Journal of Cleaner Production 234 (2019) 54–70

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Research and application of the hybrid forecasting model based on secondary denoising and multi-objective optimization for air pollution early warning system

Jianzhou Wang [a], Lu Bai [a, *], Siqi Wang [a], Chen Wang [b]

[a] School of Statistics, Dongbei University of Finance and Economics, Dalian, 116023, PR China
[b] School of Mathematics & Statistics, Lanzhou University, Lanzhou, 730000, PR China

ARTICLE INFO

Article history:
Received 4 November 2018
Received in revised form
6 June 2019
Accepted 17 June 2019
Available online 18 June 2019

Handling Editor: Dr. Govindan Kannan

Keywords:
Early warning system
Air pollution forecasting
Secondary denoising
Multi-objective optimization algorithm
Fuzzy synthetic evaluation

ABSTRACT

With the increasing irreversible damage caused by air pollution, an early warning system to send warning information to human beings so that they can avoid more harm caused by air pollution is required. A reliable warning system can provide valuable information to protect mankind from the effects of pollution and can act as a tool that allows regulators to implement corresponding measures to reduce air pollution. However, the previous most valuable research studies were focused on pollution forecasting and the extent to which pollution affects health, and the aim of only a few studies was to analyze pollution from an application perspective and to construct a reasonable early warning system. In this study, an air pollution early warning system was constructed, which comprises two modules: an air pollution forecasting module and an air quality evaluation module. In the forecasting module, two denoising methods and a multi-objective optimization algorithm are integrated into a novel hybrid forecasting model. In the evaluation module, fuzzy synthetic evaluation is used to evaluate air quality objectively. To verify the performance of the proposed early warning system, hourly pollutants concentration data were used in a case study of three metropolises in China and three numeric simulation experiments were conducted. The simulation results show that the forecasting performance of the $L_{2,1}$RF-ELM model used in this study is better than the traditional neural network, and the forecasting model proposed in this paper is better than the traditional statistical model ARIMA. Moreover,the early warning system performed well in terms of highly accurate forecasting and accurate evaluation in the three research areas.

1. Introduction

For nearly a century, the rapid development of industrialization and urbanization has increased the amount of energy consumed by human activities and caused serious air pollution in the world. Scholars have conducted a significant number of air pollution research studies. The results of extensive studies indicate that exposure to air pollution can cause a variety of diseases (Cohen et al., 2017; Guo et al., 2016). Moreover, air pollution can also be detrimental to the ecosystem, leading to the greenhouse effect, ozone layer destruction, acid rain, reduced solar radiation, etc. (Anwar et al., 2016; Desonie, 2007; Ramanathan and Feng, 2009). Therefore, accurate and authentic air quality information is increasingly needed to enable industries to minimize their production of pollutants and residents to adjust their activities promptly to mitigate the damage caused by major pollution.

To diminish the effects of air pollution, scholars have focused on analyzing and forecasting the concentrations of pollutants, devoting their efforts to providing highly accurate forecasting. During the past one hundred years, many forecasting methods were proposed, the most popular of which can be classified into three categories: physical, statistical, and artificial intelligence models (Bai et al., 2018).

Physical models use the physicochemical process of pollutants in the atmosphere as the entry point for forecasting pollutant concentrations. Statistical methods can be divided into causal models and time series models according to their fundamental characteristics. The assumption of causal models is that the

* Corresponding author. No. 217, Jianshan Street, Dalian, China.
E-mail address: baildctg@hotmail.com (L. Bai).

https://doi.org/10.1016/j.jclepro.2019.06.201
0959-6526/

Nomenclature

Abbreviations

AATVs	Average Actual Values
ANNs	Artificial Neural Networks
AQI	Air Quality Index
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
BPNN	Back-Propagation Neural Network
CW-SVR	Chance Weighted Support Vector Regression
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
EEMD	Ensemble Empirical Mode Decomposition
EOF	Empirical Orthogonal Function
FB	Fractional Bias
FSE	Fuzzy Synthetic Evaluation
IMFs	Intrinsic Mode Functions
$L_{2,1}$RFELM	ELM based on $L_{2,1}$-norm and Random Fourier
MAE	Mean Absolute Error
MAPE	Mean Average Percentage Error
MFs	Membership Functions
MOALO	Multi-Objective Ant Lion Optimizer
PCs	Principal Components
R	Pearson's correlation coefficient
RF	Random Fourier
RMSE	Root Mean Square Error
SLFN	Single Hidden Feedback Network
SSA	Singular Spectrum Analysis
SVD	Singular Value Decomposition
SEMR	SSA-EEMD-MOALO-$L_{2,1}$RFELM
SEME	SSA-EEMD-MOALO-ELM
MR	MOALO-$L_{2,1}$RFELM
EMR	EEMD-MOALO-$L_{2,1}$RFELM

historical relationship between dependent and independent variables will remain valid in the future. The main idea of time series models is that history will repeat itself, and the forecast values are thus based on the information obtained from the past and current data (Liu et al., 2016). Moreover, with the development of artificial intelligence arithmetic, artificial intelligence models have been widely used in many fields.

However, physical forecasting models, including mesoscale numerical models and fluid dynamic-related methods, require physical or meteorological information as input variables. These models consider large scale simulation data and consume excessive computing resources (Zhao et al., 2017). In addition, because of the inherent limitations of statistical methods, they show a poor extrapolation effect and narrow forecasting range. Moreover, artificial intelligence models also have some defects, such as overfitting and a low convergence rate, and they easily get stuck in a local optimum. Furthermore, raw air pollution data are characterized by instability and noise. When the amount of data is insufficient, individual models lead to imprecise forecasting. Finally, the concept of a comprehensive evaluation aimed at determining the degree of pollution is vague and it is difficult to find a clear boundary, and therefore, fuzzy logic is more suitable for air quality evaluation. However, the commonly used evaluation tool today is the Air Quality Index (AQI) (Rahman et al., 2015, 2017; Zhu et al., 2017).

As can be understood from the analysis above, it is necessary to develop an effective air quality early warning system for China, and even for the world. In our study, a new early warning system was developed, which contains two modules: a hybrid forecasting module and a synthetic evaluation module. In the forecasting module, where the "decomposition and ensemble" theory is applied, the most recently developed forecasting method, an extreme learning machine (ELM) based on $L_{2,1}$-norm and random fourier ($L_{2,1}$RFELM) and the most recently developed bio-inspired optimization algorithm called the multi-objective ant lion optimizer (MOALO) are integrated to form a hybrid model. The forecasting model is constructed as follows. First, the perspective of dynamic reconstruction of time series is applied combined with empirical orthogonal functions that can reduce noise and identify the components that show trends and periodic changes in the original sequence. In order to improve the forecasting accuracy and retain as much as possible of the effective information contained in the original data, by adding white noise to the sequence the signals of different scales are automatically distributed to the appropriate reference scale to decompose the signals again. The $L_{2,1}$RFELM model optimized by MOALO is used to forecast the non-major trend series. In an early warning system, air quality evaluation is also indispensable. In view of the uncertainty for evaluation, fuzzy synthetic evaluation (FSE) was applied in the second (evaluation) part. By using these methods, a more effective forecasting and accurate evaluation results can be achieved, and thus, valuable information can be provided to the public and governments.

The main contributions and innovations of this study are as follows.

(1) The original data were decomposed using two different denoising methods, which are based on the decomposition and ensemble data-preprocessing technique. The raw air pollution data were transformed into a filtered time series after being decomposed and reconstructed. Data preprocessing could decrease the irregularity of the air pollution data and the noise that they contain. In this study, pollution forecasting accuracy was effectively improved.

(2) A novel forecasting method based on EML was successfully applied in the proposed hybrid forecasting model. The forecasting method in the forecasting module is a novel forecasting algorithm, which is used for the first time in the field of pollution forecasting. The results of the experiments conducted in this study proved that the air pollution forecasting performance of the novel forecasting method is good.

(3) Three main pollutants were selected in each city included in the study by using FSE based on the average of pollutant concentrations. Experiments on the three main pollutants were then conducted. The experimental results indicate that these three pollutants could be used to estimate the air quality in each city, and thus, it is possible to reduce the program run time and achieve air pollution forecasting that is more individualized for different cities.

(4) An air quality early warning system that is feasible and operable was established. The system comprises accurate forecasting and effective evaluation. This system was tested using air pollution data for Beijing, Shanghai, and Guangzhou in China. All the experimental results reflect the excellent performance of the system. It is reasonable to believe that this early warning system can be used for daily life guidance to reduce the impact of environmental pollution.

And for the convenience of the readers, all abbreviations in this

paper are concluded in a table as follow:

The rest of this paper is organized as follows. Section 2 outlines the methodologies that were used in the air pollution forecasting module. Section 3 introduces the theory of FSE and its application for selecting three main pollutants. The experiments and evaluation results are reported in Section 4. Section 5 concludes this article.

2. Preprocessing of air pollution concentration data

The methods that were used in the process of preprocessing are introduced in this section, including singular spectrum analysis (SSA), Ensemble empirical mode decomposition (EEMD). Like most time series, the original air pollution data is unstable and contains noise. These noises and instability are useless or even disadvantageous to forecast. In order to improve the accuracy of forecasting, it is necessary to pre-process the data and remove the interference information from the original time series.

2.1. Initial denosing

SSA is a typical technique for analyzing time series data when no statistical hypotheses exist about the signal or noise (Ma et al., 2017). This technique can obtain a reconstructed series by refining and identifying trends or quasi-periodic components and reducing noise in the original signal (Wang et al., 2017b). Although the main principle of SSA is simple, the function of this method is powerful. In practical applications, SSA performs better than nonparametric methods, such as empirical mode decomposition, which potentially suffer from the mode-mixing problem. EEMD cannot entirely neutralize the added white noise(Du et al., 2016).

In this section, we introduce the principle of SSA to explain its implementation and usage. The process of SSA comprises two stages and each stage includes two steps.

Stage 1 Decomposition. The main purpose of the decomposition stage is to decompose the series. This stage consists of two steps.

Step 1 Embedding. This step can transfer the original time series, a one-dimensional time series, into a multi-dimensional time series.

The original series is defined as

$$\mathbf{X} = (x_1, x_2, \cdots, x_T) \tag{1}$$

Then, the window length is denoted by $L(2 \le L \le T)$. The selection rules of L are as follows.

(1) It is preferable that L be less than 1/3 of the length of the original series;
(2) If the raw data are periodic, Lshould optimally be an integer multiple of the period.

After the window length has been determined, the multi-dimensional series $\mathbf{Y}$ is also determined. Y is frequently called a trajectory matrix and can be expressed as:

$$\mathbf{Y} = [\mathbf{Y}_1\mathbf{Y}_2\cdots\mathbf{Y}_K] = \begin{bmatrix} x_1x_2x_3\cdots x_K \\ x_2x_3x_4\cdots x_{K+1} \\ \vdots\vdots\vdots \\ x_Lx_{L+1}x_{L+2}\cdots x_T \end{bmatrix} \tag{2}$$

where$K = T - L + 1$. $\mathbf{Y}$ has two important properties (Gao et al., 2016):

(1) The elements of matrix $\mathbf{Y}$ are from the original series and every row and column is a subseries of $\mathbf{X}$.
(2) $Y$is a Hankel matrix, because the elements on the anti-diagonals are equal.

Step 2 Singular value decomposition (SVD). The main operation in this step is the decomposition of the trajectory matrix $\mathbf{Y}$.

First, the matrix $\mathbf{B} = \mathbf{Y}\mathbf{Y}^T$ is calculated.

Then, a corresponding orthogonal matrix $\mathbf{U} = [u_1, u_2, \cdots, u_L] \in \mathbb{R}^{L\times L}$of the matrix B and $\mathbf{V} = [v_1, v_2, \cdots, v_L] \in \mathbb{R}^{L\times L}$ is defined. We can establish the formula

$$\mathbf{B} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{3}$$

From Eq. (4), we can obtain the eigenvalues matrix $\mathbf{S}$ of $\mathbf{B}$:

$$\mathbf{S} = [diag(\lambda_1, \lambda_2, \cdots, \lambda_L), \mathbf{O}] \quad \lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_L \ge 0 \tag{4}$$

where $\mathbf{O}$ is the zero matrix, $\mathbf{U}$ is the corresponding eigenvector, also called the empirical orthogonal function (EOF), and V represents the principal components (PCs).

