Research Article

A Novel Hybrid Method to Predict PM$_{2.5}$ Concentration Based on the SWT-QPSO-LSTM Hybrid Model

Meng Du$^{1,2,3}$, Yixin Chen$^4$, Yang Liu$^4$, and Hang Yin$^5$

$^1$Department of Finance, Shandong Technology and Business University, Yantai, Shandong 264005, China
$^2$Collaborative Innovation Center for Financial Service Transformation and Upgrading, Yantai 264005, China
$^3$School of Land Science and Technology, China University of Geosciences (Beijing), Beijing 100083, China
$^4$School of Economics and Management, Dalian University, Dalian 116622, China
$^5$Dalian Central Sub-Branch, The People's Bank of China, Dalian 116001, China

Correspondence should be addressed to Meng Du; dumeng@dlu.edu.cn

Received 10 November 2021; Revised 17 June 2022; Accepted 4 July 2022; Published 16 August 2022

Academic Editor: Sheng Du

Copyright © 2022 Meng Du 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.

PM$_{2.5}$ concentration is an important indicator to measure air quality. Its value is affected by meteorological factors and air pollutants, so it has the characteristics of nonlinearity, irregularity, and uncertainty. To accurately predict PM$_{2.5}$ concentration, this paper proposes a hybrid prediction system based on the Synchrosqueezing Wavelet Transform (SWT) method, Quantum Particle Swarm Optimization (QPSO) algorithm, and Long Short-Term Memory (LSTM) model. First, the original data are denoised by the SWT method and taken as the input of the prediction model. Then, the main parameters of the LSTM model are optimized by global search based on the QPSO algorithm, which solves the problems of slow convergence and local extremum of traditional parameter training algorithms. Finally, the PM$_{2.5}$ daily concentration data of Chengdu, Shijiazhuang, Shenyang, and Wuhan are predicted by the proposed SWT-QPSO-LSTM model, and the prediction results are compared with those of single prediction models and hybrid prediction models. The experimental results show that the proposed model achieves higher prediction precision and lower prediction error than other models.

1. Introduction

The problem of air pollution is increasingly significant as industrialization and urbanization progress, and it has attracted attention worldwide. PM$_{2.5}$ (fine particles with aerodynamic diameter $\leq 2.5 \mu m$) mainly comes from the exhaust emission of urban traffic and the exhaust gas generated by various industrial activities. As the primary component of atmospheric pollution particles, PM$_{2.5}$ is the culprit of haze weather. The epidemiological studies at home and abroad show that long-time exposure to a high concentration of PM$_{2.5}$ will cause blood pressure rise, arrhythmia, and even stroke, myocardial infarction, atherosclerosis, and other diseases, seriously threatening human health [1]. Besides, PM$_{2.5}$ also has adverse effects on climate change. It leads to abnormal rainfall, exacerbates the greenhouse effect, affects traffic visibility, and improves the traffic accident rate, thus hindering the development of the transportation industry and regional economic development [2]. Therefore, gathering PM$_{2.5}$ concentration data in real time is critical for managing air pollution, preventing cardiovascular and cerebrovascular diseases, and improving regional climate, ecological balance, and regional economic development. Because of the serious negative impact of PM$_{2.5}$ on people's life, more and more individuals are paying attention to PM$_{2.5}$ concentration predictions. PM$_{2.5}$ concentration prediction is considered to be an important and effective method to alleviate the negative impact of PM$_{2.5}$.

Physical models and data-driven models are two types of PM$_{2.5}$ concentration prediction methodologies. The physical model simulates the diffusion, propagation, and elimination process of PM$_{2.5}$ through physical models. The advantage of this method is that it can illustrate the pollutant propagation and transformation process of PM$_{2.5}$, which allows for better
knowledge of PM$_{2.5}$ pollution [3]. The typical physical models include the multiscale air quality model CMAQ developed by the U.S. Environmental Protection Agency (USEPA), the mesoscale weather prediction model chemical module WRF-Chem developed by the National Center for Atmospheric Research (NCAR) and the National Oceanic and Atmospheric Administration (NOAA), and the China Unified Atmospheric Chemical Environment (CUACE/Haze-fog) developed by the China Meteorological Administration. Because the diffusion mechanism of PM$_{2.5}$ and other pollutants is very complex, the accuracy of the physical model is affected by many meteorological elements, such as temperature, humidity, wind direction, and season. Also, the physical model is often sensitive to the initial and boundary conditions and has high uncertainty. Besides, due to the time-consuming simulation process and the high calculation cost, the application of this model is greatly limited. In comparison, the data-driven model adopts statistical and machine learning methods to predict the PM$_{2.5}$ concentration by mining the historical information and related influencing factors of PM$_{2.5}$ concentration in meteorological observation stations. This model can obtain high-accuracy prediction results.

The data-driven air quality prediction models can be divided into three types: statistical prediction models, machine learning prediction models, and hybrid prediction models. The common statistical prediction models for PM$_{2.5}$ concentration are ARIMA Model, grey prediction model (G, M), Markov model, and ANFIS [4–7]. However, the statistical prediction models are relatively simple and the prediction accuracy is relatively low. The prediction models based on machine learning mainly include Artificial Neural Network (ANN), Support Vector Machine (SVM), Support Vector Regression (SVR), and Least Squares Support Vector Machine (LS-SVM) [8–11]. However, SVM, SVR, and LS-SVM have high requirements for parameter selection and cannot deal with the problem of big data. ANN needs a lot of data for training and it is prone to overfitting in the process of data training and prediction.

