MultiStep Ahead Forecasting for Hourly PM10 and PM2.5 Based on Two-Stage Decomposition Embedded Sample Entropy and Group Teacher Optimization Algorithm

Feng Jiang 1,*, Yaqian Qiao 1, Xuchu Jiang 1,2 and Tianhai Tian 3

1 School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan 430073, China; yqqiao@stu.zuel.edu.cn (Y.Q.); xuchujiang@zuel.edu.cn (X.J.)
2 Hubei Province Key Laboratory of Systems Science in Metallurgical Process, Wuhan University of Science and Technology, Wuhan 430081, China
3 School of Mathematical Science, Monash University, Melbourne Clayton 3800, Australia; tianhai.tian@monash.edu
* Correspondence: fjiang@zuel.edu.cn

Abstract: The randomness, nonstationarity and irregularity of air pollutant data bring difficulties to forecasting. To improve the forecast accuracy, we propose a novel hybrid approach based on two-stage decomposition embedded sample entropy, group teaching optimization algorithm (GTOA), and extreme learning machine (ELM) to forecast the concentration of particulate matter (PM10 and PM2.5). First, the improvement complementary ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) is employed to decompose the concentration data of PM10 and PM2.5 into a set of intrinsic mode functions (IMFs) with different frequencies. In addition, wavelet transform (WT) is utilized to decompose the IMFs with high frequency based on sample entropy values. Then the GTOA algorithm is used to optimize ELM. Furthermore, the GTOA-ELM is utilized to predict all the subseries. The final forecast result is obtained by ensemble of the forecast results of all subseries. To further prove the predictable performance of the hybrid approach on air pollutants, the hourly concentration data of PM2.5 and PM10 are used to make one-step-, two-step- and three-step-ahead predictions. The empirical results demonstrate that the hybrid ICEEMDAN-WT-GTOA-ELM approach has superior forecasting performance and stability over other methods. This novel method also provides an effective and efficient approach to make predictions for nonlinear, nonstationary and irregular data.

Keywords: ICEEMDAN; wavelet transform; group teaching optimization algorithm; extreme learning machine; sample entropy

1. Introduction

The prediction of future events from the noisy and nonstationary time series data is a challenging problem. This type of problem has been found in a wide range of research areas, including finance, biological sciences, environment, ecology, and engineering [1–5]. Although a number of statistical inference methods and machine learning algorithms have been proposed to make predictions, it is still a substantial challenge to make more accurate and reliable forecasting for time series data with high volatility.

The forecasting of atmospheric pollution is one of the important problems in the analysis of noise and nonstationary time series. In recent years, atmospheric pollution has become one of the most important issues all over the world [6,7]. The particulate matter of aerodynamic diameter less than or equal to 2.5 and 10 microns is the major air pollutant, which also causes significant influence on human health and daily activities [8]. Therefore, a precise short-term forecasting model should be developed to enhance the prediction performance of air-pollutant concentrations, which not only can give valuable support for
decision-making of relevant departments, but also can provide helpful information for air-quality monitoring systems. That will have great significance on guiding daily activities of people for alleviating public health risks based on the real-time prediction. There is an increasing demand to develop data-driven algorithms based on hourly observations time series for making reliable short-term forecasting of particulate matter (PM) concentrations [9,10]. However, due to the nonstationary, nonlinear and complex characteristic of PM2.5 and PM10, it is difficult to make a precise forecasting for it.

In the previous study, the forecasting approach for PM10, PM2.5 and other air-pollutant concentration forecasting can be divided into traditional statistical approach, machine learning approach, and hybrid approach. The traditional statistical approaches, such as multiple linear regression (MLR) [11] and autoregressive moving integrated average (ARIMA) [12] are usually employed to predict the concentration of air pollutants. With the maturity and development of data mining technology, machine learning approaches such as artificial neural network (ANN) [13–15], support vector machine (SVM), and intelligent algorithm [16–18] have been widely used for forecasting. Park et al. [19] presented an ANN model to predict the PM10 concentration in six subway stations based the information of outdoor PM10 concentration, the number of subway trains, and ventilation rate; the forecasting accuracy was 67–80%.

In addition to the statistical and machine learning approach, the hybrid models based on data preprocessing have become more and more popular, especially for the time series with high volatility and nonlinearity characteristics. Mahajan et al. [20] introduced wavelet transform (WT) to decompose original PM2.5 concentration. The autoregressive integrated moving average (ARIMA) and neural network autoregression (NNAR) are utilized to forecast the subseries. Jiang et al. [21] proposed a novel hybrid approach that combines wavelet packet decomposition (WPD), improved pigeon-inspired optimization algorithm and extreme learning machine (ELM); for the forecasting of AQI, the prediction performance of a single model has been greatly improved. In addition, Wang et al. [22] developed a new hybrid approach based on secondary decomposition and the back propagation neural network (BPNN) model optimized by differential evolution (DE) algorithm. WT is first used to decompose the original time series of PM2.5 concentration into several subseries, and variational mode decomposition (VMD) is further utilized to decompose the subseries; the DE algorithm is employed to optimize the weights and thresholds of the BPNN. Yang et al. [23] applied the complementary ensemble empirical mode decomposition (CEEMD) to decompose the concentration data of PM2.5 and PM10; then the back propagation neural network BPNN, extreme learning machine (ELM) and double exponential smoothing (DES) are used to predict the subseries, respectively.

Using single decomposition algorithm can greatly enhance the forecasting performance. However, it still has some problems, such as it cannot completely extract the future of nonstationarity and irregular time series probably. There may be some complicated patterns hidden in the result of first decomposition. It is also not very efficient for the secondary decomposition in that all the subseries obtained from first decomposition are decomposed again. Consequently, this paper proposed a novel hybrid approach based on two-stage decomposition with sample entropy embedded in. Sample entropy is utilized to select the subseries with high frequency, and wavelet transform (WT) is employed to conduct the secondary decomposition of the selected subseries. The novel metaheuristic algorithm of group teaching optimization algorithm (GTOA) is introduced to optimize the original weights and thresholds of ELM. The improved ELM approach of GTOA-ELM is utilized to forecast all the subseries.

