

Research
Smart Environment Forecasting—Article

A New Model Using Multiple Feature Clustering and Neural Networks for Forecasting Hourly PM_{2.5} Concentrations, and Its Applications in China



Hui Liu*, Zhihao Long, Zhu Duan, Huipeng Shi

Institute of Artificial Intelligence and Robotics (IAIR), Key Laboratory of Traffic Safety on Track of Ministry of Education, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China

ARTICLE INFO

Article history:

Received 23 June 2019

Revised 14 February 2020

Accepted 25 May 2020

Available online 10 June 2020

Keywords:

PM_{2.5} concentrations forecasting

PM_{2.5} concentrations clustering

Empirical wavelet transform

Multi-step forecasting

ABSTRACT

Particulate matter with an aerodynamic diameter no greater than 2.5 μm (PM_{2.5}) concentration forecasting is desirable for air pollution early warning. This study proposes an improved hybrid model, named multi-feature clustering decomposition (MCD)–echo state network (ESN)–particle swarm optimization (PSO), for multi-step PM_{2.5} concentration forecasting. The proposed model includes decomposition and optimized forecasting components. In the decomposition component, an MCD method consisting of rough sets attribute reduction (RSAR), *k*-means clustering (KC), and the empirical wavelet transform (EWT) is proposed for feature selection and data classification. Within the MCD, the RSAR algorithm is adopted to select significant air pollutant variables, which are then clustered by the KC algorithm. The clustered results of the PM_{2.5} concentration series are decomposed into several sublayers by the EWT algorithm. In the optimized forecasting component, an ESN-based predictor is built for each decomposed sublayer to complete the multi-step forecasting computation. The PSO algorithm is utilized to optimize the initial parameters of the ESN-based predictor. Real PM_{2.5} concentration data from four cities located in different zones in China are utilized to verify the effectiveness of the proposed model. The experimental results indicate that the proposed forecasting model is suitable for the multi-step high-precision forecasting of PM_{2.5} concentrations and has better performance than the benchmark models.

© 2020 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

With the growth of urban industry in developing countries and regions, air pollution has become a difficult issue that is attracting attention all over the world. In recent years, hazy weather has appeared in most regions of China, and air quality has become a national strategic problem. Particulate matter with an aerodynamic diameter no greater than 2.5 μm (PM_{2.5}) contains a large number of toxic and harmful substances [1]. PM_{2.5} is the most common air pollutant and has a negative impact on human health and air quality [2]. Previous studies have shown that PM_{2.5} pollution has a direct impact on the respiratory and cardiovascular systems, and is closely related to the incidence and mortality of lung cancer [3]. In addition, PM_{2.5} has a bad influence on the weather and climate. For example, PM_{2.5} may cause abnormal rainfall and aggravate the greenhouse effect [4–7].

Given the serious negative impact of PM_{2.5} on people's lives, increasing attention is being paid to PM_{2.5} concentration forecasting. PM_{2.5} concentration forecasting is considered to be an important and effective method for alleviating the negative effect of PM_{2.5} [8]. This method is also important for applications of urban big data in the development of smart cities [9].

1.1. Related works

PM_{2.5} concentration forecasting methods can be divided into four types: physical models, statistical models, artificial intelligence models, and hybrid models.

Physical models focus on understanding the potentially complex emissions, transport, and conversion processes of meteorological and chemical factors [10]. The physical method yields accurate prediction results. However, physical models require sufficient emission information on air pollution [11] and their calculation cost is high [12]. Statistical models overcome the disadvantage of physical methods, as they require simple samples and have a fast calculation speed [13]. However, statistical models do not

* Corresponding author.

E-mail address: csulihui@csu.edu.cn (H. Liu).

sufficiently consider the covariance among various influential factors, because they are generally based on limited samples. A single artificial intelligence model can describe the rule of the nonlinear system and has great advantages in dealing with big data [14]. However, the disadvantage of such a model lies in the calculation costs of the neural network, which are greater than those of a statistical model. Moreover, the training process of the neural network has a certain volatility, so its output may not be the optimal result [15].

Considering the limitations of the above methods, hybrid models have been widely used in air pollution prediction. Hybrid models usually combine three parts: data preprocessing, feature selection, and a predictor. Data preprocessing can sort out complex data relationships in the original data and make it more stationary. Feature selection can improve the input data structure and reduce the difficulty of model training caused by a too-high dimension. Hybrid models can integrate the advantages of each algorithm to achieve better model performance. Many related works have shown that hybrid models tend to have better predictive performance [16–20]. Table 1 [16–28] lists some cutting-edge research on hybrid PM_{2.5} concentration forecasting to better illustrate the application of hybrid methods in PM_{2.5} concentration forecasting.

Feature selection is rarely used in the current hybrid PM_{2.5} concentration forecasting models listed in Table 1. However, if the input of a PM_{2.5} concentration forecasting model includes many features, such as PM_{2.5}, PM₁₀, sulfur dioxide (SO₂), and ozone (O₃), it may cause difficulties in the PM_{2.5} concentration forecasting model training and increase the training time. This also affects the robustness of the PM_{2.5} concentration forecasting model [29]. At the same time, complex input data may lead to overfitting of the model and may reduce the accuracy of the model [30]. At present, common feature selection algorithms include the principal components analysis (PCA), phase space reconstruction (PSR), and gradient-boosted regression tree (GBRT). However, these methods may be unsuitable for air pollutant concentration sequences because they assume a linear system, which may lead to problems such as not achieving global optimal reduction. The rough sets attribute reduction (RSAR) algorithm, which is based on fuzzy theory, has the advantages of clear stop criteria and no parameters [31]. RSAR can obtain the important attribute set of the target attribute through the dependency between different attributes. The RSAR algorithm is a hot research topic [32]. Clustering algorithms are commonly used in data mining and analysis [33]. Various clustering methods exist, such as *k*-means clustering (KC) [34], possibilistic *c*-means (PCM) [35], cure clustering [36], and so forth. Compared with others, the KC algorithm has the advantages of a simple principle, fast computing speed, and excellent clustering results; thus, the KC algorithm is the most widely used clustering algorithm at present. Combining the RSAR algorithm and the KC algorithm makes it possible to use RSAR to provide reasonable clustering objects for the KC algorithm, which is a valuable research point.

Decomposition mainly focuses on the wavelet theory method in Table 1. The decomposition algorithm can divide the original data into a series of more stable sublayers according to the different time scales. Compared with empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), and complex empirical mode decomposition (CEMD), the empirical wavelet transform (EWT) algorithm can adaptively divide the Fourier spectrum and select the appropriate wavelet filter banks [37]. The clustering method can also be employed for decomposition in PM_{2.5} concentration forecasting fields. The clustering algorithm can classify the original data according to different air pollutant scenarios. The influence of sample diversity on the model training can be reduced by clustering. However, few studies use a

Table 1
Main studies on PM_{2.5} concentration forecasting in the past four years.

Reference	Feature selection	Decomposition	Predictor
[16]	None	CEEMD	SVR optimized by GWO
[21]	None	None	The CAMS
[22]	None	EMD	LS-SVR
[23]	PCA	None	LS-SVR optimized by CSO
[24]	None	None	WRF-Chem
[25]	None	CEEMD	SVR optimized by CPSOGA
[26]	None	VMD	SVR optimized by GWO
[17]	PSR	WPD and CEEMD	LS-SVR optimized by CPSO-GA
[20]	GBRT	WPD	MLP optimized by LPBoost
[27]	None	WPD	BPNN optimized by PSO
[28]	None	None	NAQPMS
[18]	None	WT	ANN and SVM
[19]	None	WD and VMD	LSTM

PCA: principal component analysis; PSR: phase space reconstruction; GBRT: gradient-boosted regression tree; CEEMD: complete ensemble empirical mode decomposition; EMD: empirical mode decomposition; VMD: variational mode decomposition; WPD: wavelet packet decomposition; WT: wavelet transform; WD: wavelet decomposition; SVR: support vector regression; GWO: grey wolf optimization; CAMS: Copernicus Atmosphere Monitoring Service; LS-SVR: least-squares support vector regression; CSO: cuckoo search optimization; WRF-Chem: weather research and forecasting model with chemistry; CPSO-GA: combined particle swarm optimization with genetic algorithm; MLP: multi-layer perceptron; LPBoost: linear programming boosting; BPNN: backpropagation neural network; NAQPMS: Nested Air Quality Prediction Model System; ANN: artificial neural network; SVM: support vector machine; LSTM: long-short term memory network.

clustering algorithm with a decomposition algorithm in hybrid PM_{2.5} concentration forecasting.

The predictors shown in Table 1 are commonly used physical methods, machine learning, and artificial neural networks (ANNs). Although the Copernicus Atmosphere Monitoring Service (CAMS), weather research and forecasting model with chemistry (WRF-Chem), and Nested Air Quality Prediction Model System (NAQPMS) have accurate prediction results, they require complex data and an understanding of a variety of physical and chemical relationships. Therefore, these methods require a great deal of preparatory work and a high level of professional knowledge. Support vector machines (SVMs), support vector regression (SVR), and least-squares support vector regression (LS-SVR) are very demanding in their choice of parameters, and cannot handle problems with large data. Traditional neural networks such as the backpropagation neural network (BPNN) and evolutionary neural network (ENN) need a great deal of training to build complex neural relationships, and are easy to overfit. The echo state network (ESN) has a unique reservoir structure that consists of recurrently connected units. As a result, the training process of the ESN is simple and effective, which is suitable for nonlinear systems such as PM_{2.5} concentration data [38]. The ESN has been used in other fields such as wind speed prediction [38]. Therefore, application of the ESN model in hybrid PM_{2.5} concentration forecasting is very appropriate.

