frontiers in Environmental Science

ORIGINAL RESEARCH
published: 15 December 2021
doi: 10.3389/fenvs.2021.761287

Hybrid Data Mining Forecasting System Based on Multi-Objective Optimization and Selection Model for Air pollutants

Yanwen Huang[1], *Yuanchang Deng*[1]*, *Chen Wang*[1]* *and Tonglin Fu*[2,3,4]

[1]School of Intelligent System Engineering, Sun Yat-Sen University, Shenzhen, China, [2]School of Mathematics and Statistics LongDong University, Qingyang, China, [3]Shapotou Desert Research and Experiment Station, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, China, [4]University of Chinese Academy of Sciences, Beijing, China

The air quality index (AQI) indicates the short-term air quality situation and changing trend of the city, which includes six air pollutants: $PM_{2.5}$, PM_{10}, CO, NO_2, SO_2 and O_3. Due to the diversity of pollutants and the fluctuation of single pollutant time series, it is a challenging task to find out the main pollutants and establish an accurate forecasting system in a city. Previous studies primarily focused on enhancing either forecasting accuracy or stability and failed to analyze different air pollutants at length, leading to unsatisfactory results. In this study, a model selection forecasting system is proposed that consists of data mining, data analysis, model selection, and multi-objective optimized modules and effectively solves the problems of air pollutants monitoring. The proposed system employed fuzzy C-means cluster algorithm to analyze 13 original AQI series, and fuzzy comprehensive evaluation is used to find out the main air pollutants in each city. And then multiple artificial neural networks are used to forecast the main air pollutants for each category and find the optimal models. Finally, the modified multi-objective optimization algorithm is used to optimize the parameters of optimal models and model selection to obtain final forecasting values from optimal hybrid models. The experiment results of datasets from 13 cities in the Beijing–Tianjin–Hebei Urban Agglomeration demonstrated that the proposed system can simultaneously obtain efficient and reliable data for air quality monitoring.

Keywords: air quality index, data analysis, data mining, artificial neural networks, model selection

OPEN ACCESS

Edited by:
Yan Hao,
Shandong Normal University, China

Reviewed by:
Qingli Dong,
Dalian University of Technology, China
Xiaofei Luo,
University of Aizu, Japan

***Correspondence:**
Yuanchang Deng
dengych@mail.sysu.edu.cn
Chen Wang
wangch339@mail.sysu.edu.cn

Specialty section:
This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science

Received: *19 August 2021*
Accepted: *04 October 2021*
Published: *15 December 2021*

Citation:
Huang Y, Deng Y, Wang C and Fu T (2021) Hybrid Data Mining Forecasting System Based on Multi-Objective Optimization and Selection Model for Air pollutants. Front. Environ. Sci. 9:761287. doi: 10.3389/fenvs.2021.761287

INTRODUCTION

In recent years, air pollution has received increasing attention due to the negative effects, such as respiratory diseases, that it has on human health (Jiang et al., 2017). Simultaneously, air pollution is a growing environmental concern, responsible for approximately 2 million premature deaths per year worldwide (World Health Organization, 2008). A report issued by the World Health Organization (WHO) acknowledges that air pollution is one of the biggest health risks (Xu et al., 2016). Since the industrial revolution, many countries have focused on economic development while ignoring air quality, and incidents that cause harm are everywhere. In 1930, the Mas Valley event in Belgium caused nearly 60 deaths in a week. In the 1940s, the smog incident in Los Angeles caused many people to have red eyes, pharyngitis, respiratory disease deterioration, and even confusion and pulmonary edema. In 1948, the American Donora incident caused 5,911 people to become violent.

The most serious is the well-known London smog event of 1952—more than 4,000 deaths in 4 days and more than 8,000 deaths in 2 months. In addition, air pollution in China is also quite serious. The previous results in 2009 showed that the air quality index (AQI) in 107 cities of China did not meet the country's national air quality standards (NAAS) (Zhou et al., 2014). In addition, 7 of the 10 most polluted cities in the world are in China. According to the World Bank, China loses 10% of its gross domestic product each year due to air pollution. Air pollution is also associated with elevated rates of mortality, causing between 350,000 and 500,000 premature deaths each year in China (Shanshan et al., 2014). Air pollution has become the fourth leading health risk factor for China after smoking, diet, and obesity (Zhang et al., 2018). In order to reduce the losses caused by air pollution, several health and governmental institutions gather and publish data regarding what is known as AQI to inform people about the state of air pollution. For instance, the European Environment Agency (EEA) and the European Commission (EC) have launched, in 2017, an online platform that provides information about current air quality situation based on measurements from more than 2,000 air quality monitoring stations across Europe (Akyüz and Çabuk, 2009). In addition, China's environmental supervisors have also issued some plans and programs, including EIA (Environmental Influence Assessment) and Emergency Response for reducing air pollution. Since 2013, China has also begun to evaluate the quality of air through AQI values and graded the city's air quality by AQI values. AQI is an important evaluation indicator that comprehensively reflects the air pollution status related to human health. Through the use of the AQI it was possible to synthesize, in a single daily value, concentrations of major pollutants in urban areas (NO_2, O_3, CO, SO_2, $PM_{2.5}$, PM_{10}) for the entire period (Feng et al., 2015). The greater the AQI value, the more serious the air pollution. But real-time air quality monitoring can no longer meet people's needs. Like weather forecasts, people also long for air quality prediction to arrange their activities and take protective measures in advance (Hao et al., 2021).

Obviously, if we can provide early warning before the hazard occurs, based on a good air quality early-warning system, these losses might be avoided by taking effective corresponding protection measures. In order to establish an effective air warning system, observation and control of air quality is the key issue for authorities. The most significant point in any kind of air pollution control system is to be able to detect increasing (deterioration) or decreasing (improvement) trends (Hao and Tian, 2018). Unfortunately, because air quality data is obtained in limited time and space, its incompleteness and non-stationarity may result in low accuracy and poor stability of the forecasting results (Hao et al., 2019). Therefore, the prediction of AQI or other pollution indicators is a challenging task.

In recent years, many studies on air quality have focused on the prediction of atmospheric pollutant concentrations. From the angle of methodology, various quantitative prediction methods of the atmosphere pollutant concentrations can be classified into two categories, including deterministic models and empirical models (Steffens et al., 2017). The deterministic model is mainly the chemical transport model (CTM), which is based on the fundamental principles of simulating atmospheric physics and chemistry that involve transportation, emissions, and conversion processes in air pollution (Rivas et al., 2018). The forecasts are used to support flight planning by enabling the representation of important three-dimensional (3-D) atmospheric chemical structures (such as dust storm plumes, polluted air masses originated by large cities, and widespread biomass burning events) and their time evolution, which are often research targets to be detected and investigated through specific flight plans (Latif et al., 2018). Various models have been proposed to identify the interactions between various air pollutants and their emission sources (Yang and Wang, 2017). Nonetheless, due to the incomplete knowledge and understanding of the sources, dispersion and sinks of pollutants, transport processes, and atmospheric chemicals, there are some significant uncertainties in the models, resulting in air pollutant concentrations being among the most difficult to forecast accurately using CTMs (Liu et al., 2008). Therefore, CTM forecasts are less accurate than empirical air quality predictive models that are trained with local meteorological data and air quality.

A large number of empirical models include statistical models and machine learning models for the forecast of atmospheric pollutant concentrations. Common statistical models for air quality prediction include autoregressive (AR) models, moving average (MA) models, autoregressive integrated moving average (ARIMA) models, and multiple linear regression (MLR) models. For example, Zhang et al. (2018) applied the RIMA model to predict the concentration of $PM_{2.5}$ based on time series air quality data covering two warm periods and two cold periods and concludes that $PM_{2.5}$ concentration is higher in the cold period and lower in the warm period. MLR models are applied by Mehmet Akyüz et al. (Pereira et al., 2018) to forecast the concentration of individual pollutants. The study also considers the effects of contaminant concentrations and other meteorological parameters. Although Box–Jenkins Time Series (ARIMA) and MLR models have been applied to air quality forecasting in urban areas, they have limited accuracy owing to their inability to predict extreme events, and they are not applicable when performing long-term prediction and nonlinear sequence prediction.

On the contrary, artificial neural networks (ANNs) are more popular for their no-linear systems, especially when it is difficult to determine the theoretical models (Lanzafame et al., 2015). Díaz-Robles et al. (2008) combined a new hybrid model of ARIMA and ANN to improve the prediction accuracy of areas with limited air quality and meteorological data. Xiao Feng and Qi Li et al. (Feng et al., 2015) combined air mass trajectory analysis and wavelet transform and proposed that ANN predicts the daily average concentration of pollutants 2 days in advance, improves the accuracy of prediction, and is superior to other models. However, they also have certain shortcomings that may fall into local optimum or over-fitting, which may result in poor prediction.

Any model has its inevitable shortcomings, and due to the advent of the world's big data era, data mining techniques such as

decomposition methods (Güçlü et al., 2019), feature selection techniques (Pan et al., 2011), and optimization algorithms (Liu et al., 2019) combined with artificial intelligence technology are more operational. Therefore, with consideration of forecast accuracy, hybrid models which combine a new method with artificial intelligence are of great significance in air quality forecasting field (D'Allura et al., 2011). Although the construction of the combined model is usually based on actual problems to achieve the expected test objectives, there are still some problems that most of the past studies have focused on improving the prediction accuracy of the model while ignoring the stability of the model prediction. Many optimization algorithms inspired by nature including cuckoo algorithm (Urbancok et al., 2017), firefly algorithm (Bessagnet et al., 2019), bat algorithm (Liu et al., 2018), and particle swarm optimization algorithm (Kumar et al., 2019) have been developed to solve single-objective problems in recent years. However, real-world optimization problems always involve multiple objectives and so-called multi-objective optimization, which means, in this case, the solutions for a multi-objective problem, which is the main focus of the algorithm, represent the trade-offs between the objectives due to the nature of such problems (Shenfield and Rostami, 2015). The developed multi-objective optimization algorithm has been applied more and more widely in the fields of finance (Li et al., 2019) and mechanical engineering (Dhiman and Kumar, 2018). The atmosphere is a highly complex dynamic system. The air quality data sequence usually has characteristics such as non-stationarity and nonlinearity; thus, the multi-objective optimization algorithm is a suitable choice.

Furthermore, air quality assessment algorithms are developed to assess air quality and protect human health from air pollution and play a vital role in air quality warning systems. The early-warning system can increase the environmental consciousness of society and protect the public against hazardous air quality. It can also aid the relevant departments to better control air pollution and avoid negative social, economic, and environmental impacts. According to the aforementioned analysis, developing a novel and robust air quality early-warning system has become highly desirable for society. Therefore, a variety of models are employed in air quality assessment, including mobile monitoring (Li et al., 2018), CFD-RANS simulation (Lauriks et al., 2020), principal component regression (PCR), sensitivity analysis (Kim et al., 2018), Bayesian models (Han et al., 2021), support vector machines (SVM) (Leong et al., 2019), ANNs (Davood et al., 2021), and fuzzy techniques (Dass et al., 2021). However, although the air quality warning system has important practical significance to the public in other fields, China's research in this field is still relatively small.

Looking back at the previous literature on air quality forecasting research, the shortcomings of the traditional air quality forecasting models are summarized as follows: 1) the large amount of information required by the CTM model leads to uncertainty in the forecasting. 2) The single statistical models with low forecasting accuracy cannot meet the requirements of air quality forecasting. 3) In the past, many air quality studies focused on eliminating the effects of noise on data processing and less on the feature extraction of data. 4) It is easy for single-objective optimization algorithm commonly used to fall into local optimum and over-fitting, resulting in poor stability. 5) In addition, previous studies on air quality have focused on air quality forecasting, while the research on air quality assessment was relatively rare.

Based on the above analysis, it is necessary to overcome these deficiencies and develop a novel and robust air quality warning system. The evaluation–forecast system developed in this study consists of two parts: evaluation and forecasting. The evaluation part involves feature extraction and finding out the main air pollutants; in the forecasting part, a new metric is developed to find the optimal model in each category, and optimal forecasting models are optimized with modified gray wolf optimization (DEGWO) optimization algorithm and leave-one-out deciding weight strategy to improve the accuracy of forecasting results and provide support for early warning systems. The specific implementation steps of the hybrid forecasting system are as follows: First, the feature extraction of the original data is performed to find similar attributes of AQI time series according to the relevant theory of fuzzy C-mean cluster.

