

Research Article

A Decomposition-Ensemble Approach with Denoising Strategy for PM_{2.5} Concentration Forecasting

Guangyuan Xing,¹ Er-long Zhao,² Chengyuan Zhang,³ and Jing Wu² 

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

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

³School of Economics and Management, Beihang University, Beijing 100191, China

Correspondence should be addressed to Jing Wu; wu_jing_xjt@126.com

Received 7 January 2021; Revised 12 March 2021; Accepted 2 April 2021; Published 19 April 2021

Academic Editor: Giancarlo Consolo

Copyright © 2021 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.

To enhance the forecasting accuracy for PM_{2.5} concentrations, a novel decomposition-ensemble approach with denoising strategy is proposed in this study. This novel approach is an improved approach under the effective “denoising, decomposition, and ensemble” framework, especially for nonlinear and nonstationary features of PM_{2.5} concentration data. In our proposed approach, wavelet denoising approach, as a noise elimination tool, is applied to remove the noise from the original data. Then, variational mode decomposition (VMD) is implemented to decompose the denoised data for producing the components. Next, kernel extreme learning machine (KELM) as a popular machine learning algorithm is employed to forecast all extracted components individually. Finally, these forecasted results are aggregated into an ensemble result as the final forecasting. With hourly PM_{2.5} concentration data in Xi'an as sample data, the empirical results demonstrate that our proposed hybrid approach significantly performs better than all benchmarks (including single forecasting techniques and similar approaches with other decomposition) in terms of the accuracy. Consequently, the robustness results also indicate that our proposed hybrid approach can be recommended as a promising forecasting tool for capturing and exploring the complicated time series data.

1. Introduction

PM_{2.5} refers to particulate matter with a diameter of less than or equal to 2.5 microns in the atmosphere, also known as particulate matter that can enter the lungs [1–3]. Although PM_{2.5} is only a small fraction of the Earth's atmospheric composition, it has an important impact on air quality [4]. It will exert a negative influence on society life, such as increasing the risk of disease and impeding economic development [5–10]. However, more air pollutants come with the development of industry and the increase in the number of fuel-powered cars (Maji et al. 2018). Accurate forecasting of PM_{2.5} concentration has a very important guiding significance for people. It can enable people to make correct decisions, reduce economic losses, and benefit people's life and health. Therefore, it is necessary to predict the PM_{2.5} concentration efficiently and accurately.

At present, many scholars have made a lot of research studies in this field. Accordingly, these studies are roughly divided into four categories: time series model [11–14], econometric model [15–17], artificial intelligence (AI) model [18–20], and the hybrid model [21, 22]. In particular, the time series approach, as the traditional forecasting approach, is often used to predict PM_{2.5} concentration. For example, moving average (MA), autoregressive (AR), autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA) are often used for predicting PM_{2.5} and PM₁₀ concentration [14, 23]. The econometric models are also often used as a forecasting of PM_{2.5} due to its better interpretation [16].

However, due to the related multiple factors' impact, the PM_{2.5} concentration data exist the nonstationary and nonlinear features. Thus, the artificial intelligence methods, which can simultaneously capture and approach the

complicated data feature, have been extensively utilized to forecast the $PM_{2.5}$ [18]. Meanwhile, a variety of machine learning techniques are used for training and modelling the data pattern in this field [24]. Furthermore, the deep learning can well find the mapping relation of $PM_{2.5}$ concentration through nonlinear learning, which may get a good predictive performance.

In addition, considering the limitations of a single model, some scholars proposed the mixed models to make a better forecasting accuracy, which combined the advantages of different models to obtain more stable and accurate results [20, 25, 26]. Similarly, the hybrid approach involves different combination means among the forecasting models. Hybrid integration approach combines the advantages of different models to get better results. In general, it can be divided into finding different components through decomposition and then can be combined with the forecasting model [21, 27–29]. Some scholars have added some parameter optimization algorithms [30, 31].

As mentioned above, although many scholars have adopted different forecasting approaches or proposed different hybrid models in past research, most of the existing combined models cannot comprehensively consider the problems of data noise processing, data feature capture, and forecasting techniques. Particularly, as for data noise processing, many previous studies used different denoising methods for processing the original data, such as singular spectrum analysis [32], Fourier transform, and wavelet transform [33–38]. Among them, the wavelet analysis method can deal the original data more convenient and efficient. Meanwhile, the advantage of wavelet transform method is that the noise is almost completely suppressed and the characteristic peak points reflecting the original signals are well preserved [39, 40]; Dimitriou and Kassomenos 2014. As for data feature capture, the common approaches include EMD family and the VMD approach, and VMD has been widely applied to decompose the time series of different fields, such as wind speed prediction and power load prediction [41–44] and achieved relatively good results. Moreover, Compared with EMD, VMD can solve the divergence problem of EMD endpoint and better extract the components of $PM_{2.5}$ concentration data. As for forecasting techniques, the machine learning approaches are popularly incorporated into the predicted approach for forecasting in the complex system, especially in the field of $PM_{2.5}$ [30]. And, the KELM has the advantages of less training parameters, fast learning speed, and strong generalization ability and especially can get better forecasting results in nonlinear data sets [45].

Considering the noise caused by the way of $PM_{2.5}$ concentration data collection and the denoising characteristics of wavelet transform, firstly we choose wavelet denoising and secondly we choose VMD decomposition due to the complexity of time series data. Finally, considering the validity of KELM's data fitting, KELM was selected as the final prediction model. Therefore, this study proposes a novel hybrid forecasting method, WAV-VMD-KELM, for improving the forecasting accuracy by systematically considering noise processing, data feature capture, and

forecasting techniques. Accordingly, first, wavelet transform was used to denoise the data, then variational mode decomposition (VMD) was adopted to decompose the data, kernel extreme learning machine (KELM) forecasting approach was employed to predict the decomposed components, and finally the results were restored. Therefore, a new hybrid approach denoising-decomposition-ensemble algorithm is proposed based on the comprehensive consideration of the cause of noise and the nonlinear, nonstationarity of $PM_{2.5}$ concentration. And, the two main contributions of this study are as follows:

- (1) In this study, focusing on the problems of noise processing, data feature capture, and forecasting technique selection of the $PM_{2.5}$ concentration, a novel hybrid forecasting approach, i.e., Wav-VMD-KELM, is proposed for improving the forecasting accuracy by denoising, decomposing, individual forecasting, and ensemble results
- (2) A novel decomposition-ensemble approach with denoising strategy is proposed to solve the forecasting approach of hourly data of $PM_{2.5}$ concentration in Xi'an. The experimental results show that the new hybrid approach forecasting method proposed in this study makes a better forecasting performance compared with the benchmarks and can significantly improve the forecasting level of $PM_{2.5}$.

The rest of this paper is summarized as follows. Section 2 introduces the approach used in this research. And, several experiments and analyses and discussion of the forecasting results were performed in Section 3. Finally, Section 4 discusses the conclusions of this study.

2. Methodology

Section 2.1 gives an overview of the proposed decomposition-ensemble approach with denoising strategy, and Sections 2.2–2.4, respectively, describe the related techniques, wavelet denoising, EMD, VMD, and KELM.

