A novel particulate matter 2.5 concentration prediction model based on double-layer decomposition and feedback of model learning effect

HE WANG, AND HONGFENG CHEN
School of Technology, Beijing Forestry University, Beijing 100083, China
Corresponding author: He Wang (e-mail: wanghe@bjfu.edu.cn).

This work was supported in part by the Central Universities under Grant BLX2014-05 and in part by the National Natural Science Foundation of China under Grant 52107069.

Abstract: Accurate and effective particulate matter 2.5 (PM2.5) concentration prediction can provide early warning information for decision-making departments, so as to take governance and preventive measures. A combined model based on double-layer decomposition (DLD) and feedback of model learning effect for PM 2.5 concentration prediction is proposed in this paper. Firstly, ensemble empirical mode decomposition (EEMD) and variational mode decomposition (VMD) were used for double-level decomposition for the PM2.5 concentration series, to reduce the nonstationarity and nonlinearity of the concentration series and improve the predictability; Secondly, a Wavelet neural network (WNN) prediction model based on the feedback of model learning effect is established for the subsequence obtained by double-layer decomposition. Finally, the prediction results of each subsequence are superimposed to obtain the final prediction results. The case study shows that the prediction model proposed in this paper is scientific.

INDEX TERMS: Double-layer decomposition; ensemble empirical mode decomposition; variational mode decomposition; feedback of model Learning effect; particulate matter 2.5 concentration; wavelet neural network

I. Introduction

A. Motivation
With the rapid development of industry, the problem of particulate matter pollution in the air has become a global problem. Compared with coarser atmospheric particulate matter, fine particulate matter (PM2.5) has a smaller particle size, a larger area, and strong activity. Substances (for example, heavy metals, microorganisms, etc.), and their long residence time in the atmosphere and long transportation distance, thus have a greater impact on human health and atmospheric environmental quality [1].

Therefore, If the concentration of PM2.5 can be accurately predicted, the air quality monitoring department can understand the future air conditions and change trends in advance, analyze the causes of changes, and take effective and feasible prevention and treatment measures. At the same time, the prediction information can also remind people to avoid going out when serious air pollution is about to occur or take some protective measures in advance to reduce the harm of pollutants to the body. Therefore, it is of great significance to carry out accurate and effective PM2.5 concentration prediction and early warning in areas with serious air pollution.

B. Related works about prediction models
Researchers can predict PM2.5 concentration prediction models into three categories: one is a prediction model based on the mechanism of PM2.5 occurrence, this is called a mechanism-driven model. The second type of prediction model is based on the data-driven prediction model, which is called the data-driven model in this article. The third type is the combined prediction model. The combined prediction model has a combination of mechanism prediction and statistical prediction, as well as a combination of multiple data-driven prediction models.
The mechanism prediction model is based on the study of the occurrence and development mechanism of PM2.5, combined with the process of chemical substance transfer, and the sources and changes of chemical substances. Combined with high-resolution aerosol products and weather research and prediction model meteorological field, a regional scale regression model is established to predict PM2.5 concentration [1]. Five three-dimensional chemical transport models were applied to simulate PM2.5 concentration in Central Europe [2]. The chemical transport model is used to predict PM10 in Beijing [3].

This kind of model effectively analyzes the movement of PM2.5 occurrence and development, Mechanism prediction modeling effectively analyzes the movement mechanism of PM2.5 occurrence and development, so that people can better understand the generation and development process of PM2.5. However, the modeling of the mechanism prediction model needs more analysis processes, the model is more complex, the calculation time is more time-consuming, and the simulation results are also vulnerable to the impact of pollutant emissions and selected variables [4].

The data-driven prediction model is built by analyzing the characteristics of PM2.5 concentration data. The data-driven prediction model is divided into statistical model and machine learning prediction model. The advantages of the statistical prediction method, such as simple modeling process and easy operation, make it applied to air pollution prediction. The common models of statistical prediction methods include the Markov model [5], the Lagrange coupling model [6], Arima [7,8]. The concentration series of PM2.5 has not only linear characteristics, but also some nonlinear and non-stationary characteristics.

Facing these problems, the machine learning prediction model has attracted people's attention. Through analyzing and training a large number of historical data with machine learning technology, the machine learning prediction method obtains the prediction model that can track the changing trend of the PM2.5 concentration series. The typical representatives include shallow and deep artificial neural network model (ANN) [9-12], support vector machine model (SVM) [13] and least squares support vector machine model (LSSVM) [14]. Among various machine learning models, WNN has the advantages of multi-resolution of wavelet transform and the self-learning ability of the neural network. However, it has not been found to be applied to PM2.5 concentration prediction.

There are mainly three ways to combine the combined forecasting model, one is to combine the machine model and the data-driven model [15]. The second is the combination of multiple models [16,17]. The third is the combination of data analysis and data modeling [18-24].

C. Related works about data preprocessing

One of the reasons for the high prediction accuracy of the third model is that the effective decomposition of PM2.5 concentration series data with a certain scale or fluctuation trend can reduce the nonstationarity of the signal and improve the predictability of the original series. Data processing methods can be divided into EMD is based methods [19-21] and VMD is based methods [22,23].

Among various decomposition methods, ensemble empirical mode decomposition is an effective improvement on the traditional empirical mode decomposition method, which solves the mode aliasing problem of the traditional empirical mode decomposition [19-21]. The subsequence obtained by EEMD decomposition performs well in the low-frequency range, but the volatility in the high-frequency range is also too large, which is not conducive to prediction. VMD uses a non-recursive form to decompose the original sequence into a series of eigenmode functions [22-24] which has a better effect in restraining high-frequency fluctuations. However, it is relatively general in low-frequency areas. Therefore, this paper explores a two-layer decomposition method integrating the advantages of EEMD and VMD for data preprocessing.