Moreover, air quality evaluation based on the forecasting results of air pollutant concentration plays a crucial role in the development of the air quality warning system.

In this paper, in view of the uncertainty and ambiguity of each air pollutant, the fuzzy comprehensive evaluation is applied in AQI. According to the implementation of fuzzy comprehensive evaluation results, finding out the main pollutants in each city is another important part of this work. Next, we use long short-term memory (LSTM), backpropagation neural network (BPNN), adaptive network-based fuzzy inference system (ANFIS), generalized regression neural network (GRNN), and SVM models to forecast the main air pollutants time series, and a developed new metric is used to select optimal forecasting model. Finally, all these individual forecasting models' predictors based on the leave-one-out deciding weight strategy are optimized by the DEGWO optimization algorithm, and the final forecasting results are obtained. Therefore, the combination of these methods will result in more accurate forecasts and assessments performance, providing significant advantages for the construction and implementation of early warning systems for detecting air quality. The main contributions of this paper are as follows:

1) The fuzzy comprehensive evaluation is established for six air pollutants, which calculates the fuzzy membership degree of each pollutant and determines the main pollutants of each city.
2) A model selection index is established to select the optimal forecasting model from different neural network models. Based on model selection, the established weighted information criterion can select the optimal forecasting model for $PM_{2.5}$, PM_{10}, and NO_2 forecasting.
3) The forecasting performance of the optimal single model is improved. In the forecasting process, an improved multi-objective optimization algorithm is used to optimize the parameters of the single forecasting model, which not only

improves the prediction accuracy but also improves the stability of the single model

4) The model selection index is used to select the optimal forecasting value from the optimal hybrid model.

METHODOLOGY

In this subsection, the relative methods are presented in detail, including the data mining technique, forecasting model, and the DEGWO) algorithm. Subsequently, the marching process of our developed combined model is demonstrated.

Forecasting Model

Five typical models, namely, the multilayer perceptron (MLP) (You et al., 2017), ANFIS (Jang, 1993), LSTM (Muzaffar and Afshari, 2019), SVM (Brereton and Lloyd, 2010), and GRNN (Land and Schaffer, 2020), have been widely used for air pollutants forecasting because of their robustness, efficiency, and accuracy.

Modified Gray Wolf Optimization (DE-GWO)

For the DE algorithm and gray wolf optimization (GWO) algorithm, the defects of prematurity, poor stability, and ease in falling into local optimum will occur when solving the optimization problem separately. Combining the advantages and disadvantages of the two algorithms, a more efficient hybrid optimization algorithm, DEGWO algorithm, is proposed to improve global search capabilities.

Firstly, in order to avoid the phenomenon in which the population is iteratively reduced to a certain area, the crossover and selection operations of the DE algorithm are used to maintain the diversity of the population. Then, as the initial population of the GWO algorithm, the objective function value of the individual is calculated. The optimal three individuals X_α, X_β, and X_δ are selected to update the positions of other gray wolf individuals. Then, the position of the gray wolf individual is updated by the intersection and selection operations of DE, and the iterative update is repeated until the optimal one is selected. The target function value is output.

The hybrid algorithm not only improves the global search ability but also effectively avoids the defects of early maturity stagnation and falling into local optimum. The specific implementation steps of the algorithm are as follows:

Step 1: Set the relevant parameters of the hybrid optimization algorithm, population size N, maximum iteration number t_{max}, crossover probability CR, search dimension D, search range ub, lb, and scaling factor range F.

Step 2: The parameters a, A, and C are initialized, and the DE variant operation is performed on the population individual according to **Eq. 1** to generate an intermediate; an initial population and the number of iterations is set to $t = 1$.

$$v_{ij}(g+1) = \begin{cases} h_{ij}(g), & rand(0,1) \le CR \quad or \quad j = rand(1,n) \\ x_{ij}(g), & rand(0,1) > CR \quad or \quad j \ne rand(1,n) \end{cases} \tag{1}$$

Then the competition selection operation is performed according to **Eq. 2** to generate.

$$x_i(g+1) = \begin{cases} v_i(g), & f[v_i(g+1)] < f[x_i(g+1)] \\ x_i(g), & f[v_i(g+1)] \ge f[x_i(g+1)] \end{cases} \tag{2}$$

Step 3: Calculate the objective function value of each gray wolf individual in the population, sort according to the size of the objective function value, and select the optimal first three individuals as X_α, X_β, and X_δ, respectively.

Step 4: Calculate the distance between other gray wolf individuals in the population and the optimal X_α, X_β, and X_δ according to **Eqs 3–5**.

$$D_\alpha = |C_1 X_\alpha(t) - X(t)| \tag{3}$$

$$D_\beta = |C_2 X_\beta(t) - X(t)| \tag{4}$$

$$D_\delta = |C_3 X_\delta(t) - X(t)| \tag{5}$$

Finally, update the current position of each gray wolf individual according to **Eqs 6–9**.

$$X_1(t+1) = X_\alpha(t) - A_1 D_\alpha \tag{6}$$

$$X_2(t+1) = X_\beta(t) - A_2 D_\beta \tag{7}$$

$$X_3(t+1) = X_\delta(t) - A_3 D_\delta \tag{8}$$

$$X_p(t+1) = \frac{X_1 + X_2 + X_3}{3} \tag{9}$$

Step 5: Update the values of a, A, and C in the algorithm, cross-operate the position of the individual population according to **Eq. 1**, retain the better components, then perform **Eq. 2** to select new individuals and calculate the objective function values of all gray wolf individuals.

Step 6: Update the positions of the top three gray wolf individuals X_α, X_β, and X_δ.

Step 7: Determine the count value. If the maximum iteration number t_{max} is reached, the algorithm exits and, based on **Eq. 10**, outputs the multi-objective function value of the global optimal X_α; otherwise, let $t = t + 1$, and then go to Step 3 to continue execution.

$$\min f = f_1(x) + f_2(x)$$
$$s.t \begin{cases} f_1(x) = \frac{1}{n}\sum_{i=1}^{n} (error_i)^2 \\ f_2(x) = \frac{1}{n}\sum_{i=1}^{n} (error_i - \overline{error})^2 \end{cases} \tag{10}$$

Fuzzy C-Means Clustering

Fuzzy C-means clustering (FCM), known as fuzzy ISODATA, is a clustering algorithm that uses membership degrees to determine the extent to which each data point belongs to a certain cluster. In 1973, Bezdek proposed the algorithm as an improvement to the early hard C-means clustering (HCM) method (Gayen and Biswas, 2021). The clustering steps are as follows:

Step 1: Initialize the membership matrix U with a random number whose value is between 0 and 1, so that it satisfies the constraint in **Eq. 11**.

$$\sum_{i=1}^{c} u_{ij} = 1, \forall_j = 1, ..., n \tag{11}$$

Step 2: Calculate c cluster centers c_i ($i = 1, c$) using **Eq. 12**.

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}} \tag{12}$$

Step 3: Calculate the value function according to **Eq. 13**. If it is less than a certain threshold, or if the amount of change from the value of the last value function is less than a certain threshold, the algorithm stops.

$$\mathrm{J}(\mathrm{U}c_1, ..., c_c) = \sum_{\mathrm{i}=1}^{\mathrm{c}} J_i = \sum_{i=1}^{c} \sum_{j}^{n} u_{ij}^{m} d_{ij}^{2} \tag{13}$$

Step 4: Calculate the new U matrix with **Eq. 14**. Go back to Step 2.

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{(m-1)}}} \tag{14}$$

Fuzzy Synthetic Evaluation Theory

The process of establishing a fuzzy synthetic evaluation (FSE) system is as follows (Lu et al., 2011).

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 considered.

Step 2. The evaluation rank standard is determined.

The evaluation rank set is described as $\mathrm{V} = \{v_1, v_2, \cdots, v_n\}$. In our study, the air pollution degrees were divided into five levels. The pollutants grading standard according to AAQS is shown in **Supplementary Appendix S1**.

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.

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.

Step 5. The evaluation results are output.

The objective function of the DEGWO algorithm is based on stability and accuracy, in which MSE is the standard to measure accuracy and the variance of error is the standard to measure stability. **Algorithm 1** briefly outlines the process of the MODEGWO.

Algorithm 1. *MMODA*

Input: Objective function Min $fitness(x) = f_1 + f_2$

$$min \begin{cases} f_1(x) = MSE = \dfrac{1}{N}\sum_{i=1}^{N} (\hat{x}_i - x_i),\ i = 1, 2, ..., N \\ f_2(E) = Var(E) = \dfrac{1}{N}\sum_{i=1}^{N} \left(E_i - \bar{E}\right)^2,\ i = 1, 2, ..., N \end{cases}$$

Note: E_i is the test error; the calculation equation is $E_i = \hat{x}_i - x_i$

$\hat{x}_i$ and x_i are the actual data and output data by each model

Parameters of DEGWO

CR is crossover probability: 0.2;

MaxGen is the maximum number of the iteration: 500;

F is the scaling factor: [0.2, 0.8];

$\mathrm{p_{size}}$ is population size: 50.

Output: The optimal solution and the best objective function value.

Initialize a parent population, mutant population, and child population of gray wolf with a random position in a feasible region using equation;

$$x_p^k = x_p^k(low) + \left(x_p^k(up) - x_p^k(low)\right) \times rand(0, 1)$$

Note: $x_p^k(low)$ is the lower bound of the pth component of the kth individual.

$x_p^k(up)$ is the upper bound of the pth component of the kth individual.

rand(0, 1) represents a random number in [0, 1].

$p = 1, 2\cdots$, d. $k = 1, 2\cdots, \mathrm{p_{size}}$

Initialize crossover probability Pc and scaling factor F;

Initialize a, A, and C;

Evaluate f for all individuals in the parent population;

Sort the parent population in a non-decreasing order, according to the objective function value;

X_α is the best individual in the parent population of gray wolves;

X_β is the second individual in the parent population of gray wolves;

X_δ is the third individual in the parent population of gray wolves;

While ($t < MaxGen$)

for each individual in the parent population of gray wolves

Update the position using the following equation;

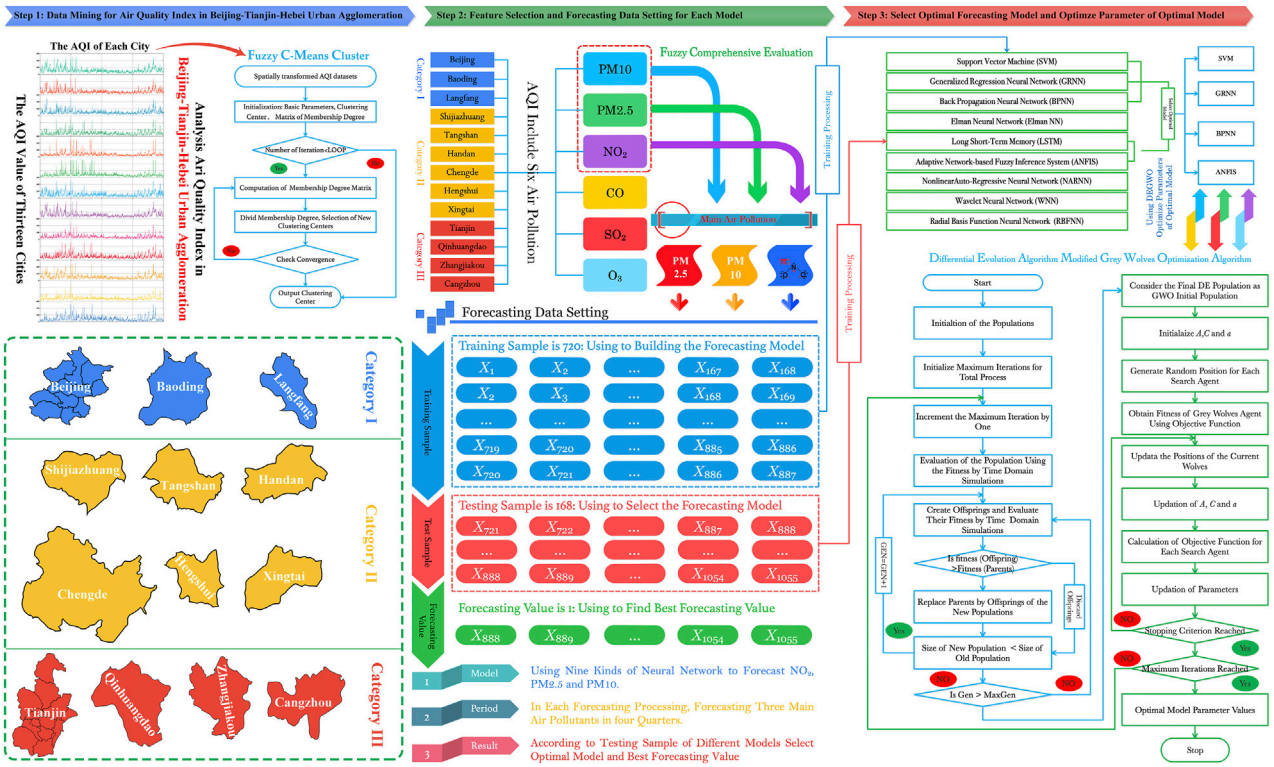

FIGURE 1 | Flowchart of air quality index forecasting system for Beijing–Tianjin–Hebei Urban Agglomeration.

