

Article

Prediction of Air Pollutant Concentration Based on One-Dimensional Multi-Scale CNN-LSTM Considering Spatial-Temporal Characteristics: A Case Study of Xi'an, China

Hongbin Dai ¹ , Guangqiu Huang ¹, Jingjing Wang ^{2,*}, Huibin Zeng ¹ and Fangyu Zhou ³

¹ School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, China; daihongbin@xauat.edu.cn (H.D.); gqhuang@xauat.edu.cn (G.H.); zenghuibin@xauat.edu.cn (H.Z.)

² College of Vocational and Technical Education, Guangxi Science & Technology of Normal University, Laibin 546199, China

³ School of Applied English, Chengdu Institute Sichuan International Studies University, Chengdu 611844, China; suansuanjunya@gmail.com

* Correspondence: wangjingjing@gxstnu.edu.cn; Tel.: +86-152-7710-7077

Abstract: Air pollution has become a serious problem threatening human health. Effective prediction models can help reduce the adverse effects of air pollutants. Accurate predictions of air pollutant concentration can provide a scientific basis for air pollution prevention and control. However, the previous air pollution-related prediction models mainly processed air quality prediction, or the prediction of a single or two air pollutants. Meanwhile, the temporal and spatial characteristics and multiple factors of pollutants were not fully considered. Herein, we establish a deep learning model for an atmospheric pollutant memory network (LSTM) by both applying the one-dimensional multi-scale convolution kernel (ODMSCNN) and a long-short-term memory network (LSTM) on the basis of temporal and spatial characteristics. The temporal and spatial characteristics combine the respective advantages of CNN and LSTM networks. First, ODMSCNN is utilized to extract the temporal and spatial characteristics of air pollutant-related data to form a feature vector, and then the feature vector is input into the LSTM network to predict the concentration of air pollutants. The data set comes from the daily concentration data and hourly concentration data of six atmospheric pollutants (PM_{2.5}, PM₁₀, NO₂, CO, O₃, SO₂) and 17 types of meteorological data in Xi'an. Daily concentration data prediction, hourly concentration data prediction, group data prediction and multi-factor prediction were used to verify the effectiveness of the model. In general, the air pollutant concentration prediction model based on ODMSCNN-LSTM shows a better prediction effect compared with multi-layer perceptron (MLP), CNN, and LSTM models.

Keywords: ODMSCNN; LSTM; atmospheric pollutant concentration prediction; deep learning; temporal and spatial characteristics



Citation: Dai, H.; Huang, G.; Wang, J.; Zeng, H.; Zhou, F. Prediction of Air Pollutant Concentration Based on One-Dimensional Multi-Scale CNN-LSTM Considering Spatial-Temporal Characteristics: A Case Study of Xi'an, China. *Atmosphere* **2021**, *12*, 1626. <https://doi.org/10.3390/atmos12121626>

Academic Editors: Jim Kelly and Antonio Donateo

Received: 25 October 2021

Accepted: 4 December 2021

Published: 6 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The problem of the atmospheric environment has received widespread public attention. The quality of the air has a greater impact on human health and the ecological environment. An increasingly serious air pollution issue has played a role in every corner of affecting people's daily lives. The air quality index (AQI) is an index system that quantitatively describes the air quality status. The higher the value and level, the more serious the air quality and pollution. It mainly includes fine particulate matter (PM_{2.5}), inhalable particulate (PM₁₀), sulfur dioxide (SO₂), nitrogen dioxide (NO₂), ozone (O₃), and carbon monoxide (CO). According to different limits of pollutants, it is converted into an air quality index according to different target concentrations. With urbanization and industrialization developments, air pollution in many cities hides an alarming reality throughout China. In 2019, WHO announced the top ten threats to global public health, of

which air pollution is considered the biggest threat [1]. In 2016, the number of premature deaths caused by $PM_{2.5}$ in the world was about 4.2 million [2]. Haze is one of the most serious environmental problems in China, the Chinese government plans to implement strict control measures to reduce the $PM_{2.5}$ concentration [3]. CO can indirectly aggravate the greenhouse effect and participate in the formation of near-surface photochemical smog, which is an important pollution component to measure the regional atmospheric environment [4]. A high concentration of O_3 will affect the human respiratory tract, cardiovascular and immune system, leading to asthma, respiratory tract infection, stroke, and arrhythmia [5–7]. PM_{10} , with a long retention time in the atmosphere and a relatively large specific surface area, is easy to carry with a large number of toxic and harmful substances. PM_{10} can enter human alveoli through the respiratory tract and even participate in blood circulation, which will cause more obvious harm to human body [8]. When NO_2 enters the alveoli, it causes bronchitis, pneumonia, emphysema, and SO_2 with high concentration [9]. It is estimated that the number of deaths caused by indoor and outdoor air pollution in China is 2.5 million people per year [10]. The reasonable assessment and control of air quality can help reduce the adverse effects of air pollution. Therefore, it is necessary to accurately predict the concentration of air pollutants to help management departments and potentially hazardous groups in various regions to reduce the impact of air pollutants.

In recent years, machine learning and deep learning models have been gradually applied to predict the concentration of air pollutants. Feng et al. [11] used artificial neural networks and wavelet transform to predict $PM_{2.5}$ concentrations based on geographic models. Ke et al. proposed a stack selection ensemble algorithm for $PM_{2.5}$ prediction [12], Zhou et al. exerted the seasonal gray model to predict the air quality indicators in the Yangtze River Delta of China [13]. Zhang et al. applied the gray multivariate convolution model to predict the daily $PM_{2.5}$ and PM_{10} concentrations in Shijiazhuang City [14]. Nouri et al. wielded principal component analysis and artificial neural network (ANN) to predict $PM_{2.5}$ concentration in Urmia, Iran [15]. Zhou et al. fused the multivariate correlation function to the Bayesian model averaging method (CBMA) combined with ANN for $PM_{2.5}$ prediction [16]. Du et al. utilized the multi-objective Harris Hawk optimization (MOHO) algorithm to predict the $PM_{2.5}$ and PM_{10} hour concentrations in Jinan, Nanjing, Chongqing [17]. Li et al. made use of integrated reinforcement learning to predict the daily concentration of $PM_{2.5}$ [18]. Guo et al. used Lagrangian and Bayesian methods to predict the hourly concentrations of PM_{10} and $PM_{2.5}$ in Xingtai [19].

Aiming at air quality prediction, the main models used in the existing research include the linear regression model [20] and the generalized weighted mixed model [21]. With the development of computer technology, machine learning (including deep learning) methods are increasingly used in concentration estimations due to their strong nonlinear modeling ability, such as k-nearest neighbor (KNN) [22], random forest (RF) [23], long-term memory network (LSTM) [24], and convolution neural network (CNN) [25]. These models all show better performance than traditional statistical models in predicting $PM_{2.5}$ concentration and have stronger nonlinear expression capabilities.

Many scholars have begun to try to use deep learning models for prediction. Guo et al. used deep learning methods such as recurrent neural network (RNN) and LSTM to predict $PM_{2.5}$ hourly concentration [26]. Sahin et al. predicted daily $PM_{2.5}$ and SO_2 concentrations in Istanbul using convolutional neural network (CNN) [27]. Sayeed et al. established a prediction model of ozone concentration 24 h in advance using a CNN [28]. WANG et al. combined a chi-square test (CT) and LSTM network model to predict AQI levels in Shijiazhuang, Hebei Province [29]. Liu et al. used industrial data to establish a factory aware attentional LSTM neural network (FAALSTM) model to predict $PM_{2.5}$ [30]. Pak et al. used the CNN and LSTM models to predict the daily average concentration of $PM_{2.5}$ in Beijing on the second day [31]. Wen et al. established a spatio-temporal convolutional short-term memory neural network expansion model (C-LSTM) to predict the $PM_{2.5}$ hourly concentration in Beijing and China [32].

The above research verifies the effectiveness of deep learning methods such as CNN and LSTM in the prediction of atmospheric pollutants, and also shows that combined prediction is beginning to be favored by many scholars. Many scholars have begun to pay attention to the influence of other factors on the prediction of atmospheric pollutant concentration. Nourani et al. used temperature, wind speed, humidity, pollutant concentration and other factors as input variables, and used ANN and adaptive neuro-fuzzy inference system (ANFIS) combined with the prediction of CO pollutant concentration [33]. Heydari et al. proposed a hybrid intelligent model based on LSTM and multiple optimization algorithm (MVO) to predict NO₂ and SO₂ in air pollutants [34]. Chen et al. predicted PM_{2.5} concentration in Zhejiang Province and found that meteorological factors such as temperature, air pressure, evaporation, and humidity have a significant correlation with PM_{2.5} concentration [35]. Zhang et al. found that O₃ hourly mass concentration is related to temperature and sun. There is a positive correlation among radiation, visibility, and wind speed, and a negative correlation with relative humidity and atmospheric pressure. The concentration of NO₂ is positively correlated with relative humidity and atmospheric pressure [36], Precipitation [37], season [38], precipitation [39], sunshine time [40], road transportation [41] and other factors have a significant impact on the concentration of air pollutants. Through the above research, it is found that meteorological factors have a significant impact on the concentration of air pollutants.

In summary, through combing the existing literature, it is found that the research has the following shortcomings: (1) most of the above research focuses on the prediction of a single or two atmospheric pollutants such as PM_{2.5} and PM₁₀, and such models were yet to find the relationship between multiple factors and atmospheric pollutants, so that the predictive performance of atmospheric pollutants cannot be fully utilized; (2) it is difficult for a single LSTM network to mine the relationship and characteristic information among the data; (3) the existing air pollutant concentration prediction models mostly start from a single or two pollutant concentrations to establish corresponding air pollutant concentration prediction models; (4) in the prediction of atmospheric pollutant data, most of the prediction objects are hourly concentration and daily concentration, but the current model fails to consider both predictions, and the prediction accuracy of some prediction models needs to be improved; (5) the existing research on air pollutant concentration prediction still has certain limitations. Most studies only select some important variables in the pollutant concentration data set, and then model the time information to predict the pollutant concentration. There is a lack of deep learning models for predicting the concentration of air pollutants using time information and multivariate time series data in the prediction of pollutant concentration. Therefore, an appropriate algorithm is necessary to be selected to model the irregular temporal information of concentration data and the spatial information of all variables [42,43].

The main contributions of this article are as follows:

(1) To establish an air pollutant prediction model that considers the temporal and spatial characteristics of pollutants and the combination of multiple factors. By combining the advantages of the two algorithms of ODMSCNN and LSTM, ODMSCNN has a strong ability to automatically extract features, and LSTM has a strong ability to deal with time series problems. ODMSCNN-LSTM-based air pollutant concentration prediction model is proposed;

(2) PM_{2.5}, PM₁₀, NO₂, SO₂, O₃, CO concentration data are selected as the characteristics of atmospheric pollutants to predict. The temperature (TEM), evaporation (EVP), minimum relative humidity (MI-RHU), maximum wind speed (MM-WIN), precipitation (PRE), sunshine duration (SSD), average wind speed (AV-WIN), and other factors are used as meteorological features;

(3) Perform daily concentration prediction and hourly concentration prediction and compare the performance of each model based on grammar correction MLP, CNN, LSTM, and ODMSCNN-LSTM models;

(4) The proposed model is compared from the perspective of grouped data and multivariate factors, and the performance of ODMSCNN-LSTM in different grouped data sets. The influence of atmospheric pollutant factors and meteorological factors on the prediction of atmospheric pollutant concentration are analyzed.

