstly, one-dimensional CNN is used for convolution calculation.  $k$  filters are set, and the kernel size is  $1k$ .  $T$  represents the range covered by attention, usually set to  $T=w$ . The convolution is calculated along the row vector of the hidden state matrix  $H$  using the above kernel to extract the time pattern matrix  $H_{i,j}^C$  of the variable within the convolutional kernel range, as shown in Equation (1).

$$H_{i,j}^C = \sum_{l=1}^{\omega} C_{j,T-\omega+l} * H_{i,(t-\omega-1+l)} \quad (1)$$

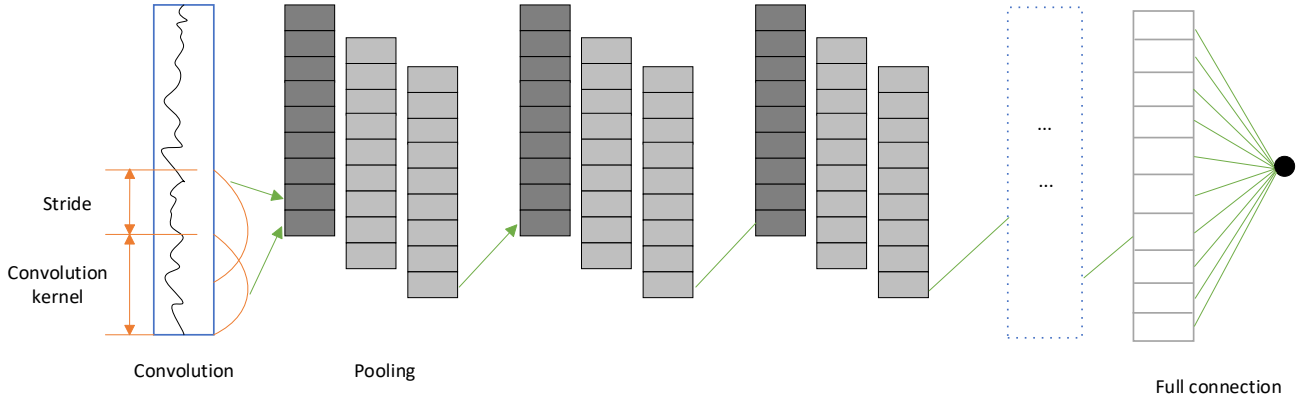
variable  $x$ . The output value of the TPA-CNN-LSTM model is obtained by adding the attention vector  $v_t$  linearly mapped with the hidden state  $h_t$  of the time step  $t$ , as shown in equation (5).

$$y_{t-1+\Delta} = W_h (W_h h_t + W_v v_t) \quad (5)$$

In the formula,  $y_{t-1+\Delta}$  represents the output value,  $\Delta$  represents the predicted time length, and  $W_h$ ,  $W_h$ , and  $W_v$  represent the weight parameter matrices of the corresponding variables.

### 2.3. Convolutional neural network

Convolutional Neural Network (CNN) have achieved excellent performance in the fields of computer vision and natural language processing, thanks to their powerful feature extraction and recognition capabilities. Typically, one-dimensional convolution is used for processing time data, two-dimensional convolution is used for spatial convolution in images, and three-dimensional convolution is used for spatial convolution in three-dimensional space. In this paper, we believe that the convolution kernel in CNN is a one-dimensional structure. By using CNN, multi-variable time series data such as meteorological data, air pollution data, and data from adjacent stations needed for  $PM_{2.5}$  prediction can be input through different channels to maximize information retention. CNN mainly consists of three modules: it is a feedforward neural network. Its basic structure consists of an input layer, a convolutional layer, a pooling layer, a fully connected layer (FC), and an output layer, as shown in Figure 3.



**Figure 3.** Structure of one-dimensional convolutional neural network

CNN has the ability to automatically learn data features and has characteristics such as local connectivity, weight sharing, pooling operations, and multi-layer structure. These features significantly reduce the complexity of the PM<sub>2.5</sub> concentration prediction model and reduce overfitting through gradient descent optimization, thus improving generalization ability. The calculation formulas for the convolution layer, pooling layer, and fully connected layer are:

$$H_i = f(H_{i-1} * W_i + b_i) \quad (6)$$

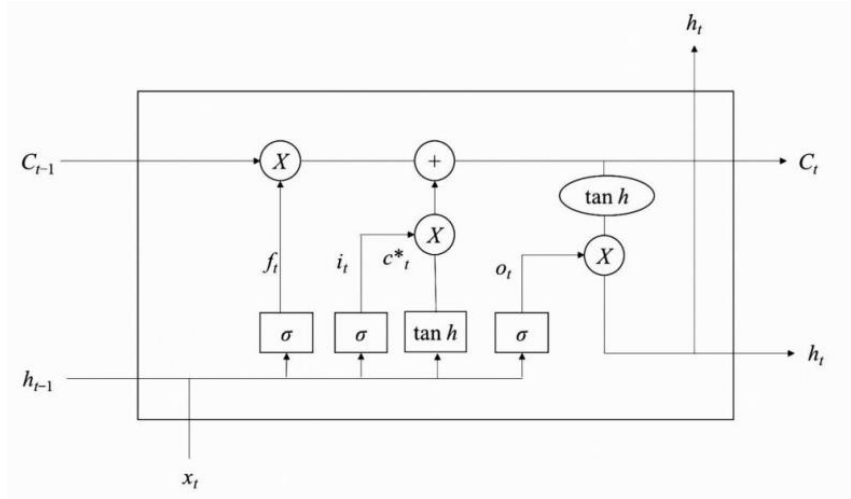
$$Z_i = \text{subsampling}(Z_{i-1}) \quad (7)$$

$$Y(i) = (L = l_i | H_0; (W, b_a)) \quad (8)$$

In the formula,  $H_i$  represents a convolutional layer,  $W_i$  represents a weight vector,  $*$  represents convolution operation,  $b_i$  represents bias parameters,  $b_a$  represents offset vector,  $f(\cdot)$  represents activation function,  $Z_i$  represents upsampling layer, *subsampling* represents sampling process,  $Y(\cdot)$  represents output,  $L$  represents loss function, and  $l_i$  represents class labels.

## 2.4. LSTM model

The LSTM controls the transmission state through a gate structure, divided into forgetting gates, selective memory gates, and output gates, as shown in Figure 4.



**Figure 4.** LSTM network structure diagram

The forgetting gate determines the extent to which the previous unit state is forgotten and can be expressed as:

$$G_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (9)$$

In the formula:  $G_t$  is the state matrix of the forget gate at time  $t$ ;  $h_{t-1}$  is the state matrix of the previous time step's state unit;  $x_t$  is the input value at time  $t$ ;  $\sigma(\cdot)$  is the

activation function;  $W_f$  is the weight matrix of the forget gate;  $b_f$  is the constant parameter matrix of the forget gate.

We choose a memory gate and an activation function to control the range of new information being added. By using the joint effect of the forget gate and the memory gate output, we update the unit state of this unit, which can be expressed as:

$$i_t = \sigma(W_s \cdot [h_{t-1}, x_t] + b_s) \quad (10)$$

$$\hat{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (11)$$

$$C_t = G_t C_{t-1} + i_t \hat{C}_t \quad (12)$$

where:  $i_t$ ,  $\hat{C}_t$ ,  $C_t$  represent the state matrices of the input gate, the intermediate quantity input, and the tuple state at time t,  $\tanh$  respectively; are hyperbolic tangent activation functions;  $W_s$ ,  $W_c$  represent the weight matrices of the input gate and the intermediate quantity output, respectively;  $b_s$ ,  $b_c$  are the memory gate constant parameter matrices.

The output gate controls the degree to which the current unit is filtered and can be expressed as:

$$O_t = \sigma[W_o [h_{t-1}, x_t] + b_o] \quad (13)$$

$$h_t = O_t \tanh(C_t) \quad (14)$$

Where:  $O_t$  is the state matrix of the output gate at time t;  $W_o$  is the weight matrix of the output gate;  $b_o$  is the output gate constant parameter matrix.

### 3. TPA-CNN-LSTM model design

#### 3.1. Model Parameter Design

The parameters of the prediction model were set as follows: the number of convolutional layers was 1, the number of filters was 16, the convolutional kernel size was 5 5, the learning rate was 0.001, the epoch was 100, the batch size was 32, and the optimizer was Adam. the parameter selection method was the grid search method.

#### 3.2. Selection of Evaluation Indicators

The root mean square error (RMSE) (15) and mean

absolute error (MAE) (16) were chosen as metrics to evaluate the prediction performance of the TPA-CNN-LSTM model at each site and to compare it with other models.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (15)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (16)$$

where,  $y_i$ ,  $\hat{y}_i$  and N denote the observed, predicted and evaluated sample sizes for each station respectively. y is the average of the data from the N observation samples. RMSE and MAE are common evaluation metrics used in time series forecasting. The lower their values are, the better the performance of the model.