$$X_p(t+1) = \frac{X_1 + X_2 + X_3}{3}$$

end for

Obtain a mutant population of gray wolves using the following equation;

$$V^i(g) = X^{r1} + F \cdot \left(X^{r2}(g) + X^{r3}(g)\right)$$

Note: g is the generation number,

F is the scaling factor, $g = 0, 1, 2, \cdots$, *MaxGen*,

MaxGen is the maximum number of the iteration

Obtain a child population of gray wolves using the following equation;

$$v_{ij}(g+1) = \begin{cases} h_{ij}(g), rand(0,1) \le CR & j = rand(1,n) \\ x_{ij}(g), rand(0,1) > CR & j = rand(1,n) \end{cases}$$

CR represents the crossover probability

for each individual $Parent_i$ in a parent population of gray wolves

If $f(Child_i) < f(Parent_i)$

Replace $Parent_i$ with $Child_i$

end if

Update A, C, and a;

Sort the parent population of gray wolves in a nondecreasing order;

Update X_α, X_β, X_δ;

$t = t + 1$;

End while

Return *Parentα* and *f*(*Parentα*).

Formulation of the Hybrid Model

The hybrid AQI forecasting system in this paper is composed of the above three parts. A flow chart of the hybrid model is presented in **Figure 1**.

From the above, we can see that the AQI forecasting step using the hybrid forecasting system proceeds as follows:

Step 1: Data Mining

1. Collect the original data in the proposed hybrid forecasting model. Specifically, the average hourly AQI and six air pollutants are utilized as experiment data in this work.
2. Using data mining technology, 13 cities in Beijing–Tianjin–Hebei Urban Agglomeration (BJ-TJ-HE) are clustered, the characteristics of each category are summarized, and each category is further analyzed.

The AQI and six air pollution time series with missing points is filling processed by shape-preserving piecewise cubic spline interpolation, which maintains the continuity of each time series.

Step 2: Feature selection and data setting for each model

1. Feature selection: According to the result of cluster, establish fuzzy comprehensive evaluation for six pollutants and find out the main air pollutants of each in the same category.
2. Data setting: Each main air pollutants time series can be divided into three parts: training sample and testing samples for the forecasting values. The training sample is used to construct and train the ANNs, which in this work consist of a BPNN, SVM, GRNN, LSTM, and ANFIS. In addition, the testing sample is used to select the optimal model. For this, the WIC values of the ANNs are calculated, and the best model in terms of the WIC is selected. The input data are used to train the ANNs before calculating the forecasting value, with 1–6 input nodes and 1–30 hidden nodes. According to the value of WIC, the best forecasting model and best structure are chosen.

Step 3: Optimize the parameters of the best forecasting model.

To ensure the forecasting performance, a modified optimization algorithm is used to further optimize the parameters of the best forecasting model (expect LSTM). Finally, the main air pollutants forecasting results are obtained and compared with those of different hybrid forecasting models.

EXPERIMENT DESIGN AND ANALYSIS

In this section, the specific information of experiment datasets in BJ-TJ-HE are described in detail. Eight performance metrics are applied to assess the performance of the proposed model. The experiments conducted in this study were implemented on Matlab 2018a, and the specifications of the hardware were as follows: Intel Core i9-7920X 2.90 GHz CPU and 32 GB RAM.

Data Description

The BJ-TJ-HE is the national capital region of the People's Republic of China. It is the biggest urbanized megalopolis region in Northern China, where Beijing, Tianjin, Baoding, and Langfang are the central core areas of BJ-TJ-HE. In this paper, the 13 cities of BJ-TJ-HE are evaluated to develop an early warning indicator for air quality. The datasets of hourly concentrations of the six major air pollutants used in this study are retrieved from the website of the China National Environment Monitoring Centre (http://www.cnemc.cn/sssj/). The first dataset includes AQI hourly concentrations collected from January 1, 2017, to December 31, 2018, in BJ-TJ-HE. **Figure 2** shows the result of fuzzy C-mean cluster, which displays the construction of a fuzzy matrix based on the attributes of AQI in 13 cities and objectively and accurately cluster (Category I: Beijing, Baoding, Langfang; Category II: Shijiazhuang, Tangshan, Handan, Chengde, Hengshui, Xingtai; Category III: Tianjin, Qinhuangdao, Zhangjiakou, Cangzhou). The result of fuzzy comprehensive evaluation is shown in **Table 1**, which found that the main air pollutants are PM_{10}, PM_2, and NO_2 in 13 cities.

According to the analysis in **Table 1**, the main air pollutants from statistical analysis of BJ-TJ-HE are NO_2, $PM_{2.5}$, and PM_{10} shown in **Table 2**, in which the average value of the main air pollutants shows obvious differences among the 13 cities. The average value of NO_2 in the different cities is between 22.2525 and 49.4348 $\mu g/m^3$, in which the average value in Xingtai is higher than in the other cities. At the same time, the $PM_{2.5}$ and PM_{10} average values in Xingtai are 69.6938 and 135.8368 $\mu g/m^3$, which are also higher than in the other cities. The maximum values of NO_2, $PM_{2.5}$, and PM_{10} were in Hengshui, Baoding, and Zhangjiakou, with values of 215, 402, and 1581 $\mu g/m^3$, and the minimum values of the three main air pollutants were in Zhangjiakou, Beijing, and Zhangjiakou, with values of 1, 3, and 12 $\mu g/m^3$.

In terms of skewness, all data sets are rightward, with values of skewness are greater than 0. For the values of kurtosis, only three data sets of NO_2 were less than 3, which meant that these three sets (Qinhuangdao, Shijiazhuang, and Xingtai) had a fat tail. At the same time, the other data sets had a thin tail.

Forecasting Metric

This section focuses on the efficiency of the different forecasting model with respect to computational performance. Eight evaluation criteria are applied to estimate the forecasting performance, namely, mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), Theil U statistic 1 (U1), and Theil U statistic 2 (U2) were calculated for all the fits; the goodness of forecasting fit (R^2) and the standard of forecasting error (STDE) indicates the stability of the forecasting models; and the direction accuracy (DA) evaluates the optimal decision-making, often relying on correct forecasting directions or turning points between the actual and forecasting values. These performance metrics are defined in **Table 3**.

Experiment Preparation: Model Selection

In the forecasting processing, there is no model that can be applied to all time series in the process of forecasting. Therefore, in this paper we developed a new metric, which measures accuracy of each hybrid model testing set and determines whether the model can provide the optimal forecasting value. The process of model selection is as follows:

Each model data is divided into 840 training samples, 168 testing samples, and one forecasting value. The accuracy of the testing sample is calculated by using the WIC. In order to eliminate the difference of the order of magnitude of forecasting metric, the MAE, MAE RMSE, MAPE, STDE, U1, and U2 are normalized. The calculation formula is as follows:

$$WIC = NMAE + NRMSE + NMAPE + NSTDE + U1 + U2 + R2 + (1 - \mathrm{DA})$$

For the first forecasting, the 1st to 840th samples are the training samples, the 841st to 1008th samples are the testing samples, and the 1009th sample is the forecasting value. At the end of the forecasting, the WIC value of the testing sample is calculated. If the WIC of the *i*th model is the smallest, the forecasting value of the *i*th model provides the optimal forecasting value.

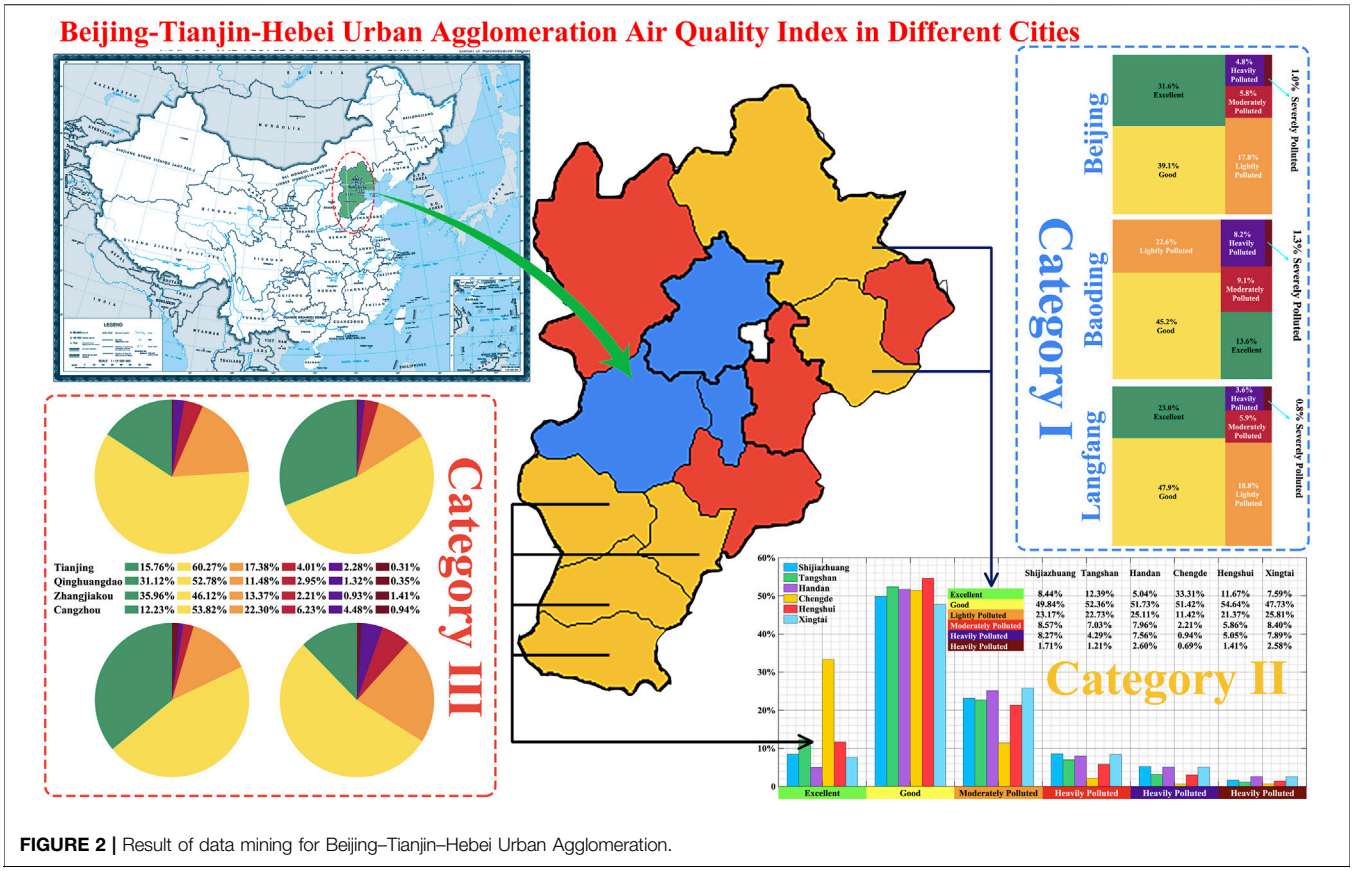

FIGURE 2 | Result of data mining for Beijing–Tianjin–Hebei Urban Agglomeration.

