

Research Article

A New Decomposition Ensemble Learning Approach with Intelligent Optimization for PM_{2.5} Concentration Forecasting

Guangyuan Xing,¹ Shaolong Sun ,² and Jue Guo ²

¹School of Economics and Finance, Xi'an Jiaotong University, Xi'an 710061, China

²School of Management, Xi'an Jiaotong University, Xi'an 710049, China

Correspondence should be addressed to Shaolong Sun; sunshl13@lzu.edu.cn

Received 4 November 2019; Revised 14 January 2020; Accepted 26 February 2020; Published 22 March 2020

Academic Editor: Seenith Sivasundaram

Copyright © 2020 Guangyuan Xing et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this study, we focus our attention on the forecasting of daily PM_{2.5} concentrations. According to the principle of “divide and conquer,” we propose a novel decomposition ensemble learning approach by integrating ensemble empirical mode decomposition (EEMD), artificial neural networks (ANNs), and adaptive particle swarm optimization (APSO) for forecasting PM_{2.5} concentrations. Our proposed decomposition ensemble learning approach is formulated exclusively to deal with difficulties in quantitating meteorological information with high volatility, irregularity, and complicity. This decomposition ensemble learning approach mainly consists of three steps. First, we utilize EEMD to decompose original time series of PM_{2.5} concentrations into a specific amount of independent intrinsic mode functions (IMFs) and residual term. Second, the ANN, whose connection parameters are optimized by APSO algorithm, is employed to model IMFs and residual terms, respectively. Finally, another APSO-ANN is applied to aggregate the forecast IMFs and residual term into a collection as the final forecasting results. The empirical results show that the forecasting of our decomposition ensemble learning approach outperforms other benchmark models in terms of level accuracy and directional accuracy.

1. Introduction

With the development at the technological level and the improvement of people's living standards, environment pollution becomes more and more serious, especially in developing countries. PM_{2.5} refers to the particles having the diameter of 2.5 micrometers or smaller, which can go directly to the alveoli of the lungs. Compared to the PM₁₀ (size of 10 microns or less) and TSP (size of 100 microns or less), PM_{2.5} is more likely to absorb hazardous and noxious substances. It is the carrier of all sorts of toxic substances in the air. Some scientific research shows that nitrogen oxides and sulfur dioxide emissions may be separately transformed to PM_{2.5} nitrate ion and sulfate ion in particular environmental conditions. Human exposure to PM_{2.5} can lead to a variety of adverse health impacts, such as cardiovascular and respiratory problems [1–3]. Based on the effect on environment and human health, PM_{2.5} pollution hierarchies

have been divided into six grades from excellent to serious pollution, which are described in Table 1. During the past two decades, some epidemiological studies have demonstrated that the major air pollutant impacting human health is particulate matter [4]. The adverse health impacts of particulate matter have become a well-known problem in our daily life. Except the accumulation of dust and the reduction of visibility, the direct effect on human health via inhalation is a severe problem [5, 6].

Due to the worse harm of PM_{2.5}, it became the study object and chief pollutant for rigorous control in the world, especially in the developed countries, during recent years. Air quality monitoring systems have gain large amounts of pollutant concentration data hourly or daily; it is necessary for us to analyze the data through appropriate methods [7, 8]. However, due to the serious environmental pollution, air quality monitoring systems show that many areas do not conform to the standards, which may lead to serious health

TABLE 1: PM_{2.5} pollution hierarchies.

PM _{2.5} daily concentration	Pollution levels
0–50 $\mu\text{g}/\text{m}^3$	I (excellent)
50–100 $\mu\text{g}/\text{m}^3$	II (good)
100–150 $\mu\text{g}/\text{m}^3$	III (mild pollution)
150–200 $\mu\text{g}/\text{m}^3$	IV (moderate pollution)
200–300 $\mu\text{g}/\text{m}^3$	V (high pollution)
>300 $\mu\text{g}/\text{m}^3$	VI (serious pollution)

problems, with ecological and economic effects. However, very few countries have a real-time air quality forecasting (RT-AQF) platform. In the United States, the public can learn about the future air quality index (AQI) through television, newspapers, radio, Internet, and other media, including air pollutant concentration and its associated health risks [9]. Therefore, PM_{2.5} concentration forecasting is obviously necessary, which will be able to give early pollution warnings and nip the pollution in the bud, so precaution and governing can get in progress as early as possible. Recently, a lot of efforts have been made in the research of PM_{2.5} concentration forecasting.

A lot of mathematical models are applied for forecasting PM_{2.5} concentrations. According to their fundamental principles and math representation, the mathematical models are mainly classified into two types: empirical models and deterministic models [10, 11]. Empirical model is the use of statistics or big data technology to quantify the relationship between AQI observed by air quality monitoring systems and that observed by meteorological parameters. Deterministic model is to estimate the air quality index based on simulating physical and chemical reaction, which uses mathematical models to understand how chemical processes occur in the transmission and transformation process of the atmosphere, and then test these models to see if they can create the desired results [11]. However, because of the complexity of meteorological parameters and the difficulty of quantitative estimation, there exists a vast amount of uncertainty which causes PM_{2.5} concentration forecasting to differ from reality. Therefore, compared with deterministic model, empirical model has higher precision of forecasting and better adaptability. Many empirical models, such as Autoregressive Integrated Moving Average (ARIMA), multilinear regression (MLR), and artificial neural network (ANN) models, have been applied to PM_{2.5} concentration forecasting [12–15]. As a traditional statistical model, ARIMA needs historical data continuity, because it is better at capturing the linear pattern of a time series, especially seasonal pattern. Similarly, MLR is more suitable for the linear pattern, but it is difficult in capturing extreme values. Additionally, artificial neural networks (ANNs), as a new machine learning technique which has great versatility, can recognize noise and nonlinear patterns that include extremes in the original data [16]. Moreover, some researchers have found that, compared with the single model, hybrid empirical models can better capture linear and nonlinear patterns of the time series and deal with extreme value effectively, ultimately improving the forecast accuracy [17–19].

Because of the computational efficiency and forecasting accuracy, ANN model has been widely used [20, 21]. Three types of artificial neural network models and a linear model have been chosen to forecast daily PM_{2.5} concentrations in El Paso (USA) and Ciudad Juarez (Mexico) [22]. Zhu et al. put forward a hybrid model optimized by particle swarm optimization (PSO) algorithm and obtained good performance in PM_{2.5} concentrations forecasting [23]. However, even considering the meteorological and geographical data, combination of linear and nonlinear models cannot meet the complexity of air quality data [24].

Fortunately, some problems we mentioned above can be partially solved by the principle of “divide and conquer” [25]. The purpose of “divide” is to simplify the forecasting difficulties by decomposing a task into some relatively easy subtasks, while its overall goal is to formulate a consensus forecasting result for the original data [26]. Therefore, recently based on this principle, some hybrid ensemble approaches had been put forward to solve some difficult forecasting problems, such as the forecasting of international crude oil price, and empirical results show that hybrid ensemble approaches are better than individual forecasting models [27, 28]. In fact, previous research has already demonstrated the advantage of “divide and conquer” principle. For instance, while integral models may ignore some value properties and thus lead to evaluation errors, Fischer has proved that decomposition method can analyze problems and their intrinsic properties and make them more comprehensive and clear [28]. Likewise, Kleinmuntz argued that individuals have the bounded ability to deal with the information, which may become invalid in the face of a large and complex system [29].