Finally, the SVD of the trajectory matrix Ycan be expressed as:

$$\mathbf{Y} = \sum_{i=1}^{d} \mathbf{Y}_i \quad d = max(i, \lambda_i > 0) = rank\,\mathbf{Y} \tag{5}$$

where $\mathbf{Y}_i = \sqrt{\lambda_i}u_iv_i^T \quad i = 1, 2, \cdots, d$ is the i-th component of the SVD, $v_i = Y^Tu_i \big/ \sqrt{\lambda_i}$ represents the singular vector of singular values $\sqrt{\lambda_i}$of $\mathbf{Y}$, and the set of $\{\sqrt{\lambda_i}\}$ is the spectrum of $\mathbf{Y}$.

Stage 2 Reconstruction. In this stage, the signals represented by each eigenvalue are analyzed and combined to reconstruct a new time series. This stage also consists of two steps.

Step 1 Grouping. Y is divided into n groups, which are non-connected. First, the set of indices $\mathbf{I} = \{1, \cdots, \mathrm{d}\}$ are split into several groups $\mathbf{I}_1, \cdots, \mathbf{I}_n$; then, Y_i within each group is summed into Y_I. The result of this step is expressed as (Golyandina et al., 2001):

$$\mathbf{Y} = [\mathbf{Y}_{I1}, \mathbf{Y}_{I2}, \cdots, \mathbf{Y}_{In}], \text{ Where } \mathbf{Y}_{I_n} = \sum_{i\in I_n} Y_i \tag{6}$$

Step 2 Diagonal averaging. The purpose of this step is to convert the matrix obtained in the previous step into a sequence of length N. Suppose Z is an $L \times K$ matrix, and $L^* = min(L, K)$ and $K^* = max(L, K)$. If $L < K$, then, $z_{ij}^* = z_{ij}$; else, $z_{ij}^* = z_{ij}^T = z_{ji}$. The restructured sequence $RC = (rc_1, rc_2, \cdots, rc_N)$ can be converted to a series using

$$rc_k = \begin{cases} \dfrac{1}{k}\displaystyle\sum_{q=1}^{k+1} z^*_{q,k-q+2} & 1 \le k \le L^* \\ \dfrac{1}{L^*}\displaystyle\sum_{q=1}^{L^*} z^*_{q,k-q+2} & L^* \le k \le K^* \\ \dfrac{1}{N-k}\displaystyle\sum_{q=k-K^*+2}^{k-K^*+1} z^*_{q,k-q+2} & K^* \le k \le N \end{cases} \tag{7}$$

In the SSA technique, the first r constituents can be considered the vital information and the remaining N-r constituents can be considered the noise of the original data (Du et al., 2016).

2.2. Secondary denoising

In order to overcome the shortcomings of Fourier spectral analysis and other traditional data analysis methods, Huang et al. (1998) devised a novel time-frequency analysis algorithm, called empirical mode decomposition (EMD) (Huang et al., 1998). EMD was effectively applied in many fields soon after it was proposed (Boudraa and Cexus, 2007; Heng et al., 2016). However, EMD also cannot process data effectively because of its weaknesses, such as the mode-mixing problem. Therefore, in order to improve the data processing capabilities of EMD, Wu et al. (2009) proposed an improved method named ensemble empirical mode decomposition (EEMD) (Wu and Huang, 2009). As compared to EMD, EEMD can offset noise by adding paired white noise. The main steps of EEMD are as follows.

Step 1. White noise is added to the original data $X(t)$, such as positive white noise $w(t)$ and negative noise $-w(t)$. The white noise amplitude is given as 0.1–0.4 times the standard deviation of the original data.

$$\begin{cases} \boldsymbol{P}(t) = X(t) + w(t) \\ \boldsymbol{N}(t) = X(t) - w(t) \end{cases} \tag{8}$$

Step 2. ($\boldsymbol{P}(t)$,$\boldsymbol{N}(t)$) is decomposed into K intrinsic mode functions (IMFs), expressed as$I_j(t)$ $(j = 1, 2, \cdots, K)$by using the EMD technique.

Step 3. Steps 1 and 2 are repeated M times and different white noise components are added. We obtain the i-th series with white noise:

$$\begin{cases} \boldsymbol{P}_i(t) = \sum_{j=1}^{K} \boldsymbol{I}_{ij}^{+}(t) \\ \boldsymbol{N}_i(t) = \sum_{j=1}^{K} \boldsymbol{I}_{ij}^{-}(t) \end{cases} \quad i = 1, 2, \cdots, M; j = 1, 2, \cdots, K \tag{9}$$

where $P_i(t)$ and $N_i(t)$ are the i-th series after white noise is added and $\boldsymbol{I}_{ij}(t)$ is the j-th IMF of the i-th trial.

Step 4. The ensemble mean of the corresponding IMFs of the decomposition is calculated as the data processing result (Wang et al., 2017a).

$$\boldsymbol{X}(t) = \boldsymbol{I}_j(t) \tag{10}$$

$$\boldsymbol{I}_j(t) = \frac{1}{2M} \sum_{i=1}^{M} \left(\boldsymbol{I}_{ij}^{+}(t) + \boldsymbol{I}_{ij}^{-}(t) \right) \tag{11}$$

3. Introduction of multi-objective optimization algorithm and forecasting algorithm

The air pollution forecasting model proposed in this paper not only combines the idea of secondary denoising, but also uses multi-objective optimization algorithm to find the best parameters of the forecasting algorithm in order to improve the accuracy of the forecasting model. This section will introduce the optimization algorithm and forecasting algorithm used in the air pollution forecasting model.

3.1. Multi-objective ant lion optimizer

The multi-objective ant lion optimizer (MOALO) is a multi-objective version of the ant lion optimizer (ALO), proposed by Mirjalili in 2016. It was inspired by the foraging behavior of ant lion larvae in nature, and the implementation of this algorithm is mainly through ant random walks, building traps, trapping ants into traps, and reconstructing traps. In contrast to those of ALO, the results of MOALO are a Pareto optimal solution set, and the fitness of the ant lion is determined by using the roulette wheel mechanism. The main steps and mathematical modeling as presented by (Du et al., 2017; Mirjalili et al., 2017) are provided below.

Step 1. In nature, large ant lions set large traps to increase their chances of survival, and therefore, the MOALO algorithm uses a roulette mechanism to ensure that strong ant lions are able to capture as much prey as possible. Moreover, to improve the distribution of the Pareto optimal solutions in the archive, two mechanisms are considered.

(1) The ant lions are selected based on the least populated neighborhood strategy. The probability of selecting a solution in the archive can be calculated as

$$\mathbf{P}_i = \frac{k}{N_i} 1 < k < N_i \tag{12}$$

where N_i is the number of solutions of the i-th solution in the vicinity and k is a constant.

(2) When the archive is full, it is necessary to remove the solution with the most populated neighborhood. The probability of a solution being removed can be expressed as

$$\mathbf{P}_\mathbf{r} = \frac{1}{\mathbf{P}_i} = \frac{N_i}{k} \tag{13}$$

Step 2. The random walks of ants can be simulated as

$$\boldsymbol{X}(t) = [0, cumsum(2f(t_1) - 1), cumsum(2f(t_2) - 1), \cdots, cumsum(2f(t_n) - 1)] \tag{14}$$

$$f(t) = \begin{cases} 1 & if\ rand > 0.5 \\ 0 & if\ rand \le 0.5 \end{cases} \tag{15}$$

where *cumsum* calculates the cumulative sum, n is the maximum iteration number, t represents the step of the random walk (iteration in this algorithm), $f(t)$ is a stochastic function, and *rand* is a random number generated with uniform distribution in the range [0,1].

Then, we normalize the random walks by

$$X_i^t = \frac{(X_i^t - a_i) \times \left(d_i^t - c_i^t \right)}{b_i - a_i} + c_i^t \tag{16}$$

where c_i^t and d_i^t are respectively the minimum and maximum of the i-th variables at the t-th iteration and a_i and b_i are respectively the minimum and maximum in the i-th variables. Normalization is aimed at controlling the random walk within the boundaries of the

search space and avoiding the problem of the ants overshooting.

Step 3. The ants fall into the trap as a result of random walks around the ant lion, while the random walk of the ants is affected by the ant lion trap. The mathematical expression of the simulated ant random walk around the ant lion trap is

$$c_i = Antlion_j^t + c^t \tag{17}$$

$$d_i = Antlion_j^t + d^t \tag{18}$$

where c^t and d^t are respectively the minimum and maximum of all variables at the t-th iteration and c_i and d_i respectively represent the minimum and maximum of all the variables for the i-th ant. $Antlion_j^t$ represents the position of the j-th ant lion at the t-th iteration.

Step 4. The boundaries of random walks are adaptively decreased based on the true activities of the ant. The equations are:

$$c^t = \frac{c^t}{I} \tag{19}$$

$$d^t = \frac{d^t}{I} \tag{20}$$

$$\boldsymbol{I} = 1 + 10^w \frac{t}{\boldsymbol{T}} \quad w = \begin{cases} 2 & if \ \ t > 0.1T \\ 3 & if \ \ t > 0.5T \\ 4 & if \ \ t > 0.75T \\ 5 & if \ \ t > 0.9T \\ 6 & if \ \ t > 0.95T \end{cases} \tag{21}$$

where I is a ratio and w is used to adjust the exploration level. c^t and d^t are respectively the minimum and maximum of all the variables at the t-th iteration, t is the current iteration number, and T is the maximum iteration number.

Step 5. This step comprises mainly modeling the processes of catching the prey and re-building the traps of the ant lions. To simulate this, we use

$$Antlion_j^t = Ant_i^t \quad if \ f\left(Ant_i^t\right) < f\left(Antlion_j^t\right) \tag{22}$$

where $Antlion_j^t$ is the position of the j-th antlion at the t-th iteration and Ant_i^t represents the position of the i-th ant at the t-th iteration.

Step 6. The best ant lion in the optimization process is called the elite ant lion. Elite ant lions affect all the ants at each stage. The random walks of each ant lion are related to the selected ant lion and elite ant lion. After considering these two effects, Eq. (23) should be satisfied:

$$Ant_i^t = \frac{R_A^t + R_E^t}{2} \tag{23}$$

where Ant_i^t is the position of the selected j-th ant at the t-th iteration. $R^t{}_A$ and $R^t{}_E$ indicate the random walk around the selected antlion and the elite ant lion at the t-th iteration, respectively.

3.2. $L_{2,1}$-norm and random fourier mapping-based extreme learning machine

In this sub-section, we provide a brief introduction to $L_{2,1}$-norm and ELMs and then describe the construction of $L_{2,1}$RF-ELM.

3.2.1. $L_{2,1}$-norm of a matrix

Definition 1. In (Ding et al., 2006; Nie et al., 2010), the $L_{2,1}$-norm of a matrix is described as

$$\|\mathbf{M}\|_{2,1} = \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{m} m_{ij}^2} = \sum_{i=1}^{n} \|m_i\|_2 \tag{24}$$

where M is an arbitrary matrix and m_{ij} represents the elements of M. m_i represents the i-th row of M. Moreover, for any rotational matrix R, $\|\mathbf{MR}\|_{2,1} = \|\mathbf{M}\|_{2,1}$.

$L_{2,1}$-norm was first proposed for solving the problem of multi-task learning. It is a perfect sparse inducing norm (Zhou et al., 2016). This norm is defective for noisy classification learning tasks and suitable for unifying the extreme learning system.

3.2.2. Extreme learning machine

An ELM, proposed by Huang et al. (2004), is a feed-forward neural network with a single hidden layer (Huang et al., 2004). Unlike the traditional feed-forward neural network algorithm, an ELM requires only that the number of hidden layer nodes be set. Parameters, such as the weights connected to the hidden layer, are randomly generated at the beginning. As compared with a feed-forward back-propagation neural network (BPNN), ELMs have advantages such as fast learning and less parameter adjustment. The related theory is described as follows.

First, we define the notations that are used in this sub-section.

In a standard single hidden feedback network (SLFN) with n-L-m, the input set X and output set Y of Q samples are

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nq} \end{bmatrix} \tag{25}$$

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1q} \\ y_{21} & y_{22} & \cdots & y_{2q} \\ \vdots & \vdots & & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mq} \end{bmatrix} \tag{26}$$

The weight of the input layer and hidden layer is denoted by w:

$$\boldsymbol{w} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & & \vdots \\ w_{l1} & w_{l2} & \cdots & w_{ln} \end{bmatrix} \tag{27}$$

The weight of the hidden layer and output layer is defined as

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{11}\ \beta_{12} \cdots \beta_{1m} \\ \beta_{21}\ \beta_{22} \cdots \beta_{2m} \\ \vdots\ \ \vdots\ \ \ \vdots \\ \beta_{l1}\ \beta_{l2} \cdots \beta_{lm} \end{bmatrix} \tag{28}$$

The threshold of the neurons in the hidden layer is denoted by b:

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_l \end{bmatrix} \tag{29}$$

The mathematical equation of a neuron can be expressed as

$$\mathbf{Y} = g(w \cdot x + b) \tag{30}$$

where $g(x)\ \ R \rightarrow R$ is an activation function, which is non-linear.

