



Article

Day-Ahead PM_{2.5} Concentration Forecasting Using WT-VMD Based Decomposition Method and Back Propagation Neural Network Improved by Differential Evolution

Deyun Wang ^{1,2,*}, Yanling Liu ¹, Hongyuan Luo ¹, Chenqiang Yue ¹ and Sheng Cheng ^{1,2}

¹ School of Economics and Management, China University of Geosciences, Wuhan 430074, China; yanlingliu1223@163.com (Y.L.); lhyuan2016@163.com (H.L.); yuechenqiang@sina.cn (C.Y.); chengsheng@cug.edu.cn (S.C.)

² Mineral Resource Strategy and Policy Research Center, China University of Geosciences, Wuhan 430074, China

* Correspondence: wang.deyun@hotmail.com; Tel.: +86-027-6784-8527

Received: 11 May 2017; Accepted: 7 July 2017; Published: 12 July 2017

Abstract: Accurate PM_{2.5} concentration forecasting is crucial for protecting public health and atmospheric environment. However, the intermittent and unstable nature of PM_{2.5} concentration series makes its forecasting become a very difficult task. In order to improve the forecast accuracy of PM_{2.5} concentration, this paper proposes a hybrid model based on wavelet transform (WT), variational mode decomposition (VMD) and back propagation (BP) neural network optimized by differential evolution (DE) algorithm. Firstly, WT is employed to disassemble the PM_{2.5} concentration series into a number of subsets with different frequencies. Secondly, VMD is applied to decompose each subset into a set of variational modes (VMs). Thirdly, DE-BP model is utilized to forecast all the VMs. Fourthly, the forecast value of each subset is obtained through aggregating the forecast results of all the VMs obtained from VMD decomposition of this subset. Finally, the final forecast series of PM_{2.5} concentration is obtained by adding up the forecast values of all subsets. Two PM_{2.5} concentration series collected from Wuhan and Tianjin, respectively, located in China are used to test the effectiveness of the proposed model. The results demonstrate that the proposed model outperforms all the other considered models in this paper.

Keywords: PM_{2.5} concentration forecasting; wavelet transform; variational mode decomposition; differential evolution; back propagation neural network

1. Introduction

Over the past few decades, with the rapid development of industrialization and urbanization, the occurrence of haze pollution episodes has become more frequent and more severe in China [1,2]. According to the statistics of China's National Development and Reform Commission, since early 2013, many areas including the north China, Huanghuai, Jianghuai, Jiangnan, south of the Yangtze River and the north of southern China have suffered severe and continuous haze weather. Haze pollution brings serious adverse effects on the environment, climate, ecological systems, economy and public health, thus causes great harm to the human production and life on a global scale [3,4]. Even though the mechanism of haze formation is still not clear [5], the high level concentrations of fine particles with aerodynamic diameter of 2.5 μm or less (PM_{2.5}) was inferred as the main reason of haze pollution episodes, and thus attracted widespread public concerns [6,7]. Compared to the PM₁₀ (particulate matter with aerodynamic diameter below 10 μm), PM_{2.5} has smaller diameter and stronger adsorption capacity of hazardous materials such as heavy metal and organic pollutants, and therefore has more

serious adverse effects on human health and atmosphere quality [8]. With long-term exposure to PM_{2.5}, the incidence of associated diseases such as lung diseases, heart diseases and premature death in human beings will be significantly increased [9,10]. The severity of adverse effects caused by PM_{2.5} makes it an urgent need for researchers to simulate and forecast its concentration. However, due to the diverse sources, strong regional and long-range transport, multiple influencing factors, along with complex formation mechanisms, the previous studies have illustrated the difficulty in accurate prediction of PM_{2.5} concentration [11,12].

In the previous studies, many methods and tools have been proposed for predicting the concentration of PM_{2.5} or other air pollutants. These methods can be generally divided into the following three categories: (1) deterministic methods; (2) causal relationships forecasting methods; and (3) time series based forecasting methods. Deterministic methods such as community multi-scale air quality (CMAQ) model can simulate the process of discharge, accumulation, diffusion and transfer of a pollutant by employing meteorological, emission and chemistry models, and are often used for predicting the concentrations of air pollutants [13]. For instance, Djalalova et al. proposed a new post-processing method based on the CMAQ model for surface PM_{2.5} predictions [14]. However, the scale and quality of the emission data adopted make great influence on the forecast accuracy of deterministic methods [15]. Moreover, the deterministic methods need much computational time to accomplish the forecasting process [16]. Therefore, many researchers integrated the deterministic methods with intelligent or statistical methods in order to enhance the forecast accuracy. For instance, Konovalov et al. proposed a hybrid model for PM₁₀ concentration forecasting based on the deterministic model and statistical model, where the statistical model was utilized to correct predictions made by the deterministic model [17]. Song et al. developed an adaptive neuro-fuzzy model to implement deterministic forecasting of PM based on the data of hourly and 12 h averaged air pollutants within Yangtze River Delta region of China [18]. Causal relationships forecasting methods work on the causal relationships between independent variables and forecast values. For instance, Sun and Sun developed a novel hybrid model based on principal component analysis (PCA) and least squares support vector machine (LSSVM) optimized by cuckoo search (CS) for daily PM_{2.5} concentration prediction. In their proposed model, the PCA is firstly adopted to obtain the most important influencing factors of PM_{2.5} concentration, then based on the selected influencing factors, the PM_{2.5} concentration is predicted using CS-LSSVM model [19]. However, as mentioned before, since the PM_{2.5} concentration is influenced by many complicated factors such as temperature, wind speed, precipitation which are very difficult to be determined accurately and quantificationally in real-world applications, thus, it is really a big challenge to build an accurate forecasting model using causal relationships forecasting method. Therefore, many researchers turn to forecasting PM_{2.5} concentration based upon the time series forecasting methods which rely on the historical series.

The most frequently used time series forecasting methods in this study area can be further divided into the following two categories: statistical models and hybrid models. Statistical models are easy, quick and efficient, and mainly involve autoregressive integrated moving average model (ARIMA), multiple linear regression (MLR), generalized autoregressive conditional heteroskedasticity (GARCH), grey model (GM), Markov models and artificial intelligence (AI) based models. For example, Jian et al. utilized ARIMA model to investigate the effect of meteorological factors on submicron particle concentrations under busy traffic conditions [20]. Stadlober et al. developed a MLR model combining information of the present day with meteorological forecasts of the next day to forecast daily PM₁₀ concentrations for sites located in Bolzano, Klagenfurt and Graz [21]. Kumar and Ridder developed a hybrid model integrated GARCH modeling technique and FFT-ARIMA to forecast daily maximum O₃ concentration, and to make probabilistic forecasts of ozone episodes at four urban sites of two major European cities (London and Brussels); the results revealed the good forecasting performance of the proposed model [22]. Pai et al. proposed seven types of first-order and one-variable grey differential equation model to predict hourly PM concentrations in Banciao city of Taiwan, the results showed that GM(1,1) is an efficient early-warning tool for providing PM information to the

inhabitants [23]. Sun et al. utilized hidden Markov models (HMMs) to forecast daily average PM_{2.5} concentration, the comparisons between different distributions used in HMMs showed that the closer the distribution employed in HMMs is to the observation sequence, the better the model prediction performance [24]. Ordieres et al. compared the forecasting capability of three different neural networks for PM_{2.5} concentration forecasting, including multilayer perceptron (MLP), radial basis function (RBF) and square multilayer perceptron (SMLP) [25]. Voukantsis et al. constructed an ANN (multi-layer perceptron) model to accomplish the forecasting of daily PM₁₀ and PM_{2.5} concentrations [26].

In order to improve the forecast accuracy, hybrid models which combine some single models have been widely used in air quality forecasting field. For example, Lin et al. forecasted concentrations of air pollutants including PM₁₀, NO_x and NO₂ by logarithm SVM with immune algorithms (IA) [27]. Perez developed a combination model of ANN and a nearest neighbor method to conduct PM₁₀ forecasting [28]. Antanasijević et al. applied the artificial intelligence model combining the genetic algorithm (GA) and ANN to forecast PM₁₀ emission [29]. However, the multiple frequency components existed in PM_{2.5} concentration series are always the challenging parts in forecasting, making the models which work on the original time series cannot handle them appropriately. Thus, many decomposition techniques such as WT, EMD, VMD and their variants have been applied for dealing with the multiple frequency components before forecasting. For instance, Bai et al. utilized wavelet transform (WT) technique and back propagation neural network (BPNN) to forecast daily air pollutants (PM₁₀, SO₂ and NO₂) concentrations, and experiment results showed that the WT-BPNN model has better forecasting performance than BPNN model in terms of the forecast accuracy [30]. Zhou et al. developed a hybrid model based on EEMD and a general regression neural network (GRNN) to predict PM_{2.5} concentration, and obtained accurate results [2]. Liu et al. presented four different hybrid models by combining four signal decomposing algorithms (e.g., wavelet decomposition (WD)/wavelet packet decomposition (WPD)/EMD/fast ensemble empirical mode decomposition (FEEMD)) and ELM model to complete the multi-step-ahead wind speed forecasting, and the experiments indicated that all the proposed hybrid models have better performance than the single ELM model [31]. Wang et al. [32] proposed a hybrid model based on two-layer decomposition method and BP neural network optimized by firefly algorithm for multi-step electricity price forecasting, and the experimental results illustrated the superior performance of the proposed model.

Although the hybrid models integrated single decomposition techniques can improve the forecasting ability to some extent, while since all the single decomposition techniques have the drawback of mode mixing with different levels, which makes the multiple frequency components existed in the PM_{2.5} concentration series cannot be effectively extracted, and consequently leads to an inferior forecasting performance. Therefore, this paper proposes a novel hybrid model based on WT-VMD decomposition method and BP neural network optimized by DE algorithm for one day ahead PM_{2.5} concentration forecasting. Firstly, WT is employed to decompose the PM_{2.5} concentration series into a number of subsets with different frequencies. Since the problem of frequency aliasing exists between the subsets may increase the forecasting difficulty, therefore, VMD is further applied to conduct the secondary decomposition of each subset generated by WT, and a set of modes are obtained. Next, BP model optimized by DE is utilized to forecast the modes generated by VMD. Finally, the forecast value of each subset is obtained through aggregating the forecast values of all modes generated from this subset by VMD, and the forecast series of PM_{2.5} concentration is obtained by adding up the forecast values of all subsets. The proposed model is tested using two PM_{2.5} concentration series collected respectively from Wuhan and Tianjin located in China. The results demonstrate that the proposed model outperforms all the other considered models in this paper.

The rest of this paper is organized as follows. Section 2 describes in detail the methodology adopted in the paper. Section 3 develops the proposed hybrid model. In Section 4, two PM_{2.5} concentration series collected from Wuhan and Tianjin, respectively, located in China are taken for conducting the empirical study. Section 4 is the concluding remarks.