The main contribution of this study is to establish a more accurate approach to forecast PM_{2.5} concentrations, and evaluate the forecasting performance of the approach. PM_{2.5} concentrations are influenced by a lot of factors, but the influence law is uncertain, so we just care about PM_{2.5} concentrations. Based on the principle of “divide and conquer”, this study proposes a novel decomposition ensemble learning approach by integrating EEMD, ANN, and APSO optimization algorithm for PM_{2.5} concentrations forecasting at Lanzhou city in China. Generally, because of the complexity and irregularity of PM_{2.5} concentration time series, the principle of “divide and conquer” is established to deal with this problem. Therefore, a novel framework of decomposition ensemble learning approach integrating EEMD, ANN, and APSO is presented. In the proposed approach, a difficult forecasting task has been divided into several relatively simple subtasks; the process of adding such a decomposition process can make it easier to solve the problem of forecasting, thus improving the forecasting performance. The main reason for selecting Lanzhou city as the research area is that it has significant characteristics in terms of climate, topography, and population. In addition, the study verified how well the approach we presented performs in different circumstances.

The remaining parts of this article are organized as follows. Section 2 will illustrate research data collection and preprocessing. Then, Section 3 will briefly introduce the

related methods used in this study. The accuracy of forecasting results and validity of the proposed approach are discussed in Section 4. Finally, the paper is concluded in Section 5.

2. Data Collection and Preprocessing

The research, analysis, and results of this paper are all based upon the data of $PM_{2.5}$ concentrations in Lanzhou, which is the capital city of Gansu province and has specific location and climatic conditions. Lanzhou is located on the upper reaches of Yellow River and at the geometric center of China's continental territory. With the Yellow River going through, the city is sandwiched by mountains on the northern and southern banks. The average altitude of Lanzhou is 1520 m, and it is located 36 degrees 3 minutes north latitude and 103 degrees 40 minutes east longitude and situated in the temperate zone with a semiarid climate.

The $PM_{2.5}$ concentration data used in this study are obtained from the Ministry of Ecology and Environment of China (<http://www.mee.gov.cn/>). The $PM_{2.5}$ concentration daily data covers the period from January 1, 2017, to October 31, 2019, with a total of 1004 observations.

3. Related Methods

As we all know the $PM_{2.5}$ concentration has high volatility, nonlinearity, and irregularity. In this study, we propose a new decomposition ensemble learning approach to forecast $PM_{2.5}$ concentrations in terms of the principle of "divide and conquer". The general framework of our proposed decomposition ensemble learning approach is as follows: decomposition, single forecasting, and ensemble forecasting. First of all, some decomposition methods can be utilized to decompose the original $PM_{2.5}$ concentrations data into several meaningful component consequences. Then, some optimized forecasting methods are employed to forecast each component, respectively. Finally, the forecasts results of each component can be aggregated into the final forecasting results by means of the ensemble approaches [30].

In summary, different data decomposition methods, intelligent optimization algorithms, forecasting models, and ensemble approaches can formulate different decomposition ensemble learning approaches. In this study, firstly, we utilize ensemble empirical mode decomposition (EEMD) to decompose original data of $PM_{2.5}$ concentrations into a specific amount of independent intrinsic mode functions (IMFs) and a residual term. Secondly, artificial neural network (ANN) optimized by adaptive particle swarm optimization (APSO) is applied to forecast all IMFs and residual term, respectively. Finally, another APSO-ANN is employed to aggregate the forecasting results of IMFs and residual term into a collection as the final forecasting results. This is called EEMD-based APSO-ANN ensemble learning approach. The overall formulation process of our proposed decomposition ensemble learning approach is as follows.

3.1. Ensemble Empirical Mode Decomposition. Empirical model decomposition (EMD) was initially proposed by

Huang et al. [31]. In order to overcome the shortcoming of the mode mixing problem in EMD, Wu and Huang presented the ensemble empirical mode decomposition (EEMD) [32]. EMD and EEMD are self-adaptive algorithms compared with other traditional decomposition methods, such as wavelet decomposition and Fourier decomposition. The specific effect of local feature can identify all modes; hence, EMD and EEMD decompose signals into several intrinsic mode functions according to its characteristic of time scale.

In recent years, EMD and EEMD have been widely applied to decompose complex time series and some complex system modeling [31, 33–35]. This study chooses EEMD as the data decomposition method. The EMD and EEMD method will be introduced as follows.

EMD method is a kind of adaptive time series decomposition technique which is used to process nonlinear and nonstationary signals and is based on Hilbert-Huang transform (HHT) [36]. Because of the complexity of the data, the method assumes that data may have different modes of oscillations simultaneously. Tested signals are decomposed into a number of intrinsic mode functions (IMFs) by using local wave method, and time-frequency spectrums of IMFs are acquired by means of Hilbert transformation, which must meet the following requirements: (1) they function within the entire time domain, in which the number of local extreme value points and zero crossings must be equal, or at most by one; (2) the local maximum envelope and the local minimum envelope must be zero on average at any time point.

We can define meaningful IMFs through those two conditions. According to the definition, we can decompose any complicated data series x_t ($t = 1, 2, \dots, T$); the process is presented as follows.

- (1) Find out all the local extrema of original data series x_t .
- (2) Use cubic spline interpolation to create the upper and lower envelopes $x_{up,t}$ and $x_{low,t}$, respectively, and calculate the average of the upper and lower envelope: $m_t = (x_{up,t} + x_{low,t})/2$.
- (3) Subtract the envelope mean from original time series and define it as c_t , $c_t = x_t - m_t$. Inspect whether c_t meets the above two basic conditions of IMF; if c_t is not an IMF, replace x_t with c_t and repeat the above two steps.
- (4) Extract an IMF and replace x_t with the residual $r_t = x_t - c_t$. Repeat Steps 1–3 until the stop criterion is satisfied.

Using this screening process, the original data series x_t can finally be decomposed into a sum of IMFs and a residue term:

$$x_t = \sum_{j=1}^n c_{j,t} + r_{n,t}, \quad (1)$$

where n is the number of IMFs, $r_{n,t}$ is the final residue term, and $c_{j,t}$ ($j = 1, 2, \dots, n$) is the j th IMF.

Even though EMD is a fully data-driven and self-adaptive data decomposition method, there is also an obvious disadvantage, such as the mode mixing. In order to address the mode mixing problem, EEMD technique was proposed by Wu and Huang [32]. EEMD takes the method of EMD as the basis and successfully solves the mode mixing problem caused by intermittent noise by adding white noise to the original time series before decomposition. EEMD method can not only reserve the information of the original data, but also overcome the drawback of mode mixing. The sifting steps of EEMD are as follows:

- (1) Add a group of white noise w_t to the original data series x_t to acquire X_t : $X_t = x_t + w_t$.
- (2) Employ EMD method to decompose X_t , and obtain a series of IMFs: $X_t = \sum_{j=1}^n c_j + r_n$.
- (3) Add different white noise series to the original data, repeat the above steps: $X_{i,t} = x_{i,t} + w_{i,t}$, and obtain corresponding IMF components: $X_{i,t} = \sum_{j=1}^n c_{i,j} + r_{i,n}$.
- (4) The final results are the ensemble averages of corresponding IMFs: $c_j = \sum_{i=1}^N c_{i,j}/N$.