The output of the network is defined as

$$\mathbf{T} = [t_1, t_2, \cdots, t_q]_{m\times q} \tag{31}$$

$$\boldsymbol{t}_j = \begin{bmatrix} \boldsymbol{t}_{1j} \\ \boldsymbol{t}_{2j} \\ \vdots \\ \boldsymbol{t}_{mj} \end{bmatrix}_{m\times l} = \begin{bmatrix} \sum_{i=1}^{l} \beta_{i1} g(w_i x_j + b_i) \\ \sum_{i=1}^{l} \beta_{i2} g(w_i x_j + b_i) \\ \vdots \\ \sum_{i=1}^{l} \beta_{im} g(w_i x_j + b_i) \end{bmatrix} \quad (i = 1,2,\cdots,l; j = 1,2,\cdots,q) \tag{32}$$

$$\boldsymbol{w}_i = [w_{i1}, w_{i2}, \cdots, w_{in}] \tag{33}$$

$$x_j = [x_{1j}, x_{2j}, \cdots, x_{nj}]^T \tag{34}$$

The mathematical formulation of $\mathbf{T} = [t_1, t_2, \cdots, t_q]_{m\times q}$ is $\mathbf{H}\beta = \mathbf{T}'$, where T' is a transposed matrix of T and H is the output matrix of the hidden layer.

Moreover, to understand the ELM algorithm better, we introduce two theorems proposed by Huang et al. (2006)

Theorem1. Given a standard SLFN with n input nodes, L hidden nodes, and m output nodes, and given a test group $\{x_i, t_i\}_{i=1}^{N}$, where x_i is an input set and t_i is an output set and $x_i \in R^n, t_i \in R^m$. If activation function $g: R \rightarrow R$ is infinitely differentiable in any interval, for any $w_i \in R^n$, $b_i \in R$, the output matrix of hidden layer H is invertible and $\|\mathbf{H}\beta - \mathbf{T}'\| = 0$.

Theorem2. Given an input-output sample set $\{x_i, t_i\}_{i=1}^{N}$ and any small positive number $\varepsilon > 0$, if the activation function $g: R \rightarrow R$ is infinitely differentiable in any interval, there is always a standard SLFN with n input nodes, L hidden nodes, and m output nodes. For any $w_i \in R^n b_i \in R$, there exists $L \leq N$ such that $\|\mathbf{T}' - \mathbf{H}\beta\| \leq \varepsilon$ with probability 1.

Based on Theorems 1 and 2, Huang et al. (2006) proposed an ELM to train SLFNs. From Theorem 1, we can understand that if $L = Q$, for any w and b an SLFN can approach training samples with zero error. The mathematical expression is

$$\sum_{j=1}^{q} \left\| t_j - y_j \right\| = 0 \ \left(y_j = \left[y_{1j}, y_{2j}, \cdots, y_{mj} \right]^T \right) \tag{35}$$

When $L < Q$, according to Theorem 2, the training error of an SLFN can approach any small positive number $\varepsilon > 0$, which is expressed mathematically as.

$$\sum_{j=1}^{q} \left\| t_j - y_j \right\| < \varepsilon \tag{36}$$

Therefore, when activation function $g: R \rightarrow R$ is infinitely differentiable, w and b can be randomly selected before the training progresses, and they do not change during the training. β can be obtained by solving the least squares solution of

$$\min_{\beta} \|\mathbf{H}\beta - \mathbf{T}'\| \tag{37}$$

The solution to this equation is $\beta = \mathbf{H}^{+}\mathbf{T}'$, where the Moore-Penrose generalized inverse of H is $\mathbf{H}^{+}$.

The main steps of ELM are summarized as follows.

Step 1. The connecting weight between the input layer and the hidden layer, w_i, is randomly generated and the bias between the hidden layer and the output layer b_i, where $i = 1, \cdots, L$, is calculated.

Step 2. The output matrix of hidden layer H is calculated;

Step 3. β is calculated.

3.2.3. Modling an forecating algorithm

Although an ELM has many advantages, no effective "training" method exists for it, and in general, the greater the number of hidden neurons, the smaller the training error. However, the number of hidden neurons cannot be excessively large, and therefore, the number of hidden neurons needs to be changed constantly to find the most suitable number of neurons. Therefore, Zhou et al. (2016) proposed an ELM algorithm based on random Fourier mapping and $L_{2,1}$-Norm. In this algorithm, random fourier (RF) ELM is an explicit approximation of the implicit Gaussian kernel ELM, and $L_{2,1}$RF-ELM can prune the redundant and irrelevant hidden neurons automatically, in order to form a discriminative and compact hidden layer (Xue et al., 2017).

The objective function of this algorithm is

$$\min_{\beta,\xi} \frac{1}{2}\|\beta\|_{2,1} + \frac{1}{2} C \sum_{i=1}^{N} \|\xi_i\|^2 \tag{38}$$

$$s.t.\ \ h(x_i)\beta = y_i^T - \xi_i^T \quad i = 1, \cdots, N \tag{39}$$

Here, β is the output weight matrix, ξ_i represents the training error vector, and C represents the penalty coefficient. For the detailed solution process of this function, please see (Zhou et al., 2016).

According to (Xue et al., 2017), we chose a two-step strategy to train the $L_{2,1}$RF-ELM.

Step 1. Neuron selection. The different K hidden nodes are selected based on the norm of the output weights, where the norm proportion is less than or equal to a given threshold value. The threshold is 96% for classification and 96.5% for regression. Moreover, K is determined by cross validation of different data sets. The best combination of weight variance and penalty coefficient is searched within a given range. The search range of the weight variance is $\{2^{-24}, 2^{-22}, \cdots, 2^{8}\}$ and the search range of the penalty coefficient is $\{2^{-20}, 2^{-19}, \cdots, 2^{20}\}$.

Step 2. Output weight calculation. After the hidden nodes are selected, the output weights can be obtained and an ELM with a

finely pruned hidden layer can be trained. The parameter search range is $\{2^{-20}, 2^{-19}, \cdots, 2^{20}\}$

3.3. Optimization of $L_{2,1}$RF-ELM algorithm

In order to improve the forecasting accuracy of $L_{2,1}$RF-ELM, we selects the multi-object ant lion algorithm to optimize the parameters of the forecasting algorithm. The optimized parameters include the number of input nodes and the number of hidden layer nodes of the ELM. And there are two optimization objects, one is mean squared error, which measures the forecasting accuracy of the model, and the other is the standard deviation of the forecasting error that measures the stability of the prediction model.

The optimization process of MOALO-$L_{2,1}$RF-ELM usually contains five steps:

Step 1. Input initial parameters of MOALO and $L_{2,1}$RF-ELM, such as iterations number, ant lion population size, number of ants and each ant lion and ant initial position, etc..

Step 2. Determining the objective function of the optimization algorithm, the objective functions of this study show as

$$\boldsymbol{min} = \begin{cases} O1 = MSE = \dfrac{1}{n}\sum_{i=1}^{n}(\widehat{y}_i - y_i)^2 \\ O2 = std(\widehat{y}_i - y_i) \end{cases} \quad i = 1, 2, \ldots, n \tag{40}$$

Since the stability and accuracy of the forecasting are equally important, we set the weights of the two objective functions to 0.5 respectively to ensure the accuracy and stability of the forecasting can be met simultaneously.

Step 3. Continuously update the position information of ant lions and ants according to the value of the objective function.

Step 4. Search for the optimal solution set and update the archive on each iteration. If reached the maximum number of iterations or expected errors, the next step is launched; otherwise, back to step 4 to continue the iteration.

Step 5: Stop the iteration and get the optimal parameters of the $L_{2,1}$RF-ELM algorithm. Finally, input the data after preprocessing to the optimized $L_{2,1}$RF-ELM to obtain the forecasting value.

4. Air quality assessment and application of air pollution early warning system

The most used air quality evaluation tool is Air Quality Index (AQI), but “pollution degree” is a relatively vague concept, and hard to find clear boundaries. Therefore, fuzzy logic is more suitable for air quality assessment.

FSE is an evaluation method based on fuzzy mathematics. One of its advantages is that its mathematical model is simple. Moreover, it is effective for multi-factor and multi-level complex problem assessment. In this section, first we introduce the theory, and then, according to fuzzy evaluation theory the application to the study areas is described in the second sub-section. Following the sub-section 4.2, we introduced five model performance metrics in sub-section 4.3. And sub-section 4.4 descriptions the data we used in study. The last two sub-sections introduce the experimental process and experimental results.

4.1. Fussy synthetic evaluation theory

The process of establishing an FSE system is as follows.

Step 1. The set of factors for the evaluation object is determined.

The selected factors should possess the traits of representativeness, feasibility, and system. Air quality evaluation relies on the concentration levels of the main air pollutants. Therefore, in this study, the indicators were chosen according to China's ambient air quality standards (AAQS: GB3095-2012). Moreover, different geographical areas have different topographic and economic characteristics, and consequently, the different key pollutants in the study areas should be also taken into account. The factors set can be given as

$$\mathbf{U} = \{u_1, u_2, \cdots, u_m\} \tag{41}$$

Step 2. The evaluation rank standard is determined.

The evaluation rank set is described as $\mathbf{V} = \{v_1, v_2, \cdots v_n\}$. In our study, the air pollution degrees were divided into five levels, and therefore, the rank set is given as

$$\mathbf{V} = \{Excellent, Good, Moderate, Heavy, Serious\} \tag{42}$$

The pollutants grading standard according to AAQS is shown in Table 1.

Step 3. Index fuzzification.In this step, the membership functions (MFs) corresponding to each index are obtained. The process of fuzzification constitutes the process of membership calculation by using MFs. In this study, we used the trapezoidal membership to calculate the membership value.

$$r_{ij} = \begin{cases} 0 & x_i \geq u_2 \\ \dfrac{u_2 - x_i}{u_2 - u_1} & u_1 < x_i < u_2 \quad (j = 1) \\ 1 & x_i \leq u_1 \end{cases} \tag{43}$$

$$r_{ij} = \begin{cases} 0 & x_i \geq u_3 \ \ or \ \ x_i \leq u_1 \\ \dfrac{u_3 - x_i}{u_3 - u_2} & u_2 < x_i < x_3 \\ \dfrac{x_i - u_1}{u_2 - u_1} & u_1 \leq x_i \leq u_2 \end{cases} \quad (j = 2) \tag{44}$$

Table 1
Threshold of pollutant concentration (μg/m^3;mg/m^3) corresponding to air quality of different levels.

Factors	Excellent	Good	Moderate	Heavy	Serious
	I	II	III	IV	V
$PM_{2.5}$	15	35	75	120	165
PM_{10}	40	70	140	210	320
SO_2	20	60	100	140	180
NO_2	20	40	60	80	100
CO	2	4	6	8	10
O_3	100	160	220	280	340

$$r_{ij} = \begin{cases} 0 & x_i \geq u_3 \ \ or \ \ x_i \leq u_2 \\ \dfrac{u_4 - x_i}{u_4 - u_3} & u_3 < x_i < x_4 \\ \dfrac{x_i - u_3}{u_3 - u_2} & u_2 \leq x_i \leq u_3 \end{cases} \quad (j = 3) \tag{45}$$

$$r_{ij} = \begin{cases} 0 & x_i \geq u_4 \ \ or \ \ x_i \leq u_3 \\ \dfrac{u_5 - x_i}{u_5 - u_4} & u_4 < x_i < x_5 \\ \dfrac{x_i - u_4}{u_4 - u_3} & u_3 \leq x_i \leq u_4 \end{cases} \quad (j = 4) \tag{46}$$

$$r_{ij} = \begin{cases} 0 & x_i \leq u_4 \\ \dfrac{x_i - u_4}{u_5 - u_4} & u_4 < x_i < x_5 \\ 1 & x_i \geq u_5 \end{cases} \quad (j = 5) \tag{47}$$

where x_i is the actual value of index i, j is the air quality rank, and u represents the threshold value of the air quality rank for different pollutants. According to the MF, we can obtain the single factor fuzzy relationship matrix R, which reflects the degree of affiliation of each pollution factor to each level of air quality.

$$\mathbf{R} = (r_{ij})_{m\times n} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} \tag{48}$$

r_{ij} in this matrix indicates the membership degree of the factor u_i corresponding to v_j.

Step 4. The factor weight is calculated.