D. Contributions of this study

Considering the above analysis, this paper proposes a new PM2.5 concentration prediction model based on double-layer decomposition and feedback of model learning effect of wavelet neural network (FMLE-WNN). Firstly, the wind speed series is processed by EEMD to obtain a series of subsequences. Secondly, based on the sample entropy theory, the subsequences whose sample entropy value is greater than the average value of all subsequences are superimposed to form a new sequence. The new sequence is decomposed by VMD to obtain a series of subsequences. Then, FMLE-WNN is used to model and predict the subsequence. Finally, all wavelet neural network prediction results are superimposed to obtain the final prediction result. Correspondingly, the main contributions of this study are listed as follows:

1) A data processing method of double-layer decomposition is proposed. Based on EEMD and VMD, EEMD is used to decompose the first layer data to obtain a series of subsequences, and then VMD is used to decompose the subsequences in which the sample entropy is higher than the average value. The double-layer decomposition is refined to mine PM2.5 sequence characteristics.

2) WNN is applied to the concentration prediction of PM2.5. WNN has the advantages of wavelet transform and artificial neural networks. It has a strong learning ability for sequence. However, it has not been found to be used in PM2.5 concentration prediction.

3) Model learning effect feedback is used to improve the prediction ability of WNN, and the model frame structure affects the prediction ability of WNN. This paper explores using model learning effect feedback to determine the network frame structure of WNN.

4) Compared with other models mentioned in Section 4, the model proposed in this paper achieves superior performance with relatively high accuracy.
The framework of the study is as follows. Section II introduces the Methodology of the proposed model. Section III describes the flow chart and the proposed model. Section IV presents the PM 2.5 concentration prediction experiments and the test results of the proposed model. In this section, two datasets are used for analysis and discussion. Section V concludes this paper.

II. Methodology

A. Ensemble empirical mode decomposition (EEMD)

Empirical mode decomposition is essentially an adaptive signal screening method, which can screen out the trends of different features existing in the original sequence step by step to obtain the intrinsic mode function (IMF) with the same characteristics \( \{x(t)\} \). EMD steps are as follows [19]:

1) Find all the maximum and minimum values in the sequence \( x(t) \). A cubic spline function is used to interpolate the upper envelope \( b_{\text{max}} \) and the lower envelope \( b_{\text{min}} \).

2) Calculate the average value of upper and lower envelope \( m(t) \), where \( m(t) = \frac{b_{\text{max}}(t) + b_{\text{min}}(t)}{2} \). Extract \( h(t) = x(t) - m(t) \), and judge whether \( h(t) \) meets the condition of the natural modal component. If yes, \( h(t) \) is the first IMF. If not, \( h(t) \) is taken as the original series.

3) Repeat step (1) and step (2) until the difference \( h_{n}(t) \) after times of filtering meets the inherent modal component. If not, \( h_{n}(t) \) is a monotonic function.

4) According to Eq. 1, \( c_{t}(t) \) is obtained from the signal \( x(t) \), and the remaining component \( r_{1}(t) \) is obtained:

\[
r_{1}(t) = x(t) - c_{t}(t) \tag{1}
\]

5) Repeat the above steps for \( r_{1}(t) \) to obtain the remaining IMF components, and terminate when the residual \( r_{n}(t) \) is a monotonic function.

Compared with the traditional wavelet decomposition algorithm, EMD reduces the influence of human factors on the decomposition results, and it has a certain degree of advancement. However, the above algorithm will appear modal aliasing in some cases. Set EEMD can effectively avoid this problem by using noise characteristics. The steps of set EEMD are as follows [25,26]:

1) Add the white noise sequence to the PM 2.5 concentration sequence \( \{x(t)\} \);

2) Use EMD to decompose the PM 2.5 concentration sequence after adding white noise into several intrinsic modal components \( c_{n}(t) \) and a residual component \( r_{n}(t) \);

3) Repeat steps 1 and 2 a total of \( s \) \( r \) times, and the amplitude of the white noise sequence added each time is different;

4) Take the overall average of the IMF obtained by \( s \) \( r \) decomposition and use it as the final IMF component of the original signal;

The white noise sequence added in the above steps should follow the normal distribution of \( (0, (\alpha \varepsilon)^2) \), where \( \alpha \) is the noise intensity parameter and \( \varepsilon \) is the standard deviation of the signal. When \( r \) is 100 and \( \alpha \) is selected from \( [0.1, 0.3] \), a better decomposition result can be obtained. Therefore, in this article, \( r \) is selected to be equal to 100, and \( \alpha \) is equal to 0.25.

B. Variational Mode Decomposition (VMD)

Variational mode decomposition is a completely non-recursive variational mode decomposition model [27,28]. Its purpose is to decompose a signal \( f(t) \) into discrete patterns \( u(k)(k = 1, 2, 3, \cdots K) \). For each mode \( u(k) \) mainly surrounds a center frequency \( \omega(k) \)

Decompose sequence \( f(t) \) into a series of finite bandwidth modal functions, the main process is as follows:

1) For each mode \( u(k) \), the Hilbert transform is used to calculate the correlation analysis signal to obtain the one-sided spectrum.

2) For each mode \( u(k) \), the spectrum of the mode is transferred to the baseband and an index is tuned to the respective estimated center frequency.