Wu and Huang demonstrated that the effect of the added noise is strictly controlled via the following statistical criteria [32]:

$$\varepsilon_n = \frac{\varepsilon}{\sqrt{N}}, \quad (2)$$

where N is the total number, ε is the amplitude of the added noise, and ε_n is the final standard deviation of error between original data series and the corresponding IMFs. In practice, the total number N is often set to 100 and ε_n of white noise series is set to 0.1 or 0.2 [33].

3.2. Artificial Neural Networks. Artificial neural networks (ANNs) are widely applied in air pollution forecasting, which can build flexible model for various nonlinear problems. Relative to other types of nonlinear models, ANNs are universal approximators with a high reliability and accuracy in estimating a large class of functions. Additionally, ANNs are largely determined by the characteristics of data in the model building process; hence these techniques do not require prior assumption. The neural network architecture usually consists of the input layer, the hidden layer, and the output layer [37]. The input layer accepts the data imported to the network, and the output layer realizes the output of evaluation results. The hidden layer, which is between the input and output layer, consists of a number of neurons or hidden units placed parallel to each other. From the viewpoint of mathematics, the hidden neuron h_j is described as the following mathematical expressions [38]:

$$\begin{aligned} h_j &= \varphi(z_j), \\ z_j &= \sum_{i=1}^l w_{ij}x_i + b_j, \end{aligned} \quad (3)$$

where $\varphi(\cdot)$ is the activation function that is usually chosen as the logistic sigmoid function $\varphi(x) = 1/(1 + e^{-x})$, w_{ij} is the

weight of input x_i at neuron j , and b_j is the bias of neuron j . The relationship between the output $f(x)$ and the inputs $(x_i, y_i)_{i=1}^l$ is presented as the following mathematical expressions:

$$f(x) = w_0 + \sum_{j=1}^q w_j h_j, \quad (4)$$

where w_j ($j = 0, 1, 2, \dots, q$) is the connection weights, q is the number of hidden nodes, and $f(\cdot)$ is a function determined by the network structure and the connection weights. In this study, the architecture of ANN selects backpropagation neural network (BPNN) that is one of the most popular and effective forecasting techniques. BPNN is a three-layered feedforward architecture based on backpropagation (BP) algorithm. The details of BPNN can be found in [38].

3.3. Adaptive Particle Swarm Optimization. Particle swarm optimization (PSO) is a heuristic search algorithm based on swarm intelligence and has been widely used to solve various problems. The principle of PSO is to simulate the characteristics that the birds update location in searching food. First of all, it initializes a group of particles in the solution space, each of which denotes a potential optimal solution. The characteristics of the particles are measured by three indicators: location, speed, and fitness. Particles update timing position by tracking individual extremum (P_{best}) and global extremum (G_{best}). PSO algorithm can easily cause early maturing; in order to address this problem, a novel adaptive particle swarm optimization (APSO) algorithm has been proposed to solve the problem of low precision and avoid premature phenomena of basic PSO algorithm [39].

Supposing that there is a population in a D -dimensional search space, which consists of n particles $X = (X_1, X_2, \dots, X_n)$. Among them, $X_i = [x_{i1}, x_{i2}, \dots, x_{iD}]^T$ denotes the position of i th particle in the D -dimensional search space, also on behalf of a potential solution of the problem. According to the objective function, we can calculate the corresponding fitness value of each particle. The speed of the position change for particle is $V_i = [V_{i1}, V_{i2}, \dots, V_{iD}]^T$, $P_i = [P_{i1}, P_{i2}, \dots, P_{iD}]^T$ denotes the best previous position which gives the best fitness value of the i th particle, and $P_g = [P_{g1}, P_{g2}, \dots, P_{gD}]^T$ denotes the best position among all the particles of the population. The specific formula of adaptive adjustment is as follows:

$$V_{id}^{t+1} = w^t V_{id}^t + c_1 r_1 (P_{id}^t - X_{id}^t) + c_2 r_2 (P_{gd}^t - X_{id}^t),$$

$$X_{id}^{t+1} = X_{id}^t + V_{id}^{t+1},$$

$$w^t = \begin{cases} \lambda \cdot w^{t-1} + \theta \frac{f(P_{gd}^t) - f(X_{id}^t)}{f(P_{gd}^t) - f(X_{\min}^t)}, & w^t \geq w_{\min}, \\ w_{\min}, & w^t < w_{\min}, \end{cases} \quad (5)$$

where w^t is the adaptive inertia weight, c_1 and c_2 are nonnegative constants, which are called acceleration factors, and r_1 and r_2 are random numbers distributed in $[0, 1]$. λ and θ are constraint factors in the range $[0, 1]$. w_{\min} is the minimum inertia weight. $f(\cdot)$ is the fitness function; it is defined as follows in this study:

$$f = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (6)$$

where y_i and \hat{y}_i are the actual value and the forecast value of $\text{PM}_{2.5}$ concentration, respectively [37].

3.4. The Framework of Our Proposed Approach. Given that $x_t (t = 1, 2, \dots, T)$ is a time series, we could purpose a proactive mechanism to make m -step ahead forecasting, i.e., x_{t+m} . In this study, it is worth reminding that we apply iterative forecasting method, which can be represented as follows:

$$\hat{x}_{t+m} = f(x_t, x_{t-1}, \dots, x_{t-(l-1)}), \quad (7)$$

where \hat{x}_{t+m} is the forecast value, x_t is the actual value, and l denotes the lag orders. In ANN, the initial weights and thresholds have significant meaning and play an important role in learning and optimizing the neural network [40]. However, these parameters are randomly generated in the beginning and then adjusted in the whole training process. Hence, APSO algorithm is applied to determine the threshold and weight values of artificial neural network, as shown in Figure 1. Meanwhile, the time series data inevitably adulterates some noise or worthless and meaningless information. Therefore, our proposed EEMD-based APSO-ANN ensemble learning approach has been established to forecast $\text{PM}_{2.5}$ concentrations at Lanzhou city in China.

According to the framework in Figure 2 and the previous research, this study will establish a novel decomposition ensemble learning approach by integrating EEMD and APSO-ANN for $\text{PM}_{2.5}$ concentrations forecasting. As shown in Figure 2, our proposed EEMD-APSO-ANN-APSO-ANN decomposition ensemble learning approach is generally composed of the following three main steps:

- (1) The original $\text{PM}_{2.5}$ concentrations time series $x_t (t = 1, 2, \dots, T)$ is decomposed into n IMFs $c_{j,t} (j = 1, 2, \dots, n)$ and one residual component $r_{n,t}$ by EEMD method.
- (2) APSO-ANN is regarded as a single forecasting technique to model the decomposed IMF components and the residue component, respectively. As a result, all components can obtain the corresponding forecasting results.
- (3) Finally, the final forecasting \hat{x}_{t+m} for the original $\text{PM}_{2.5}$ concentrations time series x_t is obtained by integrating the forecasting results of the IMFs and residue components, using another APSO-ANN technique as an ensemble approach.