2. Methodology

2.1. Wavelet Transform (WT)

WT is a powerful technique for processing the non-periodic, non-stationary and transient signals [33]. WT decomposes a time series into different components at different frequency levels: one low frequency approximation subset which shows the general trend of the signal and several high frequency detail subsets which are related to the noise and disturbance [34]. As shown in Figure 1, the process of m -level decomposition by WT for time series $S(t)$ can be defined as follows:

$$S(t) = d_1 + d_2 + \dots + d_m + a_m \tag{1}$$

Compared to the original signal, these subsets generated by WT usually have some better behaviors such as more stable variance and fewer outliers, which facilitates the prediction task and therefore improves the overall prediction accuracy [35]. WTs can be divided into the following two categories: (1) continuous wavelet transform (CWT); and (2) discrete wavelet transform (DWT).

The CWT of a signal $f(t)$ is defined as follows:

$$CWT_f(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \psi^* \left(\frac{t - b}{a} \right) dt \tag{2}$$

where a and b are the parameters of scale and translation, respectively; $*$ represents the complex conjugate; and the mother wavelet $\psi(t)$ denotes the transforming function.

The DWT of a signal $f(t)$ is defined as follows:

$$DWT_f(m, n) = 2^{-\frac{m}{2}} \sum_{t=0}^{T-1} f(t) \psi^* \left(\frac{t - n2^m}{2^m} \right) \tag{3}$$

where the integer m is the scale factor (decomposition level), the integer n is the sampling time, T is the length of signal $f(t)$, and t is the discrete time index.

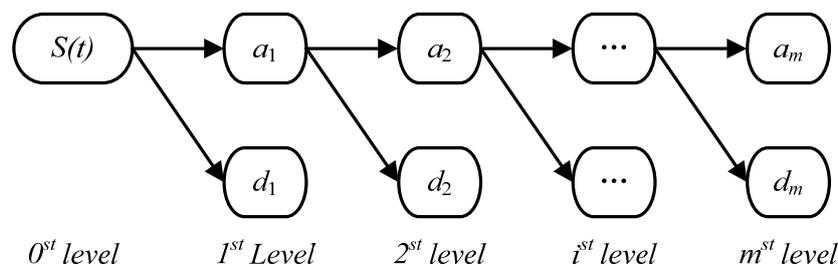


Figure 1. The process of m -level decomposition by WT for time series $S(t)$.

The number of decomposition levels and selection of the mother wavelet have considerable effects on the characteristics of subsets and thus influence significantly the overall prediction error. In the decomposition process of WT, more levels will result in more stationary subsets; however, large number of levels might cause decomposition information loss and thus low prediction accuracy [34]. Based on the above considerations, this paper adopts a three-level DWT with mother wavelet of the Daubechies wavelet of order 4 (Db4), which has the ability of providing a balance between wavelength and smoothness [35].

2.2. Variational Mode Decomposition (VMD)

VMD is an effective signal decomposition method proposed by Dragomiretskiy and Zosso in 2014 [36]. VMD can decompose a real valued signal into a discrete set of band-limited modes (denoted

by y_k) which have specific sparsity properties when producing main signal. It is assumed that each mode y_k generated by VMD can be compressed around a center pulsation ω_k which is determined along with the decomposition process. In order to obtain the bandwidth of each mode, the following procedures should be accomplished: (1) for each mode μ_k , compute the associated analytic signal with the benefit of Hilbert transform to obtain a unilateral frequency spectrum; (2) mix with an exponential tuned to the respective estimated center frequency in order to shift the mode's frequency spectrum to baseband; and (3) estimate the bandwidth of each model through Gaussian smoothness of the demodulated signal. Then, the constrained variational problem can be provided as follows:

$$\min_{\mu_k, \omega_k} = \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * \mu_k(t) \right] e^{-j\omega_k t} \right\|_2 \right\} \tag{4}$$

Subject to

$$\sum_k \mu_k = f \tag{5}$$

where f is the original signal, μ is its mode, ω is the frequency, δ is the Dirac distribution, t is time script, k is the number of modes, and $*$ denotes convolution. Recall that, in the VMD framework, the original signal f is decomposed into a set of modes denoted μ (see Equation (5)) each having a bandwidth in Fourier domain (see Equation (4)) and compacted around a center pulsation ω_k . The solution to the original minimization problem (see Equation (4)) is the saddle point of the following augmented Lagrangian (L) expression:

$$L(\mu_k, \omega_k, \lambda) = \alpha \sum_k \left\| \partial_t \left[\delta(t) + \frac{j}{\pi t} * \mu_k(t) \right] \right\|_2^2 + \|f - \sum \mu_k\|_2^2 + \langle \lambda, f - \sum \mu_k \rangle \tag{6}$$

where λ is the Lagrange multiplier and α represents the balancing parameter of the data-fidelity constraint. Consequently, the solutions for u and ω can be obtained based on the following two equations:

$$\mu_n^{n+1} = \left(f - \sum_{i \neq k} \mu_i + \frac{\lambda}{2} \right) \frac{1}{1 + 2\alpha(\omega - \omega_k)^2} \tag{7}$$

$$\omega_n^{n+1} = \frac{\int_0^\infty \omega |\mu_k(\omega)|^2 d\omega}{\int_0^\infty |\mu_k(\omega)|^2 d\omega} \tag{8}$$

where n is the number of iterations.

2.3. The DE-BP Model

2.3.1. Back Propagation (BP) Neural Network

Artificial neural networks (ANNs) include a family of intelligent models that mimic the biological neural networks. The BP neural network including one or more hidden layers is one of the ANN models, which has a relative simple structure and thus can be realized easily. Since the distinguish performance of the BP neural network, it has been popularly used in many practical fields such as wind speed forecasting [37], plastic injection molding [38], natural gas load forecasting [39] and so forth. The BP neural network used in this study has a three-layer network consisting of an input layer, a hidden layer, and an output layer (see Figure 2). The BP neural network distinguishes itself by the presence of hidden layers whose computation nodes are correspondingly called hidden neurons. The function of hidden neurons is to connect the input and the network output. Given a training set of input-output data, the most common learning rule for multi-layer perceptron (MLP) neural networks is the back-propagation algorithm which involves two following phases: the first one is a feed-forward phase in which the external input information at the input nodes is propagated forward

to compute the output information signal at the output unit; the second one is a backward phase in which modifications to the connection signals are made based on the differences between the computed and observed information signals at the output units. In this study, a tangent sigmoid function is used as the neuron transfer function.

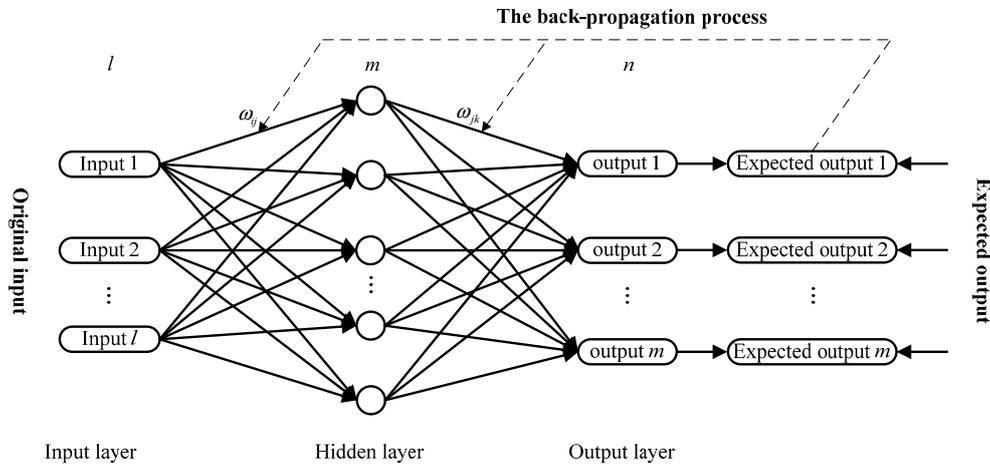


Figure 2. The basic structure of a three-layer BP neural network.

After determination of the network topology and initialization of the associated network parameters, the BP neural network should be trained and tested through the following three steps:

Step 1: Calculate the output of the *j*th node in the hidden layer using the following equation.

$$H_j = f\left(\sum_{i=1}^n \omega_{ij}x_i - a_j\right) \tag{9}$$

where *i* is the index of neuron in the input layer, *n* is the number of neurons in the input layer, ω_{ij} is the connection weights between input layer and hidden layer, x_i is the *i*th input value, a_j is threshold value, and H_j and f represent the output of hidden layer and the incentive function of neurons, respectively.

Step 2: Calculate the fitted value or forecasting value of the *k*th node in the output layer using the following equation.

$$O_k = \sum_{j=1}^l H_j \omega_{jk} - b_k \quad k = 1, 2, \dots, m \tag{10}$$

where ω_{jk} is the connection weights between the output layer and hidden layer, b_k is threshold value, and *m* is the number of neurons in the output layer.

Step 3: Calculate the fitting error e_k based on the fitted value and expected output, and update the weight factor and threshold value by the following formula.

$$\omega_{ij} = \omega_{ij} + \eta H_j (1 - H_j) x(i) \sum_{k=1}^m \omega_{jk} e_k \quad i = 1, 2, \dots, n; j = 1, 2, \dots, l \tag{11}$$

$$\omega_{jk} = \omega_{jk} + \eta H_j e_k \quad j = 1, 2, \dots, l; k = 1, 2, \dots, m \tag{12}$$

$$a_j = a_j + \eta H_j (1 - H_j) x(i) \sum_{k=1}^m \omega_{jk} e_k \quad j = 1, 2, \dots, l \tag{13}$$

$$b_k = b_k + e_k \quad k = 1, 2, \dots, m \tag{14}$$

where η denotes the learning rate, $x(i)$ is the *i*th input value.

The training process of the BP neural network is stopped when one of the following two conditions is satisfied: (1) the maximum number of iterations is reached; and (2) the fitting accuracy meets the requirement.

2.3.2. Differential Evolution (DE) Algorithm

DE algorithm proposed by Storn and Price in 1997 is a stochastic, population-based and direct search algorithm which has the characteristics of simple structure, less control parameters, fast convergence, and strong robustness, and therefore has significant advantages for dealing with the non differentiable, nonlinear and multimodal functions [40]. As shown in Figure 3, the standard DE algorithm consists of four main operations: initialization, mutation, crossover and selection.

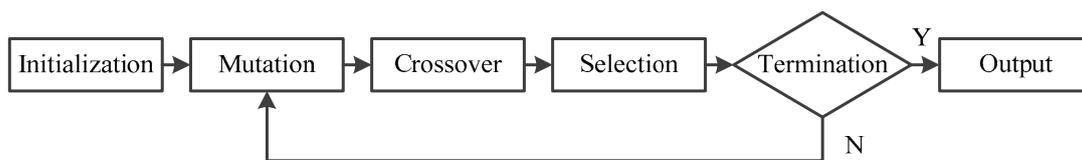


Figure 3. The flowchart of DE algorithm.

