Novel analysis–forecast system based on multi-objective optimization for air quality index

Hongmin Li[a], Jianzhou Wang[a*], Ranran Li[a], Haiyan Lu[b]

[a] *School of Statistics, Dongbei University of Finance and Economics, Dalian, China*

[b] *School of Software, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia*

[*] *Corresponding author. Address: School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China*

Tel.:+8613130476286

E-mail address: wangjz@dufe.edu.cn

Abstract

The air quality index (AQI) is an important indicator of air quality. Owing to the randomness and non-stationarity inherent in AQI, it is still a challenging task to establish a reasonable analysis–forecast system for AQI. Previous studies primarily focused on enhancing either forecasting accuracy or stability and failed to improve both aspects simultaneously, leading to unsatisfactory results. In this study, a novel analysis–forecast system is proposed that consists of complexity analysis, data preprocessing, and optimize–forecast modules and addresses the problems of air quality monitoring. The proposed system performs a complexity analysis of the original series based on sample entropy and data preprocessing using a novel feature selection model that integrates a decomposition technique and an optimization algorithm for removing noise and selecting the optimal input structure, and then forecasts hourly AQI series by utilizing a modified least squares support vector machine optimized by a multi-objective multi-verse optimization algorithm. Experiments based on datasets from eight major cities in China demonstrated that the proposed system can simultaneously obtain high accuracy and strong stability and is thus efficient and reliable for air quality monitoring.

Key words: Air quality index forecasting; Analysis–forecast system; Feature selection; Multi-objective optimization algorithm; Forecasting accuracy and stability

1. Introduction

Air pollution has become a serious problem in many parts of the world [1], and air pollution control is a high priority in government policy agenda [2]. Owing to industrialization and urbanization, the emission of waste gas from factories and automobiles has become the major source of air pollution [3]. Moreover, with the population growth due to economic prosperity, the environmental burden has increased. It is worth mentioning that the presence of harmful substances including particulate

matter (PM), carbon monoxide (CO), ozone (O_3), and nitrogen oxides (NO_x) in the atmosphere at higher concentrations than their normal ambient rates will cause damage to human health, ecosystems and the environment [4]; furthermore, there is evidence that long-term exposure to atmospheric pollutants can lead to respiratory and cardiovascular diseases [5–7]. Effective information management in urban areas can facilitate daily life and raise awareness for reducing vehicular traffic. Thus, it is necessary to accurately mine and monitor air quality information.

To effectively evaluate air quality, a large number of air quality indexes have been used. China National Environmental Monitoring Centre (CNEMC) started using the air pollution index (API) in 2000 to evaluate air quality. However, it was found that API cannot represent air quality well. Recently, the Ministry of Environmental Protection (MEP) of the People's Republic of China began applying a novel index, namely the "air quality index" (AQI) in January 2013 [8]. AQI is an important indicator for comprehensively representing the state of air pollution, which is related to human health [9]. It simplifies the concentrations of several pollutants into a single numerical form that can be calculated by new ambient air quality standards (GB3095-2012) [10]. The definition of AQI is given in **Appendix A**. It shows the changing trend of air pollution and provides data interpretation for air pollution mitigation measures and air quality control. Unfortunately, the randomness and non-stationarity of AQI often lead to poor forecasting accuracy and low stability. Therefore, it is urgent to develop a proper analysis–forecast system for AQI. The atmosphere is a highly complex dynamic system that is quite difficult to model. Therefore, it is a challenging task to forecast AQI or any other pollution index. A variety of models have been developed to forecast the concentration of $PM_{2.5}$, PM_{10}, and other pollutants [11–16]. They can be roughly divided into three categories: chemical transport models (CTMs), statistical models, and machine learning models.

CTMs are widely used as deterministic models in air quality forecasting [17]. They have the advantage of forecasting air pollutant concentrations without requiring site monitoring and a large quantity of historical data. However, forecasting accuracy is significantly affected by the quality and the scale of the emission data [18–19]. Furthermore, as Stern et al. pointed out, owing to the limited knowledge regarding pollutant sources and the incomplete representation of physicochemical processes, there may be rather strong biases in concentration prediction based on CTMs [20].

Statistical models are commonly applied in air quality forecasting, among them, autoregressive (AR) models [21], autoregressive integrated moving average (ARIMA) models [22, 23], and multiple linear regression (MLR) models [24]. However, these approaches have limitations. Namely, modeling is based on certain assumptions, they are unsuitable for long-term prediction and nonlinear series analysis [25, 26].

Machine learning models, including artificial neural networks (ANNs), fuzzy logic models, and support vector machines (SVMs) are supposed to be the methods of choice, as they can overcome the disadvantages of conventional models that cannot successfully handle nonlinear series [13, 27, 28]. For example, Bai et al. applied a back propagation neural network (BPNN) with a wavelet transform model to forecast daily air pollutant concentration and achieved high forecasting performance [29]. Li et al.

proposed a new extended long short-term memory neural network model that inherently considers spatiotemporal correlations for air pollutant concentration prediction and achieved superior performance [30]. Wang et al. developed an air quality early warning system based on a least squares support vector machine (LSSVM) [31]. Based on the above analysis, ANN models outperform other models with the advantage of high forecasting accuracy in nonlinear series forecasting. However, they also have certain disadvantages, as potential convergence to a local minimum and over-fitting may lead to insufficient accuracy [32].

No single model is flawless; thus, it has been attempted to integrate novel methods with artificial intelligence for constructing strong hybrid algorithms [33]. Owing to the world-wide information explosion, data mining techniques, such as phase-space reconstruction [34], Gaussian process regression [35], decomposition methods [36], feature selection techniques [37], and optimization algorithms [27], are preferable for integration with artificial intelligence. Accordingly, owing to their superior performance, hybrid models present a promising choice.

Although the components of a hybrid model are usually selected so as to achieve specific goals based on actual application problems, there are still several issues. Most approaches are based on decomposing the original series into several intrinsic mode functions (IMFs), then removing the noise part, and finally using the remaining IMFs as input variables for the forecasting model; however, the effectiveness of the model also depends on the structure of input variables, considering the one aspect is unreasonable. In addition, most previous studies concentrated only on enhancing accuracy, whereas few studies simultaneously considered forecasting stability, as only one forecasting factor is insufficient.

Over the past few years, some algorithms inspired by nature, such as particle swarm optimization (PSO) [38], the firefly algorithm (FA) [39], differential evolution (DE) [40], and the bat algorithm (BA) [41] have been successfully developed to solve single-objective problems. However, in multi-objective problems, the objective function of single-objective optimization algorithms cannot be balanced simultaneously, leading to unreasonable results. Fortunately, multi-objective optimization algorithms have been developed in various fields, such as mechanical engineering [42], finance [43], and electric power systems [44]. However, few studies have applied multi-objective optimization algorithms to enhance both accuracy and stability for air quality index forecasting. As air quality index series are quite nonstationary and nonlinear [45], multi-objective optimization algorithms are preferable.

Based on the above analysis on air quality monitoring and forecasting, some drawbacks and deficiencies of conventional air pollution forecasting models can be summarized as follows: (1) Traditional single models cannot meet the forecasting accuracy requirements for time series, whereas hybrid models are superior. (2) Previous studies focused on eliminating the impact of noise, whereas feature selection was overlooked. (3) Most recent studies use single-objective algorithms, and thus the accuracy and stability requirements can hardly be met simultaneously. To overcome these deficiencies, a novel analysis–forecast system is developed for hourly AQI series. It consists of three modules: complexity analysis, data preprocessing, and optimize–

forecast modules. Pertaining to the first module, a complexity analysis module based on sample entropy (SE) is constructed to analyze the nonlinearity and uncertainty in the AQI series. With respect to the data preprocessing module, a Hampel filter is first applied to remove the outliers; then, to further enhance accuracy, a novel feature selection model integrating complementary ensemble empirical mode decomposition (CEEMD) and a binary chaotic crow search algorithm (BCCSA) is proposed. As noise may reduce accuracy, an advanced decomposition approach (CEEMD) is utilized to decompose the AQI series, and then a search agent based on chaotic maps (BCCSA) is used to optimize the input structure, with two techniques combined to achieve feature selection. Regarding the optimize–forecast module, a modified LSSVM based on multi-objective multi-verse optimization (MOMVO) is proposed to forecast AQI series with both high accuracy and strong stability. To verify the effectiveness and stability of the proposed model, the Wilcoxon rank-sum and stability tests were conducted. To assess the performance of the proposed analysis–forecast system, AQI data from eight target cities in China were selected as datasets, and a variety of performance metrics were considered.

The main contributions of this study are as follows:

- A comprehensive analysis–forecast system is developed comprising complexity analysis, data preprocessing, and optimize–forecast modules. Several numerical implementations demonstrate the effectiveness and the reliability of the proposed system.
- A novel feature selection model is successfully constructed to remove noise and optimize the input structure.
- In the optimize–forecast module, a modified LSSVM with parameters optimized by MOMVO outperforms other models with high accuracy and strong stability.
- Experiments based on eight datasets were conducted, and the excellent performance of the analysis–forecast system demonstrated that it can provide accurate warning information and scientific support for decision-makers.
- Further discussions based on the system's forecasting ability, stability, and application confirm its superiority.

The remainder of this paper is organized as follows: the related approaches are briefly introduced in Section 2; the proposed system is presented in Section 3; experiments and results are described in Sections 4 and 5; further discussions are presented in Section 6; finally, in Section 7 the paper is concluded.

2. System development

In this section, the framework of the analysis–forecast system is methodically introduced. Moreover, two test methods are utilized to evaluate its performance.

2.1 System design

The framework of the analysis–forecast system is shown in Fig. 1. It comprises the complexity analysis module, the data preprocessing module, and the optimize–forecast module. Detailed descriptions are as follows:

Ⅰ. **Complexity analysis module**: Exhaustive analysis and mining of original data facilitate more accurate processing and modeling. Moreover, AQI sample data, with their strong volatility and complexity, render forecasting a challenging task. Therefore, sample entropy is applied to analyze the complexity of AQI and better understand the data characteristics, thus providing an analytical basis for modeling. Moreover, the changes in complexity before and after preprocessing can be compared and analyzed.

II. Data preprocessing module: Outliers can interfere with model learning, resulting in poor forecasting accuracy. Thus, a Hampel filter is utilized to eliminate outliers in AQI, which has the advantages that the outlier at which the interference occurs need not be known a priori and that series are not distorted after their processing. Moreover, as noise in AQI reduces accuracy, a novel feature selection approach is proposed integrating the CEEMD and BCCSA algorithms. CEEMD is applied to reduce noise as much as possible and extract the valuable information component in the original AQI and decomposed series into several IMFs. Furthermore, to better characterize and select the input series, the BCCSA algorithm with a high-speed and accurate search mechanism is employed to select the reasonable IMFs after CEEMD decomposed and reconstructed them into input series.

III. Optimize–forecast module: LSSVM regression forecasting with RBF kernel function is widely used in forecasting. However, the forecasting results are often unsatisfactory owing to the sensitivity of the model parameters, namely, the regularization parameter (gam: γ) and the squared kernel bandwidth parameter (sig^2: δ). Accordingly, the MOMVO algorithm with a leader selection mechanism in Pareto archived evolution strategy is employed to optimize the parameters of the LSSVM, which is essential for enhancing accuracy as well as forecasting stability.

