

## Research Article

# A Novel Method for Regional NO<sub>2</sub> Concentration Prediction Using Discrete Wavelet Transform and an LSTM Network

Bingchun Liu <sup>1</sup>, Lei Zhang,<sup>1</sup> Qingshan Wang <sup>2</sup>, and Jiali Chen<sup>1</sup>

<sup>1</sup>School of Management, Tianjin University of Technology, Tianjin 300384, China

<sup>2</sup>School of Humanities, Tianjin Agricultural University, Tianjin 300384, China

Correspondence should be addressed to Qingshan Wang; tigermountain@yeah.net

Received 13 December 2020; Revised 4 March 2021; Accepted 26 March 2021; Published 8 April 2021

Academic Editor: Paolo Gastaldo

Copyright © 2021 Bingchun Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Achieving accurate predictions of urban NO<sub>2</sub> concentration is essential for effectively control of air pollution. This paper selected the concentration of NO<sub>2</sub> in Tianjin as the research object, concentrating predicting model based on Discrete Wavelet Transform and Long- and Short-Term Memory network (DWT-LSTM) for predicting daily average NO<sub>2</sub> concentration. Five major atmospheric pollutants, key meteorological data, and historical data were selected as the input indexes, realizing the effective prediction of NO<sub>2</sub> concentration in the next day. Firstly, the input data were decomposed by Discrete Wavelet Transform to increase the data dimension. Furthermore, the LSTM network model was used to learn the features of the decomposed data. Ultimately, Support Vector Regression (SVR), Gated Regression Unit (GRU), and single LSTM model were selected as comparison models, and each performance was evaluated by the Mean Absolute Percentage Error (MAPE). The results show that the DWT-LSTM model constructed in this paper can improve the accuracy and generalization ability of data mining by decomposing the input data into multiple components. Compared with the other three methods, the model structure is more suitable for predicting NO<sub>2</sub> concentration in Tianjin.

## 1. Introduction

With the development of urbanization and industrialization, per capita energy consumption increases year by year. In addition to natural sources such as dust storms, bush fires, and volcanic eruptions, increased nitrogen dioxide (NO<sub>2</sub>) from vehicle exhaust and boiler exhaust has become one of the major environmental problems facing most countries in the world [1]. In China, the number of air quality standards and the average number of days with good air quality in 338 cities have increased year by year recently. However, in the Beijing-Tianjin-Hebei region and surrounding areas, the concentration of six air pollutants (PM<sub>2.5</sub>, PM<sub>10</sub>, O<sub>3</sub>, SO<sub>2</sub>, NO<sub>2</sub>, and CO) decreased the least, except ozone (O<sub>3</sub>). And, the same is true for Tianjin. In addition, the annual average concentration of NO<sub>2</sub> in Tianjin in 2018 was 47 micrograms per cubic meter, exceeding the national annual average concentration standard (40 micrograms per cubic meter). NO<sub>2</sub> has become the most important pollutant affecting the

air quality of Tianjin. All of these data above are obtained from the website <http://www.mee.gov.cn/hjzl/zghjzkqb/lnzghjzkqb/>.

In fact, nitrogen dioxide dissolves in water in the air to form acids, which may lead to the occurrence of acid rain and react with ultraviolet radiation to form photochemical smog. In addition, human exposure to NO<sub>2</sub> of different concentrations may lead to lung function damage of different degrees, seriously affecting industrial production and social activities [2]. Therefore, the effective detection and accurate prediction of NO<sub>2</sub> and the establishment of a highly accurate and stable prediction model can provide an early warning of air pollution emergencies and guide the release of NO<sub>2</sub> control measures and public health protection work.

The existing research on the prediction of atmospheric pollutant concentration can be roughly divided into three categories. The first category is the deterministic method based on the physical and chemical change model of the atmosphere [3–5]. The second category is the use of

computational methods based on regression and neural networks. The third category is the optimal combination model based on the second category. The deterministic approaches with no need for a large amount of historical data require a complete knowledge of pollution sources, timely emission quantities, main chemical reactions of the gaseous pollutants, and spatiotemporal physical transformation processes. The second and third types of computational methods usually require a large amount of historical measurement data under various meteorological conditions.

Wang et al. [6] used the Weather Research and Forecasting model coupled with Chemistry (WRF—Chem) for a serious pollution incident in Beijing in December 2016. The accompanying sensitivity analysis in this paper could capture the influence of emission sources on the concentration of target pollutants in different regions and different time periods, which provided a good reference for formulating effective emission reduction measures and regional air pollution prevention and control. However, due to the limitations of the resolution of existing models and other issues, the pollution events studied cannot be further utilized for other pollution events of different seasons and types. Baykara et al. [7] applied CMAQ (5.2), based on the heating activity data of local residents in Istanbul, to explore the influence of emissions from the residential heating sector on the level of environmental particulate matter. They thought that winter was the time when residential heating sector mainly affects regional air quality. A possible reason for this was the increase in coal burning that produces sulfur dioxide emissions but not for other man-made emissions.

The data-driven model mainly conducts a statistical analysis of air quality data and related factors to obtain scientific conclusions. In terms of linear correlation analysis, scholars have put forward a series of methods, such as geographic weighted regression (GWR), geographical and time-weighted regression (GTWR), and land use regression (LUR). Hinojosa-Baliño et al. [8] used meteorological, demographic, geographical, and social data and mixed geographic information system (GIS) and LUR to generate the prediction model and spatial distribution of PM<sub>2.5</sub> air pollution. Alahmadi et al. [9] used a local GWR model in the GIS environment to describe and quantify the contribution of transportation sector emissions to the NO<sub>2</sub> concentration in the Red Sea, but the limitation is the unavailability of some data. Further, Mirzaei et al. [10] used the GTWR model to study the spatial-temporal variability between PM<sub>2.5</sub> concentration at ground monitoring stations and satellite aerosol optical depth (AOD) data. However, in warm season, there were defects in the retrieval algorithm when detecting the low value of particulate matter, which lead to a decrease in the prediction accuracy of the model, and thus affected the simulation output. The application of machine learning algorithms, such as random forests, support vector machines (SVMs), and artificial neural networks (ANNs), takes the nonlinear relationship into consideration and improves the accuracy of prediction [11–13]. Masih [14] employed an integrated data mining tool that used random forests to predict the concentration of nitrogen dioxide in the atmosphere taking the emission inventory and meteorological

parameter monitoring data set as input prediction factors, and compared them with M5P and SVM, demonstrating the superiority of the model. Liu et al. [15] used support vector regression (SVR) to make a collaborative prediction of the Chinese urban air quality index (AQI). Experiments showed that when there was a strong interaction and correlation between air quality characteristic attributes and the air quality index, the MAPE (Mean Absolute Percentage Error) value of the multicity multidimensional regression model decreased. Cabaneros et al. [16] applied a mixed artificial neural network to the prediction of urban road NO<sub>2</sub>. Mishra and Goyal [17] developed an NO<sub>2</sub> concentration prediction model based on an artificial intelligence neuro-fuzzy model. However, these prediction models cannot capture both long-term and short-term characteristics, so Long Short-Term Memory (LSTM) is often used to predict air quality and pollutant concentration with time series characteristics. Li et al. [18] used LSTM layers to automatically extract inherent useful features of atmospheric pollutant data to predict the PM<sub>2.5</sub> concentration in Beijing in the next hour. Following this, Reddy et al. [19] extended the prediction from a single time step to the next 5 to 10 hours based on the time series data of pollution and meteorological information of Beijing.