The lack of a single appropriate method encourages the growth of hybrid models. To increase the model’s prediction performance, a hybrid model integrates data preprocessing with various optimization techniques. The decomposition algorithm is commonly used in data preprocessing to reduce noise from the original data and extract the data’s effective features. Through the decomposition algorithm, the original data can be transformed into a form suitable for machine learning model training and prediction. The decomposition algorithms used in PM$_{2.5}$ concentration prediction mainly include empirical mode decomposition (EMD), variable mode decomposition (VMD), and wavelet decomposition [12–14]. EMD and VMD algorithms lack a mathematical basis, so they suffer from several problems such as wrong signals or pattern confusion. For the wavelet packet decomposition algorithm, there is no reliable and accurate expression of nonsmooth signals. An optimization algorithm is usually adopted to increase the accuracy of model prediction by optimizing the initial values and weights of machine learning models. Common optimization algorithms include the genetic algorithm, particle swarm optimization, and grey wolf algorithm [15]. Some studies combine the optimization algorithm with a machine learning prediction model to predict PM$_{2.5}$ concentration. For example, Niu et al. [16] combined comprehensive ensemble empirical mode decomposition (CEEMD), SVR, and Grey Wolf Optimizer (GWO) to predict PM$_{2.5}$ concentration. Sun and Sun [17] constructed a hybrid model that combines principal component analysis, LSSVM, and Cuckoo algorithm to predict PM$_{2.5}$ concentration. Cheng et al. [18] believed that the prediction accuracy of hybrid models such as Wavelet-ANN, Wavelet-ARIMA, and Wavelet-SVM for PM$_{2.5}$ is significantly higher than that of ARIMA, ANN, and SVM.

In recent years, the application of deep learning methods has achieved great success in numerous fields, such as image classification, natural language processing, and time series prediction. Some scholars have begun to use deep learning models to predict air quality, and the most successful prediction model is Long Short-Term Memory (LSTM). LSTM has the characteristics of long-term and short-term memory for time series data, so the prediction methods based on this model have good generalization ability and fault tolerance. Chang et al. [19] considered that the aggregated LSTM model achieves a better prediction effect for PM$_{2.5}$ concentration than SVR, GBTR (Gradient Boosted Tree Regression), LSTM, and other models. Dhakal et al. [20] used Kathmandu Valley as a case study and discovered that the LSTM model predicted PM$_{2.5}$ better than the SARIMA model. Some scholars construct hybrid models combining data preprocessing algorithms with LSTM models to predict PM$_{2.5}$ concentrations. Yan et al. [21] proposed a hybrid model combining LSTM with the stationary wavelet transform, which is superior to SVR, LSTM, and convolutional neural network combining Long Short-Term Memory (CNN-LSTM). Zhang et al. [22] combined VMD and bidirectional LSTM (BiLSTM) to predict PM$_{2.5}$ concentration, VMD algorithm is better than EMD algorithm in preprocessing, and, compared with SVR, MLP, and LSTM model, the prediction effect of BiLSTM has obvious advantages. Jin et al. [23] constructed a hybrid model that combines wavelet transform and nested LSTM to forecast PM$_{2.5}$ concentration; the results show that the performance of the proposed method outperforms those of other models such as Decision Tree, Random Forest, MLP, EMD-LSTM, and VMD-LSTM. However, the above-mentioned papers do not use evolutionary algorithms to optimize the hyperparameters of the LSTM model, which may affect the prediction results of the models.

In this paper, a new hybrid model is developed to predict PM$_{2.5}$ concentration in China using Synchrosqueezing Wavelet Transform (SWT), Long-Short Term Memory (LSTM), and Quantum Particle Swarm Optimization (QPSO). The contributions of this work are summarized as follows:

1. In data preprocessing, the SWT is used to decompose and reconstruct the data. As a classical time-frequency analysis method, the wavelet transform
method has a solid mathematical theory basis, and it can accurately express the time-frequency local properties of signals. SWT is developed on the basis of wavelet transform theory; compared with other wavelet transforms, SWT reinstatistics energy only in the frequency direction and preserves the temporal resolution of the signal, which can improve the time-frequency resolution of the whole process without changing the clarity of the instantaneous points; these features of the SWT method are suitable for reconstructing PM$_{2.5}$ data.

(2) QPSO is applied to optimize the LSTM model. The selection of LSTM parameters affects the prediction accuracy of PM$_{2.5}$. Based on the PSO algorithm and the principle of quantum mechanics, QPSO makes the motion state of particles follow random rules so that the optimization search range of particles can be expanded to the whole feasible solution space. Meanwhile, the global optimal solution can be found faster and more accurately. The experimental results indicate that the proposed hybrid prediction model may successfully improve the PM$_{2.5}$ concentration prediction effect.