1.2. The innovation of this study

To summarize the references described above, the decomposition-based clustering algorithm, nonlinear fuzzy theory algorithm, and ESN are rarely studied in PM_{2.5} concentration forecasting. This study aims to apply these algorithms for hybrid PM_{2.5} concentration forecasting. The proposed hybrid PM_{2.5} prediction model combines three methods: multi-feature clustering decomposition (MCD), ESN, and particle swarm optimization (PSO). In MCD, the RSAR algorithm is adopted to select significant air pollutant concentrations. Then the KC algorithm is used to divide the original PM_{2.5} concentration data into several groups according to the results of the RSAR algorithm. The clustered results for the

PM_{2.5} concentration series are automatically decomposed into several sublayers by the EWT algorithm. An ESN-based predictor is built for every decomposed sublayer in each clustered group to complete the multi-step forecasting computation. The forecasted results from every sublayer are then further integrated to form the final predicting values. The PSO algorithm is utilized to optimize the initial parameters of the ESN-based predictor. The experimental results show that the proposed hybrid model can accurately predict average hourly concentrations of PM_{2.5}. The details of the proposed model are explained in Section 2.

2. Methodology

2.1. Framework of the proposed model

The construction procedure of the hybrid MCD–ESN–PSO model is as follows:

Part A: MCD

This part consists of the RSAR, KC, and EWT algorithms. The raw air quality data are filtered by the RSAR algorithm, and the filtered attribute data are clustered by the KC algorithm. The raw data are divided into several clusters by this step. Then the clustered data of each cluster are decomposed into sublayers by the EWT algorithm. Finally, the sublayers are utilized to establish different ESN models. Through the MCD method, the RSAR algorithm and KC algorithm can act on the raw data together to achieve the clustering of features. Through the decomposition processing of the EWT algorithm, the original time series is finally decomposed into more and better sublayers. The details of the RSAR, KC, and EWT algorithms are introduced in Sections 2.2–2.4, respectively.

Part B: ESN

The ESN is a basic predictor, which forecasts the decomposed PM_{2.5} concentration data. The ESN is composed of an input layer, reserve pool, and output layer. The main idea of the ESN is to use the reserve pool to simulate a complex dynamic space that can change with the input. Referring to Ref. [38], the updated equation and output state equation of the ESN can be expressed as Eqs. (1) and (2):

$$x(t+1) = f[W_{in} \times u(t+1) + W_{back} \times x(t)] \quad (1)$$

$$y(t+1) = f\{W_{out} \times [u(t+1); x(t+1)]\} \quad (2)$$

where x is input data from reserve pool to output layer; y is output; t is time; u is input data from input layer to reserve pool; f is the function of ESN; W_{in} represents the connection weights of $x(t-1)$ to $x(t)$; $u(t+1)$ is the input data; W_{back} represents the connection weights of the input layer to the reserve pool; and W_{out} represents the connection weights of $y(t-1)$ to $x(t)$.

Part C: PSO

Unlike the traditional ESN model, the ESN model proposed in this paper is combined with the PSO algorithm. In the ESN–PSO algorithm, the relevant parameters of the ESN model, such as input scaling, spectral radius, internal unit number, and connectivity, are optimized by the PSO algorithm.

Finally, the forecasting results of each sublayer are combined with the corresponding original sublayer. The prediction results of each sublayer are added to obtain the final prediction results.

2.2. Rough sets attribute reduction

RSAR can be used to remove useless information while maintaining the quality of the sorting of the existing information [31]. The obtained information is referred to as “reducts.” In an information system, a set of objects are described by a set of attributes [31]. A knowledge information system is defined as follows:

$$S = (U, V, A, h) \quad (3)$$

where U is a finite nonempty set of objects; V is a nonempty set of values; A is a finite nonempty set of attributes; and h is an information function that maps an object in U to exactly one value in V .

In this study, A is a set of all attributes such as $A = \{PM_{10}, CO, SO_2, NO_2, O_3, PM_{2.5}\}$, and V is their value. f is the dependency function that is used to obtain γ , and γ is the dependence of the set, which is calculated in the process of RSAR.

The reduct should maintain the quality of sorting (γ), defining a condition attributes set $C \subseteq A$ and an attribute set $P \subseteq C \subseteq A$. Sometimes, an information table may have more than one reduct; the intersection of all the reducts is called a “core” of a decision-making table and is expressed as core (P); this is the most important attributes set for an information system.

2.3. k-means clustering

KC is a simple iterative clustering algorithm, using distance as a similarity index [34]. Its final purpose is to find k clusters in a group of given datasets. The center of each cluster is according to the value of all clusters in which each cluster is described by the clustering center. The process of the KC algorithm is as follows:

(1) Select the k object in the data space as the initial center; each object represents a cluster center.

(2) Divide the data objects in the sample into the corresponding classes according to the nearest clustering center, according to the Euclidean distance between them and these cluster centers.

$$\text{Distance}(x_i, x_j) = \sqrt{\sum_{d=1}^D (x_{i,d} - x_{j,d})^2} \quad (4)$$

where x_i is the i th sample in the j th cluster; x_j is the center of the j th cluster; and D represents the number of attributes of a data object.

(3) Update the clustering center: taking the mean values of all objects in each cluster as the clustering center, calculating the value of the objective function.

(4) Judge whether the values of the cluster centers and objective functions are equal. If they are equal, output results; otherwise, return to step (2).

2.4. Empirical wavelet transform

In this paper, the EWT algorithm is used for data preprocessing. The EWT, proposed by Gilles [37], is a novel signal-processing technique that builds the wavelets adaptively. The EWT is based on the theoretical framework of wavelet transform but overcomes the shortage of EMD theory and the problem of signal aliasing. The EWT adaptively divides the Fourier spectrum and selects the appropriate wavelet filter banks. The empirical scaling function and the empirical wavelets can be expressed as Eqs. (5) and (6), respectively:

$$\hat{\phi}_n(\omega) = \begin{cases} 1 & \text{if } |\omega| \leq (1-\tau)\omega_n \\ \cos\left(\frac{\pi}{2}\beta\left\{\frac{1}{2\tau\omega_n}[|\omega| - (1-\tau)\omega_n]\right\}\right) & \text{if } (1-\tau)\omega_n \leq |\omega| \leq (1+\tau)\omega_n \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$$\hat{\psi}_n(\omega) = \begin{cases} 1 & \text{if } (1+\tau)\omega_n \leq |\omega| \leq (1-\tau)\omega_{n+1} \\ \cos\left(\frac{\pi}{2}\beta\left\{\frac{1}{2\tau\omega_{n+1}}[|\omega| - (1-\tau)\omega_{n+1}]\right\}\right) & \text{if } (1-\tau)\omega_{n+1} \leq |\omega| \leq (1+\tau)\omega_{n+1} \\ \sin\left(\frac{\pi}{2}\beta\left\{\frac{1}{2\tau\omega_n}[|\omega| - (1-\tau)\omega_n]\right\}\right) & \text{if } (1-\tau)\omega_n \leq |\omega| \leq (1+\tau)\omega_n \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where n is divided interval; ω is frequency; β is any function in the interval $[0,1]$ that satisfies the derivative of the order, τ is frequency

coefficient; $\beta(x) = x^4(35 - 84x + 70x^2 - 20x^3)$ and $\tau < \min n[(\omega_{n+1} - \omega_n)/(\omega_{n+1} + \omega_n)]$.

2.5. Particle swarm optimization

The PSO algorithm consists of position z , speed v , and the adaptation function. Each particle in the algorithm represents a candidate solution in the solution space. The fitness function is set according to the optimization goal. During the training of the PSO, each particle in the algorithm updates its own position by combining its current movement experience with the movement experience of the neighboring particles. The solution is realized by putting one's own position close to the target position. The calculation formula is as follows [27]:

$$v_i(m + 1) = q \times v_i(m) + c_1 \times r_1(p_i^{best} - z_i) + c_2 \times r_2(g_i^{best} - z_i) \quad (7)$$

$$z_i(m + 1) = z_i(m) + v_i(m + 1) \quad (8)$$

where m presents the number of iterations; $v_i(m)$ represents the current velocity of the i th particle; c_1 and c_2 represent constants; r_1 and r_2 represent random numbers between 0 and 1; p represents the weight of the particles; p_i^{best} represents the individual optimal value from the beginning to the current number of iterations; and g_i^{best} represents the group optimal value from the beginning to the current number of iterations.

3. Case study

3.1. Study area

Related literature studies show that the distribution of PM_{2.5} concentrations ranges widely in China. It is mainly concentrated in North China and Central China [39,40]. In order to ensure the diversity of experimental data, the selected data should include different working conditions such as serious PM_{2.5} pollution and weak PM_{2.5} pollution. Beijing in the North China Plain, Guangzhou in the Pearl River Delta, Changsha in Central China, and Suzhou in the Yangtze River Delta are typical cities that are used to verify the effectiveness of the proposed model. The selected samples are spatially representative and contain PM_{2.5} concentration data under different geographic and climatic environments, which can well verify the effectiveness of the proposed model.

Monitoring stations record the average concentrations of six kinds of air pollutants (PM_{2.5}, PM₁₀, NO₂, SO₂, O₃, and CO). Fig. 1 shows the selected datasets and related introduction.