However, the above model may lead to insufficient accuracy due to its potential convergence to local minima and overfitting [20]. In recent years, with the development of artificial intelligence and big data analysis, hybrid methods based on various information processing methods and deep learning methods have been widely used [21–26]. Kordestani and Samadi used distributed neural network and Bayesian algorithm to predict the remaining service life of Multifunctional Spoilers (MFS). Taking the data of LJ2000 series fighter data as samples, the hybrid prediction method was evaluated with relative accuracy, and it was found that the prediction effect of distributed structure was better [27]. Rezamand et al. constructed a hybrid prediction method based on real-time Supervisory Control and Data Acquisition (SCADA) and vibration signals to predict the Remaining Useful Life (RUL) of wind turbine bearings, made an empirical analysis of the hybrid model, and concluded that the prediction accuracy of this method was higher than that of the Bayesian algorithm [28]. Chen, Zhang, and Vachtsevanos proposed a prediction method of machine health condition based on Neural-Fuzzy Systems (NFSS) and Bayesian algorithm. Two examples of a cracked bearing plate and a faulty bearing were used to verify the effectiveness of the hybrid prediction method. The experimental results show that the hybrid method can effectively predict the running condition of the machine [29]. Bai et al. [30] proposed a neural network with long- and short-term memory (E-LSTM) to predict PM<sub>2.5</sub> concentration per hour and added mode decomposition (EMD) to the LSTM foundation, effectively improving the prediction accuracy. Zhao et al. [31] proposed a data-driven model called the LSTM-FC neural network, which uses historical air quality data, meteorological data, and weather forecast data to predict PM<sub>2.5</sub> pollution over 48 hours at a particular air quality monitoring station. Other researchers, such as Pak et al. [32], proposed that a mixed model convolutional

neural network (CNN) combined with long- and short-term memory (CNN-LSTM) has better seasonal stability and prediction performance compared with the single LSTM model. Wu and Lin [33] developed a hybrid model called VMD-SE-LSTM which applies VMD (Variational Mode Decomposition) technique to decompose the AQI data and employed SE (Sample Entropy) to recombine these components and train each recombinant subsequence with an LSTM neural network. This model not only improves the precision but also has good generalization ability.

Although the data decomposition and deep learning model have been combined in existing studies, the combination model of wavelet transform and deep learning method has not been applied to the prediction of air pollutant concentration. The novelties of this study is to take Tianjin as an example and build a DWT-LSTM combined model to predict the NO<sub>2</sub> concentration of the city in the future. In addition, through an empirical analysis of the NO<sub>2</sub> concentration in Tianjin, it is concluded that the prediction effect of the DWT-LSTM model constructed in this paper is better than that of the SVR, GRU, and single LSTM models. At the same time, the prediction results can provide early warning for air pollution emergencies in Tianjin and guide the introduction of NO<sub>2</sub> control measures and public health protection.

The above research is of certain use for the prediction of NO<sub>2</sub> concentration in the regional atmosphere. In this study, a data processing method based on discrete wavelet decomposition combined with an LSTM deep learning algorithm achieves the purpose of relatively accurate prediction of NO<sub>2</sub> concentration in Tianjin. The main contributions of this study are as follows: (1) the use of wavelet decomposition to carry out dimensional processing of data, optimize input variables, and improve the prediction accuracy of the LSTM model; (2) the development of the DWT-LSTM prediction model; (3) the consideration of the correlation between traditional LSTM prediction results and wavelet-LSTM results and actual data, verifying that data can improve the prediction accuracy and stability of the LSTM model through wavelet decomposition.

## 2. Research Area and Data

The geographical area of this study is Tianjin, China, located in the north China plain (117:10e39:10n, Figure 1). As of July 1, 2019, the Tianjin environmental air quality monitoring network has been established, covering the central urban area, the four districts around the city, the new Binhai area and other districts. Each point has six regular air pollutant monitoring capabilities, including PM10, PM2.5, SO<sub>2</sub>, NO<sub>2</sub>, CO, and O<sub>3</sub>. Since 2013, the Tianjin environmental air quality GIS platform has been used to release environmental air quality information for all monitoring points in Tianjin to the public. Now there are 16 national control stations and 11 municipal control stations with a total of 27 testing stations.

We collected the daily average data of PM2.5, PM10, NO<sub>2</sub>, SO<sub>2</sub>, O<sub>3</sub>, and CO on January 1, 2014, solstice to June 30, 2019 (2007 days) in Tianjin and used the latest available data

to correct the missing data of each type of air pollutant. In addition, we also downloaded meteorological observation data from the Chinese meteorological website platform established by the China meteorological administration (CMA), including wind speed, temperature, and weather conditions.

The output predictor of the experiment is the daily average concentration of NO<sub>2</sub> in Tianjin, which is shown in Figure 2. In Figure 2(e), it can be seen that the NO<sub>2</sub> concentration exhibits obvious periodicity, namely, high concentrations in the winter and summer concentration is low. Figures 2(a)–2(d) show the graph of the concentrations of PM2.5, PM10, SO<sub>2</sub>, and CO, respectively, over the same period. It can be seen that the four kinds of pollutant and NO<sub>2</sub> exhibit the same periodic variation. Figure 2(f) shows the O<sub>3</sub> concentration graph, where it can be seen that O<sub>3</sub> has periodic changes that are opposite to the other five pollutants. This may be because low-altitude O<sub>3</sub> is usually prone to produce and exceeds the standard in high temperature seasons [34]. The wind force was quantified according to the method of Bai et al. [35]. We graded the weather conditions according to how good or bad they were and quantified the weather indicators. The data (five pollutants, temperature, weather, wind data, and historical NO<sub>2</sub>) from January 2, 2014, to May 26, 2018, were used for training. The data from May 27, 2018, to June 30, 2019, were used for testing, also in conjunction with the NO<sub>2</sub> historical data.

Statistical descriptions of six pollutants are given in Table 1, where O<sub>3</sub> is the 8-hour average concentration.

## 3. Methodologies

**3.1. Long Short-Term Memory.** The LSTM neural network is a popular recursive neural network algorithm, which was first proposed by Hochreite and Schmidhuber to improve the memory of long (static) and short (cyclic) dynamic features of time series [36]. Similar to the traditional recurrent neural network model, this approach models time data by mining the circular connections between neurons and mining the internal connections between time series data. However, unlike traditional circular neural network models, it has a unique neuron structure called a “memory unit.” The hidden layer of an LSTM network constructed by this approach can store time information of any length to obtain a more accurate time series model.

The memory unit structure of the LSTM network is shown in Figure 3 [37]. The fixed length window of the time series is generated and input into the LSTM network. Multiple LSTMs can be superimposed to learn more complex patterns of sequential information [38]. The memory module consists of an input gate, forgetting gate, output gate, and a loop unit. Its core idea is to control the switch of each gate by a nonlinear function, to protect and control the state of the memory unit, so as to control the increase or decrease of information [39]. Therefore, the key of an LSTM network is to store data information through the state of the storage unit for a long time. In general, the output value of the three gates is 0~1, and the sigmoid function is used to

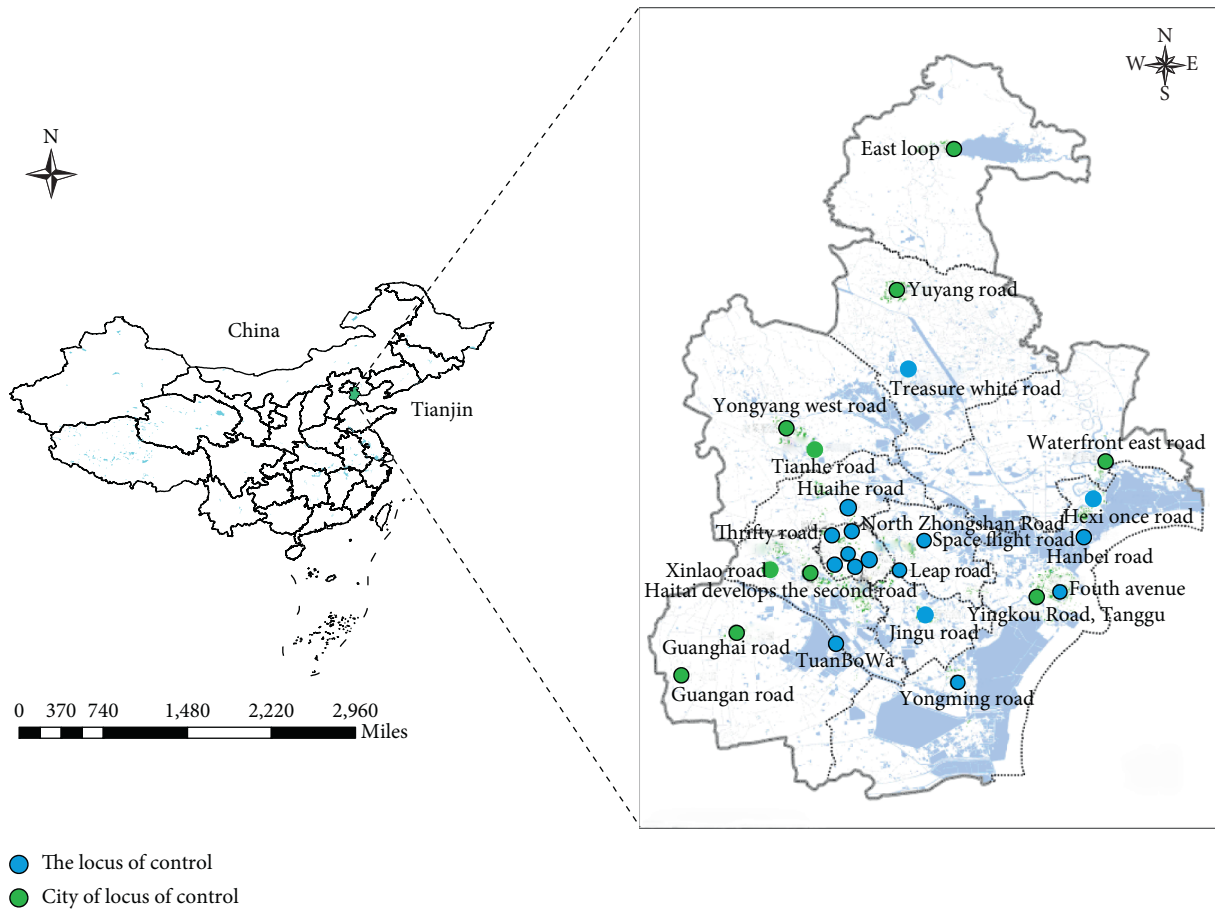


FIGURE 1: Location of the environmental monitoring stations.