The basic steps of DE algorithm are illustrated as follows:

Step 1: Population initialization. Initializing population of DE algorithm based on the following formula.

$$x_{j,i,0} = x_{j,\min} + rand_{i,j}(0,1) \times (x_{j,\max} - x_{j,\min}) \tag{15}$$

where $x_{j,i,0}$ denotes the value of i th individual in the 0 th generation and j th dimension.

Step 2: Mutation. Based on the randomly selected three indices, m, i and $j, m \neq i \neq j$, a mutant vector $V_{k,G}$ is generated based on the following formula.

$$V_{k,G} = X_{m,G} + F(X_{i,G} - X_{j,G}) \tag{16}$$

where $k \neq m \neq i \neq j, F$ is a scaling factor and $F \in [0, 2], X_{m,G}$ is the base vector.

Step 3: Crossover. Crossover operation is introduced into DE algorithm in order to improve the multiplicity of the perturbed parameter vectors. The trial point $U_{j,k,G+1}$ is established from its parents $V_{j,k,G+1}$ and $X_{j,k,G}$ by the following formula.

$$U_{j,k,G+1} = \begin{cases} V_{j,k,G+1} & \text{if } (rand\ b(j) \leq C_R) \text{ or } j = rnbr(i) \\ X_{j,k,G} & \text{if } (rand\ b(j) > C_R \text{ and } j \neq rnbr(i)) \end{cases} \quad j = 1, \dots, D, \tag{17}$$

where C_R is crossover probability and $C_R \in [0, 1], rnbr(i)$ is a randomly selected index in the set of $\{1, 2, 3, \dots, D\}$, which ensures that $U_{j,k,G+1}$ obtains at least one parameter from $V_{j,k,G+1}$. The trial vector is formed of both current parameter vectors and mutant vector parameters (see formula (17)).

Step 4: Selection. The trial vector $X_{i,G+1}$ can be obtained by comparing the fitness value of the vector obtained through mutation and crossover, and the process can be denoted as follows:

$$X_{i,G+1} = \begin{cases} U_{i,G} & \text{if } f(U_{i,G}) \leq f(X_{i,G}) \\ X_{i,G} & \text{otherwise} \end{cases} \tag{18}$$

Step 5: Iterative computing and stop the DE algorithm if the result satisfies the error requirement or the maximum number of iterations is reached. Otherwise, return to Step 2.

2.3.3. The DE-BP Model

In the BP neural network, the two kinds of training parameters of weight matrices (ω_{ij} and ω_{jk}) and thresholds (a_j and b_k) have significant influences on the prediction accuracy. In order to improve the function approximation ability of the BP neural network, especially on the catastrophe points, in this study, DE algorithm is utilized to optimize the weight matrices and thresholds, see Figure 4. The fitness function of DE algorithm used in this study is the RMSE of forecast results, and is defined as follows:

$$F_{fitness} = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{X}(t) - X(t))^2} \quad (19)$$

where $\hat{X}(t)$ denotes the forecast value at time t , $X(t)$ represents the actual value at time t , and N is the total number of data. The individual owning the minimal fitness value is the global best point, which can be used to determine the parameters of the BP neural network.

The steps of DE-BP model are described as follows:

- Step 1:** Initialization. Determine the network topology of the network and initialize the parameters of DE algorithm including population size, maximum iteration number, probabilities of mutation and crossover operators. The initial population is generated using Equation (15).
- Step 2:** Calculate the fitness value of each individual using Equation (19). The DE algorithm is stopped when the stop criterion is satisfied, and go to Step 4.
- Step 3:** Update the population of DE algorithm based on mutation, crossover and selection operators. Go to Step 2.
- Step 4:** The optimal individual obtained from DE algorithm is adopted as the initial connection weights and thresholds of the BP neural network.
- Step 5:** Train and test the BP neural network based on the training and testing samples.

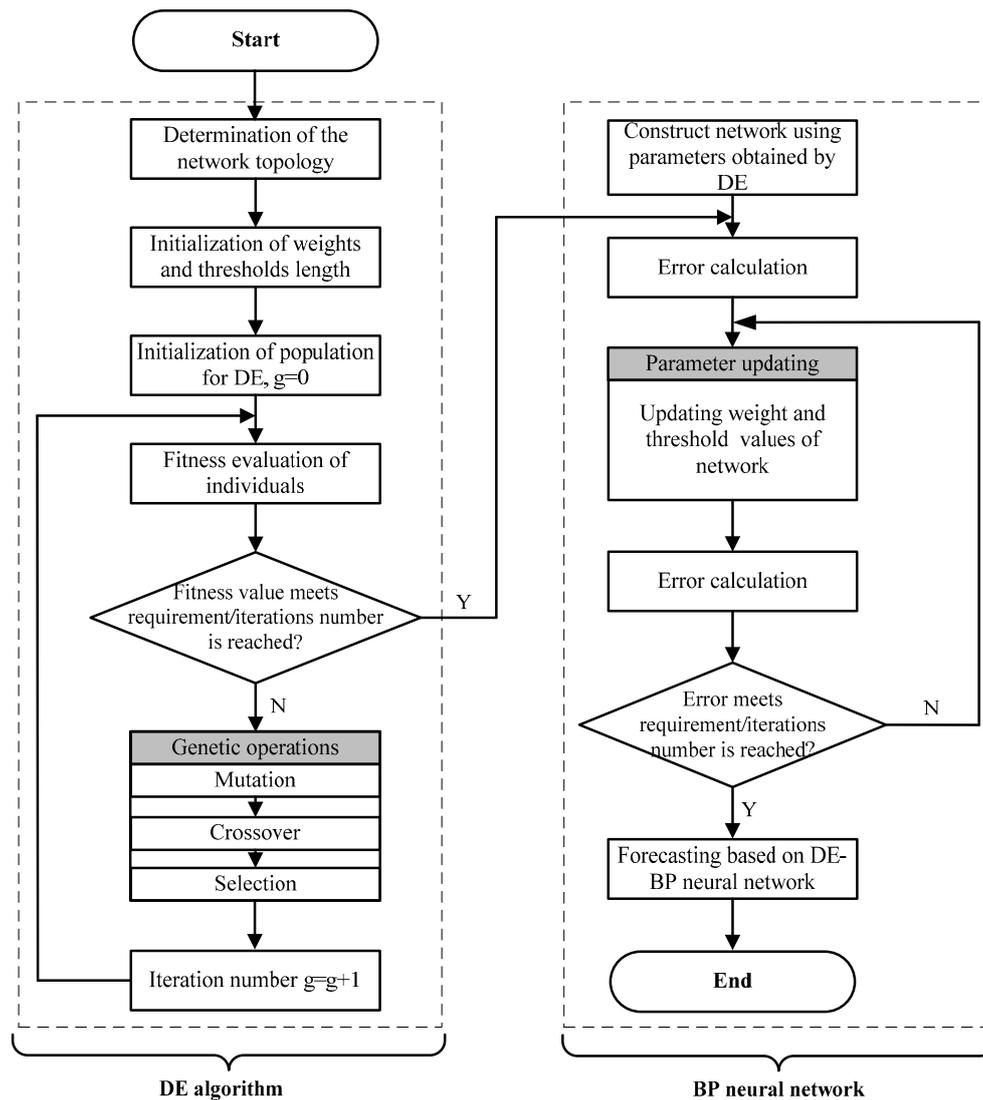


Figure 4. The flowchart of DE-BP model.

2.3.4. Hybrid WT-VMD-DE-BP Forecasting Model

In this section, the proposed WT-VMD-DE-BP model is established for daily PM_{2.5} concentration forecasting. As shown in Figure 5, the basic structure of the hybrid forecasting method includes the following five steps:

- Step 1:** First decomposition. The WT decomposition technique is utilized to decompose the PM_{2.5} concentration series into one low frequency approximation subset and several high frequency detail subsets.
- Step 2:** Second decomposition. In order to increase the forecasting accuracy, the VMD technique is further employed to conduct the secondary decomposition of each subset generated by WT, and consequently a number of VMs are obtained.
- Step 3:** Individual forecasting. Each VM generated by VMD is forecasted using DE-BP model.
- Step 4:** First summation. The forecast value of each subset generated by WT is obtained by adding up all the forecast values of VMs generated by VMD decomposition of this subset.
- Step 5:** Second summation. The forecast series of PM_{2.5} concentration is obtained by aggregating the forecast result of each subset.

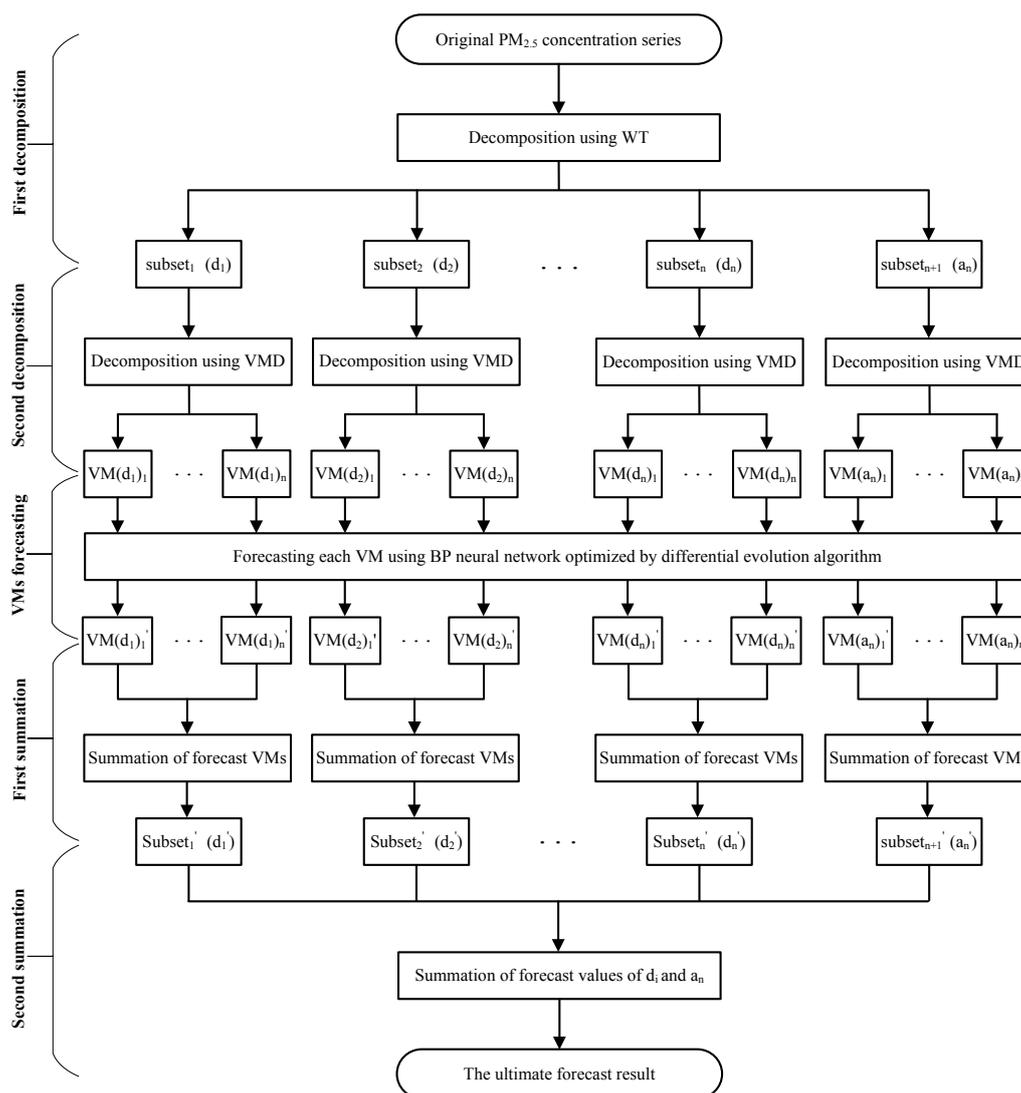


Figure 5. The flowchart of WT-VMD-DE-BP model.