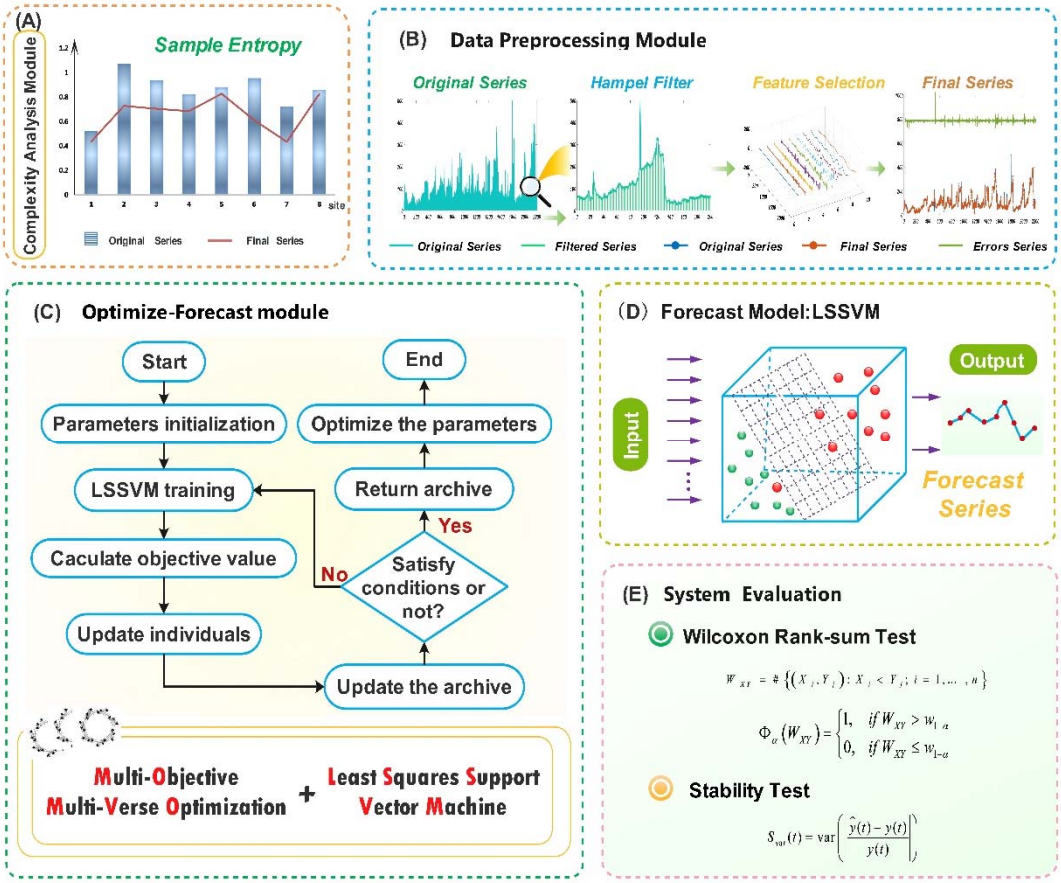

Fig.1. The framework of the proposed analysis-forecast system

2.2 Preliminaries

In this section, complexity analysis used sample entropy and a filter outlier model are introduced. Then, a novel feature selection method, namely CBCCSA, is proposed. Finally, the forecasting model LSSVM optimized by MOMVO is briefly put forth.

2.2.1 Complexity analysis with sample entropy

AQI, which directly reflects air quality, is characterized by randomness, uncertainty, and complexity. Thus, sample entropy is applied in this study to analyze the complexity of AQI series.

Sample entropy is a nonlinear analysis approach with satisfactory consistency and reflects nonlinear characteristics by measuring the complexity of the signal [46]. Lower sample entropy implies higher signal self-similarity, i.e., a simpler signal, and vice versa [47]. The specific steps of the algorithm are as follows:

(1) For an n-point time series $\{u(j)\,|\,1 \le j \le N\}$, construct m vector sequences $X_m(i) = \{u(i), u(i+1), \cdots, u(i+m-1)\}$ for $i = 1, 2, \ldots, N-m+1$.

(2) Define the distance between any two m vectors.

$$d\left[X_m(i), X_m(j)\right] = \max_{k=0,\ldots,m-1}\left(\left|x(i+k) - x(j+k)\right|\right) \tag{1}$$

(3) Given a threshold r, for each i count the number $N^m(i)$ of $d\left[X_m(i), X_m(j)\right]$ less than $r \times SD$ and calculate the ratio to the total number of distances.

$$C_i^m(r) = N^m(i)/(N-m+1) \tag{2}$$

(4) Compute the average value of $C_i^m(r)$:

$$C^m(r) = (N-m)^{-1}\sum_{i=1}^{N-m} C_i^m(r) \tag{3}$$

(5) For dimension $m+1$

$$C^{m+1}(r) = (N-m)^{-1}\sum_{i=1}^{N-m} C_i^{m+1}(r) \tag{4}$$

(6) Finally, sample entropy can be defined as:

$$SampEn(m,r) = \lim_{N\to\infty}\left\{-\ln\left[C^{m+1}(r)/C^m(r)\right]\right\} \tag{5}$$

which is estimated by the statistic

$$SampEn(m,r,N) = -\ln\left[C^{m+1}(r)/C^m(r)\right] \tag{6}$$

2.2.2 Data preprocessing with feature selection

In view of the randomness and complexity of the AQI series, a Hampel filter is initially used before analysis and prediction to eliminate outliers in the series. Furthermore, an appropriate selection of the input variables of the forecasting model can simplify the model and enhance forecasting accuracy. In this study, a novel feature selection model based on CEEMD and BCCSA is proposed to determine the IMFs for the forecasting model input. The related theory is now briefly presented.

2.2.2.1 Hampel filter

The Hampel filter described by Allen [48] is an offline frequency-domain filtering method for removing spectral outliers. In this method, the Fast Fourier Transform (FFT) is utilized to transform the time-domain signal into the frequency domain, outliers are detected by the Hampel identifier in the real spectrum and the imaginary spectrum, these outliers are replaced with interpolated values, and the cleaned spectrum is converted back to the time domain through the inverse FFT [49].

In essence, the Hampel identifier defines outliers as the data points in **Eq. 7** whose absolute difference from the median is greater than a predetermined threshold t, where x^* denotes the median of the data sequence of length N, and S is the median absolute difference (MAD) scale estimator, as defined by **Eq. 8**, which quantifies the statistical difference between the median absolute deviation of a data sequence and its median

value.

$$\left|x_j - x^*\right| > tS \quad j = 1 \ to \ N \tag{7}$$

$$S = 1.4286 \ median\left\{\left|x_j - x^*\right|\right\} \tag{8}$$

The time domain Hampel filter was developed by Pearson [50]. It applies MAD scale estimation to locate outliers occurring in sliding windows for process modeling and identification. If outliers are detected in the window, they are replaced by the median of the window data. Therefore, the length of the moving window L and the threshold t are two tunable parameters in the filter. It can be seen that as t increases, fewer data points are detected as outliers, whereas a reduction in t results in an increase in the number of outliers. In fact, the value of t can be set to ~5 with $N > 10$ [51].

2.2.2.2 Feature selection

A proper feature selection method may facilitate discarding irrelevant features in the selection of input variables. In the data preprocessing stage, the focus was usually on the decomposition of the original series into a finite number of oscillatory modes arranged from low to high frequencies, which are called IMFs. Then the high frequency noise was eliminated for data cleaning. The detection and removal of high-frequency IMFs was often overlooked. Reckless rejection of the highest-frequency IMF is unreasonable and results in suboptimal forecasting accuracy. Thus, in this study, a novel feature selection approach is developed whereby BCCSA is applied to determine and eliminate the IMFs with a negative impact on forecasting after CEEMD.

First, removing noise from the data series significantly improves the forecasting accuracy of the model, and CEEMD is applied to decompose the original series into a limited number of IMFs. The original signal $x(t)$ is decomposed as follows:

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{9}$$

where $c_i(t)$ is the i-th IMF with different frequency bands from high to low, and $r_n(t)$ is the n-th residue, which denotes the mean trend of $x(t)$.

By improving empirical mode decomposition (EMD) and ensemble empirical mode decomposition (EEMD), CEEMD adds white noise in positive–negative pairs to the original time series, so that the decomposition is more effective and has less computation cost [47]. The procedure of CEEMD is briefly described as follows:

(1) $x_m^+(t)$ and $x_m^-(t)$ are generated by adding a pair of opposite-phase white noise processes $n_m(t)$ with identical amplitudes to the original signal $x(t)$.

$$\begin{bmatrix} x_m^+(t) \\ x_m^-(t) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x(t) \\ n_m(t) \end{bmatrix}, \quad m = 1,2,\ldots,M \tag{10}$$

(2) $c_{i,m}^+$ and $c_{i,m}^-$ are obtained by performing EEMD to $x_m^+(t)$ and $x_m^-(t)$, respectively.

(3) The ensemble means of all corresponding IMFs are computed,

$$c_i = (1/2M)\sum_{m=1}^{M}(c_{i,m}^+ + c_{i,m}^-) \tag{11}$$

where c_i is the ith IMF component derived by CEEMD.

After CEEMD, BCCSA is employed, which has proved a promising method for finding an optimal feature subset that maximizes the classification performance and minimizes the number of selected features [52]. It is assumed that there are M crows in a D-dimensional search space, and $y^{j,t}$ is the position of j crow at iteration t in the search space. Each crow knows where the food is hidden, and $N^{j,t}$ represents the position of the hiding place at iteration t for crow j; moreover, $N^{j,t}$ is defined as the current best position obtained so far by crow j. At iteration t, crow j wants to track the hiding place of crow z. Therefore, the position updating equation is defined as follows:

$$y^{j,t+1} = \begin{cases} y^{j,t} + R_j \times fl^{j,t} \times \left(N^{z,t} - y^{j,t}\right), & R_z \geq qAP^{j,t} \\ Choose \quad a \quad rand \quad position, & otherwise \end{cases} \tag{12}$$

where R_z is a random number in [0,1] and $AP^{j,t}$ denotes the awareness probability of crow z at t iteration. AP plays a significant role in controlling the balance between exploration and exploitation. Small values of AP lead to local research (exploitation), whereas large values lead to global search (exploration). Each crow updates its memory as follows:

$$N^{j,t+1} = \begin{cases} y^{j,t+1}, & Fn\left(y^{j,t+1}\right) \text{ is better than } Fn\left(N^{j,t}\right) \\ N^{j,t}, & otherwise \end{cases} \tag{13}$$

where *Fn()* is defined as the objective function.

In BCCSA, chaotic rather than random variables are used for updating the crow position. A chaotic sequence generated by chaotic maps is used to update the crow position, which has a significant influence on the optimal solution and convergence rate. When the iteration termination criteria are satisfied, the crow's best position is obtained as the optimal solution. The pseudo code of BCCSA is given in **Algorithm 1**.

Algorithm 1. BCCSA

1: /*Set the initial values of *M*, *AP*, *fl*, and *tMax*.*/

2: /*Initialize the crow position *y* randomly.*/

3: /*Evaluate the fitness function of each crow *Fn(y)*.*/

4: /*Initialize the memory of search crow N.*/
5: **WHILE**($t<tMax$) **DO**
6: **FOR EACH** j=1:j($j \le M$) **DO**
7: /*Get value of chaotic map C.*/
8: **IF** $C_z \ge AP^{z,t}$ **THEN**
9: $y^{j,t+1}=y^{j,t}+C_j \times fl^{j,t} \times (N^{z,t}-y^{j,t})$
10: **ELSE**
11: $y^{j,t+1}$=A random position of the search space
12: **END IF**
13: $$y^{j,t+1}=\begin{cases}1 & if\,(s(y^{j,t+1})) \ge rand() \\ 0 & Otherwise\end{cases}$$
14: **END FOR**
15: /*Check the feasibility of $y^{j,t+1}$.*/
16: /*Evaluate the new position of crow $Fn(y^{j,t+1})$.*/
17: /*Update the crow's memory $N^{j,t+1}$.*/
18: $t=t+1$
19: **END WHILE**
20: /*Produce the best solution N.*/

2.2.3 Least squares support vector machine

LSSVMs improve SVMs by transforming complicated quadratic programming problems into linear problems, which has the advantages of generalization and low computational cost [53]. They have been successful applied in nonlinear time series forecasting, such as air quality forecasting [54], gas consumption demand forecasting [55], and wind power forecasting [53]. However, AQI series have strong nonlinear and stochastic characteristics, and thus accurate forecasting is challenging. Considering the superiority of LSSVMs in nonlinear time series forecasting, in this study, they are applied to forecast AQI series. The theory of LSSVMs is now briefly presented.

Technically, the training set is defined as $\{(x_k, y_k) \mid k=1,2,\ldots,n\}$, where $x_k \in R^n$ is the input variable and $y_k \in R^n$ is the corresponding output variable. The specific form of LSSVM model is as follows:

$$y(x)=w^T \square \phi(x)+b \tag{14}$$

where w and b represent the weights and the bias, respectively, and $\phi(x)$ is the nonlinear mapping function that maps the training set into a higher-dimensional linear feature space. More details about LSSVMs can be found in [56].

2.2.4 Multi-objective multi-verse optimization algorithm

Multi-verse optimizers (MVOs) have the advantages of simple construction and fewer controlling parameters; moreover, they require less complicated computations. To converge to the best solutions in a multi-objective search space, MOMVOs, based on the same concepts as MVOs, have recently been proposed [57].

2.2.4.1 Multi-verse optimizer

The MVO algorithm was proposed by Mirjalili et al. [58] inspired by physics theories on the existence of parallel universes, which claim that objects in the universe transferred from white holes to black holes through tunnels and wormholes also have the capacity to transfer objects without the need for white or black holes. In the MVO algorithm, the white holes and the black holes are used for exploration, whereas the wormholes are utilized for the exploitation process [59]. The inflation rate is the growing speed of a universe and plays a significant role in the improvement of solutions over the course of generations. Higher inflation rate leads to more white holes and fewer black holes, which ensures the flow of variables from better universes to worse ones. The mutation of MVOs is not absolutely random, as the wormholes establish tunnels between the optimal solution obtained so far and any solutions in the population. Elitism and the mechanism of saving the best current solution improve the exploitation of MVOs. The process can be described mathematically as follows:

$$x_i^j = \begin{cases} \begin{cases} x_j + TDR \times \left(\left(ub_j - lb_j \right) \times r_4 + lb_j \right) & r_3 < 0.5 \\ x_j - TDR \times \left(\left(ub_j - lb_j \right) \times r_4 + lb_j \right) & r_3 \geq 0.5 \end{cases} & r_2 < WEP \\ x_i^j & r_2 \geq WEP \end{cases} \tag{15}$$

where x_i^j denotes the j-th variable in the i-th solution, x_j indicates the j-th variable in the best solution, lb_j and ub_j are the lower and the upper bound of the j-th dimension, respectively, $r_2 - r_4$ are random values in [0,1], and finally WER denotes the wormhole existence rate and TDR the travelling distance rate. These two variables change adaptively during iterations to balance exploration and exploitation.