2.1. Framework. As shown in Figure 1, the entire process includes wavelet denoising, multiscale analysis, $PM_{2.5}$ concentrations forecasting, and evolution of approaches which are summarized as follows:

- (1) *Wavelet Denoising.* As for the original $PM_{2.5}$ concentration time series, an effective approach of denoising-wavelet denoising is adopted to process the nonstationary time series. Moreover, Section 2.2 gave the detailed process of wavelet denoising.
- (2) *Multiscale Analysis.* VMD is applied to decompose the above denoising sequence into several modes, and these modes reveal the different characteristics which are hidden in the $PM_{2.5}$ concentration time series by their low and high frequency.
- (3) *$PM_{2.5}$ Concentrations Forecasting.* According to the different components, the artificial intelligence tool, KELM, is selected as the forecasting technique for

modelling the corresponding components. Next, every forecasting result of the different components is integrated as the final forecasting results.

2.2. Wavelet Denoising. Wavelet denoising method was first proposed by [34]. It is a kind of nonlinear denoising method, which can be approximately optimal in the sense of minimum mean square error. Figure 2 shows the wavelet denoising method, and each step of wavelet denoising is described in detail as follows.

Wavelet denoising is based on wavelet decomposition, so some concepts of wavelet decomposition are introduced first. The meaning of wavelet transform is one of the functions called basic wavelet transform displacement τ , then at different scales α , with signal to be analyzed $X(t)$ for inner product, that is,

$$WT_x(\alpha, \tau) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{+\infty} x(t) \varphi^*\left(\frac{t-\tau}{\alpha}\right) dt. \quad (1)$$

In the formula, $\alpha > 0$ is called the scale factor, and its effect is on the basic wavelet $\varphi(t)$. Function scaling, t reflects the displacement, it can be positive or negative, and also, α and t are continuous variables.

As mentioned above, the noisy time series data can be expressed as $y(t) = f(t) + \varepsilon(t)$, where $f(x)$ is real signal and $\varepsilon(t)$ is white noise. And, the wavelet transform of both sides of the equation can be obtained as follows:

$$WT_y(a, b) = WT_f(a, b) + WT_\varepsilon(a, b). \quad (2)$$

According to the properties of the wavelet transform, the wavelet transform of the actual measured signal is equal to the sum of the wavelet transforms of multiple signals. After orthogonal wavelet transform, the correlation of signal $y(t)$ can be removed to the greatest extent, and most of the energy can be concentrated on a small number of wavelet coefficients with relatively large amplitude. After the wavelet transform, noise $\varepsilon(t)$ will be distributed on all time axes in various scales, and the amplitude is not very large. Based on this principle, the wavelet coefficients of noise are reduced to the greatest extent in each scale of wavelet transform, and then the signal is reconstructed by using the processed wavelet coefficients, so as to achieve the purpose of suppressing noise [37].

So, the threshold denoising process of one-dimensional signals can be divided into three steps: selection of the appropriate wavelet transform, threshold processing of wavelet coefficients, and wavelet reconstruction.

(1) Selection of the appropriate wavelet transform

Wavelet Decomposition of the Signal. Select a wavelet and determine the level N of a wavelet decomposition and then perform n -layer wavelet decomposition calculation on the signal. Generally, the selection of wavelet basis functions should be considered comprehensively from the aspects of support length, vanishing moment, symmetry, regularity, and similarity. Because the wavelet basis function has its own characteristics in signal processing and no one wavelet

basis function can get the optimal denoising effect for all kinds of signals. In general, Daubechies (dbN) wavelet and Symlets (symN) wavelet are two groups of wavelet bases that are often used in speech denoising. In wavelet decomposition, the selection of decomposition layers is also a very important step. The larger the value is, the more obvious the different characteristics of noise and signal are, and the more conducive it is to the separation of the two. On the other hand, the larger the number of decomposition layers, the larger the reconstructed signal distortion will be, which will affect the final denoising effect to some extent. Therefore, a suitable scale of decomposition is selected after comprehensive consideration in application.

(2) Threshold processing of wavelet coefficients

An important factor directly affecting the denoising effect is the selection of threshold, and different thresholds will have different denoising effects. The threshold function is a rule to correct the wavelet coefficient, and different inverse functions reflect different strategies to deal with the wavelet coefficient. There are two most common types of threshold functions: hard and soft. There is also a Garrote between the soft and hard threshold functions.

The hard threshold function is superior to the soft threshold method in the sense of mean square error. In this study, hard threshold is selected to denoise the data, and the details are as follows: when the absolute value of the wavelet coefficient is less than the given threshold value, let it be 0; If it is greater than the threshold, it stays the same, the mathematical expression is as follows:

$$\omega_\lambda = \begin{cases} \omega, & |\omega| \geq \lambda, \\ 0, & |\omega| < \lambda. \end{cases} \quad (3)$$

Threshold value:

$$\lambda = \sigma \sqrt{2 \log(N)}, \quad (4)$$

where s is the standard deviation of the noise signal and N is the length of the data.

(3) Wavelet reconstruction

The processed wavelet coefficients are reconstructed. According to the low frequency coefficients of the n th layer of wavelet decomposition and the quantized high frequency coefficients of the first to the n th layer, the signal is reconstructed by wavelet. The details are the estimation value of the original signal is obtained by signal reconstruction based on the low frequency coefficients of the n th layer of wavelet decomposition and the high frequency coefficients of the first to the n th layer after processing.

2.3. Variational Mode Decomposition. EMD is a commonly used decomposition method. In this paper, EMD is used as a

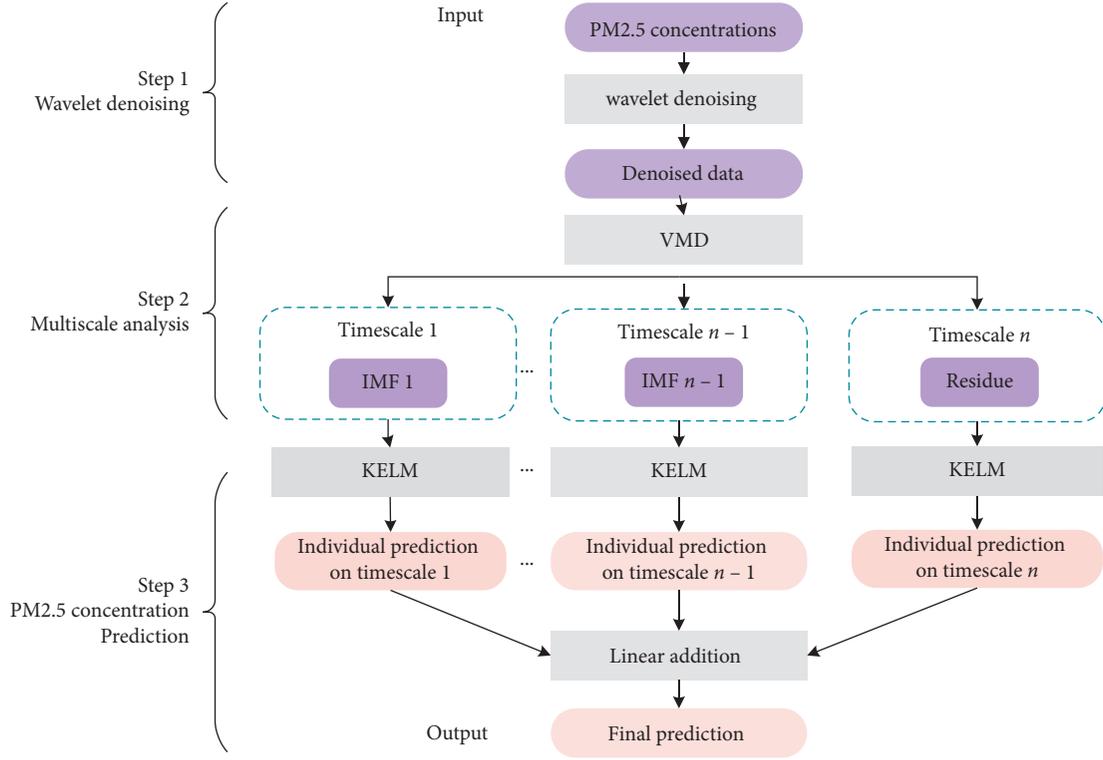


FIGURE 1: The framework of the hybrid approach.