Weight reflects the importance of each factor in synthetic evaluation and directly affects the outcome of the evaluation. Many methods exist for determining the weight, such as weighted statistics, Coefficient of variation method, the Delphi method, and entropy methods. In our study, the weight was calculated by Fuzzy Weighting Method, the formula is shown as (Xu et al., 2017a)

$$w_i = \frac{na_i}{\sum_{j=1}^{n} r_{ij}} \tag{49}$$

where a_i indicates the average of the observed concentration of the i-th factor and n represents the number of grade levels.

Step 5. The evaluation results are output.

The result B is computed as

$$\mathbf{B} = \{b_1, b_2, \cdots b_6\} = \mathbf{W} \times \mathbf{R} \tag{50}$$

4.2. Application of fuzzy synthetic evaluation

The weight of each pollutant in different cities can be obtained according to the theory of FSE. Moreover, based on the weight, the three most important pollutants in different cities are obtained for subsequent analysis, and a macro is obtained according to the average concentration of pollutants in different cities.

Table 2
Fuzzy weights and fuzzy synthetic evaluation results of pollutants in the study areas.

Area	Indices	Pollutants(μg/m³;mg/m³)						
		$PM_{2.5}$	PM_{10}	SO_2	NO_2	CO	O_3	Level
Beijing	AATVS	65.97	105.36	10.46	45.65	1.04	67.59	III
	W	0.28	0.24	0.04	0.06	0.27	0.11	
	B	0.21	0.38	0.42	–	–	–	
Shanghai	AATVS	42.22	60.20	12.12	44.10	0.75	80.58	II
	W	0.23	0.17	0.05	0.06	0.33	0.16	
	B	0.33	0.56	0.11	–	–	–	
Guangzhou	AATVS	36.49	58.53	11.07	52.98	0.89	46.38	II
	W	0.20	0.17	0.05	0.07	0.41	0.10	
	B	0.28	0.45	0.27	–	–	–	

The results of applying the fuzzy weight of pollutants are shown in Table 2, where the average actual values (AATVs) for the three cities included in the study are also presented.

From the results of the FSE listed in Table 2, some conclusions can be drawn:

1) The air quality of Beijing belongs to Level III with a 42% probability, to Level II with a 38% probability, and to Level I with a 21% probability. Therefore, according to the maximum membership degree principle, the air quality in Beijing belongs to Level III.
2) The air quality of Shanghai and Guangzhou belongs to Level II. Although both cities are metropolises, the air quality varies with geography, climate, population, and economic development.
3) To compare the weights of pollutants in the study areas, three main pollutants of cities were selected. For Beijing, the three main air pollutants were $PM_{2.5}$, CO, and PM_{10}. In Shanghai, CO, $PM_{2.5}$, and PM_{10} are the top three pollutants and in Guangzhou, the three most influential pollutants are also CO, $PM_{2.5}$, and PM_{10}.
4) An analysis of the weight of the same pollutant in the different cities shows that CO, $PM_{2.5}$, and PM_{10} are the most important pollutants in each city. The difference is that the weight of the three pollutants in Beijing is similar, while a heavier weight is assigned to CO in Shanghai and Guangzhou.

From the analysis above, it is very clear that CO, $PM_{2.5}$, and PM_{10} are the first three primary pollutants. Therefore, there is an urgent need to establish an early warning system and provide accurate information about the concentrations of CO, $PM_{2.5}$, and PM_{10} in the next hour.

4.3. Model performance metrics

According to the above introduction, in the following subsections we describe three experiments designed to evaluate the performance of the proposed early warning system in three metropolises in China. The experiments were conducted in Beijing, Shanghai, and Guangzhou, in that order temporally. A comparison of the forecasting models and pollution evaluation was included in each experiment.

To evaluate the forecasting performance of our method, five indexes were selected: the mean average percentage error (MAPE), the forecasting mean absolute error (MAE), the forecasting root mean square error (RMSE), the fractional bias of the forecasting results (FB), and Pearson's correlation coefficient (R). The details of these error criteria are shown in Table 3.

Table 3
Description and equations of the model performance metric rules.

Metric	Description	Equation
MAPE	Evaluates the forecasting accuracy	$\mathbf{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{A_i - F_i}{A_i}\right\| \times 100\%$
MAE	Measures the difference between the forecast value and the actual value	$\mathbf{MAE} = \frac{1}{N}\sum_{i=1}^{N}\|F_i - A_i\|$
RMSE	Measures the difference between the forecast value and the actual value; it is more sensitive to extreme values than MAE	$\mathbf{RMSE} = \sqrt{\frac{1}{N}\times\sum_{i=1}^{N}(F_i - A_i)^2}$
FB	Determines the amount of "over" and "under" forecasts	$\mathbf{FB} = 2(\overline{A} - \overline{F})/(\overline{A} + \overline{F})$
R	Measures the correlation of the actual value and the forecast value	$R = \frac{N\sum F_i A_i - \sum F \sum A}{\sqrt{N\sum F_i^2 - (\sum F_i)^2}\sqrt{N\sum A_i^2 - (\sum A_i)^2}}$

4.4. Data description

In this study, three metropolises in China were selected for developing and testing the proposed system. These three cities in the similar scale, but different regions mean different climatic conditions, topography and geomorphology, which lead to differences in urban layout and economic development characteristics. The combination of these factors will lead to different pollution characteristics. In addition, the data quality of these three cities is high. Therefore, selecting these three cities can test the feasibility and generalization ability of the model from the data perspective.

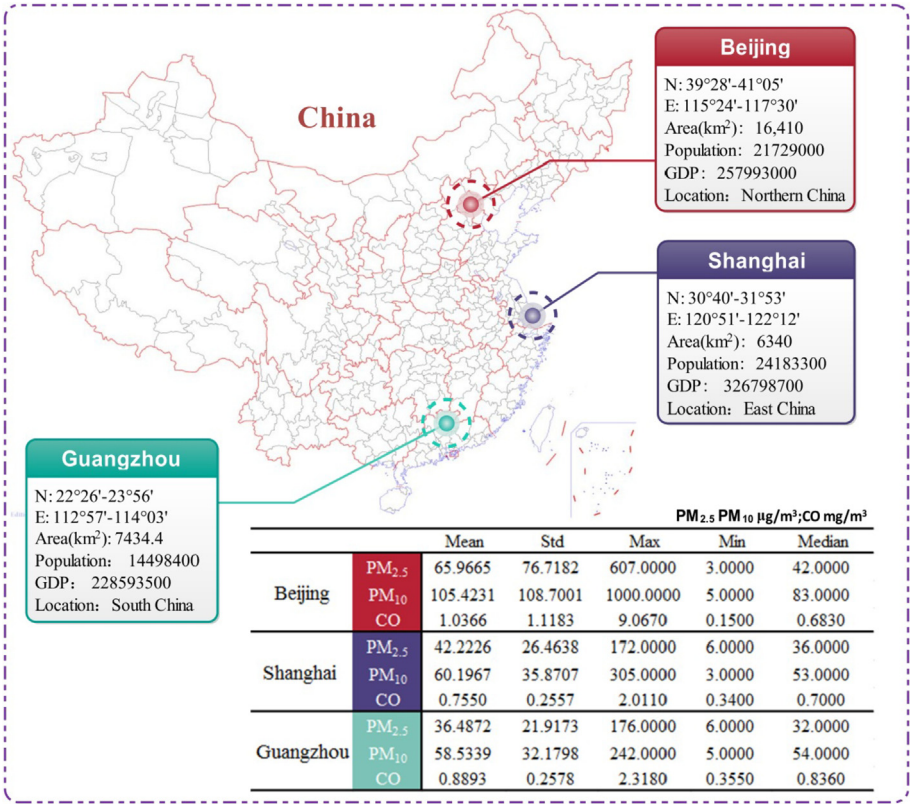
The related information of the selected study sites is shown in Fig. 1. The datasets of hourly concentrations of air pollutants were taken from the Chinese Website of air quality real-time publishing platform (http://113.108.142.147:20035/emcpublish/). The pollutants of the three cities that are analyzed are $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, and O_3. The hourly concentrations data were collected from 1 January 2017 to 1 July 2017. The descriptive statistics (maximum, minimum, mean, and standard deviation) of these pollutants are shown in Fig. 1. These datasets were divided into two sub-datasets: a training dataset, consisting of 3480 data points, and a testing dataset, consisting of 888 points.

4.5. Experiment I: the case of Beijing

This section is divided into two sub-sections. In the first, the forecasting abilities of the proposed model and the benchmark models are compared, and in the second, the FSE of the forecasting results is presented.

4.5.1. Forecasting models comparison

In order to verify the forecasting ability of the proposed hybrid model (SSA-EEMD-MOALO-$L_{2,1}$RFELM, SEMR), the following models were chosen as the comparative models in this experiment:



$PM_{2.5}$ PM_{10} μg/m³;CO mg/m³

		Mean	Std	Max	Min	Median
Beijing	$PM_{2.5}$	65.9665	76.7182	607.0000	3.0000	42.0000
	PM_{10}	105.4231	108.7001	1000.0000	5.0000	83.0000
	CO	1.0366	1.1183	9.0670	0.1500	0.6830
Shanghai	$PM_{2.5}$	42.2226	26.4638	172.0000	6.0000	36.0000
	PM_{10}	60.1967	35.8707	305.0000	3.0000	53.0000
	CO	0.7550	0.2557	2.0110	0.3400	0.7000
Guangzhou	$PM_{2.5}$	36.4872	21.9173	176.0000	6.0000	32.0000
	PM_{10}	58.5339	32.1798	242.0000	5.0000	54.0000
	CO	0.8893	0.2578	2.3180	0.3550	0.8360

Fig. 1. Information of study areas and the statistics of pollution data.

SSA-EEMD-MOALO-ELM (SEME), EEMD-MOALO-$L_{2,1}$RFELM (EMR), MOALO-$L_{2,1}$RFELM (MR), and $L_{2,1}$RFELM.

For each forecasting model, the training set consisted of data from 1 January 2017 to 25 May 2017. These data were tested on the data from 26 May 2017 to 1 July 2017. The performance of each model was evaluated using the selected six performance metrics. The results are presented in Part D of Fig. 2. Fig. 2 also shows in graphics the forecast results for Beijing, China. Parts A, B, and C of Fig. 2 illustrate the forecasting error of the proposed forecasting model for the three pollutants. The error line of SEMR in Fig. 2 is closest to zero with the lowest volatility.

In Part D of Fig. 2, the figures in bold are the models that perform well according to the various evaluation criteria. The following conclusions can be drawn from Fig. 2:

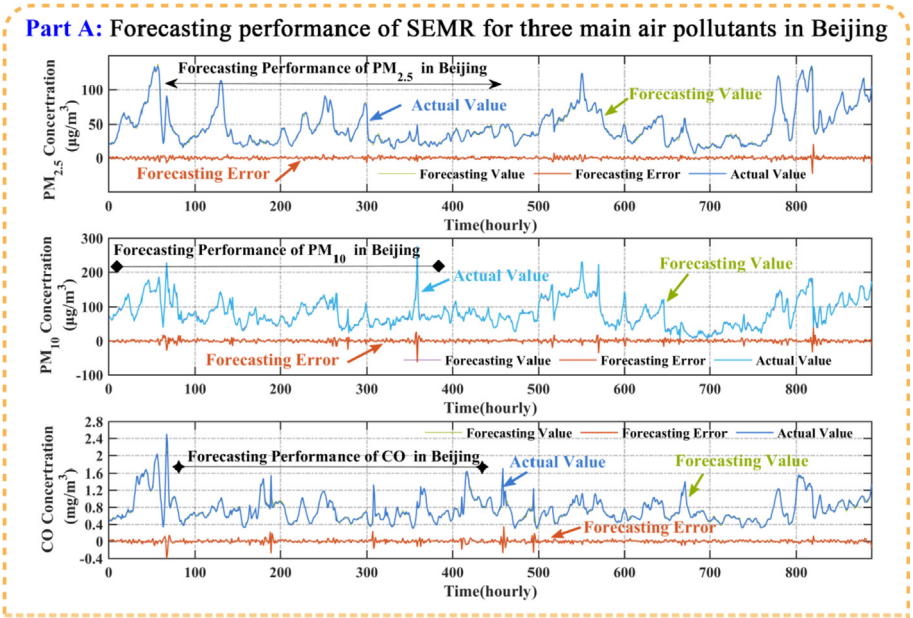
(1) The optimized forecasting model outperforms the non-optimized model. From a comparison of the forecasting results of $L_{2,1}$RFELM model and the MR model for different pollutants, it is easy to conclude that the multi-objective optimized model has a better forecasting capability. The MAPE value of MR is approximately 2.3% lower than that of $L_{2,1}$RFELM on average, and the other metrics (MAE, RMSE, FB, and R) also show that the performance of the MR forecasting model is better than that of the $L_{2,1}$RFELM forecasting model.