In short, in view of the principle of “divide and conquer”, our proposed EEMD-based APSO-ANN ensemble learning

approach can be described as a general framework of “EEMD (decomposition)–APSO-ANN (single forecasting)–APSO-ANN (ensemble forecasting)”. In order to verify the effectiveness of our proposed EEMD-APSO-ANN-APSO-ANN decomposition ensemble learning approach, $\text{PM}_{2.5}$ concentrations data collected from Lanzhou city is used as the test target. For more details, we will discuss in the next section.

4. Empirical Study

In this study, the sample data are divided into two subsets: training subset and testing subset. We treat data from January 1, 2017, to September 30, 2019, as training subset with 974 observations used for model training. Similarly, data from October 1, 2019, to October 31, 2019, with 31 observations are treated as the testing subset to evaluate the forecasting performance of the model. Additionally, data of the past 1 day (lag order 1), 2 days (lag order 2), 3 days (lag order 3), 4 days (lag order 4), and 5 days (lag order 5) are utilized as initial input form to forecast the following daily $\text{PM}_{2.5}$ concentrations, respectively, and finally the input form with minimum forecasting error is chosen as optimal input structure.

4.1. Evaluation Criteria of Forecasting Performance. In this study, two evaluation criteria are utilized to evaluate the forecasting performance of our proposed decomposition ensemble learning approach. They are mean square error (MSE) and mean absolute percent error (MAPE). The smaller the index value is, the better the forecasting performance will be. The formulas of criteria are as follows [41]:

$$\text{MSE} = \frac{1}{N} \sum_{t=1}^N |Y_t - \hat{Y}_t|, \quad (8)$$

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^N \frac{|Y_t - \hat{Y}_t|}{Y_t} * 100\%,$$

where N is the number of observation points, Y_t represents the value of actual $\text{PM}_{2.5}$ concentrations for a time period t , and \hat{Y}_t is the forecast value for the same period.

Additionally, we also consider the directional forecasting accuracy; it can be expressed by

$$D_{\text{Stat}} = \frac{1}{N} \sum_{i=1}^N a_i, \quad (9)$$

where $a_i = 1$ if $(y_{i+1} - y_i)(\hat{y}_{i+1} - y_i) \geq 0$ or $a_i = 0$ and 2 otherwise.

4.2. Empirical Results. In our proposed EEMD-APSO-ANN-APSO-ANN decomposition ensemble learning approach, the first step is to apply EEMD method to decompose the original $\text{PM}_{2.5}$ concentration data series into several independent IMF components and one residue term. In this study, the ensemble member is set to 100, and the standard deviation of added white noise in each ensemble

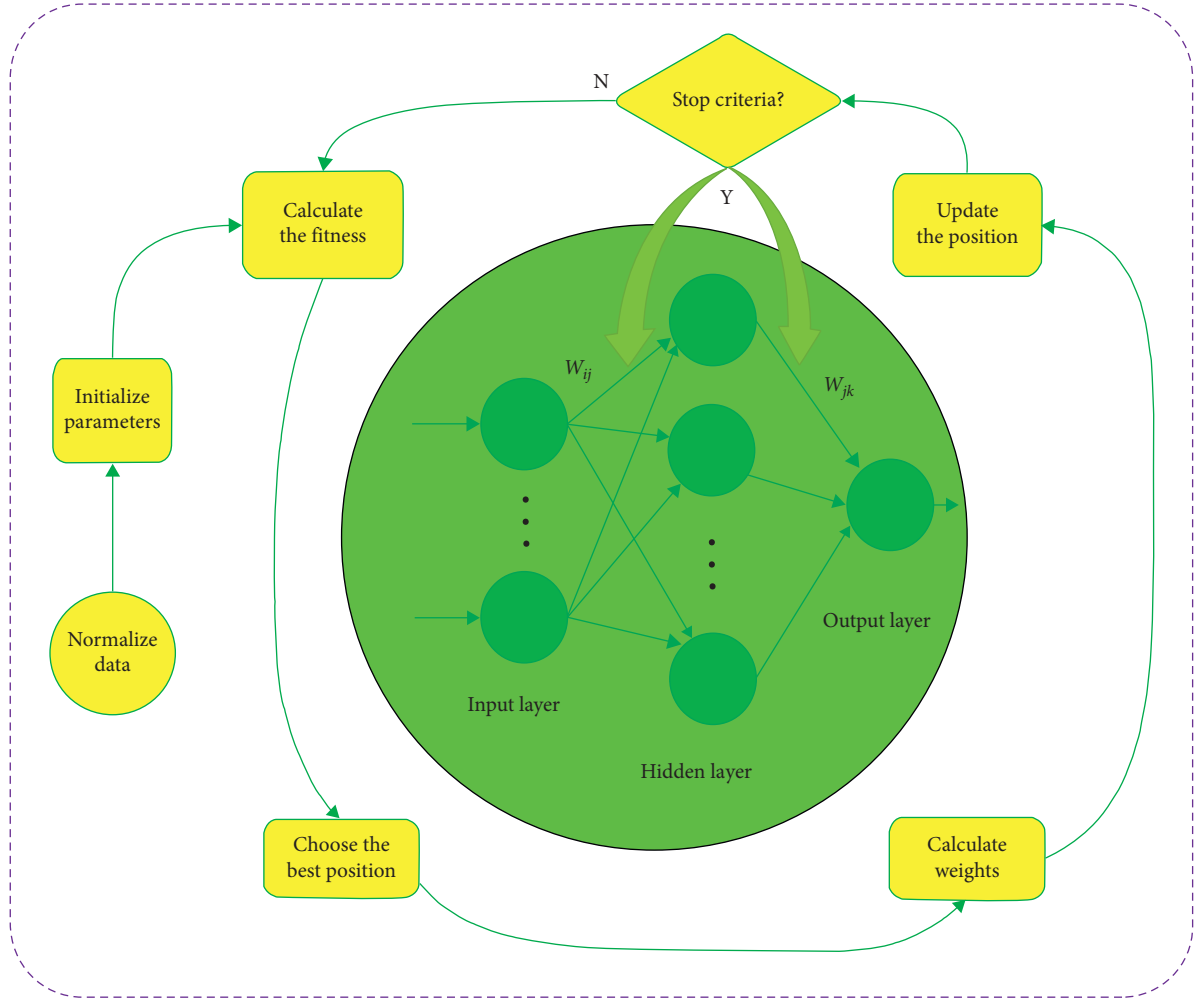


FIGURE 1: The flowchart of APSO-ANN.

member is 0.2. All IMF components are sorted from the highest to the lowest according to the frequency, and the last one is the residue term. The decomposition results of original $PM_{2.5}$ concentrations at Lanzhou city in China are shown in Figure 3. It is easy to find that the original $PM_{2.5}$ concentrations time series is decomposed into nine independent components.