The remainder of this paper is organized as follows. Related methodologies used in this paper are detailed in Section 2. Section 3 shows the empirical outcomes and the comparative results of the hybrid forecasting approach. Horizontal precision, directional precision, and stability are also examined by using statistical indicators and statistical tests. Finally, the corresponding conclusion of this paper is given in Section 4.
2. Methods

2.1. ICEEMDAN

Empirical mode decomposition (EMD) is a kind of adaptive data analysis method utilized for the nonlinear and nonstationary signal that was proposed by Huang et al. [24], and it can decompose a complex signal into several intrinsic mode functions that contain the local characteristics of the original signals at different time scales. Based on EMD, ensemble empirical mode decomposition (EEMD) [25] is extended from EMD by adding white Gaussian noise to solve the problem of mode mixture in EMD. The intrinsic mode functions are obtained by EMD decomposition of a signal added with Gaussian white noise. However, the reconstruction error cannot be eliminated absolutely with limited times of iterations. In order to overcome the drawback of EEMD, the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [26] was developed. Compared with EEMD, CEEMDAN is able to reduce the reconstruction error under the low computational cost. Nevertheless, the modes of CEEMDAN still have some residual noise. The improved CEEMDAN (ICEEMDAN) was developed by Colominas et al. [27] to achieve the perfect reconstruction and avoid spurious modes, which can provide more physical meaning to the IMF and have better convergence and stability [28]. The ICEEMDAN technique is used to decompose the time series $G(t)$ into a number of IMFs and one residual component as follows:

1. The EMD algorithm is utilized to calculate the local means of n realizations $G_i(t)$:

$$G_i(t) = G(t) + A_0 I_1(\kappa_i(t))$$

where the operator $I_j$ is the jth mode of the time series $G(t)$ decomposed by the EMD, $\kappa_i$ represents Gaussian white noise with variance ranging from 0 to 1. $A = \phi_0 \text{std}(G(t))/\text{std}(I_1(\kappa_i(t)))$, which is utilized to remove the fraction of noise energy action at the stage of algorithm begins. std represents the standard deviation.

2. Calculate the first residue $r_1(t)$, and the first mode can be obtained based on it. The operator $E(\cdot)$ produces the local mean.

$$r_1(t) = \langle E(G_i(t)) \rangle$$

$$\text{IMF}_1(t) = G(t) - r_1(t)$$

3. $r_2(t)$ can be estimated by Equation (4), and then compute the second mode $\text{IMF}_2(t)$.

$$r_2(t) = \langle E(r_1(t) + A_1 I_2(\kappa_i(t))) \rangle$$

$$\text{IMF}_2(t) = r_1(t) - r_2(t)$$

4. IMF$_l(t)$ and $r_l(t)$ for $l = 3, 4, \ldots, L$ can be calculated as follows:

$$r_l(t) = E(r_{l-1}(t) + A_{l-1} I_l(\kappa_i(t)))$$

$$\text{IMF}_l(t) = r_{l-1}(t) - r_l(t)$$

5. Step 4 is repeated until obtain all the IMFs.

2.2. Wavelet Transform (WT)

Wavelet transform (WT) is a kind of powerful signal processing technique. WT can be utilized to decompose a nonlinear and nonstationary signal into several components at different resolution levels, including a low-frequency approximation subset and several high-frequency subseries. The low-frequency approximation subset that can reflect the tendency of the original signal and those high-frequency subseries represent the components of turbulence and noise [29]. As an effective data preprocessing method, it can reduce the
difficulty of feature learning and enhance the prediction performance. Thus, WT has been widely used in many fields [30].

2.3. Sample Entropy

Sample entropy (SE) is an algorithm to measure the complexity of time series, which was proposed by Richman and Moorman [31]. The greater the value of sample entropy, the higher the complexity of time series. As an improvement of approximate entropy, SE is independent of data length and has better consistency. The value of sample entropy can be represented as \( SE(m, r, N) \), \( r \) is the similarity tolerance and the value of it was set to \( r = 0.2 \) standard deviations, the reconstruction dimension \( m \) was set as 2 in this paper. The steps to calculate the sample entropy are shown below.

Step 1: For a given time series, and m-dimensional vectors is comprised based on it.

\[
Y_i = [y_i, y_{i+1}, \cdots, y_{i+m-1}], i = 1, 2, \cdots, N - m + 1
\]  

(8)

Step 2: The absolute distance of corresponding elements between vectors is calculated, and define distance between \( Y_i \) and \( Y_j \).

\[
D_m(Y_i, Y_j) = \max\{|y_{i+l}, y_{j+l}|, l = 1, 2, \cdots, m + 1\}
\]  

(9)

Step 3: For a given threshold \( r \), count the total number of \( D_m(Y_i, Y_j) < r \), calculate the ratio \( B_i^m(r) \) and the mean value of \( B_i^m(r) \)

\[
B_i^m(r) = \frac{1}{N - m} \text{num}\{D_m(Y_i, Y_j) < r\}
\]  

(10)

\[
B^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} B_i^m(r)
\]  

(11)

Step 4: Set \( m \) to \( m+1 \), repeat steps1-3, and then obtain the mean value of \( B^{m+1}(r) \).

\[
B^{m+1}(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} B_i^{m+1}(r)
\]  

(12)

Step 5: Calculate SE value.

\[
SE(m, r) = \lim_{N \to \infty} \{ -\ln\left(\frac{B^{m+1}(r)}{B^m(r)}\right) \}
\]  

(13)

Step 6: For a finite time series, the result of SE value can be estimated through Equation (14).

\[
SE(m, r, N) = -\ln\left(\frac{B^{m+1}(r)}{B^m(r)}\right)
\]  

(14)

2.4. Group Teaching Optimization Algorithm (GTOA)

The group teaching optimization algorithm (GTOA) proposed by Zhang and Jin [32] in 2020 is a novel metaheuristic optimization algorithm that was inspired by the group teaching mechanism. Less control of parameters is the best advantage of the GTOA algorithm. In addition to the basic parameter of population size and termination criteria, it does not need any extra control parameters. Furthermore, the GTOA algorithm has the advantages of fast convergence speed and strong global optimal searching ability. The GTOA algorithm mainly contains four phases, including ability grouping phase, teacher phase, student phase and teacher allocation phase. The detailed description of four phases is given as follows [32].