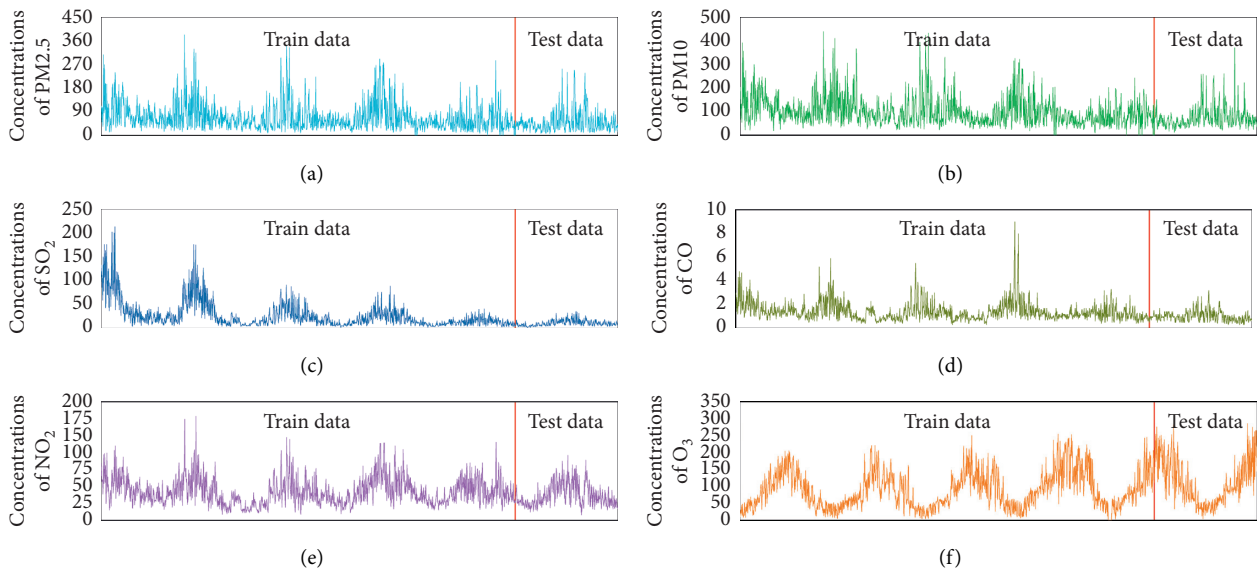


FIGURE 2: Daily mean concentration of 5 air pollutants and  $\text{NO}_2$  concentration from the previous day during the period 1/11/2014–30/6/2019.

TABLE 1: Statistical descriptions of the main indicators.

	PM <sub>2.5</sub>	PM <sub>10</sub>	SO <sub>2</sub>	CO	NO <sub>2</sub>	O <sub>3</sub>
Unit	μg/m <sup>3</sup>	μg/m <sup>3</sup>	μg/m <sup>3</sup>	mg/m <sup>3</sup>	μg/m <sup>3</sup>	μg/m <sup>3</sup>
Mean	66.49	106.18	23.87	1.33	47.39	93.07
Std.	49.71	65.21	25.53	0.74	21.98	56.67
Min	13	38	6	0.6	19	10
Max	383	483	23	9	176	286

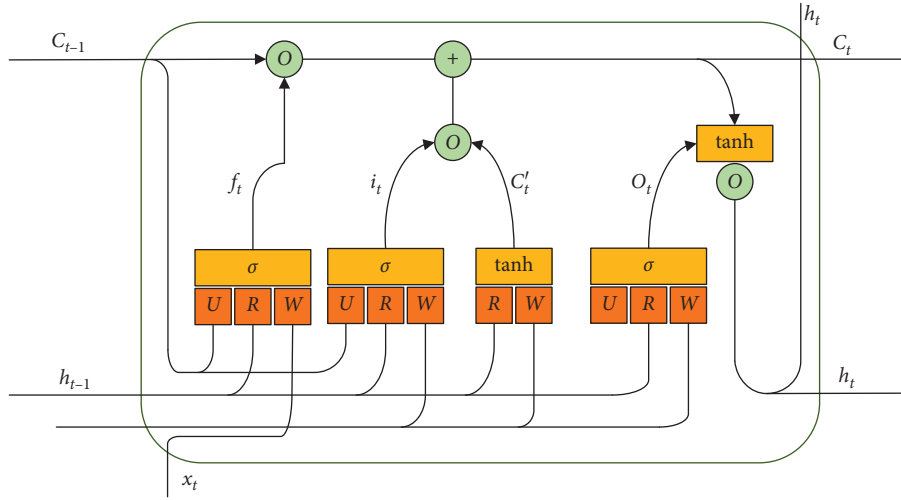


FIGURE 3: Structure of an LSTM neural network.

determine how much information can be input to the memory location. The main formula is as follows:

$$i_t = \sigma(W^i x_t + R^i h_{t-1} + U^i \circ c_{t-1} + b^i), \quad (1)$$

$$f_t = \sigma(W^f x_t + R^f h_{t-1} + U^f \circ c_{t-1} + b^f), \quad (2)$$

$$c'_t = \tanh(W^c x_t + R^c h_{t-1} + b^c), \quad (3)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ c'_t, \quad (4)$$

$$o_t = \sigma(W^o x_t + R^o h_{t-1} + U^o \circ c_t + b^o), \quad (5)$$

$$h_t = o_t \tanh(c_t), \quad (6)$$

where  $\circ$  represents the Hadamard product and  $\tanh$  is used as the activation function.  $x_t$ ,  $c_t$ , and  $h_t$  are the input, storage unit state, and output of the LSTM at time  $t$ , respectively, while  $i_t$ ,  $o_t$ , and  $f_t$  are the function values of the input gate, output gate, and forgetting gate, respectively.  $c'_t$  is the input modulate gate, which determines how much new information can be received.  $\sigma(\cdot)$  is the sigmoid function,  $R$ ,  $W$ , and  $U$  are weight matrixes,  $i$  is the input,  $f$  is the forget,  $c$  is the cell structure,  $o$  is the output.  $W_s$  represents the weight matrixes, and its superscript represents the two variables connected by the matrix. For example,  $W^i$  is the weight matrix between the input and the input gate and  $b$  is the deviation of the gate.

**3.2. Discrete Wavelet Transform.** Considering that the concentration of NO<sub>2</sub> in the atmosphere is related to the other five major pollutants, wind force, weather conditions, temperature difference, and other factors, the daily average NO<sub>2</sub> concentration series is nonstationary, volatile, and time-ordered. The above factors have different effects on NO<sub>2</sub> concentration. The input signal contains various frequency components: the contributions of low-frequency and high-frequency components to the dynamic characteristics of wind power data are different. If the components of these different frequencies can be learned by independent LSTMs, it will improve the performance of data mining. Therefore, the divide and conquer strategy requires that the original wind power data be decomposed into low-frequency and high-frequency signals through appropriate decomposition algorithms. In this paper, a discrete wavelet transform is used to decompose the original input data. It makes use of the time scale function to analyze the data and makes the wavelet transform have a multiscale resolution and time-shift characteristics. The scaling operation can observe signals of different scales. Therefore, a wavelet transform is very suitable for dealing with nonstationary time series including air pollutant data.