3.2. Data description and partitioning

Experimental data are collected from four cities: Beijing, Guangzhou, Changsha, and Suzhou. As Shi et al. [41] have indicated, the spatial representation of surface site observation is often 0.5–16 km², with the most frequent values being around

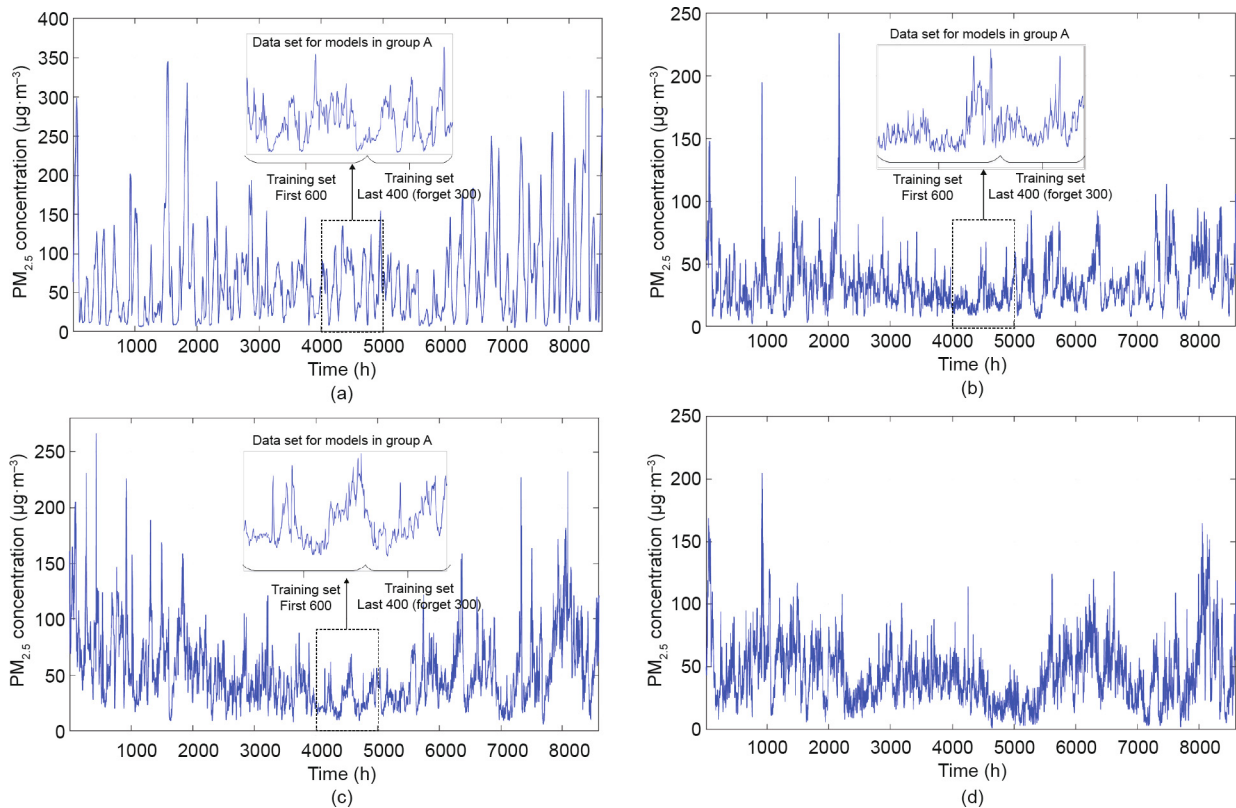


Fig. 1. Locations of the air quality monitoring stations. (a) Beijing. Beijing is the capital of China, located at the north end of the North China Plain. It has a typical warm temperate semi-humid continental monsoon climate, hot and rainy in summer, cold and dry in winter, short in spring and autumn. The average annual temperature is 10–12 °C and the average annual rainfall is more than 600 mm. (b) Changsha. Changsha is an important city in the middle reaches of the Yangtze River. It is a subtropical monsoon climate, mild climate, abundant precipitation, hot and rainy at the same time. Its average annual temperature is 17.2 °C and the average annual precipitation is 1361.6 mm. (c) Guangzhou. Guangzhou is located in the southeastern part of China, the northern edge of the Pearl River Delta, and the Pearl River passes through the urban area. It belongs to the tropical monsoon climate with high temperature, rainfall, and low wind speed. (d) Suzhou. Suzhou is located in the southeast of Jiangsu Province and the middle of the Yangtze River Delta. It is a subtropical monsoon marine climate, four distinct seasons, abundant rainfall throughout the year. Group A: models without RSAR-KC, including the ESN, LSTM, ESN-PSO, and EWT-ESN-PSO model.

3 km². The data of a single monitoring station cannot represent the air quality of the whole city. Therefore, each set of data is the mean value of all air quality monitoring stations in the corresponding city, so that the samples can represent the air quality of the whole city. For convenience, these datasets are named D1 (Beijing), D2 (Guangzhou), D3 (Changsha), and D4 (Suzhou). The length of the sample data is set to one year so that the data can cover a complete set of four seasons. In this study, the sample data in 2016 are randomly selected. All experimental data includes one-hour average concentrations of PM_{2.5}, PM₁₀, NO₂, SO₂, O₃, and CO collected from 1 January 2016 to 31 December 2016. All data are retrieved from the website of the China National Environmental Monitoring Center.[†]

Missing value filtering and outlier checking are implemented before data partitioning. Through the sample set analysis, it is concluded that there are 220 missing pieces of data in dataset D1. Dataset D2 is missing 158 pieces of data, dataset D3 is missing 158 pieces of data, and dataset D4 is missing 157 pieces of data. Since the number of missing samples is less than 2.5% of the total sample set, direct elimination will not have a significant impact. Through visual inspection of the outliers in Fig. 1, it is found that the outliers are mostly concentrated from January to March and from October to December. In order to ensure the training effect of the model, outliers are regarded as normal and retained.

After removing the missing samples, D1 has 8540 samples, D2 has 8602 samples, D3 has 8602 samples, and D4 has 8603 samples. The 4001st–4600th samples of each dataset (other data are utilized for KC in group B) are used to train the models in group A (models without RSAR–KC, including the ESN, long–short term memory network (LSTM), ESN–PSO, and EWT–ESN–PSO model), which only use PM_{2.5} concentrations. The 4601st–5000th samples are the testing set; to ensure the prediction effect, the 4601st–4900th samples are abandoned. All of the experimental data of each station is used in the RSAR and KC to preprocess the data in group B (models with RSAR–KC, including the RSAR–KC–ESN, the MCD–LSTM–PSO, and the RSAR–KC–EWT–ESN–PSO model). To ensure the effectiveness of error evaluating, each cluster is used to train an ESN model and then reconstruct the predicted results for the 4901st–5000th samples.

In order to study the influence of different sampling processes on model accuracy, the 3001st–4000th (S1) sample points and 6001st–7000th (S2) sample points in D1 are used for comparison experiments. Fig. 2 shows the distribution of datasets S1 and S2.

In order to further verify the effectiveness of the model, an additional dataset named D4 (which contains 8603 samples) is used in the experiments. The dataset D4 selects monthly data from the spring, summer, autumn, and winter for testing; these data are named T1 (1000th–1999th samples), T2 (3100th–4099th samples), T3 (5000th–5999th samples), and T4 (6000th–6999th samples). They are shown in Fig. 3. Table 2 shows the descriptive statistics of the related PM_{2.5} concentrations.

3.3. Results and discussion

3.3.1. Results of RSAR

The RSAR and KC are used to preprocess the original data. The attribute decision tables of each dataset are established according to the international PM_{2.5} classification system. According to the international PM_{2.5} concentration index classification standard, PM_{2.5} concentration data are classified and discretized. Like the method of classifying PM_{2.5} concentration levels, the concentrations of the other five air pollutants are discretized according to the levels. Table 3 shows the attribute reduction table for this study. By calculating the positive region values of the other five kinds of air pollutant concentrations and PM_{2.5} concentrations, it can be determined that the significance degrees of PM₁₀, NO₂, CO, O₃, and SO₂ are 0.0825, 0.0948, 0.0531, 0.2189, and 0.1843, respectively. O₃ and SO₂ have great significance and are judged to be the core attributes of the established information decision system.

It should be noted that if the correlation between the reduction attribute and the decision attribute is too strong, there is no distinction between the two. If the correlation between the reduction attributes and decision attributes is too weak, there is no correlation between them. The reduction attributes in both cases are redundant. Therefore, in order to ensure the diversity of input samples, the selection of reduction attributes needs to comprehensively consider the correlation and independence between the reduction attributes and decision attributes. For this reason, this paper uses covariance to evaluate the relationship between PM_{2.5} concentrations and other pollutant concentrations, as shown in Table 4. The cov(PM_{2.5}, PM₁₀), cov(PM_{2.5}, NO₂), cov(PM_{2.5}, CO), and cov(PM_{2.5}, SO₂) are positive. The cov(PM_{2.5}, O₃) is negative. However, the absolute value of cov(PM_{2.5}, PM₁₀), cov(PM_{2.5}, NO₂), and cov(PM_{2.5}, CO) is much larger than that of cov(PM_{2.5}, SO₂) and cov(PM_{2.5}, O₃). In terms of ensuring the independence of input attributes, the result of the RSAR algorithm can be verified to be

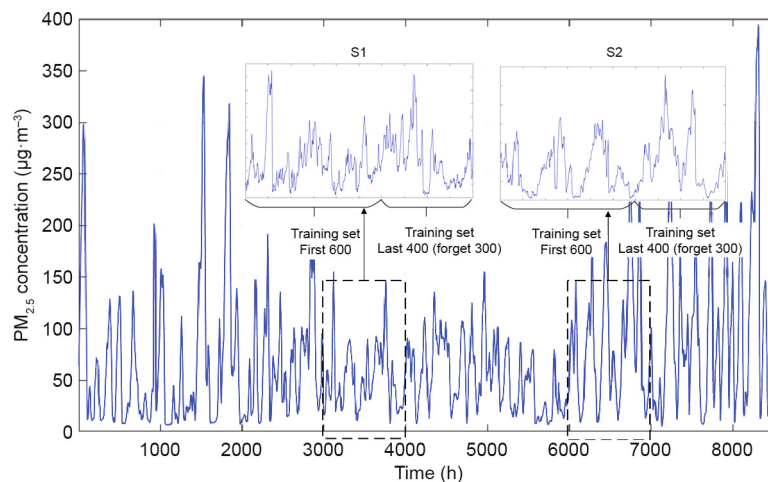


Fig. 2. PM_{2.5} concentrations series of S1 and S2.