In the phase of ability grouping, the knowledge of whole class is following the normal distribution. Equation (15) shows the knowledge distribution of students. Here \( x \) is the
knowledge of the students, $\mu$ is the mean value of the whole class, and $\sigma$ is standard deviation of knowledge among students. In the GTOA algorithm, all the students can be divided into two groups based on the ability to accept knowledge: outstanding group and average group.

\[
pop(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)
\]

(15)

In the teacher phase, students acquire knowledge from the teacher. And the students of the outstanding group and the average group update their knowledge by Equations (16) and (17), respectively.

\[
X_{t,j}^{l+1} = X_{t,j}^l + a \times (T^l - F \times (\beta \times M^l + \gamma \times X_{t,j}^l)) \quad \text{(16)}
\]

\[
X_{t,j}^{l+1} = X_{t,j}^l + 2 \times r \times (T^l - X_{t,j}^l) \quad \text{(17)}
\]

where $X_{t,j}^{l+1}$ represents the knowledge of student $j$ learned from the teacher at time $l + 1$, $X_{t,j}^l$ is the knowledge of student $j$ at time $l$, $T^l$ is the knowledge of teacher at time $l$, and the $M^l$ is the average knowledge of the group at time $l$. $a$, $\beta$, $\gamma$, and $\lambda$ are the random numbers between 0 and 1. Then, based on the following equation, the student should decide whether to accept the knowledge learned from the teacher.

\[
X_{t,j}^{l+1} = \begin{cases} 
X_{t,j}^l \times f(X_{t,j}^{l+1}) < f(X_t^l) \\
X_{t,j}^l, f(X_{t,j}^{l+1}) \geq f(X_t^l) 
\end{cases} \quad \text{(18)}
\]

In the student phase, students are able to gain the new knowledge through self-learning and cooperation with other students. The updated knowledge is expressed as follows.

\[
X_{s,j}^{l+1} = \begin{cases} 
X_{s,j}^l + a \times (X_{s,j}^{l+1} - X_{s,j}^l) + b \times (X_{s,j}^{l+1} - X_{t,j}^{l+1}), f(X_{s,j}^{l+1}) < f(X_{s,j}^{l+1}) \\
X_{s,j}^l - a \times (X_{s,j}^{l+1} - X_{s,j}^l) + b \times (X_{s,j}^{l+1} - X_{t,j}^{l+1}), f(X_{s,j}^{l+1}) \geq f(X_{s,j}^{l+1}) 
\end{cases} \quad \text{(19)}
\]

where $X_{s,j}^{l+1}$ is the knowledge of student $j$ at time $l + 1$ through learning from the student phase, $a$ and $b$ are the random numbers between 0 and 1. Similar to the teacher phase, the student should decide whether to accept the knowledge gained through the student phase according to the Equation (20).

\[
X_{j}^{l+1} = \begin{cases} 
X_{j}^{l+1} \times f(X_{j}^{l+1}) < f(X_{s,j}^{l+1}) \\
X_{s,j}^{l+1}, f(X_{s,j}^{l+1}) \geq f(X_{s,j}^{l+1}) 
\end{cases} \quad \text{(20)}
\]

In the teacher allocation phase, the teacher allocation can be expressed by the following equation for accelerating the convergence as well as avoiding local minimum.

\[
T^l = \begin{cases} 
X_{1}^{l} \times f(X_{1}^{l}) < f(X_{1}^{l} + X_{2}^{l} + X_{3}^{l}) \\
X_{1}^{l} + X_{2}^{l} + X_{3}^{l}, f(X_{1}^{l}) \geq f(X_{1}^{l} + X_{2}^{l} + X_{3}^{l}) 
\end{cases} \quad \text{(21)}
\]

where $X_{1}^{l}$, $X_{2}^{l}$, and $X_{3}^{l}$ denote the first, second, and third best students.

2.5. Modified Extreme Learning Machine

The extreme learning machine (ELM), proposed by Huang et al. [33] is a single-hidden layer feedforward neural network (SLFN) with an effective learning algorithm. The ELM is able to choose the weights and thresholds randomly and determine the output weights analytically, and thus has the advantage of fast training speed. Furthermore, it has better nonlinear mapping capability and is widely used in the field of prediction. However, due to the random selection of weights and thresholds, the generalization ability and stability may be influenced. In this paper, the ELM is modified by using the GTOA algorithm to
optimize the initial weights and thresholds. The mean absolute error (MAE) of ELM is set as the objective function of the GTOA algorithm. And the GTOA-ELM is utilized to enhance the forecasting performance, stability, and generalization ability of the ELM.

\[
\text{MAE} = \frac{1}{T} \sum_{t=1}^{T} |\hat{g}(t) - g(t)| \tag{22}
\]

2.6. Multistep Prediction

The PM10 and PM2.5 concentrations are no doubt affected by the concentrations of previous several days greatly. Thus, many forecasting models utilized the data of previous several days as input and predicted the latter one [22]. Consequently, in this paper, the n-step ahead forecasting formula is presented as Equation (23).

\[
\hat{y}_{l+n} = f(y_{l-h}, y_{l-h+1}, \ldots, y_{l-1}, y_{l}), l = 1, 2, \ldots, L \tag{23}
\]

where \(\hat{y}_{l+n}\) is the forecasting result of original time series at time \(l+n\), it is the result of n-step forecasting; \(y_{l-h}\) represents the actual value of time series at time \(l-h\), we will use the data of previous seven days to predict one-step-, two-step-, and three-step-ahead concentrations of PM10 and PM2.5.

2.7. Framework of the Proposed Hybrid Approach

In order to enhance the forecasting performance of PM2.5 and PM10 concentrations, a novel hybrid approach is developed based on a data-preprocessing technique that embedded sample entropy into two-stage decomposition and the modified ELM with GTOA. The structure of the proposed approach is shown as Figure 1, which contains three main steps:

Step 1: Data preprocessing. In order to reduce the prediction difficulty and errors, the ICEEMDAN algorithm is first used to decompose the original time series of PM10 and PM2.5 concentration into several IMFs. There are still complex subseries in IMFs. Sample entropy was adopted to measure the complexity of IMFs. For fully extracting the complex features of data, the WT algorithm is further utilized to decompose IMFs with higher sample entropy value.

Step 2: Individual forecasting. In order to improve the forecasting performance, stability and generalization ability of ELM, a new metaheuristic optimization algorithm GTOA is used to optimize the initial weights and thresholds of ELM. Then, the modified ELM of GTOA-ELM is employed to predict all the subseries.