Assuming that  $x(t)$  squared can be integrated,  $x(t)$  can be expanded under the wavelet basis function. This operation is called the continuous wavelet transform of  $x(t)$ . The mathematical definition of the wavelet basis function is

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \Phi\left(\frac{t-b}{a}\right), \quad a \neq 0, b \in \mathbb{R}. \quad (7)$$

The mathematical definition of the  $x(t)$  continuous wavelet transform is given as

$$W_x(a, b) = \langle x(t), \Psi_{a,b}(t) \rangle = \frac{1}{\sqrt{|a|}} \int_{\mathbb{R}} x_t(t) \overline{\Phi\left(\frac{t-b}{a}\right)} dt. \quad (8)$$

In equations (7) and (8),  $\psi(t)$  can be considered as the parent wavelet function.  $a$  is the scale parameter and  $b$  is the time center parameter. When  $a$  and  $b$  change continuously, the whole transformation process is called a continuous wavelet transform. However, in practical applications, continuous transformation greatly increases the computational complexity, application cost, and implementation difficulty and is usually replaced by a small step discrete wavelet (DWT) [40].

A discrete wavelet transform makes the application of a wavelet transform easy to realize. The exponential discretization of parameters  $a$  and  $b$  reduces the computational complexity and avoids the information redundancy brought by a continuous wavelet transform. The discrete wavelet transform of  $x(t)$  is defined as

$$W_x(j, k) = \langle x(t), \Psi_{j,k}(t) \rangle = \frac{1}{\sqrt{|a|}} \int_{\mathbb{R}} x_t(t) \overline{\Phi\left(\frac{t-b}{a}\right)} dt. \quad (9)$$

In equation (9),  $a$  and  $b$  are discrete.  $a = a_0^j, b = ka_0^j b_0, a_0 > 1, b_0 > 0, j \in \mathbb{Z}, k \in \mathbb{Z}$ .

The discrete wavelet transform is the Mallat algorithm proposed in 1988 [41]. This is actually a signal decomposition method. For the multiresolution characteristics of wavelets, variable  $j$  is used to determine the resolution at different scales. Specifically, the main outline of the original signal is observed on a large scale, and the detailed information of the original signal is observed on a small scale. Finally, with gradually increasing  $j$  the results come out: one approximation value (i.e., low-frequency component) and  $n$  (which needs to be artificially set) detailed signals (i.e., high-frequency)  $d_n, d_{n-1}, d_{n-2}, \dots, d_1$  [42]. The original signal and two kinds of subsignals are satisfied by the following formula:

$$x(t) = a_n + d_n + d_{n-1} + \dots + d_1. \quad (10)$$

The schematic diagram of the discrete wavelet transform decomposition is shown in Figure 4.

**3.3. Overview of DWT-LSTM.** The LSTM neural network model is used to identify data patterns, and wavelet decomposition is used to decompose the input data. The prediction model combined with the wavelet transform and LSTM neural network consists of the following stages:

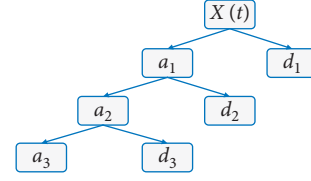


FIGURE 4: Three-layer wavelet decomposition.

Step 1: add the original data set  $\{M_1, M_2, \dots, M_n\}$  which is normalized, and the experimental data set  $\{D_1, D_2, \dots, D_n\}$ ;

Step 2: set  $\{D_1, D_2, \dots, D_n\}$  can be decomposed through  $m$  layers to obtain the high-dimensional input information set  $\{X'_1, X'_2, \dots, X'_t\}$ , where  $X'_i = (A_{mi-1}, D_{li-1}, \dots, D_{mi-1})$ ,  $i = 1, 2, \dots, t$ . The decomposition result at time  $t+1$  is  $X'_{t+1} = (A_{mt}, D_{lt}, \dots, D_{mt})$ ;

Step 3: use the new data set  $\{(X'_i, Y_i)\}_{i=1}^t$  to train the LSTM model. Through repeated data training and data testing, adjust parameters and get the optimal prediction model  $f(X_i)$ , as shown in Figure 5.

Step 4: the predicted value  $f(X'_{t+1})$  of the concentration of air pollutants in stage  $t+1$  can be measured by using the prediction model obtained above and according to the input vector  $X_{t+1}$  obtained in stage  $t+1$ .

Step 5: repeat steps 1–4 to obtain the predicted results  $f(X'_1), \dots, f(X'_{t+1})$ .

As the purpose of this study is to predict the Tianjin daily average concentration of  $\text{NO}_2$ , the input index includes two kinds of data, pollutant concentration, and meteorological factors. First, the meteorological factor information is quantified for further processing. The quantitative data are used for integration with other numeric data, but due to abnormal fluctuations, it will seriously affect the prediction ability. The consolidated data are normalized using Min-Max methods [43]:

$$d_{ij} = \frac{m_{ij} - \text{Min}(m_{i1}, m_{i2007})}{\text{Max}(m_{i1}, m_{i2007}) - \text{Min}(m_{i1}, m_{i2007})}, \quad (11)$$

$$i = 1, 2, \dots, 9; j = 1, 2, \dots, 2007.$$

The normalized data are used as the original signal for wavelet decomposition. As shown in Figure 6, a group of low-frequency subsignals and three groups of high-frequency subsignals of each original data were selected as input data and used for training and validation by LSTM neural networks. The best prediction model is obtained by adjusting the parameters and the structure of the design model.

**3.4. Model Parameters and Performance Indicator.** The prediction model proposed in this paper was implemented using *Python 2.7* in *Matlab 2017a* and the *Linux* system



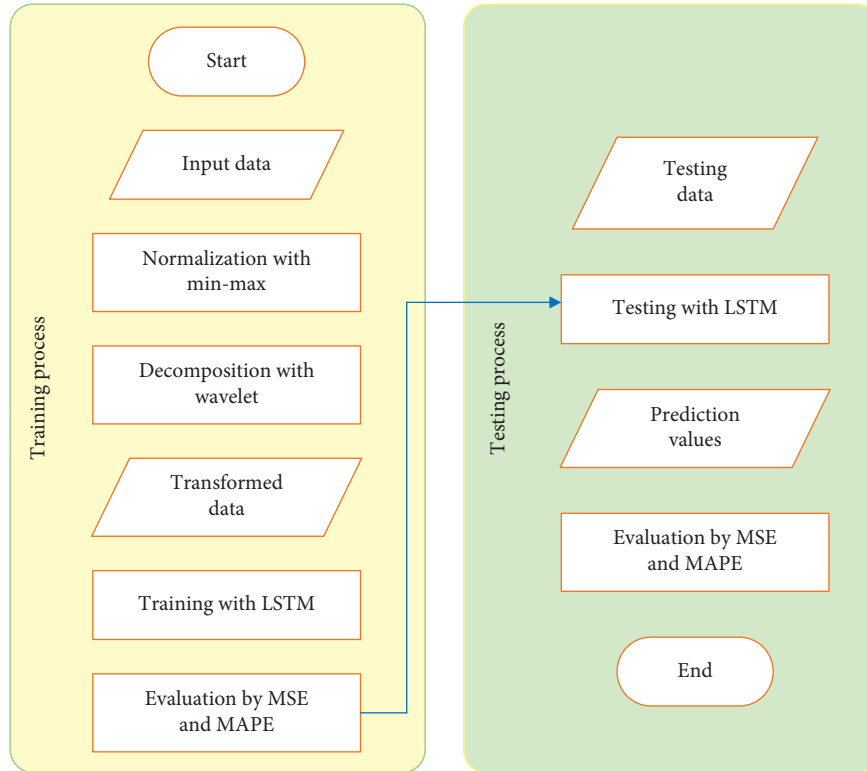


FIGURE 5: Prediction experiment process.

environment. The DWT-LSTM model adopts 3-layer wavelet decomposition and the Daubechies (DB) wavelet basis function. In network parameter setting, the primary parameters of LSTM include learning rate, max epochs, batch size, number of hidden layers, and time step. In the best model, learning rate is 0.0001, max epoch is taken as 500, number of hidden layers is 32, and time step is 3. The selection of the relevant parameters in the model targets mean absolute percentage error (MAPE) minimization [44], and this is an important indicator to measure prediction accuracy in the statistical field and is also widely used in the prediction of air pollutant concentrations. The MAPE index was used to measure the error of the prediction algorithm and compare it with other algorithms. Not only was the error between the predicted value and the true value considered but also the ratio between the error and the true value was considered [45]. The following equation gives the calculation of MAPE:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - y_i^*|}{y_i^*}, \quad (12)$$

where  $y_i^*$  is the observed  $\text{NO}_2$  concentration,  $y_i$  is the predicted  $\text{NO}_2$  concentration, and  $n$  is the number of detected samples.