2.2.4.2 Multi-objective multi-verse optimizer

MOMVOs have a similar search mechanism to that of MVOs, and their solutions are improved using white holes, black holes, and wormholes. To select solutions from the archive and create tunnels between solutions, a leader selection mechanism is employed. Specifically, the crowding distance between each solution in the archive is first selected, and the number of solutions in the neighborhood is counted as a measure of coverage or diversity in the approach. Then, to improve the distribution of solutions in the archive across all objectives, a roulette wheel from the less populated regions of the archive is applied to select solutions using the following equation:

$$p_i = c / N_i \tag{16}$$

where c is a constant greater than 1, and N_i denotes the number of solutions near the i-th solution.

With high probability, this equation transfers the solutions to less populated areas and ultimately increases the coverage of solutions in the Pareto optimal front obtained.

Obviously, the archive can accommodate a limited number of solutions and may completely fill in the optimization process. Thus, there should be a mechanism for removing unsatisfactory solutions from the archive. As an unsatisfactory solution usually has several adjacent solutions, it should be eliminated from the archive. **Eq. 17**, which is the inverse of **Eq. 16**, is used to eliminate unsatisfactory solutions with high probability:

$$P_i^{'} = N_i / c \tag{17}$$

where c is a constant greater than 1, and N_i denotes the number of solutions near the i-th solution.

Through the MOMVO mechanism, it is possible to store Pareto optimal solutions in the archive and improve them by iterating. The algorithm for comparing and adding a solution to the archive complies with the following principles [60]:

a. If a new solution dominates any solutions in the archive, it should be replaced by them immediately.

b. If a new solution does not dominate a solution in the archive, it should be deleted and not allowed to enter the archive.

c. If a solution is non-dominated with respect to all solutions in the archive, it should be added to the archive.

d. If the archive is full, an unsatisfactory solution should be deleted and a new non-dominated solution should be added to the archive.

Algorithm 2: MOMVO-LSSVM

Obiective functions:

$$\min \begin{cases} obf_1(y) = MSE = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2 \\ obf_2(y) = std\left(\hat{y}_i\right), i = 1,2,\ldots,N \end{cases}$$

Input:

$y_t^{(0)} = \left(y^{(0)}(1), y^{(0)}(2), \ldots, y^{(0)}(q)\right)$ —the training samples

$y_f^{(0)} = \left(y^{(0)}(q+1), y^{(0)}(q+2), \ldots, y^{(0)}(q+l)\right)$—the test samples

Output:

$\hat{y}_f^{(0)} = \left(y^{(0)}(q+1), y^{(0)}(q+2), \ldots, y^{(0)}(q+l)\right)$—the forecasting data

Parameters:

iter—the current iteration
Max iter—the maximum number of iterations
p—the exploitation accuracy
N_i—the number of solutions in the vicinity of the *i-th* solution.
c—a constant greater than one
x_j—indicates the *j-th* variable in the best solution
ub_j—upper bound of *j-th* dimension
lb_j—lower bound of *j-th* dimension
r—random values

/*Set the parameters of the MOMVO algorithm.*/
/*Initialize the positions of universes randomly.*/
/*Initialize *iter*=1.*/
WHILE (*iter* <*Max iter*) **DO**
    /*Update WEP and TDR by applying equations.*/
    $$WEP = \min + iter \cdot \left(\frac{\max - \min}{MAX\_iter} \right) \text{and} TDR = 1 - \frac{(iter)^{\frac{1}{p}}}{MAX\_iter^{\frac{1}{p}}}$$
    FOR EACH *i*=1:*n* **DO**
        /*Boundary checking (to bring back the universes inside search
        space if they go beyond the boundaries).*/
        /*Calculate the inflation rate (fitness) of universes.*/
    END FOR
    /*Sort fitness values.*/
    /*Find the non-dominated solutions.*/
    /*Normalized inflation rates by applying equations.*/
    /*Update the archive in regard to the obtained non-dominated solutions.*/
    IF the archive is full **DO**
        /*Delete some solutions from the archive to hold the new solutions.*/
        Applying Roulette wheel and $p_i^{'} = N_i / c$
    END IF
    /*Update the Position of universes by utilizing the equations.*/
    $$x_i^j = \begin{cases} \begin{cases} x_j + TDR \times \left(\left(ub_j - lb_j \right) \times r_4 + lb_j \right) & r_3 < 0.5 \\ x_j - TDR \times \left(\left(ub_j - lb_j \right) \times r_4 + lb_j \right) & r_3 \geq 0.5 \end{cases} & r_2 < WEP \\ x_i^j & r_2 \geq WEP \end{cases}$$
    IF any new added solutions to the archive are outside boundaries **DO**
        /*Update the boundaries to cover the new solution(s).*/
    END IF
    iter = *iter* +1
END WHILE
RETURN best universe
Obtain **X***=best universe
Set parameters of LSSVM according to **X***.
Use x_i to train and update the parameters of LSSVM.

32: Input the historical data into LSSVM to obtain the forecasting value $\hat{y}_f$

3. System evaluation

To evaluate the performance of the proposed forecasting model in terms of forecasting accuracy and stability, two well-known tests were conducted: Wilcoxon rank-sum test and stability test.

3.1 Wilcoxon rank-sum test

To compare the proposed model compared with other benchmarks, a Wilcoxon rank-sum test was conducted. It is more sensitive than the *t*-test, as it assumes proportionality of differences between two pair samples. Moreover, it is safer than the *t*-test, as it does not assume normal distributions [61]. Assuming that *X* and *Y* are absolutely continuous random variables with distribution functions *F* and *G*, respectively, then let $X_m = (X_1, \cdots, X_m)$ and $Y_n = (Y_1, \cdots, Y_n)$ be independent random samples from *F* and *G*, respectively. Assuming that the null hypothesis is that *X* and *Y* are equal in distribution, the alternative hypothesis is that *Y* is stochastically strictly greater than *X*, i.e.,

$$H_0 : Y \underline{d} X \quad vs. \quad H_1 : Y >_{st} X \tag{18}$$

The statistic (written in the Mann–Whitney form) is expressed as follows:

$$W_{XY} = \#\left\{\left(X_i, Y_j\right) : X_i < Y_j ; i = 1, \ldots, n\right\} \tag{19}$$

For a given level α, the Wilcoxon rank-sum test is

$$\Phi_\alpha\left(W_{XY}\right) = \begin{cases} 1, & if\ W_{XY} > w_{1-\alpha} \\ 0, & if\ W_{XY} \le w_{1-\alpha} \end{cases} \tag{20}$$

where w_q is the *q*-quantile of the null distribution of W_{XY}. The null distribution W_{XY} depends only on the sample sizes *m* and *n*. The approximated values of W_{XY} can be obtained by the well-known normal approximation as follows:

$$\frac{W_{XY} - mn/2}{\left(mn(m+n+1)/12\right)^{1/2}} \square N(0,1) \tag{21}$$

3.2 Stability test

The forecasting stability of the model affects its application value. A model with high stability can better withstand external interference and ensure the reliability of the forecasting results. Therefore, a suitable stability test was devised in this study. The detailed formula is as follows:

$$S_{\mathrm{var}}(t) = \mathrm{var}\left(\left|\frac{\hat{y}(t) - y(t)}{y(t)}\right|\right) \tag{22}$$

where $\hat{y}(t)$ represents the forecasted series and $y(t)$ denotes the original series. The relative error is the ratio of the absolute error to the true value and better reflects the credibility of the measurement. To avoid the impact of positive and negative offset and better perceive the fluctuations of the relative error, its absolute value is considered.

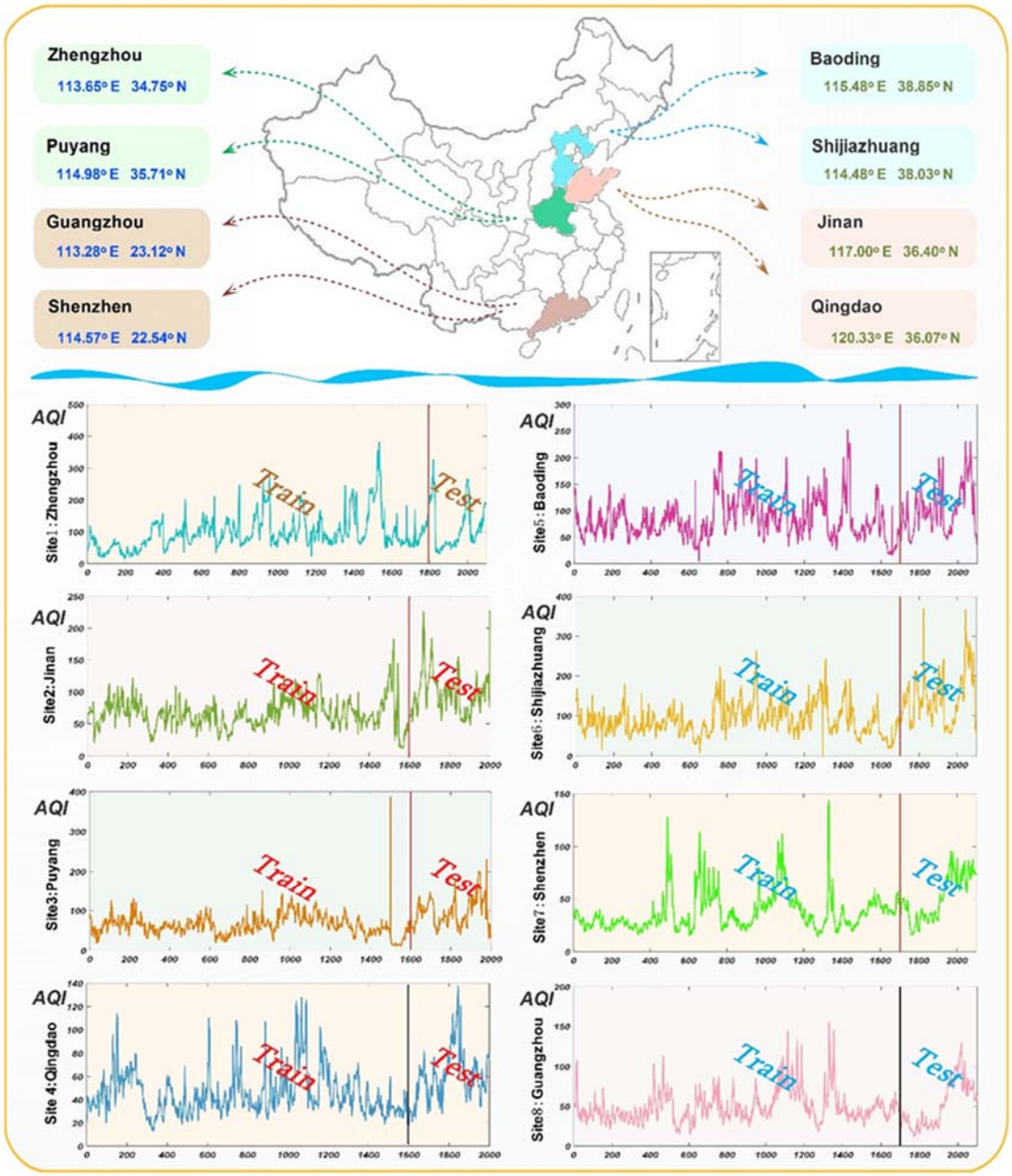

Fig.2. Sites and data description

4. Study region and data description

In this section, the specific information of validation datasets and the location

information of selected sites are described in detail. To further assess the performance of the proposed forecasting model, three performance metrics are provided. The experiments conducted in this study were implemented on Matlab2016a, running on a Windows 8.1 Professional operating system. The specific hardware parameters were as follows: Intel (R) Core i5-4590 3.30 GHz CPU and 8 GB RAM.