2. Methodology

2.1. Principle of Synchrosqueezing Wavelet Transform. Traditional wavelet analysis is widely used in signal denoising because of its good time-frequency characteristics and multiresolution characteristics. SWT is a time-frequency domain redistribution method proposed by Daubechie et al. [24], which is a further development of wavelet transform (WT). Based on the continuous wavelet transform (CWT), this method adds and recombines wavelet coefficients with the same instantaneous frequency; that is, the wavelet coefficients near the central frequency are squeezed to improve the fuzzy phenomenon in the scale direction. Meanwhile, the time-frequency curve is thinned so that there is no cross term, which can significantly improve the time-frequency resolution and reconstruct the signal without distortion. Using the SWT method, the energy in the time-scale plane is redistributed according to the module size of each element, and the plane is transformed into a time-frequency plane through special mapping. This method achieves good results in various applications, such as climate analysis, mechanical fault diagnosis, signal denoising, civil engineering structures, harmonic and interharmonic detection, and seismic signal extraction.

SWT is based on continuous wavelet transform (CWT), and the phase of the signal in the frequency domain processed by wavelet transform is not affected by scale transformation. CWT is applied to a signal sequence $s(t)$, and the expression of wavelet coefficient can be obtained as follows:

$$W_s(a, b; \phi) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) \phi(t - b/a) dt,$$  

(1)

where $\phi(t - b/a)$ is the complex conjugate of the wavelet function; $a$ is the scale parameter (also known as the stretching parameter) and $a > 0$; $b \in R$ represents the translation parameter, and the wavelet basis function is

$$\phi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \phi\left(\frac{t - b}{a}\right).$$  

(2)

According to equations (1) and (2), wavelet transform can be regarded as the convolution of wavelet function and signal after telescopid translation transform, and the convolution coefficients are arranged in the time-scale plane.

In practice, the collected signal sequence usually contains various noise and human factors. When the coefficient $|W_s|$ obtained by wavelet transform is close to 0, the phase of $W_s$ will be relatively unstable. For a threshold $\gamma$, if $|W_s| \leq \gamma$, this part is filtered out. According to different noise levels of wavelet coefficients, the threshold of each layer of wavelet coefficients can be adaptively obtained by using the following equation:

$$\gamma = \sigma_n \sqrt{2 \log n},$$

\[ \sigma_n = \text{median} \left\{ |W_s(a_{k-1}, b)| - \text{median} \left| W_s(a_{k-1}, b) \right| \right\} / 0.6745, \]

(3)

where $n$ represents the number of sampling points of the original signal; $\sigma_n^2$ is the noise variance; and $W_s(a_{k-1}, b)$ is the wavelet coefficient on the length of the scale factor.

In fact, the obtained wavelet coefficients spectrum always diffuses in the scale direction, and the focusing performance is unsatisfactory. This makes the time-frequency diagram blurred. However, the phase of wavelet coefficients is not affected by scale transformation, so the instantaneous frequency $W_s(a, b)$ can be calculated according to wavelet coefficients $W_s(a, b; \phi)$ ($|W_s| > \gamma$):

$$W_s(a, b) = \frac{-i}{2\pi n W_s(a, b; \phi)} \frac{\partial W_s(a, b; \phi)}{\partial b}$$

(4)

Through the calculation of equation (4), the time-frequency domain redistribution is realized, and the time-scale plane is transformed into a time-frequency plane. This is the basic idea of compressed wavelet transform. Based on this, the value of the neighborhood interval $[\omega_l - 1/2\Delta \omega, \omega_l + 1/2\Delta \omega]$ of any frequency $\omega_l$ can be compressed to $\omega_l$. The formula of synchronous compression transformation can be expressed as

$$T_s(\omega_l, b) = (\Delta \omega)^{-1} \sum_{a_l : |a_l - b| / \omega_l \leq \Delta \omega / 2} W_s(a_l, b; \phi)a_l^{3/2} \Delta a_k,$$

(5)

where $a_k$ is the discrete scale and $k$ is the number of scales. When the signal is in the discrete state, the scale coordinate $\Delta a_k$ and frequency coordinates $\Delta \omega$ are discrete values:

$$\Delta a_k = a_k - a_{k-1},$$

$$\Delta \omega = \omega_l - \omega_{l-1}.$$  

(6)
SWT is a time-frequency rearrangement algorithm. Different from other rearrangement algorithms in the past, SWT can not only increase the time-frequency resolution ratio but also reconstruct the characteristic segment of the signals in the time-frequency diagram. Thus, SWT is reversible, and its inverse transform (ISWT) can be denoted as

\[ s(t) = 2\text{Re}\left[ C_{\psi}^{-1} \sum_{f} T_{\psi}(\omega, b) (\Delta \omega) \right]. \]

\[ C_{\psi} = \int_{0}^{\infty} \frac{\psi^{*}(\xi)}{\xi} d\xi. \]

Signal reconstruction can be realized by equation (7), where \( \psi^{*}(\xi) \) represents the conjugate Fourier transform of a basic wavelet function equation.

To sum up, the extraction of signal sequences based on SWT consists of the following three steps:

1. Use CWT to transform signal \( s(t) \) to obtain the wavelet coefficients \( W_{\psi}(a, b; \phi) \)
2. Obtain the instantaneous frequency from the wavelet coefficients \( |W_{\psi}| > \gamma \) and obtain the synchronous compression coefficient \( T_{\psi}(\omega, b) \) by synchronous compression
3. Use the effective signal for synchronous compression transformation and reconstruction to realize signal extraction

2.2. Principle of Long Short-Term Memory Neural Network. Recurrent neural network (RNN) is often used to handle the data with sequence changes. RNN is a network containing loops, which allows information persistence. Every neural network module of RNN transmits the message to the next module through replication. However, when the interval between the predicted position and relevant information magnifies to a certain extent, RNN will lose its learning ability because it cannot connect such far information. LSTM is a specific RNN proposed by Hochreiter and Schmidhuber [25], which can solve the problem of gradient disappearance or explosion in long sequence training. LSTM achieves better performances on training longer sequences than conventional RNN.