2. Study Area and Dataset

2.1. Study Area

The study area is located in Xi'an, a node city on the Fenwei Plain. The specific latitude and longitude are 107.40 to 109.49 degrees east and 33.42 to 34.45 degrees north. According to China's ecological environment status bulletin, the bulletin demonstrates that among 169 cities at the prefecture level and above in China, Xi'an was ranked seventh from the bottom in 2017 [44] and twelfth from the bottom in 2018 [45]; however, Xi'an ranked first from the bottom among national central cities in China in terms of air quality. On 25 January 2021, the People's Government of Xi'an City promulgated the Emergency Plan for Heavy Pollution Weather in Xi'an City [46]. As shown in Figure 1, the monthly air quality data for the region from 2014 to 2020 are calculated. The data indicate that the proportion of slight pollution in the last month is 45.2% and that moderate pollution in the last month has accumulated to 13 months.

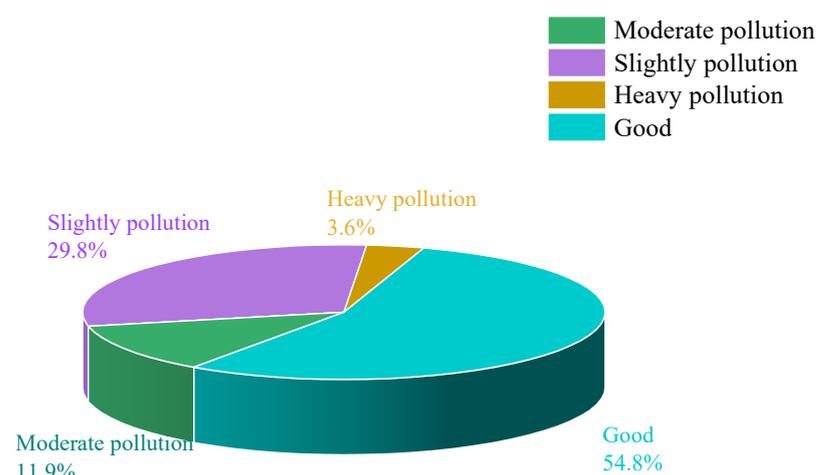


Figure 1. Distribution of air quality in Xi'an from 2014 to 2020.

2.2. Study Data

2.2.1. Air Quality Data

Since December 2013, China Environmental Protection Agency (EPA) has published the open-air quality observation data of China's ground monitoring stations. The research data of this paper is from the daily concentration data set of air pollutants ($PM_{2.5}$, PM_{10} , NO_2 , SO_2 , O_3 , CO) in Xi'an from 2 December 2014, to 30 December 2020, and the hourly concentration data of the above six air pollutants from 1 January 2019 to 30 December 2020.

2.2.2. Meteorological Data

The meteorological data in this paper comes from the Chinese weather website platform. Through data preprocessing, 15 kinds of meteorological factors are listed in this paper, and they are as follows: evaporation, daily average surface temperature, daily maximum surface temperature, daily minimum surface temperature, daily average wind speed, the wind direction of day and day wind speed, day and night wind speed and direction, daily precipitation, daily average pressure, daily maximum pressure, daily minimum pressure, sunshine hours, daily average relative humidity and daily minimum relative humidity, and the seasons.

2.3. Analysis of Main Data Characteristics

Through the Pearson correlation analysis of the collected air pollution concentration data, as shown in Figure 2, the coefficient between $PM_{2.5}$ and PM_{10} is 0.89, which proves that $PM_{2.5}$ and PM_{10} are highly correlated. CO and SO_2 are highly correlated with a coefficient of 0.81. $PM_{2.5}$ and PM_{10} are highly correlated with CO, with a coefficient of 0.75. SO_2 is highly correlated with $PM_{2.5}$ (0.6) and PM_{10} (0.64). NO_2 is highly correlated with $PM_{2.5}$ and PM_{10} , and both coefficients are 0.65. O_3 has a moderate correlation with $PM_{2.5}$, SO_2 , and CO, and a weak correlation with PM_{10} and NO_2 .

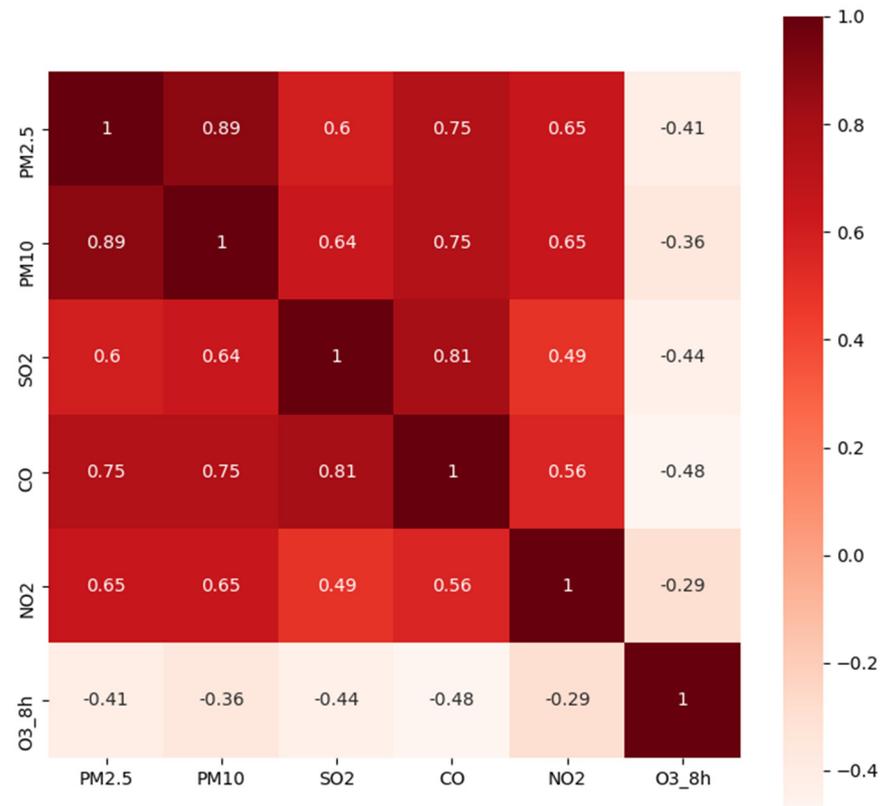


Figure 2. Pearson Analysis of six air pollutants.

2.4. Data Processing

2.4.1. Division of Data Set

The data set needs to be trained by the divided input model, otherwise the prediction model will have no additional data for effect evaluation, and the training results may be overfitted due to training on all data. In the experiment, each data set is divided into training set and test set, and then the training set is divided into training set and verification set. The data ratio of the training set, test set and verification set is 6:2:2. Among them, the training set mainly learns sample data sets, and builds a classifier by matching some parameters. The establishment of a classification method is mainly used to train the model. The validation set is used to determine the network structure or the parameters that control the complexity of the model, and to select the number of hidden units in the neural network. The test set is used to test the performance of the final selected optimal model. It is mainly to test the resolution ability (recognition rate, etc.) of the trained model.

2.4.2. Raw Data Processing

① Identification and Processing of Abnormal Data

The occurrence of abnormal data may be due to errors in the process of collecting and recording data. Abnormal data will affect the prediction accuracy of the model, so it is necessary to identify and process the abnormal data. Abnormal data is found through outlier detection. Here, the statistical quartile analysis method is used to identify the abnormal data. The first quartile and the third quartile of the variable are solved first. If any value is less than the first quartile or greater than the third quartile, the value is judged as an outlier. Then, we use the horizontal processing method to correct abnormal data [47].

The calculation formula of the horizontal processing method is shown in Formulas (1) and (2):

If,

$$\begin{cases} |y_i - y_{i-1}| > \varepsilon_a \\ |y_i - y_{i+1}| > \varepsilon_a \end{cases} \quad (1)$$

Then,

$$y_t = \frac{y_{t+1} + y_{t-1}}{2} \quad (2)$$

Among them, y_i is the concentration of air pollutants in a certain day or hour, y_{i-1} is the concentration of air pollutants in the previous day or hour, and y_{i+1} is the concentration of air pollutants in the next day or hour, ε_a is the threshold value.

② Data Normalization

Due to the different meanings and dimensions of physical quantities such as air pressure and evaporation, the input to the prediction model will have an impact, so it is necessary to normalize such data. The input of normalized data into the prediction model can effectively reduce the training time of the model, accelerate the convergence speed of the model, and further improve the prediction accuracy of the model. The normalized calculation formula of the data is shown in formula (3). This method realizes the equal scaling of the original data [48]:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

Among them, x_{norm} is the normalized value, x is the original data, x_{min} is the minimum value in the original data, x_{max} is the maximum value in the original data, and the size of the normalized data is constrained to be between [0,1].

2.4.3. Data Encoding

Through the above data feature analysis, it can be concluded that the forecast of atmospheric pollutant concentration is affected by temperature, weather factors, and human factors. Among them, temperature, weather conditions, date types, and heating are qualitative characteristics. These qualitative indicators should be mapped to [0,1] interval, converted into quantitative data, as shown in Table 1.

Table 1. Quantification of candidate input variables for air pollutant concentration prediction model.

Type	Variable	Unit	Min	Max	Transformed
Air quality data	PM _{2.5}	µg/m ³	6	296	[0,1]
	PM ₁₀	µg/m ³	11	581	[0,1]
	SO ₂	µg/m ³	3	44	[0,1]
	CO	mg/m ³	1	2.7	[0,1]
	NO ₂	µg/m ³	8	111	[0,1]
	O _{3_8h}	µg/m ³	6	245	[0,1]
Meteorological data	Average air pressure	0.1 hpa	9487	9947	[0,1]
	Maximum daily pressure	0.1 hpa	9509	9981	[0,1]
	Lowest daily pressure	0.1 hpa	9465	9918	[0,1]
	Precipitation at 20-8 o'clock	0.1 mm	0	472	[0,1]
	Precipitation at 8-20 o'clock	0.1 mm	0	434	[0,1]
	Cumulative precipitation from 20-20 o'clock	0.1 mm	0	698	[0,1]
	Mean surface temperature	0.1 °C	−39	417	[0,1]
	Average relative humidity	1%	14	99	[0,1]
	Sunshine duration	0.1 h	0	130	[0,1]
	Average temperature	0.1 °C	−67	346	[0,1]
	Average wind speed	0.1 m/s	3	80	[0,1]
	Small scale evaporation	0.1 mm	0	98	[0,1]
	Large scale evaporation	0.1 mm	0	270	[0,1]
	Season	-	1	4	[0,1]
Heating	-	0	1	[0,1]	

3. Method

3.1. One-Dimensional Multi-Scale Convolutional Neural Network (ODMSCNN)

The convolutional neural network is successfully applied to the direction of image recognition, verifying that the network has a powerful effect on the feature extraction ability of feature maps, and this article needs to extract the spatial and temporal features of atmospheric pollutant concentration data and meteorological factors. This paper analyzes the data set and finds that the characteristics of the data are multi-features, which are expressed in the form of numerical values instead of feature maps. Therefore, this paper first preprocesses the data, merges the features of the data into a feature map, and finally inputs them into the convolution neural network to extract the spatial and temporal features.