#### 3.3. TPA-CNN-LSTM prediction model construction

In order to overcome the limitation of CNN in capturing only local information, this paper proposes a model that combines TPA with CNN. The model first performs feature extraction on the input time series sample and then uses it as the input to the CNN convolutional layer. This combination can reduce the computational complexity of CNN and speed up the feature extraction process. After the CNN performs the second feature extraction, the output time series is processed for dimensionality reduction. Considering the complex model structure of LSTM, when inputting long time series, the model training time is long. However, using the time series extracted and dimensionality-reduced by TPA-CNN as the input to LSTM can improve the accuracy of LSTM prediction and reduce its operating time. The prediction process of the entire TPA-CNN-LSTM model is shown in Figure 5.

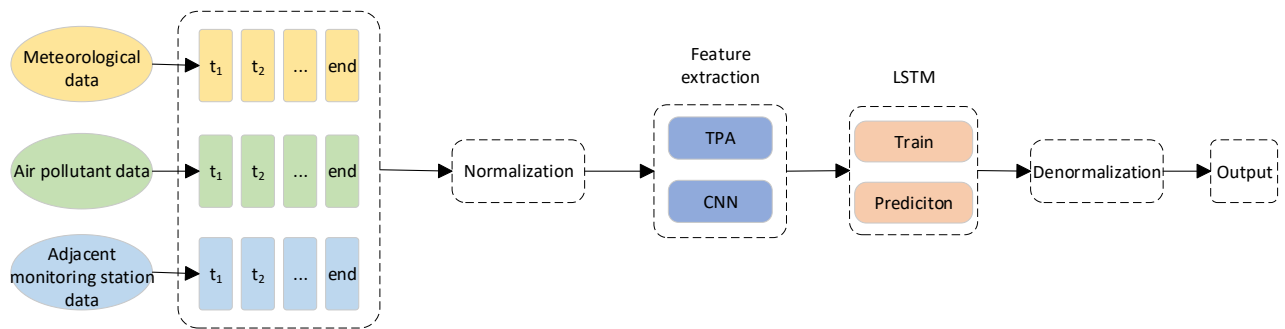


Figure 5. Flow chart of TPA-CNN-LSTM model

The main steps of the PM<sub>2.5</sub> concentration prediction model based on TPA-CNN-LSTM are as follows:

Step 1: Correlation analysis. Analyze the factors affecting the PM<sub>2.5</sub> concentration changes and determine the input sequence.

Step 2: Data preprocessing. Handle missing values and outliers, and normalize the input time series.

Step 3: Feature extraction. Use TPA-CNN to extract important features from the normalized time series.

Step 4: PM<sub>2.5</sub> concentration prediction. Input the TPA-CNN processed time series into the LSTM network for model training.

Step 5: Output. Reverse normalize the predicted data and output the results.

Through the above steps, the goal of improving prediction accuracy and computational efficiency is achieved.

## 4. Experimental Results and Analysis

### 4.1. Correlation Analysis

This paper uses the Pearson correlation coefficient analysis method to analyze the correlation between various variables and PM<sub>2.5</sub> concentration. The Pearson correlation coefficient is used to describe the correlation between two variables and is often expressed as an R<sup>2</sup> score.

$$R^2 Score = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{\delta_X} \right) \left( \frac{Y_i - \bar{Y}}{\delta_Y} \right) \quad (17)$$

Where  $\left( \frac{X_i - \bar{X}}{\delta_X} \right)$ ,  $\bar{X}$ , and  $\delta_X$  are respectively the

standardized scores, sample mean, and sample standard deviation of sample  $X_i$ . The value of R ranges from [-1,1], and the closer its value is to 1 or -1, the stronger the correlation. The closer its value is to 0, the weaker the correlation.

The PM<sub>2.5</sub> concentration value of a region can be analyzed from two aspects: historical concentration values and current meteorological conditions. Historical concentration values will have a certain impact on the concentration value of the next moment, while other pollutants and meteorological factors will have a greater impact on the concentration value of the next moment.