3. Empirical Study

3.1. Study Area and Data Description

In this paper, two PM_{2.5} concentration series respectively collected from Wuhan and Tianjin located in China are adopted for testing the validity of the proposed model. The specific locations of the two study areas are briefly depicted in Figure 6. Wuhan, situated in the middle-lower Yangtze Plain and the eastern part of Jiangnan Plain (30° N and 114° E), has been regarded as China’s important industrial base, integrated transportation hub, and science and education base. The Yangtze River, which is the third longest river in the world, and the largest tributary of the Han River meet at this city, making Wuhan become a very important inland river port. Wuhan has a sub-tropical monsoon humid climate with abundant rainfall, abundant sunshine and four distinct seasons. Tianjin, the largest coastal city in northern China, is located along the west coast of the Bohai Gulf (39° N and 117° E). Tianjin has become a new growth pole and a hub of advanced industry and financial activity in China. Tianjin has a sub-humid warm temperate monsoon climate that characterized by significant winds and four distinct seasons. With rapid development of urbanization in the past several decades, both Wuhan and Tianjin become two megalopolises with a population of more than ten million people. Simultaneously, because of the development of industrialization and increase of motor vehicles, the

occurrence of haze weather in these two megalopolises becomes more frequent and more severe, which makes it an urgent need for researchers and relevant government departments to simulate and forecast the $PM_{2.5}$ concentration in order to protect public health and atmospheric environment.

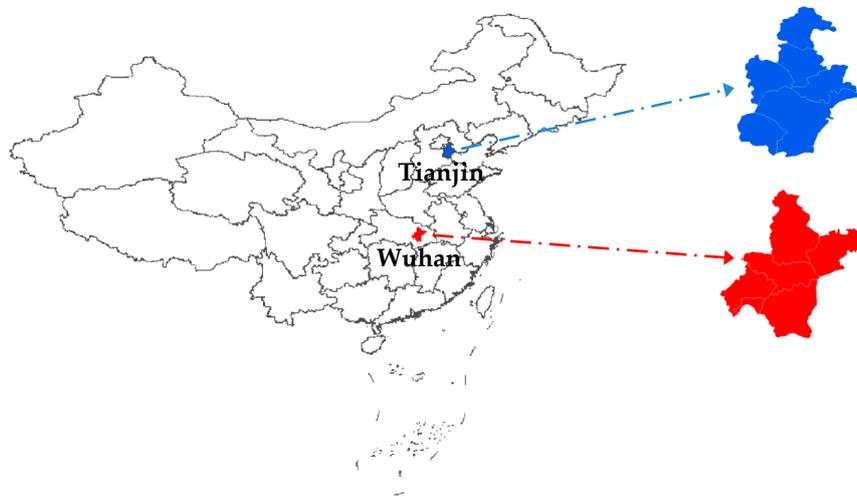


Figure 6. Geographical locations of the study areas.

In this paper, the two original daily $PM_{2.5}$ concentration series from 1 January 2014 to 30 June 2016 with a total of 912 observations in Wuhan and Tianjin are collected from China's online air quality monitoring and analysis platform (<http://www.aqistudy.cn/>), as shown in Figure 7. In Figure 7, it can be seen that the two $PM_{2.5}$ concentration series share some common features, for example, both Wuhan and Tianjin have high level of $PM_{2.5}$ concentration in winter (roughly between November and February of each year). However, since Wuhan and Tianjin have different geographical positions, climatic characteristics and industrial structures, the $PM_{2.5}$ concentration series associated with the two megalopolises appear to be different, and therefore are suitable for testing the effectiveness and practicability of the proposed forecasting model. Specifically, in each $PM_{2.5}$ concentration series, the 1st–882nd data (1 January 2014–31 May 2016) and 883rd–912th data (1 June 2016–30 June 2016) are adopted as the training and testing samples, respectively. This study selects four accuracy measures presented in Section 3.2 in order to evaluate the performance of the proposed forecasting model. In addition, it should be noted that all considered models adopted in this paper are coded in MATLAB R2010a.

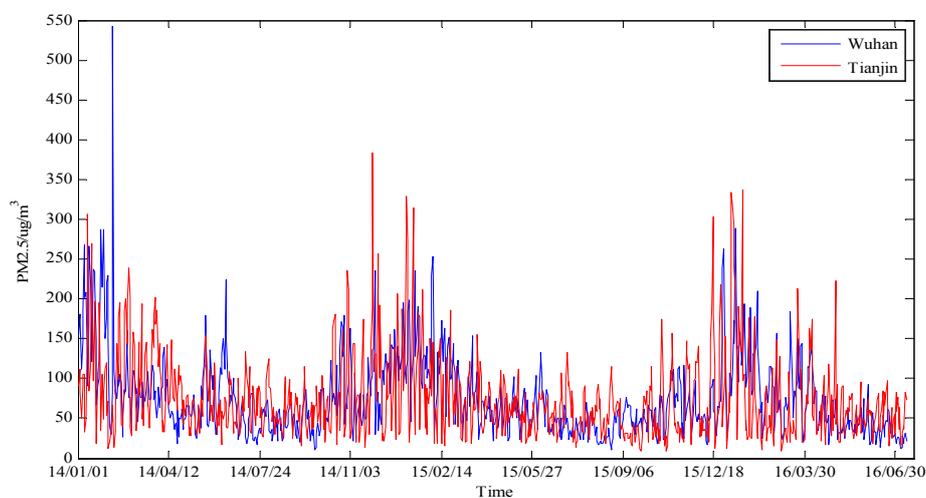


Figure 7. The original $PM_{2.5}$ concentration series of Wuhan and Tianjin.

3.2. Performance Criteria of Forecasting Accuracy

This study adopts the following four error metrics to testify the effectiveness and practicability of the proposed forecasting model: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE) and Theil's inequality coefficient (TIC). The performance measures of MAE, RMSE and MAPE are utilized to quantify the errors of forecast values, and the smaller they are, the better the prediction accuracy is. TIC is employed to evaluate the predictive capability of different forecasting models, and the smaller it is, the better the forecasting capability that the model has.

The computational formulas of these four performance measures are provided as follows:

$$MAE = \frac{1}{n} \sum_{t=1}^n |\hat{x}(t) - x(t)| \quad (20)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{x}(t) - x(t))^2} \quad (21)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{x}(t) - x(t)}{x(t)} \right| \quad (22)$$

$$TIC = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{x}(t) - x(t))^2}}{\sqrt{\frac{1}{n} \sum_{t=1}^n x(t)^2 + \frac{1}{n} \sum_{t=1}^n \hat{x}(t)^2}} \quad (23)$$

where n is the number of observed $PM_{2.5}$ concentration values; and $\hat{x}(t)$ and $x(t)$ are the forecast and observed values of $PM_{2.5}$ concentration at time t , respectively.

3.3. $PM_{2.5}$ Concentration Forecasting in Wuhan

3.3.1. Analysis of Decomposition Results

The multiple frequency components in $PM_{2.5}$ concentration series are always the challenging parts in forecasting, making the models which work on the original time series cannot handle them appropriately. In order to improve the forecast accuracy, in this study, WT is firstly employed to divide the $PM_{2.5}$ concentration series collected from Wuhan into four components including one low frequency approximation subset and three high frequency detail subsets (see Figure 8). The four components are denoted respectively as d_1 , d_2 , d_3 and a_3 , where a_3 is the low frequency approximation subset which illustrates the general trend of the $PM_{2.5}$ concentration series, and d_1 , d_2 , and d_3 indicate the high frequency detail subsets, which are related to the noise and disturbance.

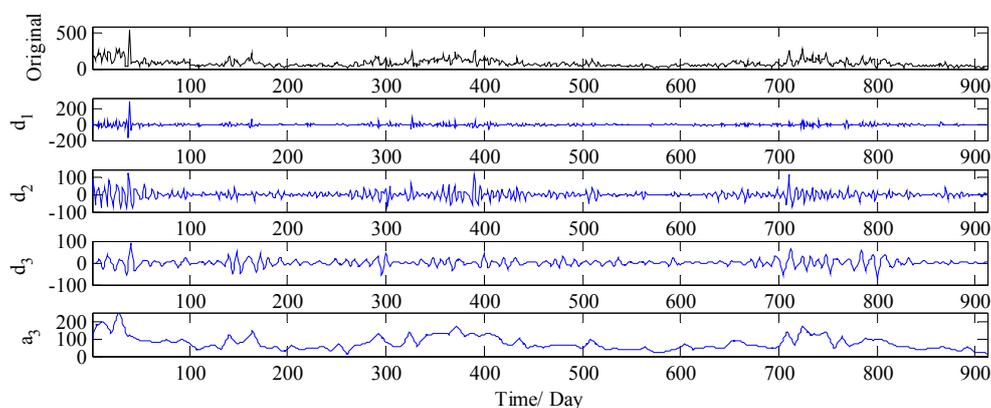


Figure 8. Decomposition results of $PM_{2.5}$ concentration series by WT (Wuhan).

After decomposition of original $PM_{2.5}$ concentration series by WT, DE-BP model is utilized to forecast all the subsets. There is no doubt that the $PM_{2.5}$ concentrations of previous several days have a great influence on the latter ones. Therefore, in this study, a certain number of previous $PM_{2.5}$ concentration data are taken as the input of DE-BP model for forecasting the latter one. After several simulations and predictions, the optimal length of predicted series is set as eight in DE-BP model in order to obtain the higher accuracy. In DE algorithm, the parameter settings are listed as follows: population size: $N_p = 100$, scaling factor: $F = 0.5$, crossover probability: $C_R = 0.5$, max iterations: $G_M = 100$. The above parameter settings and input determination method are used in all tests throughout the paper in order to ensure fair and valid comparisons between different forecasting models.