Taking single-factor SO₂ as an example, the spatio-temporal feature extraction of SO₂ is shown in Figure 3. The feature map is traversed from left to right on the data feature axis through a one-dimensional multi-scale convolution kernel (1 × 3, 1 × 5, 1 × 7) to complete the convolution operation; the number of steps is 1, with different convolution kernels. The output feature vectors are spliced and fused, and a single-factor spatial feature relationship is obtained for this purpose. On the time axis, as the convolution kernel traverses from top to bottom to complete the convolution operation, the number of steps is 1, and the local trend of single factor changes over time can be obtained. Finally, the spliced and fused feature vectors are merged in the data feature direction, and the spatio-temporal features of the multi-site SO₂ are output.

The following is the formula derivation of ODMSCNN's convolution operation on the special whole. The feature map contains *N* sample data and *N* air pollutant factors. Then the feature map formula of single factor *i* is as follows:

$$X_i = [x_i^1, x_i^2, x_i^3, \dots, x_i^N]^T \tag{4}$$

$$X_i^{t:t+T-1} = [x_i^t, x_i^{t+1}, x_i^{t+2}, \dots, x_i^{t+T-1}]^T \tag{5}$$

In the formula, $X_i^t = [x_i^t, x_i^{t+1}, x_i^{t+2}, \dots, x_i^{t+T-1}] \in R$ is the vector of the single factor i at time t , $X_i^{t:t+T-1}$ represents the T group vector of X_i in the time zone $[t, t + T - 1]$, and T represents the matrix transpose.

The convolution operation multiplies the weight matrix W_j and $X_i^{t:t+T-1}$

(1) Single-factor spatial feature relationship: multiply W_j by $X_i^{t:t+T-1}$ on the data feature axis;

(2) Single factor time change feature: multiply W_j by $X_i^{t:t+T-1}$ on the time feature axis.

When the first convolution kernel traverses the entire feature map on the time axis, and the number of steps is 1, the feature vector a_i^j is obtained, the size of which is $N - T + 1$, and the feature vector obtained by multiple convolution kernels Z is merged with the size of $[N - T + 1] \times Z$ in the data feature direction A_i , A_i represents the single-factor spatio-temporal characteristic matrix.

$$a_i^j = [a_{t+T-1}^j, a_{t+T}^j, a_{t+T+1}^j, \dots, a_N^j] \tag{6}$$

$$A_i = [a_n^1, a_n^2, a_n^3, \dots, a_n^Z] \tag{7}$$

So far, the single-factor spatio-temporal feature extraction is completed, but the data set also contains other features, such as NO_2 , O_3 , CO , etc. Our process includes M factors, so we can extract the M factors through the same operation as above, and then they can be extracted. Single-feature spatio-temporal feature matrix, and then linear splicing and fusion of them, and finally forming a multi-factor fusion spatio-temporal feature matrix A , as shown in Equation (8):

$$A = [A_1, A_2, A_3, \dots, A_M] \tag{8}$$

Based on the ODMSCNN convolution neural network, the space-time characteristics of air quality data are extracted. This method makes a simple transformation of the two-dimensional feature map to form a side-by-side one-dimensional feature map, which makes the network training show better generalization ability. The method of automatic feature extraction by convolutional neural network replaces the traditional manual feature selection method, which makes feature extraction have a more comprehensive and deeper effect.

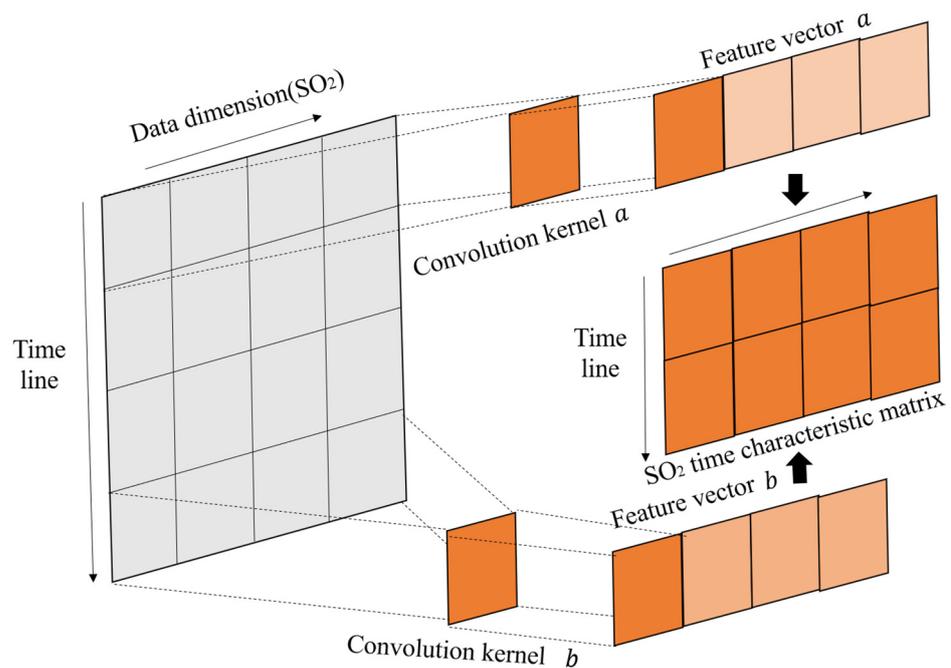


Figure 3. One-dimensional convolution feature extraction process diagram.

3.2. Long- and Short-Term Memory Neural Network

LSTM is an improvement on the RNN. There is a long-term dependence on the problem of RNN's disappearance and the explosion of gradients during training. LSTM can effectively solve this problem, it introduces a gate mechanism, which makes LSTM have a longer-term memory than RNN and can learn more effectively. In LSTM, each neuron is equivalent to a memory cell (cell, c_t). LSTM controls the state of the memory cell through a "gate" mechanism, increasing or deleting the information in it. The structure of LSTM is shown in Figure 4.

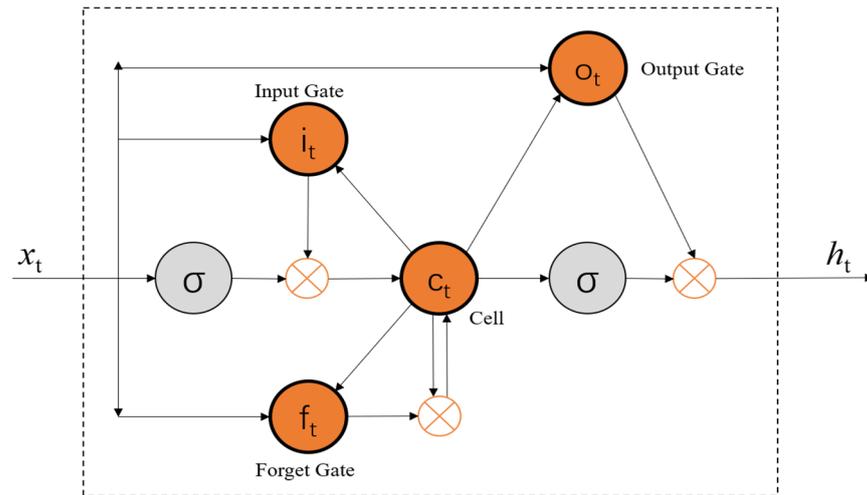


Figure 4. LSTM unit structure.

In the LSTM cell structure, the Input Gate (i_t) is used to determine what information is added to the cell, and the Forget Gate (f_t) is used to determine what information is deleted from the cell. The output gate (o_t) is used to determine what information is output from the cell. The complete training process of LSTM is that at each time t , the three gates receive the input vector x_t at time t and the hidden state h_{t-1} of the LSTM at time $t - 1$ and the information of the memory unit c_t , and then perform the received information Logical operation, the logical activation function σ decides whether to activate it, and then synthesize the processing result of the input gate and the processing result of the forgetting gate to generate a new memory unit c_t , and finally obtain the final output result h_t through the nonlinear operation of the output gate. The calculation formula for each process is as follows.

Input gate calculation formula:

$$i_t = \sigma(W_{xi}^T x_t + W_{hi}^T h_{t-1} + b_i) \tag{9}$$

Forget Gate calculation formula:

$$f_t = \sigma(W_{xf}^T x_t + W_{hf}^T h_{t-1} + b_f) \tag{10}$$

Output Gate calculation formula:

$$o_t = \sigma(W_{xo}^T x_t + W_{ho}^T h_{t-1} + b_o) \tag{11}$$

Memory unit calculation formula, Hidden state:

$$c_t = f_t \times c_{t-1} + i_t \times \tanh(W_{xc}^T x_t + W_{hc}^T h_{t-1} + b_c) \tag{12}$$

Hidden state calculation formula:

$$h_t = o_t \tanh(c_t) \tag{13}$$

Among them, σ is generally a nonlinear activation function, such as a sigmoid or tanh function. $W_{xi}, W_{xf}, W_{xo}, W_{xc}$ are the weight matrices of nodes connected to the input vector W_i for each layer, $W_{hi}, W_{hf}, W_{ho}, W_{hc}$ are the weight matrices connected to the previous short-term state $ht-1$ for each layer, b_i, b_f, b_o, b_c are the offset terms of each layer node.

In short, the input gate in LSTM can identify important inputs, and the forget gate can reasonably retain important information and extract it when needed. Therefore, this feature of LSTM can effectively identify long-term patterns such as time series, making training convergence faster.

3.3. ODMSCNN-LSTM-Based Atmospheric Pollutant Concentration Prediction Model

As shown in Figure 5, the model is composed of two parts: ODMSCNN and LSTM. The temporal and spatial features of multiple variable data are extracted through ODMSCNN, and then pass to the LSTM layer. The LSTM layer models the spatio-temporal feature information input by the ODMSCNN layer, and then ODMSCNN and LSTM pass through the connection layer and predict the concentration of air pollutants.

First of all, the starting layer of the model is composed of ODMSCNN to accept multiple variable inputs of atmospheric pollutant accumulation data, such as relevant atmospheric pollutant factors and meteorological factors. The factor variables are input into the convolutional neural network, and the spatio-temporal feature relationship among each of them is extracted through ODMSCNN. There are multiple hidden layers involved in the feature extraction process of ODMSCNN. The deeper the number of layers, the deeper the network depth. The extracted features are used as the input of the LSTM layer. Secondly, the hidden layer content of CNN consists of a convolutional layer, an activation function and a pooling layer. The convolution operation simulates the response of a single neuron to visual stimulation, that is, when processing the time series, each neuron processes the received data and extracts the data features. The convolution operation can reduce the number of parameters and make the CNN-LSTM network deeper. Finally, dimensionality reduction and feature extraction are performed on the data through the CNN convolutional layer, and the single-factor spatio-temporal feature relationship extracted by ODMSCNN is simply linearly spliced and fused to obtain the mutual spatio-temporal feature relationship of multiple factors. Avoid over-fitting phenomenon, improve the robustness in the feature extraction process, and apply LSTM to solve the problem of long-term data memory, and realize the source data fusion perception and multi-layer perception.