#### 4.1.1. Correlation Analysis of PM<sub>2.5</sub> Concentrations with Other Pollutants

Figure 6 analyzes the correlation between PM<sub>2.5</sub> and other pollutants. It can be seen that PM<sub>10</sub>, NO<sub>2</sub>, CO, and SO<sub>2</sub> are positively correlated with PM<sub>2.5</sub>, while O<sub>3</sub> is negatively correlated with PM<sub>2.5</sub>. This indicates that the concentration of PM<sub>2.5</sub> increases with the increase of PM<sub>10</sub>, NO<sub>2</sub>, CO, and SO<sub>2</sub> concentrations, and decreases with the increase of O<sub>3</sub> concentration.

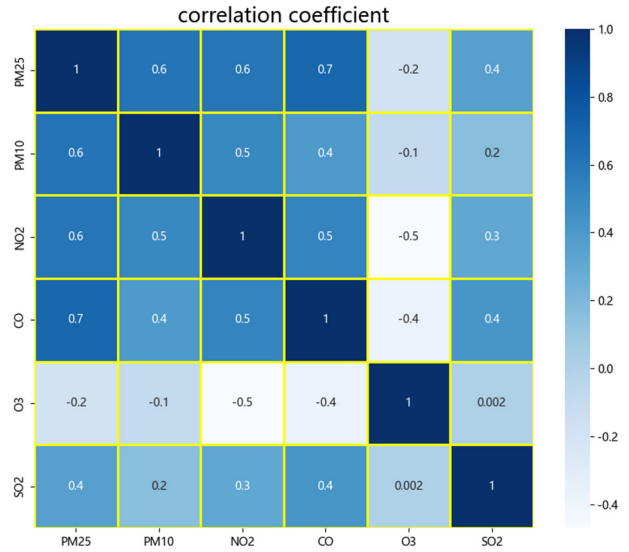


Figure 6. Correlation coefficient of PM<sub>2.5</sub> with other pollutants

#### 4.1.2. Correlation analysis between PM<sub>2.5</sub> concentration and meteorological factors

Figure 7 analyzes the correlation between PM<sub>2.5</sub> and meteorological factors. It can be seen from the graph that there is only a certain correlation between PM<sub>2.5</sub> and meteorological factors. However, the correlation between weather, humidity, wind speed, and PM<sub>2.5</sub> is relatively strong, while the correlation between other meteorological factors and PM<sub>2.5</sub> is weak. Among them, weather, humidity, and PM<sub>2.5</sub> are positively correlated, while wind speed and PM<sub>2.5</sub> are negatively correlated. Although the correlation between other factors and PM<sub>2.5</sub> is small, there is still a certain correlation. Therefore, this paper will select these data for training.

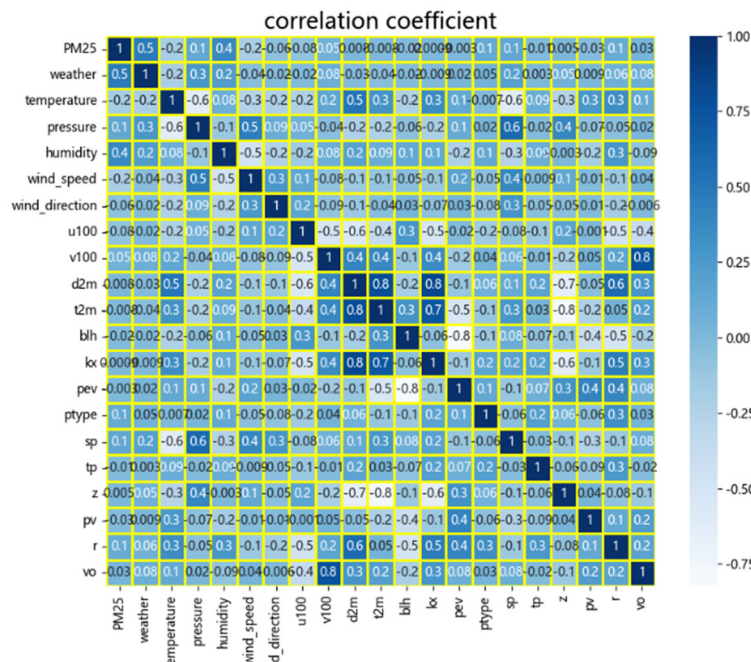
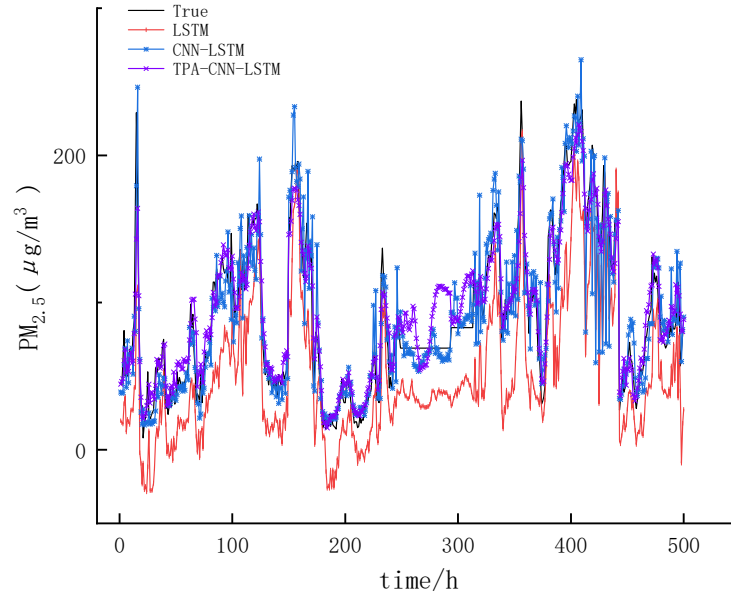