Based on the above parameter settings, each subset is forecasted using DE-BP model, and the forecast results are illustrated in Figure 9. It is obvious that all the four subsets cannot be forecasted with high accuracy, especially the subsets d_1 , d_2 , and d_3 , which are related to the noise and disturbance. Therefore, it can be concluded that the single decomposition process by WT cannot effectively extract the multiple frequency components existed in the $PM_{2.5}$ concentration series, and therefore leads to a relatively inferior forecasting performance.

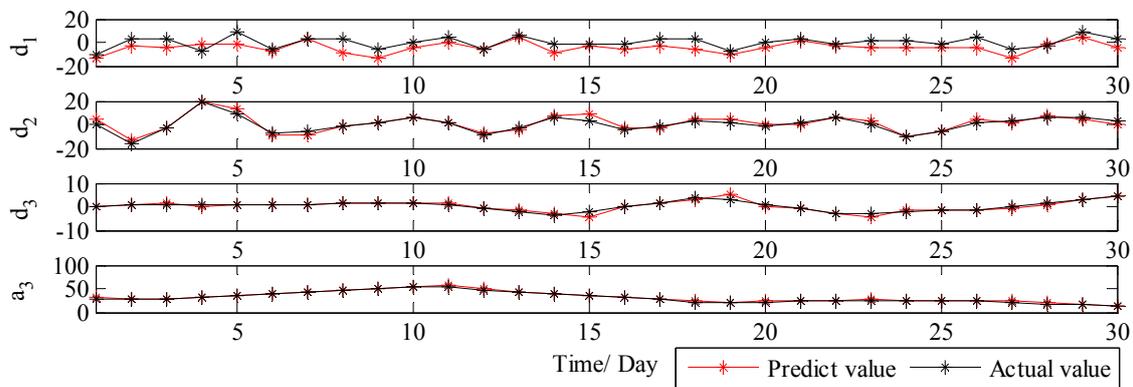


Figure 9. The forecast results of each subset by WT-DE-BP model (Wuhan).

Therefore, in order to solve the drawback of WT and further improve the forecast accuracy, VMD is further applied to conduct the secondary decomposition of each subset generated by WT. In this study, each subset is decomposed into eight VMs, and the decomposition results of d_1 , d_2 , d_3 and a_3 by VMD are illustrated in Figure 10. Then, DE-BP model is employed to forecast all VMs based on the rolling technology. Next, the forecasting value of each subset is obtained by aggregation of the forecast values of all the VMs generated by VMD decomposition of this subset. Finally, the ultimate forecast result of $PM_{2.5}$ concentration series can be obtained by adding up the forecast values of each subset. The ultimate forecast results and the corresponding MAE, RMSE, MAPE and TIC of WT-VMD-DE-BP model for all the four subsets are presented in Figure 11 and Table 1. From Figure 11 and Table 1, it is obvious that after secondary decomposition by VMD, the forecast accuracies of all the four subsets are significantly improved, which confirms the effectiveness of the hybrid decomposition technology proposed in this study.

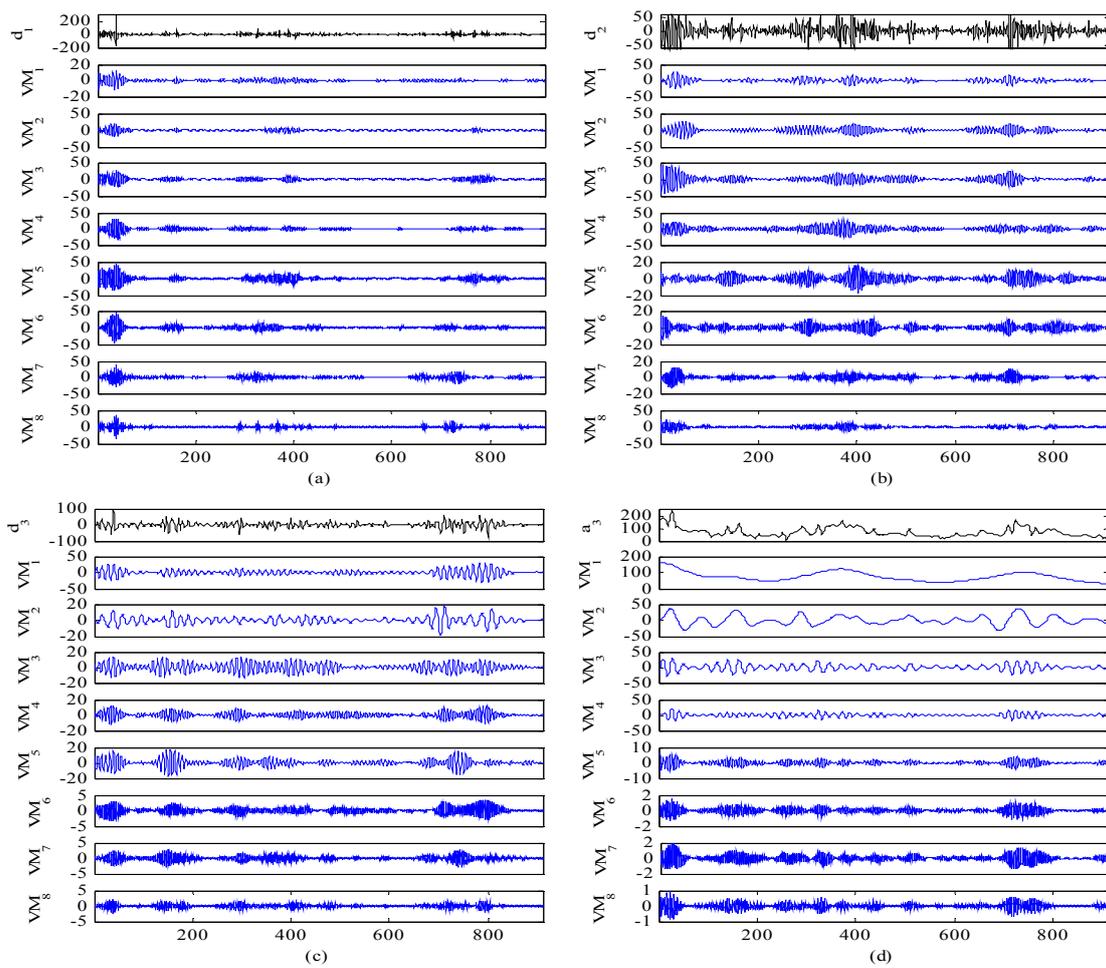


Figure 10. Decomposition results of each subset by VMD (Wuhan).

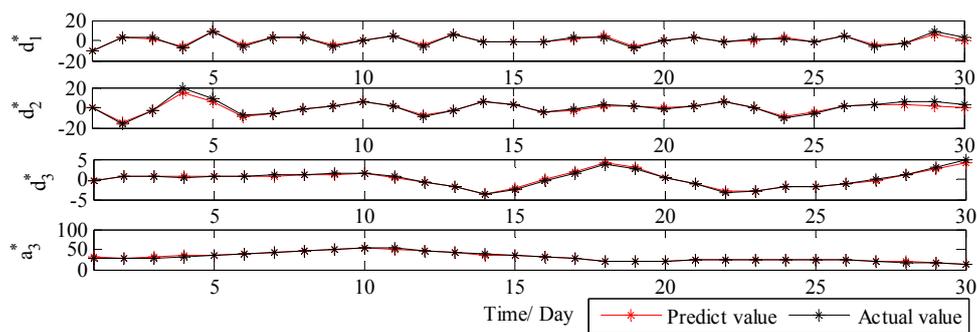


Figure 11. The forecast results of each subset using WT-VMD-DE-BP (Wuhan).

Table 1. Forecast errors of WT-DE-BP and WT-VMD-DE-BP for each subset.

Index	Forecast Errors of WT-DE-BP				Forecast Errors of WT-VMD-DE-BP			
	d_1	d_2	d_3	a_3	d_1	d_2	d_3	a_3
MAE	4.97	1.83	0.45	1.38	0.85	1.08	0.17	0.95
RMSE	5.79	2.31	0.75	1.89	1.05	1.58	0.21	1.20
MAPE (%)	222.38	357.01	56.56	5.42	33.75	36.04	22.83	3.80
TIC	0.52	0.17	0.18	0.03	0.12	0.13	0.05	0.02

3.3.2. Results and Discussions

In this section, to verify the superiority of the proposed WT-VMD-DE-BP model in forecasting capability, forecasting models of BP, DE-BP, WT-DE-BP, VMD-DE-BP and WT-VMD-DE-BP are adopted as the benchmark models. Four error measurements, MAE, RMSE, MAPE and TIC, are employed for evaluating the performance of all the forecasting models.

The forecast results of all considered models are shown in Figure 12, and the forecast errors including MAE, RMSE, MAPE and TIC of the proposed model and benchmark models are presented in Table 2 where the smallest value of each row is marked in boldface. As shown in Table 2, the error values of MAE, RMSE, MAPE and TIC of the proposed model are all smallest compared with all the other benchmark models, which confirms that the proposed hybrid WT-VMD-DE-BP model based on WT-VMD decomposition technique has the best forecasting performance. In order to present the comparison more intuitively, the error figures MAE, RMSE, MAPE and TIC of different models are also provided in Figure 13.

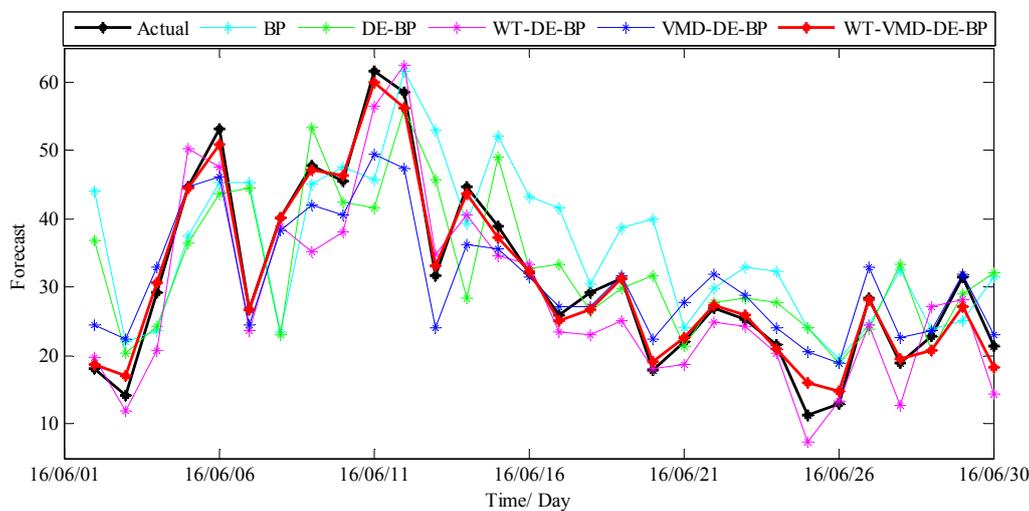


Figure 12. The ultimate forecast result of different models (Wuhan).