(2) The forecasting precision can be improved after data preprocessing. Let us take CO as an example to compare the MAPE values in Part D of Fig. 2. The MAPE value of the MR forecasting model is 8.5376% and of the model after a single denoising process is 4.3217%, whereas that of the proposed hybrid model (with two denoising processes) it is 4.1337%.

(3) The models where two denoising methods are applied perform better. From the experimental results of SEMR, EMR, and SEME, it can be seen that the MAPE values of different pollutants for the models that include two denoising methods (SEMR and SEME) are almost smaller than that which includes a single denoising method (EMR). In particular, the MAPE value for SEMR is smaller than that for SEME. According to this analysis, the forecasting ability of the forecasting model proposed in this paper is good.

(4) The forecasting model proposed in this paper is significantly better than the traditional statistical model ARIMA for forecasting air pollutant concentration, and the forecasting performance of the $L_{2,1}$RF-ELM model used in this study is better than that of the traditional neural network, BPNN. According to the data in Part D of Fig. 2, the metric values of the forecasting model proposed in this paper are significantly lower than those of the ARIMA model. In addition, from the comparison of $L_{2,1}$RF-ELM and BPNN it can be seen that the BPNN metrics are higher than the $L_{2,1}$RF-ELM index values. It can thus be considered that BPNN's forecasting ability is inferior to that of the $L_{2,1}$RF-ELM model.

***Remark**: We compared the MAPE, MAE, RMSE, FB, and R values in this study. The proposed hybrid model, SEMR, performs better according to all metrics. It is proved that the proposed model is superior to other models in terms of air pollution forecasting. In addition, the proposed hybrid forecasting model fits well the data of pollutants with high volatility and inhomogeneity. Therefore, the model can be used as part of an air quality warning system (Xu et al., 2017b).*



Part B: Indicators analysis of different models in Beijing

	$PM_{2.5}$					PM_{10}					CO				
	MAPE (%)	MAE	RMSE	FB	R	MAPE (%)	MAE	RMSE	FB	R	MAPE (%)	MAE	RMSE	FB	R
SEMR	4.2708	1.5716	2.2401	-0.0013	0.9963	5.7845	3.5938	5.7444	0.0004	0.9901	3.9248	0.0281	0.0459	0.0001	0.9887
ARIMA	8.5973	3.3780	5.5317	0.0026	0.9777	15.0867	9.6475	16.666	0.0021	0.9143	8.7982	0.0659	0.1216	0.0017	0.9196
SEME	4.5730	1.6245	2.2763	-0.0026	0.9963	5.9497	3.7793	6.0234	-0.0009	0.9890	4.1337	0.0295	0.0470	-0.0018	0.9880
EMR	4.6190	1.6514	2.1932	-0.0024	0.9965	6.5861	3.9841	6.6309	-0.0018	0.9863	4.3217	0.0303	0.0477	-0.0037	0.9873
MR	8.2804	3.2394	5.2144	-0.0059	0.9803	14.9794	9.3296	15.7233	-0.0093	0.9224	8.5376	0.0637	0.1164	-0.0109	0.9236
$L_{2,1}$RF-ELM	8.4735	3.3391	5.5529	-0.0060	0.9777	15.1561	9.4870	16.3943	-0.0071	0.9165	8.4616	0.0643	0.1204	-0.0057	0.0197
BPNN	13.9274	5.3842	13.488	0.0104	0.9813	15.3855	16.6669	40.9539	0.0084	0.9660	12.9307	0.0642	0.1411	-0.0350	0.8872

Fig. 2. Forecasting error and forecasting result of the proposed model for three main pollutants in Beijing.

Table 4
Air quality evaluation results for Beijing.

Data Point	Evaluation results based on the forecasting data						Evaluation results based on the actual data					
	Fuzzy evaluation result B					Level	Level	Fuzzy evaluation result B				
26/05/2017	0.0808	0.6866	0.2326	0.0000	0.0000	II	II	0.2315	0.5251	0.2433	0.0000	0.0000
27/05/2017	0.1212	0.6268	0.2520	0.0000	0.0000	II	II	0.1390	0.4704	0.3905	0.0000	0.0000
28/05/2017	0.1032	0.0123	0.7523	0.1322	0.0000	III	III	0.1457	0.3060	0.4683	0.0799	0.0000
29/05/2017	0.2126	0.6382	0.1491	0.0000	0.0000	**II**	I	0.5346	0.3873	0.0781	0.0000	0.0000
30/05/2017	0.1062	0.8395	0.0543	0.0000	0.0000	II	II	0.3158	0.6563	0.0279	0.0000	0.0000
31/05/2017	0.0785	0.3355	0.5860	0.0000	0.0000	II	II	0.1902	0.4327	0.3771	0.0000	0.0000
01/06/2017	0.5731	0.4269	0.0000	0.0000	0.0000	I	I	0.6538	0.3462	0.0000	0.0000	0.0000
02/06/2017	0.4653	0.5347	0.0000	0.0000	0.0000	II	II	0.4673	0.5327	0.0000	0.0000	0.0000
03/06/2017	0.5226	0.4774	0.0000	0.0000	0.0000	I	I	0.5243	0.4757	0.0000	0.0000	0.0000
04/06/2017	0.0851	0.6630	0.2518	0.0000	0.0000	II	II	0.2330	0.5437	0.2233	0.0000	0.0000
05/06/2017	0.0911	0.2973	0.6116	0.0000	0.0000	III	III	0.1998	0.2082	0.5920	0.0000	0.0000
06/06/2017	0.6126	0.3755	0.0118	0.0000	0.0000	I	I	0.6363	0.3612	0.0025	0.0000	0.0000
07/06/2017	0.1961	0.6646	0.1393	0.0000	0.0000	II	II	0.4499	0.4748	0.0754	0.0000	0.0000
08/06/2017	0.5890	0.4110	0.0000	0.0000	0.0000	I	I	0.4008	0.3490	0.2502	0.0000	0.0000
09/06/2017	0.2141	0.6465	0.1394	0.0000	0.0000	II	II	0.2923	0.4559	0.2518	0.0000	0.0000
10/06/2017	0.3661	0.6257	0.0082	0.0000	0.0000	I	I	0.5605	0.4392	0.0003	0.0000	0.0000
11/06/2017	0.1650	0.7402	0.0947	0.0000	0.0000	II	II	0.2834	0.5677	0.1489	0.0000	0.0000
12/06/2017	0.1846	0.7584	0.0570	0.0000	0.0000	II	II	0.3229	0.5941	0.0830	0.0000	0.0000
13/06/2017	0.2301	0.7077	0.0623	0.0000	0.0000	II	II	0.3316	0.6345	0.0339	0.0000	0.0000
14/06/2017	0.2549	0.7253	0.0198	0.0000	0.0000	II	II	0.3184	0.5840	0.0976	0.0000	0.0000
15/06/2017	0.0903	0.7825	0.1272	0.0000	0.0000	II	II	0.1779	0.6277	0.1944	0.0000	0.0000
16/06/2017	0.0642	0.3007	0.6352	0.0000	0.0000	II	II	0.0825	0.5791	0.3383	0.0000	0.0000
17/06/2017	0.0676	0.0000	0.8024	0.1300	0.0000	III	III	0.0991	0.3069	0.5141	0.0799	0.0000
18/06/2017	0.0753	0.1089	0.8157	0.0000	0.0000	III	III	0.1504	0.3598	0.4899	0.0000	0.0000
20/06/2017	0.2928	0.7072	0.0000	0.0000	0.0000	II	II	0.3103	0.6897	0.0000	0.0000	0.0000
21/06/2017	0.1111	0.6368	0.2521	0.0000	0.0000	II	II	0.2349	0.6372	0.1279	0.0000	0.0000
22/06/2017	0.5668	0.4332	0.0000	0.0000	0.0000	I	I	0.7540	0.2460	0.0000	0.0000	0.0000
23/06/2017	0.9291	0.0709	0.0000	0.0000	0.0000	I	I	0.9689	0.0311	0.0000	0.0000	0.0000
24/06/2017	0.8826	0.1174	0.0000	0.0000	0.0000	I	I	0.9533	0.0467	0.0000	0.0000	0.0000
25/06/2017	0.6677	0.3323	0.0000	0.0000	0.0000	I	I	0.6975	0.3025	0.0000	0.0000	0.0000
26/06/2017	0.7533	0.2467	0.0000	0.0000	0.0000	I	I	0.6433	0.3567	0.0000	0.0000	0.0000
27/06/2017	0.0799	0.6385	0.2816	0.0000	0.0000	II	II	0.1055	0.5436	0.3509	0.0000	0.0000
28/06/2017	0.0909	0.0980	0.6074	0.2037	0.0000	**III**	II	0.1288	0.3797	0.3676	0.1238	0.0000
29/06/2017	0.0976	0.5158	0.3866	0.0000	0.0000	II	II	0.2046	0.5921	0.2032	0.0000	0.0000
30/06/2017	0.0809	0.7064	0.2126	0.0000	0.0000	II	II	0.0857	0.5183	0.3960	0.0000	0.0000
01/07/2017	0.0766	0.0504	0.7142	0.1589	0.0000	III	III	0.0746	0.3945	0.4343	0.0966	0.0000

4.5.2. Fuzzy synthetic evaluation of air quality in Beijing

In our experiment, we selected three major air pollutants, $PM_{2.5}$, PM_{10}, and CO. In order to verify the effectiveness of the FSE system, the daily average forecast values of these three major pollutants were used for FSE of air pollution, and then compared with the actual pollution level. The daily average forecast data were calculated from the hourly average, and the actual pollution level was obtained by using FSE based on the actual concentration of six pollutants. The comparison results are shown in Table 4.

The left hand side of Table 4 shows the results of fuzzy

Table 5
Forecasting results of the proposed model and other models in Shanghai.

		MAPE(%)	MAE	RMSE	FB	R
$PM_{2.5}$	SEMR	3.6465	1.0598	1.4796	0.0007	0.9984
	ARIMA	10.5252	3.0923	4.4770	−0.0007	0.9848
	SEME	3.7818	1.1066	1.5661	−0.0005	0.9982
	EMR	3.7076	1.0804	1.6058	−0.0008	0.9981
	MR	9.6955	2.9843	4.4070	−0.0045	0.9853
	$L_{2,1}$RF-ELM	9.8328	2.9938	4.4955	−0.0039	0.9846
	BPNN	7.9557	2.3406	3.3361	−0.0106	0.9812
PM_{10}	SEMR	4.5328	1.8670	3.0369	−0.0019	0.9946
	ARIMA	12.7248	5.1509	8.3233	−0.0097	0.9589
	SEME	4.6276	1.8701	3.0780	−0.0026	0.9946
	EMR	4.7576	1.9299	3.0601	−0.0037	0.9946
	MR	12.2929	4.9899	8.0953	−0.0122	0.9609
	$L_{2,1}$RF-ELM	12.3966	5.0582	8.2988	−0.0107	0.9585
	BPNN	8.7630	5.3363	9.9142	0.0051	0.9796
CO	SEMR	1.6953	0.0107	0.0156	−0.0006	0.9973
	ARIMA	4.5847	0.0299	0.0436	0.0008	0.9778
	SEME	1.7259	0.0109	0.0156	−0.0015	0.9973
	EMR	1.7906	0.0113	0.0156	−0.0026	0.9972
	MR	4.6822	0.0300	0.0439	−0.0081	0.9779
	$L_{2,1}$RF-ELM	4.7413	0.0303	0.0440	−0.0068	0.9773
	BPNN	4.2084	0.0293	0.0427	−0.0086	0.9685

Table 6
Forecasting results of the proposed model and other models in Guangzhou.

		MAPE(%)	MAE	RMSE	FB	R
$PM_{2.5}$	SEMR	2.6965	0.5028	0.6698	−0.0023	0.9976
	ARIMA	5.4247	1.0810	1.5657	−0.0028	0.9867
	SEME	2.7312	0.5112	0.6822	−0.0045	0.9975
	EMR	2.7503	0.5191	0.6755	−0.0055	0.9976
	MR	5.5808	1.1004	1.6095	−0.0161	0.9867
	$L_{2,1}$RF-ELM	5.6780	1.1164	1.6233	−0.0150	0.9864
	BPNN	9.2161	2.3874	3.7394	−0.0325	0.9800
PM_{10}	SEMR	3.3672	1.1505	1.5153	−0.0034	0.9955
	ARIMA	9.2703	3.2478	4.4011	−0.0013	0.9605
	SEME	3.3815	1.1564	1.5206	−0.0039	0.9955
	EMR	3.5642	1.2107	1.5818	−0.0047	0.9950
	MR	9.2728	3.2244	4.3838	−0.0177	0.9617
	$L_{2,1}$RF-ELM	9.4138	3.2802	4.4535	−0.0179	0.9608
	BPNN	9.1559	4.2441	6.2811	−0.0043	0.9738
CO	SEMR	1.3851	0.0099	0.0127	−0.0020	0.9959
	ARIMA	3.6532	0.0264	0.0354	−0.0012	0.9666
	SEME	1.4564	0.0104	0.0135	−0.0022	0.9953
	EMR	1.4283	0.0101	0.0131	−0.0039	0.9956
	MR	3.9516	0.0284	0.0372	−0.0108	0.9647
	$L_{2,1}$RF-ELM	4.0622	0.0291	0.0382	−0.0116	0.9629
	BPNN	4.0867	0.0336	0.0454	−0.0090	0.9691

comprehensive evaluation based on the forecast values of the three major pollutants and the right hand side shows the results of the fuzzy neutralization evaluation based on the real values of six pollutants. The corresponding air quality level is also shown. The bold words (2/37) represent misevaluations; that is, the evaluation results based on the forecast value are different from those based on the real value.