comparison of VMD. EMD is a method proposed in [46]. The main purpose of this algorithm is to decompose signals into characteristic modes. The advantage of this method is that it does not use any defined function as the basis, but generates the inherent mode function adaptively based on the analyzed signal. It can be used to analyze nonlinear and nonstationary signal sequences with high signal-to-noise ratio and good time-frequency focusing.

VMD is a novel nonrecursive and adaptive signal decomposition method that can accommodate much more sampling and noise than some popular decomposition methods such as empirical mode decomposition. The main goal of VMD is to decompose a time series into a discrete set of band-limited modes u_k , where each mode u_k is considered to be compact around a center pulsation ω_k , which is determined during the decomposition.

For example, the time series $f(x)$ is decomposed into a set of modes u_k around a center pulsation ω_k according to the following constrained variational problem [39, 40]:

$$\min_{\mu_k, \omega_k} \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * \mu_k(t) \right] e^{-j\omega_k t} \right\|_2^2, \quad (5)$$

subject to

$$\sum_k u_k = f, \quad (6)$$

where k is the number of modes, δ and $*$ represent the Dirac distribution and the convolution operator. $\{u_k\}$ and $\{\omega_k\}$ represent the set of modes $\{u_1, u_2, \dots, u_k\}$ and the set of center pulsations, respectively.

The above constraint variational problem can be headed with an unconstrained variational problem according to Lagrange multipliers λ , which is given as follows:

$$L(\mu_k, \omega_k, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * \mu_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k \mu_k(t) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_k \mu_k(t) \rangle, \quad (7)$$

where α represents a balance parameter, λ represents the Lagrange multipliers, and $\|f(t) - \sum_k \mu_k(t)\|_2^2$ denotes a quadratic penalty term for the accelerating rate of convergence.

Furthermore, the solution to the equation can be solved by the alternative direction method of multipliers (ADMM) by means of finding the saddle point of the augmented Lagrangian function L in a sequence of iterative

suboptimizations. Therefore, the solutions for u_k , ω_k , and λ can be obtained as follows:

$$\begin{aligned} \widehat{\mu}_k^{n+1}(\omega) &= \frac{\widehat{f}(\omega) - \sum_{i \neq k} \widehat{\mu}_i(\omega) + (\widehat{\lambda}(\omega)/2)}{1 + 2\alpha(\omega - \omega_k)^2}, \\ \omega_k^{n+1} &= \frac{\int_0^\infty \omega |\widehat{\mu}_k(\omega)|^2 d\omega}{\int_0^\infty |\widehat{\mu}_k(\omega)|^2 d\omega}, \\ \widehat{\lambda}^{n+1}(\omega) &= \widehat{\lambda}^n(\omega) + \tau \left(\widehat{f}(\omega) - \sum_k \widehat{\mu}_k^{n+1}(\omega) \right), \end{aligned} \quad (8)$$

where $\widehat{f}(\omega)$, $\widehat{\mu}_i(\omega)$, $\widehat{\lambda}(\omega)$, $\widehat{\lambda}^n(\omega)$, and $\widehat{\mu}_k^{n+1}(\omega)$ represent the Fourier transforms of $f(\omega)$, $\mu_i(\omega)$, $\lambda(\omega)$, $\lambda^n(\omega)$, and $\mu_k^{n+1}(\omega)$, and the number of iterations is n .

The number of modes k needs to be determined, before the VMD method. There is no theory on the optimal selection of the parameter k .

2.4. Kernel Extreme Learning Machine. For a single hidden layer neural network, assume N arbitrary samples (X_i, t_i) , among the rest,

$$\begin{aligned} X_i &= [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n, \\ t_i &= [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m, \end{aligned} \quad (9)$$

and, for a single hidden layer neural network with L hidden layer nodes, it can be expressed as follows:

$$\sum_{i=1}^L \beta_i g(W_i \cdot X_j + b_i) = o_j, \quad (10)$$

where $g(x)$, $j = 1, 2, \dots, N$, is the activation function, $W_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ is the input weight, β_i is the output weights, and b_i is the bias of the i th hidden layer element. $W_i \cdot X_j$ is W_i and X_j do the inner product. The objective of single hidden layer neural network learning is to minimize the error of output, which can be expressed as follows:

$$\sum_{j=1}^N o_j - t_j = 0. \quad (11)$$

That is to say, there exist β_i , W_i , b_i , making

$$\sum_{i=1}^L \beta_i g(W_i \cdot X_j + b_i) = t_j, \quad j = 1, 2, \dots, N. \quad (12)$$

It can be represented by a matrix:

$$H\beta = T, \quad (13)$$

where H is the output of the hidden layer node, β is the output weight, and T is the expected output.

$$\begin{aligned} H(W_1, \dots, W_L, b_1, \dots, b_L, X_1, \dots, X_L) &= \begin{bmatrix} g(W_1 \cdot X_1 + b_1) & \dots & g(W_L \cdot X_1 + b_N) \\ \vdots & \dots & \vdots \\ g(W_1 \cdot X_N + b_1) & \dots & g(W_L \cdot X_N + b_N) \end{bmatrix}_{N \times L}, \\ \beta &= \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}, \\ T &= \begin{bmatrix} T_1^T \\ \vdots \\ T_L^T \end{bmatrix}_{L \times m}. \end{aligned} \quad (14)$$

In order to train single hidden layer neural network, we need to get \widehat{W}_i , \widehat{b}_i , and $\widehat{\beta}_i$, making

$$\left\| H(\widehat{W}_i, \widehat{b}_i) \widehat{\beta}_i - T = \min_{W, b, \beta} H(W_i, b_i) \beta_i - T \right\|, \quad (15)$$

where $i = 1, 2, \dots, L$, which the same thing as minimizing the loss function

$$E = \sum_{j=1}^N \left(\sum_{i=1}^L \beta_i g(W_i \cdot X_j + b_i) - t_j \right)^2. \quad (16)$$

Traditional gradient-based algorithms can be used to solve such problems, but the basic gradient-based learning algorithm needs to adjust all the parameters in the process of

iteration [45]. In the ELM algorithm, once the input weight W_i and the bias b_i of the hidden layer are randomly determined, the output matrix H of the hidden layer is uniquely determined. Training a single hidden layer neural network can be transformed to solve a linear system $H\beta = T$. And, the output weight β can be determined:

$$\widehat{\beta} = H^+ T, \quad (17)$$

where H^+ is the matrix's Moore–Penrose generalized inverse matrix. And, it can be proved that the norm of the solution is minimal and unique.