Step 3: Prediction result aggregating. the prediction outcomes of IMFs is obtained via summarizing the forecasting results of subseries decomposed by WT, and then we further summarize the results of IMFs to get the final prediction results.
Figure 1. The structure of the proposed improvement complementary ensemble empirical mode decomposition with adaptive noise (ICEEMDAN)-wavelet transform (WT)-group teaching optimization algorithm (GTOA)-extreme learning machine (ELM) hybrid approach. For $i = 1, \ldots, N$, IMF($i$) is the sub-sequence of the data PM10 and PM2.5 decomposed by ICEEMDAN.

3. Empirical Results and Analysis

In this section, we will use the hybrid ICEEMDAN-WT-GTOA-ELM model to make one-step-, two-step-, and three-step-ahead predictions for hour PM2.5 and PM10 concentrations of Beijing.

3.1. Data Set and Evaluation Criteria

In this study, the data comprises hour concentrations of fine particulate matter (PM10 and PM2.5) of Beijing, the time series includes 10,510 observations from 10 May 2018 to 1 August 2019. The concentrations of PM10 and PM2.5 are shown in Figures 2 and 3. In addition, in the PM2.5 concentration series and PM10 concentration series, the data of the last three months are set as a testing set, and the rest of the data are a corresponding training set.
3. Empirical Results and Analysis

In this section, we will use the hybrid ICEEMDAN-ELM model to make forecasts of fine particulate matter (PM2.5) and coarse particulate matter (PM10) of Beijing. The training set includes hourly observations from May 2018 to August 2019. The concentration series are set as \( G_i \) and \( \hat{G}_i \) represent the actual value and the predicted result, respectively. The horizontal errors, which can be used to assess the horizontal deviation between the predicted result and the actual value [34]. Ds is utilized to measure the capability of forecasting direction. The calculation method of those statistical indicators are explained as follows:

\[
\text{MAE} = \frac{1}{T} \sum_{t=1}^{T} |G(t) - \hat{G}(t)|
\]

\[
\text{MAPE} = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{G(t) - \hat{G}(t)}{G(t)} \right| \times 100\%
\]
\[ \text{NRMSE} = \frac{100}{\bar{G}(t)} \sqrt{\frac{1}{T} \sum_{t=1}^{T} (G(t) - \hat{G}(t))^2} \times 100\% \quad (26) \]

\[ \text{TIC} = \frac{\sqrt{\frac{1}{T} \sum_{t=1}^{T} (G(t) - \hat{G}(t))^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^{T} G^2(t) \sqrt{\frac{1}{T} \sum_{t=1}^{T} \hat{G}^2(t)}}} \quad (27) \]

\[ \text{Ds} = \frac{1}{T} \sum_{t=2}^{T} w(t) \times 100\% \right) = \left\{ \begin{array}{ll}
1, & (G(t) - G(t-1))(\hat{G}(t) - G(t-1)) \geq 0 \\
0, & \text{otherwise}
\end{array} \right. \quad (28) \]

where \( T \) is the total number of observations in the testing set, \( G(t) \) represents the original PM10 or PM2.5 concentration series, \( \hat{G}(t) \) indicates the predicted value of PM10 or PM2.5 concentration. \( \bar{G}(t) \) is the mean value of actual concentrations.

In order to further evaluate the forecasting performance of different models, the Diebold-Mariano (DM) test [35] is utilized to analyze the statistically significant differences between the proposed model and other benchmark models. In this paper, mean square error is taken as the loss function of the DM test. The null hypothesis is that the performance of the benchmark model and the tested model is not significant; the alternative hypothesis is that the forecasting performance of the test model is better than the benchmark models.

3.2. Result and Analysis

In this section, the concentrations of fine particulate matter PM2.5 and PM10 are used to verify the superiority of the proposed model. For a comparison purpose, ELM, DE-ELM, GTOA-ELM, WT-ELM, WT-GTOA-ELM, ICEEMDAN-ELM, and ICEEMDAN-GTOA-ELM models are employed for comparison from different points. Multi-step forecasting results and comparative analysis for PM10 and PM2.5 are shown in the following subsections.

3.2.1. Subsubsection

Figures 4–8 show the MAPE, MAE, NRMSE, TIC, and Ds values of the eight forecast models of one-step-, two-step-, and three-step-ahead predictions. The DM test of multistep predictions is given in Tables 1–3. From Figures 4–8, we can conclude that: (1) The proposed model based on sample entropy and two types of data decomposition technology achieve the highest forecast performance on the horizontal and directional forecasting—the MAE, MAPE, NRMSE, and TIC of it are all the smallest in one-step-, two-step-, and three-step-ahead forecasting. The values for the directional forecasting of the ICEEMDAN-WT-GTOA-ELM are 97.02%, 90.09%, and 88.53%, which are higher than other benchmark models. (2) The models that embedded with sample entropy selection and two-stage decomposition technology display better performance than those models that only include one type of decomposition technology or not. For example, the MAPE of ICEEMDAN-WT-GTOA-ELM is 3.11% in one-step-ahead forecasting and 4.06%, 6.79%, respectively, in ICEEMDAN-GTOA-ELM and WT-GTOA-ELM; the same conclusion can be draw in two-step and three-step forecasting. (3) The forecasting error of models with decomposition technology are all lower than those models without decomposition, which demonstrates that the decomposition technology can effectively extract the feature of nonstationary PM10 concentration time series, further improving the forecasting performance of the single model. (4) The forecasting errors of GTOA-ELM are less than DE-ELM and ELM. The forecasting performance of DE-ELM is also better than ELM. It can be seen that the optimization algorithm can improve the prediction accuracy of the single model, and the GTOA algorithm is more suitable for optimizing the weights and thresholds of ELM than DE optimization algorithm.
that only include one type of decomposition technology or not. For example, the MAPE of ICEEMDAN-WT-GTOA-ELM is 3.11% in one-step-ahead forecasting and 4.06%, 6.79%, respectively, in ICEEMDAN-GTOA-ELM and WT-GTOA-ELM; the same conclusion can be drawn in two-step and three-step forecasting. (3) The forecasting error of models with decomposition technology are all lower than those models without decomposition, which demonstrates that the decomposition technology can effectively extract the feature of non-stationary PM10 concentration time series, further improving the forecasting performance of the single model. (4) The forecasting errors of GTOA-ELM are less than DE-ELM and ELM. The forecasting performance of DE-ELM is also better than ELM. It can be seen that the optimization algorithm can improve the prediction accuracy of the single model, and the GTOA algorithm is more suitable for optimizing the weights and thresholds of ELM than DE optimization algorithm.