Unlike the standard RNN with only one transmission state \( (h^{t}) \) and a single tanh layer, LSTM has two transmission states \( (c^{t} \) and \( h^{t} \)), and its duplicating module includes four layers that interact in a special way. Besides, LSTM adds a “processor” named cell to the algorithm, which can estimate whether the input information is serviceable. There are three gates in a cell: input gate, forgetting gate, and output gate. These gates protect and control the state of the cells. When information enters the LSTM network, the cell determines whether this information is useful according to the rules. Meanwhile, the information consistent with the algorithm identification is left, and the inconsistent information is forgotten through the forgetting gate.

The processing process of LSTM is mainly divided into three steps. The first step discards the irregular information in the cell state through the forgetting gate according to the rules. The gate reads \( h_{t-1} \) and \( x_{t} \) and assigns each cell a value between 0 and 1 in the cell state \( C_{t-1} \), where 1 indicates “completely reserved” and 0 indicates “completely discarded.” At this time, the cell status output is

\[ f_{t} = \sigma(W_{f} \cdot [h_{t-1}, x_{t}] + b_{f}). \]

The second step determines the new information stored in the cell state according to the rules. The Sigmoid layer is called “input layer,” and it resolves the value that needs to be updated. The tanh layer creates a new candidate value vector, and \( \tilde{C}_{t} \) is involved in the conditions. The formula for input door is as follows:

\[ i_{t} = \sigma(W_{i} \cdot [h_{t-1}, x_{t}] + b_{i}), \]
\[ \tilde{C}_{t} = \tanh(W_{c} \cdot [h_{t-1}, x_{t}] + b_{c}). \]

The state of an old cell is updated so that \( C_{t-1} \), is updated to \( C_{t} \). The expression is as follows:

\[ C_{t} = f_{t} \cdot C_{t-1} + i_{t} \cdot \tilde{C}_{t}. \]

The third step decides the final output value based on the state of the cells. The output gate determines the information output by the cells, and its formula is as follows:

\[ o_{t} = \sigma(W_{o} \cdot [h_{t-1}, x_{t}] + b_{o}), \]
\[ h_{t} = o_{t} \cdot \tanh(C_{t}). \]

Here, \( W \) is the weight measurement of every door, and \( b \) is the deviation.

2.3. Principle of Quantum Particle Swarm Optimization. QPSO is an improved PSO algorithm proposed by Sun et al. [26] to handle the problems of premature convergence and local extremum of standard PSO algorithm.

In the PSO algorithm, particles have two attributes: location \( (X) \) and velocity \( (V) \). In a \( D \)-dimensional solution space, the state of particle \( i \) is expressed as \( X_{i} = (x_{i1}, x_{i2}, \ldots, x_{iD}) \) and \( V_{i} = (v_{i1}, v_{i2}, \ldots, v_{iD}) \). The fitness function is calculated to determine the fitness to express the advantages and disadvantages of the particle. Each time the particle is updated, its fitness needs to be recalculated. The particle with the best fitness is selected to update the optimal positions of individuals \( P_{i} = (p_{i1,1}, p_{i1,2}, \ldots, p_{i1,D}) \) and the group optimal position \( P_{g} = (p_{g1}, p_{g2}, \ldots, p_{gD}) \).

The update of the speed and position of particles at a certain time depends on the state of particles at the previous time. Thus, the speed and position of particles at any time are not random, and the particle optimization search cannot cover the whole solution space and falls into the local extremum easily. That is, the PSO algorithm is not a global convergence algorithm, and the final optimal solution obtained by this algorithm is the local optimal solution instead of the global optimal solution. With the expansion of applications, the PSO algorithm suffers from some problems, such as early convergence and dimension explosion.

Following the principle of quantum mechanics, QPSO algorithm uses the quantum behavior of particles for
optimization. Specifically, the particles follow the random rule of quantum motion. The motion state of the particles at present is no longer affected by that of the previous moment. Thus, the particle search scope can be extended to the whole feasible solution space, and the global optimum solution can be obtained faster and better. QPSO algorithm improves the iterative behavior and optimization strategy of particle swarm optimization, and it has the advantages of fewer parameters, high efficiency, and global optimization.

When a particle following the quantum motion rules moves, its velocity and position are uncertain. In this case, the motion state of the particle is expressed by the probability density function (PDF) of the particle at a certain point in the solution space. In our study, PDF is replaced by the square of the wave function, and the exact position of the particle is simulated by the Monte Carlo method. The update formula of particle position is

\[
\begin{align*}
    x_i^{t+1} &= x_i^t + \beta \left[ m_{\text{best}} - x_i^t \right] * \ln \left( \frac{1}{\mu} \right), \quad \mu > 0.5 \\
    &\quad - \left[ m_{\text{best}} - x_i^t \right] * \ln \left( \frac{1}{\mu} \right), \quad \mu \leq 0.5
\end{align*}
\]  