Kernel function has strong nonlinear mapping ability, which can overcome the dimension disaster. For the problem of linear inseparability, kernel function can be

mapped to high-dimensional space to make it linearly separable [45]. In KELM, the feature map of the hidden layer $h(x)$ remains unknown, replaced by its corresponding kernel function $K(u, v)$. The number of hidden layer nodes L also does not need to be set. The detailed structure of KELM can be referred to Figure 3.

3. Experimental Results and Analysis

PM_{2.5} concentration data are hourly data collected from Xi'an, Shaanxi province, in this research. And, the data collected from (<http://www.cnemc.cn/>). These data serve as our experimental data, and then two popular evaluation criteria of forecasting approaches are used to verify the forecasting result of the hybrid approach. In this section, we verified the effectiveness of our hybrid approach with the experiment of PM_{2.5} forecasting in Xi'an. Also, MATLAB R2018a environment running on Windows 10 with a 64-bit 2.00 GHz AMD Ryzen 5 2500U with Radeon Vega Mobile Gfx was performed in this experiment.

3.1. Data Description. Xi'an, located in the west of China, is a famous tourist city with a long history of civilization, and Xi'an is also an important city for One Belt and One Road. However, the large amount of automobile exhaust and industrial exhaust in Xi'an has caused serious pollution to the air, the haze continues to exist in Xi'an, and the air quality becomes increasingly serious. The quality of air quality not only affects tourists' travel experience but also has a potential impact on people's health. Hourly PM_{2.5} datasets are collected from January 1, 2019, to December 31, 2019, as illustrated in Figure 4. Before prediction, some basic descriptive statistical analysis were used to make a preliminary exploration of the data. For instance, Table 1 lists mean, standard deviation (std.), and min-max of the PM_{2.5} concentration data. As depicted in Table 1 and Figure 4, the PM_{2.5} concentration time series is complex data with nonlinear and nonstationary characteristics, which is embodied in the maximum and minimum values and the continuity of the data. More importantly, we found some data less than 3 and were 10 times different from the adjacent data. For such data, we use the mean value of the data before and after the data. And, the former data from January 1, 2019, 0 o'clock to 24 hours, 24 December 2019, were selected as the training data set in the sample and the remaining data from 0 o'clock December 25, 2019, to December 31, 2019, were taken as the forecast data set out of sample.

In this study, a common time series forecasting method is used. History observations $\{x_{t-1}, x_{t-2}, \dots, x_{t-p}\}$ as inputs to calculate predicted value x_{t+h-1} , where h and p indicate forecasting horizon and lagging order. Figure 5 details the criteria for dividing data by this approach. In this study, lagging order $p = 12$ and forecasting horizon $h = 1$.

3.2. Evaluation Criteria. Next, we use two commonly used evaluation criteria, the mean absolute error (MAE) and the root mean square error (RMSE), to evaluate the forecasting effect of the approach. MAE can be used to reflect the overall

level of forecasting errors; then, RMSE represents the degree among the actual values and the predicted results:

$$\begin{aligned} \text{MAE} &= \frac{1}{N} \sum_{i=1}^N |y_{\text{per}(i)} - y_{\text{act}(i)}|, \\ \text{RMSE} &= \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{\text{per}(i)} - y_{\text{act}(i)})^2}, \end{aligned} \quad (18)$$

where N is the numbers of testing sample, $y_{\text{act}(i)}$ is the i th observed values, and $y_{\text{per}(i)}$ represents the i th forecasting values.

3.3. Benchmarking Models. In the field of deep learning as air pollution forecasting, SVR, BP, ELM, and KELM are popular artificial intelligence techniques forecasting approaches. In addition, the above single forecasting techniques are often combined with denoising and decomposition integration methods to improve the forecasting performance. For the proposed WAV-VMD-KELM approach, four single benchmarks, i.e., SVR, BP, ELM, and KELM, and four denoising-forecasting benchmarks, i.e., WAV-SVR, WAV-BP, WAV-ELM, and WAV-KELM, are all formed for comparing the forecasting accuracy, from the perspectives of the denoising method and the forecasting techniques. In these above hybrid method for ensemble approaches, the first step responds to the denoising process, and the second step for the individual forecasting process. Finally, seven denoising decomposition-ensemble benchmarks are WAV-EMD-SVR, WAV-EMD-BP, WAV-EMD-ELM, WAV-EMD-KELM, WAV-VMD-SVR, WAV-VMD-BP, and WAV-VMD-ELM.

3.4. Empirical Results. A comprehensive comparison of the proposed decomposition-ensemble approach with denoising algorithm with other popular forecasting approaches and similar denoising, ensemble approaches is conducted. For a clear discussion, the comparison results are analyzed from the following four perspectives. First, the four single benchmarking models are compared with each other. Second, the four denoising single benchmarking models are compared with last experiment, to verify the effectiveness of wavelet denoising. Third, the proposed WAV-VMD-KELM is compared with other similar denoising-decomposition-ensemble counterparts, to verify its superiority. Finally, major findings of the empirical study are concluded.

3.4.1. Performance Comparison of Single Methods. Four single benchmarking models, i.e., SVR, BP, ELM, and KELM, are performed to forecast the PM_{2.5} concentrations. Forecasting performances of the four single models are evaluated via MAE and RMSE, as the results reported in Table 2. Compared with the other three artificial intelligence (AI) models, KELM has the optimal effect under the evaluation standard of accuracy.

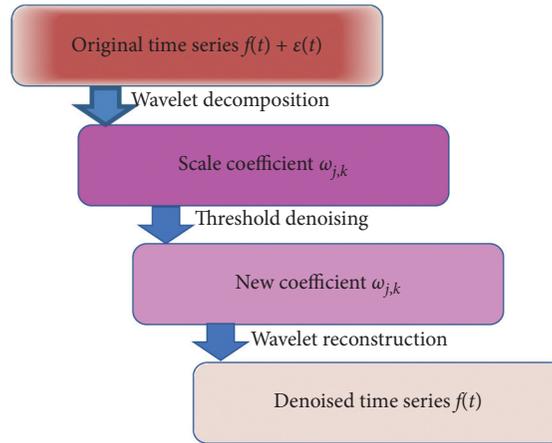


FIGURE 2: Wavelet denoising flow chart.

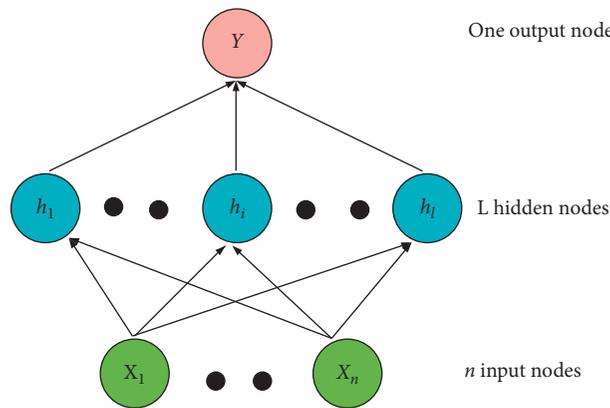


FIGURE 3: The topology diagram of the KELM network.