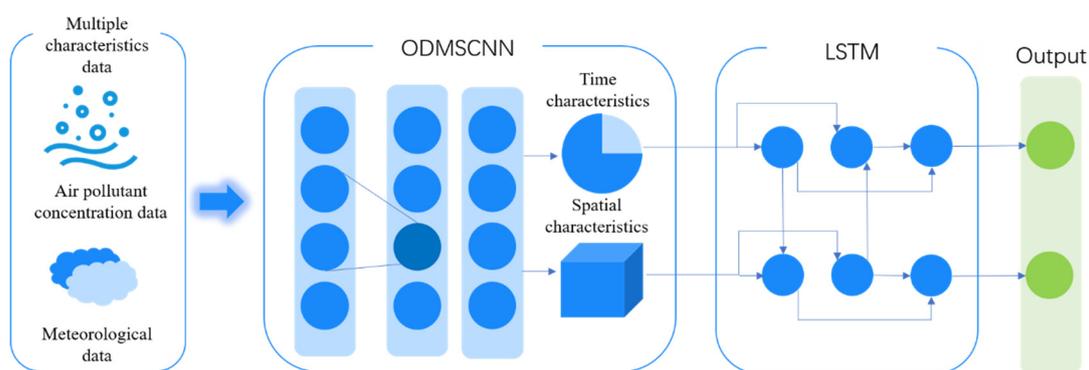


Figure 5. ODMSCNN-LSTM model structure.

3.4. Prediction Process of Atmospheric Pollutant Concentration

Figure 6 is the forecasting process of atmospheric pollutant concentration, including five parts of data pre-processing, extract features, nodal features, model test and data prediction.

(1) Data processing preparation. Data preprocessing is performed in the original data of atmospheric pollutant concentration prediction, and the data set is divided into

training set, verification set, and test set. The training set is mainly used to train the model. The validation set is used to adjust the parameters of the concentration prediction model, determine the network structure of ODMSCNN and LSTM, and select the number of hidden units. The test set is used to verify the performance of the model;

(2) Feature extract. Input the training set into ODMSCNN to extract spatio-temporal features and input the extracted spatio-temporal features into LSTM for training to find the optimal structure and parameters of the model;

(3) Model tuning. Use training set for model tuning. The experiment process adopts the control variable method. Firstly, determine the initial number of layers and parameters of the ODMSCNN-LSTM model, fix the structure of LSTM, and the number of layers and parameters of ODMSCNN. Secondly, after the best structure of ODMSCNN is determined, adjust the number of layers and parameters of LSTM. Finally, choose the best structure of ODMSCNN-LSTM, and judge whether loss has converged ($loss = mae$). If it has converged, move to the next step. However, if it has not converged, continue with the previous step until the optimal structure and parameters of the model are found;

(4) Model test. Put the validation set into the trained model for prediction. The evaluation index is used to evaluate the prediction result to determine whether the prediction effect is the best, that is, whether the predicted value achieved the best fitting effect with true value. If the fitting result is the best, the prediction is completed. However, if the fitting effect is not good, go back to the second step to continue iterating the model parameters, and verify the performance and accuracy of the method, analyze and compare the experimental results;

(5) Data prediction. Use the test set data to make predictions. The predicted value is compared with the true value to determine whether the optimal value is obtained. When the optimum is reached, the predicted value is output.

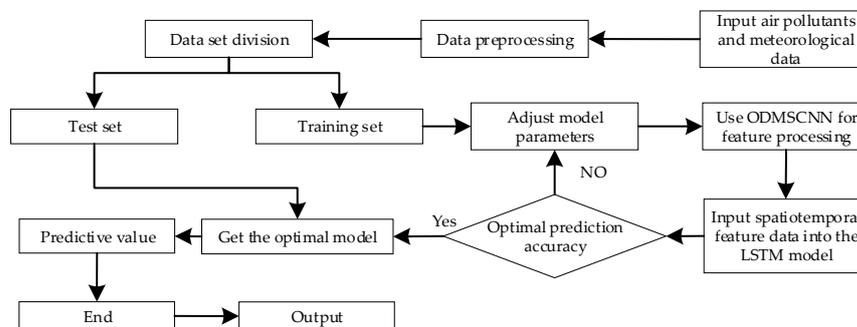


Figure 6. Flow chart of air pollutant concentration prediction model.

3.5. Model Evaluation Indicators

Root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and goodness of fit R^2 are two commonly used cross-validation indicators. This study uses these four indicators to evaluate the model.

The specific formula derivation of the LSTM is as follows:

$$RMSE = \sqrt{\sum_{i=1}^N \frac{(\hat{y}_i - y_i)^2}{N}} \tag{14}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |X_i - \bar{X}|^2 \tag{15}$$

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{X_i - \bar{X}_i}{X_i} \right| \tag{16}$$

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y} - y_i)^2} \quad (17)$$

where \hat{y}_i is the predicted values of air pollutants, y_i is the true values of air pollutants, and N is the number of test samples. Generally, the larger the value of RMSE and MAE, the greater the error and the lower the prediction accuracy of the model. MAPE is the most intuitive criterion for prediction accuracy. When MAPE tends to 0%, it means the model is perfect. When MAPE tends to 100%, it means that the model is inferior. Generally, it can be considered that the prediction accuracy is higher when the MAPE is less than 10% [49].

4. Results

4.1. Daily Concentration Data Prediction

The daily concentration data from January 2014 to December 2020 is selected. Table 2 shows the daily air pollutant concentration prediction results of different models. In terms of prediction, the order of RMSE is arranged in descending order of MLP, CNN, LSTM, and ODMSCNN-LSTM.

As shown in Figure 7, the four prediction models are used to predict the concentrations of six kinds of air pollutants, respectively. The forecast results of the six air pollutant concentrations in four different models are drawn. The green and red lines of the prediction results represent the actual and ODMSCNN-LSTM predicted air pollutant concentrations, respectively. By comparing the predicted values of the four models with their corresponding true values, it is found that the predicted value of ODMSCNN-LSTM shows its more accurate prediction performance compared with the other three models. It can be seen from the six air pollutant concentration prediction curves that the constructed ODMSCNN-LSTM prediction model well reflects the timeliness and nonlinearity of the air pollutant concentration distribution. This model responds quickly both to short-term and long-term predictions, and it performs well even when the six kinds of air pollutant concentration suddenly change. Due to the great change of air pollutant concentration between 2014 and 2020, the RMSE value and the maximum average error value of the ODMSCNN-LSTM model for daily concentration prediction of air pollutants are greater than those of the subsequent hourly data, but the prediction effect of the ODMSCNN-LSTM model is better than other prediction models.

Table 2. Results of different models for daily prediction of atmospheric pollutants (RMSE, MAPE, MAE).

Model	Metric	PM _{2.5}	PM ₁₀	NO ₂	SO ₂	O ₃	CO
MLP	RMSE	32.95	35.4	15.54	4.4	18.74	0.39
	MAE	21.27	25.67	12.53	3.6	14.88	0.33
	MAPE	0.48	0.28	0.26	0.39	0.44	0.48
CNN	RMSE	35.61	35.82	14.64	3.59	27.06	0.42
	MAE	25.37	28.61	11.82	2.9	21.28	0.33
	MAPE	0.56	0.32	0.25	0.33	0.74	0.47
LSTM	RMSE	40.8	34.66	13.62	4.39	24.55	0.42
	MAE	28.3	25.96	10.85	3.76	19.72	0.35
	MAPE	0.61	0.25	0.25	0.41	0.58	0.48
ODMSCNN-LSTM	RMSE	16.14	31.69	11.8	2.23	18.49	0.19
	MAE	11.93	23.83	8.19	1.39	14.82	0.14
	MAPE	0.3	0.23	0.19	0.15	0.39	0.2

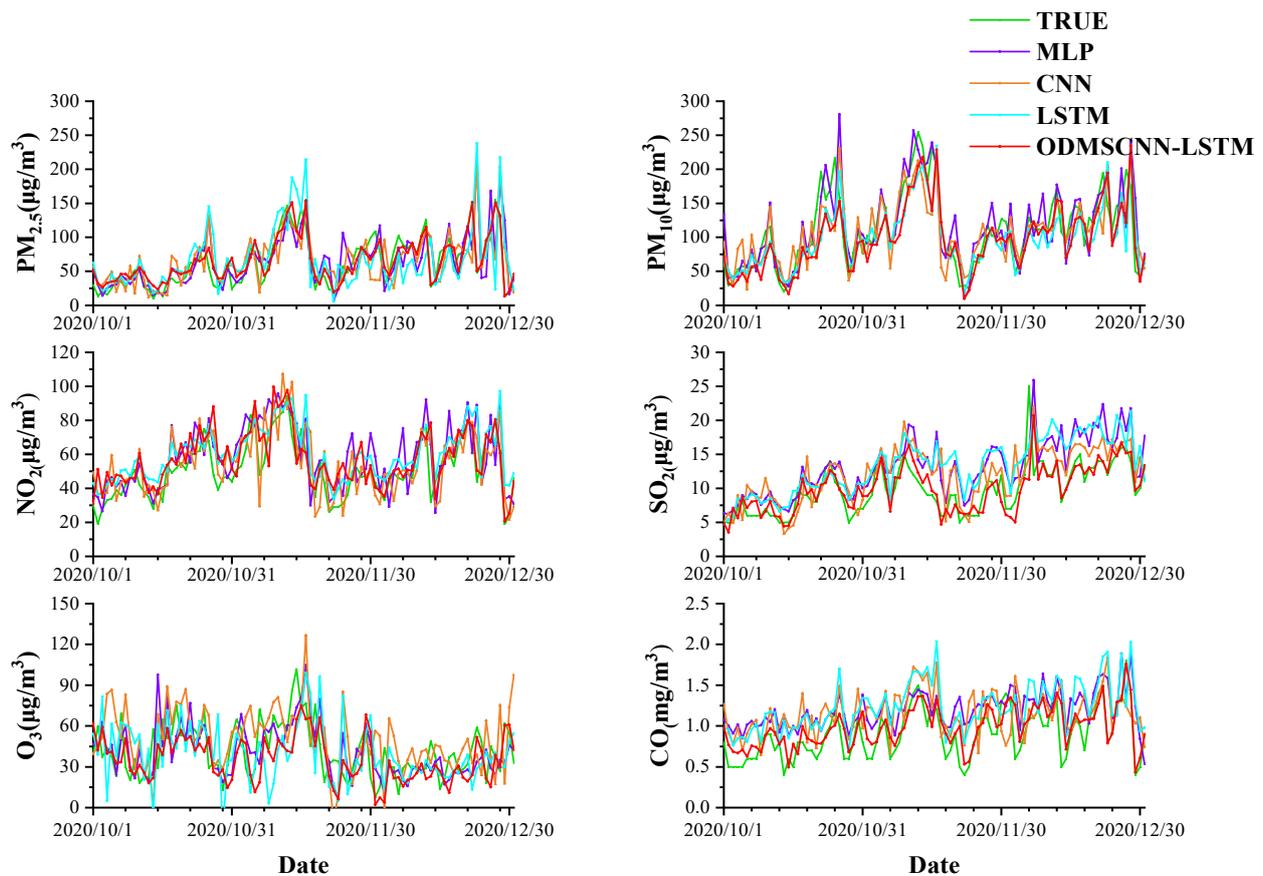


Figure 7. Comparison of daily data results of four different prediction models.