TABLE 1 | The result of fuzzy comprehensive evaluation

City	Air pollution					
	PM_{10}	$PM_{2.5}$	NO_2	CO	SO_2	O_3
Beijing	**0.91237**	**0.98072**	**0.63530**	0.43071	0.39424	0.38603
Tianjin	**0.86953**	**0.90659**	**0.64191**	0.30671	0.37710	0.48547
Shijiazhuang	**0.75937**	**0.70576**	**0.67125**	0.38993	0.43133	0.41205
Tangshan	**0.91044**	**0.74470**	**0.64615**	0.37757	0.35546	0.43152
Qinhuangdao	**0.87371**	**0.89213**	**0.72628**	0.41083	0.30361	0.41777
Handan	**0.82031**	**0.98267**	**0.72983**	0.48603	0.31468	0.49503
Baoding	**0.87530**	**0.95838**	**0.75981**	0.34638	0.41152	0.39322
Zhangjiakou	**0.90838**	**0.99560**	**0.63953**	0.30755	0.42579	0.31045
Chengde	**0.81195**	**0.73604**	**0.82316**	0.40926	0.48029	0.31550
Langfang	**0.75893**	**0.81218**	**0.87910**	0.47526	0.42354	0.30550
Cangzhou	**0.84830**	**0.82908**	**0.86664**	0.49499	0.31973	0.41617
Hengshui	**0.75544**	**0.93064**	**0.97252**	0.35830	0.40872	0.33812
Xingtai	**0.90721**	**0.93224**	**0.70024**	0.47492	0.44532	0.31803

Note: if the result of fuzzy comprehensive evaluation is greater than 0.5, the pollutant is the main pollutant. The bold values are main Air Pollution, which the fuzzy comprehensive evaluation results are greater than 0.5.

For the second forecasting, the 2nd to 841st samples are the training samples, the 842nd to 1009th samples are the testing samples, and the 1010th sample is the forecasting value. At the end of the forecasting, the WIC value of the testing sample is calculated. If the WIC of the ith model is the smallest, the forecasting value of the ith model provides the optimal forecasting value.

For the kth forecasting, the kth to $(840 + k - 1)$th samples are the training samples, the $(840 + k)$th to $(1008 + k - 1)$th samples are the testing samples, and the $(1008 + k)$th sample is the forecasting value. At the end of the forecasting, the WIC value of the testing sample is calculated. If the WIC of the ith model is the smallest, the forecasting value of the ith model provides the optimal forecasting value.

TABLE 2 | Statistic of each main air pollutant in different cities.

City	Pollutant	Min	Max	Mean	Skewness	Kurtosis	Mode
Beijing	NO_2	16	158	39.1947	0.7894	3.2759	18
	$PM_{2.5}$	3	273	50.0068	1.7017	6.5652	5
	PM_{10}	56	1,058	83.0744	5.7125	63.4085	27
Tianjin	NO_2	17	195	49.4348	0.8042	3.5754	36
	$PM_{2.5}$	11	237	49.9759	2.0621	9.2214	40
	PM_{10}	23	321	85.0560	1.7136	7.1605	68
Shijiazhuang	NO_2	4	145	44.0150	0.6088	2.7786	16
	$PM_{2.5}$	12	343	69.6938	1.7761	6.5778	43
	PM_{10}	29	461	132.2141	1.3934	4.9017	80
Tangshan	NO_2	5	200	54.2055	0.6325	3.5206	36
	$PM_{2.5}$	12	350	60.4803	2.4753	12.7626	44
	PM_{10}	26	484	114.9052	1.8083	8.0477	76
Qinhuangdao	NO_2	4	130	43.2894	0.6124	2.7408	17
	$PM_{2.5}$	6	250	38.2879	2.1722	9.8110	21
	PM_{10}	15	392	78.6504	1.7126	7.8222	54
Handan	NO_2	4	153	40.4909	0.9210	3.4780	17
	$PM_{2.5}$	8	321	69.4114	1.9836	7.2348	45
	PM_{10}	24	513	135.8368	1.6375	5.8557	85
Baoding	NO_2	4	159	44.7402	0.8947	3.3395	15
	$PM_{2.5}$	11	402	66.6175	1.9971	8.9785	33
	PM_{10}	24	527	115.9434	1.5658	6.3524	65
Zhangjiakou	NO_2	1	136	22.2525	1.9035	8.0002	19
	$PM_{2.5}$	7	184	31.1680	2.2524	10.1861	16
	PM_{10}	12	1,581	88.7293	6.2068	64.1126	31
Chengde	NO_2	2	126	30.6221	0.7458	3.5114	6
	$PM_{2.5}$	7	189	31.7188	2.2798	10.0002	13
	PM_{10}	14	561	81.6457	2.6593	14.4003	36
Langfang	NO_2	7	206	45.0315	1.0029	3.9963	20
	$PM_{2.5}$	9	282	51.8249	1.9651	7.9869	15
	PM_{10}	21	428	100.0498	1.6177	6.3231	69
Cangzhou	NO_2	5	170	41.2205	1.0111	4.0626	22
	$PM_{2.5}$	10	381	58.8355	2.5108	12.7637	35
	PM_{10}	16	509	104.4355	1.8495	8.5249	75
Hengshui	NO_2	2	215	32.2611	1.1356	4.8037	15
	$PM_{2.5}$	14	310	61.7583	2.3912	10.0981	40
	PM_{10}	19	391	103.5030	1.7636	6.3377	68
Xingtai	NO_2	3	155	47.4996	0.5950	2.9794	18
	$PM_{2.5}$	11	369	69.1419	1.9618	7.4487	35
	PM_{10}	17	524	135.3799	1.3911	5.2029	88

TABLE 3 | Definition of the performance metrics.

Metric	Definition	Equation
MAE	The average absolute forecasting results error of *n*	$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$
RMSE	The mean absolute percentage error of *n* forecasting results	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
MAPE	The root mean square error of *n* forecasting results	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\lvert\frac{y_i - \hat{y}_i}{y_i}\rvert \times 100\%$
STDE	The standard of error of *n* forecasting results	$STDE = \sqrt{Var(y_n - \hat{y}_n)}$
U1	Theil U statistic 1	$U1 = \frac{\sqrt{\frac{1}{T}\sum_{T=1}^{T}(y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{T}\sum_{T=1}^{T}(y_i)^2} + \sqrt{\frac{1}{T}\sum_{T=1}^{T}(\hat{y}_i)^2}}$
U2	Theil U statistic 2	$U2 = \frac{\sqrt{\frac{1}{T}\sum_{T=1}^{T}(\frac{y_{i+1} - y_{i+1}}{y_i})^2}}{\sqrt{\frac{1}{T}\sum_{T=1}^{T}(\frac{\hat{y}_{i+1} - y_{i+1}}{y_i})^2}}$
R^2	The goodness of forecasting fit	$\frac{R^2 = \sum_{n=1}^{N}(y_n - \bar{y}) - \sum_{n=1}^{N}(y_n - \hat{y}_n)}{\sum_{n=1}^{N}(y_n - \bar{y})}$
DA	Directions or turning points between actual and forecasting values	$DA = \frac{1}{T}\sum_{i=1}^{T} a_i,\ a_i = \begin{cases} 1 & if\ (y_{i+1} - \hat{y}_i)(\hat{y}_{i+1} - y_i) > 0 \\ 0 & otherwise. \end{cases}$

Note: y_i is the actual value, $\hat{y}_i$ is the forecasted value, and T *is the total number of data items.*

TABLE 4 | The result of model selection for main air pollutants in different seasons.

Category	Air pollutants	Selection	First season				Second season				
I	NO_2	Model	SVM	GRNN	BPNN	ANFIS	SVM	GRNN	BPNN	LSTM	
		Percentage	67.86%	21.43%	7.14%	3.57%	39.29%	46.43%	10.71%	3.57%	
II		Model	SVM	GRNN	BPNN	LSTM	SVM	GRNN	BPNN	ANFIS	LSTM
		Percentage	53.57%	35.71%	3.57%	7.14%	64.29%	17.86%	7.14%	3.57%	7.14%
III		Model	SVM	GRNN	BPNN		SVM	GRNN	BPNN	LSTM	
		Percentage	53.57%	35.71%	10.71%		32.14%	50.00%	10.71%	7.14%	
I	$PM_{2.5}$	Model	SVM	BPNN			SVM	BPNN			
		Percentage	42.86%	57.14%			57.14%	42.86%			
II		Model	SVM	BPNN	ANFIS	LSTM	SVM	BPNN	ANFIS	LSTM	
		Percentage	64.29%	21.43%	3.57%	10.71%	64.29%	17.86%	4.29%	3.57%	
III		Model	SVM	BPNN	ANFIS	LSTM	SVM	BPNN	ANFIS	LSTM	
		Percentage	71.43%	17.86%	3.57%	7.14%	78.57%	14.29%	3.57%	3.57%	
I	PM_{10}	Model	SVM	BPNN	ANFIS		SVM	BPNN			
		Percentage	82.14%	14.29%	3.57%		78.57%	17.86%			
II		Model	SVM	BPNN			SVM	BPNN			
		Percentage	76.79%	16.07%			78.57%	21.43%			
III		Model	SVM	BPNN			SVM	GRNN	BPNN		
		Percentage	82.14%	17.86%			75.00%	14.29%	10.71%		

Category	Air pollutants	Selection	Third season					Fourth season			
I	NO_2	Model	SVM	GRNN	BPNN			SVM	GRNN	BPNN	LSTM
		Percentage	57.14%	25.00%	17.86%			60.71%	28.57%	5.36%	5.36%
II		Model	SVM	GRNN	BPNN			SVM	GRNN	BPNN	LSTM
		Percentage	60.71%	32.14%	7.14%			51.79%	32.14%	8.93%	7.14%
III		Model	SVM	GRNN	BPNN	ANFIS	LSTM	SVM	GRNN	BPNN	
		Percentage	46.43%	25.00%	14.29%	7.14%	7.14%	46.43%	39.29%	10.71%	
I	$PM_{2.5}$	Model	SVM	BPNN	LSTM			SVM	BPNN	ANFIS	
		Percentage	60.71%	17.86%	21.43%			50.00%	33.93%	14.29%	
II		Model	SVM	BPNN	ANFIS	LSTM		SVM	BPNN	ANFIS	
		Percentage	71.43%	17.86%	7.14%	3.57%		71.43%	21.43%	7.14%	
III		Model	SVM	BPNN	ANFIS	LSTM	GRNN	SVM	BPNN	ANFIS	LSTM
		Percentage	53.57%	28.57%	7.14%	7.14%	3.57%	64.29%	21.43%	10.71%	3.57%
I	PM_{10}	Model	SVM	BPNN				SVM	BPNN	ANFIS	
		Percentage	82.14%	17.86%				71.43%	21.43%	7.14%	
II		Model	SVM	BPNN	ANFIS			SVM	BPNN		
		Percentage	71.43%	17.86%	10.71%			67.86%	21.43%		
III		Model	SVM	BPNN	ANFIS			SVM	BPNN	ANFIS	
		Percentage	78.57%	17.86%	3.57%			76.79%	17.86%	5.36%	

In the whole forecasting process, 168 optimal forecasting values are generated, and 168*5 WIC values are generated at the same time. **Table 4** only shows the optimal model and the percentage of optimal forecasting value for three main air pollutants.

From **Table 4**, it can be seen that SVM provides more optimal forecasting value for the three main pollutants at different times, especially in the PM_{10} forecasting process; the optimal forecasting value for the first quarter and the third quarter is 82.14% (138 optimal forecasting value), and the other four models also provide corresponding optimal forecasting value.

Experiment I: Forecasting Processing for Three Categories of NO_2 by Each Model in the First Season

In this portion, the hourly NO_2 time series for 13 cities in three categories were utilized as the testing data for the five hybrid models with one-step-ahead forecasting. Beyond that, with the purpose of comprehensively comparing the precision of the modeling forecasting, this experiment consisted of two parts: the multi-step forecasts demonstrated in **Table 4** and, for the local analysis horizon, the local forecasts presented in **Table ′5** and **Figure 3**, which focus on first season. **Table 5** and **Figure 3** demonstrate the following:

1) Focusing on Category I, the new proposed model based on model selection realizes excellent results on the eight evaluation indices in the first season forecasting. On the contrary, DEGWO-ANFIS has the lowest effectiveness. The maximum reduction of MAPE for the proposed model compared with the other hybrid models is approximately 71.18% in Beijing's NO_2 forecasting, 53.93% in Baoding's NO_2 forecasting, and 61.61% in Langfang's NO_2 forecasting, respectively. The reduction was about 62.39% and 76.49% for one-step forecasting and 2.79%, 6.10%, and 19.33% for the three cities at the hourly interval

TABLE 5 | The forecasting results of each model for NO_2 in three categories.