At the MAE and RMSE evaluation level, it can be found that there is one interesting conclusion. First, it is obvious found that KELM to be the best of all AI models, and the possible reason is that the KELM model might be inapplicable to the complex time series of $PM_{2.5}$ concentration.

3.4.2. Performance Comparison of Denoising Forecasting Model. In the actual process of signal acquisition, the collected signal will inevitably be disturbed by noise or environment and other factors. The main reason lies in sources of the data. First, wavelet denoising is carried out on the data, and then the single approach mentioned above is used to predict the denoised data and then wavelet denoising using the aforementioned single algorithms as individual forecasting tools are then performed and compared with each other.

(1) Denoising results

In the first step of the proposed work, wavelet denoising is employed to remove white noise generated by the collection process in $PM_{2.5}$ data. Figure 6 shows the contrast between the denoising results and the original data. From the results, it can be obviously seen that, after wavelet denoising, the

data becomes flat where the change is sudden. This process of smoothing the data can help restore the true state of the data and also help the models' forecasting.

The parameters corresponding to wavelet denoising are as follows: wavelet basis function: db6; layer: 3; decompose the filter into three layers using hard thresholds and threshold 4.88. RMSE of the denoised data and the original data is 5.88.

(2) Forecasting results

In the second step, the single models, i.e., SVR, BP, ELM, and KELM, are employed as individual forecasting tools. Here, to keep the comparison consistent before and after the experiment. The parameters used are the same as before.

Table 3 shows the comparison results of the four denoising and forecasting in terms of MAE and RMSE. One important conclusion can be obviously seen that the proposed wavelet denoising is verified for $PM_{2.5}$ concentration data forecasting, in terms of accuracy from these results.

In terms of MAE and RMSE, two important findings can be obtained. First, wavelet denoising has obvious

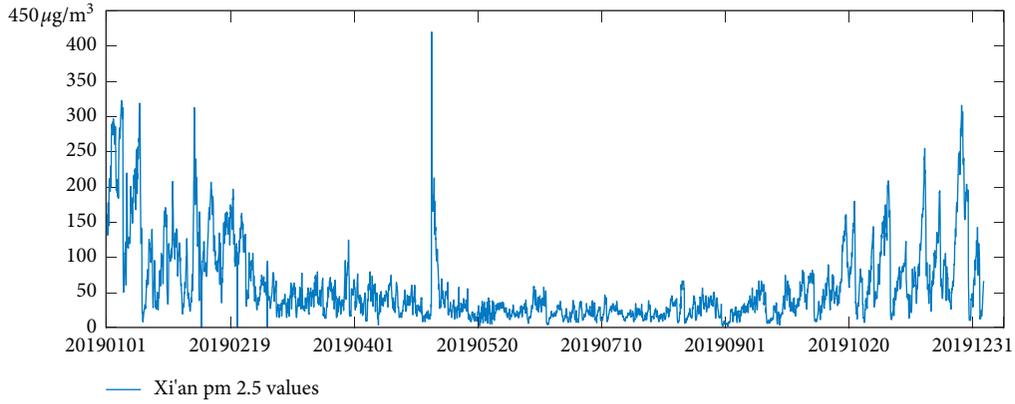


FIGURE 4: The distribution of the original data.

TABLE 1: Basic descriptive statistics of $PM_{2.5}$ concentration data.

| Statistical index | Mean | Std | Min | Max |
|-------------------|-------|-------|------|--------|
| Observations | 58.70 | 55.92 | 1.01 | 420.00 |

improvement for all single models, which also proves that wavelet denoising is effective for this kind of sensor data. The main reason is that $PM_{2.5}$ data is mainly collected through sensors, which may produce some noise due to equipment. Wavelet denoising designed in this study can effectively remove such noise; in this way, the result can be greatly improved. Second, from the perspective of the accuracy of the two evaluation criteria, KELM's forecasting is the best, which may be mainly because the KELM approach can better adapt to such non-stationary time series data.

3.4.3. Performance Comparison of Denoising-Decomposition-Ensemble Models. The seven different denoising-decomposition-ensemble approaches use the denoising-forecasting technology as individual denoising and forecasting tools, and then these approaches compared and performed with each other. In particular, EMD and VMD are selected as two decomposition methods to compare the effectiveness of decomposition.

(1) Decomposition results

In the first step of the previous research, EMD and VMD are used to decompose the denoised $PM_{2.5}$ concentration data. Figures 7 and 8 partially show the decomposition results of EMD and VMD. The IMFs are listed from the highest to the lowest frequencies, while the last one is the residue variable in Figure 7. From the reveal of Figure 7, it is obvious that the complex $PM_{2.5}$ concentration values (also see Figure 5) can be divided via EMD into a number of simple components, and these components further help to make modelling easier. In addition, it is obvious that IMFs 1–3 follow a random walk between 0 and 100, while IMFs 4–6 reveal a regular feature of periodicity with different cycles. Moreover,

MFs 7 and the residue appear smooth central tendencies. Based on these components, some simple models can characterize the characteristics of the data, which helpfully enhances forecasting accuracy. At the same time, in Figure 8, the components are described from the highest to the lowest frequencies in order. From the results of Figure 8, it can be obviously seen that the complex $PM_{2.5}$ concentration data can be divided via VMD into simple components, which simple models can get better results.

(2) Forecasting results

In this step, the denoising-decomposition-ensemble models, WAV-EMD-BP, WAV-EMD-SVR, WAV-EMD-ELM, WAV-EMD-KELM, WAV-VMD-BP, WAV-VMD-SVR, and WAV-VMD-ELM, are used as forecasting tools to respectively model the previous extracted components. The parameters here are the same as before.

Table 4 displays the comparison results of the seven hybrid models in terms of MAE and RMSE. An important conclusion can be obviously seen that the decomposition technique, especially, VMD is verified for $PM_{2.5}$ concentration data forecasting, in terms of accuracy from these results. In terms of MAE and RMSE, three important findings can be obtained.

First, the decomposition technique to improve the forecasting accuracy, which proves that the decomposition of integration technology is effective on such time series forecasting, and the main reason is that the sequence is divided into different components, and each component has a certain regularity. It is under the same approach forecasting model which can be compared with no decomposition of the data to achieve better results. Then VMD effect is better than that of EMD, mainly because the advantage of VMD is that it

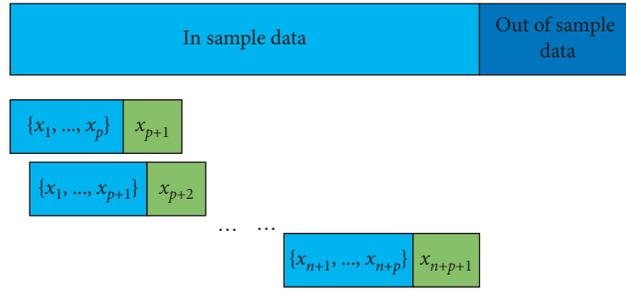


FIGURE 5: Data partitioning diagram.