4.1 Validation data sets

AQI is used to monitor air quality and reflect its trends, which plays an important role in air quality control. In this study, case studies involving hourly average AQI data from Qingyue Open Environmental Data Center were conducted to evaluate the performance of the proposed mode [62]. Thus, the datasets from eight sites in China located in ZZ (Zhengzhou), JN (Jinan), PY (Puyang), QD (Qingdao), BD (Baoding), SJZ (Shijiazhuang), SZ (Shenzhen) and GZ (Guangzhou) were tested. Among them, ZZ and PY are major cities in Henan Province. Henan Province has a warm temperate to subtropical, humid to semi-humid monsoon climate. It has four distinct seasons, simultaneous rain and heat, and complex and diverse features. The population of Henan Province is relatively dense, and the level of air pollutants is high. In addition to the geographic location, the suspended particles in the air are difficult to disperse into the atmosphere and are blocked at low altitude and near the ground. Therefore, hazy weather is frequent. JN and QD are located in Shandong Province. The climate in Shandong is of warm temperate monsoon type. Owing to industrial structure problems and the geographic location, Shandong Province is also affected by hazy weather. BD and SJZ are in Hebei Province. The frequent hazy weather in Hebei Province is related not only to economic factors but also to the geographic location. Hebei Province is located in the North China Plain. The air is not particularly cold, the wind is weak, and the atmospheric structure is stable. Owing to evaporation and precipitation, the relative humidity of the near-surface air is so large that the pollutants are not easily diffused in the atmosphere. SZ and GZ are located in Guangdong Province, which is located in the southernmost part of mainland China. Most of it is in the subtropical zone, and a small part in the tropics. This unique geographical location also makes the winter climate of Guangdong warm and springy. The haze in Guangdong Province is seriously exacerbated by industrial emissions and automobile exhausts. In view of their hazy weather, eight cities from the above provinces were selected in this study. Forecasting AQI not only is important for verifying the performance of the proposed system but also provides significant information support for future air quality monitoring and decision management in these cities. Fig. 2 shows the basic information of the original data and the location information of the example sites. Table 1 shows the statistical characteristics of the original data. Each data set was divided into two parts: a training set and a testing set. Considering that the total number of samples and the ratio of the size of the training set to the size of the testing set have a considerable impact on the experimental results, eight case studies with different number of observations were conducted. The testing sets contained 400 observations to ensure that the aforementioned ratios are different. Details are in Table 1. Hourly AQI data were

selected for experimental analysis. In the selected eight datasets, the test set contained 400 sample observation points, that is, AQI series were forecasted 16 days and 16 hours ahead. This is significant for monitoring and early warning.

4.2 Performance evaluation

To assess the effectiveness of the proposed forecasting model, three widely used performance metrics, i.e., mean absolute percent error (MAPE), mean absolute error (MAE), and root mean square error (RMSE), were adopted. They are defined as follows:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{y}(t)-y(t)}{y(t)}\right|\times 100\% \tag{23}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}(t)-y(t)\right| \tag{24}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}(t)-y(t)\right)^2 \tag{25}$$

Table 1
The statistics characteristics of the original data

Site	City	Time Period	All Samples	Train Samples	Test Samples	Min	Max	Std	SE
1	**Zhengzhou**	2017.10.01-2017.12.31	2200	1800	400	16	500	66.0223	0.5244
2	**Jinan**	2017.08.09-2017.10.31	2000	1600	400	12	228	28.3885	1.0722
3	**Puyang**	2017.08.09-2017.10.31	2000	1600	400	11	387	30.5903	0.9361
4	**Qingdao**	2017.08.09-2017.10.31	2000	1600	400	13	221	21.3019	0.8220
5	**Baoding**	2017.08.05-2017.10.31	2100	1700	400	5	252	40.3750	0.8792
6	**Shijiazhuang**	2017.08.05-2017.10.31	2100	1700	400	8	369	48.6820	0.9546
7	**Shenzhen**	2017.08.05-2017.10.31	2100	1700	400	14	144	19.9698	0.7221
8	**Guangzhou**	2017.08.05-2017.10.31	2100	1700	400	12	155	21.5709	0.8605

Table 2
Experimental results of the first four sites

Models	MAPE					MAE					MSE				
	ZZ	**JN**	**PY**	**QD**	**AV**	**ZZ**	**JN**	**PY**	**QD**	**AV**	**ZZ**	**JN**	**PY**	**QD**	**AV**
ENN	7.53	5.64	10.76	6.18	**7.53**	8.91	6.53	11.47	3.60	**7.63**	206.97	179.54	364.60	25.08	**194.05**
GRNN	11.94	8.56	13.29	8.46	**10.56**	16.54	9.46	13.91	5.01	**11.23**	710.53	273.66	474.87	49.62	**377.17**
RBF	10.84	16.77	17.36	10.62	**13.90**	16.35	29.65	20.91	7.91	**18.70**	2219.32	58175.00	5979.79	714.73	**16772.21**
SVM	30.25	15.78	23.38	11.33	**20.19**	57.54	19.95	27.40	7.82	**28.18**	8934.71	1232.35	1756.40	190.10	**3028.39**
LSSVM	17.96	12.31	18.44	9.39	**14.52**	33.53	15.97	22.30	6.32	**19.53**	3878.31	967.31	1375.15	125.93	**1586.68**
DP*	11.58	9.37	14.79	6.54	**10.57**	21.06	12.66	18.73	4.48	**14.23**	1951.00	780.47	1167.38	75.25	**993.52**
MM*	6.58	9.66	9.16	5.83	**7.81**	8.73	10.39	10.06	3.35	**8.13**	214.86	650.78	351.44	21.24	**309.58**
SSAMM*	5.72	4.79	8.53	5.13	**6.04**	7.10	5.22	9.39	2.96	**6.17**	143.16	109.59	309.06	16.96	**144.69**
EMDMM*	5.26	8.62	6.69	3.82	**6.10**	6.52	9.34	6.90	1.76	**6.13**	125.22	664.40	133.56	5.32	**232.13**
Proposed	3.72	2.70	4.89	3.09	**3.60**	4.57	3.10	4.85	1.78	**3.57**	52.66	60.42	59.94	5.68	**44.67**
	BD	**SJZ**	**SZ**	**GZ**	**AV**	**BD**	**SJZ**	**SZ**	**GZ**	**AV**	**BD**	**SJZ**	**SZ**	**GZ**	**AV**

ENN	6.92	7.73	4.81	12.84	**8.07**	6.97	12.42	2.03	3.85	**6.32**	93.32	469.96	8.08	24.56	**148.98**
GRNN	9.43	10.10	6.71	11.14	**9.35**	9.36	16.48	2.95	4.60	**8.35**	167.27	766.55	14.68	49.69	**249.55**
RBF	11.12	24.77	9.83	10.09	**13.95**	11.90	49.99	6.68	5.37	**18.49**	759.73	141303.31	268.22	218.76	**35637.51**
SVM	18.33	27.78	5.59	20.53	**18.06**	23.78	51.24	3.04	7.64	**21.42**	1505.48	5606.26	21.38	200.12	**1833.31**
LSSVM	15.55	14.30	4.71	16.00	**12.64**	19.50	27.42	2.07	5.30	**13.57**	1037.58	2749.42	8.50	96.00	**972.87**
DP*	9.95	11.01	3.96	11.92	**9.21**	11.54	21.53	1.63	4.28	**9.74**	362.16	2081.34	5.23	58.06	**626.70**
MM*	6.52	12.20	3.43	8.85	**7.75**	6.67	22.87	1.72	4.03	**8.82**	92.46	2098.81	6.36	55.26	**563.22**
SSAMM*	6.09	7.02	3.11	8.26	**6.12**	6.25	11.32	1.57	3.73	**5.72**	82.72	390.17	5.21	44.55	**130.66**
EMDMM*	6.12	5.72	3.31	5.71	**5.21**	7.52	8.74	1.78	2.76	**5.20**	89.45	260.12	6.75	18.74	**93.77**
Proposed	5.86	3.58	2.24	5.42	**4.28**	6.13	5.79	1.08	2.37	**3.84**	79.48	160.27	2.18	11.06	**63.25**

Note: AV represents the average value of examples, DP represents data preprocessing, MM represents multi-objective multi-verse optimizer,* represents LSSVM

Table 3

Experimental results of the last four sites

Models	**MAPE**					**MAE**					**MSE**				
	ZZ	**JN**	**PY**	**QD**	**AV**	**ZZ**	**JN**	**PY**	**QD**	**AV**	**ZZ**	**JN**	**PY**	**QD**	**AV**
DPCS*	5.40	5.07	9.56	3.50	**5.88**	9.05	6.98	12.67	2.19	**7.72**	485.73	434.07	841.91	17.64	**444.84**
DPFA*	5.96	4.92	9.42	3.43	**5.93**	9.82	6.79	12.52	2.19	**7.83**	593.14	428.62	839.01	19.15	**469.98**
DPBA*	9.74	4.57	8.93	3.82	**6.77**	18.88	6.38	11.93	2.67	**9.96**	1865.41	419.66	810.06	36.94	**783.02**
Proposed	3.72	2.70	4.89	3.09	**3.60**	4.57	3.10	4.85	1.78	**3.57**	52.66	60.42	59.94	5.68	**44.67**
	BD	**SJZ**	**SZ**	**GZ**	**AV**	**BD**	**SJZ**	**SZ**	**GZ**	**AV**	**BD**	**SJZ**	**SZ**	**GZ**	**AV**
DPCS*	6.26	8.17	2.82	10.15	**6.85**	6.58	16.27	1.27	3.89	**7.00**	96.15	1584.86	2.87	45.53	**432.36**
DPFA*	7.88	8.24	3.25	9.61	**7.24**	8.69	16.57	1.64	3.92	**7.70**	196.46	1651.92	5.32	50.92	**476.16**
DPBA*	6.15	7.90	2.64	9.53	**6.56**	6.44	15.80	1.21	3.81	**6.82**	93.24	1540.99	2.59	45.31	**420.53**
Proposed	5.86	3.58	2.24	5.42	**4.28**	6.13	5.79	1.08	2.37	**3.84**	79.48	160.27	2.18	11.06	**63.25**

Note: AV represents the average value of examples, DP represents data processing,* represents LSSVM

5. Experimental results

In this section, the effectiveness of the proposed system is experimentally evaluated.

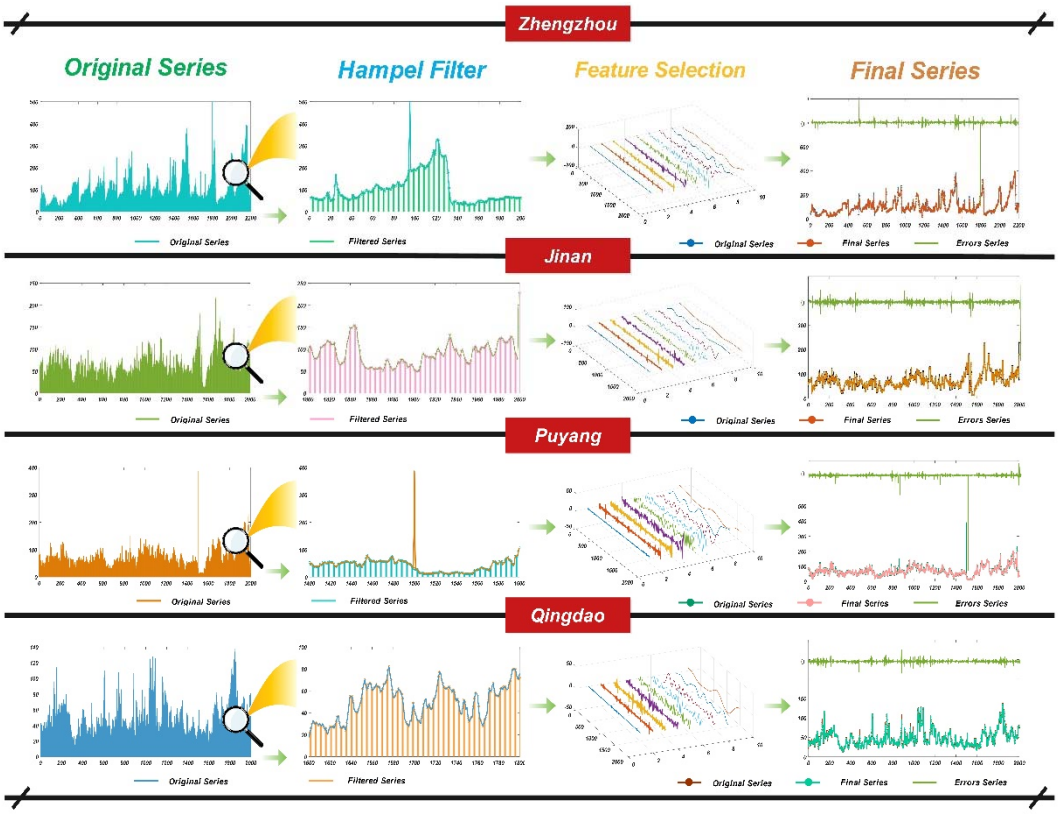

Fig.3. The results of complexity analysis and data preprocessing in the first four sites

5.1 Complexity analysis

AQI is an important indicator reflecting the quality of air, calculated from the concentration of major pollutants. It may be influenced by human activities and spatiotemporal conditions. Therefore, it has uncertainty, volatility, and randomness. Moreover, the operation and the decision support for an air monitoring system rely on a large number of field-measured data. However, distortion of the measured data inevitably occurs during the operation of the monitoring system, which also causes the randomness of AQI. Therefore, complexity analysis is critical to the prediction of AQI. The original AQI series from eight sites are illustrated in Fig. 2, where the large uncertainty, volatility, and randomness are clearly seen. SE was employed to analyze the complexity of the original series, where relatively high values indicate high complexity. The values of SE from eight sites are shown in Fig. 3 and Table 1. It can be seen that the series of site2 have the largest complexity, whereas the series of site1 have the smallest complexity. However, this does not imply that the air quality of site1 is better than that of site2; rather, the series of site1 are less complex. The values of SE

are significantly reduced after outlier filtering and feature selection, indicating that data preprocessing can reduce the complexity of the series and thus facilitate the establishment of the model.