Figure 4. Mean absolute error (MAE) of PM10 forecasting.

Figure 5. Mean absolute percentage error (MAPE) of PM10 forecasting.

Figure 6. Normalized root mean square error (NRMSE) of PM10 forecasting.
Figure 6. Normalized root mean square error (NRMSE) of PM10 forecasting.

Figure 7. Theil’s coefficient (TIC) of PM10 forecasting.

Figure 8. Direction statistics (Ds) of PM10 forecasting.

Table 1. DM test results of one-step ahead forecasting of PM10.

| Tested Model      | Benchmark Model | ECMEN-ELM | ICEEMDAN-GTOA-ELM | ICEEMDAN-ELM | WT-GTOA-ELM | WT-ELM | GTOA-ELM | DE-ELM | ELM |
|-------------------|-----------------|-----------|--------------------|--------------|-------------|--------|----------|--------|-----|
| ICEEMDAN-WT-GTOA-ELM | (0.0005) | -3.5039 | -6.7207 | -6.0325 | -3.5972 | -6.5332 | -7.4301 | -8.6697 |
| ICEEMDAN-GTOA-ELM | (0.0000) | -3.8489 | -1.4691 | -3.0902 | -6.1659 | -7.0459 | -8.4643 | -8.4683 |
| ICEEMDAN-ELM | (0.0001) | 0.9991 | -2.6930 | -5.8667 | -6.7564 | -7.1801 | -8.1801 | -8.4683 |
| WT-GTOA-ELM | (0.3179) | -3.0345 | -6.2075 | -7.0796 | -7.0796 | -8.4683 | -8.1801 | -8.4683 |
| WT-ELM | (0.0024) | -4.1550 | -4.6741 | -7.2905 | -7.2905 | -8.1801 | -8.4683 | -8.1801 |
| GTOA-ELM | (0.0000) | -2.1288 | -8.1789 | -7.2905 | -7.2905 | -8.4683 | -8.1801 | -8.4683 |
| DE-ELM | (0.0000) | -6.7950 | -6.7950 | -6.7950 | -6.7950 | -6.7950 | -6.7950 | -6.7950 |
The DM test is utilized to test the statistical difference between the test model and the benchmark models. The null hypothesis is that the performance of the benchmark models and the tested model is not significant, and the alternative hypotheses is that the forecasting performance of the tested model is better than the benchmark models. Tables 1–3 report DM test results for one-step-, two-step-, and three-step-ahead predictions. The values in brackets are p-values of the DM test, and the digits above p-values are the values of the DM statistic. From the results of the DM test, several conclusions can be obtained: (1) When the proposed ICEEEMDAN-WT-GTOA-ELM is set as the tested model, the p-values are all smaller than 0.01; we can conclude that the data preprocessing technology with sample entropy embedded into two-stage decomposition can effectively reduce the forecasting difficulty of PM10 concentration and significantly improve the prediction accuracy of the hybrid model. (2) GTOA-ELM is outperforming ELM model at the significance level of 5% in one-step-, two-step- and three-step-ahead forecasting, indicating that the GTOA algorithm is an effective optimizer tool. (3) The predictive capability of GTOA-ELM is also better than DE-ELM, which shows that the GTOA algorithm can significantly improve the forecasting precision of ELM compared with the DE algorithm.
3.2.2. Comparison Analysis of PM10 Forecasting

In this subsection, the new dataset of hourly PM2.5 concentrations is utilized to test the stability of the new hybrid model. The multistep forecasting results of PM2.5 via the proposed hybrid model are also compared with the results of other benchmark models.

The forecasting errors of all the models are shown in Table 4, the optimal values are highlighted in bold, and it can be seen clearly in Figures 9–13. Tables 5–7 provide the DM test results of one-step, two-step, and three-step forecasting.

Table 4. Forecasting error of PM2.5 concentration.

| Multistep Ahead | Model                | MAE (µg/m³) | MAPE (%) | NRMSE | TIC (%) | Ds (%) |
|-----------------|----------------------|-------------|----------|--------|---------|--------|
| One-step ahead  | ELM                  | 4.40        | 19.61    | 16.27  | 0.07    | 62.49  |
|                 | GTOA-ELM             | 3.40        | 13.24    | 13.54  | 0.06    | 68.00  |
|                 | DE-ELM               | 3.82        | 15.03    | 14.76  | 0.06    | 56.98  |
|                 | WT-ELM               | 1.71        | 6.71     | 6.62   | 0.03    | 89.17  |
|                 | WT-GTOA-ELM         | 1.26        | 4.78     | 4.95   | 0.02    | 93.71  |
|                 | ICEEMDAN-ELM        | 1.11        | 4.33     | 4.03   | 0.02    | 94.00  |
|                 | ICEEMDAN-GTOA-ELM   | 0.60        | 2.35     | 2.42   | 0.01    | 98.10  |
|                 | ICEEMDAN-WT-GTOA-ELM| 0.49        | 1.86     | 1.90   | 0.01    | 98.88  |
| Two-step ahead  | ELM                  | 6.69        | 25.75    | 27.21  | 0.11    | 49.54  |
|                 | GTOA-ELM             | 5.56        | 20.43    | 23.52  | 0.10    | 52.07  |
|                 | DE-ELM               | 5.72        | 21.38    | 23.74  | 0.10    | 48.07  |
|                 | WT-ELM               | 2.91        | 11.13    | 11.00  | 0.05    | 80.23  |
|                 | WT-GTOA-ELM         | 2.48        | 9.33     | 9.42   | 0.04    | 83.89  |
|                 | ICEEMDAN-ELM        | 1.97        | 7.59     | 7.77   | 0.03    | 85.46  |
|                 | ICEEMDAN-GTOA-ELM   | 1.31        | 4.87     | 5.25   | 0.02    | 92.92  |
|                 | ICEEMDAN-WT-GTOA-ELM| 1.13        | 4.31     | 4.31   | 0.02    | 94.39  |
| Three-step ahead| ELM                  | 8.91        | 37.23    | 34.62  | 0.14    | 48.24  |
|                 | GTOA-ELM             | 7.66        | 28.59    | 31.41  | 0.13    | 47.61  |
|                 | DE-ELM               | 7.99        | 32.80    | 32.18  | 0.13    | 46.92  |
|                 | WT-ELM               | 4.10        | 15.45    | 15.45  | 0.06    | 74.17  |
|                 | WT-GTOA-ELM         | 3.57        | 13.32    | 13.63  | 0.06    | 77.64  |
|                 | ICEEMDAN-ELM        | 2.83        | 10.84    | 11.41  | 0.05    | 78.17  |
|                 | ICEEMDAN-GTOA-ELM   | 1.42        | 5.39     | 6.01   | 0.03    | 92.92  |
|                 | ICEEMDAN-WT-GTOA-ELM| 1.32        | 5.11     | 5.41   | 0.02    | 93.51  |

Figure 9. MAE of PM2.5 forecasting.
Figure 9. MAE of PM2.5 forecasting.