TABLE 2: The forecasting results of the original data.

| Models | RMSE | MAE |
|--------|-------|------|
| SVR | 8.31 | 4.45 |
| BP | 8.24 | 4.36 |
| ELM | 11.38 | 7.29 |
| KLEM | 8.24 | 4.2 |

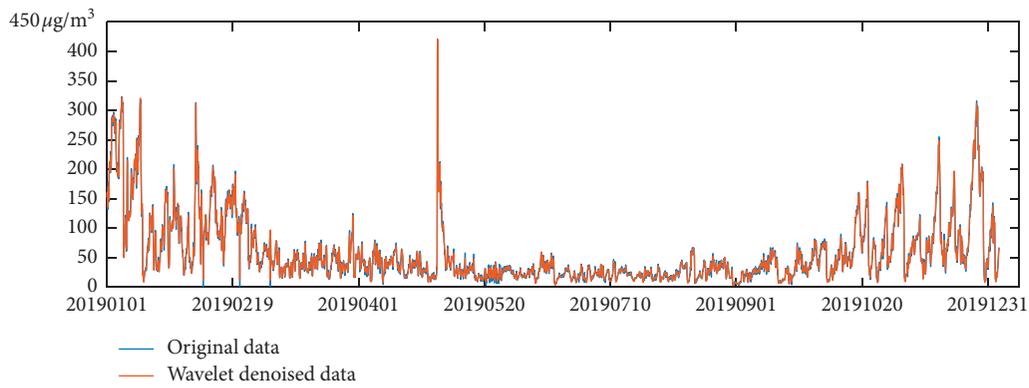


FIGURE 6: Contrast diagram before and after wavelet denoising.

TABLE 3: The forecasting results after wavelet denoising.

| Models | RMSE | MAE |
|----------|-------|------|
| WAV-SVR | 7.13 | 3.77 |
| WAV-BP | 6.83 | 3.64 |
| WAV-ELM | 10.07 | 6.36 |
| WAV-KLEM | 6.96 | 3.58 |

can solve the endpoint divergence problem of EMD, which can find the characteristics of PM_{2.5} concentration data, Thirdly, the novel hybrid approach proposed achieves the best effect in this experiment.

3.4.4. Discussion. The effectiveness of proposed model is discussed in this section.

The experimental results show that, firstly, wavelet denoising is effective in processing the data information collected by the sensor, and the decomposition effect after

EMD is better than that without decomposition, and the decomposition effect of VMD is better than that of EMD.

Besides, it is apparent that, compared with wavelet, the WAV-VMD-KELM approach has lower forecasting errors in terms of MAE and RMSE, which demonstrates the superiority of VMD over EMD. In Figures 9 and 10, it is clear that the developed hybrid approach has the best forecasting result which can explain the effectiveness of the combination method of the proposed approach and its advantage forecasting performance. In order to conveniently show the forecasting availability of the VMD-KELM approach, Figure 11 shows the forecasting curve

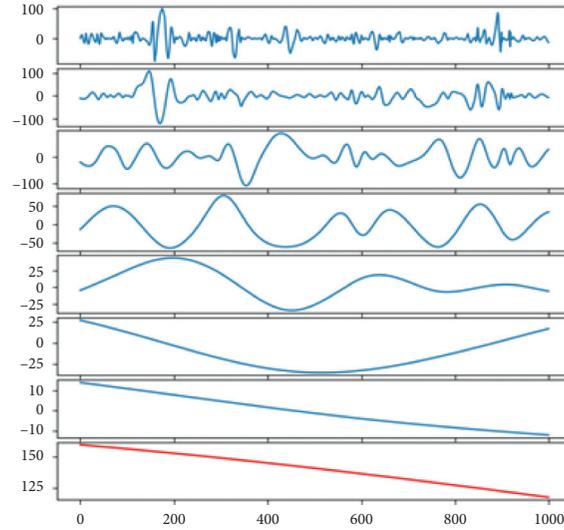


FIGURE 7: The decomposition result diagram of EMD.

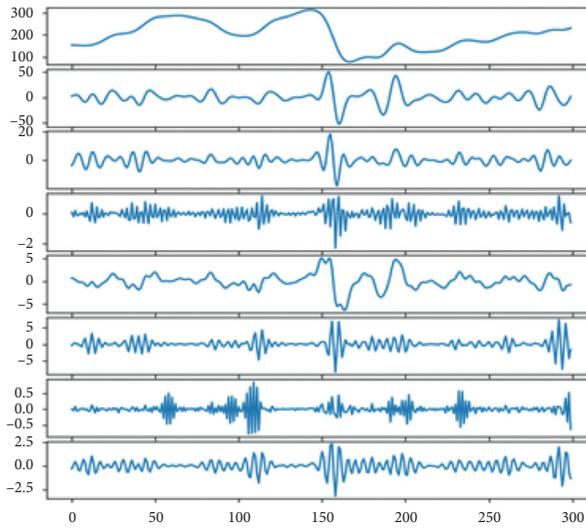


FIGURE 8: The decomposition result diagram of VMD.

TABLE 4: The forecasting results of the original data.

| Models | RMSE | MAE |
|--------------|------|------|
| WAV-EMD-BP | 3.45 | 6.56 |
| WAV-EMD-SVR | 3.8 | 6.76 |
| WAV-EMD-ELM | 5.08 | 8.57 |
| WAV-EMD-KELM | 3.62 | 6.77 |
| WAV-VMD-BP | 2.38 | 3.13 |
| WAV-VMD-SVR | 2.55 | 3.22 |
| WAV-VMD-ELM | 2.97 | 3.82 |
| WAV-VMD-KELM | 2.29 | 3.02 |

of the $PM_{2.5}$ concentration of the WAV-VMD-KELM approach. As shown in Figure 11 and Table 4 that the developed hybrid forecasting approach can get the highest forecasting result among these approaches considered in this research.

It can be summarized that this WAV-VMD-KELM approach is obviously better than the comparison approaches with its smaller error when comparing with other approaches considered in this research.

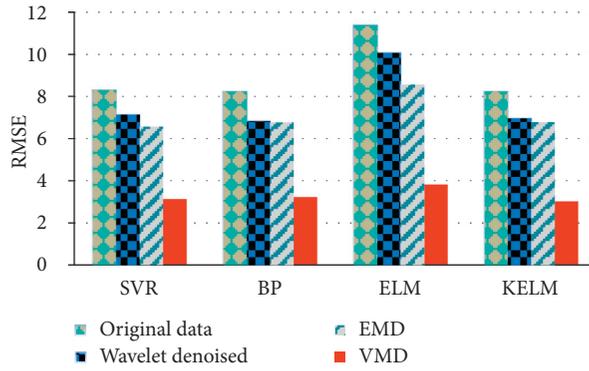


FIGURE 9: RMSE comparison of different approaches.

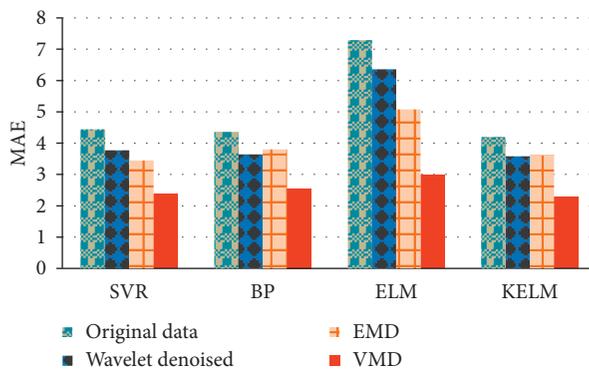


FIGURE 10: MAE comparison of different approaches.