Figure 7. Correlation coefficient of PM<sub>2.5</sub> with other pollutants

## 4.2. Model Comparison Experiment Results

To validate the stronger learning capability and more accurate prediction results of the TPA-CNN-LSTM model

proposed in this paper, the LSTM model and CNN-LSTM model were used to predict PM<sub>2.5</sub> concentration on the same dataset, and the results of the three predictions were compared and analyzed, as shown in Figure 8.



**Figure 8.** Overall Comparison of Results of Different Prediction Models

According to Figure 8, the three prediction models have consistent trends with the actual values, with the TPA-CNN-LSTM model having predicted values that are closer to the real values, while the LSTM model has a larger deviation in its predicted values. CNN can perform network learning from the original input sequence, avoiding the accumulation of errors caused by manually extracting features. By adding the TPA network, important information can be prioritized and the input data of the LSTM network can be optimized. Therefore, the TPA-CNN-LSTM network can more effectively extract features of the time series of PM<sub>2.5</sub>

concentration and related influencing factors and improve the prediction accuracy.

Table 2 and Table 3 summarize the evaluation indicators for the three prediction models. Although the prediction accuracy of these three models decreases over time, it is noteworthy that the TPA-CNN-LSTM model performs better in the prediction results for each hour, and its prediction accuracy is higher than that of the LSTM and CNN-LSTM models. This result is also confirmed at the average prediction level of 6 hours.

**Table 2.** The performances of the different models in terms of root-mean-square error (RMSE)

time/h	RMSE/( $\mu\text{g}/\text{m}^3$ )		
	LSTM	CNN-LSTM	TPA-CNN-LSTM
t+1	23.83	21.77	20.65
t+2	28.34	27.40	23.51
t+3	36.51	32.26	27.34
t+4	38.82	36.18	30.97
t+5	41.62	38.53	33.62
t+6	44.18	40.72	37.72
average	35.55	32.81	28.97

**Table 6.** The performances of the different models in terms of mean absolute error(MAE)

time/h	MAE/( $\mu\text{g}/\text{m}^3$ )		
	LSTM	CNN-LSTM	TPA-CNN-LSTM
t+1	14.07	12.78	11.09
t+2	18.65	16.59	14.87
t+3	21.32	19.42	18.54
t+4	24.18	22.42	20.94
t+5	27.52	24.58	22.43
t+6	29.84	26.41	24.67
average	22.60	20.37	18.76

From the table, it can be seen that in the multi-hour prediction task, the TPA-CNN-LSTM model has the lowest MAE and RMSE. The predicted values of the TPA-CNN-LSTM model in the multi-scale prediction task are closer to the true values. In addition, in the one-hour PM<sub>2.5</sub> prediction task in Figure 8, the R<sup>2</sup> of the TPA-TCN-LSTM model is the highest. After adding the time pattern attention mechanism, the performance of TPA-CNN-LSTM in the multi-scale prediction task is better than that of LSTM and CNN-LSTM. The results indicate that the proposed TPA-CNN-LSTM model can effectively learn the spatiotemporal correlations of air pollutants and is suitable for related tasks of predicting PM<sub>2.5</sub> concentration.