## 4. Results and Discussion

**4.1. Data Description.** This paper selects six air pollutants of  $\text{PM}_{2.5}$ ,  $\text{PM}_{10}$ ,  $\text{NO}_2$ ,  $\text{SO}_2$ ,  $\text{O}_3$ , and  $\text{CO}$  and three meteorological observation factors of wind speed, maximum

temperature, and minimum temperature, as input indexes. In order to evaluate the accuracy of the  $\text{NO}_2$  concentration prediction model, the index data from January 1, 2014 to June 30, 2019, with a total of 2007 points, were selected in this paper. The original data sample was divided into two data sets: 80% of the original data (1606 data points) were used as the training sample, and the remaining 20% of the original data (401 data points) were used as the test sample to evaluate the prediction performance of the model.

**4.2. Results of the Wavelet Transform.** The LSTM method is suitable for time series prediction as it has good prediction performance. Also, the LSTM model can effectively represent the nonlinear relationship between the input vector and prediction target through the use of a kernel function. Appropriate high-dimensional input vectors can describe the information in features more effectively and accurately and express the meaning of the data. Therefore, the prediction performance depends largely on the choice of input vector in model design. In this study, when the LSTM model is used to predict pollutant concentration, in order to make the prediction results more accurate and stable, the structural transformation of the input variables can be determined to obtain a new set of input variables. By using wavelet decomposition, the data are promoted from one-dimensional data to high-dimensional data, which fully represents the trend of data change and improves the prediction accuracy. In this study, wavelet decomposition is based on the wavelet basis function of Daubechies (DB). Daubechies has low-pass and high-pass filtering

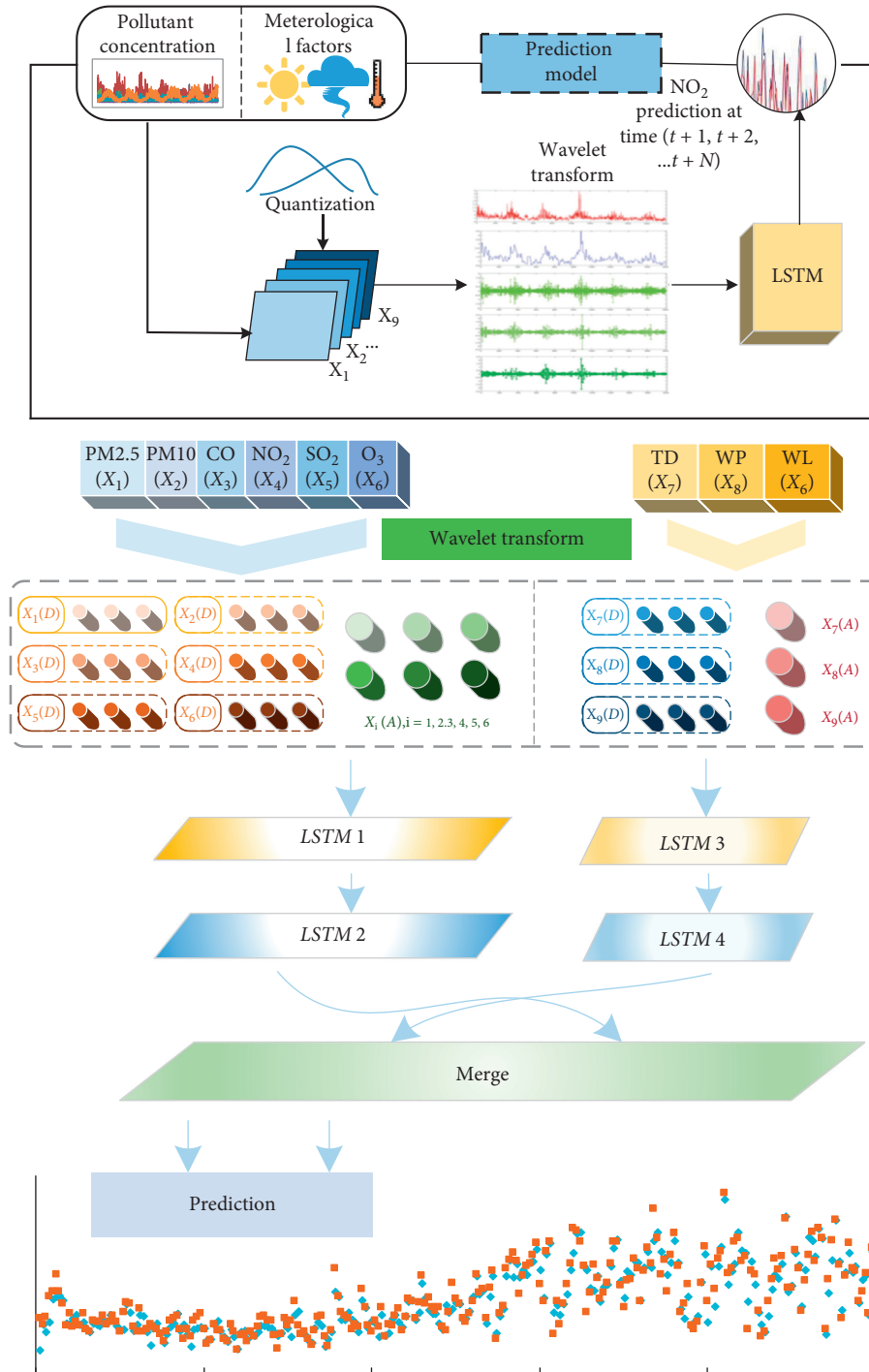


FIGURE 6: Architecture of the LSTM network with discrete wavelet transform (DWT-LSTM).

characteristics and is suitable for feature selection. Due to its inherent orthogonality, the Daubechies wavelet can be used widely and shows good performance in analyzing applied time series data.

Using the Matlab tool, low-frequency approximate information and high-frequency information obtained by wavelet decomposition transformation are taken as another new input vector group of the LSTM model to form a new prediction data set of the six kinds of air pollutants (PM10,

PM<sub>2.5</sub>, NO<sub>2</sub>, SO<sub>2</sub>, O<sub>3</sub>, and CO). The transformation results are provided in Figure 7, which shows the high-frequency information group and the low-frequency information group. The set of wavelet decomposition transforms the density time series data of the three input characteristic variables to generate high-dimensional input vectors, which effectively increases the amount of data representation information and significantly improves the prediction stability of the model.