Figure 10. MAE of PM2.5 forecasting.

Table 4. Forecasting error of PM2.5 concentration.

| Prediction Horizon | Model | MAE (μg/m$^3$) | MAPE (%) | NRMSE | TIC (%) |
|--------------------|-------|----------------|-----------|--------|---------|
| One-step ahead     | ELM   | 4.40           | 19.61     | 16.27  | 0.07    |
|                    | GTOA-ELM | 3.40         | 13.24     | 13.54  | 0.06    |
|                    | DE-ELM  | 3.82           | 15.03     | 14.76  | 0.06    |
|                    | WT-ELM  | 1.71           | 6.71      | 6.62   | 0.03    |
|                    | WT-GTOA-ELM | 1.26       | 4.78      | 4.95   | 0.02    |
|                    | ICEEMDAN-ELM | 1.11       | 4.33      | 4.03   | 0.02    |
|                    | ICEEMDAN-GTOA-ELM | 0.60     | 2.35      | 2.42   | 0.01    |
|                    | ICEEMDAN-WT-GTOA-ELM | 0.49     | 1.86      | 1.90   | 0.01    |
| Two-step ahead     | ELM   | 6.69           | 25.75     | 27.21  | 0.11    |
|                    | GTOA-ELM | 5.56         | 20.43     | 23.52  | 0.10    |
|                    | DE-ELM  | 5.72           | 21.38     | 23.74  | 0.10    |
|                    | WT-ELM  | 2.91           | 11.13     | 11.00  | 0.05    |
|                    | WT-GTOA-ELM | 2.48       | 9.33      | 9.42   | 0.04    |
|                    | ICEEMDAN-ELM | 1.97       | 7.59      | 7.77   | 0.03    |
|                    | ICEEMDAN-GTOA-ELM | 1.31     | 4.87      | 5.25   | 0.02    |
|                    | ICEEMDAN-WT-GTOA-ELM | 1.13     | 4.31      | 4.31   | 0.02    |
| Three-step ahead   | ELM   | 8.91           | 37.23     | 34.62  | 0.14    |
|                    | GTOA-ELM | 7.66         | 28.59     | 31.41  | 0.13    |
|                    | DE-ELM  | 7.99           | 32.80     | 32.18  | 0.13    |

Figure 11. NRMSE of PM2.5 forecasting.

Figure 12. TIC of PM2.5 forecasting.
Figure 13. Ds of PM2.5 forecasting.

Table 5. DM test results of one-step ahead forecasting of PM2.5.

| Tested Model                  | Benchmark Model                  |      |      |      |      |      |      |
|------------------------------|----------------------------------|------|------|------|------|------|------|
| ICEEMDAN-ELM                | ICEEMDAN-GTOA-ELM                |      |      |      |      |      |      |
| WT-ELM                      | WT-GTOA-ELM                      |      |      |      |      |      |      |
| DE-ELM                      | ELM                              |      |      |      |      |      |      |
| ICEEMDAN-WT-GTOA-ELM        | ICEEMDAN-GTOA-ELM                | −2.7175 | −13.5408 | −10.7244 | −10.3035 | −10.2518 | −12.9119 | −11.8119 |
| (0.0066)                    | (0.0000)                         | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| ICEEMDAN-GTOA-ELM           | ICEEMDAN-ELM                     | −9.3974 | −8.6997 | −9.2922 | −10.1336 | −12.7361 | −11.6786 |          |
| (0.0000)                    | (0.0000)                         | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| ICEEMDAN-ELM                | WT-GTOA-ELM                      | −4.0939 | −6.7991 | −9.6258 | −12.2484 | −11.2966 |          |          |
| (0.0000)                    | (0.0000)                         | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| WT-ELM                      | DE-ELM                           | −9.0503 | −6.3891 | −9.4224 | −12.0698 | −11.2966 |          |          |
| (0.0000)                    | (0.0000)                         | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| GTOA-ELM                    | ELM                              | −4.5612 | −6.3891 | −9.4224 | −12.0698 | −11.2966 |          |          |
| (0.0000)                    | (0.0000)                         | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| DE-ELM                      |                                  | −3.0523 | −8.6997 | −10.4871 | −10.4871 | −11.0846 |          |          |
| (0.0023)                    |                                  | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |

Table 6. DM test results of two-step ahead forecasting of PM2.5.