3) Estimate the bandwidth using the \( H^1 \) Gaussian smoothness of the demodulated signal. Then, the constraint variation problem can be expressed as follows:

\[
\begin{align*}
\min_{\{u_k\}, \{\omega_k\}} & \sum_{k=1}^{K} \left\| \left( \delta(t) + \frac{j}{\pi t} \right) u(k) \right\|_{2}^2 \\
\text{s.t.} & \sum_{k=1}^{K} u(k) = f(t) \\
\end{align*}
\]

where \( u(k) \) indicates the \( k \) mode; \( \omega(k) \) is the center frequency; \( K \) is the total number of modes; \( \delta(t) \) is the Dirac distribution; \( f(t) \) is the original signal. In order to eliminate the constraints in Eq. (2), a quadratic penalty term and Lagrange multiplier are introduced to transform constrained optimization into unconstrained optimization:

\[
\begin{align*}
L([u_k], \{\omega_k\}, \lambda) &= \alpha \sum_{k=1}^{K} \left\| \delta(t) + \frac{j}{\pi t} \right\|_{2}^2 \\
&+ \left\| f(t) - \sum_{k=1}^{K} u_k \right\|_{2}^2 \\
&+ \sum_{k=1}^{K} \lambda(t, f(t) - \sum_{k=1}^{K} u_k(t)) \\
\end{align*}
\]

Where \( \alpha \) is the penalty parameter; \( \lambda \) is the Lagrange multiplier; \( \otimes \) is the convolution operator.

Therefore, the original minimization problem is transformed into an optional model, which can be solved by the alternating allocation multiplier method (ADMM) method. According to ADMM, the update formulas of \( u(k) \) and \( \omega(k) \) are as follows.

1) Minimize \( u(k) \)

\[
u(k) = \frac{\tilde{f}(\omega) - \sum_{i=\xi} \bar{a}_{i}^{n} + \bar{\delta}^{n}(\omega)}{1 + 2\alpha(\omega - \omega_{k}^{n})^{2}} \tag{4}
\]

2) Minimize \( \omega(k) \)

\[
\omega(k) = \frac{\int_{0}^{\infty} \omega |\bar{a}_{k}^{n+1}(\omega)|^2 d\omega}{\int_{0}^{\infty} |\bar{a}_{k}^{n+1}(\omega)|^2 d\omega} \tag{5}
\]

where \( \omega \) represents the symbol of the signal changing from the time domain \( t \) to the frequency domain. n
represents the number of iterations: \( \hat{u}_k^{n+1}, \hat{f}(\omega), \hat{u}(\omega) \) and \( \hat{\lambda}(\omega) \) are the Fourier transforms of \( u(t) \), \( f(t) \), \( u(t) \) and \( \lambda(t) \) respectively [27].

C. Sample entropy (SE)

Sample Entropy (SE) [29] is similar in physical meaning to approximate entropy. Both measure the sequence complexity by measuring the probability of generating a new pattern in the signal. Compared with approximate entropy, sample entropy has two advantages: 1) The calculation of entropy does not depend on the data length; 2) The sample entropy has a better consistency. The higher the sample entropy value, the more difficult the prediction. Sample entropy is currently used to assess the complexity of physiological time series (EEG[30], SEMG[31], etc.) and to diagnose pathological conditions. The specific sample entropy calculation process can be found in the literature.

D. Error evaluation index

Scientific and effective error evaluation standards are of great significance for judging the performance of the model. However, at present, there is no uniform error evaluation standard for PM2.5 concentration prediction. This article selects the following two commonly used error standards to judge the model performance, including Mean absolute error (MAE) (\( \mu g/m^3 \)) and Mean Square Percent Error (MSPE) (%), which are expressed as follows:

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |\bar{y}(i) - y(i)|
\]

\[
MSPE = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{|\bar{y}(i) - y(i)|}{y(i)} \right)^2 \times 100\%
\]

Where \( \bar{y}(i) \) is the predicted value, \( y(i) \) is the actual value.

E. Wavelet neural network (WNN)

Wavelet neural network is the combination of wavelet theory and artificial neural network. It has the advantage of multi-resolution of wavelet transform and the self-learning ability of neural networks [32]. The basic idea is to use the wavelet basis function as the excitation function of the neuron to establish the relationship between wavelet transform and the neural network. Because the wavelet neural network inherits the characteristics of wavelet decomposition, it can learn in the loose area of sequence distribution at low resolution, and in the dense area of sequence distribution at high resolution. These characteristics are conducive to the neural network to "capture" the internal law between input and output data more easily. Therefore, the wavelet neural network is more intelligent, efficient than the traditional neural network.

Let \( \psi(t) \in L^2(\mathbb{R}) \) be a wavelet basis function, and \( \psi(t) \) satisfy the admissibility condition:

\[
\int_{-\infty}^{\infty} \frac{|\psi(\omega)|^2}{|\omega|} d\omega < +\infty
\]

Where \( \psi(\omega) \) is the Fourier transform of \( \psi(t) \). \( \psi(t) \) generates a set of wavelet basis functions after translation and expansion:

\[
\psi_{a,t}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-t}{a}\right)
\]

Where \( a \) is the scale factor, \( \tau \) is the expansion factor. For signal \( x(t) \in L^2(\mathbb{R}) \), the wavelet transform is:

\[
W_x(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \psi\left(\frac{t-t}{a}\right) dt
\]

Figure 1: Topological structure of wavelet neural network

The topological structure of the three-layer wavelet neural network model is shown in Fig. 1.