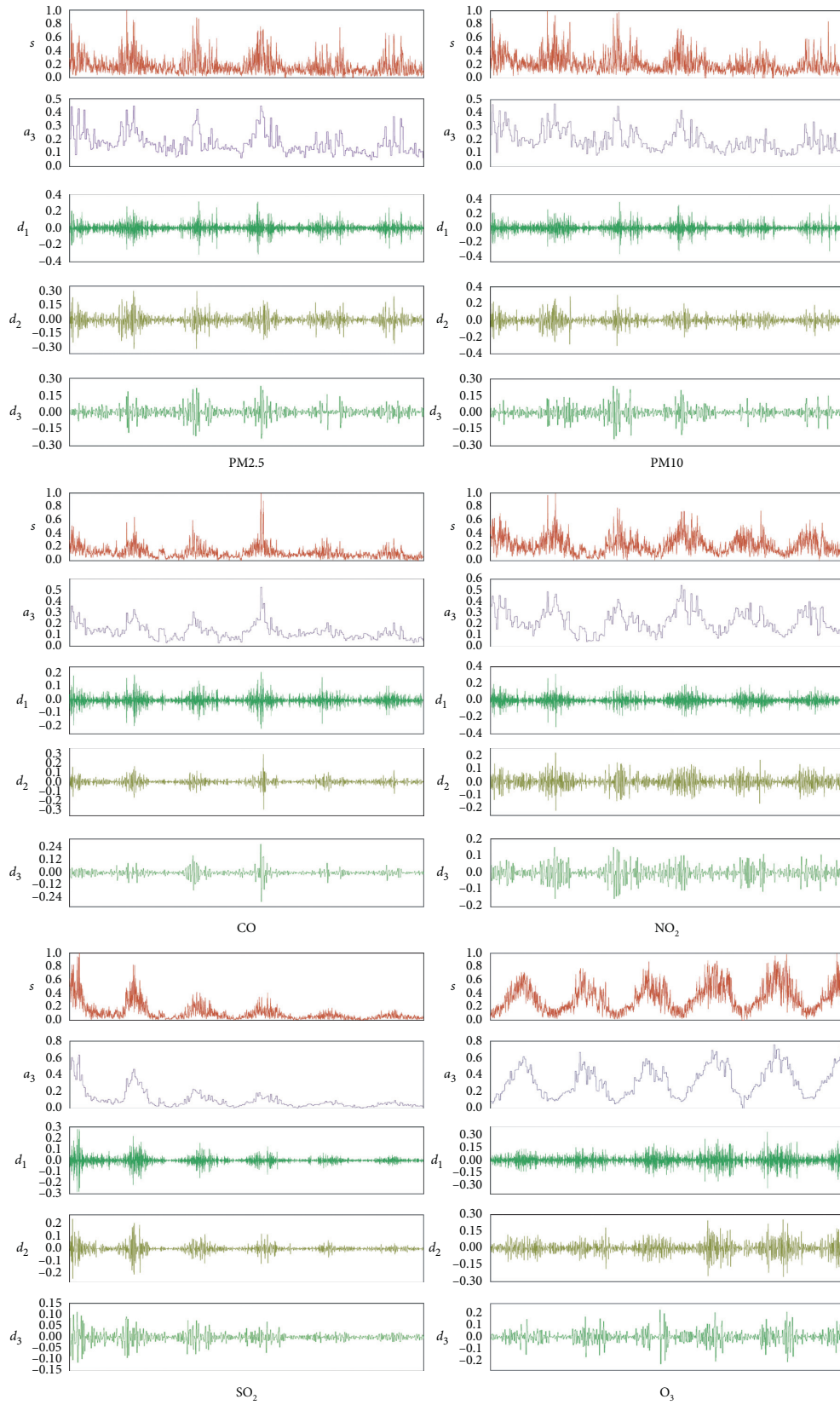


FIGURE 7: The result of the wavelet transform.

**4.3. Result of Prediction.** We compared the performance of the proposed DWT-LSTM model with that of SVR, GRU and the single LSTM model and trained and tested these models with the same training and test set applicable to the DWT-LSTM model. In order to evaluate the effectiveness of this method, we added two indexes: root mean square error (RMSE) and average absolute error (MAE). These indicators can be expressed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2}, \quad (13)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - y_i^*|.$$

Figure 8 intuitively shows the experimental results of the four prediction models. Through visual analysis, it can be seen that the prediction curve of the SVR model is relatively flat, it is difficult to accurately predict the fluctuation of data, and it presents a fluctuation trend opposite to the target value in some time periods. GRU, a variant of the LSTM, algorithmically combines forgetting and input gates into a single update gate, as well as a mixture of cellular and hidden states, and other changes. Although it has better performance in some experiments [46], its performance in this experiment is not as good as that of the single traditional LSTM model; especially it cannot predict outliers well. The LSTM model performs well regarding outliers (maxima and minima). For example, the prediction accuracy of a single LSTM model is better than that of the other three models on the two maxima of day 185 and day 206 and the two minima of day 155 and day 259. In order to objectively evaluate the performance of the four models, we calculated the predicted results according to the above formula, and the results are shown in Table 2. The evaluation results show that the performance of the DWT-LSTM model is better than that of the other three neural network models. Although the performance of the predicted outliers is not as good as that of the single LSTM model, the overall prediction accuracy is the highest.

The value of MAE and RMSE can explain the above phenomena, and the average absolute error and the average error of the LSTM model are greater than the DWT-LSTM model, which shows that the predictive value of the LSTM model is large, so it is also more likely to approximate the real value in the case of abnormal values. But the average absolute error and the average mean and mean error are relatively small, and the relatively small number of changes is far from the real value, which is higher than the other models and can more effectively predict the change of the concentration of  $\text{NO}_2$ , which can be more effective in the prediction of other areas or other pollutants and more effectively guide the prevention and control of air pollution.

**4.4. Analysis of Influencing Factors.** In order to explore the relationship between various factors and  $\text{NO}_2$  concentration, we changed the input indexes and conducted a series of experiments:

- (1) To investigate whether the weather conditions, wind power, and temperature difference are related to the concentration of  $\text{NO}_2$ , we eliminate the meteorological index and only used  $\text{NO}_2$  historical data and other 5 pollutants as input indicators for fitting. The results showed that the MAPE increased from 11.58% to 13.54%, proving that the meteorological index correlated with the concentration of air pollutants. Then, on the basis of the above experiments, we successively added weather condition indicators, temperature difference, and wind force in the experiments, and the MAPES were 13.63%, 13.61%, and 11.53%. It is proved that among the three meteorological factors, wind power has the largest effect on  $\text{NO}_2$  concentration; weather condition and temperature difference do not have much effect.
- (2) In order to explore the relationship between pollutant concentration and  $\text{NO}_2$  concentration, we firstly forecast the historical data of  $\text{NO}_2$  as a single input, and the MPAE is 16.61%. This proves that the historical data alone cannot effectively predict the future  $\text{NO}_2$  concentration. As a result, we will explore which major air pollutants have the greatest impact on  $\text{NO}_2$  in the future. On the basis of the historical data of  $\text{NO}_2$ , we successively add PM2.5, PM10,  $\text{SO}_2$ , CO, and  $\text{O}_3$  to carry out five groups of experiments, and the MAPE of the results is 14%, 14.96%, 15.39%, 12.68%, and 16.8%. It is shown that the effect of CO on  $\text{NO}_2$  concentration is the largest, followed by PM2.5 and PM10,  $\text{SO}_2$ , and  $\text{O}_3$  having little effect on  $\text{NO}_2$  concentration.

**4.5. Analysis of  $\text{NO}_2$  Concentration Change.** From 2014 to the first half of 2019, the  $\text{NO}_2$  concentration in Tianjin showed an overall downward trend. The seasonal periodic change rule remained unchanged, and the peak value in winter also showed a downward trend every year. This shows that Tianjin has paid more attention to the ecological environment during the 13th five-year plan period, and its specific work has achieved results.

- (1) Since 2011, Tianjin has been working in the development of new energy vehicles and has issued a series of related documents. By 2015, energy saving and new energy vehicles in the city increased to a total of more than 60000. In 2017 and 2018, this increased to 79000 vehicles. The promotion of new energy vehicles resulted in 36% of the total transport buses being new energy buses, bringing the total that has been put into operation to 3670. By the end of April 2019, the number of new energy vehicles in the city had reached 125,000. As a result, the coal consumption and oil consumption in Tianjin also tended to decline, and the concentration of  $\text{NO}_2$  in the air also decreases accordingly.
- (2) During this period, Tianjin carried out the optimization and upgrading of traditional industries and forced the closure of a series of enterprises with

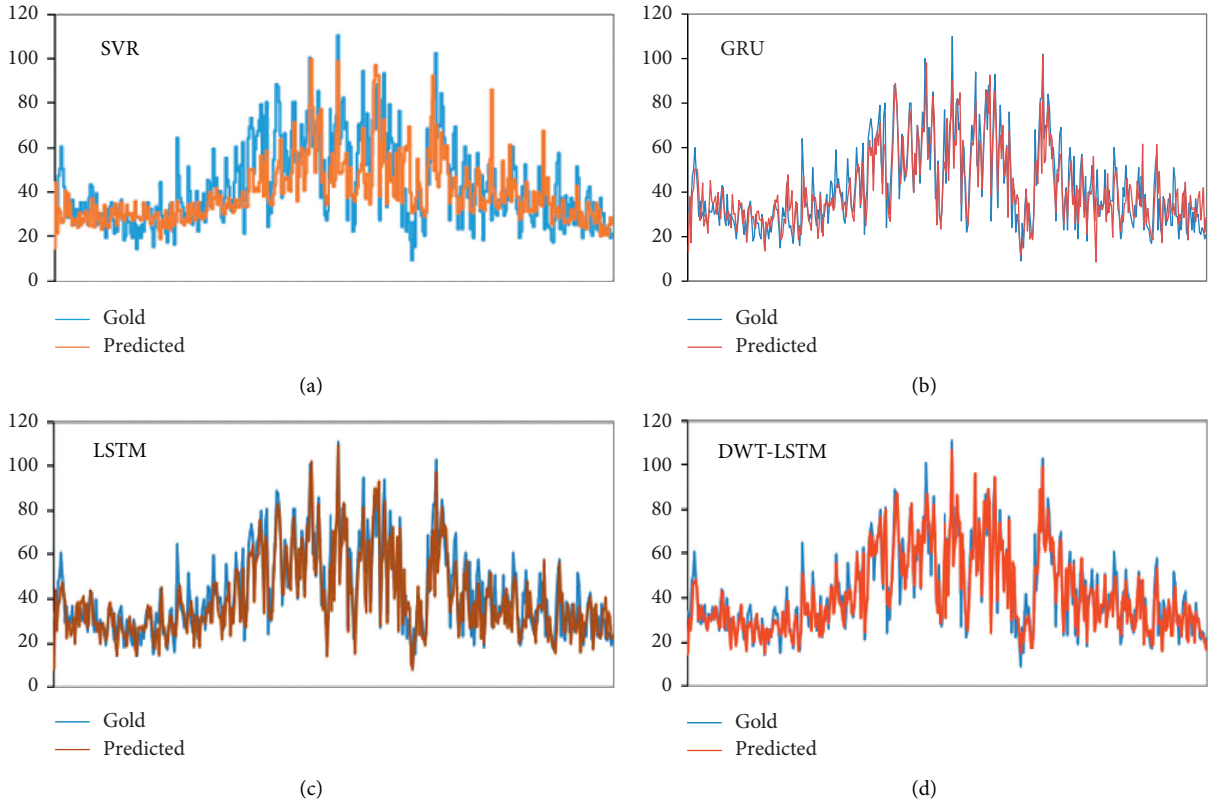


FIGURE 8: The forecasting results of different models.