4.2. Hourly Concentration Data Prediction

In order to better verify the validity and prediction ability of the model, the hourly concentration data from 1 January 2019 to 31 December 2020 were selected in this paper. The data from the test set were used as a new prediction range and the performance of the air pollutant prediction model was evaluated again.

According to the data in Table 3, in terms of model performance, the average values of RMSE, MAE, and MAPE, evaluation indexes of six kinds of air pollutants in four prediction models, are arranged in the order of CNN, MLP, LSTM, MSCNN-LSTM. Then, the three evaluation indexes RMSE, MAE, and MAPE of the four models for hourly data are compared. It is found that the all the three indexes provided by the MSCNN-LSTM model for all the six air pollutants including $PM_{2.5}$ are the lowest.

As shown in Figure 8, the four prediction models are used to predict the concentration of six air pollutants, and their respective prediction results are plotted. The figure shows the concentration prediction results of the six air pollutants for 6 days (144 h in total). The green and red lines represent the actual and ODMSCNN-LSTM predicted air pollutant concentrations, respectively. It can be observed that the ODMSCNN-LSTM model performs best among the four models. Through the comparison of the prediction results of the four models in Figure 8 the hourly concentration prediction model based on ODMSCNN-LSTM well predicts the 6-day air pollutant concentration distribution and shows good performance. Comparing the hourly concentration data with the daily concentration data. The MSCNN-LSTM model shows a better prediction effect because the average values of its three indexes RMSE, MAE and MAPE for the six air pollutants have increased by 68.87%, 66.36%, and 63.09%, respectively. The reason is that the daily concentration data has fewer samples than hourly data, and the daily concentration data fluctuates greatly, and ODMSCNN-LSTM performs better for large samples of continuous time series data.

Table 3. Using hourly concentration data to predict, the prediction performance comparison of MLP, CNN, LSTM and ODMSCNN-LSTM models.

Model	Metric	PM _{2.5}	PM ₁₀	NO ₂	SO ₂	O ₃	CO
MLP	RMSE	7.357	9.302	3.699	1.216	7.482	0.077
	MAE	5.994	6.98	2.817	0.957	6.405	0.059
	MAPE	0.167	0.074	0.094	0.073	0.52	0.071
CNN	RMSE	6.143	21.9	5.727	1.98	4.567	0.104
	MAE	4.488	20.47	4.423	1.672	3.374	0.074
	MAPE	0.09	0.26	0.173	0.123	2.22	0.071
LSTM	RMSE	5.498	10.004	4.694	1.236	5.047	0.088
	MAE	3.952	7.986	3.821	1.007	4.226	0.072
	MAPE	0.112	0.09	0.106	0.079	0.347	0.102
ODMSCNN-LSTM	RMSE	4.966	7.223	3.588	1.091	4.129	0.06
	MAE	3.487	5.287	2.712	0.832	3.155	0.0506
	MAPE	0.066	0.0628	0.076	0.063	0.219	0.068

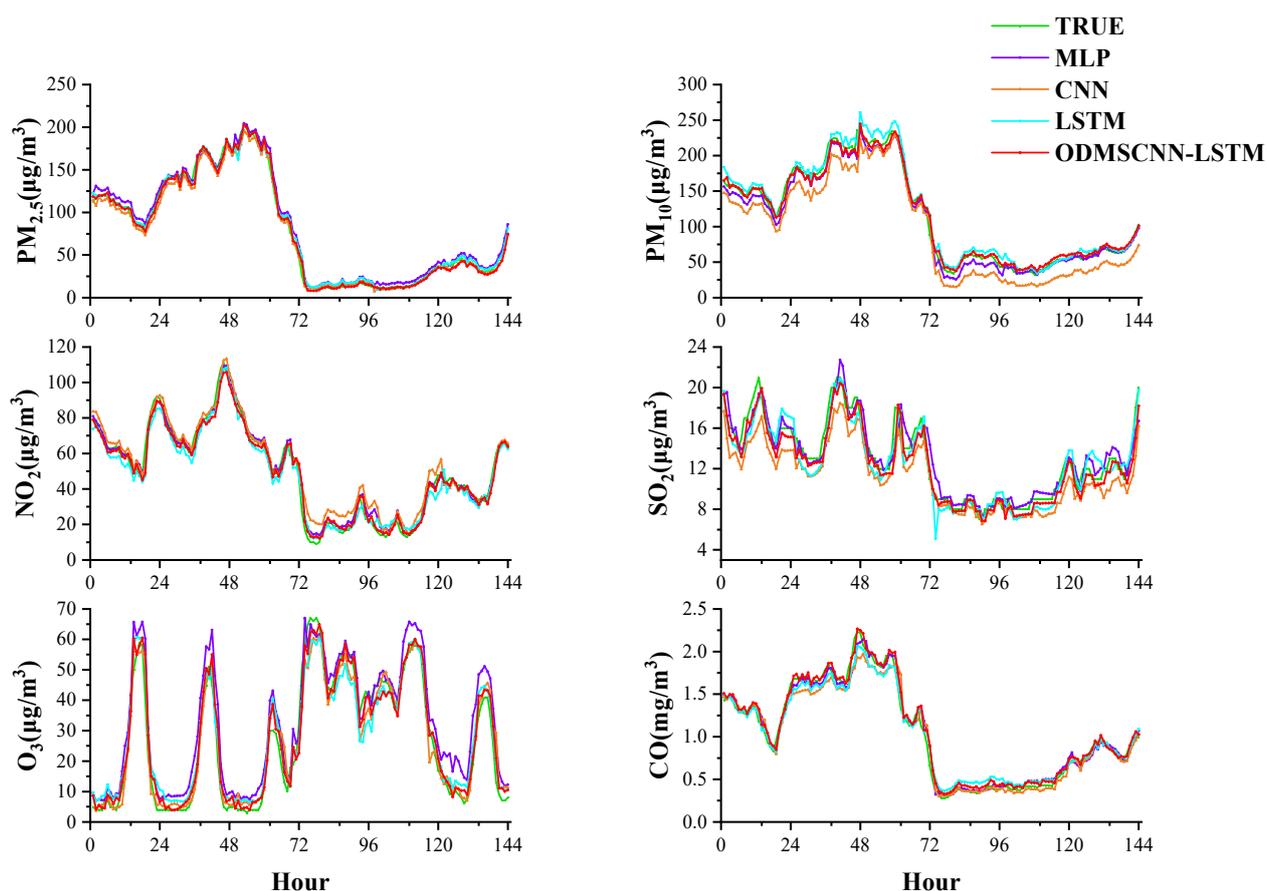


Figure 8. Comparison of time data results of four different prediction models.

4.3. Grouped Data Comparison

According to the above research, it is found that the overall accuracy and error of the hourly data of air pollutants are better than those of the daily concentration data. Therefore, in order to test the performance of ODMSCNN-LSTM in the hourly concentration prediction, the grouped data comparison method is used to divide all the data into three

groups to jointly verify the model. The first group of data includes the data set from 1 January 2019 to 31 December 2019, the second group of data consists of the data set from 1 January 2020 to 31 December 2020, and the third group uses data set from 1 January 2020 to 31 December 2020. The three groups of data are divided into the training set, validation set, and test set according to the ratio of 6:2:2. Additionally, the training set data is used to verify the prediction effect of each group of data.

The study found that the prediction performance of the same data in every model is different. In terms of prediction performance, the accuracy of the first group and the overall prediction performance of the training and validation sets are sorted in ascending order of MLP, CNN, LSTM, and ODMSCNN-LSTM. The prediction accuracy of the second group is MLP, CNN, LSTM, and ODMSCNN-LSTM from low to high. The accuracy of the third group is slightly different, arranged in ascending order of MLP, CNN, ODMSCNN-LSTM, and LSTM. Compared with the prediction performance of overall data set in the third group, the first and the second group have similar patterns with the overall prediction model. Using four deep learning models for training and verification of prediction accuracy, the results show that ODMSCNN-LSTM has the highest prediction accuracy on most test sets, and its prediction performance is better than other models for verification.

As shown in Figures 9–11, the RMSE, MAE, and MAPE values of the six air pollutants in the four models are shown, respectively. The prediction performance of the same pollutant using different data sets is different in the same model. This study selects the best-performing ODMSCNN-LSTM as an example. RMSE and MAE of the four air pollutants PM_{2.5}, PM₁₀, CO and SO₂ are arranged in descending order of 2019, 2020, 2019–2020. The RMSE and MAE of O₃ are arranged in descending order of 2019, 2020, 2019–2020, and those of NO₂ are arranged in descending order of 2020, 2019, 2019–2020. The four atmospheric pollutants MAPE, PM₁₀, CO, O₃, and SO₂ are arranged in descending order of 2019, 2020, and 2019–2020.

In general, there may be many reasons why the prediction performance of the daily concentration data set in 2019 is worse than that in 2020. The concentration of air pollutants in 2019 is usually higher and there is a large fluctuation, which may lead the prediction model to make underestimation or overestimation. Therefore, RMSE, MAE and MAPE increase in the annual data prediction. Relevant laws, regulations and policies formulated by the Chinese government and the Xi'an municipal government may be one reason for that. Another reason is that there are less company-made and man-made emissions since the outbreak of COVID-19. In 2020, the concentration of air pollutants decreased, the overall air quality became better, and the peak value fluctuated slower. The spatial variation of air pollutant data set in 2020 tends to be stable compared with that in 2019, so prediction on air pollutants in 2020 is relatively easier. Which may be the main reason for smaller error of data prediction in 2020. Another reason for the overall good performance of 2019–2020 data set is related with certain performance of the deep learning model, which reflects that the more the data, the better the performance.