The results of observation revealed that the air quality that was misevaluated was in fact one level lower than the air quality according to the evaluation. For example, the actual air quality of 29/05/2017 was Level I, while according to the evaluation based on the forecast value it was Level II. This can prevent human health from harm caused by the fact that the actual air quality is lower than the forecast air quality.

Remark: *The concentrations of pollutants change rapidly with the weather and human activity. There is no means of exactly assessing the contamination. The ultimate purpose of the so-called pollution evaluation system is to provide advice on avoiding major pollution incidents and on guiding human industrial production and life activities. If the actual air quality is worse than the forecasting air quality, it may lead to incorrect opinions being put forward and lead to the occurrence of serious pollution accidents.*

4.6. Experiment II: the case of Shanghai and Guangzhou

A good forecasting model should have a high generalization ability, and therefore, the objective of this experiment was to verify the model's generalizability for air pollution concentration forecasting and air quality evaluation in different cities. This section is divided into two sub-sections: in the first, the forecast results are compared and in the second an FSE is presented.

4.6.1. Forecasting models comparison in Shanghai and Guangzhou

The experimental results are presented in Tables 5 and 6, where the values in bold are the best values of each evaluation metric of the model. A comparison of the values of the evaluation criteria

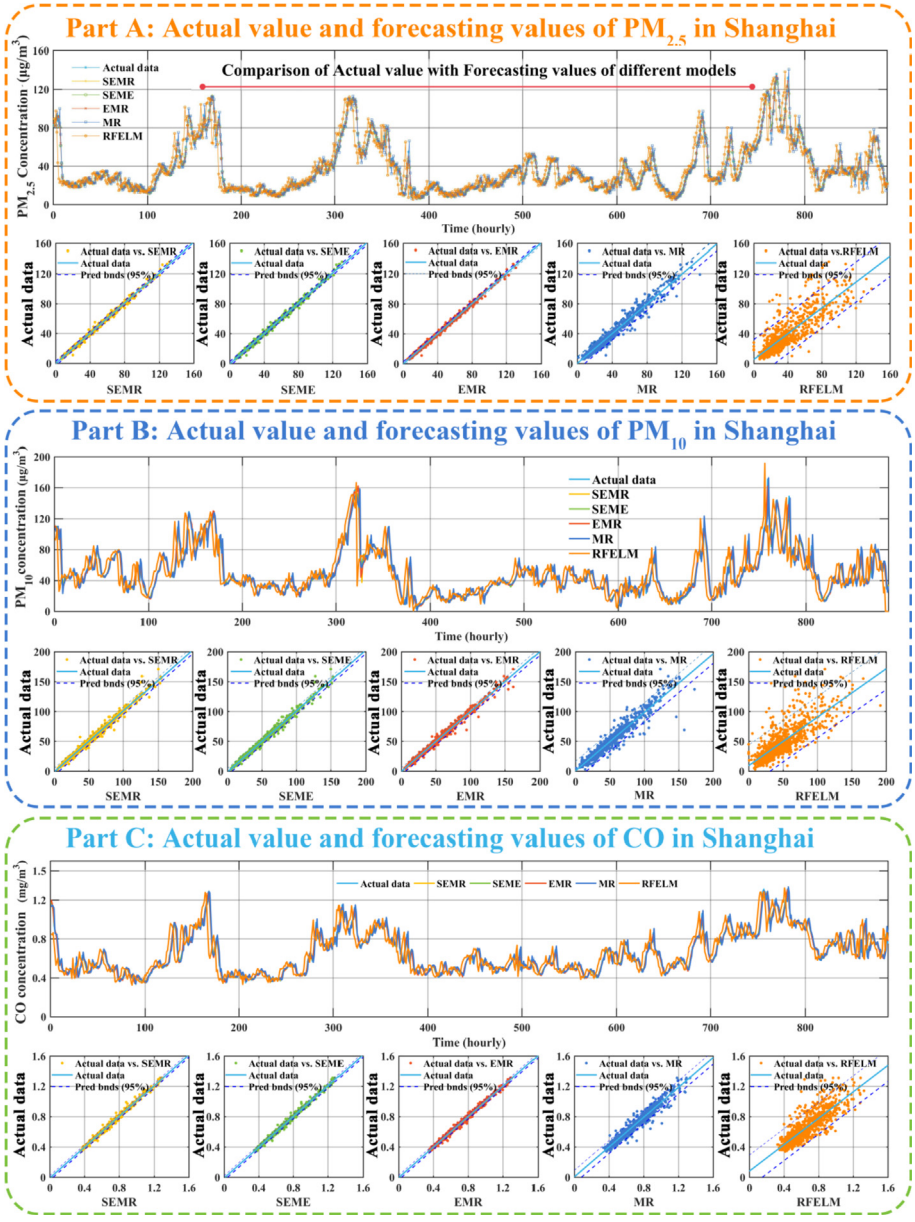


Fig. 3. Forecasting results and forecasting errors for three pollutants in Shanghai.

shows that the performance of the proposed model is almost better than that of the other models; in particular in Shanghai, the metric values of the proposed model are the best.

From the data in Tables 5 and 6, we can reach the same conclusion as that drawn from the results of Experiment I. The hybrid model proposed in this paper is superior to the traditional statistical model ARIMA, and the $L_{2,1}$RF-ELM forecasting algorithm used in this study has a better forecasting ability than the BPNN.

So that the comparison results can be understood more intuitively, they are presented in Figs. 3 and 4. It can clearly be observed that the proposed model performs well and is superior to the individual model and the single denoising model.

Parts A, B, and C of Fig. 3 describe the forecasting performance of the different models for $PM_{2.5}$, PM_{10}, and CO, respectively, in Shanghai, China. Each part consists of two types of figure: the upper figure shows the forecast values of the different models and the actual data and the lower figure shows the linear interpolant fitting between the actual data and different forecast values. The more concentrated the points, the better is the forecasting performance. From this figure, it can be concluded that the proposed forecasting model performs well, because the forecasting values are close to the true values and the errors are small.

The forecasting performance for Guangzhou is shown in Fig. 4. The figure is divided into three parts, where the first part shows the forecast values and the actual value of $PM_{2.5}$ in Shanghai. This part also shows a comparison of the different model performance values. Parts 2 and 3 show the forecasting performance for PM_{10} and CO, respectively. Their composition is similar to that of Part 1, showing the actual data and the forecast values, as well as the forecasting performance.

The results of Experiment II also account for the conclusions that were drawn from the results of Experiment I. In other words, the hybrid model proposed in this paper can in general be used in the forecasting of air pollution. In Part D of Fig. 2, Table 5, and Table 6, it can be seen that the MAPE values of the proposed model for $PM_{2.5}$ in Beijing, Shanghai, and Guangzhou are 4.2708%, 3.6465%, and 2.6965%, respectively. However, before preprocessing by SSA and EEMD, the MAPE values for the model increase by 0.9388%, 1.6589%, and 1.0696%, respectively, as compared with the proposed model. In comparison, SEMR is an effective hybrid model and superior to

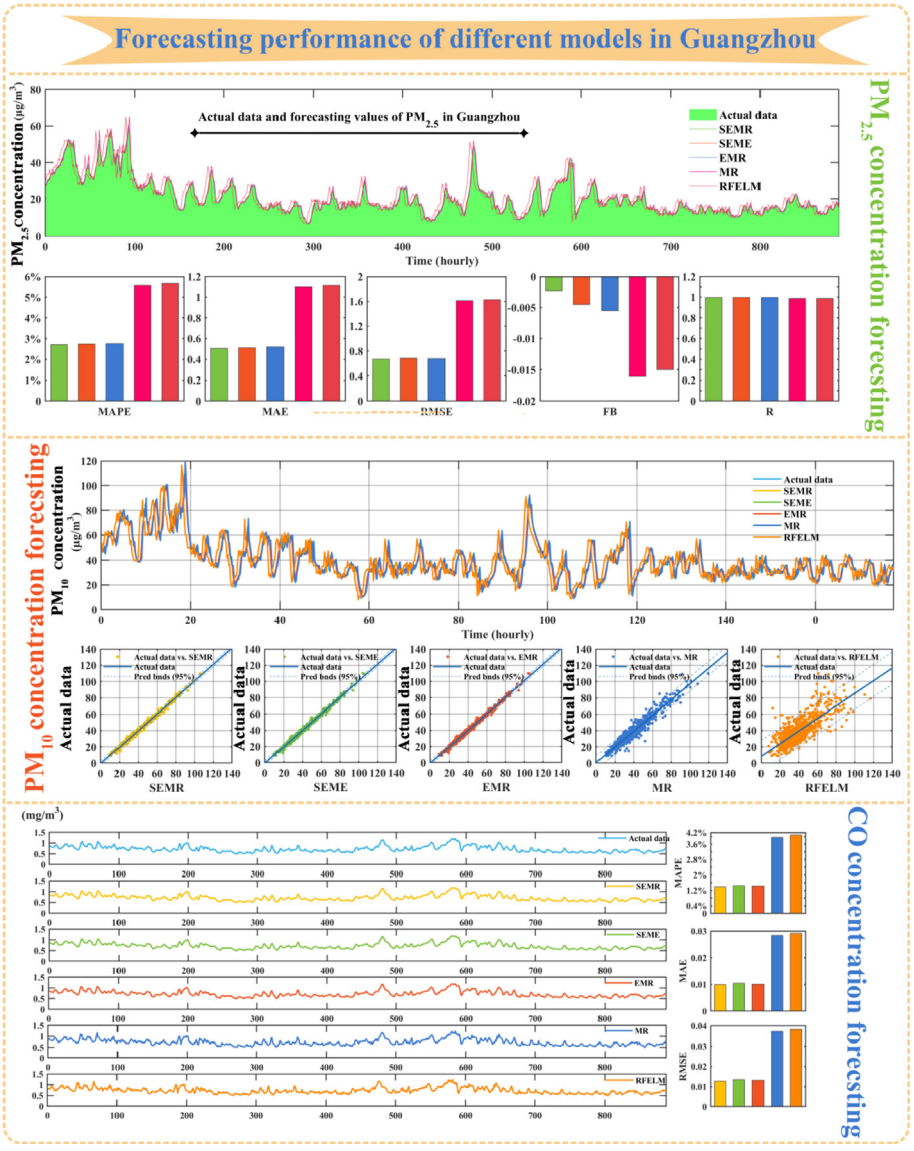


Fig. 4. Forecast results and boxplot for forecasting error for three pollutants in Guangzhou.

the other models. Typical criteria were used to evaluate the performance of the models. The results of all the experiments for the three cities show the excellent performance of the proposed model. The proposed model yields the lowest MAPE, MAE, RMSE, and R values, and the highest FB value. In summary, the novel hybrid model is not only effective and practical but also can be generalized. It yields a high accuracy level and effective forecasting results for air quality evaluation.

Remark: *A good forecasting model is the foundation of an early warning system. According to the analysis of the forecasting model proposed in this paper, the $L_{2,1}$-RFELM model can effectively be used in air pollution forecasting. The "decomposition and ensemble" process and the optimized process effectively improve the precision of the pollution concentration forecasting, and therefore, in our opinion the proposed hybrid forecasting model can be applied to construct an air pollution early warning system.*

4.6.2. Fuzzy synthetic evaluation in Shanghai and Guangzhou

In order to verify the generalizability of the proposed FSE system, it was applied to air pollution in Shanghai and Guangzhou. The results are divided into two parts: the evaluation results based on the actual value and the evaluation results based on the forecast value. Detailed information about the results for Shanghai and Guangzhou is shown in Tables 7 and 8, respectively.