(13)

where \( P_i^t = (P_{i1}^t, P_{i2}^t, \ldots, P_{iN}^t) \) is the random point where the particle moves in the feasible space; \( \beta \) is the contraction expansion control coefficient; \( \mu \) is a stochastic number within the range of \([0, 1] \); \( m_{\text{best}} \) represents the average value of the historical optimal position of particles, and its computing formula is as follows:

\[
m_{\text{best}} = \frac{1}{N} \sum_{i=1}^{N} P_{\text{best},i}^t = \left( \frac{1}{N} \sum_{i=1}^{N} P_{\text{best},1}^t, \frac{1}{N} \sum_{i=1}^{N} P_{\text{best},2}^t, \ldots, \frac{1}{N} \sum_{i=1}^{N} P_{\text{best},D}^t \right),
\]

(14)

where \( N \) is the total number of particles in the population; \( D \) is the particles’ dimension; \( P_{\text{best},i}^t \) is the individual optimal position of the \( i \)-th particle in the current number of iterations. The optimal placement of individual particles interacts with the optimal location of the global particle swarm during particle motion, and \( P_i \) is constantly updated. The iterative equation of the \( i \)-th particle is

\[
P_i^t = \varphi P_{\text{best},i}^t + (1 - \varphi) P_{\text{gbest},i}^t, \quad (16)
\]

where \( P_{\text{gbest}}^t \) is the historical optimal location of the group and \( \varphi \) is a random number with \([0, 1] \). The procedure of the QPSO algorithm is illustrated in Algorithm 1.

3. The Hybrid Model of PM$_{2.5}$

Concentration Prediction

3.1. Structure of the Prediction Model. The PM$_{2.5}$ concentration prediction model studied in this paper mainly consists of the four following components:

(a) Data Selection. PM$_{2.5}$ is the main component of smog particles. Exposed to high concentrations of PM$_{2.5}$ for a long time, people will suffer from an increased risk of respiratory, cardiovascular, and cerebrovascular diseases and even cancer. Therefore, the PM$_{2.5}$ daily concentration data of a certain area is taken as the experimental data in this paper.

(b) Data Preprocessing. There are some problems in the collected data, such as missing values, anomalies, and inconsistent dimensions. Therefore, missing value processing and standardization need to be performed on the experimental data before simulation analysis. In the experiment, the SWT method is used for data preprocessing.

(c) Neural Network. The traditional neural network falls into the local extremum easily. Also, its convergence speed is slow, and the prediction precision needs to be improved. This paper uses QPSO to optimize the parameters of the LSTM neural network to construct a PM$_{2.5}$ concentration prediction model.

(d) Analysis of Prediction Results. In this paper, the prediction precision of the simulation experiment is analyzed by calculating the values of RMSE, MAE, and MAPE. The structure of the prediction model is illustrated in Figure 1.

3.2. SWT-QPSO-LSTM Hybrid Prediction Model. The key parameters in the SWT-QPSO-LSTM hybrid prediction model include the number of neurons in LSTM \((L1, L2)\), learning rate \((\epsilon)\), and training iterations \((k)\). The four key parameters are taken as the characteristics of particle optimization, and the LSTM model is adjusted and optimized by the QPSO algorithm. The overall process of the algorithm design and optimization is as follows:

(a) Standardizing the PM$_{2.5}$ historical concentration data processed by the SWT method. Because the LSTM neural network model is sensitive to the scale of input data, a large data scale will affect the training effect of the model. The daily PM$_{2.5}$ concentration data of 36 months before the sample data is taken as the training set, and the PM$_{2.5}$ concentration data of 12 months after the sample data is taken as the testing set. The data can be standardized by the following equation:

\[
y = (x - x_{\text{mean}}) * \left( \frac{y_{\text{std}}}{x_{\text{std}}} \right) + y_{\text{mean}},
\]

(16)

where the default value of \( y_{\text{mean}} \) is 0 and the default value of \( y_{\text{std}} \) is 1; \( x \) is the original PM$_{2.5}$ concentration data; \( x_{\text{mean}} \) is the mean value of the original data; \( x_{\text{std}} \) is the standard deviation of the original data, and \( y \) is the standardized PM$_{2.5}$ concentration data with a mean value of 0 and standard deviation of 1.

(b) Initializing the particle swarm parameters, including the population number, the learning rate, the maximum number of iterations, and the value range of particle position and velocity.

(c) According to experience, this paper selects the initial values of \( L1, L2, \epsilon, \) and \( k \) to establish the LSTM
model. Then, the model is trained on the training set, and the results are compared with those obtained on the testing set. Finally, a set of superparameters are obtained by searching to minimize the prediction error of LSTM. This paper takes the mean square error (MSE) as the fitness function, which is calculated as follows:

\[
MSE = \frac{1}{M} \sum_{m=1}^{M} (y_m - \bar{y}_m)^2. \tag{17}
\]

In the previous formula, \(y_m\) and \(\bar{y}_m\) represent the real value and predicted value of the sample.

(d) The global optimal position gbest and local optimal position pbest are determined by the initial fitness value of the particles, and they are regarded as the historical optimal positions. The speeds and positions of the particles are updated. Then, the corresponding particle fitness value is calculated and compared with the local and global optimal solution to improve the accuracy.

(e) Judging whether the termination condition of the algorithm is satisfied (the fitness value of the particles tends to be stable or the maximum number of iterations is reached). If the termination condition is satisfied, the optimal parameter values of the SWT-QPSO-LSTM model are obtained; otherwise, go to (d).