Table 2. Comparison of prediction performances of different models (Wuhan).

Index	BP	DE-BP	WT-DE-BP	VMD-DE-BP	WT-VMD-DE-BP
MAE	9.61	8.12	4.05	4.53	1.34
RMSE	11.68	10.08	4.83	5.54	1.79
MAPE (%)	39.50	31.94	13.84	17.88	5.95
TIC	0.16	0.14	0.07	0.08	0.03

Note: The smallest value of each row is marked in boldface.

In order to further analyze the effects of the decomposition technique and DE optimization algorithm on the proposed model, the following three categories of comparisons are conducted in this experiment. The first category of comparison (Comparison I), which is designed for testing the positive effects of single decomposition techniques, is conducted between the forecasting models embedded with single decomposition techniques (WT-DE-BP model and VMD-DE-BP model) and the forecasting models without any decomposition techniques (DE-BP model). The second category of comparison (Comparison II), which is designed for proving the advantages of hybrid WT-VMD decomposition technique, is carried out between the proposed model and the forecasting models embedded with single decomposition techniques (WT-DE-BP model and VMD-DE-BP model). The third category of comparison (Comparison III), which is designed for confirming the contribution of DE algorithm on the BP model, is conducted between the DE-BP model and BP model. The comparison results of

Comparison I, Comparison II and Comparison III are all presented in Table 3. Based on the results listed in Table 3, the following three categories of findings can be obtained.

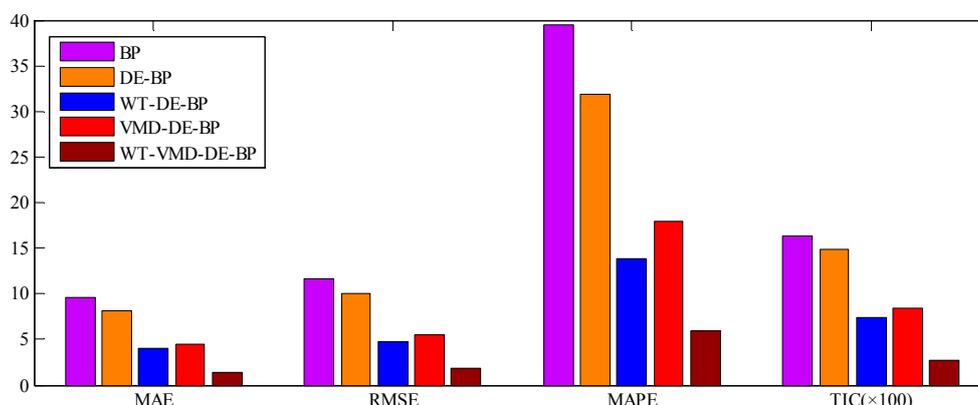


Figure 13. Error graphics of different models (Wuhan).

Table 3. The comparison results of Comparisons I, II and III (Wuhan).

Index	The Proportion of Reduction				
	WT-VMD-DE-BP vs. WT-DE-BP	WT-VMD-DE-BP vs. VMD-DE-BP	WT-DE-BP vs. DE-BP	VMD-DE-BP vs. DE-BP	DE-BP vs. BP
	MAE (%)	66.91	70.42	50.12	44.21
RMSE (%)	62.94	67.69	51.08	45.04	13.69
MAPE (%)	57.01	66.72	56.67	44.02	19.14
TIC (%)	57.14	62.50	50.00	42.86	14.29

(1) Findings of Comparison I (WT-DE-BP vs. DE-BP and VMD-DE-BP vs. DE-BP)

In Table 3, it is obvious that the values of MAE, RMSE, MAPE and TIC of DE-BP model have been reduced by 50.12%, 51.08%, 56.67% and 50.00%, respectively, via integrating the WT decomposition technique into DE-BP model, and have been decreased by 44.21%, 45.04%, 44.02% and 42.86%, respectively, through combining the VMD decomposition technique into DE-BP model. Based on the above comparison results, it can be concluded that through decomposing the PM_{2.5} concentration series into a set of subsets with different frequencies, the single decomposition technique (WT and VMD) can decrease the characteristics of non-linearity and non-stability existed in the original PM_{2.5} concentration series to some extent, and thus is benefit for improving the forecasting ability of DE-BP model.

(2) Findings of Comparison II (WT-VMD-DE-BP vs. WT-DE-BP and WT-VMD-DE-BP vs. VMD-DE-BP)

In Table 2, it can be found that the values of MAE, RMSE, MAPE and TIC of WT-VMD-DE-BP model decrease by 66.91%, 62.94%, 57.01% and 57.14%, respectively, compared with those of WT-DE-BP model, and 70.42%, 67.69%, 66.72% and 62.50%, respectively, compared with those of VMD-DE-BP model. Therefore, based on the above analysis, it can be easily found that the proposed WT-VMD-DE-BP model can significantly decrease the errors including MAE, RMSE, MAPE and TIC of WT-DE-BP and VMD-DE-BP models. Thus, it can be concluded that the hybrid WT-VMD decomposition technique proposed in this paper is very effective for improving the forecast accuracy. The reason lies in that the single decomposition techniques (WT and VMD) have the drawback of mode mixing problem with different levels, which makes the multiple frequency components existed in the PM_{2.5} concentration series cannot be effectively extracted, and therefore leads to an inferior forecasting performance.

(3) Findings of Comparison III (DE-BP vs. BP)

In Table 2, it is obvious that the values of MAE, RMSE, MAPE and TIC of BP model has been reduced by 15.50%, 13.69%, 19.14% and 14.29%, respectively, via integrating the DE algorithm into BP model. Thus, it can be concluded that through optimizing weight matrices and thresholds using DE algorithm, the BP model obtains stronger approximation ability. In addition, it can also be seen that the DE algorithm cannot effectively decrease the forecast errors without decomposition techniques (WT and VMD), which confirms that the multiple frequency components existed in the $PM_{2.5}$ concentration series have remarkable influence on the forecast accuracy.

3.3.3. $PM_{2.5}$ Concentration Forecasting in Tianjin

In order to further systematically and comprehensively testify the validity and applicability of the proposed WT-VMD-DE-BP model, the $PM_{2.5}$ concentration series collected in Tianjin (see Figure 7) is also taken as another study case. Similar to the case in Wuhan, the decomposition results of original $PM_{2.5}$ concentration series using WT decomposition method are shown in Figure 14. For each subset, the forecast result of WT-DE-BP model is shown in Figure 15. The decomposition results of each subset based on VMD decomposition method are depicted in Figure 16. In addition, for each subset, the forecast result of WT-VMD-DE-BP model is provided in Figure 17. Finally, the ultimate $PM_{2.5}$ concentration forecast results of different models are illustrated in Figure 18. The forecast errors, MAE, RMSE, MAPE and TIC, of all the forecasting models are also calculated and displayed, respectively, in Tables 4 and 5 and Figure 19.

Based on the simulation results, it is obvious that the similar conclusions to the case in Wuhan can be obtained. As it is shown, the proposed WT-VMD-DE-BP model owns the best performance compared with all the other considered models including BP, DE-BP, WT-DE-BP and VMD-DE-BP in this paper once again, which further confirms that the proposed model is suitable for $PM_{2.5}$ concentration forecasting, and owns a highly application ability. Similarly, it has been demonstrated that the performance of the proposed model based on the hybrid decomposition technique (WT-VMD-DE-BP model) is much better than the models embedded with single decomposition technique (WT-DE-BP model and VMD-DE-BP model). It has also been verified that the decomposition technique (WT and VMD) can improve the forecasting ability of DE-BP model, and the DE algorithm has a positive effect on the BP model.

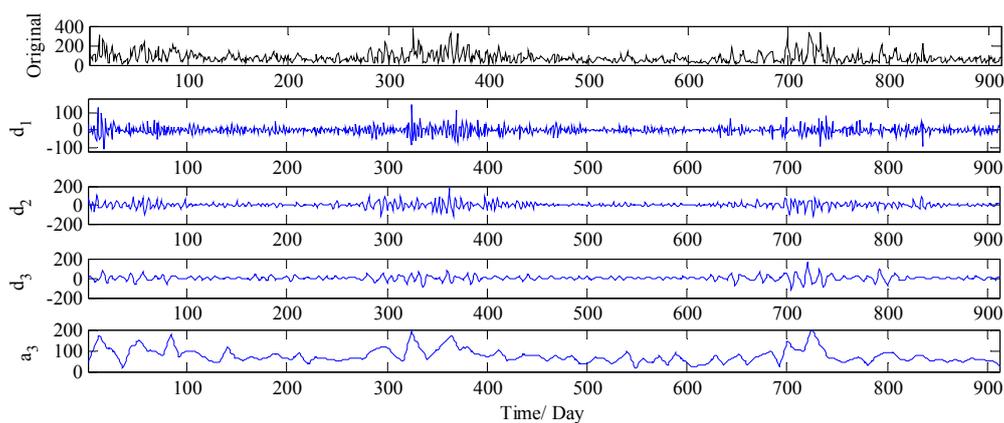


Figure 14. Decomposition results of $PM_{2.5}$ concentration series by WT (Tianjin).

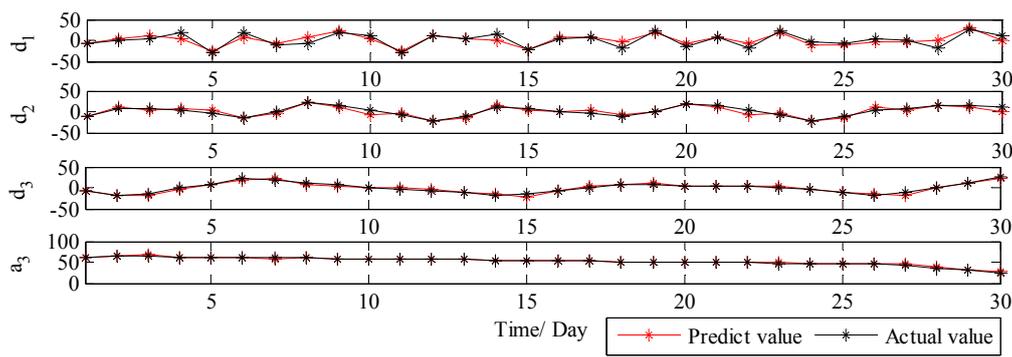


Figure 15. The forecast results of each subset by WT-DE-BP model (Tianjin).