| Tested Model                  | Benchmark Model                  |      |      |      |      |      |      |
|------------------------------|----------------------------------|------|------|------|------|------|------|
| ICEEMDAN-ELM                | ICEEMDAN-GTOA-ELM                |      |      |      |      |      |      |
| WT-ELM                      | WT-GTOA-ELM                      |      |      |      |      |      |      |
| DE-ELM                      | ELM                              |      |      |      |      |      |      |
| ICEEMDAN-WT-GTOA-ELM        | ICEEMDAN-GTOA-ELM                | −7.2714 | −8.5699 | −15.3825 | −14.6345 | −10.2056 | −10.5647 | −11.1624 |
| (0.0000)                    | (0.0000)                         | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| ICEEMDAN-GTOA-ELM           | ICEEMDAN-ELM                     | −8.0793 | −13.4719 | −13.5744 | −10.1615 | −10.5384 | −11.1396 |          |
| (0.0000)                    | (0.0000)                         | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| ICEEMDAN-ELM                | WT-GTOA-ELM                      | −4.4998 | −7.8214 | −10.0408 | −10.4871 | −10.8486 | −10.8486 |          |
| (0.0000)                    | (0.0000)                         | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| WT-ELM                      | GTOA-ELM                         | −6.7670 | −8.9025 | −9.2585 | −10.1599 | −9.7565 |          |          |
| (0.0000)                    | (0.0000)                         | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| DE-ELM                      |                                  | −0.7343 | −8.3488 | −8.6986 | −9.7565 | −9.7565 |          |          |
| (0.2314)                    |                                  | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| GTOA-ELM                    |                                  | −0.7343 | −8.3488 | −8.6986 | −9.7565 | −9.7565 |          |          |
| (0.2314)                    |                                  | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| DE-ELM                      |                                  | −6.8786 |          |          |          |          |          |          |
| (0.0000)                    |                                  | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
From Table 4 and Figures 9–13, it is obvious that similar conclusions can be summarized that: (1) In most cases, the proposed model ICEEMDAN-WT-GTOA-ELM performs better than the other benchmark models. With one-step-ahead prediction as an example, the MAE, MAPE, NRMSE and TIC values of the proposed model can be reduced more than 80% compared with GTOA-ELM. It is further demonstrated that the data preprocessing technology with sample entropy and two-stage decomposition can lead to a significant enhancement of forecasting performance. (2) Comparing the two types of decomposition models that use different decomposition techniques, the proposed hybrid models based on ICEEMDAN decomposition technique perform better than the models based on WT decomposition technique. (3) Based on the values of direction statistics, the hybrid model of embedded decomposition algorithm has good performance on directional forecasting, while the other models cannot forecast the variation directions precisely.

Tables 5–7 show the DM test results for one-step-, two-step-, and three-step-ahead predictions. The proposed model ICEEMDAN-WT-GTOA-ELM is also significantly better than the other benchmark models; the horizontal and directional accuracy can be greatly improved by data preprocessing technology that embedded sample entropy into the two-stage decomposition technique. The p-value between GTOA-ELM and DE-ELM in the two-step-ahead forecasting is greater than 0.05, but it is smaller than 0.01 in one-step- and three-step-ahead forecasting, and the forecasting errors of GTOA-ELM are all smaller than DE-ELM. The GTOA algorithm is still considered superior to the DE algorithm.

4. Conclusions and Discussion

In this paper, a hybrid approach of GTOA and ELM based on data preprocessing technology to predict the concentration of PM10 and PM2.5 is proposed. This hybrid ICEEMDAN-WT-GTOA-ELM prediction model is able to achieve precise prediction for PM10 and PM2.5 concentrations. The original concentration data of PM10 and PM2.5 are decomposed by ICEEMDAN. In order to completely extract the future of nonstationarity and irregular time series, WT is employed to conduct the secondary decomposition of subseries with high frequency, which possesses greater sample entropy value. The data preprocessing technology that embedded sample entropy into two-stage decomposition can reduce the volatility and prediction difficulty of the initial data significantly. GTOA is used to optimize the initial weights and thresholds of ELM, and the modified ELM is adopted to predict the concentration of subseries. For testing the superiority of the hybrid ICEEMDAN-WT-GTOA-ELM model, we utilized concentration series of PM10 and PM2.5 to make one-step-, two-step-, and three-step-ahead predictions. Empirical analysis results demonstrate that the proposed hybrid approach has superior performance in forecasting...
accuracy and stability for air-pollutant forecasting compared with other benchmark models, which indicates that the data preprocessing technology that embedded sample entropy into two stage-decomposition is a powerful tool for complex series forecasting and the heuristic algorithms of GTOA can efficiently optimize weights and thresholds of ELM.

Although the proposed approach has made substantial progress to make reliable forecasting using irregular and nonstationary time series, there are still a number of issues that need to be addressed. First, this paper considers the univariate time series prediction; some extra factors also can be taken into account. In addition, the deep learning methods with powerful ability of feature extraction have appeared. As a new machine learning method, it will be more efficient in promoting the prediction accuracy of air pollutant concentration. In the future, we will take the meteorological factors and other influence factors into consideration and utilize deep learning models that combine with other heuristic algorithms to further improve the prediction accuracy. Furthermore, it would be important to apply our improved approach to analyze the observed data in real time for making timely predictions. In addition, the proposed model can be also considered to forecast the complex time series in other fields such as finance, biological sciences, ecology, and engineering. All of these issues will be the topics of our further research.

**Author Contributions:** Conceptualization, F.J. and Y.Q.; methodology, F.J.; formal analysis, Y.Q.; data curation, F.J.; writing—original draft preparation, Y.Q. and X.J.; writing—review and editing, Y.Q. and T.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China, grant number 61773401; Hubei Province Key Laboratory of Systems Science in Metallurgical Process (Wuhan University of Science and Technology) grant number Y202001 and Natural Science Foundation of Hubei Province, China, grant number 2020CFB180).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data and methods used in the research have been presented in sufficient detail in the paper.

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

**References**

1. Wen, M.; Li, P.; Zhang, L.; Chen, Y. Stock market trend prediction using high-order information of time series. *IEEE Access* 2019, 7, 28299–28308. [CrossRef]
2. Jeff, T. Ocean scientists work to forecast huge plankton blooms in arabian sea. *Nature* 2018, 555, 569–570. [CrossRef]
3. Kim, K.; Kim, D.K.; Noh, J.; Kim, M. Stable Forecasting of Environmental Time Series via Long Short Term Memory Recurrent Neural Network. *IEEE Access* 2018, 6, 75216–75228. [CrossRef]
4. Dippner, J.W.; Krncke, I. Ecological forecasting in the presence of abrupt regime shifts. *J. Mar. Syst.* 2015, 150, 34–40. [CrossRef]
5. Qiao, W.; Yang, Z. Forecast the electricity price of U.S using a wavelet transform-based hybrid model. *Energy* 2020, 193, 511–530. [CrossRef]
6. Zou, B.; Li, S.; Zheng, Z.; Zhan, F.B.; Yang, Z.; Wan, N. Healthier routes planning: A new method and online implementation for minimizing air pollution exposure risk. *Comput. Environ. Urban Syst.* 2020, 80, 101456. [CrossRef]
7. Ma, X.; Longley, I.; Gao, J.; Salmond, J. Evaluating the effect of ambient concentrations, route choices, and environmental (in)justice on students dose of ambient no2 while walking to school at population scales. *Environ. Sci. Technol.* 2020, 54, 12908–12919. [CrossRef]
8. Afghan, F.R.; Patidar, S.K. Health impacts assessment due to PM2.5, PM10 and NO2 exposure in National Capital Territory (NCT) Delhi. *Pollution* 2020, 6, 115–126. [CrossRef]
9. Sharma, E.; Dee, R.C.; Prasad, R.; Parisi, A.V. A hybrid air quality early-warning framework: An hourly forecasting model with online sequential extreme learning machines and empirical mode decomposition algorithms. *Sci. Total Environ.* 2020, 709, 135934. [CrossRef]
10. Gualtieri, G.; Carotenuto, F.; Finardi, S.; Tartaglia, M.; Toscano, P.; Gioli, B. Forecasting PM10 hourly concentrations in northern Italy: Insights on models performance and PM10 drivers through self-organizing maps. *Atmos. Pollut. Res.* 2018, 9, 1204–1213. [CrossRef]
11. Venkataraman, V.; Usmanulla, S.; Sonnappa, A.; Sadashiv, P.; Mohammed, S.S.; Narayanan, S.S. Wavelet and multiple linear regression analysis for identifying factors affecting particulate matter PM2.5 in Mumbai City. *India Int. J. Qual. Reliab. Manag.* 2019, 36, 1750–1783. [CrossRef]