[†] <http://www.cnemc.cn/>

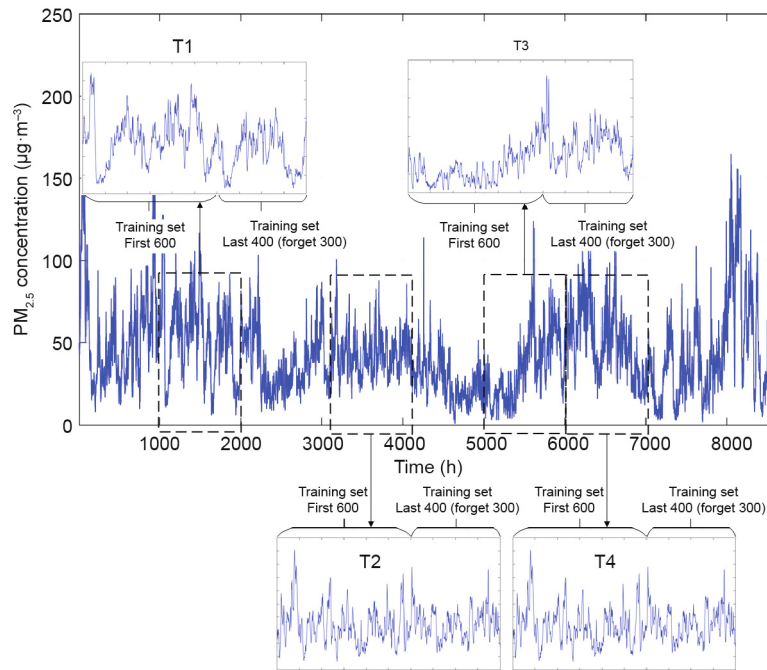


Fig. 3. PM_{2.5} concentrations series of T1–T4.

Table 2
Descriptive statistics of PM_{2.5} concentrations.

Dataset	City	Group	PM _{2.5} concentration (µg·m ⁻³)				Skewness	Kurtosis
			Mean	Minimum	Maximum	SD		
D1	Beijing	B	71.40	3.00	692.00	71.00	2.01	8.64
		A	65.82	5.00	223.00	40.04	0.64	3.39
		S1	52.92	4.00	190.00	34.52	1.26	4.85
		S2	87.57	5.00	328.00	64.81	0.86	3.16
D2	Guangzhou	B	34.48	2.00	234.00	20.72	1.92	11.09
		A	22.58	9.00	68.00	10.03	1.53	5.90
D3	Changsha	B	52.93	6.00	409.00	33.58	1.96	11.20
		A	28.39	9.00	69.00	12.11	0.76	2.76
D4	Suzhou	T1	52.48	6.00	128.00	24.21	0.27	2.88
		T2	43.86	16.00	101.00	13.02	0.77	4.18
		T3	34.77	3.00	124.00	20.24	0.68	3.30
		T4	33.40	2.00	124.00	19.48	0.74	3.67

SD: standard deviation.

Table 3
Attribute reduction table.

Object <i>U</i>	Condition attributes					Decision attributes PM _{2.5}
	PM ₁₀	NO ₂	CO	O ₃	SO ₂	
1	4	5	4	5	6	5
2	6	6	5	6	6	6
3	5	4	6	6	5	6
4	1	2	6	1	1	2
5	1	1	6	2	2	1
...

U: a finite nonempty set of objects.

Table 4
Covariance table.

Covariance	Value
cov(PM _{2.5} , PM ₁₀)	4157.93
cov(PM _{2.5} , NO ₂)	1247.56
cov(PM _{2.5} , CO)	51375.21
cov(PM _{2.5} , O ₃)	-119.34
cov(PM _{2.5} , SO ₂)	339.80

effective. In order to avoid difficulty in model training caused by dimensional disaster, these more relevant attributes are selected as the core attributes, and other data with weak correlation become the reduction attributes.

3.3.2. Results of KC

After the attribute reduction, the original dataset becomes an $N \times 3$ sample space; three-dimensional (3D) KC is used to divide

it into several similar clusters. The sum of the squared errors (SSE) [42] and the silhouette coefficient (SC) [43] are used to choose the best value of k . Because the clustering results of the three datasets are very similar, D1 is used as an example to show the results.

Fig. 4 shows different SSE and SC when choosing different k . The k value ranges from 1 to 15, and the SSE value decreases with the increase of the k value. When $k = 3$, the SC value is the largest, but the SSE value is larger at the same time. Considering SSE and SC together in Fig. 4, the final k value is 7. Finally, the data of station D1 are divided into seven clusters.

When $k = 7$, the original data D1 are divided into seven groups; the results are shown in Fig. 5. In this figure, (a) shows the results for $PM_{2.5}$, while the results for SO_2 and O_3 are presented separately in (b) and (c). The results of the KC for $PM_{2.5}$ shown in Fig. 5(a) are the key part of this paper. Intuitively, the amplitude of cluster C1 is between 0 and $200 \mu\text{g}\cdot\text{m}^{-3}$, and it fluctuates gently. The amplitude of C2 is between 0 and $55 \mu\text{g}\cdot\text{m}^{-3}$, and fluctuates violently. However, the short period of C2 is obvious. The amplitude of C3 is between 0 and $400 \mu\text{g}\cdot\text{m}^{-3}$. The fluctuation of C3 is smooth but the periodicity is not obvious. The amplitude of C4 is between 50 and $150 \mu\text{g}\cdot\text{m}^{-3}$, and its periodicity and symmetry are good. The amplitude of C5 is between 0 and $200 \mu\text{g}\cdot\text{m}^{-3}$, but it fluctuates more violently than that of C1. The amplitude of C6 is between 160 and $240 \mu\text{g}\cdot\text{m}^{-3}$, fluctuates violently, and has strong symmetry. The amplitude of C7 is between 0 and $100 \mu\text{g}\cdot\text{m}^{-3}$, and the period is obvious but the symmetry is weak. Overall, compared with the original data in Fig. 1, each kind of data becomes more stable after clustering. The peaks and troughs of each type of data show different intervals.

The above description is only a subjective analysis; in order to draw a more convincing conclusion, descriptive statistics are used to further analyze the clustering results of $PM_{2.5}$ concentration data. Table 5 shows the descriptive statistics of D1 for the different clusters.

The mean values of the seven groups of data are 71.54, 24.00, 285.74, 91.47, 83.90, 177.00, and $34.85 \mu\text{g}\cdot\text{m}^{-3}$, respectively. Combined with the maximum and minimum values of each group of data, the seven groups of data after clustering are concentrated according to the value size, thereby reducing the fluctuation range of the data within the group. This is consistent with the amplitude distribution of each group of data in Fig. 5.

The standard deviation (SD) reflects the dispersion degree among individuals in a group. The standard deviation values of the seven groups of data after clustering are 37.02, 14.25, 47.30, 20.96, 32.70, 29.81, and $19.42 \mu\text{g}\cdot\text{m}^{-3}$, respectively, which are all smaller than the $71.00 \mu\text{g}\cdot\text{m}^{-3}$ before clustering. The data of each group after clustering are closer to their average values. This trend is reflected in Fig. 5: The symmetry of the upper and lower fluctuations of the data curves of each group is stronger.

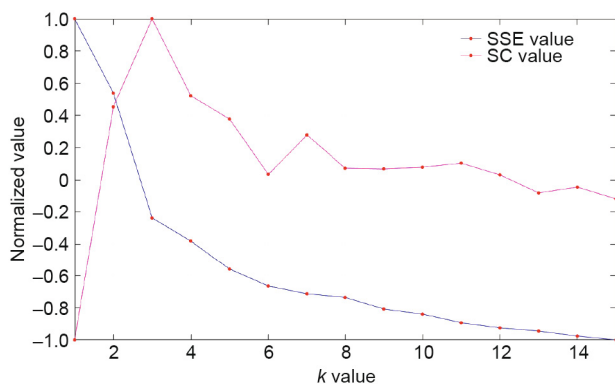


Fig. 4. SSE and SC with different k values.

The skewness values of the seven groups of data after clustering are 0.70, 0.72, 0.74, 0.21, 1.01, 0.22, and 0.88, respectively, which are all smaller than the 2.01 before clustering. The wave peak symmetry of the data after clustering is stronger; that is, the cycle rule is more obvious. The kurtosis values of the seven groups of data after clustering are 3.23, 2.45, 2.50, 1.98, 4.00, 1.82, and 3.12, respectively, which are all smaller than the 8.64 before clustering. The extreme distribution of data in each group of data after clustering is reduced. In Fig. 5, the fluctuation of each group of data is smooth, and there is no obvious outlier. In other words, Fig. 5 and Table 5 show that the clustered data have the above advantages.

The MCD-ESN model is used to analyze the series length in each cluster. In order to ensure the validity of the error evaluation, the first 80% of the data in each cluster is selected for model training and the last 20% is used for model prediction performance analysis. Table 6 shows the error evaluation of each cluster.