For comparison, we choose some other popular forecasting models as benchmarks to be compared with our proposed EEMD-APSO-ANN-APSO-ANN decomposition ensemble learning approach. According to previous literature, five single forecasting models, ANN, GA-ANN, PSO-ANN, APSO-ANN, and ARIMA, and three groups of decomposition ensemble learning approaches are chosen as benchmark models. For the purpose of consistency, the parameters of the decomposition ensemble learning approaches are the same as single forecasting models.

To clearly analyze data, the empirical results consisted of two parts. In the first part, we will compare the results of five single forecasting models and then choose the optimum model as a single forecasting and ensemble model for decomposition ensemble learning approach. In the second part, the forecasting performance of our proposed EEMD-

APSO-ANN-APSO-ANN decomposition ensemble learning approach is compared with the other three decomposition ensemble learning approaches.

4.2.1. Performance Comparison of Single Models. In this subsection, we compare five single forecasting models, ANN, GA-AND, PSO-ANN, APSO-ANN, and ARIMA. For the ANN techniques, the numbers of inputs and hidden layer nodes are determined using the trial-and-error method, and the active function of hidden layer is sigmoid function. Table 2 shows the forecasting errors by means of MSE, MAPE, and D_{stat} . The forecasting results indicate that the APSO-ANN has a high forecasting accuracy, followed by PSO-ANN.

From Table 2, it is clearly seen that all of the ANN techniques are superior to the traditional ARIMA model, and the optimal lag order of inputs is 3. Compared with ARIMA, ANN, GA-ANN, and PSO-ANN, the MAPE, MSE, and D_{stat} of APSO-ANN are 25.22%, 0.1491, and 61.29%, respectively. On the contrary, the MAPE, MSE, and D_{stat} of ARIMA are 36.27%, 0.2231, and 51.61%. It is obvious that ANN technique whose parameters are optimized by PSO or

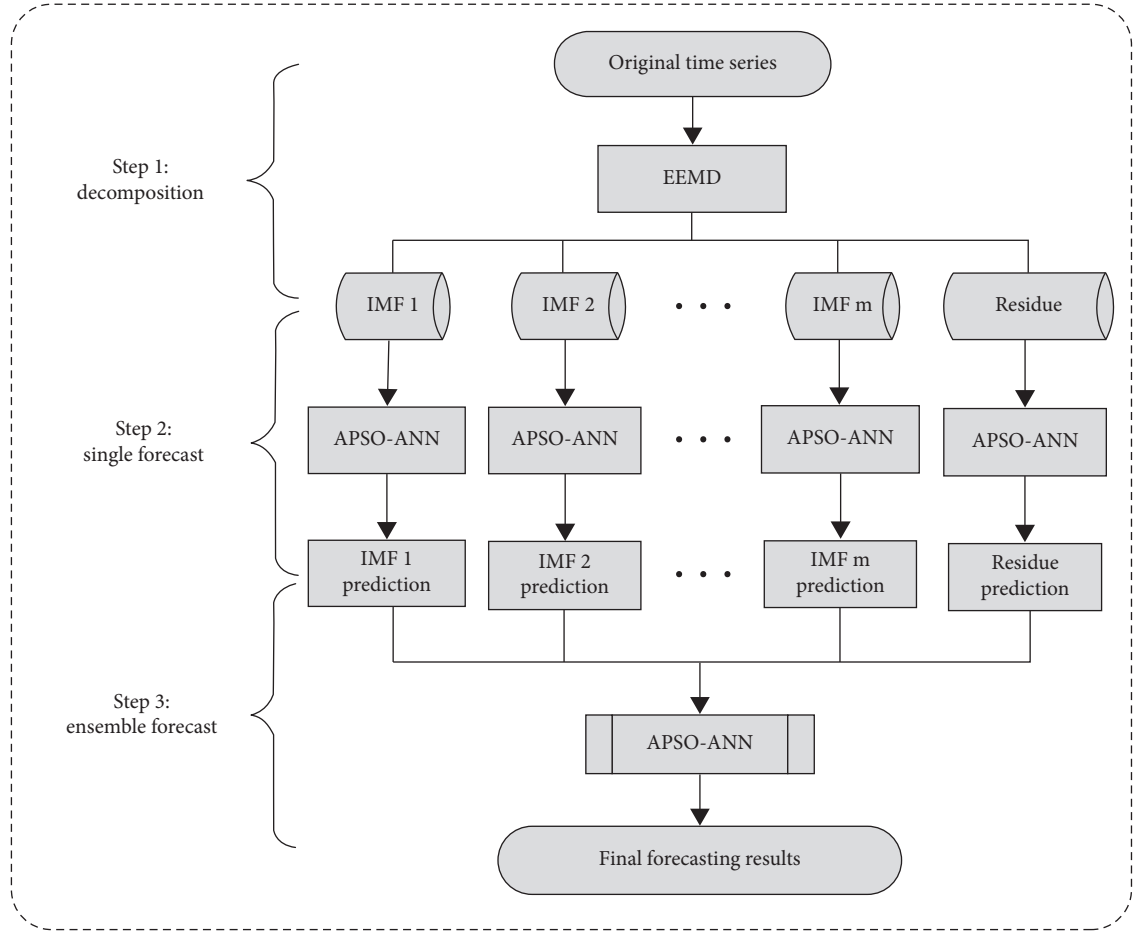


FIGURE 2: The general framework of our proposed decomposition ensemble approach.

APSO algorithm is better than ANN without any optimization scheme, so the optimized ANN technique is regarded as the single forecasting and ensemble forecasting method in our proposed decomposition ensemble learning approach.

4.2.2. Performance Comparison of Decomposition Ensemble Approaches. This subsection focuses on the forecasting performance comparison of three groups of decomposition ensemble learning approaches. Some variants of decomposition ensemble learning approaches with other decomposition methods (e.g., EMD method) and other ensemble approaches (e.g., simple addition (ADD)) are also employed as decomposition ensemble learning benchmarks to be compared with our proposed EEMD-APSO-ANN-APSO-ANN decomposition ensemble learning approach. Therefore, we select three groups of decomposition ensemble learning approaches, i.e., [EMD-APSO-ANN-ADD, EMD-APSO-ANN-PSOANN, EMD-APSO-ANN-APSO-ANN], [EEMD-PSOANN-ADD, EEMD-PSOANN-PSOANN, EEMD-PSOANN-APSO-ANN], and [EEMD-APSO-ANN-ADD, EEMD-APSO-ANN-PSOANN, EEMD-APSO-ANN-APSO-ANN]. Table 3 provides the forecasting results of different decomposition ensemble learning approaches.

For the above different decomposition ensemble learning approaches, we firstly discuss forecasting performance of decomposition ensemble learning approaches with different decomposition methods. We can clearly see that the EEMD-based decomposition ensemble learning approaches can obtain better forecasting accuracy than the corresponding EMD-based decomposition ensemble learning approaches. That is, the EEMD is much more efficient than EMD in data decomposition of $PM_{2.5}$ concentrations. Secondly, the forecasting performance of APSO-ANN-based decomposition ensemble learning approaches is mostly better than ADD-based decomposition ensemble learning approaches in terms of MSE, MAPE, and D_{stat} criteria. This indicates that APSO-ANN is a powerful ensemble learning method. Thirdly, we compare single forecasting models; it is clearly seen that the forecasting accuracy of APSO-ANN is better than that of PSO-ANN and GA-ANN.