12. Zhang, L.; Lin, J.; Qui, R.; Hu, X.; Zhang, H.; Chen, Q.; Tian, H.; Lin, D.; Wang, J. Trend analysis and forecast of PM2.5 in Fuzhou, China using the ARIMA model. *Ecol. Indic.* 2018, 95, 702–710. [CrossRef]

13. Xiao, Q.; Zeng, Z. Scale-limited lagrange stability and finite-time synchronization for memristive recurrent neural networks on time scales. *IEEE Trans. Cybern.* 2017, 47, 2984–2994. [CrossRef] [PubMed]

14. Zhou, Y.; Chang, F.J.; Chang, L.C.; Kao, I.F.; Wang, Y.S.; Kang, C.C. Multi-output support vector machine for regional multi-step-ahead PM2.5 forecasting. *Sci. Total Environ.* 2019, 651, 230–240. [CrossRef]

15. Xayasouk, T.; Lee, H.M.; Lee, G. Air pollution prediction using long short-term memory (LSTM) and Deep Autoencoder (DAE) Models. *Sustainability* 2020, 12, 2570. [CrossRef]

16. Kouziokas, G.N. SVM kernel based on particle swarm optimized vector and Bayesian optimized SVM in atmospheric particulate matter forecasting. *Appl. Soft Comput. J.* 2020, 93, 106410. [CrossRef]

17. Zhou, Y.; Chang, F.J.; Chang, L.C.; Kao, I.F.; Wang, Y.S.; Kang, C.C. Multi-output support vector machine for regional multi-step-ahead PM2.5 forecasting. *Sci. Total Environ.* 2019, 651, 230–240. [CrossRef]

18. Zheng, N.; Luo, M.; Zou, X.; Qui, X.; Lu, J.; Han, J.; Wang, S.; Wei, Y.; Zhang, S.; Yao, H. A novel method for the recognition of air visibility level based on the optimal binary tree support vector machine. *Atmosphere* 2018, 9, 481. [CrossRef]

19. Park, S.; Kim, M.; Kim, M.; Namgung, H.G.; Kim, K.T.; Cho, K.H.; Kwon, S.B. Predicting PM 10 concentration in Seoul metropolitan subway stations using artificial neural network (ANN). *J. Hazard. Mater.* 2018, 341, 75–82. [CrossRef]

20. Mahajan, S.; Liu, H.M.; Tsai, T.C.; Chen, L.J. Improving the accuracy and efficiency of PM2.5 forecast service using cluster-based hybrid neural network model. *IEEE Access* 2018, 6, 19193–19204. [CrossRef] [PubMed]

21. Jiang, F.; He, J.; Tian, T. A clustering-based ensemble approach with improved pigeon-inspired optimization and extreme learning machine for air quality prediction. *Appl. Soft Comput. J.* 2019, 85, 105827. [CrossRef]

22. Wang, D.; Liu, Y.; Luo, H.; Yue, C.; Cheng, S. Day-Ahead PM2.5 concentration forecasting using WT-VMD based decomposition method and back propagation neural network improved by differential evolution. *Int. J. Environ. Res. Public Health* 2017, 14, 764. [CrossRef] [PubMed]

23. Yang, H.; Zhu, Z.; Li, C.; Li, R. A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight. *Appl. Soft Comput. J.* 2019, 85, 105972. [CrossRef]

24. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 1998, 454, 903–995. [CrossRef]

25. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* 2009, 1, 1–41. [CrossRef]

26. Torres, M.E.; Colominas, M.A.; Schlotthauer, G.; Flandrin, P. A complete ensemble empirical mode decomposition with adaptive noise. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011; pp. 4144–4147. [CrossRef]

27. Colominas, M.A.; Schlotthauer, G.; Torres, M.E. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomed. Signal Process. Control* 2014, 14, 19–29. [CrossRef]

28. Altuve, M.; Suárez, L.; Ardila, J. Fundamental heart sounds analysis using improved complete ensemble EMD with adaptive noise. *Biocybern. Biomed. Eng.* 2020, 40, 426–439. [CrossRef]

29. Tascikaraoglu, A.; Sanandaji, B.M.; Poolla, K.; Varaiya, P. Exploiting sparsity of Interconnections inspatsio-temporal wind speed forecasting using Wavelet Transform. *Appl. Energy* 2016, 165, 735–747. [CrossRef]

30. Liu, H.; Mi, X.W.; Li, Y.F. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. *Energy Convers. Manag.* 2018, 156, 498–514. [CrossRef]

31. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 2000, 278, H2039–H2049. [CrossRef]

32. Zhang, Y.; Jin, Z. Group teaching optimization algorithm: A novel metaheuristic method for solving global optimization problems. *Expert Syst. Appl.* 2020, 148, 113246. [CrossRef]

33. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* 2006, 70, 489–501. [CrossRef]

34. Sun, S.; Wang, S.; Zhang, G.; Zheng, J. A decomposition-clustering-ensemble learning approach for solar radiation forecasting. *Sol. Energy* 2018, 163, 189–199. [CrossRef]

35. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* 2002, 20, 134–144. [CrossRef]