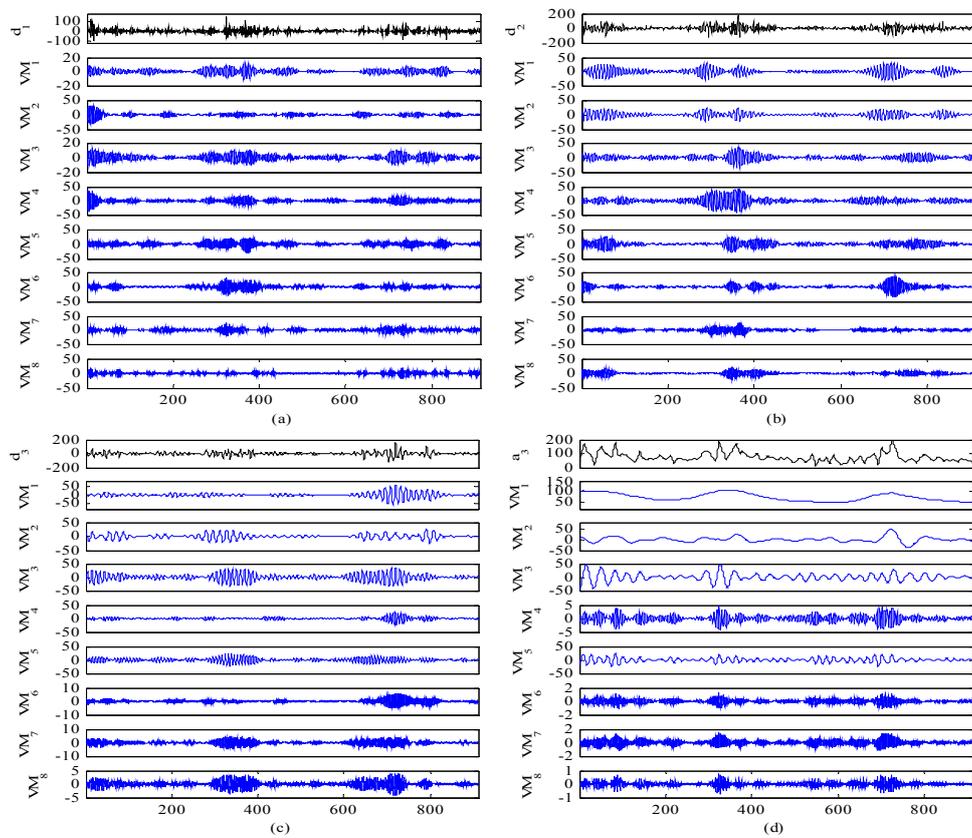


Figure 16. Decomposition results of each subset by VMD (Tianjin).

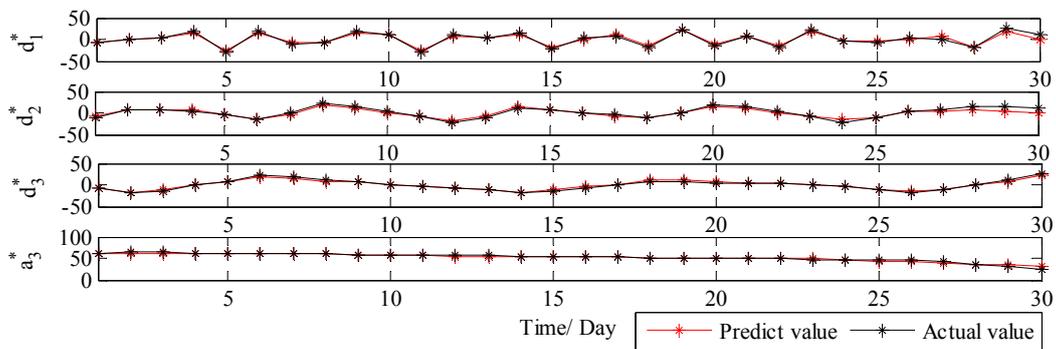


Figure 17. The forecast results of each subset by WT-VMD-DE-BP (Tianjin).

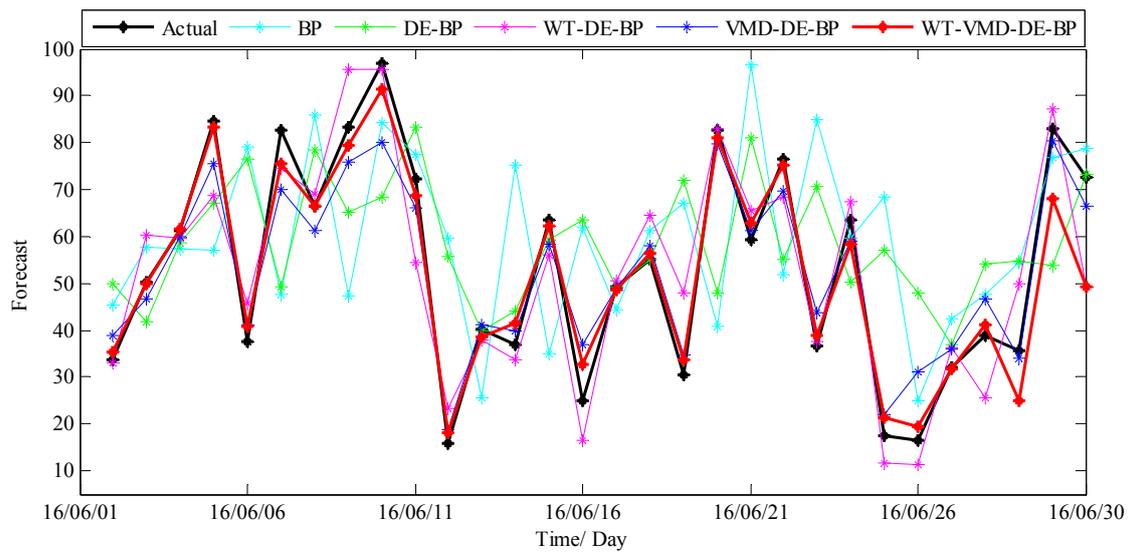


Figure 18. The ultimate forecast results of different models (Tianjin).

Table 4. Comparison of prediction performances of different models (Tianjin).

Index	BP	DE-BP	WT-DE-BP	VMD-DE-BP	WT-VMD-DE-BP
MAE	22.52	19.45	7.53	5.54	4.05
RMSE	27.28	23.81	9.50	6.79	6.25
MAPE(%)	61.03	55.51	17.49	13.66	8.88
TIC	0.22	0.20	0.08	0.06	0.05

Note: The smallest value of each row is marked in boldface.

Table 5. The comparison results of Comparisons I, II and III (Tianjin).

Index	The Proportion of Reduction				
	WT-VMD-DE-BP vs. WT-DE-BP	WT-VMD-DE-BP vs. VMD-DE-BP	WT-DE-BP vs. DE-BP	VMD-DE-BP vs. DE-BP	DE-BP vs. BP
MAE (%)	46.22	26.89	61.13	71.51	11.43
RMSE (%)	34.21	7.95	60.10	71.48	10.55
MAPE (%)	49.19	34.99	68.51	75.39	9.10
TIC (%)	37.50	16.67	60.00	70.00	9.09

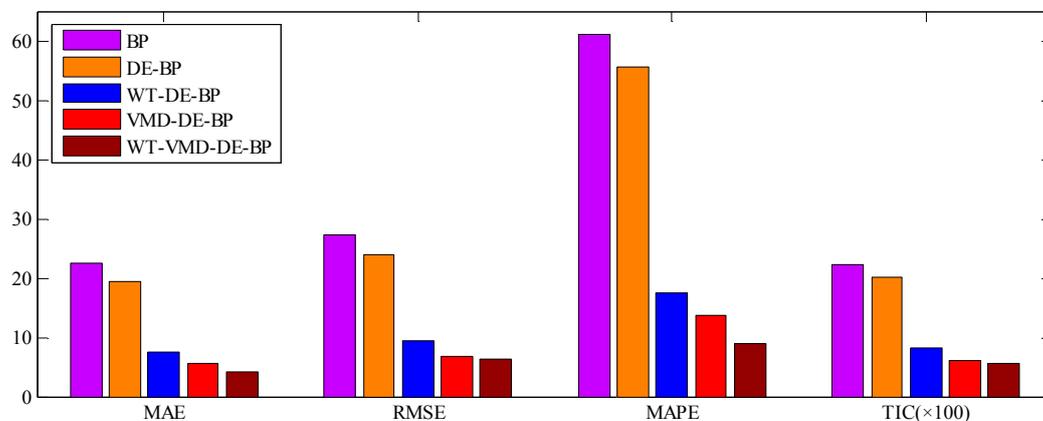


Figure 19. Error graphics of different models (Tianjin).

4. Conclusions

Accurate PM_{2.5} concentration forecasting is crucial for risk-analysis and decision-making in environmental protection departments. However, the multiple frequency components that exist in PM_{2.5} concentration series are always the challenging parts in forecasting, making models that work on the original time series unable to handle them appropriately. Thus, many researchers have been making efforts to solve this problem using different data decomposition techniques such as WT and VMD before forecasting. Since all single decomposition techniques have the drawback of mode mixing problem with different levels, which makes the multiple frequency components that exist in the PM_{2.5} concentration series unable to be effectively extracted, and consequently leading to an inferior forecasting performance. Therefore, in order to solve the mode mixing problem existed in the single decomposition technique, this paper, through combing the advantages of WT and VMD, proposes a novel hybrid WT-VMD decomposition technique, and then established a forecasting model based on WT-VMD and DE-BP model to improve the forecast accuracy of PM_{2.5} concentration.

In order to demonstrate the effectiveness and applicability of the proposed model, two PM_{2.5} concentration series collected from Wuhan and Tianjin located in China are taken for conducting the empirical study. Based on the experimental results, four main conclusions can be obtained as follows: (1) The proposed WT-VMD-DE-BP model owns the best performance compared with all the other considered benchmark models including BP, DE-BP, WT-DE-BP and VMD-DE-BP, which demonstrates that the proposed model is highly suitable for the non-stationary PM_{2.5} concentration forecasting; (2) The single decomposition techniques of WT and VMD cannot improve the forecasting ability of DE-BP model significantly due to the drawback of mode mixing problem with different levels existed in WT and VMD; (3) The hybrid WT-VMD decomposition technique performs better than the single decomposition methods of WT and VMD in extracting the multiple frequency components that exist in the PM_{2.5} concentration series, thus leading to a good forecasting performance; (4) DE algorithm has a positive effect on the BP model by optimizing the weights and thresholds between input layer and hidden layer.

However, as mentioned above, the intermittent and unstable nature of PM_{2.5} concentration series makes its forecasting become a very difficult task. Therefore, there are still several research directions left for the future. For example, some meteorological factors such as atmospheric pressure, temperature, and precipitation may be integrated into the forecasting model to improve the forecast accuracy. Furthermore, since the PM_{2.5} concentration series have some similar characteristics as other time series such as non-linearity and non-stability, the proposed model in this study can also be used for other complex time series forecasting, such as forecasting of electricity load, wind speed and stock price.

Acknowledgments: This research was supported by the National Natural Science Foundation of China (Grant No. 71301153); the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry of China; the Science Foundation of Mineral Resource Strategy and Policy Research Center, China University of Geosciences (Grant No. H2017011B); and the Natural Science Foundation of Hubei Province (Grant No. 2015CFB497).