Metric	Model	MAE	RMSE	STDE	DA	U1	U2	MAPE	R^2
Category I									
Beijing	MODEGWO-SVM	2.1754	3.1919	3.2010	80.84%	0.0464	2.0148	9.95%	0.9868
	MODEGWO-GRNN	3.0082	3.9179	3.9242	50.90%	0.0569	2.6113	15.23%	0.9800
	MODEGWO-BPNN	2.6589	3.7849	3.7857	69.46%	0.0550	2.4488	13.07%	0.9814
	MODEGWO-ANFIS	3.4548	6.0636	6.0514	64.07%	0.0889	4.1096	16.57%	0.9524
	Model select	**2.1300**	**3.1765**	**3.1847**	**77.25%**	**0.0463**	**1.9446**	**9.68%**	**0.9869**
Baoding	MODEGWO-SVM	4.7162	7.4952	7.5055	77.25%	0.0665	2.1441	12.69%	0.9620
	MODEGWO-GRNN	5.1070	7.9556	7.8716	60.48%	0.0706	3.6465	17.57%	0.9569
	MODEGWO-BPNN	6.9580	9.8870	9.5789	59.28%	0.0906	2.3774	17.26%	0.9355
	MODEGWO-ANFIS	7.0574	10.5964	10.1816	56.89%	0.2401	2.4591	18.41%	0.9223
	Model select	**4.4448**	**7.3418**	**7.3599**	**76.65%**	**0.0654**	**2.1115**	**11.96%**	**0.9628**
Langfang	MODEGWO-SVM	4.5141	7.5578	7.5731	76.05%	0.0880	2.2411	13.46%	0.9372
	MODEGWO-GRNN	3.8320	5.2949	5.3098	62.28%	0.0625	1.7523	12.52%	0.9682
	MODEGWO-BPNN	5.3965	8.4998	8.3895	62.87%	0.0980	2.9703	17.22%	0.9214
	MODEGWO-ANFIS	6.3336	11.4957	11.3948	60.48%	0.1305	3.2108	18.23%	0.8745
	Model select	**3.6594**	**5.3249**	**5.3380**	**68.86%**	**0.0627**	**1.7819**	**11.28%**	**0.9676**
Category II									
Shijiazhuang	MODEGWO-SVM	2.8162	4.8606	4.8733	82.04%	0.0498	2.3785	8.21%	0.9695
	MODEGWO-GRNN	3.5480	5.0067	5.0163	50.90%	0.0517	2.2642	10.46%	0.9674
	MODEGWO-BPNN	4.3533	6.7126	6.6382	58.68%	0.0694	2.9803	11.62%	0.9450
	Adam-LSTM	5.8291	8.3652	8.3630	55.09%	0.0859	3.5319	14.95%	0.9167
	Model select	**2.8001**	**4.0036**	**4.0095**	**74.25%**	**0.0412**	**2.0289**	**8.02%**	**0.9792**
Tangshan	MODEGWO-SVM	4.2928	6.4751	6.4876	82.63%	0.0584	2.3028	12.41%	0.9643
	MODEGWO-GRNN	3.5396	5.2423	5.2563	73.65%	0.0474	1.5259	9.14%	0.9763
	MODEGWO-BPNN	5.9179	8.2307	8.2550	68.86%	0.0743	2.3721	16.20%	0.9423
	Adam-LSTM	5.0169	6.9705	6.9845	74.25%	0.0630	2.3430	13.86%	0.9579
	Model select	**3.5604**	**5.3107**	**5.3266**	**77.84%**	**0.0481**	**1.5419**	**9.10%**	**0.9757**
Handan	MODEGWO-SVM	3.6331	5.9196	5.9315	72.46%	0.0570	2.2986	9.69%	0.9552
	MODEGWO-GRNN	4.0699	6.2263	6.2395	60.48%	0.0604	2.3158	11.35%	0.9494
	MODEGWO-BPNN	5.7352	8.4969	8.4366	55.69%	0.0832	2.3412	13.92%	0.9058
	Adam-LSTM	5.3202	7.0150	6.9787	60.48%	0.0682	2.2208	12.31%	0.9378
	Model select	**3.3618**	**5.4628**	**5.4791**	**76.05%**	**0.0527**	**2.1130**	**8.79%**	**0.9620**
Chengde	MODEGWO-SVM	2.6775	4.0365	4.0326	71.26%	0.0653	1.8783	15.35%	0.9735
	MODEGWO-GRNN	2.6787	4.0350	4.0461	62.87%	0.0658	2.0476	17.48%	0.9735
	MODEGWO-BPNN	3.4332	4.9117	4.9230	60.48%	0.0804	2.2906	20.15%	0.9605
	Adam-LSTM	4.5095	6.9772	6.9648	67.07%	0.1123	2.6262	26.28%	0.9201
	Model select	**2.4481**	**3.7646**	**3.7746**	**73.05%**	**0.0612**	**1.7081**	**13.90%**	**0.9769**
Hengshui	MODEGWO-SVM	4.7952	6.5453	6.5494	73.65%	0.0764	2.1369	15.41%	0.9405
	MODEGWO-GRNN	6.0417	8.0467	8.0648	47.90%	0.0950	3.2945	23.59%	0.9078
	MODEGWO-BPNN	10.7085	17.2255	16.7668	55.69%	0.2122	3.6978	27.02%	0.5924
	Adam-LSTM	7.5643	10.4959	9.9896	49.70%	0.1296	2.3976	22.00%	0.8544
	Model select	**4.3346**	**6.1328**	**6.1423**	**74.85%**	**0.0717**	**2.0600**	**14.27%**	**0.9477**
Xingtai	MODEGWO-SVM	3.7973	5.6732	5.6880	70.66%	0.0494	2.1280	9.22%	0.9688
	MODEGWO-GRNN	4.0825	6.1326	6.1358	65.87%	0.0535	2.3674	10.61%	0.9633
	MODEGWO-BPNN	5.5707	8.1697	8.0221	61.68%	0.0727	2.6425	12.39%	0.9361
	Adam-LSTM	7.1767	11.0393	11.0367	65.87%	0.0967	2.9015	15.29%	0.8884
	Model select	**3.5614**	**5.4338**	**5.4465**	**67.07%**	**0.0474**	**1.9570**	**8.58%**	**0.9710**
Category III									
Tianjin	MODEGWO-SVM	2.5155	3.5288	3.5317	80.84%	0.0355	1.8109	5.98%	0.9830
	MODEGWO-GRNN	2.9357	4.2368	4.2449	70.06%	0.0427	1.9437	7.51%	0.9750
	MODEGWO-BPNN	3.1418	4.3375	4.3501	77.84%	0.0436	2.0924	7.57%	0.9744
	Model select	**2.3407**	**3.3033**	**3.3130**	**87.24%**	**0.0333**	**1.7699**	**5.66%**	**0.9850**
Qinhuangdao	MODEGWO-SVM	3.8108	5.4949	5.4950	81.44%	0.0554	1.7553	10.87%	0.9744
	MODEGWO-GRNN	3.7633	5.3865	5.3965	65.87%	0.0549	2.1029	12.88%	0.9752
	MODEGWO-BPNN	5.4520	7.6113	7.3974	67.66%	0.0791	2.0834	14.56%	0.9531
	Model select	**3.4285**	**4.8590**	**4.8733**	**74.85%**	**0.0494**	**1.7261**	**10.28%**	**0.9797**
Zhangjiakou	MODEGWO-SVM	2.3096	3.5862	3.5967	75.45%	0.0945	1.8076	13.87%	0.9058
	MODEGWO-GRNN	3.1474	4.1463	4.1587	44.91%	0.1106	2.3875	20.99%	0.8669
	MODEGWO-BPNN	2.7829	4.2140	4.1925	61.08%	0.1137	2.0222	16.66%	0.8640
	Model select	2.0741	2.8440	2.8524	71.86%	0.0753	1.7182	13.10%	0.9395
Cangzhou	MODEGWO-SVM	6.1489	9.0417	9.0626	78.44%	0.0803	2.0706	15.17%	0.9373
	MODEGWO-GRNN	5.4532	7.6253	7.6448	58.68%	0.0682	2.0220	14.67%	0.9535
	MODEGWO-BPNN	8.0522	11.4542	11.4009	55.69%	0.1031	2.8158	19.38%	0.9018
	Model select	**5.3105**	**7.9582**	**7.9179**	**70.66%**	**0.0716**	**1.9162**	**12.87%**	**0.9509**

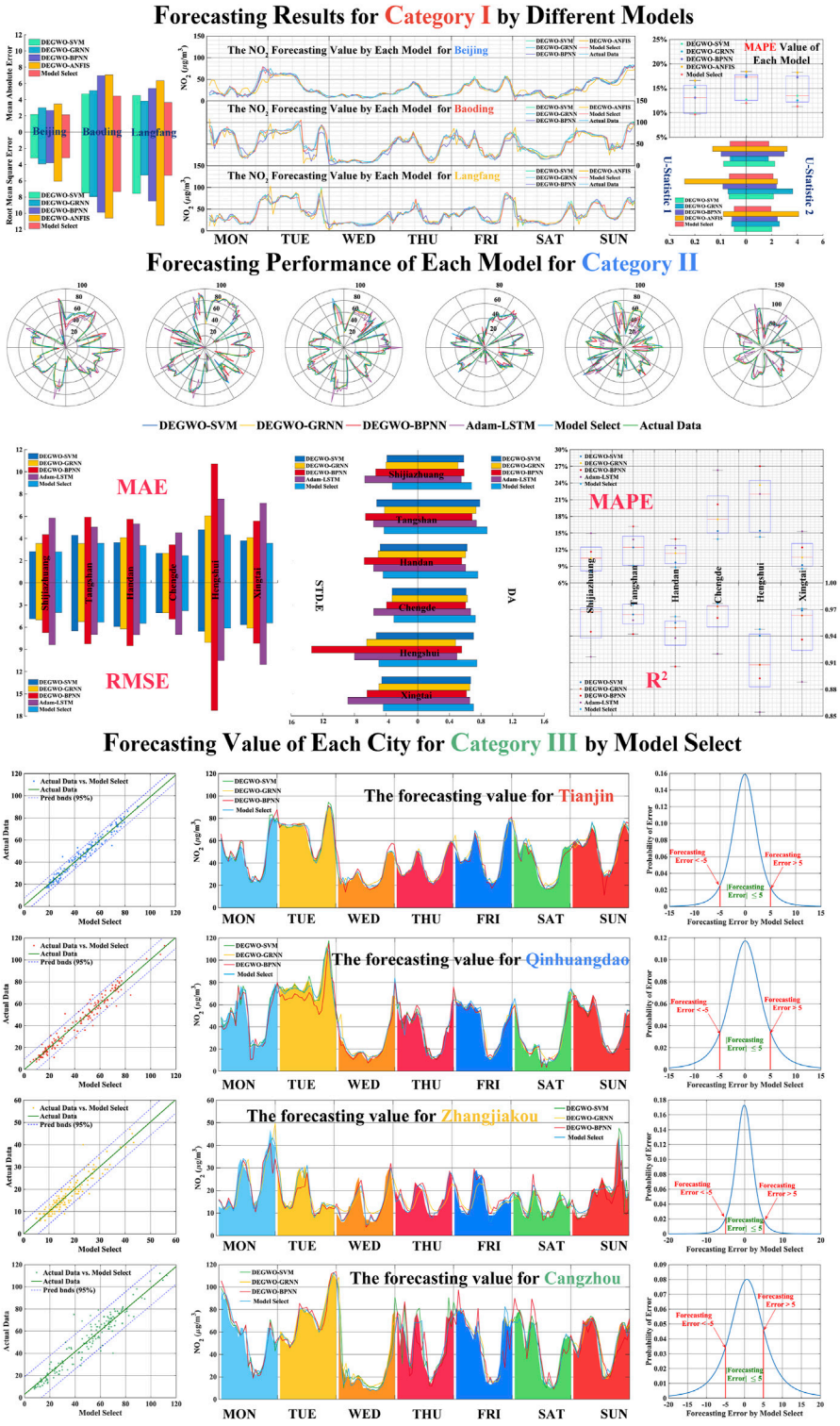

FIGURE 3 | Forecasting result of NO_2 for three categories in the first season.

NO_2 forecasting in Category I. As for the R^2, the proposed model has the best performance among the four single-hybrid models for hourly interval NO_2 time series.

2) Focusing on Category II, it is clear that proposed model based on model selection exhibits the best performance among the single fourth hybrid models implemented for all eight criteria

TABLE 6 | The $PM_{2.5}$ forecasting result for each city in Category II by different models.