(f) The prediction results on the test set are obtained by the SWT-QPSO-LSTM prediction model constructed with the optimal parameter values. The prediction results are analyzed and summarized according to the evaluation indexes.

3.3. Model Evaluation Indexes. In this study, the indexes of root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are used to evaluate the prediction performance. The lower the values of RMSE, MAE, and MAPE, the better the prediction results of the proposed model. The calculation formulas of these evaluation indexes are as follows:

\[
\text{RMSE}(y_i, \bar{y}_i) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{y}_i)^2 / n},
\]

\[
\text{MAE}(y_i, \bar{y}_i) = \frac{1}{N} \sum_{i=1}^{N} |y_i - \bar{y}_i|, \tag{18}
\]

\[
\text{MAPE}(y_i, \bar{y}_i) = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \bar{y}_i}{y_i} \right| \times 100%.
\]

where \(n\) is the number of test samples; \(y_i\) and \(\bar{y}_i\) are, respectively, the truth value and predicted value at time \(i\).

4. Simulation Analysis

4.1. Data Selection and Analysis. In order to verify the predictive validity of our constructed model, four Chinese cities with different regions and different air quality levels, Chengdu, Wuhan, Shenyang, and Shijiazhuang, are selected as samples for empirical study in this paper. Chengdu, the

| Algorithm 1: QPSO algorithm. |
|-----------------------------|

\begin{verbatim}
Input: PopNum: population number. Maxstep: maximum optimization iterated algebra. Dim: optimization dimension.
Require: x_{j,t+1}: random initialization speed and position
Output: pbest; gbest
while step:1 to Maxstep do
    calculate m_{best} by equation (14)
    for i = 1 to PopNum
        for j = 1 to Dim
            if \(\mu > 0.5\)
                \[x_{j,t+1} = P_{j,t+1}^i + \beta^* | m_{best,j} - x_{j,t}^i | \ast \ln (1/\mu)\]
            else
                \[x_{j,t+1} = P_{j,t+1}^i - \beta^* | m_{best,j} - x_{j,t}^i | \ast \ln (1/\mu)\]
        end if
    end for
    fit (i) ← fitness \(x_{j,t+1}^i\)
    if fit (i) < f_{pbest} (i)
        pbest (i,:) ← x_{j,t+1}^i
        f_{pbest} (i) ← fit (i)
    end if
    if f_{pbest} (i) < f_{gbest}
        f_{gbest} ← f_{pbest} (i)
        gbest ← pbest (i,:)
    end if
end for
end do
end
\end{verbatim}
The largest city in southwest China, has a more reasonable industrial structure and better air quality, and its air quality index is in the top of the 168 key cities in China. Wuhan, an important city in central China, has seen significant improvement in its air quality with the transformation of its industrial structure; now the Air Quality Index of Wuhan is ranked in the upper middle of China’s 168 key cities. Shenyang is an important heavy industrial city, and its long-term economic development approach has led to rising resource and energy consumption; the air pollution problem is becoming increasingly acute; the Air Quality Index of Shenyang is in the lower reaches of 168 key cities in China. Shijiazhuang is located in the core of the Beijing-Tianjin-Hebei city cluster, it has poor air quality due to its special geographical location and unreasonable industrial layout, and its air quality index is at the end of 168 key cities in China. If the constructed model can well predict the air quality for these four cities, it means that the constructed model is effective. This paper uses the daily PM$_{2.5}$ concentration data from January 1, 2018, to March 31, 2021,
Comparison of signal before and after denoise

(a)

SWT time frequency diagram noise reduction

(b)

Figure 2: Continued.
Figure 2: SWT results of the raw PM$_{2.5}$ concentration data of Chengdu.

Figure 3: The variation of the particle fitness of the SWT-QPSO-LSTM model.
the case study data to determine the effectiveness of the constructed model. The data set can be obtained from the public air quality query website (https://www.aqistudy.cn/historydata/). The PM$_{2.5}$ concentration unit is $\mu$g/m$^3$, and there are 1,186 data points. The analysis revealed the phenomenon of missing values in the data, so the interpolation method was used to anomalously process the raw data, the data is divided into training and test sets according to a ratio of 7.3, the training data are used to fit the model, the test data are used to evaluate the performance of the constructed model, and their descriptive statistics are shown in Table 1.

4.2. Experimental Process. The section is divided into two parts: data preprocessing and hybrid model prediction.

4.2.1. Data Preprocessing Results after SWT. In this section, SWT is used to reduce the noise of the PM$_{2.5}$ concentration data. Figure 2 shows the SWT results of the original data, and Figure 2(a) presents the PM$_{2.5}$ concentration data before and after denoising; Figure 2(b) presents the SWT time-frequency diagram of the original data before and after noise reduction; Figure 2(c) presents the FFT spectrum of the original data before and after noise reduction. It can be seen from Figures 2(b) and 2(c) that the noise at the low frequency of SWT is well suppressed, and the frequency component is compressed. This indicates that SWT can suppress random noise and extract characteristic frequency. Based on this, the noise interference of the original sample data is reduced by the SWT method to achieve more accurate prediction results.

4.2.2. QPSO-LSTM Prediction Model Based on the SWT Method. To increase the prediction precision of the LSTM model, the QPSO algorithm is adopted to optimize the parameters of the LSTM model established on the SWT data preprocessing method. As we need to optimize the 4 parameters of the LSTM model, the optimization dimension is set to 4, in order to ensure that the parameter optimization could converge, the number of populations is set as large as possible, the population number is set to 100, and the value range of the learning rate is determined as [0.001, 0.01].