In Fig. 1. \( X_i (i = 1,2,\ldots,t) \) is the input sequence of WNN, \( Y \) is the output sequence, \( w_{ij} \) is the connection weight between the input layer and the hidden layer, \( w_{j} \) is the connection weight between the hidden layer and the output layer. When the input signal sequence is \( x_i (i = 1,2,\ldots,k) \), the output calculation formula of the hidden layer is:

\[
g(j) = \psi_j \left[ \sum_{i=1}^{k} w_{ij} x_i - \tau_j \right] / a_j, j = 1,2,\ldots,l
\]

Where \( g(j) \) is the output of the \( j \)-th node of the hidden layer, \( \psi_j \) is the wavelet basis function, \( a_j \) is the expansion factor of \( \psi_j \), \( \tau_j \) is the translation factor of \( \psi_j \), and the calculation formula of the output layer is:

\[
Y = \sum_{j=1}^{l} w_{j} g(i) j = 1,2,\ldots,l
\]

The basic strategy of WNN neural network learning is to use the principle of error function minimization to constantly change the shape and scale of the wavelet base. The correction process of weight and wavelet basis function coefficient is as follows:

1) Calculation of WNN network fitting error

\[
e = \hat{y}(k) - Y
\]

Where \( \hat{y}(k) \) is the expected output and \( Y \) is the WNN fitting output.

2) Correction of wavelet neural network weight and wavelet basis function coefficient based on error \( e \) as follow:
Determine the input dimension and the number of neurons

Start

Iterate over all combinations of input dimensions and neurons

Model learning effect feedback comparison

Output optimal input dimension and neuron number combination

Training the best prediction model of FLE-WNN and making prediction

End

Figure 2. Flowchart of FMLE-WNN

F. Feedback of the model learning effect (FMLE)

As a kind of neural network, WNN has the characteristics of a neural network. The framework of network structure has a great influence on the prediction performance of WNN. When the WNN dimension is too small, it can't perfectly mine the data features that affect the output layer. Too much dimension of the input layer will reduce the generalization ability of the model. The determination of the dimension of the input layer of the neuron is based on the reconstruction of the phase space and the ARIMA, ARMA, and other methods based on the box Jenkins theory. Both methods can effectively mine the data features, but it is not clear about the learning effect of the model on the data, which is similar to a person providing better food, but the absorption effect of food is not clear. The number of hidden layer neurons in the model will also affect the performance of the network to a large extent. If the number of hidden layer nodes is too small, the fault tolerance of the network will be reduced, and a complex mapping relationship will not be established, even the model will not be able to train the network and results. If too many hidden layer neurons are selected, the training error will be affected, and the learning time of the network will be delayed. Long. In order to solve the problem of the input layer dimension and the number of hidden layer neurons, this paper proposes a neural network framework determination method based on the learning effect feedback of the WNN model. The flow chart is shown in Fig. 2, and the corresponding steps are as follows:

\[
\begin{align*}
\Delta w_{n,k}^{t+1} &= w_{n,k}^t + \Delta w_{n,k}^t \\
\Delta a_k^{t+1} &= a_k^t + \Delta a_k^t \\
\Delta r_k^{t+1} &= r_k^t + \Delta r_k^t
\end{align*}
\]  

(14)

Where \( \Delta w_{n,k}^{t+1}, \Delta a_k^{t+1}, \Delta r_k^{t+1} \) are calculated by the network prediction error \( e \) according to Eq. (15):

\[
\begin{align*}
\Delta w_{n,k}^{t+1} &= -\eta \frac{\partial e}{\partial w_{n,k}^t} \\
\Delta a_k^{t+1} &= -\eta \frac{\partial e}{\partial a_k^t} \\
\Delta r_k^{t+1} &= -\eta \frac{\partial e}{\partial r_k^t}
\end{align*}
\]

(15)

Where \( \eta \) is the learning efficiency. The choice of learning step size \( \eta \) in WNN is very important. If \( \eta \) is too large, it will cause instability, but the convergence speed is fast; if \( \eta \) is small, the convergence speed is slow, but instability can be avoided. To overcome this contradiction, it can be improved by adding momentum terms. The WNN learning process formula of increasing momentum term can be modified as follows:

The specific steps of the improved wavelet neural network algorithm prediction model are as follows:

1) Initialize the network. \( a_k, r_k \) and network connection weights \( w_{ij}, w_j \) are initialized randomly.

2) Calculation of fitting error. Input the training samples into the network, calculate the prediction output of the network, and calculate the error between the output and the expected output.

3) Modify the weight and wavelet basis function coefficients. Modify the network weight and wavelet function parameters according to Eq. (15).

4) Determine whether the algorithm ends. When the absolute value of the error function is less than the preset error \( \xi \) or the maximum number of iterations is reached, it ends; no, return to step 2).

The performance of the wavelet neural network prediction model is also related to the selection of the wavelet basis function. There are many kinds of wavelet bases, among which the Morlet function has strong robustness and adaptability in regression prediction. In this paper, the Morlet function is used as a wavelet basis function, and its mathematical expression is as follows:

\[
h_j = \cos(1.75x) e^{-\frac{x^2}{2}}
\]

(16)
1. The first step is to determine the input dimension and the number of neuron nodes.
2. The second step is to train the WNN model of each combination and determine the learning effect of each combination.
3. The third step is to Evaluate the learning effect of the model. At the end of model training, the training data of one ten minutes in the sample is selected randomly to judge the learning fitting effect of the model. In this paper, MSPE is used to judge the learning effect of the model.
4. The fourth step is to output the best WNN framework combination.

III. EEMD-VMD-FMLE-WNN modeling process

PM2.5 concentration series is a typical non-linear and non-stationary series, and EEMD-VMD- FMLE-WNN prediction model modeling structure flow chart is shown in Fig. 3. The modeling process is as follows:

Step 1: The original data sequence is acquired, and the training and testing data of the sequence are divided.