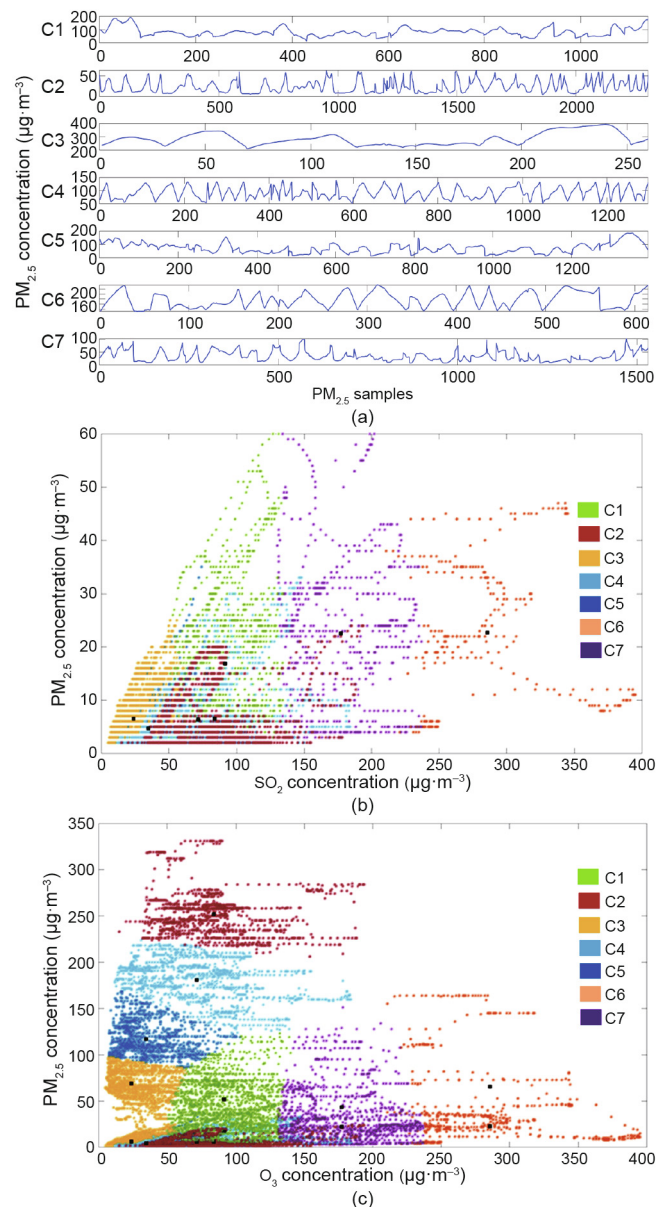


Fig. 5. (a) Results of KC for $PM_{2.5}$; (b) results of KC for $PM_{2.5}$ and SO_2 ; (c) results of KC for $PM_{2.5}$ and O_3 .

Table 5
Descriptive statistics of D1 for different clusters.

Cluster	PM _{2.5} concentration (μg·m ⁻³)				Skewness	Kurtosis
	Mean	Minimum	Maximum	SD		
C1	71.54	12.00	184.00	37.02	0.70	3.23
C2	24.00	5.00	64.00	14.25	0.72	2.45
C3	285.74	211.00	395.00	47.30	0.74	2.50
C4	91.47	50.00	140.00	20.96	0.21	1.98
C5	83.90	15.00	194.00	32.70	1.01	4.00
C6	177.00	130.00	237.00	29.81	0.22	1.82
C7	34.85	8.00	103.00	19.42	0.88	3.12

Table 6
Error evaluation of each cluster for D1 of MCD-ESN.

Cluster	Number	Step	PM _{2.5} concentration				R	IA
			MAPE (%)	MAE (μg·m ⁻³)	RMSE (μg·m ⁻³)	SDE (μg·m ⁻³)		
C1	1527	Step-1	7.47	3.07	6.33	6.30	0.93	0.95
		Step-2	14.11	6.09	9.23	8.94	0.78	0.89
		Step-3	19.29	8.16	11.80	11.30	0.52	0.72
C2	260	Step-1	3.60	11.22	16.67	13.50	0.95	0.96
		Step-2	4.30	14.75	17.30	14.17	0.90	0.92
		Step-3	9.91	33.03	38.73	34.40	0.79	0.84
C3	2307	Step-1	6.38	1.75	2.61	2.61	0.93	0.96
		Step-2	12.09	3.34	5.04	5.04	0.80	0.82
		Step-3	20.55	5.26	7.47	7.43	0.48	0.51
C4	615	Step-1	3.51	5.30	13.41	13.41	0.94	0.97
		Step-2	8.69	12.88	23.10	22.98	0.81	0.89
		Step-3	10.17	15.54	26.13	26.10	0.67	0.78
C5	1297	Step-1	4.01	3.57	4.61	4.60	0.93	0.96
		Step-2	7.67	6.66	8.68	8.66	0.84	0.92
		Step-3	11.49	9.91	12.52	12.48	0.68	0.77
C6	1394	Step-1	4.43	5.54	12.27	11.85	0.92	0.96
		Step-2	7.85	10.22	15.25	13.44	0.78	0.88
		Step-3	10.06	13.60	17.15	13.58	0.62	0.74
C7	1140	Step-1	2.81	3.08	3.38	8.30	0.96	0.97
		Step-2	5.30	5.85	12.46	12.30	0.90	0.94
		Step-3	8.28	9.20	16.37	15.99	0.83	0.91

MAPE: mean absolute percentage error; MAE: mean absolute error; RMSE: root mean square error; SDE: standard deviation of error; R: Pearson's correlation coefficient; IA: index of agreement.

When the sample size is greater than 1000, the amount of data has little effect on the prediction, as in C1, C3, C5, C6, and C7. However, when the number of samples is less than 1000, the prediction effect of the model is greatly reduced; this shows that the prediction effect of the ESN network is more sensitive to the number of low samples, as in C2 and C4. When the number of samples is small after clustering, the problem can be solved by increasing the number of samples in the original sequence. This will be a problem for future research.

3.3.3. Forecasting accuracy and analysis

In this study, six other prediction models are provided as comparison models to investigate the prediction performance of the proposed model. In addition, to investigate the multi-step prediction performance of the proposed model, all the involved models are conducted for the step-1 to step-3 predictions. The proposed model must forget a certain amount of output results due to the characteristics of the ESN algorithm [38]. Therefore, the prediction accuracy fluctuates within a certain range. In this paper, this problem is solved by averaging the results of three repeated experiments. This solution does not increase the computing time cost by too much.

The mean absolute percentage error (MAPE), mean absolute error (MAE), root mean square error (RMSE), standard deviation of error (SDE), Pearson's correlation coefficient (R), and index of agreement (IA) are utilized to analyze the experimental results of

the prediction models; the index values of the above models for D1, D2, and D3 are given in Table 7. As can be seen from Table 7, the three datasets reflect the same model performance. In order to keep the length of the paper within a reasonable range, only D1 is selected for specific analysis. The PM_{2.5} concentration forecasting results for D1 are shown in Fig. 6. The R and IA results of the six prediction models for S1, S2, and T1–T4 are given in Table 8. The MAPE, MAE, RMSE, and SDE results of the six prediction models for S1 and S2 are given in Fig. 7. The MAPE, MAE, RMSE, and SDE results of the six prediction models for T1 and T2 are given in Fig. 8. The MAPE, MAE, RMSE and SDE results of the six prediction models for T3 and T4 are given in Fig. 9. It should be noted that since the values of R and IA do not belong to the same dimension as the other four evaluation indicators, they are not shown in the form of a graph.

In Tables 7 and 8 and Figs. 6–9, the proposed model has the smallest error evaluation values, and thus achieves accurate prediction in PM_{2.5} concentration forecasting. Compared with the other six comparison prediction models, the proposed model has better prediction accuracy from step-1 to step-3 predictions. This shows that the methods used in the hybrid model interact positively.

The prediction accuracy of the ESN-PSO model is better than that of the ESN model. This phenomenon indicates that the optimal parameters selected by the PSO algorithm can help to improve the prediction accuracy of the ESN model. The prediction accuracy of

Table 7
Error evaluation for the PM_{2.5} concentrations of D1, D2, and D3.

Forecasting model	Step	PM _{2.5} concentration																	
		MAPE (%)			MAE (µg·m ⁻³)			RMSE (µg·m ⁻³)			SDE (µg·m ⁻³)			R			IA		
		D1	D2	D3	D1	D2	D3	D1	D2	D3	D1	D2	D3	D1	D2	D3	D1	D2	D3
LSTM	Step-1	14.56	8.34	9.42	8.92	1.96	5.22	15.59	2.47	6.80	15.56	2.47	6.68	0.84	0.87	0.86	0.86	0.89	0.88
	Step-2	25.31	12.72	15.36	14.58	3.21	8.31	24.88	4.13	10.58	24.85	3.94	10.26	0.76	0.80	0.78	0.82	0.84	0.83
	Step-3	35.37	16.91	20.32	21.41	4.28	10.64	33.63	4.60	14.25	32.59	5.03	13.58	0.58	0.73	0.66	0.67	0.73	0.72
ESN	Step-1	13.48	7.35	6.35	8.61	1.80	3.65	14.14	2.29	4.84	13.77	2.28	4.78	0.85	0.88	0.87	0.88	0.90	0.89
	Step-2	24.75	12.31	12.35	13.72	3.06	7.05	23.62	3.71	9.70	23.46	3.52	9.63	0.78	0.76	0.80	0.82	0.86	0.84
	Step-3	34.82	16.55	19.20	20.59	3.97	10.57	32.57	5.04	13.76	32.36	4.88	12.83	0.63	0.72	0.69	0.75	0.82	0.78
ESN-PSO	Step-1	10.72	6.98	5.95	6.54	1.69	3.44	11.96	2.21	4.66	11.95	2.20	4.61	0.88	0.91	0.90	0.89	0.94	0.93
	Step-2	21.58	11.80	11.66	12.34	2.80	6.65	21.62	3.60	9.16	21.44	3.60	9.03	0.79	0.82	0.74	0.84	0.83	0.85
	Step-3	33.20	15.04	17.67	18.72	3.65	9.55	30.22	4.47	12.81	29.80	4.38	12.56	0.66	0.75	0.65	0.77	0.79	0.80
EWT-ESN-PSO	Step-1	4.88	1.97	0.69	2.88	0.60	0.36	5.04	0.74	0.47	4.97	0.74	0.46	0.92	0.93	0.94	0.96	0.95	0.94
	Step-2	6.04	2.27	0.89	3.46	0.68	0.47	5.57	0.86	0.61	5.35	0.86	0.58	0.86	0.90	0.90	0.93	0.92	0.91
	Step-3	8.42	2.85	1.70	4.43	0.88	0.88	6.65	1.14	1.07	6.12	1.11	1.01	0.79	0.87	0.78	0.89	0.90	0.89
RSAR-KC-ESN	Step-1	9.78	4.54	3.66	6.94	0.97	1.99	10.38	1.19	2.44	10.29	1.16	2.43	0.89	0.90	0.91	0.92	0.94	0.93
	Step-2	16.58	6.18	5.64	10.35	1.39	3.05	13.66	1.61	3.77	13.29	1.52	3.75	0.82	0.79	0.76	0.89	0.85	0.86
	Step-3	23.63	8.19	8.28	14.08	1.84	4.38	17.77	2.14	5.44	17.76	1.95	5.32	0.71	0.72	0.68	0.84	0.83	0.81
MCD-LSTM-PSO	Step-1	1.78	0.66	0.24	1.06	0.19	0.25	1.43	0.31	0.12	1.27	0.06	0.11	0.95	0.96	0.97	0.98	0.97	0.96
	Step-2	5.34	3.04	0.68	2.67	0.75	0.37	3.45	0.73	0.24	2.01	0.31	0.23	0.92	0.93	0.94	0.96	0.94	0.94
	Step-3	7.89	6.22	1.76	4.99	1.96	0.83	6.17	1.59	0.69	3.97	0.87	0.69	0.88	0.90	0.92	0.91	0.92	0.92
Proposed model	Step-1	1.67	0.57	0.18	0.88	0.11	0.04	1.18	0.19	0.05	1.11	0.01	0.05	0.96	0.97	0.98	0.98	0.99	0.99
	Step-2	4.47	2.05	0.48	2.42	0.42	0.12	2.97	0.61	0.15	1.80	0.19	0.14	0.95	0.96	0.95	0.97	0.97	0.97
	Step-3	7.66	5.46	1.36	4.37	1.11	0.34	5.26	1.48	0.46	3.52	0.60	0.41	0.92	0.92	0.93	0.95	0.93	0.95