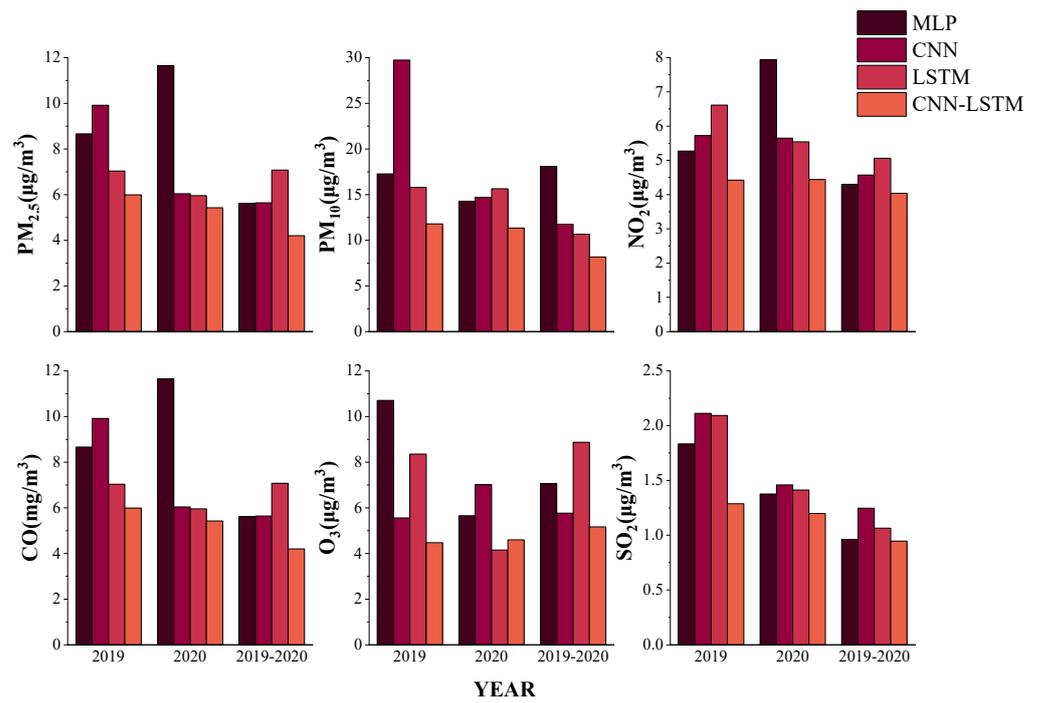


Figure 9. Comparison of RMSE predicted by four models for six air pollutants in 2019, 2020, and 2019–2020.

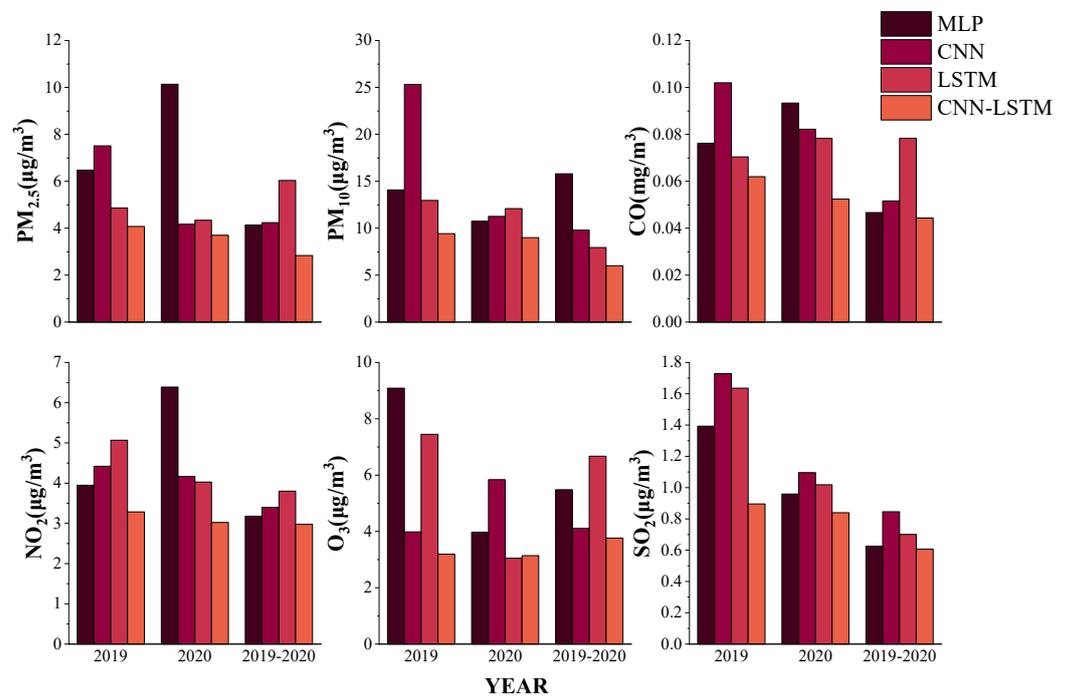


Figure 10. Comparison of the four models predicted MAE for six air pollutants in 2019, 2020, and 2019–2020.

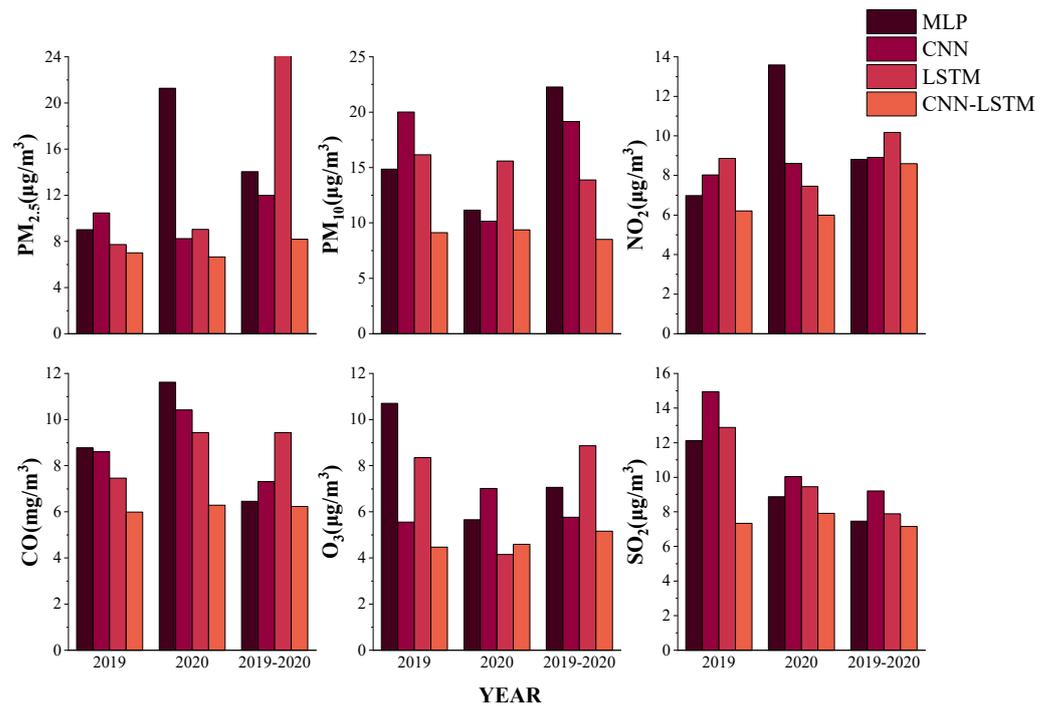


Figure 11. Comparison of the four models predicted MAPE for six air pollutants in 2019, 2020, and 2019–2020.

4.4. Comparison of Multiple Factors

After comparing the daily concentration data and hourly concentration data from the four models, it is found that ODMSCNN-LSTM model is better than others in overall performance. In order to test the efficiency of the prediction ability of the ODMSCNN-LSTM model, this paper continues to use the test set data of hourly data concentration to explore the influence of different factors on the prediction of air pollutant concentration.

The air pollutant concentration in Xi'an is affected by multiple factors. In order to better verify the ODMSCNN-LSTM model, a prediction model for different influencing factors has been established. The ODMSCNN-LSTM model was compared with air pollutants, meteorological factors, and overall factors to further verify the influence of multiple factors on the effect of the prediction model. All the data can be divided into three categories according to the types of influencing factors. The first category is to forecast only when meteorological factors are added. In the second category, only air pollutants are added as influencing factors. The third type is the whole factor which includes the meteorological factor and the air pollutant factor at the same time.

The paper draws the prediction curve of air pollutant concentration under different factors. Figure 12 shows the comparison between: the actual value and the predicted value obtained after adding meteorological factors, and Figure 13 shows the comparison between the actual value and the predicted value obtained after adding the air pollutants, and Figure 14 shows the comparison between the actual value and the predicted value obtained after adding both the meteorological factors and the air pollutants. It is found that the prediction with meteorological factors is less effective than the other two types. Prediction, with only meteorological factors added, is better than the other two types of prediction. Furthermore, the three index values of MAE, MAPE, and RMSE are lower. By comparing the three types of influencing factors, it is found that prediction accuracy varies with different air pollutants.

Different air pollutant models have different performances in reducing errors and improving the consistency of changes in different factors. Adding meteorological factors may not be able to effectively improve the prediction accuracy. The decrease in the number of features of different factors leads to a decrease in the overall data for model training,

which may limit the prediction ability of ODMSCNN-LSTM. Taking into account the annual characteristics of the concentration of air pollutants, the prediction error may be related to the type of air pollutants and the degree of dispersion of the air pollutant concentration in different seasons.

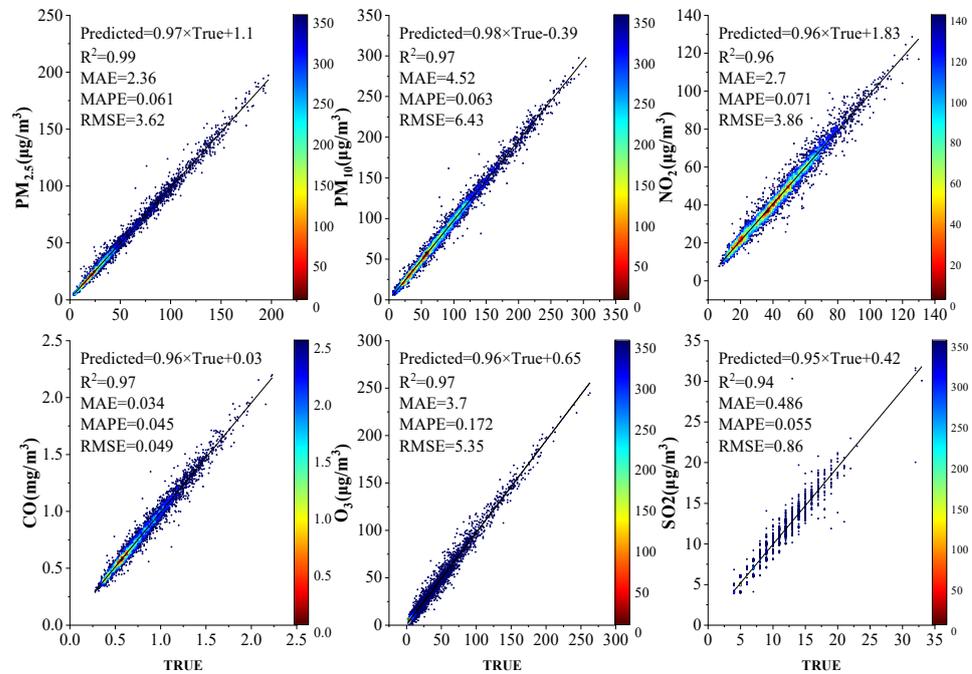


Figure 12. Comparison of actual and predicted values with meteorological factors.