Step 2: PM2.5 time series is decomposed by EEMD and a series of subsequences are obtained.

Step 3: The SE of each subsequence is calculated, and get the average value D.

Step 4: If the SE of the subsequence is less than D, it will be left. All the subsequences whose SE is greater than D are superposed to form a new sequence.

Step 5: VMD is used to decompose the new sequence formed in the previous step, and the new subsequence formed after VMD decomposition is obtained.

Step 6: The subsequences left after EEMD decomposition and VMD decomposition are combined to form a new subsequence set.

Step 7: A FMLE-WNN prediction model based on model learning effect feedback is established for all subsequences.

Step 8: The prediction results of each subsequence are superposed to get the final prediction results.

Step 9: Analyze the error of the prediction results.

IV. Case study

A. Data acquisition and double-layer decomposition

In order to verify the effectiveness and advanced nature of the model, two city PM2.5 data are selected for research and analysis. The two cities are located in Beijing and Guangzhou.
respectively. Beijing, located in North China, is the capital of the people’s Republic of China, a megacity, a national political center, cultural center, international communication center, and scientific and technological innovation center. The climate of Beijing is a typical warm temperate semi-humid continental monsoon climate, with high temperature and rain in summer, cold and dry in winter. Located in South China, Guangzhou is the capital of Guangdong Province and the political, military, economic, cultural, scientific, and educational center of South China. Guangzhou is the radiation center of Guangfu culture and the world-famous Oriental port city. Guangzhou is located on the subtropical coast. The Tropic of Cancer passes through the south-central part. It has a marine subtropical monsoon climate, characterized by warm and rainy, sufficient light and heat, long summer, and a short frost period.

Figure 4. The original PM2.5 concentration series of Beijing and Guangzhou

Figure 5. IMF obtained after EEMD decomposition (Beijing)
**Figure 6.** IMF obtained after EEMD (Guangzhou)

**Figure 7.** SE value of each IMF (Beijing)
Figure 8. SE value of each IMF (Guangzhou)

Figure 9. The new sequence is decomposed by VMD (Beijing)
Figure 10. SE value comparison of each IMF in two-layer decomposition (Beijing)

Figure 11. The new sequence is decomposed by VMD (Guangzhou)
**Figure 12.** SE value comparison of each IMF in two-layer decomposition (Guangzhou)

**Figure 13.** Beijing prediction results after the first type decomposition (DLD)

**Figure 14.** Guangzhou prediction results after the first type decomposition (DLD)
Figure 15.  Beijing prediction results after the second type decomposition (VMD)

Figure 16.  Guangzhou prediction results after the second type decomposition (VMD)

Figure 17.  Beijing prediction results after the third type decomposition (EEMD)
All the experiments in this paper are conducted in WINDOWS10, equipped with 1.80 GHz, Intel(R) core i7-8550U CPU, and 8 GB RAM. All simulation experiments were carried out with Matlab R2018b (The MathWorks, Inc., Natick, MA, USA).

The two original daily PM2.5 concentration time series data from 1 January 2017 to 31 December 2018 in Beijing city and Guangzhou city. They were from the China National Environmental Monitoring Center. They were the values of 24 hours averages. as shown in Fig. 4. There are 730 data in each city, the first 530 are model training data, and the last 200 are model testing data. The PM2.5 concentration series data of Beijing and Guangzhou were pre-processed and decomposed by EEMD. The decomposition results are shown in Fig. 5 and Fig. 6. Calculate the SE value of each subsequence obtained by EEMD decomposition, and choose the embedding dimension \( m = 2 \) when calculating the sample entropy value; similar tolerance \( r = 0.25 \times \text{std} \), where std represents the standard deviation of the calculated sequence. SE value calculation results are shown in Fig. 7 (Beijing) and Fig. 8 (Guangzhou).

As can be seen from Fig. 7, the SE values of the four sequences of IMF6-IMF9 in each IMF obtained from the decomposition of PM2.5 concentration sequences in Beijing by EEMD are greater than the average value \( D \). Too large SE values indicate poor predictability. The four subsequences of IMF6-IMF9 are superimposed to obtain a new subsequence. VMD decomposition is used for the new subsequence. Since the number of sequences before the merge is 3, the \( K \) value of the decomposition subsequence when setting VMD decomposition is 3, the new sequence and VMD decomposition. The results are shown in Fig. 11. The four subsequences after the decomposition replace the IMF6-IMF9 before the decomposition to form a subsequence set after secondary decomposition in Guangzhou. The SE values of the first six IMFs in the new sequence set have not changed. The SE values of the last three IMFs are shown in Fig. 12.

### B. Determination of FMLE-WNN prediction model framework

As can be seen from Fig. 10 and Fig. 12, after the second decomposition, the SE value in the combined new sequence set is reduced to some extent. To establish the WNN prediction model based on the feedback of the model learning effect mentioned in this article for the distribution of each subsequence, the model framework of WNN needs to be determined first. Based on the method of determining the WNN framework based on the model learning effect proposed in Section II of this article, the sequence determines the model framework. Because of the characteristics of the data, this paper considers the selected input dimension as \( p \), and sets the range of \( p \) to \( 3 \leq p \leq 8 \). When the input dimension is \( p \), the number of hidden layer nodes is \( q \), \( q \) range is \( p \leq q \leq 3p \). All frame combinations are used to learn and fit the training data of each subsequence, and the optimal frame is output according to the learning results. The optimal WNN frame combination data of each subsequence in Beijing and Guangzhou are shown in Table I. After determining the optimal frame data of each subsequence, the
training data is used again to train the WNN in preparation for prediction.