Taking the PM$_{2.5}$ concentration data of Chengdu as an example, the variation of the particle fitness with the increasing iterations of the SWT-QPSO-LSTM model is shown in Figure 3. The particle fitness finally stabilizes at 0.0070347.

The change of the maximum number of optimization iterations ($K$), learning rate ($\varepsilon$), and the number of hidden layer neurons ($L_1$ and $L_2$) with the number of model iterations is shown in Figure 4. It can be found that all four parameters converge at a certain point; the final parameter optimization result of the SWT-QPSO-LSTM prediction model is $K = 89$, $\varepsilon = 0.0075506$, $L_1 = 195$, and $L_2 = 60$. The results of parameter optimization for other models are shown in Table 2.

To verify the effectiveness of the SWT-QPSO-LSTM hybrid model proposed in this paper, the hybrid model is
compared with the single prediction model SWT-LSTM and the hybrid prediction model SWT-PSO-LSTM. Besides, to prove that the accuracy of the PM2.5 prediction model can be better improved by the SWT data processing method, the hybrid model proposed in this paper is compared with the QPSO-LSTM model without SWT processing and the QPSO-LSTM prediction model with the CEEMD data processing method. CEEMD is a decomposed method that is widely used in processing time-series data. Table 3 shows the evaluation indexes of the five prediction models on the PM2.5 concentration data of Chengdu, Wuhan, Shenyang, and Shijiazhuang. Figures 5(a)–5(d) show the prediction results obtained by five prediction models (SWT-LSTM, SWT-PSO-LSTM, SWT-QPSO-LSTM, CEEMD-QPSO-LSTM, and QPSO-LSTM) on the PM2.5 concentration data of Chengdu, Wuhan, Shenyang, and Shijiazhuang, respectively.

Data preprocessing has a greater impact on the prediction results, and the prediction results of the models with data preprocessing using either the SWT or CEEMD methods are significantly better than the prediction results of the models without data preprocessing; compared to the
Figure 5: Continued.
Figure 5: Continued.
Figure 5: Continued.
CEEMD method, data preprocessing by SWT method results have smaller prediction errors. The LSTM models, whose parameters are optimized by the swarm intelligence algorithm, have the better prediction; among them, the LSTM model optimized by QPSO algorithm predicts slightly better than those optimized by PSO. The above analysis results indicate that the SWT-QPSO-LSTM hybrid prediction model proposed in this paper achieves less prediction error than other prediction models. This verifies the feasibility and effectiveness of the established model.

5. Conclusion

Aiming at the prediction of PM$_{2.5}$ concentration, this paper proposes a hybrid prediction model based on the LSTM neural network and the QPSO optimization algorithm. The data used in this study is preprocessed by the SWT method, and the obtained data is input into the QPSO-LSTM model for simulation and prediction. The prediction results of the proposed model are analyzed, and the model is compared with other hybrid models. It is proved that the proposed method achieves better feasibility and effectiveness than other methods. The specific conclusions are as follows:

(1) The use of the SWT method to preprocess nonstationary signals can realize signal reconstruction, which improves the accuracy of subsequent data prediction.

(2) The use of the QPSO optimization algorithm to optimize the key parameters of the LSTM model can effectively avoid the impact of parameter setting on the prediction performance.

(3) The proposed SWT-QPSO-LSTM hybrid prediction model achieves better prediction performance than
QPSO-LSTM, CEEMD-QPSO-LSTM, SWT-LSTM, and SWT-PSO-LSTM, indicating that the SWT-QPSO-LSTM model can effectively predict PM$_{2.5}$ concentration.

The model proposed in this paper provides a theoretical basis for air pollution control; however, there are still some limitations. How to balance prediction accuracy and complexity of the model has always been a difficult task. Our proposed model has obvious advantages in prediction accuracy; however, the time spent on training the deep model is also worthy of attention. In the future, we consider embedding the model in the parallel framework such as Apache Spark to improve the model running efficiency. In addition, meteorological factors are not considered in this paper; we will add more influencing factors and use spatiotemporal deep learning neural network model to improve PM$_{2.5}$ prediction accuracy in the future work.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the projects of the China Postdoctoral Science Foundation (Grant no. 2020M670396), Social Science Foundation for Youths of Liaoning Province (Grant no. L20CJY001), and Liaoning Provincial Education Department Basic Research Project of China (Grant no. LJKR0574).

References

[1] J. Lelieveld, K. Klingmüller, A. Pozzer et al., "Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions," European Heart Journal, vol. 40, no. 20, pp. 1590–1596, 2019.

[2] L. Zhang, C. S. Lee, R. Zhang, and L. Chen, "Spatial and temporal evaluation of long term trend (2005-2014) of OMI retrieved NO2 and SO2 concentrations in Henan Province, China," Atmospheric Environment, vol. 154, pp. 151–166, 2017.

[3] D. Byun and K. L. Schere, "Review of the governing equations, computational algorithms, and other components of the models-3 community multiscale Air quality (CMAQ) modeling system," Applied Mechanics Reviews, vol. 59, no. 2, p. 51, 2006.

[4] L. Zhang, J. Lin, R. Qiu et al., "Trend analysis and forecast of PM$_{2.5}$ in Fuzhou, China using the ARIMA model," Ecological Indicators, vol. 95, pp. 702–710, 2018.