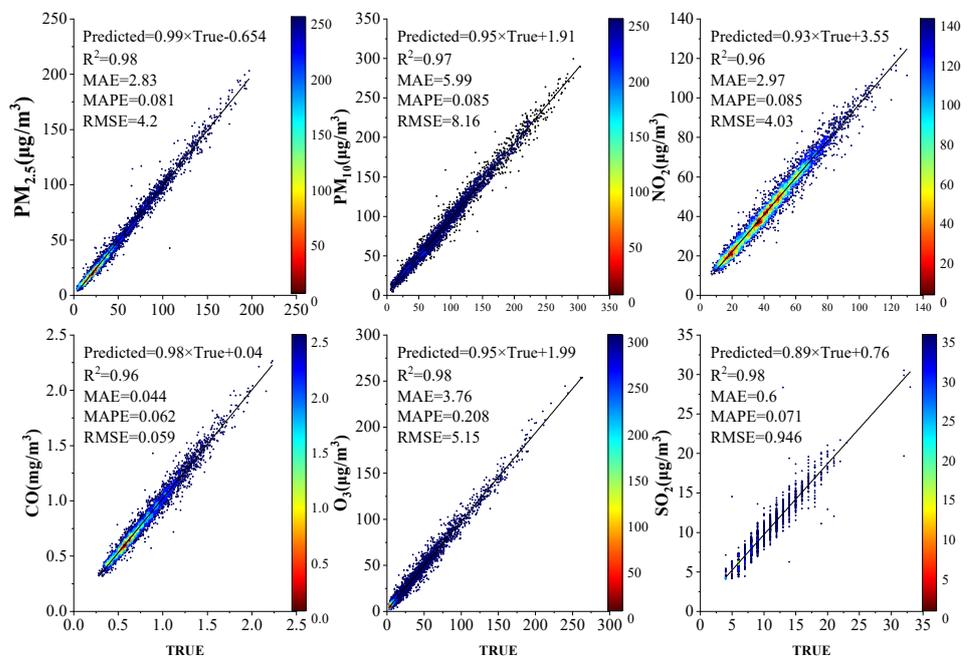


Figure 13. Comparison of actual and predicted values with air pollutant factors added.

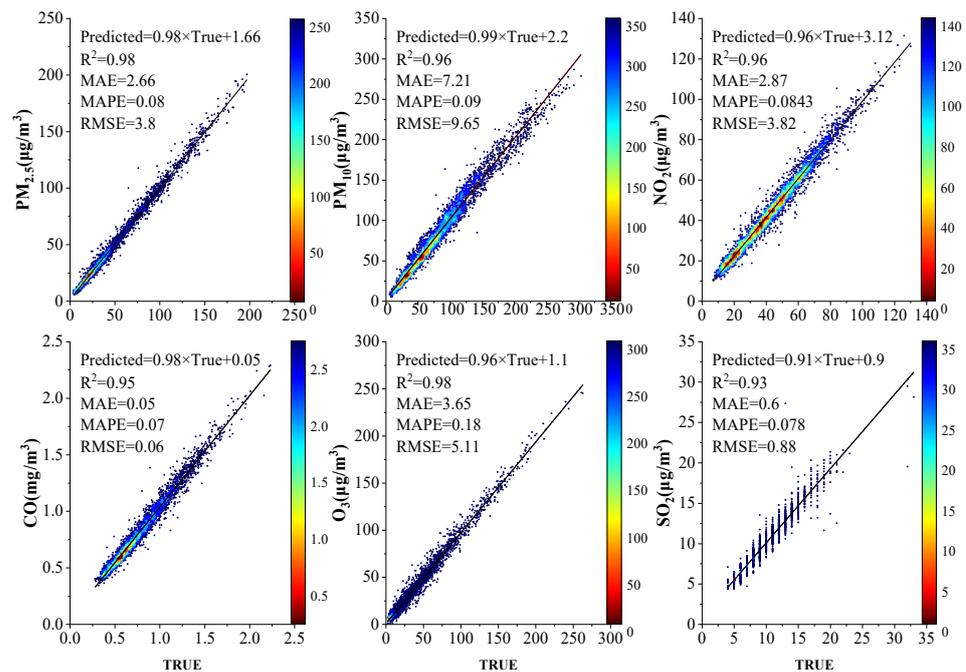


Figure 14. Comparison of actual and predicted values with meteorological and air pollutants added.

5. Discussion

In this study, an air pollutant concentration prediction model based on ODMSCNN-LSTM was developed and compared with three deep learning methods. The results show that the ODMSCNN-LSTM model performs better overall, is more suitable for daily or hourly concentration data sets and performs better on hourly data sets. This research shows that, compared with other machine learning, deep learning is a more effective method for processing big data (especially spatio-temporal data). Combining spatio-temporal data and models can improve the performance of spatio-temporal data prediction to a certain extent but adding meteorological factors does not necessarily improve the prediction performance of air pollutants. The reduction in the number of features may also affect the performance of the model. The prediction method proposed in this paper is feasible for the hourly concentration data prediction of multiple pollutants, and the method can also be applied into the air pollutant concentration prediction in multiple locations. In terms of input variables, regular monitoring data from the China National Environmental Monitoring Center and China Meteorological Administration are used. In terms of modeling methods, deep learning algorithms are combined with spatial and temporal features.

The performance of different air pollutants in the ODMSCNN-LSTM model may be due to differences in driving factors, temporal and spatial features, model types, model structures, and model development methods. At the same time, the amount of data, the degree of dispersion and spatial correlation between the concentration of air pollutants may also affect the prediction performance of the model, so it is necessary to further analyze the reasons for the difference. In addition, other factors, such as the increased output in equipment manufacturing, automobile manufacturing, electric power, heat production, gas and other industries, and the passenger and freight volume of the transportation industry need to be added in the follow-up research.

6. Conclusions

In this study, the daily concentration data of air pollutants in Xi'an from 2014 to 2020 and the hourly concentration data of air pollutants in Xi'an from 2019 to 2020 were used for prediction. A deep learning prediction model based on ODMSCNN-LSTM is established, and its performance is compared to those of MLP, CNN, and LSTM. The main research results are as follows: ODMSCNN-LSTM performs best in both daily concentration

prediction and hourly concentration prediction, and the overall performance of hourly concentration prediction is better than that of daily concentration prediction. From the perspective of multiple factors, the prediction effect under adding meteorological factors is not significantly improved compared with the prediction under all influencing factors all influencing factors, but air pollutants may affect the prediction results. In terms of grouped data, the prediction on the hourly concentration of air pollutants in 2020 is better than that of 2019, and the prediction on the hourly concentration of overall data from 2019 to 2020 performs better than that from 2019 and 2020. In general, predictions based on daily or hourly concentration are more suitable for the overall data, and the more data sets the better the model performance. In terms of overall performance, the prediction performance of the ODMSCNN-LSTM model is generally better than that of the MLP, CNN and LSTM models. Compared with other prediction methods, the prediction of air pollutant concentration based on the ODMSCNN-LSTM model has better performance in approximating the true concentration value. Especially at the inflection points and peaks and valleys of the concentration, it shows a better prediction effect.

Author Contributions: The article was written through the contributions of all authors. H.D.: conceptualization, methodology, modelling, analysis, writing original draft preparation. G.H.: conceptualization, modelling, writing reviewing and editing, revision. J.W.: conceptualization, editing, revision. H.Z.: modelling, analysis, writing original draft preparation, writing reviewing and editing. F.Z.: analysis writing reviewing and editing, revision. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is funded by the Research Foundation Ability Enhancement Project for Young and Middle-aged Teachers in Guangxi Universities(2019KY0864). The school-level scientific research fund projects (GXSZ2019YB023). The National Natural Science Foundation of China (71874134). Key Project of Basic Natural Science Research Plan of Shaanxi Province (2019JZ-30).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Availability of data and materials both air pollutant data and meteorological data come from public data provided by the Ministry of Ecology and Environment of the People's Republic of China (<http://www.mee.gov.cn/>, accessed on 25 October 2021) and China Meteorological Administration (<http://www.cma.gov.cn/>, accessed on 25 October 2021).

Acknowledgments: We sincerely appreciate the editor and the five anonymous reviewers for their valuable comments to help improve this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization. Ten Threats to Global Health in 2019. Available online: <https://www.who.int/emergencies/ten-threats-to-global-health-in-2019>. (accessed on 25 October 2021).
2. World Health Organization. Ambient (Outdoor) Air Quality and Health. Available online: [https://www.who.int/en/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/en/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health) (accessed on 25 October 2021).
3. Wang, J.; Zhao, B.; Wang, S.; Yang, F.; Xing, J.; Morawska, L.; Ding, A.; Kulmala, M.; Kerminen, V.-M.; Kujansuu, J.; et al. Particulate matter pollution over China and the effects of control policies. *Sci. Total. Environ.* **2017**, *584–585*, 426–447. [[CrossRef](#)]
4. Houweling, S.; Röckmann, T.; Aben, I.; Keppler, F.; Krol, M.; Meirink, J.F.; Dlugokencky, E.J.; Frankenberg, C. Atmospheric constraints on global emissions of methane from plants. *Geophys. Res. Lett.* **2006**, *33*, 33. [[CrossRef](#)]
5. Brauer, M.; Freedman, G.; Frostad, J.; Van Donkelaar, A.; Martin, R.; Dentener, F.; Van Dingenen, R.; Estep, K.; Amini, H.; Apte, J.; et al. Ambient Air Pollution Exposure Estimation for the Global Burden of Disease 2013. *Environ. Sci. Technol.* **2016**, *50*, 79–88. [[CrossRef](#)]
6. Li, T.; Yan, M.; Ma, W.; Jie, B.; Tao, L.; Hualiang, L.; Zhaorong, L. Short-term effects of multiple ozone metrics on daily mortality in a megacity of China. *Environ. Sci. Pollut. Res.* **2015**, *22*, 8738–8746. [[CrossRef](#)]
7. Devlin, R.B.; Duncan, K.E.; Jardim, M.; Schmitt, M.T.; Rappold, A.; Diaz-Sanchez, D. Controlled Exposure of Healthy Young Volunteers to Ozone Causes Cardiovascular Effects. *Circulation* **2012**, *126*, 104–111. [[CrossRef](#)]
8. Zanobetti, A.; Franklin, M.; Koutrakis, P.; Schwartz, J. Fine particulate air pollution and its components in association with cause-specific emergency admissions. *Environ. Health* **2009**, *8*, 58. [[CrossRef](#)]