The results shown in Tables 4, 7 and 8 indicate that the FSE system performs at an equivalent level for these cities. The evaluation accuracies for the three cities are Beijing 35/37, Shanghai 36/37, and Guangzhou 37/37. The precise evaluation results of the three cities indicate that the three major pollutants can replace six pollutants in air quality evaluation. The accurate evaluation results also prove the accuracy of the forecasting results.

According to the analysis of the experimental results, the air pollution early warning system proposed in this paper has yielded higher forecasting accuracy and precise evaluation results. The relevant governmental departments can issue the information of air quality to the public based on the evaluation results, and then, the public can take the corresponding protective measures. Moreover, the government can also adopt corresponding countermeasures based on the issued information to adjust the production of factories. Therefore, the system can provide not only early warning for air pollution, but also strong technical support for the relevant departments to formulate policies and solve potential air pollution problems to avoid serious pollution incidents.

Remark: *The performance level of the air quality evaluation based on precise forecasting results is equal for the three study areas. The results also prove the reliability of the proposed system. Therefore, citizens can take the corresponding protection measures according to the evaluation results. Moreover, the government can take corresponding measures to solve the potential air pollution problem to avoid heavy pollution incidents.*

5. Discussion

Although the results of model performance metrics used in this paper show that the novel hybrid forecasting model proposed in this paper is superior to the control models, the significance of the

Table 7
Air quality evaluation results of Shanghai.

Data Point	Evaluation results based on the forecast data						Evaluation results based on the actual data					
	Fuzzy evaluation result B					Level	Level	Fuzzy evaluation result B				
26/05/2017	0.2238	0.6685	0.1077	0.0000	0.0000	II	II	0.3125	0.5737	0.1139	0.0000	0.0000
27/05/2017	0.4125	0.5875	0.0000	0.0000	0.0000	II	II	0.4541	0.5096	0.0364	0.0000	0.0000
28/05/2017	0.3981	0.6019	0.0000	0.0000	0.0000	II	II	0.4084	0.5916	0.0000	0.0000	0.0000
29/05/2017	0.9402	0.0598	0.0000	0.0000	0.0000	I	I	0.9013	0.0987	0.0000	0.0000	0.0000
30/05/2017	0.7321	0.2679	0.0000	0.0000	0.0000	I	I	0.8503	0.1497	0.0000	0.0000	0.0000
31/05/2017	0.0996	0.7015	0.1989	0.0000	0.0000	II	II	0.2591	0.4987	0.2423	0.0000	0.0000
01/06/2017	0.0890	0.2066	0.6389	0.0654	0.0000	III	III	0.1922	0.2166	0.5539	0.0372	0.0000
02/06/2017	0.1403	0.6860	0.1737	0.0000	0.0000	II	II	0.3482	0.5704	0.0814	0.0000	0.0000
03/06/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.8290	0.1710	0.0000	0.0000	0.0000
04/06/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.9266	0.0734	0.0000	0.0000	0.0000
05/06/2017	0.9597	0.0403	0.0000	0.0000	0.0000	I	I	0.8938	0.1062	0.0000	0.0000	0.0000
06/06/2017	0.6419	0.3581	0.0000	0.0000	0.0000	I	I	0.7137	0.2863	0.0000	0.0000	0.0000
07/06/2017	0.1881	0.5334	0.2785	0.0000	0.0000	II	II	0.2686	0.5965	0.1349	0.0000	0.0000
08/06/2017	0.0805	0.6686	0.1876	0.0633	0.0000	II	II	0.0894	0.5050	0.3681	0.0374	0.0000
09/06/2017	0.0998	0.5072	0.3929	0.0000	0.0000	II	II	0.1255	0.6178	0.2566	0.0000	0.0000
10/06/2017	0.3286	0.6714	0.0000	0.0000	0.0000	II	II	0.4654	0.5346	0.0000	0.0000	0.0000
11/06/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.9720	0.0280	0.0000	0.0000	0.0000
12/06/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.9144	0.0856	0.0000	0.0000	0.0000
13/06/2017	0.9626	0.0374	0.0000	0.0000	0.0000	I	I	0.7072	0.2928	0.0000	0.0000	0.0000
14/06/2017	0.8457	0.1543	0.0000	0.0000	0.0000	I	I	0.7726	0.2274	0.0000	0.0000	0.0000
15/06/2017	0.4999	0.5001	0.0000	0.0000	0.0000	II	II	0.4514	0.5422	0.0064	0.0000	0.0000
16/06/2017	0.6006	0.3994	0.0000	0.0000	0.0000	I	I	0.5726	0.4274	0.0000	0.0000	0.0000
18/06/2017	0.6226	0.3774	0.0000	0.0000	0.0000	I	I	0.8333	0.1667	0.0000	0.0000	0.0000
19/06/2017	0.9196	0.0804	0.0000	0.0000	0.0000	I	I	0.8573	0.1427	0.0000	0.0000	0.0000
20/06/2017	0.6281	0.3719	0.0000	0.0000	0.0000	I	I	0.6244	0.3756	0.0000	0.0000	0.0000
21/06/2017	0.5433	0.4566	0.0001	0.0000	0.0000	II	I	0.5239	0.4761	0.0000	0.0000	0.0000
22/06/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.9434	0.0566	0.0000	0.0000	0.0000
23/06/2017	0.2432	0.4881	0.2687	0.0000	0.0000	II	II	0.2652	0.4670	0.2678	0.0000	0.0000
24/06/2017	0.4978	0.5022	0.0000	0.0000	0.0000	II	II	0.3806	0.4787	0.1408	0.0000	0.0000
25/06/2017	0.3160	0.5113	0.1727	0.0000	0.0000	II	II	0.2698	0.4951	0.2352	0.0000	0.0000
26/06/2017	0.0944	0.2073	0.5440	0.1542	0.0000	III	III	0.2053	0.2788	0.4210	0.0949	0.0000
27/06/2017	0.0925	0.2013	0.4754	0.2308	0.0000	III	III	0.2213	0.2723	0.3637	0.1427	0.0000
28/06/2017	0.2361	0.5360	0.2279	0.0000	0.0000	II	II	0.4174	0.4646	0.1180	0.0000	0.0000
29/06/2017	0.4972	0.5028	0.0000	0.0000	0.0000	**II**	I	0.5402	0.4598	0.0000	0.0000	0.0000
30/06/2017	0.4764	0.5236	0.0000	0.0000	0.0000	II	II	0.4217	0.5776	0.0007	0.0000	0.0000
01/07/2017	0.2735	0.5861	0.1403	0.0000	0.0000	II	II	0.3638	0.5688	0.0674	0.0000	0.0000

Table 8
Air quality evaluation results of Guangzhou.

Data Point	Evaluation results based on the forecasting data						Evaluation results based on the actual data					
	Fuzzy evaluation result B					Level	Level	Fuzzy evaluation result B				
26/05/2017	0.2797	0.6950	0.0253	0.0000	0.0000	II	II	0.3202	0.6115	0.0683	0.0000	0.0000
27/05/2017	0.2715	0.6711	0.0574	0.0000	0.0000	II	II	0.3349	0.5547	0.1104	0.0000	0.0000
28/05/2017	0.0935	0.1314	0.7751	0.0000	0.0000	III	III	0.2452	0.3601	0.3857	0.0089	0.0000
29/05/2017	0.1131	0.6843	0.2026	0.0000	0.0000	II	II	0.2056	0.7012	0.0933	0.0000	0.0000
30/05/2017	0.4244	0.5756	0.0000	0.0000	0.0000	II	II	0.4881	0.5040	0.0079	0.0000	0.0000
31/05/2017	0.6689	0.3311	0.0000	0.0000	0.0000	I	I	0.6317	0.3683	0.0000	0.0000	0.0000
01/06/2017	0.8845	0.1155	0.0000	0.0000	0.0000	I	I	0.8076	0.1924	0.0000	0.0000	0.0000
02/06/2017	0.7600	0.2400	0.0000	0.0000	0.0000	I	I	0.5393	0.4607	0.0000	0.0000	0.0000
03/06/2017	0.7378	0.2622	0.0000	0.0000	0.0000	I	I	0.5308	0.4692	0.0000	0.0000	0.0000
04/06/2017	0.8392	0.1608	0.0000	0.0000	0.0000	I	I	0.5744	0.4256	0.0000	0.0000	0.0000
05/06/2017	0.9885	0.0115	0.0000	0.0000	0.0000	I	I	0.8188	0.1812	0.0000	0.0000	0.0000
06/06/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.8801	0.1199	0.0000	0.0000	0.0000
07/06/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.6354	0.3646	0.0000	0.0000	0.0000
08/06/2017	0.9431	0.0569	0.0000	0.0000	0.0000	I	I	0.5880	0.4120	0.0000	0.0000	0.0000
09/06/2017	0.9142	0.0858	0.0000	0.0000	0.0000	I	I	0.4913	0.4055	0.1032	0.0000	0.0000
10/06/2017	0.9801	0.0199	0.0000	0.0000	0.0000	I	I	0.6665	0.3335	0.0000	0.0000	0.0000
11/06/2017	0.8998	0.1002	0.0000	0.0000	0.0000	I	I	0.7939	0.2061	0.0000	0.0000	0.0000
12/06/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.7704	0.2296	0.0000	0.0000	0.0000
13/06/2017	0.9758	0.0242	0.0000	0.0000	0.0000	I	I	0.5714	0.4286	0.0000	0.0000	0.0000
14/06/2017	0.8393	0.1607	0.0000	0.0000	0.0000	I	I	0.4371	0.2716	0.2913	0.0000	0.0000
15/06/2017	0.3743	0.6257	0.0000	0.0000	0.0000	I	II	0.4823	0.5177	0.0000	0.0000	0.0000
16/06/2017	0.9976	0.0024	0.0000	0.0000	0.0000	I	I	0.4924	0.4738	0.0338	0.0000	0.0000
17/06/2017	0.9719	0.0281	0.0000	0.0000	0.0000	I	I	0.4334	0.2008	0.3658	0.0000	0.0000
18/06/2017	0.1300	0.8700	0.0000	0.0000	0.0000	I	II	0.4739	0.5251	0.0010	0.0000	0.0000
20/06/2017	0.8553	0.1447	0.0000	0.0000	0.0000	I	I	0.4686	0.3497	0.1817	0.0000	0.0000
21/06/2017	0.9236	0.0764	0.0000	0.0000	0.0000	I	I	0.4990	0.4039	0.0971	0.0000	0.0000
22/06/2017	0.9030	0.0970	0.0000	0.0000	0.0000	I	I	0.4897	0.4204	0.0899	0.0000	0.0000
23/06/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.6893	0.3107	0.0000	0.0000	0.0000
24/06/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.7378	0.2622	0.0000	0.0000	0.0000
25/06/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.8133	0.1867	0.0000	0.0000	0.0000
26/06/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.8661	0.1339	0.0000	0.0000	0.0000
27/06/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.8801	0.1199	0.0000	0.0000	0.0000
28/06/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.8227	0.1773	0.0000	0.0000	0.0000
29/06/2017	0.9642	0.0358	0.0000	0.0000	0.0000	I	I	0.7598	0.2402	0.0000	0.0000	0.0000
30/06/2017	0.9753	0.0247	0.0000	0.0000	0.0000	I	I	0.7501	0.2499	0.0000	0.0000	0.0000
01/07/2017	1.0000	0.0000	0.0000	0.0000	0.0000	I	I	0.7596	0.2404	0.0000	0.0000	0.0000

model remains to be tested. According to relevant literature, Diebold–Mariano (DM) test is the most commonly used method to test the significance of forecasting models. In order to ensure the comparability of the results, this paper will also select the DM test to verify the significance of the proposed forecasting method.

DM test was proposed by Diebold and Marina in 1995. This method was proposed to test for whether two sets of forecast errors have equal mean value. DM test is a hypothesis test, a pair of hypotheses of which is as follows (Wu et al., 2019):

The Null hypothesis is that the loss functions of the forecast values of two models have the same unconditional expectation.

H_0:$E(dm) = 0$

The Alternative hypothesis is that the forecasting capabilities of the two models are different.

H_1:$E(dm) \neq 0$

The D-M statistic is calculated as follow formula:

$$DM = \frac{\sum_{i=1}^{m}(L(\partial_m^{(1)})\text{-}L(\partial_m^{(2)}))/m}{\sqrt{S^2/m}} S^2$$

The explanation of each variable in the formula is as follows:

1. $\partial_m^{(1)}$ represents forecasting error of first method.
2. $\partial_m^{(2)}$ is forecasting error of second method.
3. $L\left(\partial_m^{(1)}\right)$ is a loss function of first method, which used to estimate the forecasting precision. And in this paper we chose absolute deviation loss:

$$L(\partial_m^{(i)}) = \left|\partial_m^{(i)}\right|$$

4. S^2 is a variance estimator of the difference of loss function $dm = L(\partial_m^{(1)})\text{-}L(\partial_m^{(2)})$.

DM follows the standard normal distribution under null hypothesis, so the rejection region is $|DM| > z_{\alpha/2}$. It is means that if the value of DM falls in the interval $\left[-z_{\alpha/2}, z_{\alpha/2}\right]$ the null hypothesis of no difference will be accepted, else the null hypothesis will be rejected (Gao et al., 2016). In this study, the DM test results are show in the Table 9.