the EWT-ESN-PSO model is better than that of the ESN-PSO model, which shows that the model accuracy can be improved by adding the EWT decomposition algorithm. The sequence obtained by the EWT algorithm is more stable and less random. Therefore, it is better to take the decomposed sublayers as the model input to obtain the prediction results. The prediction accuracy of the RSAR-KC-ESN model is better than that of the ESN model, which shows that the model accuracy can be improved by adding the RSAR-KC algorithm. After clustering, there is a larger difference between different groups of data and higher similarity between the same groups of data, which can improve the prediction accuracy of the original model to a certain extent.

Moreover, the accuracy of each prediction model decreases as the number of steps increases in Table 7 and Figs. 6–9. With the increase of the forecasting step, the error accumulation will

become increasingly serious, resulting in decline of the prediction accuracy.

The city with the best air quality is Changsha (D3), followed by Guangzhou (D2). The air quality of Beijing (D1) is relatively poor. The forecasting accuracy in Table 7 and Fig. 6 is consistent with this order. Moreover, the data in Fig. 7 show that samples with different pollution levels in the same area have no effect on model accuracy. The PM_{2.5} concentrations of S1 are smaller than those of S2, but the prediction accuracy of S2 is higher than that of S1. Therefore, it can be concluded that the prediction accuracy of the proposed model is better in cities with better air quality than in cities with heavy pollution.

In the above analysis, Tables 7 and 8 and Figs. 6 and 7 verify the validity of the data forecasts for different cities over the same time period. In order to verify the validity of the prediction for the same

Table 8
R and IA for the PM_{2.5} concentrations of S1, S2, and T1–T4.

Forecasting model	Step	R (µg·m ⁻³)						IA					
		S1	S2	T1	T2	T3	T4	S1	S2	T1	T2	T3	T4
LSTM	Step-1	0.81	0.82	0.78	0.85	0.86	0.83	0.85	0.87	0.83	0.88	0.90	0.87
	Step-2	0.74	0.75	0.69	0.77	0.78	0.73	0.79	0.79	0.75	0.81	0.84	0.82
	Step-3	0.56	0.58	0.51	0.60	0.62	0.54	0.62	0.64	0.57	0.75	0.77	0.71
ESN	Step-1	0.83	0.83	0.79	0.86	0.87	0.84	0.86	0.87	0.84	0.88	0.91	0.88
	Step-2	0.76	0.77	0.71	0.79	0.80	0.79	0.80	0.81	0.78	0.82	0.86	0.84
	Step-3	0.58	0.60	0.53	0.63	0.64	0.57	0.64	0.66	0.58	0.76	0.78	0.73
ESN-PSO	Step-1	0.85	0.86	0.82	0.88	0.88	0.87	0.90	0.89	0.87	0.90	0.92	0.89
	Step-2	0.78	0.81	0.75	0.81	0.83	0.81	0.84	0.85	0.81	0.84	0.88	0.87
	Step-3	0.61	0.65	0.60	0.70	0.71	0.64	0.72	0.76	0.63	0.81	0.80	0.76
EWT-ESN-PSO	Step-1	0.90	0.91	0.87	0.92	0.93	0.92	0.94	0.95	0.91	0.94	0.94	0.93
	Step-2	0.84	0.87	0.81	0.86	0.88	0.88	0.89	0.89	0.87	0.91	0.92	0.90
	Step-3	0.67	0.76	0.73	0.78	0.76	0.77	0.78	0.82	0.78	0.86	0.86	0.81
RSAR-KC-ESN	Step-1	0.86	0.87	0.83	0.88	0.89	0.88	0.89	0.91	0.88	0.91	0.92	0.90
	Step-2	0.80	0.83	0.77	0.82	0.85	0.84	0.85	0.86	0.83	0.84	0.89	0.88
	Step-3	0.65	0.67	0.62	0.68	0.72	0.68	0.74	0.72	0.67	0.82	0.81	0.79
MCD-LSTM-PSO	Step-1	0.96	0.97	0.95	0.97	0.97	0.96	0.98	0.98	0.96	0.98	0.98	0.97
	Step-2	0.94	0.96	0.93	0.96	0.96	0.95	0.95	0.96	0.95	0.97	0.97	0.96
	Step-3	0.88	0.91	0.90	0.93	0.94	0.91	0.91	0.93	0.93	0.95	0.96	0.94
Proposed model	Step-1	0.98	0.98	0.96	0.99	0.99	0.98	0.99	0.99	0.97	0.99	0.99	0.99
	Step-2	0.96	0.97	0.94	0.97	0.98	0.96	0.97	0.98	0.96	0.98	0.98	0.97
	Step-3	0.92	0.93	0.91	0.94	0.95	0.92	0.94	0.95	0.94	0.96	0.96	0.94

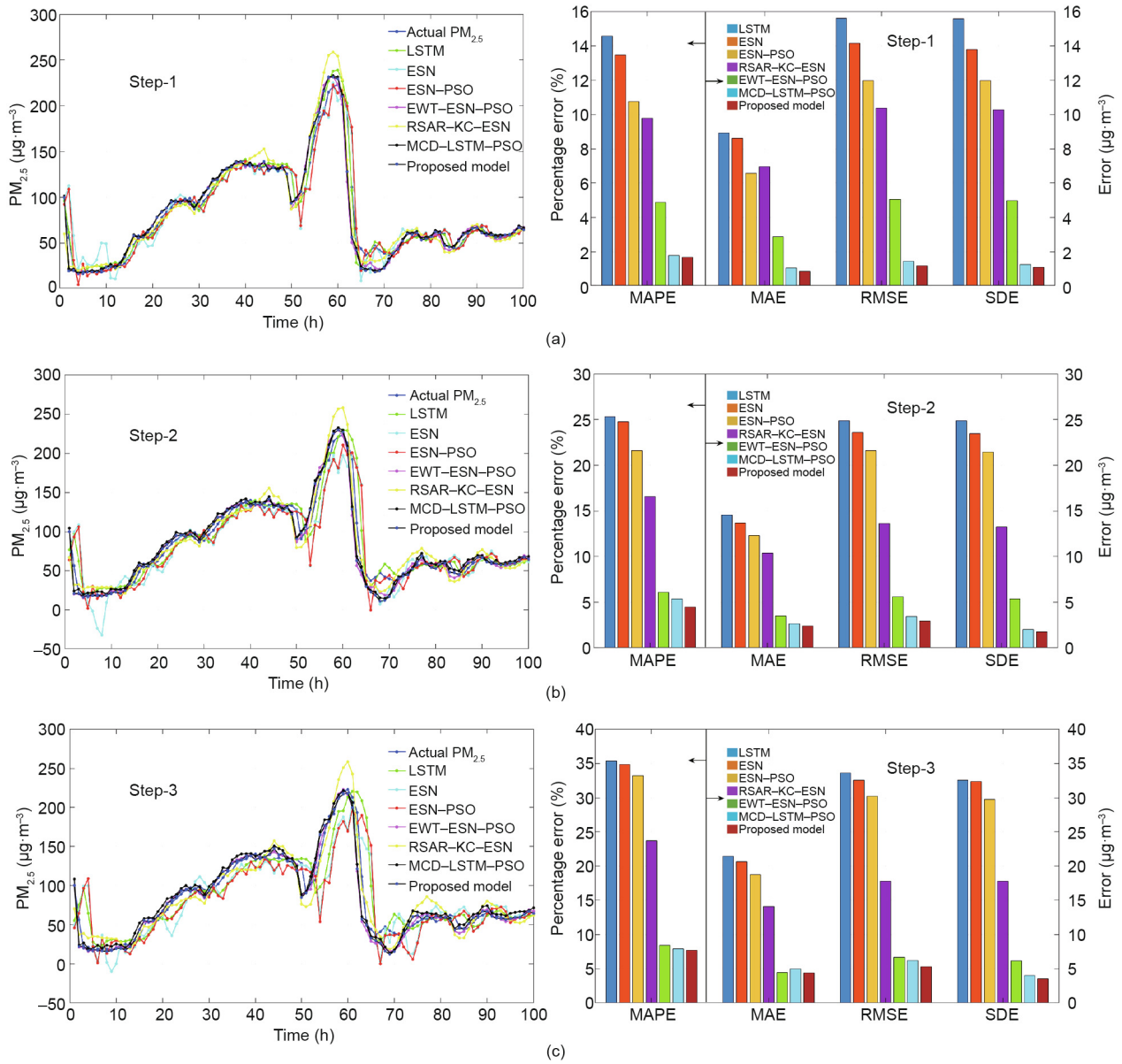


Fig. 6. Results of the various ahead-step predictions for the $PM_{2.5}$ concentrations of D1. (a) Step-1; (b) step-2; (c) step-3.