**TABLE I**

| Beijing       | IMF   | Input dimension | Number of neurons |
|---------------|-------|----------------|-------------------|
| IMF1          | 4     | 7              |                   |
| IMF2          | 3     | 6              |                   |
| IMF3          | 5     | 10             |                   |
| IMF4          | 6     | 11             |                   |
| IMF5          | 4     | 10             |                   |
| IMF6          | 6     | 13             |                   |
| IMF7          | 5     | 12             |                   |
| IMF8          | 6     | 12             |                   |
| IMF9          | 5     | 13             |                   |

**Guangzhou**

| IMF1          | 4     | 9              |                   |
| IMF2          | 4     | 7              |                   |
| IMF3          | 5     | 11             |                   |
| IMF4          | 5     | 9              |                   |
| IMF5          | 6     | 11             |                   |
| IMF6          | 5     | 11             |                   |
| IMF7          | 6     | 10             |                   |
| IMF8          | 5     | 12             |                   |
| IMF9          | 6     | 13             |                   |

**C. Prediction results**

In order to verify the performance of the prediction model proposed in this paper, this paper uses the following three types of six models for comparative study. The first type of model is based on the Double-layer Decomposition (DLD) data proposed in the previous paper, one is based on the WNN proposed in this paper, which is based on the feedback of the model learning effect (FMLE). This paper is called the DLD-LFE-WNN model; The second is to build a prediction model based on PSO to optimize the performance of LSSVM (PSO-LSSVM), which is called the DLD-PSO-LSSVM model. The prediction results of the first type of model are shown in Fig. 15 (Beijing) and Fig. 16 (Guangzhou). The third type of model is based on EEMD processing data. One is based on the WNN proposed in this paper, which is based on the feedback of the model learning effect. This is called the EEMD-FMLE-WNN model; the second is based on PSO-LSSVM, which is called the EEMD-PSO-LSSVM model. The prediction results of the third type of model are shown in Fig. 17 (Beijing) and Fig. 18 (Guangzhou). The prediction errors of all models are shown in Table II.

**TABLE II**

| Models            | MAE (ug/m³) | MSPE (%) |
|-------------------|-------------|----------|
| DLD-FMLE-WNN      | 16.491      | 9.81     |
| DLD-PSO-LSSVM     | 17.0317     | 10.51    |
| VMD-FMLE-WNN      | 17.6569     | 12.00    |
| VMD-PSO-LSSVM     | 17.5882     | 12.49    |
| EEMD-FMLE-WNN     | 18.0947     | 12.77    |
| EEMD-PSO-LSSVM    | 18.7525     | 12.93    |

**Guangzhou**

| Models            | MAE (ug/m³) | MSPE (%) |
|-------------------|-------------|----------|
| DLD-FMLE-WNN      | 6.0155      | 2.11     |
| DLD-PSO-LSSVM     | 6.0399      | 2.23     |
| VMD-FMLE-WNN      | 6.4675      | 2.27     |
| VMD-PSO-LSSVM     | 6.5892      | 2.35     |
| EEMD-FMLE-WNN     | 6.7115      | 2.85     |
| EEMD-PSO-LSSVM    | 6.8817      | 2.91     |

**D. Discussion**

It can be found from Table II that the prediction error of EM is less than that of other models, whether Beijing or Guangzhou, which shows the advanced and robustness of the model. The minimum Mae value of Beijing prediction error is 16.491 ug/m³, and the minimum MSPE is 9.81%. The best prediction results of each model in Guangzhou are 6.0155ug / m³ and 2.11%. The minimum prediction error of the two cities is quite different, which is related to the fluctuation of PM2.5 concentration in the two cities. The fluctuation range of Beijing is large and the prediction is difficult, while the fluctuation range of Guangzhou is smaller and the prediction is more predictable. This shows that the predictability of the sample has a great influence on the prediction effect of the model.

Comparing the prediction results of various methods, it is found that in terms of the data samples in this paper, no matter in Beijing or Guangzhou, the prediction effect of double-layer decomposition proposed in this paper is better than that of...
VMD and EEMD. This shows that data mining can improve predictability to a certain extent.

Comparing the prediction results of three kinds of data processing methods, it can be found that the prediction model proposed in this paper, in the 12 groups of error comparison between Beijing and Guangzhou, except for Beijing's The MAE of VMD-PSO-LSSVM model is better than that of VMD-FMLE-WNN model, and the prediction error of the other 11 groups is less than that of PSO-LSSVM, and the better prediction effect is 91.67%.

In general, the combination prediction model proposed in this paper has a better data processing effect on the one hand and better performance on the other hand. The combination of the two makes the prediction model have better prediction performance.

V. Conclusions
A novel PM2.5 concentration prediction model EEMD-VMD-FMLE-WNN is presented. For the first time, the denoising method FMLE-WNN is systematically introduced by the authors for the PM2.5 concentration prediction application. Moreover, promising feedback of model learning effect of WNN(FMLE-WNN), specialized in extracting time-series information, is applied to forecast the future PM2.5 concentration series. With the combined advantages of the two promising methods, superior performance of the proposed combined model for PM2.5 concentration prediction with relatively high accuracy, compared with the other models mentioned herein.

According to the test results of the prediction by EEMD-VMD-FMLE-WNN, it can be summarized that 1) The proposed EEMD-VMD-FMLE-WNN model can evidently achieve a much higher forecasting accuracy compared with the one layer methods; 2) It is the FMLE-WNN, instead of the other forecasting methods mentioned herein, which has the better forecasting performance when it cooperates with WNN only. 3) The EEMD-VMD-FMLE-WNN owns the smaller volatility of results in the experiments, compared to the other forecast methods.