In general, through the analysis above, we can obtain some interesting findings as follows: (1) Decomposition ensemble learning approaches are significantly better than other single models, such as ARIMA, ANN, GA-ANN, PSO-ANN, and APSO-ANN. The main reason is that the strategy of “divide and conquer” can effectively improve the performance of $PM_{2.5}$ concentrations forecasting. (2) It is

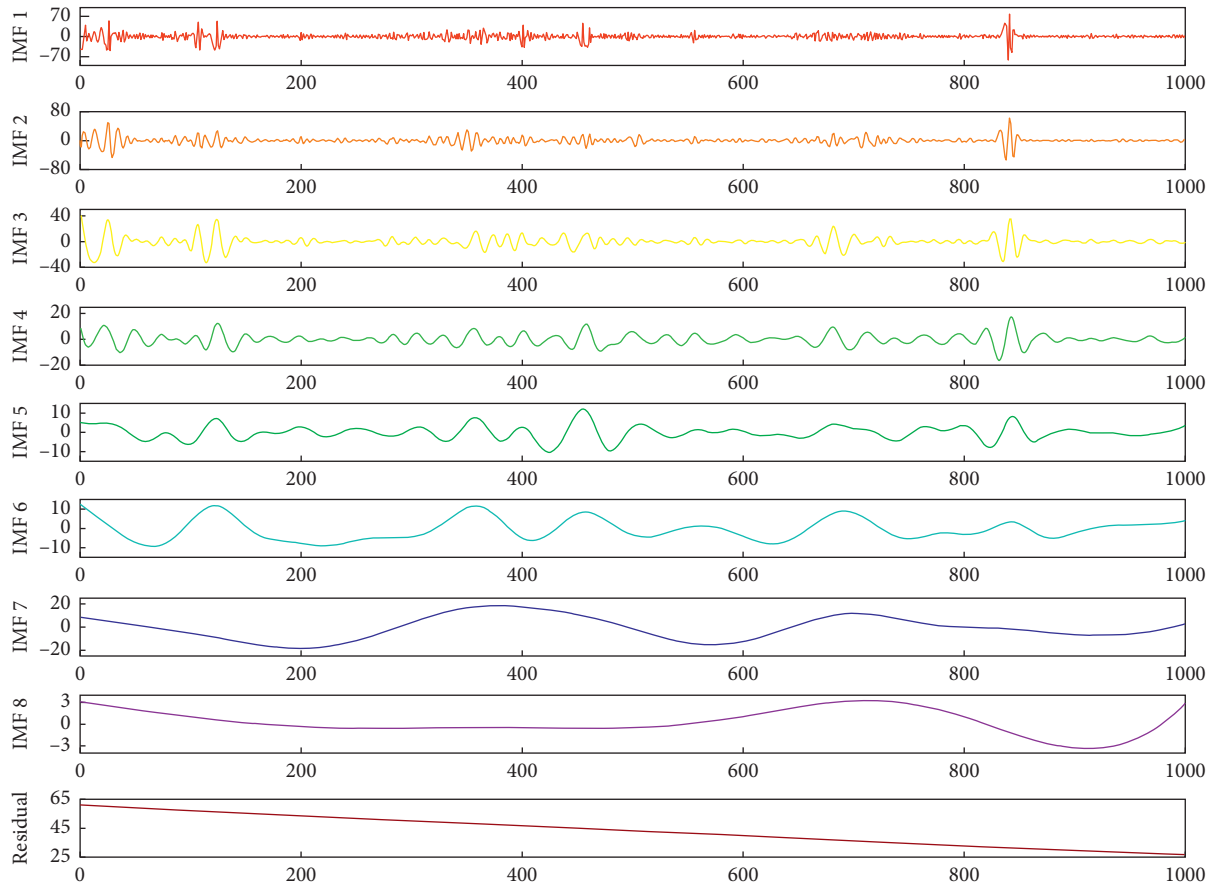


FIGURE 3: The IMFs and a residue term of $PM_{2.5}$ concentration data in Lanzhou city via EEMD algorithm.

TABLE 2: The forecasting results of single models.

Lag order	Evaluation criteria	Single forecasting models				
		ARIMA	ANN	GA-ANN	PAO-ANN	APSO-ANN
1	MSE	0.2935	0.1593	0.16951	0.1549	0.1523
	MAPE (%)	44.23	28.37	31.08	27.53	27.25
	D_{stat} (%)	54.84	54.84	58.06	58.06	58.06
2	MSE	0.1984	0.2013	0.1649	0.1531	0.1548
	MAPE (%)	35.68	28.36	29.17	27.85	27.34
	D_{stat} (%)	48.37	54.84	54.84	61.29	54.84
3	MSE	0.2231	0.1658	0.1613	0.1512	0.1491
	MAPE (%)	36.27	28.51	29.05	27.53	25.22
	D_{stat} (%)	51.61	58.06	54.84	58.06	61.29
4	MSE	0.2015	0.1702	0.1649	0.1580	0.1543
	MAPE (%)	36.87	30.32	29.41	28.19	26.94
	D_{stat} (%)	51.61	51.61	58.06	58.06	54.84
5	MSE	0.1983	0.1688	0.1671	0.1553	0.1532
	MAPE (%)	35.82	30.85	31.05	28.83	27.49
	D_{stat} (%)	48.37	51.61	58.06	54.84	54.84

clearly seen that the EEMD method performs much better than the counterpart method with EMD in terms of both level forecasting accuracy and directional forecasting accuracy. (3) After decomposition, the second and third steps of decomposition ensemble learning approach are individual forecasting and ensemble forecasting by means of

APSO-ANN and PSO-ANN with optimal weights and threshold values; empirical results show that these decomposition ensemble learning approaches are better than the other AI benchmark models. (4) Our proposed EEMD-APSO-ANN-APSO-ANN decomposition ensemble learning approach is superior to all the other benchmark models in

TABLE 3: The forecasting results of different decomposition ensemble approaches.

Lag order	Evaluation criteria	EMD-APSO-ANN			EEMD-PSOANN			EEMD-APSO-ANN		
		ADD	PSOANN	APSO-ANN	ADD	PSOANN	APSO-ANN	ADD	PSOANN	APSO-ANN
1	MSE	0.1215	0.1143	0.1126	0.1214	0.1212	0.1206	0.1211	0.1015	0.0926
	MAPE (%)	24.13	23.14	22.72	23.83	23.79	23.62	23.85	20.51	18.79
	D_{stat} (%)	70.97	74.19	77.42	67.74	70.97	77.42	70.97	74.19	80.65
2	MSE	0.0835	0.0862	0.0751	0.0782	0.0746	0.0741	0.0745	0.0756	0.0702
	MAPE (%)	16.45	16.91	15.71	16.19	15.51	15.12	15.88	15.60	14.65
	D_{stat} (%)	67.74	70.97	77.42	70.97	70.97	80.65	74.19	77.42	83.87
3	MSE	0.0774	0.0821	0.0763	0.0692	0.0663	0.0652	0.0668	0.0652	0.0611
	MAPE (%)	16.15	15.89	15.81	13.59	13.05	12.73	13.35	12.84	12.12
	D_{stat} (%)	70.91	74.19	80.65	70.97	74.19	83.87	80.65	80.65	87.10
4	MSE	0.0909	0.0883	0.0825	0.0677	0.0645	0.0639	0.0581	0.0605	0.0568
	MAPE (%)	17.68	18.16	16.57	12.95	12.71	12.40	11.59	11.48	11.33
	D_{stat} (%)	70.97	70.97	80.65	74.19	77.42	80.65	77.42	80.65	83.87
5	MSE	0.0895	0.0823	0.0853	0.0598	0.0611	0.0583	0.0591	0.0571	0.0562
	MAPE (%)	17.76	16.82	17.64	12.13	11.75	11.22	11.95	11.97	10.82
	D_{stat} (%)	70.97	74.19	83.87	77.42	80.65	83.87	80.65	87.10	90.32