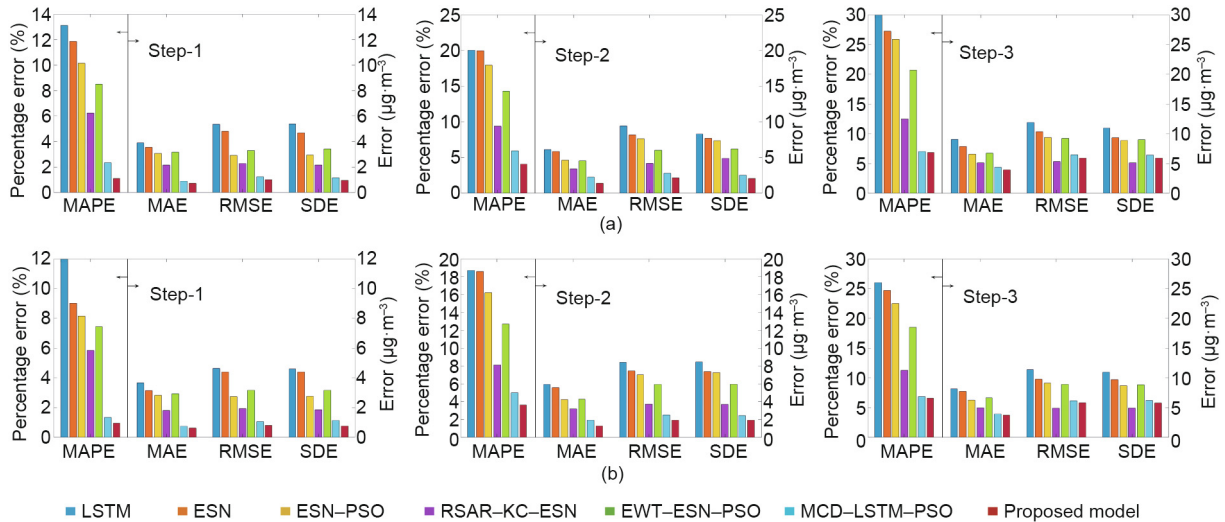


Fig. 7. Error evaluation for the $PM_{2.5}$ concentrations of (a) S1 and (b) S2.

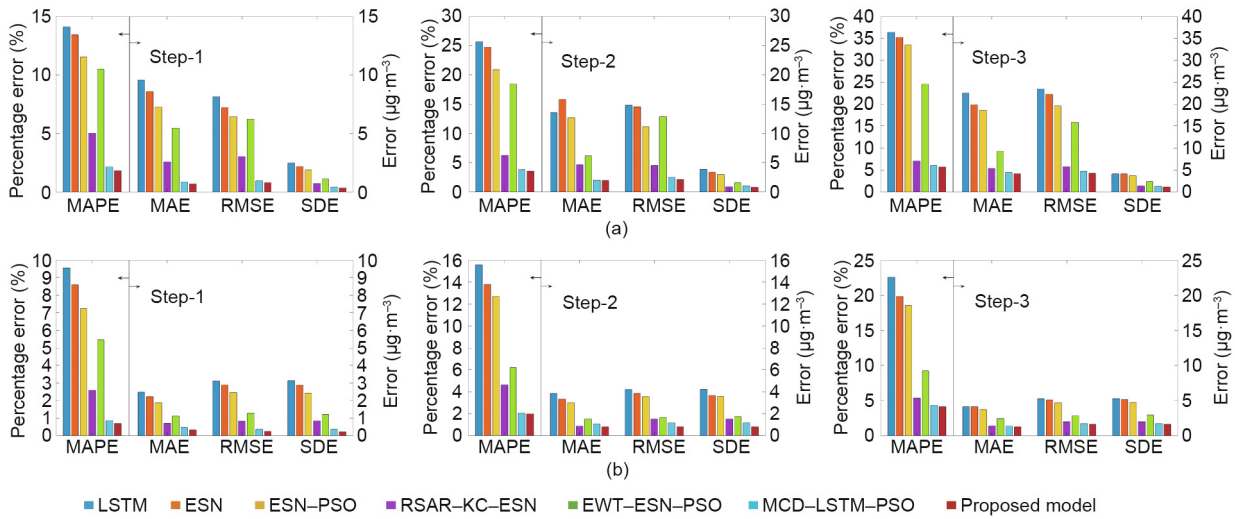


Fig. 8. Error evaluation for the PM_{2.5} concentrations of (a) T1 and (b) T2.

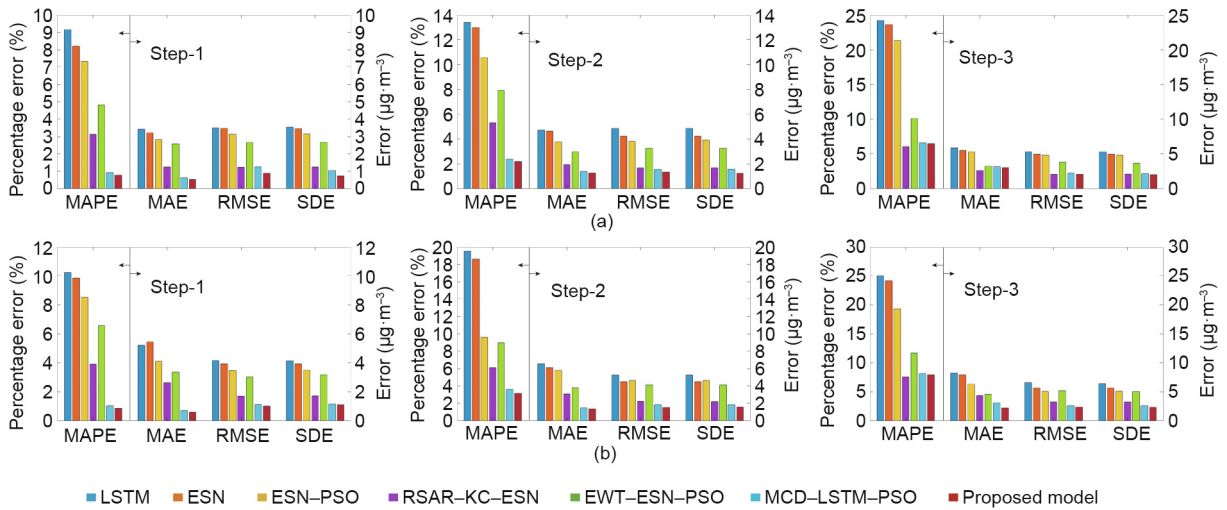


Fig. 9. Error evaluation for the PM_{2.5} concentrations of (a) T3 and (b) T4.

city over different time periods, the experiments shown in Figs. 8 and 9 were carried out. According to the data in Figs. 8 and 9, the proposed model maintained a stable prediction effect with the change of time period. In other words, the data in Figs. 8 and 9 verify the stability and validity of the proposed model over the whole year.

In this study, the following computing is implemented in the simulation environments: Intel i5-6500 CPU 3.2 GHz, RAM 8 GB. Table 9 gives the computation time of the comparison models in D1. Because both RSAR-KC and PSO are offline processing, the computation time is not compare with them.

The computation speed of the ESN is much faster than that of the LSTM, owing to the advantages of the ESN network itself. Because of the existence of the reserve pool, it is only necessary to train the output weight in the training process of the ESN network, which greatly improves the computing speed.

After adding the EWT decomposition algorithm, the computational speed of the model decreases to a certain extent. Because each decomposition layer needs to be trained and predicted, the computational speed of the original model plays a vital role here, which further reflects the superiority of the ESN.

The change in prediction steps has little effect on the calculation speed of the model. This may be because the computational capacity of the algorithm model is relatively large.

4. Conclusions

This study establishes an improved hybrid ESN forecasting model to predict and analyze the hourly average concentrations

Table 9
Computation time of the comparison models in D1.

Forecasting model	Step	Computation time (s)
LSTM	Step-1	18.00
	Step-2	18.26
	Step-3	18.15
ESN	Step-1	0.06
	Step-2	0.08
	Step-3	0.07
EWT-ESN	Step-1	1.05
	Step-2	1.03
	Step-3	1.03

of PM_{2.5} based on the MCD method and the PSO. The proposed hybrid model is compared with several benchmark models to prove its effectiveness. The attribute reduction results show that the concentrations of SO₂ and O₃ play an important role in predicting the concentrations of PM_{2.5}. Moreover, research on the influence of relevant meteorological parameters will be carried out in future studies. The clustering results show that PM_{2.5} concentration data become stationary and regular after the clustering processing. These features are useful and conducive for ESN-based deep training. The prediction results show that: ① The MCD method can improve the accuracy of models; ② the proposed hybrid model has better prediction accuracy than other relevant deep learning or single models; ③ the proposed hybrid model has achieved good experimental results with PM_{2.5} pollutant concentration data from four cities in China; and ④ the proposed hybrid PM_{2.5} forecasting framework can also be applied in other air pollution time series multi-step predictions. The forecasted results can be embedded in relevant early warning systems for urban air pollution management.

The main contributions of this study can be summarized as follows:

(1) A novel PM_{2.5} concentrations multi-step prediction model is developed based on the MCD, ESN, and PSO, which yields accurate forecasting performance for hourly average PM_{2.5} concentrations. The multi-step forecasting results can be used for the development of PM_{2.5} pollution warning systems.