Based on the two different real PM2.5 concentration datasets, it reveals that the proposed combined model has superior performance, particularly in adaptability in practice.

Appendix A. List of abbreviations

| Abbreviation | Description |
|--------------|-------------|
| ANN          | artificial neural network |
| ARIMA        | autoregressive integrated moving average |
| ARMA         | autoregressive moving average |
| DLD          | double-layer decomposition |
| EEG          | electroencephalogram |
| EEMD         | ensemble empirical modal decomposition |
| EMD          | empirical modal decomposition |
| FMLE         | feedback of model learning effect |
| LSSVM        | least-squares support vector machine model |
| MAE          | mean absolute error |
| MSPE         | mean square percent error |
| PSO          | particle swarm optimization |
| PM2.5        | particulate matter 2.5 |
| PM10         | particulate matter 10 |
| RMSE         | root mean square error |
| SE           | Sample entropy |
| SEMG         | surface Electromyography |
| SVM          | support vector machine |
| VMD          | variational modal decomposition |
| WNN          | wavelet neural network |
References

[1] Y. Sathe, S. Kulkarni, P. Gupta, A. Kaginalkar, S. Islam, P. Gargava, “Application of Moderate Resolution Imaging Spectroradiometer (MODIS) Aerosol Optical Depth (AOD) and Weather Research Forecasting (WRF) model meteorological data for assessment of fine particulate matter (PM2.5) over India,” Atmospheric Pollution Research, vol. 10, no. 2, pp. 418-434, Mar. 2019.

[2] I. Stem, P. Builjies, M. Schaap, R. Timmermans, R. Vautard, A. Hodzic, M. Memmesheimer, H. Feldmann, E. Renner, R. Wolke, A. Kerschaubma, “A model inter-comparison study focussing on episodes with elevated PM10 concentrations,” Atmospheric Environment, vol. 42, no. 19, pp. 4567-4588, Jun. 2008.

[3] X. H. Cheng, Y. L. Liu, X. D. Xu, W. You, Z. L. Zang, L. N. Gao, Y. B. Chen, D. B. Su, P. Yan, “Lidar data assimilation method based on CRTM and WRF-Chem models and its application in PM2.5 forecasts in Beijing,” Science of The Total Environment, vol. 682, pp. 541-552, Sep. 2019.

[4] H. Wen, Y. Dang, L. Li, “Short-Term PM2.5 Concentration Prediction by Combining GNSS and Meteorological Factors,” IEEE Access, vol. 99, pp. 1-11, 2020.

[5] W. Sun, H. Zhang, A. Palazoglu, A. Singh, W. D. Zhang, S. W. Liu, “Prediction of 24-hour-average PM2.5 concentrations using a hidden Markov model with different emission distributions in Northern California,” Science of The Total Environment, vol. 443, pp. 93-103, Jan. 2013.

[6] L. F. Guo, B. Z. Chen, H. F. Zhang, G. Xu, L. J. Lu, X. F. Lin, Y. W. Kong, F. Wang, Y. P. Li, “Improving PM2.5 Forecasting and Emission Estimation Based on the Bayesian Optimization Method and the Coupled FMLLEX PART-WRF Model,” Atmosphere, vol. 9, no. 11, pp. 428, Nov. 2018.

[7] U. A. Bhatti, Y. D. Yan, M. Q. Zhou, S. Ali, A. Hussain, Q. S. Hoo, Z. Y. Yu, L. W. Yuan, “Time Series Analysis and Forecasting of Air Pollution Particulate Matter (PM2.5): An SARIMA and Factor Analysis Approach,” IEEE Access, vol. 9, pp. 41019-41031, 2021.

[8] S. Mahajan, L. J. Chen, T. C. Tsai, “Short-Term PM2.5 Forecasting Using Exponential Smoothing Method: A Comparative Analysis,” Sensors, vol. 18, no. 10, pp. 3223, Oct. 2018.

[9] P. Perez, C. Menares, “Forecasting of Hourly PM2.5 in South-West Zone in Santiago de Chile,” Aerosol and Air Quality Research, vol. 18, no. 10, pp. 6266-62679, Oct. 2018.

[10] M. H. Nguyen, P. Le Nguyen, K. Nguyen, V. A. Le, T. H. Nguyen, Y. Ji, “PM2.5 Prediction Using Genetic Algorithm-Based Feature Selection and Encoder-Decoder Model,” IEEE Access, vol. 9, pp. 57338-57350, 2021.

[11] M. Opreat, S. F. Mihalache, M. Popescu, “Computational Intelligence-based PM2.5 Air Pollution Forecasting,” International Journal of Computers Communications & Control, vol. 12, no. 3, pp. 365-380, Jun. 2017.

[12] X. H. Xu, W. J. Ren, “Application of a Hybrid Model Based on Echo State Network and Improved Particle Swarm Optimization in PM2.5 Concentration Forecasting: A Case Study of Beijing, China,” Sustainability, vol. 11, no. 11, p. 3096, Jun. 2019.

[13] P. J. Nieto, F. S. Lashezas, E. Garcia-Gonzalo, F. J. D. Juez, “PM10 concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM, MLP, VARIMA and ARIMA: A case study,” Science of The Total Environment, vol. 621, pp. 753-761, Apr. 2018.

[14] W. Sun, Y. J. Sun, “Daily PM2.5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm,” Journal of Environmental Management, vol. 188, pp. 144-152, Mar. 2017.