terms of both level forecasting accuracy and directional forecasting accuracy. Therefore, our proposed EEMD-APSO-ANN-APSO-ANN decomposition ensemble learning approach can be used as an effective forecasting framework for forecasting $PM_{2.5}$ concentrations.

Additionally, we have set the length of inputs for ANN as lag order 1, 2, 3, 4, and 5, respectively, and while the input form is lag order 5, our proposed EEMD-APSO-ANN-APSO-ANN decomposition ensemble learning approach has the highest forecasting accuracy. Figure 4 shows the best forecasting results of $PM_{2.5}$ concentrations at Lanzhou city in China from October 1, 2019, to October 31, 2019.

5. Conclusions

The ascension of $PM_{2.5}$ concentration will lead to serious health, climate, and environment problems and cause respiratory and cardiovascular diseases. As a result, it is important and urgent to establish an early warning system based on the accurate $PM_{2.5}$ concentration forecasting. In order to address this hard issue, based on the principle of “divide and conquer”, this study proposes a new decomposition ensemble learning approach by integrating ensemble empirical mode decomposition (EEMD), artificial neural networks (ANNs), and adaptive particle swarm optimization (APSO) in order to improve the performance of $PM_{2.5}$ concentration forecasting. The $PM_{2.5}$ concentration data used in this study covers the period from January 1, 2017, to October 31, 2019, at Lanzhou city in China. Our proposed decomposition ensemble learning approach takes advantage of multiple methods, such as the effective self-adaptive data decomposition of EEMD and end-to-end parameters optimization of APSO, to improve the performance of $PM_{2.5}$ concentration forecasting. To verify performance of our proposed approach, three groups of decomposition ensemble learning approaches were chosen as benchmarks to be compared with our proposed EEMD-APSO-ANN-APSO-ANN decomposition ensemble learning approach. Empirical results show that our proposed EEMD-APSO-ANN-APSO-ANN decomposition ensemble learning approach significantly

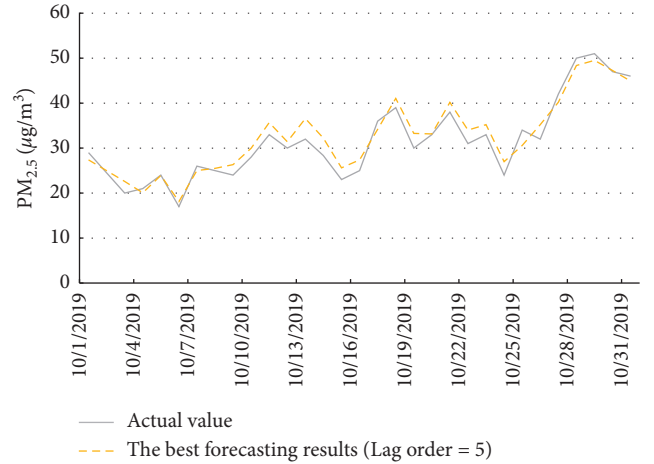


FIGURE 4: The best forecasting results of $PM_{2.5}$ concentrations at Lanzhou city in China.

improves the forecasting performance and outperforms some other benchmarks in terms of level forecasting accuracy and directional forecasting accuracy. This indicates that our proposed decomposition ensemble learning approach with effective decomposition, as well as nonlinear single and ensemble forecasting, can be used as a very promising framework to solve other complex time series forecasting problems, especially for the data characterized by high volatility and irregularity.

Additionally, our proposed EEMD-APSO-ANN-APSO-ANN decomposition ensemble learning approach can be applied to other applications such as finance forecasting and energy forecasting. Furthermore, this study mainly considers the univariate time series forecasting, while other factors affecting $PM_{2.5}$ concentrations were not taken into consideration. If those factors were incorporated into our proposed EEMD-APSO-ANN-APSO-ANN decomposition ensemble learning approach, the forecasting performance may still improve. These limitations will hopefully be addressed in future research.

Data Availability

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

Conflicts of Interest

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

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China under Grant no. 71904153 and the project funded by China Postdoctoral Science Foundation (No. 2018M53598).