## 5. Conclusions and Future Work

This paper proposes a CNN-LSTM model based on temporal pattern attention for predicting the concentration of PM<sub>2.5</sub> over multiple hours. The model uses air quality data, meteorological data, and PM<sub>2.5</sub> concentrations from neighboring monitoring stations as inputs to capture the spatiotemporal correlation and long-term dependency of PM<sub>2.5</sub>. The temporal pattern attention mechanism helps to capture the importance of different feature states and improve the prediction accuracy of the model. Experimental results show that the TPA-CNN-LSTM model performs well in multi-scale prediction tasks. The main conclusions of this paper are as follows:

Analysis of air pollution data shows that PM<sub>2.5</sub> concentrations have a strong spatiotemporal correlation. Due to the propagation of air flow, PM<sub>2.5</sub> concentrations in the prediction range are easily affected by PM<sub>2.5</sub> concentrations from neighboring monitoring stations. In addition, because PM<sub>2.5</sub> stays in the air for a long time, past feature states also affect future PM<sub>2.5</sub> concentrations. Furthermore, compared with a single model, an ensemble model combines the feature extraction capabilities of multiple models, resulting in more stable training results and stronger generalization ability. The TPA-CNN-LSTM prediction model can better capture the nonlinear relationship between each input variable and PM<sub>2.5</sub> concentration than other models, and has higher prediction accuracy.

Although the model achieved good prediction performance, there are still some limitations. The data collected in this study did not consider emissions from factories and vehicle exhaust, which are also important factors in the formation of air pollutants. Additionally, in the future, we will explore the use of this model for large-scale prediction of other air pollutants

and incorporate satellite meteorological data into the input of the prediction model.

## 6. Funding

Supported by The Innovation Fund of Postgraduate, Sichuan University of Science & Engineering under Grant Y2022184.

## References

- [1] Liang B, Li X, Ma K, et al. Pollution characteristics of metal pollutants in PM<sub>2.5</sub> and comparison of risk on human health in heating and non-heating seasons in Baoding, China[J]. *Ecotoxicology and Environmental Safety*, 2019, 170: 166-171.
- [2] KIOUMOURTZOGLOU M A, SCHWARTZ J, JAMES P, et al. PM<sub>2.5</sub> and mortality in 207 US cities modification by temperature and city characteristics[J]. *Epidemiology*, 2016, 27(2): 221-227.
- [3] ZHANG Y, GUO A, WU H, et al. Seasonal prediction of PM<sub>2.5</sub> based on the PCA-BP neural network[J]. *JOURNAL OF NANJING FORESTRY UNIVERSITY*, 2020, 44(5): 231-241.
- [4] Samal K, Babu K, Das S. Multi-directional temporal convolutional artificial neural network for PM<sub>2.5</sub> forecasting with missing values: A deep learning approach[J]. *Urban Climate*, 2021, 36, 100800:1-10.
- [5] LIU Linbo, LIU Lilong, Li Junyu, et al. Prediction of PM<sub>2.5</sub> mass concentration based on GA-BP neural network with water vapor[J]. *Journal of Guilin University of Technology*, 2019, 39(02): 420-426.
- [6] Li S, Xie G, Ren J, et al. Urban PM<sub>2.5</sub> Concentration Prediction via Attention-Based CNN-LSTM[J]. *Applied Sciences*, 2020, 10(6),1953: 1-17.
- [7] Liu X, Liu Q, Zou Y, et al. A self-organizing LSTM-based approach to PM<sub>2.5</sub> forecast[C]//Cloud Computing and Security: 4th International Conference, ICCCS 2018, Haikou, China, June 8–10, 2018, Revised Selected Papers, Part IV 4. Springer International Publishing, 2018: 683-693.
- [8] Chaloulakou, A. Kassomenos, P. Spyrellis, N. Demokritou, P. Koutrakis, P. Measurements of PM<sub>10</sub> and PM<sub>2.5</sub> particle concentrations in Athens, Greece. *Atmos. Environ.* 2003, 37, 649–660.
- [9] Hussein, T. Karppinen, A. Kukkonen, J. Härkönen, J. Aalto, P.P. Hämeri, K.; Kerminen, V.M.Kulmala, M. Meteorological dependence of size-fractionated number concentrations of urban aerosol particles. *Atmos. Environ.* 2006, 40, 1427–1440.
- [10] Chen, J, Lu, J. Avise, J.C. DaMassa, J.A. Kleeman, M.J. Kaduwela, A.P. Seasonal modeling of PM<sub>2.5</sub> in California's San Joaquin Valley. *Atmos. Environ.* 2014, 92, 182–190.