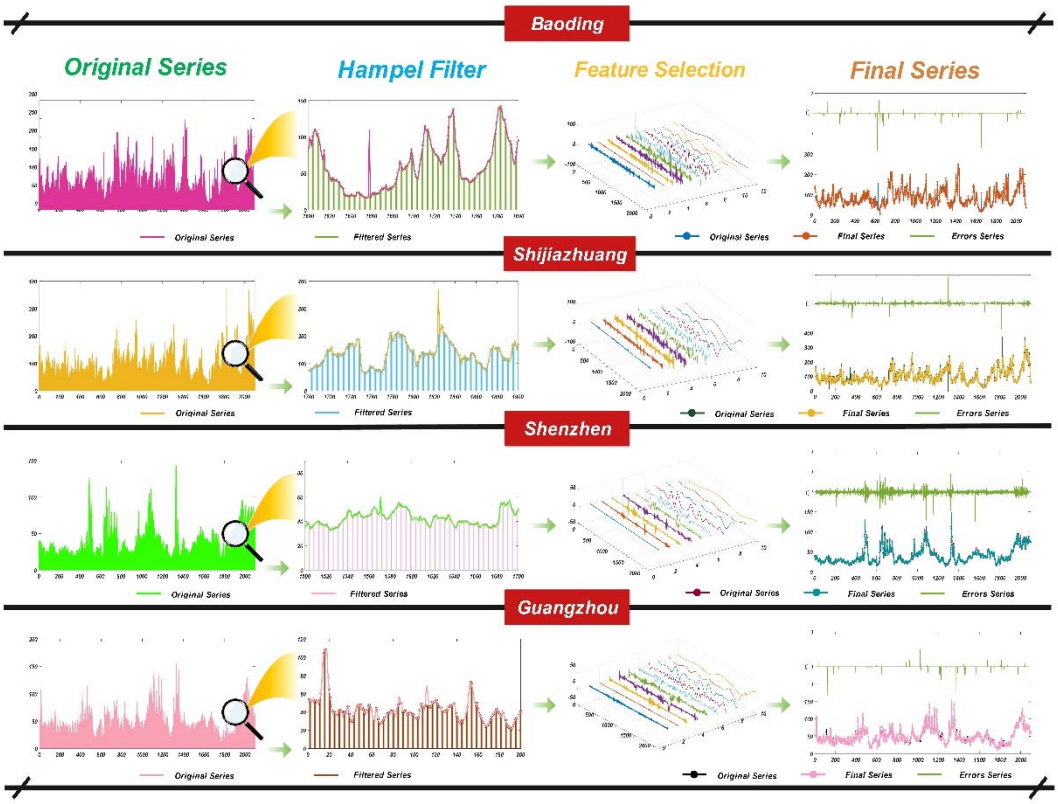

Fig.4. The results of complexity analysis and data preprocessing in the last four sites

5.2 Analysis of data preprocessing results

Occurrences of extreme air pollution occasionally result in outliers in the AQI series. Outliers can interfere with the training of the model and affect its forecasting performance. Thus, eliminating outliers is a significant step toward enhancing forecasting accuracy. The results of eliminating outliers are shown in Fig. 3-4.

After outliers were eliminated, the proposed feature selection approach was applied to determine the input variables. First, the AQI series were decomposed by CEEMD into several IMFs. In this study, the number of ensemble members was set to 200 because an excessively large number of ensemble members may increase computational cost. The standard deviation of the added white noise in each ensemble member was set to 0.2. The feature selection part in Fig. 3-4 shows the decomposition results of the AQI series, and the IMFs are listed in decreasing order of frequency. Then, BCCSA was employed to select the optimal subset of the AQI series and reconstruct them to obtain the optimal input series. From the feature selection results in Fig. 3-4, it can be seen that the quality is obviously improved compared with the original series.

Finally, compared with the changes of SE values, the complexity of the series significantly reduced and the series became cleaner. The results verified the effectiveness of data preprocessing.

5.3 Model comparison

To verify the effectiveness of the proposed forecasting model, five single models, namely, an Elman neural network (ENN), a general regression neural network (GRNN), a radial basis neural network (RBF), an SVM, and an LSSVM, and two hybrid models (DP* and MM*) were selected as benchmarks for comparison.

For single-model forecasting, performance is affected not only by different models but also by different sample numbers. Table 2 shows the experimental results with different sample numbers. The values of MAPE of the proposed model were 3.72%, 2.7%, 4.89%, 3.09%, 5.86%, 3.58%, 2.24%, and 5.42% for the eight sites, which are smaller compared with the corresponding values for the other five single models. The results demonstrated that the proposed model is superior. Under the same conditions, the values of MAE and MSE of the proposed model are almost the smallest. Moreover, compared with ENN, GRNN, RBF, SVM, and LSSVM, the decrease of the average MAPE of the proposed model was 3.93%, 6.96%, 10.30%, 16.59%, and 10.92%, respectively in the first four sites, and the condition of the last four sites was the same, as shown in Table 4. This also verified the effectiveness of the proposed model.

For hybrid-model forecasting, DP* and MM* were compared with the proposed model, which achieved the smallest MAPE, MAE, and MSE for the eight sites, as shown in Table 2-3. The experimental results demonstrated the superiority of the proposed model. Compared with LSSVM, DP*, and MM*, the reduction of the MAPE for the proposed model was 10.92%, 6.97%, and 4.21%, respectively, in first four sites, and 8.36%, 4.93%, and 3.47%, respectively, in the last four sites, as shown in Table 2. Regarding the values of MAE and MSE, the proposed model had a significant decrease, which demonstrated that it is more precise than DP* and MM*.

Remark: The experimental results demonstrated that the forecasting ability of hybrid models is greater than that of single models, which implies that the utilization of outlier filtering, feature selection, and optimization enhances forecasting. It can be concluded that the proposed model outperforms single and hybrid models in terms of forecasting accuracy.

5.4 Evaluation based on Wilcoxon rank-sum test

To test the significance of the difference between the performance of the proposed model and that of the other models, the Wilcoxon rank-sum test was conducted. The null hypothesis was that the forecasting error series from two different models have no significance difference. Statistical values were calculated to determine the case that would allow rejection of the null hypothesis at a given significance level. The Wilcoxon rank-sum statistical values are denoted by p, are calculated by **Eq. 19**, and are listed in Table 4. Some conclusions can be drawn as follows:

(a) The test results were different from different sites, but the Wilcoxon rank-sum statistical values of all benchmark models were greater than the critical value of 10%

significance level in all eight sites.

(b) The Wilcoxon rank-sum test values for the ENN, LSSVM, DP*, MM*, SSAMM*, and EMDMM* models were greater than the critical value of 5% significance level in all eight sites.

(c) The results of the Wilcoxon rank-sum test indicated that the proposed model markedly outperforms the other models at different significance levels, which demonstrates its superiority.

Table 4

The results of Wilcoxon Rank-sum Test.

Model	ZZ			JN			PY			OD		
	p	α_1	α_2	p	α_1	α_2	p	α_1	α_2	p	α_1	α_2
ENN	0.00	1	1	0.00	1	1	0.00	1	1	0.00	1	1
LSSVM	0.00	1	1	0.00	1	1	0.00	1	1	0.00	1	1
DP*	0.00	1	1	0.00	1	1	0.00	1	1	0.00	1	1
MM*	0.00	1	1	0.00	1	1	0.00	1	1	0.00	1	1
SSAMM*	0.00	1	1	0.00	1	1	0.00	1	1	0.00	1	1
EMDMM*	0.02	1	1	0.00	1	1	0.01	1	1	0.00	1	1
DPCS*	0.09	0	1	0.01	1	1	0.02	1	1	0.03	1	1
DPFA*	0.09	0	1	0.03	1	1	0.04	1	1	0.06	0	1
Model	**BD**			**SJZ**			**SZ**			**GZ**		
	p	α_1	α_2	p	α_1	α_2	p	α_1	α_2	p	α_1	α_2
ENN	0.00	1	1	0.00	1	1	0.00	1	1	0.00	1	1
LSSVM	0.00	1	1	0.00	1	1	0.00	1	1	0.00	1	1
DP*	0.00	1	1	0.00	1	1	0.00	1	1	0.00	1	1
MM*	0.00	1	1	0.00	1	1	0.00	1	1	0.00	1	1
SSAMM*	0.00	1	1	0.00	1	1	0.00	1	1	0.00	1	1
EMDMM*	0.00	1	1	0.00	1	1	0.00	1	1	0.00	1	1
DPCS*	0.03	1	1	0.01	1	1	0.02	1	1	0.02	1	1
DPFA*	0.07	0	1	0.04	1	1	0.08	0	1	0.03	1	1

Note: the α_1 indicate the 5% significance level and α_2 indicate the 10% significance level respectively.

6. Further discussion

To further analyze and evaluate the performance of the proposed model and it characteristics, the internal mechanisms of the system are further discussed.

6.1 Effectiveness of the proposed model

The analysis of the above experimental results has proved that the proposed hybrid model is superior to other comparison models in terms of forecasting performance. To fully demonstrate the effectiveness of the model, three additional metrics are introduced to provide more evidence. They are briefly presented in Table 5. Table 6 shows the improvement of the proposed model over the comparison models.

Table 5

Three additional metrics of the model.

Metric	Definition	Equation
P_{MAPE}	The improvement percentages of MAPE	$P_{MAPE}=\left\|\frac{MAPE_1-MAPE_2}{MAPE_1}\right\|\times 100\%$
P_{MAE}	The improvement percentages of MAE	$P_{MAE}=\left\|\frac{MAE_1-MAE_2}{MAE_1}\right\|\times 100\%$
P_{MSE}	The improvement percentages of MSE	$P_{MSE}=\left\|\frac{MSE_1-MSE_2}{MSE_1}\right\|\times 100\%$

By comparing the results in Table 6, the following can be concluded:

(a) It can be clearly seen that the improvement in the forecasting performance of the proposed model is obvious.

(b) For a definitive site comparison order (i.e., comparison with LSSVM, DP*, and MM*), the values of the improvement percentage for forecasting error metrics gradually decrease. That is, the forecasting performance of the benchmarks gradually increases. This also demonstrates the significant contribution of data preprocessing and multi-objective optimization algorithms to the enhancement of forecasting accuracy.

(c) It is worth noting that when different decomposition approaches and optimization algorithms are applied to integrate the hybrid model, the superiority of forecasting performance of the proposed model is unquestionable.

Table 6

The improvement percentages among the proposed model and comparison models.