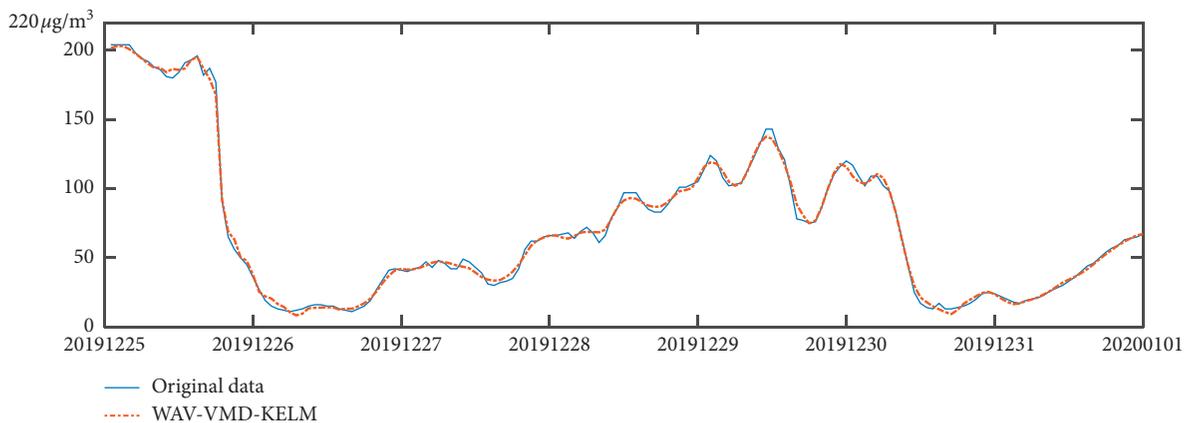


FIGURE 11: Forecasting curve of the $\text{PM}_{2.5}$ concentration.

4. Conclusions

For the past few years, there are many alternative approaches available for $\text{PM}_{2.5}$ concentrations forecasting. And, these works finally contributed to the improvement of forecasting

performances to a certain degree. Yet, there are still some deficiencies in previous articles. For instance, the importance of noise produced by equipment is often neglected and all that. Thus, to solve the problems mentioned above, a new hybrid forecasting method which combines wavelet denoising is

developed. Based on the experimental results and analysis, we can reach the conclusions as follows:

- (1) In those experiments, WAV-VMD-KELM gets the best performance among the comparison approaches
- (2) As a decomposition method, VMD has better predictive ability than EMD considering that the former is effective in capturing the various features hidden in the original datasets
- (3) Wavelet denoising can improve the accuracy of forecasting which indicates that there is necessity to remove the noise specifically when the data is obtained by the sensor

Overall, with more accurate prediction, this proposed hybrid approach presents a superior performance beyond other alternative approaches, offering a novel and feasible method in the field of $PM_{2.5}$ concentrations forecasting. Furthermore, this new and viable option can also be applied to many other complex areas, such as tourism demand forecasting, wind speed forecasting, economic growth forecasting, product sales forecasting, and traffic flow forecasting, and others. In fact, it is worth noticing that there still exist some limitations. First, this research only puts $PM_{2.5}$ time series into consideration and does not include other influencing factors. Another shortcoming is that this paper adopts single objective optimization algorithms only for forecasting $PM_{2.5}$ time series without trying the multi-objective versions. Hence, future research on $PM_{2.5}$ time series forecasting should highlight the following aspects, that is, the usage of other possible factors (meteorological factors in particular) as well as the application of the multiobjective optimization algorithm. In addition, it can also be practicable to establish deep learning-enabled approaches, which are more effective and have the potential to be another significant research direction for future work [12, 47].

Data Availability

$PM_{2.5}$ concentration data are hourly data collected from China National Environmental Monitoring Centre (CNEMC, <http://www.cnemc.cn/>) in this research.

Conflicts of Interest

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

Acknowledgments

This research work was partly supported by the Fundamental Research Funds for the Central Universities under grant no. xpt012020022, the National Natural Science Foundation of China under grant no. 71904153, and the Project Funded by China Postdoctoral Science Foundation under grant no. 2018M53598.

References

- [1] Feng, D. Gao, F. Liao, F. Zhou, and X. Wang, "The health effects of ambient $PM_{2.5}$ and potential mechanisms," *Ecotoxicology and Environmental Safety*, vol. 128, pp. 67–74, 2016.
- [2] X. Lu, C. Lin, W. Li et al., "Analysis of the adverse health effects of $PM_{2.5}$ from 2001 to 2017 in China and the role of urbanization in aggravating the health burden," *Science of The Total Environment*, vol. 652, pp. 683–695, 2019.
- [3] X. Lu, T. Yao, J. C. H. Fung, and C. Lin, "Estimation of health and economic costs of air pollution over the Pearl River Delta region in China," *Science of The Total Environment*, vol. 566–567, pp. 134–143, 2016.
- [4] L. Chen, M. Shi, S. Li et al., "Quantifying public health benefits of environmental strategy of $PM_{2.5}$ air quality management in Beijing-Tianjin-Hebei region, China," *Journal of Environmental Sciences*, vol. 57, pp. 33–40, 2017.
- [5] Y. Shi, L. Kang, Y. Wang, N.-N. Zhang, and M.-T. Ju, "Health loss attributed to $PM_{2.5}$ pollution in China's cities: economic impact, annual change and reduction potential," *Journal of Cleaner Production*, vol. 217, pp. 284–294, 2019.
- [6] P. Huang, J. Zhang, Y. Tang, and L. Liu, "Spatial and temporal distribution of $PM_{2.5}$ pollution in Xi'an City, China," *International Journal of Environmental Research and Public Health*, vol. 12, no. 6, pp. 6608–6625, 2015.
- [7] J. Sinha and N. Kumar, "Mortality and air pollution effects of air quality interventions in Delhi and Beijing," *Frontiers in Environmental Science*, vol. 7, 2019.
- [8] C. Carlsten, "Air pollution and children's respiratory health," *Canadian Respiratory Journal*, vol. 22, no. 5, p. 256, 2015.
- [9] S. Yang, D. Fang, and B. Chen, "Human health impact and economic effect for $PM_{2.5}$ exposure in typical cities," *Applied Energy*, vol. 249, pp. 316–325, 2019.
- [10] H. Yin, M. Pizzol, and L. Xu, "External costs of $PM_{2.5}$ pollution in Beijing, China: uncertainty analysis of multiple health impacts and costs," *Environmental Pollution*, vol. 226, pp. 356–369, 2017.
- [11] M. d. C. Bas, J. Ortiz, L. Ballesteros, and S. Martorell, "Evaluation of a multiple linear regression model and SAR-IMA model in forecasting 7 Be air concentrations," *Chemosphere*, vol. 177, pp. 326–333, 2017.
- [12] I. Mura, J. F. Franco, L. Bernal, N. Melo, and R. Akhavan-Tabatabaei, "A decade of air quality in Bogotá: a descriptive analysis," *Frontiers in Environmental Science*, vol. 8, p. 85, 2020.
- [13] L. Jian, Y. Zhao, Y.-P. Zhu, M.-B. Zhang, and D. Bertolatti, "An application of ARIMA model to predict submicron particle concentrations from meteorological factors at a busy roadside in Hangzhou, China," *Science of The Total Environment*, vol. 426, pp. 336–345, 2012.
- [14] L. Zhang, J. Lin, R. Qiu et al., "Trend analysis and forecast of $PM_{2.5}$ in Fuzhou, China using the ARIMA model," *Ecological Indicators*, vol. 95, pp. 702–710, 2018.
- [15] B. Lin, W. G. Cobourn, and Y. Bai, "Development of non-linear empirical models to forecast daily $PM_{2.5}$ and ozone levels in three large Chinese cities," *Atmospheric Environment*, vol. 147, pp. 209–223, 2016.
- [16] B. Wang, Q. Jiang, and P. Jiang, "A combined forecasting structure based on the L1 norm: application to the air quality,"