[15] I. A. Knopalo, M. Beckmann, F. Meulex, A. Dutot, G. Foret, “Combining deterministic and statistical approaches for PM10 forecasting in Europe,” Atmospheric Environment, vol. 43, no. 40, pp. 6425-6434, Dec. 2009.

[16] S. Mahajan, L. J. Chen, T. C. Tsai, “Short-Term PM2.5 Forecasting Using Exponential Smoothing Method: A Comparative Analysis,” Sensors, vol. 18, no. 10, p. 3223, Oct. 2018.

[17] C. J. Huang, P. H. Kao, “A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities,” Sensors, vol. 18, no. 7, p. 2220, Jul. 2018.

[18] D. Y. Wang, Y. L. Liu, H. Y. Luo, C. Q. Yue, S. Cheng, “Day-Ahead PM2.5 Concentration Forecasting Using WT-VMD Based Decomposition Method and Back Propagation Neural Network Improved by Differential Evolution,” International Journal of Environmental Research and Public Health, vol. 14, no. 7, p. 764, Jul. 2017.

[19] G. Y. Huang, X. Y. Li, B. Zhang, J. D. Ren, “PM2.5 concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition,” Science of The Total Environment, vol. 768, pp. 144516, May. 2021.

[20] X. B. Jin, N. X. Yang, X. Y. Wang, Y. T. Bai, T. L. Su, J. L. Kong, “Deep Hybrid Model Based on EMD with Classification by Frequency Characteristics for Long-Term Air Quality Prediction,” Mathematics, vol. 8, no. 2, p. 214, Feb. 2020.

[21] S. L. Zhu, X. Y. Lian, L. Wei, J. X. Che, X. P. Shen, L. Yang, X. L. Qu, X. N. Liu, W. L. Gao, X. W. Ren, J. S. Li, “PM2.5 forecasting using SVR with PSOGSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors,” Atmospheric Environment, vol. 183, pp. 20-32, Jun. 2018.

[22] H. Liu, H. P. Wu, X. W. Lv, Z. R. Ren, M. Liu, Y. F. Li, H. P. Shi, “An intelligent hybrid model for air pollutant concentrations forecasting: Case of Beijing in China,” Sustainable Cities and Society, vol. 47, p. 101471, May. 2019.

[23] S. A. A. Shah, W. Aziz, M. Almaraashi, M. S. A. Nadeem, N. Habib, S. O. Shim, “A hybrid model for forecasting of particulate matter concentrations based on multiscale characterization and machine learning techniques,” Mathematical Biosciences and Engineering, vol. 18, no. 3, pp. 1992-2009, 2021.

[24] Y. Bai, B. Zeng, C. Li, J. Zhang, “An ensemble long short-term memory neural network for hourly PM2.5 concentration forecasting,” Chemosphere, vol. 222, pp. 286-294, May. 2019.

[25] O. J. Escalona, W. D. Lynn, G. Peripinan, L. McFrederick, D. J. McEneaney, “Data-Driven ECG Denoising Techniques for Characterising Bipolar Lead Sets along the Left Arm in Wearable Long-Term Heart Rhythm Monitoring,” Electronics, vol. 6, no. 4, Dec. 2017.

[26] H. L. Song, H. B. Dong, Z. W. Yuan, J. Zhu, H. Y. Zhang, Y. J. Huang, “An EEMD-Based Electromagnetic Induction Method for Nondestructive Testing of Buried Metal Conductors,” IEEE Access, vol. 7, pp. 142261-142271, 2019.

[27] K. Dragomiretskiy, D. Zosso, “Variational Mode Decomposition,” IEEE Trans. Signal Processing, vol. 62, no. 3, pp. 531-544, Feb. 2014.

[28] H. J. Cui, Y. Guan, H. Y. Chen, “Rolling Element Fault Diagnosis Based on VMD and Sensitivity MCKD,” IEEE Access, vol. 9, pp. 120297-120308, 2021.

[29] J. S. Richman, D. E. Lake, J. R. Moorman, “Sample Entropy,” in Methods in Enzymology, vol. 384, Elsevier Science, 2004, pp. 172-184.

[30] X. Borjas, N. Lluberes, M. Noguera, A. Pez, “Sample entropy of ECG time series of fearful flyers: preliminary results,” Nonlinear Dynamics Psychology & Life Sciences, vol. 10, no. 3, pp. 301-318, Jul. 2006.

[31] V. E. Kosmidou, L. J. Hadjileontiadis, “Sign Language Recognition Using Intrinsic-Mode Sample Entropy on eEMG and Accelerometer Data,” IEEE Trans. Biomedical Engineering, vol. 56, no. 12, pp. 2879-2890, Dec. 2009.

[32] A. Subasi, M. Yilmaz, H. R. OrcaIl, “Classification of EMG signals using wavelet neural network,” Journal of Neuroscience Methods, vol. 156, no. 1-2, pp. 360-367, Sep. 2006.
HE WANG  He received a B.S degree in electrical engineering from the college of electrical and information engineering, Hunan University, Changsha, China, in 2009, and a Ph.D. degree in electrical engineering from the Wuhan University, Wuhan, China, in 2014. Since 2014, he has been a lecturer in the school of engineering, Beijing Forestry University, China. His special areas of interest are machine learning and signal processing.

HONGFENG CHEN  Received the Bachelor of Engineering from Qingdao University of Science and Technology in 2020, where he is currently pursuing a master’s degree in electronic information at Beijing Forestry University. His research direction is related to applications of data analysis and neural network technology. His research interests include the application of neural network technology in PM2.5 concentration prediction.