9. Chen, C.; Zhao, B.; Weschler, C.J. Assessing the Influence of Indoor Exposure to “Outdoor Ozone” on the Relationship between Ozone and Short-term Mortality in U.S. Communities. *Environ. Health Perspect.* **2012**, *120*, 235. [[CrossRef](#)]
10. Kulmala, M. Atmospheric chemistry: China’s choking cocktail. *Nat. Cell Biol.* **2015**, *526*, 497–499. [[CrossRef](#)]
11. Feng, X.; Li, Q.; Zhu, Y.; Hou, J.; Jin, L.; Wang, J. Artificial neural networks forecasting of PM_{2.5} pollution using air mass trajectory based geographic model and wavelet transformation. *Atmos. Environ.* **2015**, *107*, 118–128. [[CrossRef](#)]
12. Gu, K.; Xia, Z.; Qiao, J. Stacked Selective Ensemble for PM_{2.5} Forecast. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 660–671. [[CrossRef](#)]
13. Zhou, W.; Wu, X.; Ding, S.; Cheng, Y. Predictive analysis of the air quality indicators in the Yangtze River Delta in China: An application of a novel seasonal grey model. *Sci. Total. Environ.* **2020**, *748*, 141428. [[CrossRef](#)]
14. Zhang, Z.; Wu, L.; Chen, Y. Forecasting PM_{2.5} and PM₁₀ concentrations using GMCN(1,N) model with the similar meteorological condition: Case of Shijiazhuang in China. *Ecol. Indic.* **2020**, *119*, 106871. [[CrossRef](#)]
15. Nouri, A.; Lak, M.G.; Valizadeh, M. Prediction of PM_{2.5} Concentrations Using Principal Component Analysis and Artificial Neural Network Techniques: A Case Study: Urmia, Iran. *Environ. Eng. Sci.* **2020**, *38*, 89–98. [[CrossRef](#)]
16. Zho, Y.B.; Fjc, A.; Hua, C.C. Exploring Copula-based Bayesian Model Averaging with multiple ANNs for PM_{2.5} ensemble forecasts. *J. Clean. Prod.* **2020**, *263*, 121528. [[CrossRef](#)]
17. Du, P.; Wang, J.; Hao, Y.; Niu, T.; Yang, W. A novel hybrid model based on multi-objective Harris hawks optimization algorithm for daily PM_{2.5} and PM₁₀ forecasting. *Appl. Soft Comput.* **2020**, *96*, 106620. [[CrossRef](#)]
18. Li, Y.; Liu, Z.; Liu, H. A novel ensemble reinforcement learning gated unit model for daily PM_{2.5} forecasting. *Air Qual. Atmos. Health* **2021**, *14*, 443–453. [[CrossRef](#)]
19. Guo, L.; Chen, B.; Zhang, H.; Zhang, Y. A new approach combining a simplified FLEXPART model and a Bayesian-RAT method for forecasting PM₁₀ and PM_{2.5}. *Environ. Sci. Pollut. Res.* **2019**, *27*, 2165–2183. [[CrossRef](#)]
20. Baker, K.R.; Foley, K.M. A nonlinear regression model estimating single source concentrations of primary and secondarily formed PM_{2.5}. *Atmos. Environ.* **2011**, *45*, 3758–3767. [[CrossRef](#)]
21. Kloog, I.; Koutrakis, P.; Coull, B.A.; Lee, H.J.; Schwartz, J. Assessing temporally and spatially resolved PM_{2.5} exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos. Environ.* **2011**, *45*, 6267–6275. [[CrossRef](#)]
22. Yazdi, M.D.; Kuang, Z.; Dimakopoulou, K.; Barratt, B.; Suel, E.; Amini, H.; Lyapustin, A.; Katsouyanni, K.; Schwartz, J. Predicting Fine Particulate Matter (PM_{2.5}) in the Greater London Area: An Ensemble Approach using Machine Learning Methods. *Remote Sens.* **2020**, *12*, 914. [[CrossRef](#)]
23. Bi, J.; Stowell, J.; Seto, E.Y.W.; English, P.B.; Al-Hamdan, M.Z.; Kinney, P.L.; Freedman, F.R.; Liu, Y. Contribution of low-cost sensor measurements to the prediction of PM_{2.5} levels: A case study in Imperial County, California, USA. *Environ. Res.* **2020**, *180*, 108810. [[CrossRef](#)] [[PubMed](#)]
24. Dhakal, S.; Gautam, Y.; Bhattarai, A. Exploring a deep LSTM neural network to forecast daily PM_{2.5} concentration using meteorological parameters in Kathmandu Valley, Nepal. *Air Qual. Atmos. Health* **2021**, *14*, 83–96. [[CrossRef](#)]
25. Park, Y.; Kwon, B.; Heo, J.; Hu, X.; Liu, Y.; Moon, T. Estimating PM_{2.5} concentration of the conterminous United States via interpretable convolutional neural networks. *Environ. Pollut.* **2020**, *256*, 113395. [[CrossRef](#)] [[PubMed](#)]
26. Guo, C.; Liu, G.; Chen, C.-H. Air Pollution Concentration Forecast Method Based on the Deep Ensemble Neural Network. *Wirel. Commun. Mob. Comput.* **2020**, *2020*, 8854649. [[CrossRef](#)]
27. Şahin, Ü.A.; Bayat, C.; Uçan, O.N. Application of cellular neural network (CNN) to the prediction of missing air pollutant data. *Atmos. Res.* **2011**, *101*, 314–326. [[CrossRef](#)]
28. Sayeed, A.; Choi, Y.; Eslami, E.; Lops, Y.; Roy, A.; Jung, J. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Netw.* **2020**, *121*, 396–408. [[CrossRef](#)]
29. Wang, J.; Li, J.; Wang, X.; Wang, J.; Huang, M. Air quality prediction using CT-LSTM. *Neural Comput. Appl.* **2021**, *33*, 4779–4792. [[CrossRef](#)]
30. Liu, D.-R.; Hsu, Y.-K.; Chen, H.-Y.; Jau, H.-J. Air pollution prediction based on factory-aware attentional LSTM neural network. *Computing* **2021**, *103*, 75–98. [[CrossRef](#)]
31. Pak, U.; Ma, J.; Ryu, U.; Ryom, K.; Juhyok, U.; Pak, K.; Pak, C. Deep learning-based PM_{2.5} prediction considering the spatiotemporal correlations: A case study of Beijing, China. *Sci. Total. Environ.* **2020**, *699*, 133561. [[CrossRef](#)]
32. Wen, C.; Liu, S.; Yao, X.; Peng, L.; Li, X.; Hu, Y.; Chi, T. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *Sci. Total. Environ.* **2019**, *654*, 1091–1099. [[CrossRef](#)] [[PubMed](#)]
33. Nourani, V.; Karimzadeh, H.; Baghanam, A.H. Forecasting CO pollutant concentration of Tabriz city air using artificial neural network and adaptive neuro-fuzzy inference system and its impact on sustainable development of urban. *Environ. Earth Sci.* **2021**, *80*, 136. [[CrossRef](#)]
34. Heydari, A.; Nezhad, M.M.; Garcia, D.A.; Keynia, F.; De Santoli, L. Air pollution forecasting application based on deep learning model and optimization algorithm. *Clean Technol. Environ. Policy* **2021**, 1–15. [[CrossRef](#)]
35. Chen, B.H.; Jin, Q.F.; Chai, H.L. Spatiotemporal distribution and correlation factors of PM_{2.5} concentrations in Zhejiang Province. *Acta Sci. Circumstantiae* **2021**, *41*, 817–829.
36. Zhang, Z.F.; Zheng, M.D.; Zhang, Y.; Zhou, J.; Liu, H.D. The Survey and Influence Factors of Air Pollution in Ningbo. *Environ. Monit. China* **2020**, *36*, 96–103.

37. Li, L.; Li, H.; Peng, L.; Li, Y.; Zhou, Y.; Chai, F.; Mo, Z.; Chen, Z.; Mao, J.; Wang, W. Characterization of precipitation in the background of atmospheric pollutants reduction in Guilin: Temporal variation and source apportionment. *J. Environ. Sci.* **2020**, *98*, 1–13. [[CrossRef](#)] [[PubMed](#)]
38. Boleti, E.; Hueglin, C.; Grange, S.K.; Prévôt, A.S.H.; Takahama, S. Temporal and spatial analysis of ozone concentrations in Europe based on timescale decomposition and a multi-clustering approach. *Atmos. Chem. Phys. Discuss.* **2020**, *20*, 9051–9066. [[CrossRef](#)]
39. Ji, M.; Jiang, Y.; Han, X.; Liu, L.; Xu, X.; Qiao, Z.; Sun, W. Spatiotemporal Relationships between Air Quality and Multiple Meteorological Parameters in 221 Chinese Cities. *Complex.* **2020**, *2020*, 6829142. [[CrossRef](#)]
40. Wang, Z.-B.; Li, J.-X.; Liang, L.-W. Spatio-temporal evolution of ozone pollution and its influencing factors in the Beijing-Tianjin-Hebei Urban Agglomeration. *Environ. Pollut.* **2020**, *256*, 113419. [[CrossRef](#)]
41. Wang, Y.; Song, J.; Yang, W.; Fang, K.; Duan, H. Seeking spatiotemporal patterns and driving mechanism of atmospheric pollutant emissions from road transportation in china. *Resour. Conserv. Recycl.* **2020**, *162*, 105032. [[CrossRef](#)]
42. Ronao, C.A.; Cho, S.B. Recognizing human activities from smartphone sensors using hierarchical continuous hidden Markov models. *Int. J. Distrib. Sens. Netw.* **2017**, *13*. [[CrossRef](#)]
43. Koschwitz, D.; Frisch, J.; van Treeck, C. Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and NARX Recurrent Neural Network: A comparative study on district scale. *Energy* **2018**, *165*, 134–142. [[CrossRef](#)]
44. Ministry of Ecology and Environment of the People’s Republic of China. 2017 Bulletin on the State of China’s Ecological Environment. Available online: <http://www.mee.gov.cn/hjzl/sthjzk/zghjzkgb/201805/P020180531534645032372.pdf> (accessed on 25 October 2021).
45. Ministry of Ecology and Environment of the People’s Republic of China. 2018 Bulletin on the State of China’s Ecological Environment. Available online: http://www.mee.gov.cn/ywtd/tpxw/201905/t20190529_704841.shtml/W020190529619750576186.pdf (accessed on 25 October 2021).
46. Xi’an Municipal Government. Notice of the General Office of the Xi’an Municipal Government on Issuing the Emergency Plan for Heavy Pollution Weather in Xi’an. Available online: <http://www.xa.gov.cn/gk/zcfg/szbf/5fb23324f8fd1c59664812a3.html> (accessed on 25 October 2021).
47. Hong, F. Research on Fault Location of Distribution Network Based on Matrix Method. Master’s Thesis, Guangdong University of Technology, Guangzhou, China, 2020.
48. Kong, Z.; Zhang, C.; Lv, H.; Xiong, F.; Fu, Z. Multimodal Feature Extraction and Fusion Deep Neural Networks for Short-Term Load Forecasting. *IEEE Access* **2020**, *8*, 185373–185383. [[CrossRef](#)]
49. Lu, H.; Azimi, M.; Iseley, T. Short-term load forecasting of urban gas using a hybrid model based on improved fruit fly optimization algorithm and support vector machine. *Energy Rep.* **2019**, *5*, 666–677. [[CrossRef](#)]