- Journal of Environmental Management*, vol. 246, pp. 299–313, 2019.
- [17] Y. Wang, H. Wang, S. Chang, and A. Avram, “Prediction of daily $PM_{2.5}$ concentration in China using partial differential equations,” *PLoS One*, vol. 13, Article ID e0197666, 2018.
 - [18] S. M. Cabaneros, J. K. Calautit, and B. R. Hughes, “A review of artificial neural network models for ambient air pollution prediction,” *Environmental Modelling & Software*, vol. 119, pp. 285–304, 2019.
 - [19] U. Pak, J. Ma, U. Ryu et al., “Deep learning-based $PM_{2.5}$ prediction considering the spatiotemporal correlations: a case study of Beijing, China,” *Science of The Total Environment*, vol. 699, Article ID 133561, 2020.
 - [20] W. Sun and J. Sun, “Daily $PM_{2.5}$ concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm,” *Journal of Environmental Management*, vol. 188, pp. 144–152, 2017.
 - [21] H. Liu, Y. Xu, and C. Chen, “Improved pollution forecasting hybrid algorithms based on the ensemble method,” *Applied Mathematical Modelling*, vol. 73, pp. 473–486, 2019.
 - [22] J. Wang, P. Du, Y. Hao, X. Ma, T. Niu, and W. Yang, “An innovative hybrid model based on outlier detection and correction algorithm and heuristic intelligent optimization algorithm for daily air quality index forecasting,” *Journal of Environmental Management*, vol. 255, Article ID 109855, 2020.
 - [23] S. G. Ilieva and A. Ivanov, “Assaying SARIMA and generalised regularised regression for particulate matter PM_{10} modelling and forecasting,” *International Journal of Environment and Pollution*, vol. 66, no. 1–3, pp. 41–62, 2019.
 - [24] F. Biancofiore, M. Verdecchia et al., “Recursive neural network model for analysis and forecast of PM_{10} and $PM_{2.5}$,” *Atmospheric Pollution Research*, vol. 8, no. 4, pp. 652–659, 2017.
 - [25] P. Busilacchio, H. Zhang, Z. Qin, and G. Zhang, “A novel hybrid-Garch model based on ARIMA and SVM for $PM_{2.5}$ concentrations forecasting,” *Atmospheric Pollution Research*, vol. 8, no. 5, pp. 850–860, 2017.
 - [26] Y. Zhou, F.-J. Chang, L.-C. Chang, I.-F. Kao, Y.-S. Wang, and C.-C. Kang, “Multi-output support vector machine for regional multi-step-ahead $PM_{2.5}$ forecasting,” *Science of The Total Environment*, vol. 651, pp. 230–240, 2019.
 - [27] Z. Wang, L. Chen, Z. Ding, and H. Chen, “An enhanced interval $PM_{2.5}$ concentration forecasting model based on BEMD and MLPI with influencing factors,” *Atmospheric Environment*, vol. 223, Article ID 117200, 2020b.
 - [28] Q. Wu and H. Lin, “A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors,” *Science of The Total Environment*, vol. 683, pp. 808–821, 2019.
 - [29] Q. Zhou, H. Jiang, J. Wang, and J. Zhou, “A hybrid model for $PM_{2.5}$ forecasting based on ensemble empirical mode decomposition and a general regression neural network,” *Science of The Total Environment*, vol. 496, pp. 264–274, 2014.
 - [30] D. Wang, S. Wei, H. Luo, C. Yue, and O. Grunder, “A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine,” *Science of The Total Environment*, vol. 580, pp. 719–733, 2017a.
 - [31] S. Zhu, X. Lian, L. Wei et al., “ $PM_{2.5}$ forecasting using SVR with PSO-GSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors,” *Atmospheric Environment*, vol. 183, pp. 20–32, 2018.
 - [32] J. Lian, Y. Termonia, and J. Deltour, “Smoothing and differentiation of data by simplified least square procedure,” *Analytical Chemistry*, vol. 44, no. 11, pp. 1906–1909, 1972.
 - [33] R. N. Bracewell, “The Fourier transform and its applications,” *American Journal of Physics*, vol. 34, no. 8, p. 712.
 - [34] D. L. Kahn and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
 - [35] X. Feng, Q. Li, Y. Zhu, J. Hou, L. Jin, and J. Wang, “Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation,” *Atmospheric Environment*, vol. 107, pp. 118–128, 2015.
 - [36] M. Lasiewicz, M. Bogusz, and M. Kosla, “Universal system for forecasting changes in PM_{10} and $PM_{2.5}$ particulate matter air pollution concentration,” *Int J Environ Pollut*, vol. 56, pp. 94–108, 2015.
 - [37] R. Sircar, “An introduction to wavelets and other filtering methods in finance and economics *Waves in Random Media*,” vol. 12, p. 399, 2002.
 - [38] B. Vidakovic and C. B. Lozoya, “On time-dependent wavelet denoising,” *IEEE Transactions on Signal Processing*, vol. 46, no. 9, pp. 2549–2554, 1998.
 - [39] A. A. Abdoos, “A new intelligent method based on combination of VMD and ELM for short term wind power forecasting,” *Neurocomputing*, vol. 203, pp. 111–120, 2016.
 - [40] S. Lahmiri, “A variational mode decomposition approach for analysis and forecasting of economic and financial time series,” *Expert Systems with Applications*, vol. 55, pp. 268–273, 2016.
 - [41] H. Liu, X. Mi, and Y. Li, “Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, lstm network and elm,” *Energy Conversion and Management*, vol. 159, pp. 54–64, 2018.
 - [42] Z. Zhang, W. C. Hong, and J. Li, “Electric load forecasting by hybrid self-recurrent support vector regression model with variational mode decomposition and improved cuckoo search algorithm,” *IEEE Access*, vol. 99, p. 1, 2020.
 - [43] Wei, Y. Cao and Chen, Applications of variational mode decomposition in seismic time-frequency analysis,” *Geophysics*, vol. 81, pp. V365–V378, 2016.
 - [44] W. Liu, S. Cao, Z. Wang, X. Kong, and Y. Chen, “Spectral decomposition for hydrocarbon detection based on VMD and teager-Kaiser energy,” *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 4, pp. 539–543, 2017.
 - [45] G.-B. Huang, “An insight into extreme learning machines: random neurons, random features and kernels,” *Cognitive Computation*, vol. 6, no. 3, pp. 376–390, 2014.
 - [46] 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.
 - [47] R. Shen, F. Yannick, B. Murielle, M. Reto, and R. Valentin, “Exposure to fine particulate matter leads to rapid heart rate variability changes,” *Frontiers in Environmental Science*, vol. 6, p. 2, 2018.