Metrics	Models	ZZ	JN	PY	QD	BD	SJZ	SZ	GZ	AV
P_{MAPE}	**ENN**	50.60	52.13	54.55	50.00	15.32	53.69	53.43	57.79	**48.84**
	LSSVM	79.29	78.07	73.48	67.09	62.32	74.97	52.44	66.13	**69.22**
	DP*	67.88	71.18	66.94	52.75	41.11	67.48	43.43	54.53	**58.16**
	MM*	43.47	72.05	46.62	47.00	10.12	70.66	34.69	38.76	**45.42**
	SSAMM*	34.97	43.63	42.67	39.77	3.78	49.00	27.97	34.38	**34.52**
	EMDMM*	29.28	68.68	26.91	19.11	4.25	37.41	32.33	5.08	**27.88**
	DPCS*	31.11	46.75	48.85	11.71	6.39	56.18	20.57	46.60	**33.52**
	DPFA*	37.58	45.12	48.09	9.91	25.63	56.55	31.08	43.60	**37.20**
P_{MAE}	**ENN**	48.71	52.53	57.72	50.56	12.05	53.38	46.80	38.44	**45.02**
	LSSVM	86.37	80.59	78.25	71.84	68.56	78.88	47.83	55.28	**70.95**
	DP*	78.30	75.51	74.11	60.27	46.88	73.11	33.74	44.63	**60.82**
	MM*	47.65	70.16	51.79	46.87	8.10	74.68	37.21	41.19	**47.21**
	SSAMM*	35.63	40.61	48.35	39.86	1.92	48.85	31.21	36.46	**35.36**
	EMDMM*	29.91	66.81	29.71	1.14	18.48	33.75	39.33	14.13	**29.16**
	DPCS*	49.50	55.59	61.72	18.72	6.84	64.41	14.96	39.07	**38.85**
	DPFA*	53.46	54.34	61.26	18.72	29.46	65.06	34.15	39.54	**44.50**
P_{MSE}	**ENN**	74.56	66.35	83.56	77.35	14.83	65.90	73.02	54.97	**63.82**
	LSSVM	98.64	93.75	95.64	95.49	92.34	94.17	74.35	88.48	**91.61**
	DP*	97.30	92.26	94.87	92.45	78.05	92.30	58.32	80.95	**85.81**
	MM*	75.49	90.72	82.94	73.26	14.04	92.36	65.72	79.99	**71.82**
	SSAMM*	63.22	44.87	80.61	66.51	3.92	58.92	58.16	75.17	**56.42**
	EMDMM*	57.95	90.91	55.12	6.77	11.15	38.39	67.70	40.98	**46.12**
	DPCS*	89.16	86.08	92.88	67.80	17.34	89.89	24.04	75.71	**67.86**
	DPFA*	91.12	85.90	92.86	70.34	59.54	90.30	59.02	78.28	**78.42**

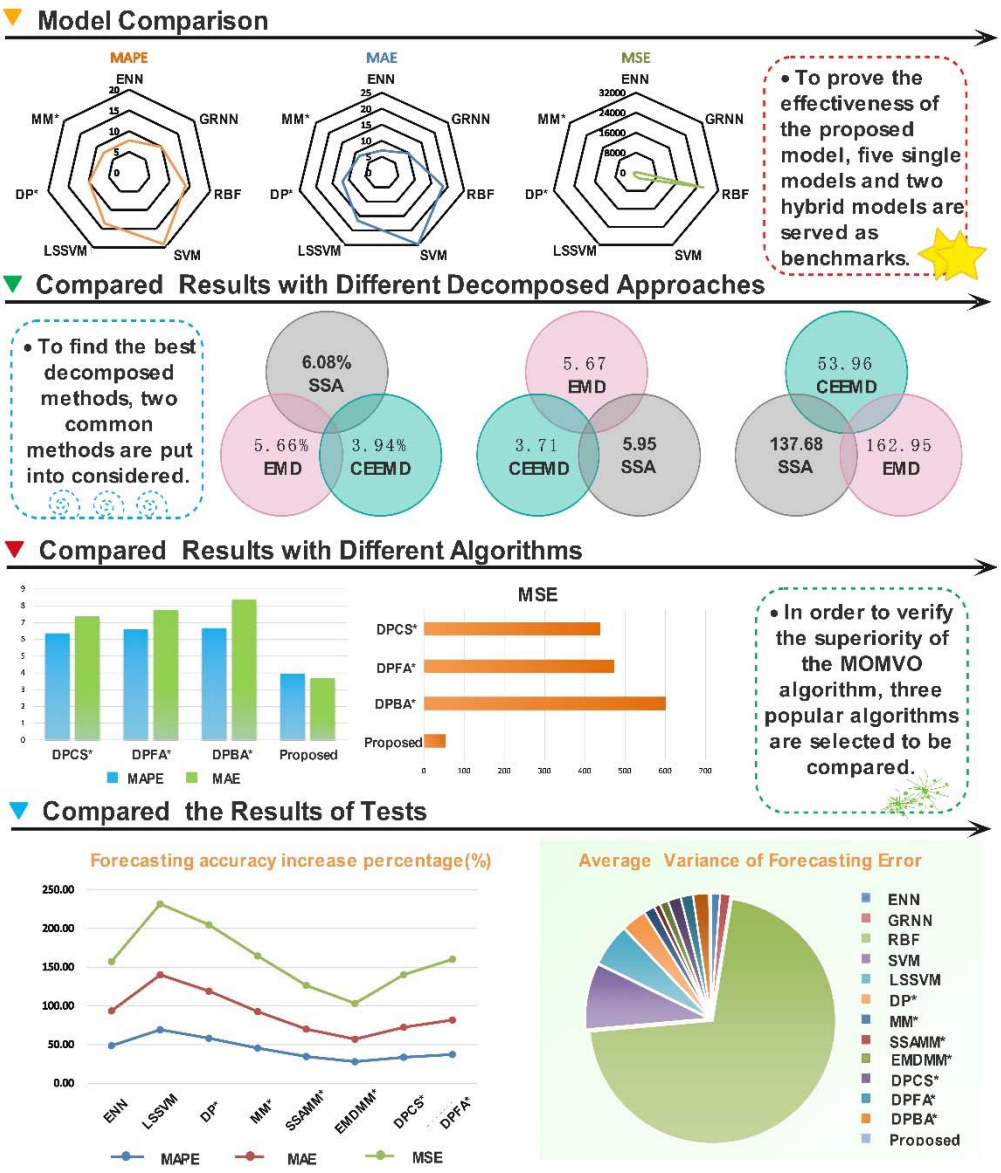

Fig.5 Comparison results of the hybrid forecasting framework

6.2 Superiority of components

The proposed hybrid framework integrates data preprocessing, multi-objective optimization, and the resulting model is applied to AQI series forecast. Each component in the hybrid forecasting framework significantly contributes to enhancing accuracy. Indeed, in this section the superiority of each component is discussed in detail. Furthermore, based on the results in Table 7 and Fig. 5, the detailed comparisons are as follows:

Table 7

The average values of metrics

Models	MAPE	MAE	MSE
ENN	7.80	6.98	171.52
GRNN	9.96	9.79	313.36

RBF	13.93	18.60	26204.86
SVM	19.13	24.80	2430.85
LSSVM	13.58	16.55	1279.78
DP*	9.89	11.99	810.11
MM*	7.78	8.48	436.40
SSAMM*	6.08	5.95	137.68
EMDMM*	5.66	5.67	162.95
DPCS*	6.37	7.36	438.60
DPFA*	6.59	7.77	473.07
DPBA*	6.67	8.39	601.78
Proposed	3.94	3.71	53.96

(a) For the first comparison, ENN, GRNN, RBF, SVM, and LSSVM were compared with the proposed model. The experimental results indicate that the hybrid model performs better than single models in forecasting AQI series.

(b) In comparison II, DP* and LSSVM had MAPE of 9.89% and 13.58%, respectively, which indicates the contribution of data preprocessing to the enhancement of accuracy. Similarly, MM* with a MAPE of 7.78% outperformed LSSVM, which demonstrates that multi-objective algorithms play a significant role in hybrid models for forecasting AQI.

(c) Comparison III demonstrated the superiority of the developed feature selection model compared with SSA and EMD. The proposed model exhibited a reduction of 2.14% and 1.72% in MAPE compared with SSAMM* and EMDMM*, respectively, which shows that the integrated CEEMD and BCCSA significantly outperformed SSA and EMD. It should be noted that owing to the randomness of the AQI series and the multiple selectivity of the model input structure, eliminating noise in the series and selecting the optimal input variables are necessary in constructing the hybrid forecasting model.

(d) Comparison IV was carried out to confirm the high quality of the multi-objective algorithm MOMVO by comparing the proposed model with DPCS*, DPFA*, and DPBA*. Based on the above analysis as well as the results in Table 7, it can be concluded that MOMVO can markedly outperform the other single-objective algorithms. The evidence supported that conclusion drawn from the reduction in the values of MAPE compared with DPCS*, DPFA*, and DPBA*, which was 2.43%, 2.65%, and 2.73%, respectively.

6.3 Performance of the MOMVO algorithm

To evaluate the effectiveness of MOMVO, four well-known testing problems described in Appendix A were considered. Four reasonable metrics were also applied. For quantifying convergence, the generational distance (GD) [63] and the inverted generational distance (IGD) [64, 67] were selected. Accordingly, a smaller value implies better performance. Furthermore, spacing (SP) [60, 65] and maximum spread (M) [66] were employed to quantify the distribution of the obtained Pareto optimal solutions. Similarly, lower values imply better coverage. The equations of these performance metrics are as follows:

$$GD = \frac{\sqrt{\sum_{i=1}^{no} d_i^2}}{no} \tag{26}$$

where no represents the number of obtained Pareto optimal solutions and d_i indicates the Euclidean distance between the *i-th* obtained Pareto optimal solution and the closest true Pareto optimal solution.

$$IGD = \frac{\sqrt{\sum_{i=1}^{nt} \left(d_i'\right)^2}}{nt} \tag{27}$$

where nt represents the number of true Pareto optimal solutions and d_i' indicates the Euclidean distance between the *i-th* true Pareto optimal solution and the closest obtained Pareto optimal solution.

$$SP = \sqrt{\frac{1}{no-1} \sum_{i=1}^{no} \left(\bar{d} - d_i\right)^2} \tag{28}$$

where $\bar{d}$ represents the average of all d_i.

$$M = \sqrt{\sum_{i=1}^{o} \max\left(d\left(a_i, b_i\right)\right)} \tag{29}$$

where *o* indicates the number of objectives, and a_i and b_i represent the maximum value and minimum value in the *i-th* objective, respectively.

For every test function, each algorithm was repeatedly simulated 50 times. 100 iterations, 40 search agents, and an archive size of 100 were used in every experiment. The final simulation results are shown in Table 8. The values of GD and IGD closely approximated zero, which demonstrated the excellent convergence properties of MOMVO; the values of SP and MS were also small, which proved that the distribution of the obtained Pareto optimal solutions is reasonable. The corresponding Pareto fronts obtained by MOMVO are shown in Fig. 6, in which it can be observed that the Pareto front obtained by MOMVO is close to the true Pareto front. According to Mirjalili [59], MOMVO is superior to traditional multi-objective optimization algorithms for various problems.

Table 8

Assessment results of MOMVO.

	GD				IGD			
	Min	**Max**	**AV**	**Std**	**Min**	**Max**	**AV**	**Std**
ADT1	0.0005	0.2466	0.0143	0.0397	0.0016	0.0033	0.0022	0.0004
ADT2	0.0004	0.0688	0.0076	0.0173	0.0015	0.0037	0.0022	0.0004
ADT3	0.0016	0.0785	0.0082	0.0182	0.0238	0.0259	0.0247	0.0004
ADT5	0.0012	0.0501	0.0073	0.0104	0.0101	0.0148	0.0118	0.0010
	SP				**MS**			
	Min	**Max**	**AV**	**Std**	**Min**	**Max**	**AV**	**Std**
ADT1	0.0072	0.8179	0.0996	0.2117	0.9338	6.5083	1.6603	1.5756
ADT2	0.0100	0.6877	0.0844	0.1696	0.9322	5.5875	1.4556	1.1591

ADT3	0.0110	0.7639	0.0585	0.1539	0.7200	3.7960	1.0853	0.6028
ADT5	0.0039	0.5008	0.0596	0.1012	0.7271	4.3023	1.3278	0.6888

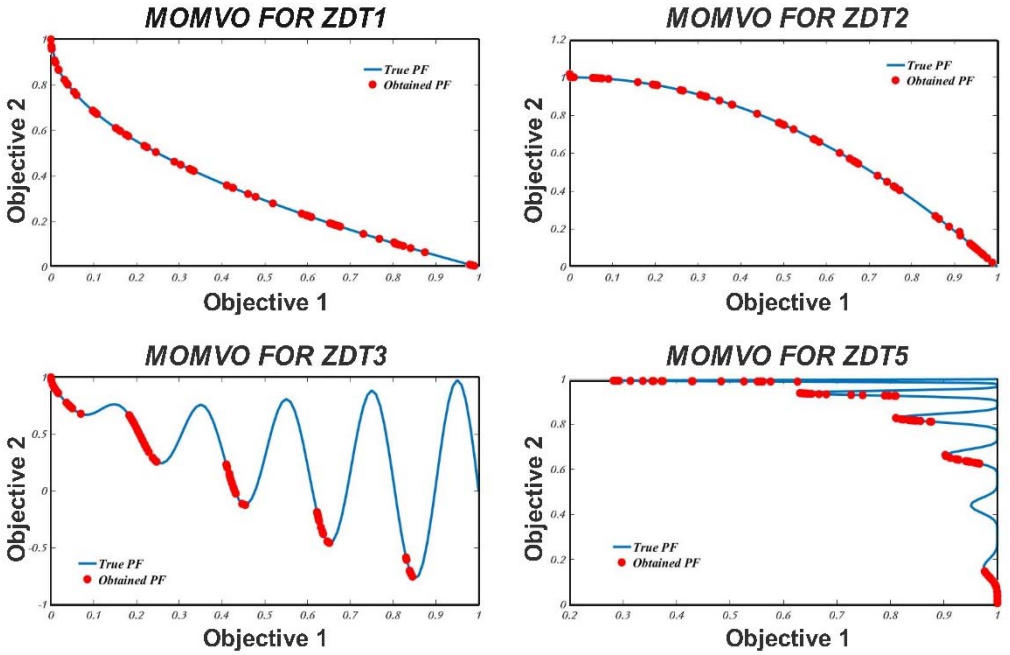

Fig.6 Test of MOMVO algorithm

6.4 Stability of the proposed model

AQI is the most direct indicator of changes in air quality. As these changes may be dramatic, a highly stable forecasting system for AQI can provide more accurate and reliable early warning information. The proposed analysis–forecast system based on the multi-objective optimization algorithm MOMVO can not only improve forecasting accuracy but also ensure forecasting stability. The above analysis has proved the forecasting performance of the system, and to evaluate its stability, a test based on **Eq. 22** was conducted. It is well known that the forecasting error is one of the most important indicators for evaluating the forecasting performance of the model. Additionally, the variance of the forecasting error can indicate the stability of the forecasting performance. Smaller variance indicates greater stability [45]. Accordingly, in this section, the variance of the forecasting errors is used to evaluate the stability of the hybrid forecasting framework and other comparison models. The results of the test are listed in Table 9. It can be seen that the variance of the proposed forecasting model is significantly smaller than that of the benchmarks, which implies that the proposed model is more stable.