Author Contributions: Deyun Wang designed the experiment for testing the proposed hybrid forecasting model and wrote the manuscript. Yanling Liu and Hongyuan Luo made the program in MATLAB and analyzed the data. Chenqiang Yue and Sheng Cheng provided critical review and manuscript editing. All authors read and approved the final manuscript.

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

References

1. Zhang, X.Y.; Wang, Y.Q.; Niu, T.; Zhang, X.C.; Gong, S.L.; Zhang, Y.M.; Sun, J.Y. Atmospheric aerosol compositions in China: Spatial/temporal variability, chemical signature, regional haze distribution and comparisons with global aerosols. *Atmos. Chem. Phys.* **2012**, *12*, 779–799. [[CrossRef](#)]

2. Zhou, Q.P.; Jiang, H.Y.; Wang, J.Z.; Zhou, J.L. A hybrid model for PM_{2.5} forecasting based on ensemble empirical mode decomposition and a general regression neural network. *Sci. Total Environ.* **2014**, *496*, 264–274. [[CrossRef](#)] [[PubMed](#)]
3. Mishra, D.; Goyal, P.; Upadhyay, A. Artificial intelligence based approach to forecast PM_{2.5} during haze episodes: A case study of Delhi, India. *Atmos. Environ.* **2015**, *102*, 239–248. [[CrossRef](#)]
4. Qiao, T.; Zhao, M.; Xiu, G.; Yu, J.Z. Simultaneous monitoring and compositions analysis of PM₁, and PM_{2.5} in Shanghai: Implications for characterization of haze pollution and source apportionment. *Sci. Total Environ.* **2016**, *557*, 286–394.
5. Ye, X.N.; Ma, Z.; Zhang, J.C.; Du, H.H.; Chen, J.M.; Chen, H.; Yang, X.; Gao, W.; Geng, F.H. Important role of ammonia on haze formation in Shanghai. *Environ. Res. Lett.* **2011**, *6*, 1–5. [[CrossRef](#)]
6. Donkelaar, A.; Martin, R.; Brauer, M.; Kahn, R.; Levy, R.; Verduzco, C.; Villeneuve, P.J. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environ. Health Perspect.* **2010**, *118*, 847–855. [[CrossRef](#)] [[PubMed](#)]
7. Lv, B.L.; Cobourn, W.G.; Bai, Y.Q. Development of nonlinear empirical models to forecast daily PM_{2.5} and ozone levels in three large Chinese cities. *Atmos. Environ.* **2016**, *147*, 209–223. [[CrossRef](#)]
8. Liu, S.K.; Cai, S.; Chen, Y.; Xiao, B.; Chen, P.; Xiang, X.D. The effect of pollutional haze on pulmonary function. *J. Thorac. Dis.* **2016**, *8*, 41–56.
9. World Health Organization (WHO). *Health Risks of Particulate Matter from Long-Range Transboundary Air Pollution*; WHO Regional Office for Europe: Copenhagen, Denmark, 2006.
10. Li, H.; Wu, H.; Wang, Q.; Yang, M.; Li, F.; Sun, Y.X.; Qian, X.; Wang, J.; Wang, C. Chemical partitioning of fine particle-bound metals on haze-fog and non-haze-fog days in Nanjing, China and its contribution to human health risks. *Atmos. Res.* **2016**, *183*, 142–150. [[CrossRef](#)]
11. Feng, X.; Li, Q.; Zhu, Y.; Hou, J.; Jin, L.; Wang, J. Artificial neural networks forecasting of PM_{2.5} pollution using air mass trajectory based geographic model and wavelet transformation. *Atmos. Environ.* **2015**, *107*, 118–128. [[CrossRef](#)]
12. Liu, D.J.; Li, L. Application study of comprehensive forecasting model based on entropy weighting method on trend of PM_{2.5} concentration in Guangzhou, China. *Int. J. Environ. Res. Public Health.* **2015**, *12*, 7085–7099. [[CrossRef](#)] [[PubMed](#)]
13. Chen, Y.Y.; Shi, R.H.; Shu, S.J.; Gao, W. Ensemble and enhanced PM₁₀ concentration forecast model based on stepwise regression and wavelet analysis. *Atmos. Environ.* **2013**, *74*, 346–359. [[CrossRef](#)]
14. Djalalova, I.; Monache, L.D.; Wilczak, J. PM_{2.5} analog forecast and Kalman filter post-processing for the Community Multi-scale Air Quality (CMAQ) model. *Atmos. Environ.* **2015**, *108*, 76–87. [[CrossRef](#)]
15. Zhang, H.; Zhang, W.D.; Palazoglu, A.; Sun, W. Prediction of ozone levels using a hidden Markov model (HMM) with gamma distribution. *Atmos. Environ.* **2012**, *62*, 64–73. [[CrossRef](#)]
16. Niu, M.; Wang, Y.; Sun, S.; Li, Y. A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM_{2.5}, concentration forecasting. *Atmos. Environ.* **2016**, *134*, 168–180. [[CrossRef](#)]
17. Konovalov, I.B.; Beekmann, M.; Meleux, F.; Dutot, A.; Foret, G. Combining deterministic and statistical approaches for PM₁₀ forecasting in Europe. *Atmos. Environ.* **2009**, *43*, 6425–6434. [[CrossRef](#)]
18. Song, Y.; Qin, S.; Qu, J.; Liu, F. The forecasting research of early warning systems for atmospheric pollutants: A case in Yangtze River Delta region. *Atmos. Environ.* **2015**, *118*, 58–69. [[CrossRef](#)]
19. Sun, W.; Sun, J.Y. Daily PM_{2.5} concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm. *J. Environ. Manag.* **2017**, *188*, 144–152. [[CrossRef](#)] [[PubMed](#)]
20. Jian, L.; Zhao, Y.; Zhu, Y.P.; Zhang, M.B.; Bertolatti, D. An application of ARIMA model to predict submicron particle concentrations from meteorological factors at a busy roadside in Hangzhou, China. *Sci. Total Environ.* **2012**, *426*, 336–345. [[CrossRef](#)] [[PubMed](#)]
21. Stadlober, E.; Hörmann, S.; Pfeiler, B. Quality and performance of a PM₁₀ daily forecasting model. *Atmos. Environ.* **2008**, *42*, 1098–1109. [[CrossRef](#)]
22. Kumar, U.; Ridder, K.D. GARCH modeling in association with FFT-ARIMA to forecast ozone episodes. *Atmos. Environ.* **2010**, *44*, 4252–4265. [[CrossRef](#)]
23. Pai, T.Y.; Ho, C.L.; Chen, S.W.; Lo, H.M.; Sung, P.J.; Lin, S.W.; Lai, W.-J.; Tseng, S.-C.; Ciou, S.-P.; Kuo, J.-L.; et al. Using seven types of GM (1.1) model to forecast hourly particulate matter concentration in Banciao City of Taiwan. *Water Air Soil Pollut.* **2011**, *217*, 25–33. [[CrossRef](#)]

24. Sun, W.; Zhang, H.; Palazoglu, A.; Singh, A.; Zhang, W.D.; Liu, S.W. Prediction of 24-hour-average PM_{2.5} concentrations using a hidden Markov model with different emission distributions in Northern California. *Sci. Total Environ.* **2013**, *443*, 93–103. [[CrossRef](#)] [[PubMed](#)]
25. Ordieres, J.B.; Vergara, E.P.; Capuz, R.S.; Salazar, R.E. Neural network prediction model for fine particulate matter (PM_{2.5}) on the US-Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua). *Environ. Modell. Softw.* **2005**, *20*, 547–559. [[CrossRef](#)]
26. Voukantsis, D.; Karatzas, K.; Kukkonen, J.; Räsänen, T.; Karppinen, A.; Kolehmainen, M. Inter comparison of air quality data using principal component analysis, and forecasting of PM₁₀ and PM_{2.5} concentrations using artificial neural networks, in Thessaloniki and Helsinki. *Sci. Total Environ.* **2011**, *409*, 1266–1276. [[CrossRef](#)] [[PubMed](#)]
27. Lin, K.P.; Pai, P.F.; Yang, S.L. Forecasting concentrations of air pollutants by logarithm support vector regression with immune algorithms. *Appl. Math. Comput.* **2011**, *217*, 5318–5327. [[CrossRef](#)]
28. Perez, P. Combined model for PM₁₀ forecasting in a large city. *Atmos. Environ.* **2012**, *60*, 271–276. [[CrossRef](#)]
29. Antanasijević, D.Z.; Pocajt, V.V.; Povrenović, D.S.; Ristić, M.D.; Perić-Grujić, A.A. PM₁₀ emission forecasting using artificial neural networks and genetic algorithm input variable optimization. *Sci. Total Environ.* **2013**, *443*, 511–519. [[CrossRef](#)] [[PubMed](#)]
30. Bai, Y.; Li, Y.; Wang, X.X.; Xie, J.J.; Li, C. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmos. Pollut. Res.* **2016**, *7*, 557–566. [[CrossRef](#)]
31. Liu, H.; Tian, H.Q.; Li, Y.F. Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms. *Energy Convers. Manag.* **2015**, *100*, 16–22. [[CrossRef](#)]
32. Wang, D.Y.; Luo, H.Y.; Grunder, O.; Lin, Y.B.; Guo, H.X. Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. *Appl. Energy* **2017**, *190*, 390–407. [[CrossRef](#)]
33. Bilgin, S.; Çolak, O.H.; Koklukaya, E.; Niyazi, A. Efficient solution for frequency band decomposition problem using wavelet packet in HRV. *Digit. Signal Process.* **2008**, *18*, 892–899. [[CrossRef](#)]
34. Tascikaraoglu, A.; Sanandaji, B.M.; Poolla, K.; Varaiya, P. Exploiting sparsity of interconnections in spatio-temporal wind speed forecasting using Wavelet Transform. *Appl. Energy* **2016**, *165*, 735–747. [[CrossRef](#)]
35. Amjady, N.; Keynia, F. Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. *Energy* **2009**, *34*, 46–57. [[CrossRef](#)]
36. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE T. Signal. Proces.* **2014**, *62*, 531–544. [[CrossRef](#)]
37. Wang, S.; Zhang, N.; Wu, L.; Wang, Y. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew. Energy* **2016**, *94*, 629–636. [[CrossRef](#)]
38. Wang, H.S.; Wang, Y.N.; Wang, Y.C. Cost estimation of plastic injection molding parts through integration of PSO and BP neural network. *Expert Syst. Appl.* **2013**, *40*, 418–428. [[CrossRef](#)]
39. Yu, F.; Xu, X. A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. *Appl. Energy* **2014**, *134*, 102–113. [[CrossRef](#)]
40. Storn, R.; Price, K. Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.* **1997**, *11*, 341–359. [[CrossRef](#)]