References

- [1] K. R. Baker and K. M. Foley, "A nonlinear regression model estimating single source concentrations of primary and secondarily formed $PM_{2.5}$," *Atmospheric Environment*, vol. 45, no. 22, pp. 3758–3767, 2011.
- [2] X. Liu, "Effect of energy development and technological innovation on $PM_{2.5}$ in China: a spatial durbin econometric analysis," *Discrete Dynamics in Nature and Society*, vol. 2018, Article ID 2148318, 12 pages, 2018.
- [3] H. Junninen, H. Niska, K. Tuppurainen, J. Ruuskanen, and M. Kolehmainen, "Methods for imputation of missing values in air quality data sets," *Atmospheric Environment*, vol. 38, no. 18, pp. 2895–2907, 2004.
- [4] H. Kan, S. J. London, G. Chen et al., "Differentiating the effects of fine and coarse particles on daily mortality in Shanghai, China," *Environment International*, vol. 33, no. 3, pp. 376–384, 2007.
- [5] S. Qin, F. Liu, J. Wang, and B. Sun, "Analysis and forecasting of the particulate matter (PM) concentration levels over four major cities of China using hybrid models," *Atmospheric Environment*, vol. 98, pp. 665–675, 2014.
- [6] X. Li, Y. Song, Z. Yao, and R. Xiao, "Forecasting China's CO_2 emissions for energy consumption based on cointegration approach," *Discrete Dynamics in Nature and Society*, vol. 2018, Article ID 4235076, 9 pages, 2018.
- [7] J. C. Ortega, J. S. Fu, G. D. Reed et al., "A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: the case of Temuco, Chile," *Atmospheric Environment*, vol. 42, no. 35, pp. 8331–8340, 2008.
- [8] J. Tian and D. Chen, "A semi-empirical model for predicting hourly ground-level fine particulate matter ($PM_{2.5}$) concentration in southern Ontario from satellite remote sensing and ground-based meteorological measurements," *Remote Sensing of Environment*, vol. 114, no. 2, pp. 221–229, 2010.
- [9] K. Yahya, Y. Zhang, and J. M. Vukovich, "Real-time air quality forecasting over the southeastern United States using WRF/Chem-MADRID: multiple-year assessment and sensitivity studies," *Atmospheric Environment*, vol. 92, pp. 318–338, 2014.
- [10] J. Li, L. Chen, Y. Xiang, and M. Xu, "Research on influential factors of $PM_{2.5}$ within the beijing- tianjin-hebei region in China," *Discrete Dynamics in Nature and Society*, vol. 2018, Article ID 6375391, 10 pages, 2018.
- [11] W. G. Cobourn, "An enhanced $PM_{2.5}$ air quality forecast model based on nonlinear regression and back-trajectory concentrations," *Atmospheric Environment*, vol. 44, no. 25, pp. 3015–3023, 2010.
- [12] P. Goyal, A. T. Chan, and N. Jaiswal, "Statistical models for the prediction of respirable suspended particulate matter in urban cities," *Atmospheric Environment*, vol. 40, no. 11, pp. 2068–2077, 2006.
- [13] E. Stadlober, S. Hörmann, and B. Pfeiler, "Quality and performance of a PM_{10} daily forecasting model," *Atmospheric Environment*, vol. 42, no. 6, pp. 1098–1109, 2008.
- [14] P. Wang, Y. Liu, Z. Qin, and G. Zhang, "A novel hybrid forecasting model for PM_{10} and SO_2 daily concentrations," *Science of the Total Environment*, vol. 505, pp. 1202–1212, 2015.
- [15] J. Doynne and J. John, "Predicting chaotic time series," *Physical Review Letters*, vol. 59, no. 8, pp. 845–848, 1987.
- [16] L. Hrust, Z. B. Klaić, J. Križan, O. Antonić, and P. Hercog, "Neural network forecasting of air pollutants hourly concentrations using optimised temporal averages of meteorological variables and pollutant concentrations," *Atmospheric Environment*, vol. 43, no. 35, pp. 5588–5596, 2009.
- [17] P. Perez and J. Reyes, "An integrated neural network model for PM_{10} forecasting," *Atmospheric Environment*, vol. 40, no. 16, pp. 2845–2851, 2006.
- [18] D. Z. Antanasijević, V. V. Pocajt, D. S. Povrenović et al., " PM_{10} emission forecasting using artificial neural networks and genetic algorithm input variable optimization," *Science of the Total Environment*, vol. 443, pp. 511–519, 2013.
- [19] A. K. Paschalidou, S. Karakitsios, S. Kleanthous, and P. A. Kassomenos, "Forecasting hourly PM_{10} concentration in Cyprus through artificial neural networks and multiple regression models: implications to local environmental management," *Environmental Science and Pollution Research*, vol. 18, no. 2, pp. 316–327, 2011.
- [20] Q. Bing, D. Qu, X. Chen, F. Pan, and J. Wei, "Short-term traffic flow forecasting method based on lssvm model optimized by GA-PSO hybrid algorithm," *Discrete Dynamics in Nature and Society*, vol. 2018, Article ID 3093596, 10 pages, 2018.
- [21] N. Kumar, A. Chu, and A. Foster, "An empirical relationship between $PM_{2.5}$ and aerosol optical depth in Delhi Metropolitan," *Atmospheric Environment*, vol. 41, no. 21, pp. 4492–4503, 2007.
- [22] J. B. Ordieres, E. P. Vergara, R. S. Capuz, and R. E. Salazar, "Neural network prediction model for fine particulate matter ($PM_{2.5}$) on the US-Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua)," *Environmental Modelling & Software*, vol. 20, no. 5, pp. 547–559, 2005.
- [23] S. Zhu, J. Wang, W. Zhao, and J. Wang, "A seasonal hybrid procedure for electricity demand forecasting in China," *Applied Energy*, vol. 88, no. 11, pp. 3807–3815, 2011.
- [24] A. Vlachogianni, P. Kassomenos, A. Karppinen, S. Karakitsios, and J. Kukkonen, "Evaluation of a multiple regression model for the forecasting of the concentrations of NO_x and PM_{10} in Athens and Helsinki," *Science of The Total Environment*, vol. 409, no. 8, pp. 1559–1571, 2011.
- [25] S. Y. Wang, L. Yu, and K. K. Lai, "Crude oil price forecasting with TEI@I methodology," *Journal of Systems Science & Complexity*, vol. 18, pp. 145–166, 2005.
- [26] J. Heo and D. Kim, "A new method of ozone forecasting using fuzzy expert and neural network systems," *Science of The Total Environment*, vol. 325, no. 1–3, pp. 221–237, 2004.
- [27] L. Yu, S. Wang, and K. K. Lai, "Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm," *Energy Economics*, vol. 30, no. 5, pp. 2623–2635, 2008.

- [28] G. W. Fischer, "Convergent validation of decomposed multi-attribute utility assessment procedures for risky and riskless decisions," *Organizational Behavior and Human Performance*, vol. 18, no. 2, pp. 295–315, 1997.
- [29] D. Kleinmuntz, *Decomposition and the Control of Error in Decision-Analytic Models*, University of Chicago, Chicago, IL, USA, 1990.
- [30] L. Tang, L. Yu, S. Wang, J. Li, and S. Wang, "A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting," *Applied Energy*, vol. 93, pp. 432–443, 2012.
- [31] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [32] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [33] X. Zhang, K. K. Lai, and S.-Y. Wang, "A new approach for crude oil price analysis based on Empirical Mode Decomposition," *Energy Economics*, vol. 30, no. 3, pp. 905–918, 2008.
- [34] Q. Xie, B. Xuan, S. Peng, J. Li, W. Xu, and H. Han, "Bandwidth empirical mode decomposition and its application," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 6, no. 6, pp. 777–798, 2008.
- [35] S. Wang, L. Yu, and L. Tang, "A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in China," *Energy*, vol. 36, no. 11, pp. 6542–6554, 2020.
- [36] D. K. Ghose, S. S. Panda, and P. C. Swain, "Prediction of water table depth in western region, Orissa using BPNN and RBFN neural networks," *Journal of Hydrology*, vol. 394, no. 3–4, pp. 296–304, 2010.
- [37] S. Kang and S. Cho, "Approximating support vector machine with artificial neural network for fast prediction," *Expert Systems with Applications*, vol. 41, no. 10, pp. 4989–4995, 2014.
- [38] G. D. Gennaro, L. Trizio, A. Di et al., "Neural network model for the prediction of PM₁₀ daily concentrations in two sites in the Western Mediterranean," *Science of the Total Environment*, vol. 463–464, pp. 875–883, 2013.
- [39] M. Niu, S. Sun, J. Wu, and Y. Zhang, "Short-term wind speed hybrid forecasting model based on bias correcting study and its application," *Mathematical Problems in Engineering*, vol. 2015, Article ID 351354, 13 pages, 2015.
- [40] U. Farooq, S. Ting-Zhi, Z. San-Yuan, and M. Imran, "Image restoration by using new AGA optimized BPNN," *Procedia Engineering*, vol. 29, pp. 3028–3032, 2012.
- [41] Z. Zhao, J. Wang, J. Zhao, and Z. Su, "Using a Grey model optimized by Differential Evolution algorithm to forecast the per capita annual net income of rural households in China," *Omega*, vol. 40, no. 5, pp. 525–532, 2012.