Metric	Shijiazhuang				
	MODEGWO-SVM	**MODEGWO-BPNN**	**MODEGWO-ANFIS**	**Adam-LSTM**	**Model select**
MAE	0.4632	0.4450	0.6554	1.2945	0.3518
RMSE	0.6106	0.5877	1.2148	2.0978	0.4497
STDE	0.6124	0.5809	1.2182	2.1038	0.4509
DA	75.45%	73.05%	70.66%	42.51%	81.44%
U1	0.0067	0.0064	0.0133	0.0221	0.0049
U2	1.4464	1.3998	3.4525	4.3147	1.2087
MAPE	1.31%	1.08%	1.85%	2.88%	0.88%
R^2	0.9991	0.9992	0.9964	0.9889	0.9995
Metric	**Handan**				
	MODEGWO-SVM	**MODEGWO-BPNN**	**MODEGWO-ANFIS**	**Adam-LSTM**	**Model select**
MAE	0.5522	0.5250	1.3025	2.1956	0.4694
RMSE	0.7024	0.7009	2.5927	3.1495	0.6125
STDE	0.6835	0.7007	2.5352	3.1450	0.6092
DA	82.63%	83.83%	71.26%	43.71%	88.02%
U1	0.0051	0.0051	0.0191	0.0232	0.0045
U2	1.3944	1.4399	2.7102	4.7186	1.2395
MAPE	0.90%	0.85%	2.08%	3.46%	0.76%
R^2	0.9993	0.9992	0.9896	0.9831	0.9994
Metric	**Hengshui**				
	MODEGWO-SVM	**MODEGWO-BPNN**	**MODEGWO-ANFIS**	**Adam-LSTM**	**Model select**
MAE	0.4649	0.5305	1.5424	2.6723	0.4392
RMSE	0.6107	0.7545	3.3746	3.8908	0.6260
STDE	0.6044	0.7515	3.3556	3.8552	0.6279
DA	82.63%	74.85%	65.87%	38.92%	83.44%
U1	0.0056	0.0070	0.0312	0.0356	0.0058
U2	1.3886	1.4988	6.8777	6.1294	1.2831
MAPE	0.92%	1.07%	3.34%	5.02%	0.88%
R^2	0.9985	0.9984	0.9744	0.9724	0.9989
Metric	**Tangshan**				
	MODEGWO-SVM	**MODEGWO-BPNN**	**MODEGWO-ANFIS**	**Adam-LSTM**	**Model select**
MAE	0.3165	0.4449	1.1059	1.5655	0.3107
RMSE	0.3984	0.5718	2.9201	2.4642	0.3948
STDE	0.3961	0.5714	2.8491	2.4622	0.3947
DA	80.24%	69.46%	62.28%	43.11%	82.04%
U1	0.0052	0.0074	0.0381	0.0321	0.0051
U2	1.0656	1.3226	11.8381	5.4107	1.0363
MAPE	0.94%	1.36%	3.85%	4.63%	0.91%
R^2	0.9996	0.9992	0.9818	0.9845	0.9996
Metric	**Chengde**				
	MODEGWO-SVM	**MODEGWO-BPNN**	**MODEGWO-ANFIS**	**Ada-LSTM**	**Model select**
MAE	0.2413	0.3283	0.6149	0.9537	0.2388
RMSE	0.3103	0.4255	1.1569	1.7033	0.3017
STDE	0.3101	0.4241	1.1353	1.7052	0.3007
DA	70.06%	61.08%	53.89%	37.13%	72.06%
U1	0.0096	0.0104	0.0281	0.0414	0.0076
U2	1.0296	0.9017	2.7485	5.1292	1.0022
MAPE	1.46%	1.79%	3.20%	4.77%	1.34%
R^2	0.9987	0.9976	0.9827	0.9663	0.9987
Metric	**Xingtai**				
	MODEGWO-SVM	**MODEGWO-BPNN**	**MODEGWO-ANFIS**	**Ada-LSTM**	**Model select**
MAE	0.4611	0.5550	0.9154	1.4376	0.4354
RMSE	0.5936	0.7151	1.4925	1.9675	0.5685
STDE	0.5905	0.7119	1.4969	1.9238	0.5691
DA	74.85%	70.06%	56.89%	33.53%	77.25%
U1	0.0049	0.0059	0.0123	0.0161	0.0047
U2	1.3588	1.4830	3.4545	4.4867	1.2932
MAPE	0.79%	0.98%	1.67%	2.47%	0.75%
R^2	0.9988	0.9983	0.9923	0.9874	0.9989

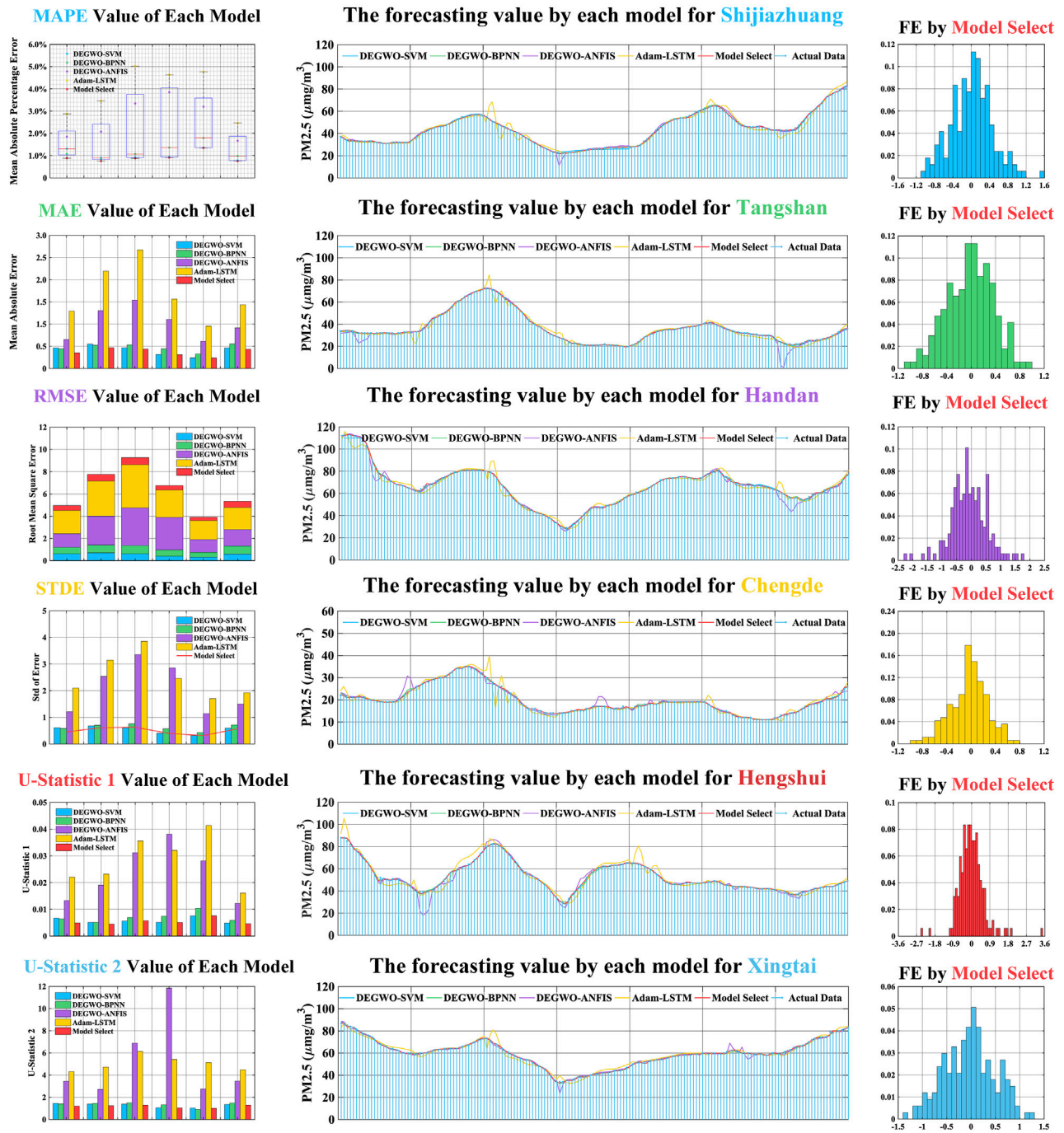

FIGURE 4 | Forecasting performance of each model for Category II in the first season.

involved. For the MAPE, there are average reductions between the proposed model and single hybrid model, by approximately 12.48%, 29.12%, 60.00%, and 66.70% in six cities for the hourly NO_2 time series forecasting, respectively. Comparing the four single hybrid models for the hourly NO_2 forecasting, the forecasting accuracy of DEGWO-SVM is higher than that of the other three hybrid models. The average reduction of MAPE among the MODEGWO-SVM and the other three hybrid models is 16.70%, 42.79%, and 50.10%, respectively. In addition, all the R^2 values of the proposed model are over 90%, which underlines the higher fitting effect on Category II.

3) The forecasting metric of the single hybrid models and the proposed model in **Table 4** indicates that the proposed model based on model selection performs better than the single hybrid model in Category III. As an example, with respect to Tianjin, the DA values of the individual hybrid models are 80.84% (MODEGWO-SVM), 70.06% (MODEGWO-GRNN), and 77.84% (MODEGWO-BPNN), while the DA values of the proposed models is 87.24%, respectively. The comparative analysis between the proposed model and the single model confirms the advantages of the hybrid forecasting model.
4) Moreover, **Table 5** displays each metric of NO_2 forecasting among the developed hybrid forecasting system and the single hybrid models. According to **Table 5**, it is obvious that the values of MAE, RMSE, MAPE, U1, and U2 of the proposed hybrid model are all smaller than the other considered models, and the values of DA and R^2 of the developed hybrid forecasting system are all greater than that of the single hybrid model, which further confirms the superiority of the presented hybrid forecasting system in terms of forecasting ability.

In summary, from the analyses above, it can be concluded that the model selection forecasting system realizes the best forecasting results compared to the single hybrid model. Model selection also gives better forecasting performance in the other season with the results shown in **Supplementary Appendices S2–S4** indicating better robustness of the model selection forecasting system.

Experiment II: The Forecasting Results of Category II $PM_{2.5}$ Forecasting in the First Season

This experiment mainly focused on the forecasting performance of each model for $PM_{2.5}$ of Category II in the first season, with the forecasting results of four different hybrid models (MODEGWO-SVM, MODEGWO-BPNN, MODEGWO-ANFIS, Adam-LSTM) and model selection represented in **Table 6** and **Figure 4**

1) For first season $PM_{2.5}$ forecasting accuracy, the final forecast results of $PM_{2.5}$ for six cities in Category II are composed of four hybrid models, which include MODEGWO-SVM, MODEGWO-BPNN, MODEGWO-ANFIS, and Adam-LSTM. Among the four models, MODEGWO-SVM and MODEGWO-BPNN have better forecasting performance, with the MODEGWO-SVM obtaining 64.29% optimal forecasting points and the MODEGWO-BPNN obtaining 21.43% optimal points for six cities in Category II. The smallest MAPE values of MODEGWO-SVM are 0.92%, 0.94%, 1.36%, and 0.79% for Hengshui, Tangshan, Chengde, and Xingtai $PM_{2.5}$ forecasting, and the MODEGWO-BPNN obtains the best MAPE (1.08% and 0.85%) value for Shijiazhuang and Handan.
2) For the goodness of fit, the R^2 values of four different hybrid models are over 0.95 for six cities in the first season, which indicates that the forecasting values obtained by these models is close to the actual value. The forecasting result of Shijiazhuang shows that the R^2 value of best hybrid modes (MODEGWO-BPNN) is 0.9993, very close to 1, which indicates that there is less difference between forecasting data and actual data, and the forecasting value is basically consistent with the actual value. Meanwhile, The DA values of the best hybrid model are over 70%, which proves the best models can effectively capture the changing trend of the actual data.
3) For the forecasting results of model selection, **Table 6** and **Figure 4** clearly show that the forecasting performance of model selection is better than the hybrid model. The MAPE value of model selection is 0.88%; compared with the optimal hybrid model the MAPE improved 22.73%, and the MAE and RMSE reduced 26.49% and 30.69%, respectively. Although the model selection can improve the forecasting accuracy, in the $PM_{2.5}$ forecasting of some cities, due to the better forecasting performance of the single hybrid model, the forecasting accuracy of the model selection is not significantly improved. For example, in the $PM_{2.5}$ forecasting of Xingtai, the MAPE value of optimal forecasting model (MODEGWO-SVM) is 0.79%, and the MAPE of the model selection is 0.76%, which shows that forecasting accuracy has no significant improvement.
4) Similar to the first season, the $PM_{2.5}$ forecasting results of Category II in the second to fourth seasons are listed in **Supplementary Appendix Table 7** in which the best forecasting performances of the hybrid model are shown by DEGWO-SVM, DEGWO-BPNN, DEGWO-ANFIS, and Adam-LSTM for $PM_{2.5}$ forecasting in each city. Compared with the optimal hybrid model, the final forecasting results obtained by the model selection is more accurate than single hybrid model, which indicates that the optimal hybrid model has good forecasting performance but cannot be applied to all the forecasting data.