(2) A novel decomposition method in hybrid PM_{2.5} concentration forecasting named MCD is developed. This method integrates the feature extraction into decomposition. Multi-dimensional KC clustering is carried out using the feature extraction results of the RSAR algorithm, which not only guarantees the effectiveness of the clustering results, but also considers the influence of multi-dimensional features. The EWT algorithm-based KC algorithm is then employed for data preprocessing. The clustering algorithm is used to group the original PM_{2.5} concentrations according to different PM_{2.5} concentration scenarios. Next, combined with the EWT decomposition algorithm, raw PM_{2.5} concentrations data are distinguished according to different characteristics in the timescale. Finally, the optimization function of the decomposition is realized.

(3) In the proposed hybrid PM_{2.5} concentration forecasting model, the ESN is employed as a predictor. The sparse connection of neurons in the reservoir of the ESN not only improves the convergence of the neural network model, but also enhances the model generalization. This characteristic can reduce the probability of the overfitting problem in the process of model training. Moreover, the ESN has good real-time performance in computing.

Acknowledgements

The study is fully supported by the National Natural Science Foundation of China (61873283), the Changsha Science & Technology Project (KQ1707017) and the Innovation Driven Project of the Central South University (2019CX005).

Compliance with ethics guidelines

Hui Liu, Zhihao Long, Zhu Duan, and Huipeng Shi declare that they have no conflict of interest or financial conflicts to disclose.

References

- [1] Ding Y, Wu P, Liu Y, Song Y. Environmental and dynamic conditions for the occurrence of persistent haze events in North China. *Engineering* 2017;3(2):266–71.
- [2] Polezer G, Tadano YS, Siqueira HV, Godoi AF, Yamamoto CI, de André PA, et al. Assessing the impact of PM_{2.5} on respiratory disease using artificial neural networks. *Environ Pollut* 2018;235:394–403.
- [3] Zheng S, Pozzer A, Cao C, Lelieveld J. Long-term (2001–2012) fine particulate matter (PM_{2.5}) and the impact on human health in Beijing, China. *Atmos Chem Phys Discuss* 2014;14:28657–84.
- [4] Zhao C, Lin Y, Wu F, Wang Y, Li Z, Rosenfeld D, et al. Enlarging rainfall area of tropical cyclones by atmospheric aerosols. *Geophys Res Lett* 2018;45(16):8604–11.
- [5] Zhao C, Yang Y, Fan H, Huang J, Fu Y, Zhang X, et al. Aerosol characteristics and impacts on weather and climate over the Tibetan Plateau. *Natl Sci Rev* 2020;7(3):492–5.
- [6] Zhao C, Garrett TJ. Effects of Arctic haze on surface cloud radiative forcing. *Geophys Res Lett* 2015;42(2):557–64.
- [7] Garrett TJ, Zhao C. Increased Arctic cloud longwave emissivity associated with pollution from mid-latitudes. *Nature* 2006;440(7085):787–9.
- [8] Huang K, Xiao Q, Meng X, Geng G, Wang Y, Lyapustin A, et al. Predicting monthly high-resolution PM_{2.5} concentrations with random forest model in the North China Plain. *Environ Pollut* 2018;242:675–83.
- [9] Pan Y, Tian Y, Liu X, Gu D, Hua G. Urban big data and the development of city intelligence. *Engineering* 2016;2(2):171–8.
- [10] Guo H, Cheng T, Gu X, Wang Y, Chen H, Bao F, et al. Assessment of PM_{2.5} concentrations and exposure throughout China using ground observations. *Sci Total Environ* 2017;601–602:1024–30.
- [11] Zhao C, Wang Y, Shi X, Zhang D, Wang C, Jiang JH, et al. Estimating the contribution of local primary emissions to particulate pollution using high-density station observations. *J Geophys Res Atmos* 2019;124(3):1648–61.
- [12] Prakash D, Christian H, Winston H, Kevin G, Jia-Yeong K, Gopal S. A retrospective comparison of model-based forecasted PM_{2.5} concentrations with measurements. *Air Repair* 2010;60(11):1293–308.
- [13] Sun W, Zhang H, Palazoglu A, Singh A, Zhang W, Liu S. Prediction of 24-hour-average PM_{2.5} concentrations using a hidden Markov model with different emission distributions in northern California. *Sci Total Environ* 2013;443(3):93–103.
- [14] Biancofiore F, Busilacchio M, Verdecchia M, Tomassetti B, Aruffo E, Bianco S, et al. Recursive neural network model for analysis and forecast of PM₁₀ and PM_{2.5}. *Atmos Pollut Res* 2017;8(4):652–9.
- [15] Jahed Armaghani D, Raja Shoib RSNB, Faizi K, Rashid ASA. Developing a hybrid PSO-ANN model for estimating the ultimate bearing capacity of rock-socketed piles. *Neural Comput Appl* 2017;28(2):391–405.
- [16] Niu M, Wang Y, Sun S, Li Y. A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM_{2.5} concentration forecasting. *Atmos Environ* 2016;134:168–80.
- [17] Gan K, Sun S, Wang S, Wei Y. A secondary-decomposition-ensemble learning paradigm for forecasting PM_{2.5} concentration. *Atmos Pollut Res* 2018;9(6):989–99.
- [18] Cheng Y, Zhang H, Liu Z, Chen L, Wang P. Hybrid algorithm for short-term forecasting of PM_{2.5} in China. *Atmos Environ* 2019;200:264–79.
- [19] Wu Q, Lin H. A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors. *Sci Total Environ* 2019;683:808–21.
- [20] Liu H, Duan Z, Chen C. A hybrid framework for forecasting PM_{2.5} concentrations using multi-step deterministic and probabilistic strategy. *Air Qual Atmos Hlth* 2019;12(7):785–95.
- [21] Stidworthy A, Jackson M, Johnson K, Carruthers D, Stocker J. Evaluation of local and regional air quality forecasts for London. *Int J Environ Pollut* 2018;64(1–3):178–91.
- [22] Zhu S, Lian X, Liu H, Hu J, Wang Y, Che J. Daily air quality index forecasting with hybrid models: a case in China. *Environ Pollut* 2017;231:1232–44.
- [23] Sun W, Sun J. Daily PM_{2.5} concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm. *J Environ Manage* 2017;188:144–52.
- [24] Reátegui-Romero W, Sánchez-Cocoylo OR, de Fatima Andrade M, Moya-Alvarez A. PM_{2.5} estimation with the WRF/Chem model, produced by vehicular flow in the lima metropolitan area. *Open J Air Pollut* 2018;7(3):215.
- [25] Zhu S, Lian X, Wei L, Che J, Shen X, Yang L, et al. PM_{2.5} forecasting using SVR with PSO-GSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors. *Atmos Environ* 2018;183:20–32.
- [26] Niu M, Hu Y, Sun S, Liu Y. A novel hybrid decomposition-ensemble model based on VMD and HGWO for container throughput forecasting. *Appl Math Modell* 2018;57:163–78.
- [27] Liu H, Jin K, Duan ZJ. Air PM_{2.5} concentration multi-step forecasting using a new hybrid modeling method: comparing cases for four cities in China. *Atmos Pollut Res* 2019;10(5):1588–600.
- [28] Wang Q, Zeng Q, Tao J, Sun L, Zhang L, Gu T, et al. Estimating PM_{2.5} concentrations based on MODIS AOD and NAQPMS data over Beijing–Tianjin–Hebei. *Sensors* 2019;19(5):1207.
- [29] Da L, Wang J, Hui W. Short-term wind speed forecasting based on spectral clustering and optimised echo state networks. *Renew Energ* 2015;78:599–608.
- [30] Xu L, Yu Y, Yu J, Chen J, Niu Z, Yin L, et al. Spatial distribution and sources identification of elements in PM_{2.5} among the coastal city group in the Western Taiwan Strait region, China. *Sci Total Environ* 2013;442(1):77–85.
- [31] Li C, Zhu Z. Research and application of a novel hybrid air quality early-warning system: a case study in China. *Sci Total Environ* 2018;626:1421–38.
- [32] Wang C, Shao M, He Q, Qian Y, Qi Y. Feature subset selection based on fuzzy neighborhood rough sets. *Knowl-Based Syst* 2016;111:173–9.
- [33] Wang S, Li Q, Yuan H, Li D, Geng J, Zhao C, et al. δ -Open set clustering—a new topological clustering method. *WIRES Data Mining Knowl Discov* 2018;8(6):e1262.

- [34] Yu S, Chu S, Wang C, Chan Y, Chang T. Two improved k -means algorithms. *Appl Soft Comput* 2018;68:747–55.
- [35] Zhang Q, Yang LT, Chen Z, Li P. High-order possibilistic c -means algorithms based on tensor decompositions for big data in IoT. *Inf Fusion* 2018;39:72–80.
- [36] Majumdar J, Udandakar S, Bai BM. Implementation of cure clustering algorithm for video summarization and healthcare applications in big data. In: *Emerging Research in Computing, Information, Communication and Applications*. Singapore: Springer; 2019. p. 553–64.
- [37] Gilles J. Empirical wavelet transform. *IEEE Trans Signal Process* 2013;61(16):3999–4010.
- [38] Chitsazan MA, Fadali MS, Trzynadlowski AM. Wind speed and wind direction forecasting using echo state network with nonlinear functions. *Renew Energy* 2019;131:879–89.
- [39] Fan H, Zhao C, Yang Y. A comprehensive analysis of the spatio-temporal variation of urban air pollution in China during 2014–2018. *Atmos Environ* 2020;220:117066.
- [40] Zhang K, Zhao C, Fan H, Yang Y, Sun Y. Toward understanding the differences of $PM_{2.5}$ characteristics among five China urban cities. *Asia-Pacific J Atmos Sci* 2019;55(2):1–10.
- [41] Shi X, Zhao C, Wang C, Jiang J, Yung Y. A method of examination about the spatial representation of $PM_{2.5}$ obtained from a network of limited surface stations. *J Geophys Res Atmos* 2018;123:3145–58.
- [42] Rokach L, Maimon O. Clustering methods. In: *Data Mining and Knowledge Discovery Handbook*. Boston: Springer; 2005. p. 321–52.
- [43] Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 1999;20(20):53–65.