TABLE 2: Comparison of forecasting performances using different models.

Method	MAPE (%)	MAE	RMSE
SVR	25.8688	10.1734	13.1987
GRU	18.3897	6.9120	8.8052
LSTM	17.8518	7.0439	9.0199
DWT-LSTM	11.5884	4.3377	5.9291

serious pollution emissions, especially internal combustion engine production enterprises and nonferrous and ferrous metal smelters. On the basis of the traditional manufacturing industry, the industrial structure has been adjusted, focusing on the development of high-end equipment, new-generation information technology, aerospace, new energy vehicles, new materials, biomedicine, new energy, energy conservation and environmental protection, modern petrochemical, modern metallurgy, and ten other industries, while paying attention to the development of the service industry. Therefore,  $NO_x$  emissions are reduced by reducing emission sources.

- (3) We can clearly find that the concentration of  $NO_2$  in spring and winter is higher than that in summer and autumn in Tianjin. This phenomenon may be caused by the low temperature in spring and winter, which is not conducive to the diffusion of  $NO_2$  and leads to an increase in concentration. For Tianjin, this may be related to the cold weather in winter and spring,

which requires a large amount of coal burning and some natural gas to heat the city. Fossil fuels such as coal produce a large amount of nitrogen dioxide, leading to an increase in the total  $NO_2$  content. In addition, although fireworks are banned in Tianjin, fireworks are still set off in rural areas due to insufficient supervision, leading to an increase in the concentration of  $NO_2$  in the city.

**4.6. Cause of Prediction Deviation.** In 2012, the ministry of environmental protection and the general administration of quality supervision, inspection, and quarantine jointly issued the environmental air quality standard (GB 3095-2012), which has been implemented nationwide since January 1, 2016. The implementation of the new standard is slightly different from that before in the way pollutants are counted, so the numerical performance is inconsistent within the statistical range. In addition, the accumulation of historical data is insufficient, the high concentration value of

heavy pollution days is less, and the limited training data brings great uncertainty to the prediction of high values.

On the other hand, DWT-LSTM is a data-based statistical model, which mainly relies on the empirical formula of meteorological parameters and historical monitoring data, and fails to consider the atmospheric chemical transformation, pollution source emission change, and regional transmission process. NO<sub>2</sub> reacts photochemically with O<sub>3</sub> and acts as a catalyst in the air to convert O<sub>2</sub> into O<sub>3</sub>. Most of the NO produced by human activities comes from the combustion of fossil fuels, such as automobiles, airplanes, internal combustion engines, and industrial kilns. It also comes from the process of producing and using nitric acid, such as nitrogenous fertilizer plant, organic intermediate plant, non-ferrous, and ferrous metal smelting plant. NO reacts with the oxygen in the air in the atmosphere, generating NO<sub>2</sub>. Therefore, NO<sub>2</sub> is associated with changes of industrial structure adjustment and pollution emissions. At the same time, external sources also contribute significantly to pollutant concentration during heavy pollution period.

## 5. Conclusions

In this study, a combined prediction model was established based on discrete wavelet decomposition and a neural network method to predict NO<sub>2</sub> concentration in Tianjin. The conclusions are as follows:

- (1) The combined prediction model uses wavelets to decompose the time series data of air pollutant concentration and takes the low-frequency and high-frequency data obtained after decomposition as input variables at the same time, thus increasing data dimensionality. Through the use of information representation of pollutant concentration time series data at different frequencies, the characteristics of the data can be better described.
- (2) The prediction model was built using an LSTM neural network, a high-dimensional nonlinear learning algorithm. When applied to the prediction of Tianjin NO<sub>2</sub> concentration, the performance was not as good as that of a single LSTM model, but the overall prediction accuracy was the highest. However, due to the low dimensionality of pollutant concentration time series data, the representation of the information is incomplete, affecting the ability of the prediction model to generalize.
- (3) The DWT-LSTM neural network method can be used to accurately predict the air pollutant concentration. Compared with the single LSTM model, the MAPE decreased from 17.85% to 11.58%; the MAE and RMSE increased to 4.3377 and 5.9291, respectively.
- (4) The practical significance of the use of a statistical prediction method for urban development and social activities is demonstrated by discussing the reasons underpinning prediction deviations and NO<sub>2</sub> concentration change.

In the process of data modeling, because of the limitation of the hardware used in the experimentation, the data cannot be fully analyzed, including the exploration of the structure of the neural network, leading to less complex model design. Research in the future will increase the dimension of data collection and carry out further experiments and explorations under the improved hardware environment.

## Data Availability

Raw data used to support the results of this study are included in the article.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Key Research and Development Program (2019YFB1803701) and the Research Program Fund for Humanities and Social Sciences of the Ministry of Education of China (20YJA630042).

## References

- [1] E. Rijnders, N. A. H. Janssen, P. H. N. Van Vliet, and B. Brunekreef, "Personal and outdoor nitrogen dioxide concentrations in relation to degree of urbanization and traffic density," *Environmental Health Perspectives*, vol. 109, no. 3, pp. 411–417, 2001.
- [2] J. S. Pandey, R. Kumar, and S. Devotta, "Health risks of NO<sub>2</sub>, SPM and SO<sub>2</sub> in Delhi (India)," *Atmospheric Environment*, vol. 39, no. 36, pp. 6868–6874, 2005.
- [3] T. Chai, G. R. Carmichael, Y. Tang et al., "Regional NO<sub>x</sub> emission inversion through a four-dimensional variational approach using SCIAMACHY tropospheric NO<sub>2</sub> column observations," *Atmospheric Environment*, vol. 43, no. 32, pp. 5046–5055, 2009.
- [4] K. M. Han, C. H. Song, H. J. Ahn et al., "Investigation of NO<sub>x</sub> emissions and NO<sub>x</sub>-related chemistry in East Asia using CMAQ-predicted and GOME-derived NO<sub>2</sub> columns," *Atmospheric Chemistry and Physics*, vol. 9, no. 3, pp. 1017–1036, 2009.
- [5] L. Wang, Y. Zhang, K. Wang, B. Zheng, Q. Zhang, and W. Wei, "Application of weather research and forecasting model with chemistry (WRF/Chem) over northern China: sensitivity study, comparative evaluation, and policy implications," *Atmospheric Environment*, vol. 124, pp. 337–350, 2016.
- [6] D. Wang, B. Jiang, F. Li, and W. Lin, "Investigation of the air pollution event in Beijing-Tianjin-Hebei region in December 2016 using WRF-chem," *Advances in Meteorology*, vol. 2018, Article ID 1634578, 12 pages, 2018.
- [7] M. Baykara, U. Im, and A. Unal, "Evaluation of impact of residential heating on air quality of megacity Istanbul by CMAQ," *Science of the Total Environment*, vol. 651, pp. 1688–1697, 2019.
- [8] I. Hinojosa-Baliño, O. Infante-Vázquez, and M. Vallejo, "Distribution of PM<sub>2.5</sub> air pollution in Mexico city: spatial analysis with land-use regression model," *Applied Sciences*, vol. 9, no. 14, p. 2936, 2019.