The results of Table 9 indicate that the proposed forecasting model is significantly superior to the other six models. For Beijing, the values of the DM statistic between the proposed model and the control models are larger than the upper limit at a 1% significance level, which means the proposed model is distinct superiority over the other models in forecasting Beijing pollutants' concentration. And for Guangzhou, the value of the DM statistic between the proposed model and SSA-EEMD-MOALO-ELM is larger than the upper limit at a 5% significance level, this result means there is a weaker significant superiority between those two models. There are three values smaller than the limit at a 1% significance level in Shanghai's results, but in general, the proposed model is significant then control models.

In summary, the proposed model outperforms other control models, and for different pollutants, the superiority is also different.

Table 9
DM test result of three pollutants in three research cities.

		Beijing	Shanghai	Guangzhou
$PM_{2.5}$	MR	3.7672*	10.4080*	6.1226*
	EMR	3.3252*	1.4984***	4.1974*
	SEME	3.6204*	2.3029**	4.5904*
	BPNN	3.8031*	11.1070*	6.4116*
	ARIMA	3.7654*	11.0366*	5.9510*
	$L_{2,1}$RF-ELM	3.4945*	9.6985*	6.0251*
PM_{10}	MR	5.8715*	3.1601*	12.4637*
	EMR	5.7991*	3.1601*	10.3919*
	SEME	6.1879*	1.5206***	2.1596**
	BPNN	5.6750*	5.0354*	12.2556*
	ARIMA	5.5420*	5.2005*	12.4395*
	$L_{2,1}$RF-ELM	5.3663*	4.6675*	11.9389*
CO	MR	5.6732*	9.4685*	13.9766*
	EMR	4.6265*	8.0884*	2.8076*
	SEME	3.8183*	8.1703*	1.9618**
	BPNN	5.2622*	9.6582*	13.7855*
	ARIMA	5.5566*	10.0692*	12.4049*
	$L_{2,1}$RF-ELM	5.3989*	9.4575*	13.4534*

*indicates the 1% significance level; ** indicates the 5% significance level; *** indicates the 15% significance level.

6. Conclusion

Air pollution has attracted worldwide attention in recent years, not only because it is harmful to people and the environment, but also because it is difficult to control. In order to solve these problems, a two-module air pollution early warning system was proposed in this paper. The results of this early warning system can not only guide humans, but also provide a basis for air pollution control. When air pollution is expected to reach a certain level, the production activities of some severely polluting factories can be halted and residents can be reminded to go outdoors less and take more protective measures.

The first module of the early warning system is the forecasting module. Its objective is to provide accurate forecasting of pollutant concentrations. In this module, an EML based on $L_{2,1}$-norm and random Fourier ($L_{2,1}$RFELM), a novel and very precise algorithm, was employed. Based on this method, a hybrid air quality forecasting model was proposed. It also integrates the principle of "decomposition and ensemble" and a multi-objective optimization algorithm. The first step of this module follows the principle of "decomposition and ensemble." "Principal components," a method from the perspective of dynamic reconstruction of time series, combined with empirical orthogonal functions, was used to decompose the original time series into two series, one of which contains most of the information in the original series and is called the trend sequence. The second series is called the main sequence. Then, the trend sequence is forecast by using a supervised learning algorithm for single hidden layer feed-forward neural networks and obtains the forecasting value of the trend sequence. Meanwhile, the main sequence is decomposed again by applying a method based on the "decomposition and ensemble" principle. After decomposition, the main sequence becomes IMFs and we use the $L_{2,1}$-RFELM optimized by multi-objective ant lion optimizer (MOALO) to forecast each IMF. Ultimately, the main forecast value and the trend forecast value are added to obtain the final forecast results. In our study, the proposed hybrid model was applied to forecast the hourly air pollution concentration of major pollutants in Beijing, Shanghai, and Guangzhou. The forecasting results of the hybrid model were compared with those of benchmark models. The comparative results indicate that the proposed hybrid model outperformed the other models for the three cities. The second module is the air quality evaluation element, in which FSE was used to assess air quality based on the forecast concentration.

Most research on air pollution forecasting is aimed at achieving a high forecasting accuracy level, but ignores further research and the applications of the forecasting results. Forecasting is meaningless if not applied to human life. Therefore, using the accurate forecasting results, a further study addressing air pollution early warning systems can be conducted. According to the information provided by the early warning system, people can adopt measures to control pollution and take precautions against air pollution. The early warning system proposed in this study provides accurate forecasting and an effective evaluation. Overall, it can guide people's activities to avoid the influence of air pollution.

Acknowledgement

This work was supported by Major Program of National Social Science Foundation of China (Grant No.17ZDA093).

References

Anwar, F., Chaudhry, F.N., Nazeer, S., Zaman, N., Azam, S., 2016. Causes of ozone layer depletion and its effects on Human : review. Atoms. Climate Sci. 6, 129–134. https://doi.org/10.4236/acs.2016.61011.

Bai, L., Wang, J., Ma, X., Lu, H., 2018. Air Pollution Forecasts: an Overview. International Journal of Environmental Research and Public Health. https://doi.org/10.3390/ijerph15040780.

Boudraa, A.O., Cexus, J.C., 2007. EMD-based signal filtering. IEEE Trans. Instrum. Meas. 56, 2196–2202. https://doi.org/10.1109/TIM.2007.907967.

Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., Pope, C.A., Shin, H., Straif, K., Shaddick, G., Thomas, M., van Dingenen, R., van Donkelaar, A., Vos, T., Murray, C.J.L., Forouzanfar, M.H., Forouzanfar, M.H., 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. Lancet 389, 1907–1918. https://doi.org/10.1016/S0140-6736(17)30505-6.

Desonie, D., 2007. Atmosphere: Air Pollution and its Effects.

Ding, C., Zhou, D., He, X., Zha, H., 2006. R1-PCA: rotational invariant L1-norm principal component analysis for robust subspace factorization. In: Proceedings of the 23rd International Conference on Machine Learning - ICML '06, pp. 281–288. https://doi.org/10.1145/1143844.1143880.

Du, P., Jin, Y., Zhang, K., 2016. A hybrid multi-step rolling forecasting model based on SSA and simulated annealing-adaptive particle swarm optimization for wind speed. Sustainability 8. https://doi.org/10.3390/su8080754.

Du, P., Wang, J., Guo, Z., Yang, W., 2017. Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting. Energy Convers. Manag. 150, 90–107. https://doi.org/10.1016/j.enconman.2017.07.065.

Gao, Y., Qu, C., Zhang, K., 2016. A hybrid method based on singular spectrum analysis, firefly algorithm, and BP neural network for short-term wind speed forecasting. Energies 9. https://doi.org/10.3390/en9100757.

Golyandina, N., Nekrutkin, V., Zhigljavsky, A., 2001. Analysis of Time Series Structure: SSA and Related Techniques, vol. 1706. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, p. 320. https://doi.org/10.1198/jasa.2002.s239.

Guo, Y., Zeng, H., Zheng, R., Li, S., Barnett, A.G., Zhang, S., Zou, X., Huxley, R., Chen, W., Williams, G., 2016. The association between lung cancer incidence and ambient air pollution in China: a spatiotemporal analysis. Environ. Res. 144, 60–65. https://doi.org/10.1016/j.envres.2015.11.004.

Heng, J., Wang, C., Zhao, X., Wang, J., 2016. A hybrid forecasting model based on empirical mode decomposition and the cuckoo search algorithm: a case study for power load. Mathematical Problems in Engineering 2016. https://doi.org/10.1155/2016/3205396.

Huang, G.-B., Zhu, Q., Siew, C.-K., 2006. Extreme learning machine: theory and applications. Neurocomputing 70, 489–501. https://doi.org/10.1016/j.neucom.2005.12.126.

Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., Yen, N., Tung, C., Liu, H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. Math. Phys. Eng. Sci. 454, 995, 903. https://doi.org/10.1098/rspa.1998.0193.

Huang, G. Bin, Zhu, Q.Y., Siew, 2004. Proceedings, C.K.B.T.-I.I.J.C. On N.N., 2005. Extreme Learning Machine: a New Learning Scheme of Feedforward Neural Networks, vol. 2, pp. 985–990.

Liu, L., Wang, Q., Wang, J., Liu, M., 2016. A rolling grey model optimized by particle swarm optimization in economic prediction. Comput. Intell. 32, 391–419. https://doi.org/10.1111/coin.12059.

Ma, X., Jin, Y., Dong, Q., 2017. A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting. App. Soft Comput. J. 54, 296–312. https://doi.

org/10.1016/j.asoc.2017.01.033.
Mirjalili, S., Jangir, P., Saremi, S., 2017. Multi-objective ant lion optimizer: a multi-objective optimization algorithm for solving engineering problems. Appl. Intell. 46, 79–95. https://doi.org/10.1007/s10489-016-0825-8.
Nie, F., Huang, H., Cai, X., Ding, C., 2010. Efficient and robust feature selection via joint l2,1-norms minimization. In: Advances in Neural Information Processing Systems, pp. 1813–1821.
Rahman, N.H.A., Lee, M.H., Suhartono Latif, M.T., 2015. Artificial neural networks and fuzzy time series forecasting: an application to air quality. Qual. Quantity 49, 2633–2647. https://doi.org/10.1007/s11135-014-0132-6.
Rahman, P.A., Panchenko, A.A., Safarov, A.M., 2017. Using Neural Networks for Prediction of Air Pollution Index in Industrial City, p. 42016.
Ramanathan, V., Feng, Y., 2009. Air pollution, greenhouse gases and climate change: global and regional perspectives. Atmos. Environ. 43, 37–50. https://doi.org/10.1016/j.atmosenv.2008.09.063.
Wang, J., Du, P., Niu, T., Yang, W., 2017a. A novel hybrid system based on a new proposed algorithm—multi-Objective Whale Optimization Algorithm for wind speed forecasting. Appl. Energy 208, 344–360. https://doi.org/10.1016/j.apenergy.2017.10.031.
Wang, J., Heng, J., Xiao, L., Wang, C., 2017b. Research and application of a combined model based on multi-objective optimization for multi-step ahead wind speed forecasting. Energy 125, 591–613. https://doi.org/10.1016/j.energy.2017.02.150.
Wu, C., Wang, J., Chen, X., Du, P., Yang, W., 2019. A novel hybrid system based on multi-objective optimization for wind speed forecasting. Renew. Energy. https://doi.org/10.1016/j.enconman.2017.07.065.
Wu, Z., Huang, N.E., 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv. Adapt. Data Anal. 1, 1–41. https://doi.org/10.1142/S1793536909000047.
Xu, Y., Du, P., Wang, J., 2017a. Research and application of a hybrid model based on dynamic fuzzy synthetic evaluation for establishing air quality forecasting and early warning system: a case study in China. Environ. Pollut. 223, 435–448. https://doi.org/10.1016/j.envpol.2017.01.043.
Xu, Y., Yang, W., Wang, J., 2017b. Air quality early-warning system for cities in China. Atmos. Environ. 148, 239–257. https://doi.org/10.1016/j.atmosenv.2016.10.046.
Xue, J., Zhou, S.H., Liu, Q., Liu, X., Yin, J., 2017. Financial time series prediction using ℓ2,1RF-ELM. Neurocomputing 277, 176–186. https://doi.org/10.1016/j.neucom.2017.04.076.
Zhao, J., Guo, Y., Xiao, X., Wang, J., Chi, D., Guo, Z., 2017. Multi-step wind speed and power forecasts based on a WRF simulation and an optimized association method. Appl. Energy 197, 183–202. https://doi.org/10.1016/j.apenergy.2017.04.017.
Zhou, S., Liu, X., Liu, Q., Wang, S., Zhu, C., Yin, J., 2016. Random Fourier extreme learning machine with ℓ2,1-norm regularization. Neurocomputing 174, 143–153. https://doi.org/10.1016/j.neucom.2015.03.113.
Zhu, S., Lian, X., Liu, H., Hu, J., Wang, Y., Che, J., 2017. Daily air quality index forecasting with hybrid models: a case in China. Environ. Pollut. 231, 1232–1244. https://doi.org/10.1016/j.envpol.2017.08.069.