Table 9
The variance of the forecasting errors.

Model	ZZ	BD	GZ	SJZ	SZ	JN	PY	QD	AV
ENN	0.0089	0.0038	0.0313	0.0071	0.0033	0.0038	0.0113	0.0027	0.0090

GRNN	0.0190	0.0073	0.0156	0.0124	0.0037	0.0063	0.0139	0.0048	0.0104
RBF	0.0850	0.0831	0.0312	2.6156	0.0333	1.1321	0.3491	0.0664	0.5495
SVM	0.0883	0.0316	0.2921	0.0428	0.0025	0.0261	0.0412	0.0125	0.0671
LSSVM	0.0733	0.0236	0.1577	0.0289	0.0032	0.0201	0.0297	0.0085	0.0431
DP*	0.0429	0.0122	0.0710	0.0250	0.0032	0.0173	0.0254	0.0049	0.0252
MM*	0.0075	0.0039	0.0192	0.0233	0.0009	0.0216	0.0106	0.0025	0.0112
SSAMM*	0.0061	0.0036	0.0179	0.0064	0.0007	0.0023	0.0090	0.0020	0.0060
EMDMM*	0.0044	0.0050	0.0025	0.0055	0.0007	0.0449	0.0049	0.0006	0.0086
DPCS*	0.0077	0.0045	0.0404	0.0195	0.0009	0.0098	0.0194	0.0013	0.0129
DPFA*	0.0105	0.0042	0.0335	0.0189	0.0007	0.0097	0.0194	0.0013	0.0123
DPBA*	0.0297	0.0090	0.0377	0.0202	0.0008	0.0096	0.0189	0.0022	0.0160
Proposed	**0.0018**	**0.0030**	**0.0028**	**0.0020**	**0.0003**	**0.0012**	**0.0025**	**0.0006**	**0.0018**

6.5 Application of the proposed analysis–forecast system

The proposed analysis-forecast system has not only the advantages of high forecasting accuracy and strong stability but also several potential applications, including analyzing the characteristics of air quality changes, warning the public before the occurrence of hazardous air pollutants, and coordinating daily life activities. Additionally, it consists of three modules: complexity analysis module, data preprocessing module, and optimize–forecast module, all of which have their own practical applications.

1) In fact, the atmosphere is an intricate dynamic system that contains a large amount of information. Therefore, the complexity analysis module can provide detailed and reliable analysis of the air quality index series, which clearly reflects changes in the complexity of the AQI and indirectly reflects changes in air quality; furthermore, it is helpful in data interpretation for decision making related to pollution mitigation measures and air quality management.
2) Based on the complexity analysis of AQI, the data preprocessing module can eliminate the outliers and the negative noise in the AQI series, and sensibly select the input variable structure for the prediction model. This module solves the problems of randomness, non-stationarity, and irregularity of series in the analysis and prediction of air quality indicators.
3) The optimize–forecast module provides accurate and reliable warning information by forecasting future AQI values. This may assist environmental policy makers in taking effective protection measures before the occurrence of hazardous air pollutants as well as provide useful daily-life guidance. It is worth noting that it also provides intuitive guidance for residents: when the concentration of pollutants exceeds a threshold, air quality supervision departments can promptly take relevant prevention and control measures, and thus residents can adjust their travel plans in time.

7. Conclusions

Forecasting the air pollution index is important in pollution control. Owning to the uncertainties and randomness inherent in AQI series, it is still a challenging task to

reasonably forecast and model the air quality index. Previous studies mainly focused on enhancing either accuracy or stability, and few studies successfully addressed the two aspects simultaneously. Therefore, developing an efficient and reliable analysis–forecast system is still desirable. In this study, a novel analysis–forecast system was proposed that overcomes the shortcomings mentioned above and is intended to be a powerful and efficient technique for air quality analysis and monitoring. In the developed system, complexity analysis based on sample entropy was proposed as the first step in system analysis to mine the information in AQI. Then, to further improve modeling accuracy and rationality of modeling, a Hampel filter was employed to eliminate outliers, and a novel feature selection approach based on CEEMD and BCCSA was developed to determine the optimal input variable structure. Finally, a modified LSSVM based on the MOMVO algorithm was developed to simultaneously obtain high accuracy and strong stability, and two tests verified its efficiency.

The developed analysis–forecast system was evaluated on hourly AQI series from eight cities, and several performance metrics (i.e., MAPE, MAE, MSE, P_{MAPE}, P_{MAE}, P_{MSE}) were calculated. The experimental results indicated that the proposed hybrid forecasting model is superior to comparison models with the smallest MAPE of 3.72%, 2.70%, 4.89%, 3.09%, 5.86%, 3.58%, 2.24%, and 5.42%, respectively, in the eight datasets. Moreover, the variance of the forecasting error of the proposed model was 0.0018, 0.0030, 0.0028, 0.0020, 0.0003, 0.0012, 0.0025, and 0.0006, respectively, in the eight datasets, indicating its strong forecasting performance. Overall, the proposed analysis–forecast system exhibits outstanding performance in analyzing and monitoring air quality. Specifically, it can not only deeply mine the information in the air quality index but also approximate the actual values with high accuracy and stability. Thus, the developed system can be a powerful tool for decision-makers in monitoring and forecasting air quality.

Conflicts of interest

The authors declare that there is no conflict of interest with regard to the publication of this paper.

Acknowledgements

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

Appendix A

A.1. Definition of AQI

AQI is a dimensionless index that quantitatively describes air quality. Larger values indicate more serious air pollution and hence greater health hazard. The major pollutants involved in air quality assessment are $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO. The equation for AQI is as follows:

$$AQI_{po} = \frac{AQI_{Hi} - AQI_{Lo}}{BP_{Hi} - BP_{Lo}}\left(C_{po} - BP_{Lo}\right) + AQI_{Lo}$$

where AQI_{po} represents the value of AQI, C_{po} is the concentration of pollutant *po*, and BP_{Hi} and BP_{Lo} indicate the high breakpoint value and low breakpoint value for C_{po}, respectively. AQI_{Hi} and AQI_{Lo} are AQI indexes corresponding to BP_{Hi} and BP_{Lo}, respectively.

Appendix B

B.1. Test functions of MOMVO

The ZDT test function set was proposed by K. Deeb [68, 69]. Owing to the complexity of the functions involved, it is difficult to solve. Therefore, it is regarded as a static multi-objective optimization problem set, which is used for testing the performance of related algorithms. These functions have the following formal definitions:

$$\begin{cases} \min F\left(X\right) = \left(f_1\left(x_1\right), f_2\left(X\right)\right) \\ s.t \quad f_2\left(X\right) = g(X)h\left(f_1\left(x_1\right), g\left(X\right)\right) \end{cases}$$

where $X = \left(x_1, x_2, \ldots, x_m\right)$.

ZDT1:

$$\begin{cases} f_1\left(x_1\right) = x_1, \\ g\left(X\right) = 1 + 9 \cdot \sum_{i=2}^{m} x_i / \left(m-1\right), \\ h\left(f_1, g\right) = 1 - \sqrt{f_1 / g} \end{cases}$$

where $1 \le m \le 30$, $x_i \in [0,1]$, $i = 1, 2, \ldots, m$. The non-inferior solution set for this function is convex in the objective function space.

ZDT2:

$$\begin{cases} f_1(x_1) = x_1, \\ g(X) = 1 + 9 \cdot \sum_{i=2}^{m} x_i / (m-1), \\ h(f_1, g) = 1 - (f_1 / g)^2 \end{cases}$$

where $1 \le m \le 30$, $x_i \in [0,1]$, $i = 1, 2, \ldots, m$. The non-inferior solution set for this function is non-convex in the objective function space.

ZDT3:

$$\begin{cases} f_1(x_1) = x_1, \\ g(X) = 1 + 9 \cdot \sum_{i=2}^{m} x_i / (m-1), \\ h(f_1, g) = 1 - \sqrt{f_1 / g} - (f_1 / g) \cdot \sin(10 \cdot \pi f_1) \end{cases}$$

where $1 \le m \le 30$, $x_i \in [0,1]$, $i = 1, 2, \ldots, m$. This function contains 21^9 local Pareto frontiers; thus, it is suitable for evaluating the ability of the algorithm to test multimodal functions.

ZDT5:

$$\begin{cases} f_1(x_1) = 1 - \exp(-4x_1) \sin^6(6\pi x_1), \\ g(X) = 1 + 9 \cdot \left(\sum_{i=2}^{m} x_i / (m-1) \right)^{0.25} \\ h(f_1, g) = 1 - (f_1 / g)^2 \end{cases}$$

where $1 \le m \le 10$, $x_i \in [0,1]$, $i = 1, 2, \ldots, m$. The non-inferior solution set for this function is non-convex and non-uniformly distributed in the objective function space.

Reference

[1] Kurt, A., Oktay, A.B., 2010. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. Expert Syst. Appl. 37, 7986–7992. https://doi.org/10.1016/j.eswa.2010.05.093

[2] Fan, Y., Hou, L., Yan, K.X., 2018. On the density estimation of air pollution in Beijing. Econ. Lett. 163, 110–113. https://doi.org/10.1016/j.econlet.2017.12.020

[3] Qin, S., Liu, F., Wang, C., Song, Y., Qu, J., 2015. Spatial-temporal analysis and projection of extreme particulate matter (PM10and PM2.5) levels using association rules: A case study of the Jing-Jin-Ji region, China. Atmos. Environ. 120, 339–350. https://doi.org/10.1016/j.atmosenv.2015.09.006

[4] Güler Dincer, N., Akkuş, Ö., 2018. A new fuzzy time series model based on robust clustering for forecasting of air pollution. Ecol. Inform. 43, 157–164. https://doi.org/10.1016/j.ecoinf.2017.12.001

[5] Kiesewetter, G., Schoepp, W., Heyes, C., Amann, M., 2015. Modelling PM2.5impact indicators in Europe: Health effects and legal compliance. Environ. Model. Softw. 74, 201–211. https://doi.org/10.1016/j.envsoft.2015.02.022

[6] Reisen, V.A., Sarnaglia, A.J.Q., Reis, N.C., Lévy-Leduc, C., Santos, J.M., 2014. Modeling and forecasting daily average PM10concentrations by a seasonal long-memory model with volatility. Environ. Model. Softw. 51, 286–295. https://doi.org/10.1016/j.envsoft.2013.09.027

[7] Vedrenne, M., Borge, R., Lumbreras, J., Rodríguez, M.E., 2014. Advancements in the design and validation of an air pollution integrated assessment model for Spain. Environ. Model. Softw. 57, 177–191. https://doi.org/10.1016/j.envsoft.2014.03.002

[8] Zheng, S., Cao, C.X., Singh, R.P., 2014. Comparison of ground based indices (API and AQI) with satellite based aerosol products. Sci. Total Environ. 488-489, 398–412. https://doi.org/10.1016/j.scitotenv.2013.12.074

[9] Ribeiro, M.C., Pinho, P., Branquinho, C., Llop, E., Pereira, M.J., 2016. Geostatistical uncertainty of assessing air quality using high-spatial-resolution lichen data: A health study in the urban area of Sines, Portugal. Sci. Total Environ. 562, 740–750. https://doi.org/10.1016/j.scitotenv.2016.04.081

[10] 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

[11] Li, R., Jin, Y., 2018. The early-warning system based on hybrid optimization algorithm and fuzzy synthetic evaluation model. Inf. Sci. (Ny). 435, 296–319. https://doi.org/10.1016/j.ins.2017.12.040

[12] Wang, J., Zhang, X., Guo, Z., Lu, H., 2017. Developing an early-warning system for air quality prediction and assessment of cities in China. Expert Syst. Appl. 84, 102–116. https://doi.org/10.1016/j.eswa.2017.04.059