- [9] S. Alahmadi, K. Al-Ahmadi, and M. Almeshari, "Spatial variation in the association between NO<sub>2</sub> concentrations and shipping emissions in the Red Sea," *Science of the Total Environment*, vol. 676, pp. 131–143, 2019.
- [10] M. Mirzaei, J. Amanollahi, and C. G. Tzanis, "Evaluation of linear, nonlinear, and hybrid models for predicting PM 2.5 based on a GTWR model and MODIS AOD data," *Air Quality, Atmosphere & Health*, vol. 12, no. 10, pp. 1215–1224, 2019.
- [11] M. Zhao and X. Li, "An application of spatial decision tree for classification of air pollution index," in *Proceedings of the 2011 19th International Conference on Geoinformatics*, pp. 1–6, IEEE, Chicago, IL, USA, November 2011.
- [12] L. Gutiérrez, R. H. Mena, and M. Ruggiero, "A time dependent Bayesian nonparametric model for air quality analysis," *Computational Statistics & Data Analysis*, vol. 95, pp. 161–175, 2016.
- [13] M. Oprea, M. Popescu, and S. F. Mihalache, "A neural network based model for PM 2.5 air pollutant forecasting," in *Proceedings of the 2016 20th International Conference on System Theory, Control and Computing (ICSTCC)*, pp. 776–781, IEEE, Sinaia, Romania, October 2016.
- [14] A. Masih, "Application of random forest algorithm to predict the atmospheric concentration of NO<sub>2</sub>," in *Proceedings of the 2019 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT)*, pp. 252–255, IEEE, Yekaterinburg, Russia, April 2019.
- [15] B. C. Liu, A. Binaykia, P. C. Chang, M. K. Tiwari, and C. C. Tsao, "Urban air quality forecasting based on multi-dimensional collaborative support vector regression (SVR): a case study of Beijing-Tianjin-Shijiazhuang," *PLoS One*, vol. 12, no. 7, Article ID e0179763, 2017.
- [16] S. M. S. Cabaneros, J. K. S. Calautit, and B. R. Hughes, "Hybrid artificial neural network models for effective prediction and mitigation of urban roadside NO<sub>2</sub> pollution," *Energy Procedia*, vol. 142, pp. 3524–3530, 2017.
- [17] D. Mishra and P. Goyal, "Neuro-fuzzy approach to forecast NO<sub>2</sub> pollutants addressed to air quality dispersion model over Delhi, India," *Aerosol and Air Quality Research*, vol. 16, no. 1, pp. 166–174, 2016.
- [18] X. Li, L. Peng, X. Yao et al., "Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation," *Environmental Pollution*, vol. 231, pp. 997–1004, 2017.
- [19] V. Reddy, P. Yedavalli, S. Mohanty, and U. Nakhat, "Deep air: forecasting air pollution in Beijing, China," 2018, <http://arxiv.org/abs/1804.07891>.
- [20] W. Z. Lu and Y. Xue, "Prediction of particulate matter at street level using artificial neural networks coupling with chaotic particle swarm optimization algorithm," *Building and Environment*, vol. 78, pp. 111–117, 2014.
- [21] J. Ma, Y. Ding, J. C. P. Cheng, F. Jiang, and Z. Wan, "A temporal-spatial interpolation and extrapolation method based on geographic Long Short-Term Memory neural network for PM<sub>2.5</sub>," *Journal of Cleaner Production*, vol. 237, Article ID 117729, 2019.
- [22] X. Song, J. Huang, and D. Song, "Air quality prediction based on LSTM-kalman model," in *Proceedings of the 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, pp. 695–699, IEEE, Chongqing, China, May 2019.
- [23] W. Yuan, K. Wang, X. Bo, L. Tang, and J. Wu, "A novel multi-factor & multi-scale method for PM<sub>2.5</sub> concentration forecasting," *Environmental Pollution*, vol. 255, Article ID 113187, 2019.
- [24] C.-J. Huang and P.-H. Kuo, "A deep CNN-LSTM model for particulate matter (PM<sub>2.5</sub>) forecasting in smart cities," *Sensors*, vol. 18, no. 7, p. 2220, 2018.
- [25] M. Rezamand, M. Kordestani, R. Carriveau, D. S. K. Ting, and M. Saif, "An integrated feature-based failure prognosis method for wind turbine bearings," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 3, pp. 1468–1478, 2020.
- [26] D. Zhang, A. D. Bailey, and D. Djurdjanovic, "Bayesian identification of hidden markov models and their use for condition-based monitoring," *IEEE Transactions on Reliability*, vol. 65, no. 3, 2016.
- [27] M. Kordestani, M. F. Samadi, and M. Saif, "A new hybrid fault prognosis method for MFS systems based on distributed neural networks and recursive bayesian algorithm," *IEEE Systems Journal*, vol. 14, no. 4, pp. 5407–5416, 2020.
- [28] M. Rezamand, M. Kordestani, M. Orchard, R. Carriveau, D. S. K. Ting, and M. Saif, "Improved remaining useful life estimation of wind turbine drivetrain bearings under varying operating conditions (VOC)," *IEEE Transactions on Industrial Informatic*, vol. 17, no. 3, pp. 1742–1752, 2020.
- [29] C. Chen, B. Zhang, and G. Vachtsevanos, "Prediction of machine health condition using neuro-fuzzy and Bayesian algorithms," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 2, pp. 297–306, 2012.
- [30] Y. Bai, B. Zeng, C. Li, and J. Zhang, "An ensemble long short-term memory neural network for hourly PM<sub>2.5</sub> concentration forecasting," *Chemosphere*, vol. 222, pp. 286–294, 2019.
- [31] J. Zhao, F. Deng, Y. Cai, and J. Chen, "Long short-term memory—fully connected (LSTM-FC) neural network for PM<sub>2.5</sub> concentration prediction," *Chemosphere*, vol. 220, pp. 486–492, 2019.
- [32] U. Pak, C. Kim, U. Ryu, K. Sok, and S. Pak, "A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction," *Air Quality, Atmosphere & Health*, vol. 11, no. 8, pp. 883–895, 2018.
- [33] Q. Wu and H. Lin, "A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors," *Science of the Total Environment*, vol. 683, pp. 808–821, 2019.
- [34] K. Xiao, Y. Wang, G. Wu, B. Fu, and Y. Zhu, "Spatiotemporal characteristics of air pollutants (PM<sub>10</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, and CO) in the inland basin city of Chengdu, southwest China," *Atmosphere*, vol. 9, no. 2, p. 74, 2018.
- [35] Y. Bai, Y. Li, B. Zeng, C. Li, and J. Zhang, "Hourly PM<sub>2.5</sub> concentration forecast using stacked autoencoder model with emphasis on seasonality," *Journal of Cleaner Production*, vol. 224, pp. 739–750, 2019.
- [36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [37] F. A. Gers and J. Schmidhuber, "Recurrent nets that time and count," in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000*, vol. 3, pp. 189–194, Neural Computing: New Challenges and Perspectives for the New Millennium, IEEE, Como, Italy, July 2000.
- [38] N. C. Petersen, F. Rodrigues, and F. C. Pereira, "Multi-output bus travel time prediction with convolutional LSTM neural network," *Expert Systems with Applications*, vol. 120, pp. 426–435, 2019.
- [39] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: a hybrid model integrating LSTM with multiple

- GARCH-type models,” *Expert Systems with Applications*, vol. 103, pp. 25–37, 2018.
- [40] H. Liu, X.-w. Mi, and Y.-f. Li, “Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network,” *Energy Conversion and Management*, vol. 156, pp. 498–514, 2018.
- [41] I. Daubechies, “Orthonormal bases of compactly supported wavelets,” *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909–996, 1988.
- [42] H. Ming, D. Huang, L. Xie, J. Wu, M. Dong, and H. Li, “Deep bidirectional LSTM modeling of timbre and prosody for emotional voice conversion,” *Interspeech*, vol. 9, no. 8, pp. 2453–2457, 2016.
- [43] P. Sugiartawan, R. Pulungan, and A. K. Sari, “Prediction by a hybrid of wavelet transform and long-short-term-memory neural network,” *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 2, pp. 326–332, 2017.
- [44] X. Li, L. Peng, Y. Hu, J. Shao, and T. Chi, “Deep learning architecture for air quality predictions,” *Environmental Science and Pollution Research*, vol. 23, no. 22, pp. 22408–22417, 2016.
- [45] Y. Xu, W. Yang, and J. Wang, “Air quality early-warning system for cities in China,” *Atmospheric Environment*, vol. 148, pp. 239–257, 2017.
- [46] M. Yuan, Y. Wu, and L. Lin, “Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network,” in *Proceedings of the 2016 IEEE International Conference on Aircraft Utility Systems (AUS)*, pp. 135–140, IEEE, Beijing, China, October 2016.