In summary, for the Category I and Category III $PM_{2.5}$ forecasting list in **Supplementary Appendices S5–S7**, the model selection forecasting system exhibits the best forecasting accuracy among the different hybrid models for four seasons. It is evident that the forecasting capacity of the model selection is robust when considering each forecasting metrics. The accuracy of model selection depends on the hybrid model, so it is necessary to increase the types of models in the modeling process which ensures that more forecasting results can be obtained, and the optimal forecasting value can be selected in the model selection process.

Experiment III: PM_{10} Forecasting Analysis for Category III in Four Seasons

For the hourly PM_{10} time series for three categories, it can be observed that the model selection forecasting system attains satisfactory results. Specifically, the lowest values of MAE are 0.4643, 0.4600, and 0.3869 and of RMSE are 0.7302, 0.7906, and 0.5561, corresponding to PM_{10} forecasting in Category I in three cities, successively. The results of Category I indicate that the smaller the MAE and MSE, the smaller the deviation between the

TABLE 7 | The forecasting result of each model in different seasons for Category III.

Location	Metric	First season			Second season					Third season				Fourth season			
		DEGWO-SVM	DEGWO-BPNN	Model selection	DEGWO-SVM	DEGWO-GRNN	DEGWO-BPNN	DEGWO-ANFIS	Model selection	DEGWO-SVM	DEGWO-BPNN	DEGWO-ANFIS	Model selection	DEGWO-SVM	DEGWO-BPNN	DEGWO-ANFIS	Model selection
Tianjin	MAE	0.4119	0.4687	0.3576	2.3288	3.1388	2.6986	6.2681	1.8811	0.7567	0.8285	1.1091	0.5870	0.6775	0.8685	1.0346	0.5747
	RMSE	0.5572	0.6236	0.4552	8.1840	7.7183	8.6510	25.5466	7.1780	0.9382	1.1333	1.5570	0.7880	0.8649	1.1016	1.4746	0.7649
	STDE	0.5472	0.6019	0.4535	8.2080	7.7255	8.6768	25.5589	7.1987	0.9259	1.1090	1.5400	0.7841	0.8531	1.0938	1.4463	0.7531
	DA	81.44%	74.25%	76.05%	76.05%	31.74%	64.67%	65.87%	76.05%	55.09%	55.09%	47.90%	68.26%	71.26%	50.30%	50.90%	71.26%
	U1	0.0037	0.0051	0.0037	0.0435	0.0412	0.0460	0.1343	0.0382	0.0064	0.0078	0.0107	0.0054	0.0063	0.0076	0.0101	0.0053
	U2	1.3126	1.4969	1.2192	1.9855	1.9275	2.0215	13.1964	1.6470	2.8029	1.8298	2.7405	1.7659	2.2110	2.2721	2.9704	2.1810
	MAPE	0.71%	0.83%	0.65%	2.65%	3.69%	3.09%	7.55%	2.12%	1.09%	1.13%	1.53%	0.84%	0.93%	1.19%	1.41%	0.83%
	R^2	0.9998	0.9996	0.9998	0.9601	0.9629	0.9534	0.7662	0.9681	0.9974	0.9965	0.9930	0.9982	0.9973	0.9969	0.9938	0.9983
Qinhuangdao	MAE	0.4140	0.5075	0.4034	0.7588	2.5923	1.0097	1.9170	0.7441	0.4645	0.6423	1.1009	0.4303	0.7377	0.8356	1.1088	0.6459
	RMSE	0.5527	0.6725	0.5440	1.2183	3.4536	1.6718	3.9976	1.2117	0.6622	0.8380	2.4277	0.5914	1.0205	1.2012	1.5364	0.9614
	STDE	0.5529	0.6736	0.5452	1.2219	3.4639	1.6274	3.9538	1.2146	0.6639	0.8317	2.3892	0.5932	1.0232	1.2029	1.5351	0.9642
	DA	82.04%	68.86%	83.23%	85.63%	32.93%	79.64%	72.46%	86.23%	81.44%	70.66%	72.46%	82.63%	79.04%	80.84%	64.67%	85.63%
	U1	0.0044	0.0054	0.0044	0.0061	0.0173	0.0083	0.0199	0.0061	0.0053	0.0067	0.0194	0.0047	0.0054	0.0063	0.0081	0.0050
	U2	1.1881	1.2933	1.1573	1.3515	4.7857	1.5998	4.0494	1.2920	1.6354	1.6295	5.0927	1.4787	1.4708	1.4926	1.9014	1.3827
	MAPE	0.81%	0.96%	0.75%	0.86%	3.22%	1.09%	2.09%	0.84%	0.96%	1.19%	2.18%	0.82%	1.00%	1.07%	1.49%	0.83%
	R^2	0.9997	0.9996	0.9997	0.9995	0.9964	0.9993	0.9959	0.9996	0.9994	0.9991	0.9925	0.9995	0.9996	0.9994	0.9991	0.9997
Zhangjiakou	MAE	0.6366	0.7970	0.4310	4.1069	7.4836	4.7017	4.8835	3.7104	0.4875	0.5652	0.8045	0.4782	0.7864	1.0681	1.7907	0.7297
	RMSE	0.7642	1.1575	0.5662	11.3281	12.9323	11.1572	12.2400	9.2301	0.6852	0.7945	1.2237	0.6836	1.0777	1.4880	2.8905	1.0128
	STDE	0.7655	1.1117	0.5666	11.3585	12.8990	11.1833	12.2350	9.2556	0.6797	0.7559	1.1982	0.6802	1.0638	1.4884	2.8853	0.9772
	DA	67.25%	62.87%	77.25%	83.83%	33.53%	74.25%	67.07%	82.04%	72.46%	70.06%	65.87%	72.46%	74.25%	66.47%	56.89%	75.45%
	U1	0.0099	0.0101	0.0049	0.0219	0.0251	0.0217	0.0237	0.0179	0.0042	0.0049	0.0075	0.0042	0.0066	0.0091	0.0177	0.0062
	U2	1.7496	2.5892	1.3923	1.6411	3.0821	1.5223	1.8703	1.3734	1.4487	1.7344	1.5603	1.4729	1.9104	2.1059	2.8750	1.8712
	MAPE	1.09%	1.93%	0.97%	2.27%	6.19%	2.51%	3.28%	2.06%	0.61%	0.72%	1.02%	0.59%	1.24%	1.73%	2.78%	1.18%
	R^2	0.9993	0.9991	0.9998	0.9979	0.9973	0.9980	0.9976	0.9986	0.9988	0.9985	0.9963	0.9988	0.9993	0.9986	0.9945	0.9994
Cangzhou	MAE	0.5447	0.5872	0.5178	4.8605	4.9067	4.8833	4.8570	2.0438	0.4579	0.5585	0.7233	0.4391	0.6380	0.8810	1.8394	0.6300
	RMSE	0.7422	0.7952	0.7083	7.2624	7.2642	7.2682	7.2725	3.8192	0.5983	0.7079	1.0983	0.5802	0.8898	1.2163	3.6287	0.8931
	STDE	0.7422	0.7963	0.7061	5.9796	5.9749	5.9578	5.9428	3.8275	0.5990	0.6926	1.0950	0.5813	0.8872	1.2080	3.6191	0.8812
	DA	81.44%	82.63%	83.23%	80.84%	73.89%	77.84%	80.84%	87.31%	80.24%	68.86%	64.07%	82.63%	85.03%	78.44%	68.86%	86.83%
	U1	0.0045	0.0048	0.0043	0.0293	0.0293	0.0294	0.0294	0.0127	0.0040	0.0048	0.0074	0.0039	0.0043	0.0059	0.0176	0.0043
	U2	1.3594	1.3911	1.3065	5.5390	5.6602	5.5440	5.5906	1.8692	1.4269	1.4604	1.5698	1.4152	1.4389	1.4227	2.7705	1.3837
	MAPE	0.72%	0.76%	0.67%	2.83%	2.90%	2.84%	2.81%	1.67%	0.65%	0.80%	0.99%	0.62%	0.66%	0.93%	1.83%	0.65%
	R^2	0.9996	0.9995	0.9996	0.9251	0.9288	0.9386	0.9472	0.9984	0.9992	0.9990	0.9975	0.9993	0.9994	0.9989	0.9904	0.9994

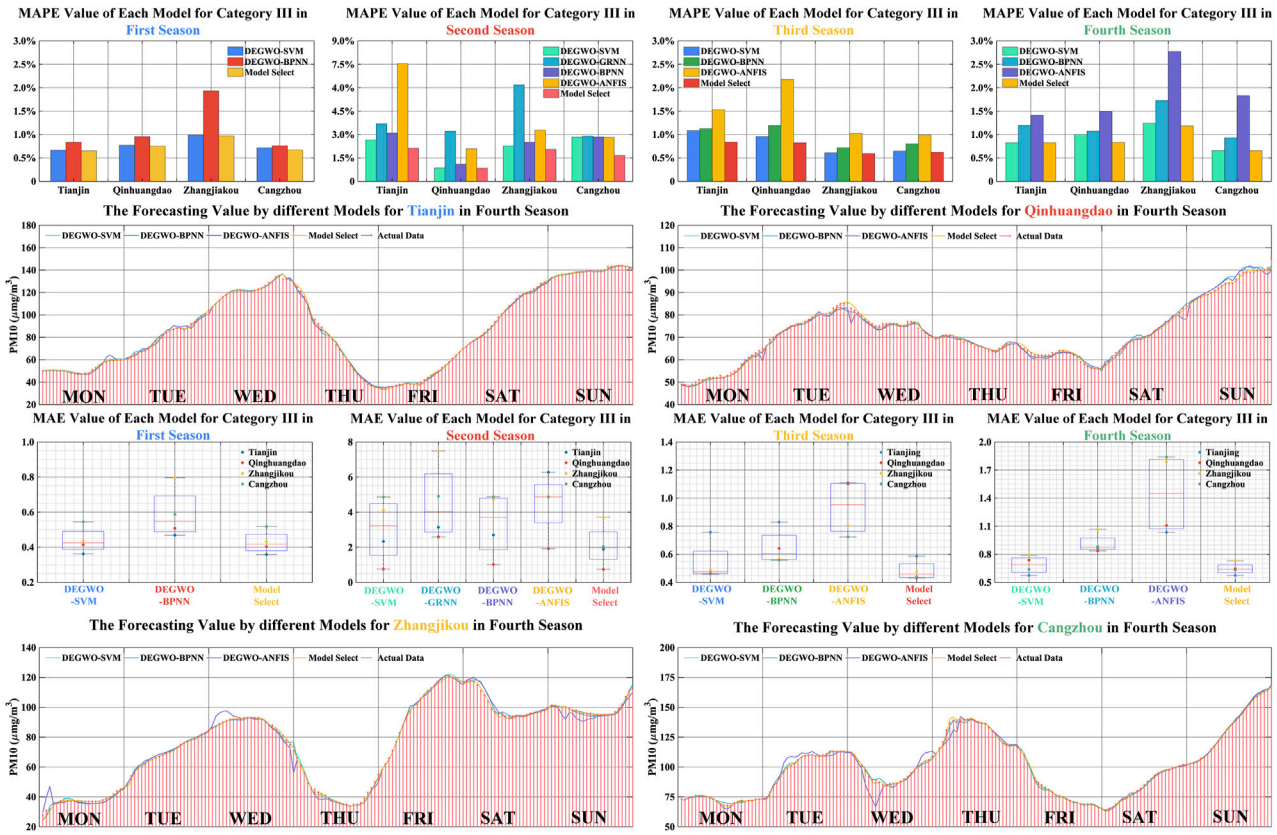

FIGURE 5 | The forecasting result of PM_{10} for Category III in different seasons.

observations and forecasting, which verifies the forecasting performance. In addition, the average MAPE values of model selection for six cities of Category II in four seasons are lower than 1%. Compared with the optimal hybrid model, the model selection is approximately reduced by 10%. The analyses reveal the forecasting superiority of the model selection system. By parity of reasoning, a similar conclusion can be reached through the analyses of the hourly PM_{10} forecasting results for Category III (the forecasting results are shown in **Table 7** and **Figure 5**).