[13] García Nieto, P.J., Sánchez Lasheras, F., García-Gonzalo, E., de Cos Juez, F.J., 2018. PM 10 concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM, MLP, VARMA and ARIMA: A case study. Sci. Total Environ. 621, 753–761.

https://doi.org/10.1016/j.scitotenv.2017.11.291
[14] Afzali, A., Rashid, M., Afzali, M., Younesi, V., 2017. Prediction of air pollutants concentrations from multiple sources using AERMOD coupled with WRF prognostic model. J. Clean. Prod. 166, 1216–1225. https://doi.org/10.1016/j.jclepro.2017.07.196
[15] Li, C., Zhu, Z., 2018. Research and application of a novel hybrid air quality early-warning system: A case study in China. Sci. Total Environ. 626, 1421–1438. https://doi.org/10.1016/j.scitotenv.2018.01.195
[16] Xu, Y., Yang, W., Wang, J., 2017. Air quality early-warning system for cities in China. Atmos. Environ. 148, 239–257. https://doi.org/10.1016/j.atmosenv.2016.10.046
[17] Konovalov, I.B., Beekmann, M., Meleux, F., Dutot, A., Foret, G., 2009. Combining deterministic and statistical approaches for PM10 forecasting in Europe. Atmos. Environ. 43, 6425–6434. https://doi.org/10.1016/j.atmosenv.2009.06.039
[18] Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., Wang, J., 2015. Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation. Atmos. Environ. 107, 118–128. https://doi.org/10.1016/j.atmosenv.2015.02.030
[19] Sun, W., Zhang, H., Palazoglu, A., Singh, A., Zhang, W., Liu, S., 2013. Prediction of 24-hour-average PM2.5 concentrations using a hidden Markov model with different emission distributions in Northern California. Sci. Total Environ. 443, 93–103. https://doi.org/10.1016/j.scitotenv.2012.10.070
[20] Stern, R., Builtjes, P., Schaap, M., Timmermans, R., Vautard, R., Hodzic, A., Memmesheimer, M., Feldmann, H., Renner, E., Wolke, R., Kerschbaumer, A., 2008. A model inter-comparison study focussing on episodes with elevated PM10 concentrations. Atmos. Environ. 42, 4567–4588. https://doi.org/10.1016/j.atmosenv.2008.01.068
[21] Slottje, D., Nieswiadomy, M., Redfearn, M., 2001. Economic inequality and the environment. Environ. Model. Softw. https://doi.org/10.1016/S1364-8152(00)00081-5
[22] Zafra, C., Ángel, Y., Torres, E., 2016. ARIMA analysis of the effect of land surface coverage on PM<inf>10</inf> concentrations in a high-altitude megacity. Atmos. Pollut. Res. https://doi.org/10.1016/j.apr.2017.01.002
[23] Wang, P., Zhang, H., Qin, Z., Zhang, G., 2017. A novel hybrid-Garch model based on ARIMA and SVM for PM2.5concentrations forecasting. Atmos. Pollut. Res. 8, 850–860. https://doi.org/10.1016/j.apr.2017.01.003
[24] Zhang, H., Zhang, W., Palazoglu, A., Sun, W., 2012. Prediction of ozone levels using a Hidden Markov Model (HMM) with Gamma distribution. Atmos. Environ. 62, 64–73. https://doi.org/10.1016/j.atmosenv.2012.08.008
[25] Qin, M., Li, Z., Du, Z., 2017. Red tide time series forecasting by combining ARIMA and deep belief network. Knowledge-Based Syst. 125, 39–52. https://doi.org/10.1016/j.knosys.2017.03.027
[26] Xu, Y., Du, P., Wang, J., 2017. 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

[27] Li, H., Wang, J., Lu, H., Guo, Z., 2018. Research and application of a combined model based on variable weight for short term wind speed forecasting. Renew. Energy 116, 669–684. https://doi.org/10.1016/j.renene.2017.09.089

[28] Elangasinghe, M.A., Singhal, N., Dirks, K.N., Salmond, J.A., 2014. Development of an ANN–based air pollution forecasting system with explicit knowledge through sensitivity analysis. Atmos. Pollut. Res. 5, 696–708. https://doi.org/10.5094/APR.2014.079

[29] Bai, Y., Li, Y., Wang, X., Xie, J., Li, C., 2016. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. Atmos. Pollut. Res. 7, 557–566. https://doi.org/10.1016/j.apr.2016.01.004

[30] Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., Chi, T., 2017. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. Environ. Pollut. 231, 997–1004. https://doi.org/10.1016/j.envpol.2017.08.114

[31] Wang, J., Niu, T., Wang, R., 2017. Research and application of an air quality early warning system based on a modified least squares support vector machine and a cloud model. Int. J. Environ. Res. Public Health 14. https://doi.org/10.3390/ijerph14030249

[32] He, H., Lu, W.-Z., Xue, Y., 2014. Prediction of particulate matter at street level using artificial neural networks coupling with chaotic particle swarm optimization algorithm. Build. Environ. 78, 111–117. https://doi.org/10.1016/j.buildenv.2014.04.011

[33] Wang, J., Yang, W., Du, P., Li, Y., 2018. Research and application of a hybrid forecasting framework based on multi-objective optimization for electrical power system. Energy 148, 59–78. https://doi.org/10.1016/j.energy.2018.01.112

[34] Yang, Y., Che, J., Li, Y., Zhao, Y., Zhu, S., 2016. An incremental electric load forecasting model based on support vector regression. Energy 113, 796–808. https://doi.org/10.1016/j.energy.2016.07.092

[35] Wang, J., Hu, J., 2015. A robust combination approach for short-term wind speed forecasting and analysis - Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. Energy 93, 41–56. https://doi.org/10.1016/j.energy.2015.08.045

[36] Niu, M., Wang, Y., Sun, S., Li, Y., 2016. A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM2.5concentration forecasting. Atmos. Environ. 134, 168–180. https://doi.org/10.1016/j.atmosenv.2016.03.056

[37] Mafarja, M., Mirjalili, S., 2018. Whale optimization approaches for wrapper feature selection. Appl. Soft Comput. J. 62, 441–453. https://doi.org/10.1016/j.asoc.2017.11.006

[38] Pham, T.X., Siarry, P., Oulhadj, H., 2018. Integrating fuzzy entropy clustering with an improved PSO for MRI brain image segmentation. Appl. Soft Comput. J. 65, 230–242. https://doi.org/10.1016/j.asoc.2018.01.003
[39] Yelghi, A., Köse, C., 2018. A modified firefly algorithm for global minimum optimization. Appl. Soft Comput. J. 62, 29–44. https://doi.org/10.1016/j.asoc.2017.10.032
[40] Liao, G.C., 2014. Hybrid improved differential evolution and wavelet neural network with load forecasting problem of air conditioning. Int. J. Electr. Power Energy Syst. 61, 673–682. https://doi.org/10.1016/j.ijepes.2014.04.014
[41] Al-Betar, M.A., Awadallah, M.A., Faris, H., Yang, X.S., Tajudin Khader, A., Alomari, O.A., 2018. Bat-inspired algorithms with natural selection mechanisms for global optimization. Neurocomputing 273, 448–465. https://doi.org/10.1016/j.neucom.2017.07.039
[42] 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
[43] Ravi, V., Pradeepkumar, D., Deb, K., 2017. Financial time series prediction using hybrids of chaos theory, multi-layer perceptron and multi-objective evolutionary algorithms. Swarm Evol. Comput. 36, 136–149. https://doi.org/10.1016/j.swevo.2017.05.003
[44] Lin, W., Jin, X., Mu, Y., Jia, H., Xu, X., Yu, X., 2017. Multi-objective Optimal Hybrid Power Flow Algorithm for Integrated Community Energy System, in: Energy Procedia. pp. 2871–2878. https://doi.org/10.1016/j.egypro.2017.03.638
[45] Wang, J., Heng, J., Xiao, L., Wang, C., 2017. 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
[46] Sun, W., Wang, Y., 2018. Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. Energy Convers. Manag. 157, 1–12. https://doi.org/10.1016/j.enconman.2017.11.067
[47] Xu, Y., Zhang, M., Zhu, Q., He, Y., 2017. An improved multi-kernel RVM integrated with CEEMD for high-quality intervals prediction construction and its intelligent modeling application. Chemom. Intell. Lab. Syst. 171, 151–160. https://doi.org/10.1016/j.chemolab.2017.10.019
[48] Allen, D.P., 2009. A frequency domain Hampel filter for blind rejection of sinusoidal interference from electromyograms. J. Neurosci. Methods 177, 303–310. https://doi.org/10.1016/j.jneumeth.2008.10.019
[49] Allen, D.P., Stegemöller, E.L., Zadikoff, C., Rosenow, J.M., MacKinnon, C.D., 2010. Suppression of deep brain stimulation artifacts from the electroencephalogram by frequency-domain Hampel filtering. Clin. Neurophysiol. 121, 1227–1232. https://doi.org/10.1016/j.clinph.2010.02.156
[50] Pearson, R.K., 2002. Outliers in process modeling and identification. IEEE Trans. Control Syst. Technol. 10, 55–63. https://doi.org/10.1109/87.974338

[51] Davies, L., Gather, U., 1993. The identification of multiple outliers. J. Am. Stat. Assoc. 88, 782–792. https://doi.org/10.1080/01621459.1993.10476339
[52] Sayed, G.I., Hassanien, A.E., Azar, A.T., 2017. Feature selection via a novel chaotic crow search algorithm. Neural Comput. Appl. 1–18. https://doi.org/10.1007/s00521-017-2988-6
[53] Yuan, X., Tan, Q., Lei, X., Yuan, Y., Wu, X., 2017. Wind power prediction using hybrid autoregressive fractionally integrated moving average and least square support vector machine. Energy 129, 122–137. https://doi.org/10.1016/j.energy.2017.04.094
[54] Lv, Y., Yang, T., Liu, J., 2015. An adaptive least squares support vector machine model with a novel update for NOx emission prediction. Chemom. Intell. Lab. Syst. 145, 103–113. https://doi.org/10.1016/j.chemolab.2015.04.006
[55] Wu, Y.H., Shen, H., 2018. Grey-related least squares support vector machine optimization model and its application in predicting natural gas consumption demand. J. Comput. Appl. Math. 338, 212–220. https://doi.org/10.1016/j.cam.2018.01.033
[56] Sun, W., Sun, J., 2017. Daily PM2.5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm. J. Environ. Manage. 188, 144–152. https://doi.org/10.1016/j.jenvman.2016.12.011
[57] WU, Z., HUANG, N.E., 2009. Ensemble Empirical Mode Decomposition: a Noise-Assisted Data Analysis Method . Adv. Adapt. Data Anal 2009, 1–41. https://doi.org/10.1142/S1793536909000047
[58] Fathy, A., Rezk, H., 2018. Multi-verse optimizer for identifying the optimal parameters of PEMFC model. Energy 143, 634–644. https://doi.org/10.1016/j.energy.2017.11.014
[59] Mirjalili, S., Jangir, P., Mirjalili, S.Z., Saremi, S., Trivedi, I.N., 2017. Optimization of problems with multiple objectives using the multi-verse optimization algorithm. Knowledge-Based Syst. 134, 50–71. https://doi.org/10.1016/j.knosys.2017.07.018
[60] Coello, C. a C., Pulido, G.T., Lechuga, M.S., 2004. Handling multiple objectives with particle swarm optimization. Evol. Comput. IEEE Trans. 8, 256–279. https://doi.org/10.1109/TEVC.2004.826067
[61] APADerrac, J., García, S., Molina, D., & Herrera, F. 2011. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm Evol Comput 1(1):3–18. https://doi.org/10.1016/j.swevo.2011.02.002
[62] https://data.epmap.org
[63] Mirjalili, S., Saremi, S., Mirjalili, S.M., Coelho, L.D.S., 2016. Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization. Expert Syst. Appl. 47, 106–119. https://doi.org/10.1016/j.eswa.2015.10.039
[64] Hong, Y.Y., Lin, J.K., 2013. Interactive multi-objective active power scheduling considering uncertain renewable energies using adaptive chaos clonal evolutionary programming. Energy 53, 212–220. https://doi.org/10.1016/j.energy.2013.02.070
[65] Schott, J.R., 1995. Fault Tolerant Design Using Single and Multicriteria Genetic Algorithm Optimization. Massachusetts Inst. Technol. Boston, MA.

[66] Zitzler, E., 1999. Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications. TIK-Schriftenreihe 30, 1–122. https://doi.org/citeulike-article-id:4597043

[67] Wang, J., Du, P., Niu, T., Yang, W., 2017. 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

[68] Zitzler, E., Deb, K., Thiele, L., 2000. Comparison of multiobjective evolutionary algorithms: empirical results. Evol. Comput. 8, 173–195. https://doi.org/10.1162/106365600568202

[69] Deb, K., 1999. Multi-objective Genetic Algorithms: Problem Difficulties and Construction of Test Problems. Evol. Comput. 7, 205–230. https://doi.org/10.1162/evco.1999.7.3.205