[5] Z. Zhang, L. Wu, and Y. Chen, "Forecasting PM$_{2.5}$ and PM$_{10}$ concentrations using GMCN(1,N) model with the similar meteorological condition: case of Shijiazhuang in China," Ecological Indicators, vol. 119, no. 3, Article ID 106871, 2020.

[6] W. Sun, H. Zhang, A. Palazoglu, A. Singh, W. Zhang, and S. Liu, "Prediction of 24-hour-average PM$_{2.5}$ concentrations using a hidden Markov model with different emission distributions in Northern California," Science of the Total Environment, vol. 443, pp. 93–103, 2013.

[7] S. Ausati and J. Amanollahi, "Assessing the accuracy of ANFIS, EEMD-GRNN, PCR, and MLR models in predicting PM$_{2.5}$," Atmospheric Environment, vol. 142, pp. 465–474, 2016.

[8] G. Goudarzi, P. K. Hopke, and M. Yazdani, "Forecasting PM$_{2.5}$ concentration using artificial neural network and its health effects in Ahvaz, Iran," Chemosphere, vol. 283, Article ID 131285, 2021.

[9] W. Liu, G. Guo, F. Chen, and Y. Chen, "Meteorological pattern analysis assisted daily PM$_{2.5}$ grades prediction using SVM optimized by PSO algorithm," Atmospheric Pollution Research, vol. 10, no. 5, pp. 1482–1491, 2019.

[10] S. Zhu, X. Lian, L. Wei et al., "PM$_{2.5}$ forecasting using SVR with PSOGSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors," Atmospheric Environment, vol. 183, pp. 20–32, 2018.

[11] Z. Tan, G. De, M. Li et al., "Combined electricity-heat-cooling-gas load forecasting model for integrated energy system based on multi-task learning and least square support vector machine," Journal of Cleaner Production, vol. 248, no. 14, Article ID 119252, 2020.

[12] H. Luo, M. Astitha, C. Hogrefe, R. Mathur, and S. T. Rao, "Evaluating trends and seasonality in modeled PM$_{2.5}$&lt;/sub&gt;&lt;/sub&gt; concentrations using empirical mode decomposition," Atmospheric Chemistry and Physics, vol. 20, no. 22, Article ID 13801, 2020.

[13] J. Chu, Y. Dong, X. Han, J. Xie, X. Xu, and G. Xie, "Short-term prediction of urban PM$_{2.5}$ based on a hybrid modified variational mode decomposition and support vector regression model," Environmental Science and Pollution Research, vol. 28, no. 1, pp. 56–72, 2020.

[14] Y. Ding, Z. Chen, W. Lu, and X. Wang, "A CatBoost approach with wavelet decomposition to improve satellite-derived high-resolution PM estimates in Beijing-Tianjin-Hebei," Atmospheric Environment, vol. 249, Article ID 118212, 2021.

[15] J. Ma, Y. Ding, J. C. P. Cheng et al., "Identification of high impact factors of air quality on a national scale using big data and machine learning techniques," Journal of Cleaner Production, vol. 244, Article ID 118955, 2020.

[16] M. Niu, Y. Wang, S. Sun, and Y. Li, "A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM$_{2.5}$ concentration forecasting," Atmospheric Environment, vol. 134, pp. 168–180, 2016.

[17] W. Sun and J. Sun, "Daily PM$_{2.5}$ concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm," Journal of Environmental Management, vol. 188, pp. 144–152, 2017.

[18] Y. Cheng, H. Zhang, Z. Liu, L. Chen, and P. Wang, "Hybrid algorithm for short-term forecasting of PM$_{2.5}$ in China," Environmental Science and Pollution Research, vol. 200, pp. 264–279, 2019.

[19] Y. S. Chang, H. T. Chiao, S. Abimannan, Y. P. Huang, Y. T. Tsai, and K. M. Lin, "An LSTM-based aggregated model for air pollution forecasting," Atmospheric Pollution Research, vol. 11, no. 8, pp. 1451–1463, 2020.

[20] S. Dhakal, Y. Gautam, and A. Bhattarai, "Exploring a deep LSTM neural network to forecast daily PM$_{2.5}$ concentration using meteorological parameters in Kathmandu Valley, Nepal," Air Quality, Atmospheric & Health, vol. 14, no. 1, pp. 83–96, 2021.

[21] K. Yan, W. Li, Z. Ji, M. Qi, and Y. Du, "A hybrid LSTM neural network for energy consumption forecasting of individual households," IEEE Access, vol. 7, Article ID 157633, 2019.
[22] Z. Zhang, Y. Zeng, and K. Yan, “A hybrid deep learning technology for PM2.5 air quality forecasting,” *Environmental Science and Pollution Research*, vol. 28, no. 29, Article ID 39409, 2021.

[23] N. Jin, Y. Zeng, K. Yan, and Z. Ji. "Multivariate air quality forecasting with nested long short term memory neural network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 8514–8522, 2021.

[24] I. Daubechies, J. Lu, and H. T. Wu, "Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 243–261, 2011.

[25] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[26] J. Sun, W. Xu, and B. Feng, “A global search strategy of quantum-behaved particle swarm optimization,” *Cybernetics and Intelligent Systems*, vol. 1, pp. 111–116, 2004.