1) For PM_{10} forecasting in the first season, the optimal hybrid models are DEGWO-SVM and DEGWO-BPNN, with which the MAPE values of the best hybrid model (DEGWO-SVM) for four cities of Category III are 0.71%, 0.81%, 1.09%, and 0.72%. The final forecasting values are obtained by model selection; based on the results of DEGWO-SVM and DEGWO-BPNN the MAPE values of model selection are 0.65%, 0.75%, 0.97%, and 0.67%. Additionally, the values of other forecasting metrics are at their best under the model selection.
2) In the second season of PM_{10} forecasting for four cities, the final forecasting results consist of four different hybrid models (DEGWO-SVM, DEGWO-GRNN, DEGWO-BPNN, and DEGWO-ANFIS). From the forecasting performance of four hybrid models, the forecasting performance of DEGWO-SVM is better than the other three models, in which the values of MAPE are 2.65%, 0.86%, 2.27%, and 2.83% for different cities' PM_{10} forecasting in Category III. Additionally, the DA value of MODEGWO-SVM is over 75%, which indicates that the hybrid model can capture future changing trends of PM_{10}. The final forecasting results are obtained by model selection, in which the forecasting accuracy is better than those of the four hybrid models. Compared with the best hybrid model, the MAPE values of model selection are reduced by 25.00%, 2.38%, 10.19%, and 69.46%, respectively.
3) According to forecasting results in **Table 7** and **Figure 5** for PM_{10} of the third season, the three kinds of hybrid models (DEGWO-SVM, DEGWO-BPNN, and DEGWO-ANFIS) are employed to forecast hourly PM_{10}; the DEGWO-SVM has the best forecasting performance among the three hybrid models in Zhangjiakou, and the MAPE is 0.61%. DEGWO-SVM has better forecasting accuracy, and model selection has little improvement on the forecasting accuracy in the final prediction results, but the MAPE has maximum improvement of 29.76% for Tianjin PM_{10} forecasting.
4) According to the results in **Table 6**, the three kinds of hybrid models are used to forecast PM_{10} for four cities of Category III in the fourth season, and the R^2 value of each model was greater than 0.99, which shows that these models have a good forecasting performance for the PM_{10}. Meanwhile, model

selection uses the predicted values of each model to form the final forecasting results, and the corresponding MAE values are 0.5747, 0.6459, 0.7297, and 0.6300 for four cities.

In summary, whether for Category III or the other categories (the results are shown in **Supplementary Appendix S8** and **Supplementary Appendix S9**) PM_{10} forecasting, the model selection system attained the best performance for 13 cities. In the comparison of various hybrid models, the forecasting performance of MODEGWO-SVM is better than other hybrid models. Additionally, it can be observed that some models have low accuracy, which can still provide some optimal forecasting values for the model selection system for hourly PM_{10} forecasting. Based on the above analysis, it can be seen that none of the models has been playing the best forecasting performance in the forecasting process, and various hybrid models are needed to make up for the shortcomings of the single hybrid model.

DISCUSSIONS

This section mainly discusses the hyperparameter related to the SVM and ANN model that would influence the forecasting performance. A large variety of machine learning models and ANNs are available for air pollution time series including three different type pollutants. Finally, compare computing in different model.

Support Vector Machine

According to the results of each experiment, SVM provided the more optimal forecast values for the three main pollutants in the four quarters of 13 cities. The reason for the favorable score produced by SVM is that SVM provides a way to avoid the complexity of high-dimensional space by directly using the inner product function of the space (which is the kernel function) and then directly solving the corresponding decision-making problem in high-dimensional space by using the solution method under the condition of linear separability. When the kernel function is known, it can simplify the difficulty of solving the problem in high-dimensional space. Meanwhile, SVM is based on the small sample statistical theory, which conforms to machine learning. In addition, support vector machine has better generalization ability than neural network. Although the time series of the three main air pollutants are neither regular nor seasonal, SVM can also effectively capture future changes of the three main air pollutants.

SVM has two very important parameters: c and g. c is the penalty coefficient, that is, tolerance of errors. The higher the c, the more intolerable the errors and easy to over-fit. The smaller the c, the less easy fitting is. If c is too large or too small, the generalization ability becomes worse. g is a parameter that comes with RBF function when it is selected as a kernel. Implicitly it determines the distribution of data after mapping to a new feature space. The bigger the g, the less support vectors; it will only act near the support vector samples. For the unknown samples, the classification effect is very poor. There is a possibility that the training accuracy can be very high, and the test accuracy is not high, that is, over-fitting. The smaller the g value is, the more support vectors there are, and the greater the smoothing effect will be; the higher accuracy of the training processing cannot be obtained, and the accuracy of the testing processing will also be affected. This paper used DEGWO algorithm and GWO algorithm to optimize the parameters of SVM (c and g). The results of two types of hybrid SVMs are shown in **Table 8**, which displays that the optimum penalty coefficients of SVM corresponding to pollutant forecasting in different cities vary widely. For example, in the forecasting processing of NO_2, the variation range of parameters is [2, 99]. However, the fluctuation range of g is small, with most variations ranging from 0 to 1. In general, the air pollutants forecasting performances of support vector machine are very dependent on the penalty coefficient. In the whole experiment it can be observed that the support vector machine has good forecasting accuracy for three main air pollutants forecasting, but it cannot provide the best forecasting value in each point. It indicates that the support vector machine is suitable for hourly air pollutants forecasting.

Artificial Neural Network: Number of Input Layer and Number of Hidden Layers

ANN as a nonlinear mapping model is used to solve the problem of time series forecasting, because the ANN model can find the optimal solution of a complex problem with the help of high-speed computing ability of the computer. In order to ensure the forecasting accuracy of the ANN model, parameters of ANN need to be elaborately configured. However, there is no effective rule for establishing the values of these parameters on air pollutants forecasting. Although there are many studies on the tuning of the parameters of the neural network, it is obvious that the selection of the whole parameter space is beyond the scope of this study. Therefore, the parameters of the neural network are set by means of simulation experiment and optimization algorithm: the experimental design is as follows:

This processing configured various input layers and a number of hidden layers to find out the influence of the usage of recent data on the performance of different ANN models. The number of input layers from 1 to 10 increases for three main air pollutants, which means there are 1,008 pieces of sample data on NO_2, $PM_{2.5}$, and PM_{10}; the train-to-verify ratio 5:1 means that 840 pieces of sample data were used as training data for building the ANN model, while 168 pieces of sample data were used as testing data for finding the training-to-testing ratio and parameter of each ANN model (the optimal number of input layers of each model and the number of hidden layers of LSTM and BPNN). **Figure 6** shows the forecasting performance with the different configurations of the optimal number of input layers and the number of hidden layers of LSTM and BPNN, in which it is difficult to find a regular correlation between the forecasting performances and the parameters. Consequently, it is difficult to find an optimal combination of ANN's parameters that brings the model to the best performance in the practical air pollutant forecasting where MAPE and R^2 are unknown.

During the experiment on configurations of ANN's parameter, the optimal parameter setting trained the networks models for each ANN model in 13 cities. The forecasting values with the best

TABLE 8 | The result of two types of SVM for three main air pollutants in different cities.

Air Pollution	Parameter	Beijing	Tianjin	Shijiazhuang	Tangshan	Qinhuangdao	Handan	Baoding	Zhangjiakou	Chengde	Langfang	Cangzhou	Hengshui	Xingtai
DEGWO-SVM (NO_2)	best.*C*	46.9568	99.5518	99.309	99.161	99.8544	9.6662	83.7412	3.5907	16.9145	2.6853	49.9809	99.1334	4.3575
	best.*g*	0.9694	0.0763	0.1945	0.3663	0.0393	0.5403	0.0101	4.0931	0.0105	1.4979	0.0105	0.4098	0.4622
	SSE	$1.74*10^3$	$1.92*10^3$	$3.82*10^3$	$7.76*10^3$	$5.30*10^3$	$7.12*10^3$	$8.78*10^3$	$2.09*10^3$	$3.06*10^3$	$8.96*10^3$	$1.27*10^4$	$7.37*10^3$	$5.62*10^3$
	min.E	0.0338	0.0186	0.0061	0.0336	0.0002	0.0446	0.0216	0.0127	0.0355	0.0077	0.0494	0.0305	0.0259
	max.E	10.7395	13.5376	36.7462	27.3748	20.6095	41.3333	38.0069	13.0612	18.5541	44.2883	34.1065	27.4761	24.1837
	Var.E	10.3228	11.1665	22.8619	46.4519	31.2395	42.479	52.0663	12.3554	18.01	53.4352	75.7013	43.7583	33.6634
	Time	467.0666	493.4147	487.456	503.9043	510.1607	497.2697	474.2788	495.2658	499.3622	474.5837	490.1128	478.4362	492.8458
GWO-SVM (NO_2)	best.*C*	78.5905	324.3692	210.5643	31.6254	212.0154	10.2694	200.0915	33.5771	17.2807	278.5659	50.1119	169.9275	401.8597
	best.*g*	0.9669	0.0765	0.1961	0.3913	0.0395	0.5486	0.01	3.9411	0.1216	1.4618	0.0105	0.4118	0.2077
	SSE	$1.86*10^3$	$1.94*10^3$	$4.12*10^3$	$7.96*10^3$	$5.70*10^3$	$7.67*10^3$	$9.29*10^3$	$2.10*10^3$	$3.33*10^3$	$9.74*10^3$	$1.32*10^4$	$7.90*10^3$	$5.76*10^3$
	min.E	0.0368	0.0187	0.0062	0.0367	0.0003	0.049	0.023	0.0134	0.0383	0.0083	0.0507	0.0321	0.0283
	max.E	11.3172	13.5905	38.0581	29.5744	20.7387	44.7556	38.7649	13.1132	19.3	45.5339	35.1327	28.8798	25.4748
	Var.E	10.612	11.2675	23.2602	50.7895	32.7046	46.1177	53.6559	13.2875	18.7424	56.4917	78.4229	46.1941	34.9107
	Time	414.2995	409.8461	397.1843	439.9873	449.712	434.443	409.6929	405.8517	405.4657	426.6385	438.0274	395.0127	436.3719
DEGWO-SVM ($PM_{2.5}$)	best.*C*	77.7686	0.079	0.1832	32.0261	0.697	4.7476	47.0616	88.2993	99.8909	2.7945	1.4898	0.0972	5.0422
	best.*g*	5.0885	2.6319	7.6937	0.01	3.736	0.01	0.01	0.016	0.0134	4.7102	3.5942	1.9463	0.01
	SSE	$8.67*10^2$	$4.91**10^2$	$7.68*10^2$	$2.42*10^2$	$1.10*10^2$	$1.61*10^3$	$1.71*10^3$	50.63	$1.63*10^2$	$7.55*10^2$	$1.69*10^3$	$1.31*10^3$	$5.85*10^2$
	min.E	0.024	0.0079	0.0308	0.0114	0.0008	0.0034	0.0016	0.0011	0.0039	0.0182	0.0711	0.0363	0.0053
	max.E	4.3326	4.8468	7.4561	3.3869	2.1798	9.1996	9.9057	1.5044	2.5344	5.0337	8.2179	8.8812	6.0958
	Var.E	3.6799	2.544	4.0871	1.4022	0.6552	7.9582	10.6013	0.2743	0.6602	3.8421	9.8701	6.7322	3.4501
	Time	78.5201	79.9691	79.8954	80.0222	81.0965	85.7327	85.5043	80.7387	85.0065	80.3694	84.4306	83.4787	83.4082
GWO-SVM ($PM_{2.5}$)	best.*C*	40.8136	93.0752	8.4678	31.86	60.0461	4.7539	52.6286	87.9288	100	92.1335	17.8947	41.9715	59.3705
	best.*g*	5.0366	1.397	8.7715	0.01	3.7298	0.01	0.01	0.0165	0.0134	4.7102	3.6199	3.306	0.01
	SSE	$8.83*10^2$	$8.08*10^2$	$8.20*10^2$	$2.43*10^2$	$1.11*10^2$	$1.63*10^3$	$1.77*10^3$	50.61	$1.57*10^2$	$7.55*10^2$	$1.68*10^3$	$2.63*10^3$	$6.07*10^2$
	min.E	0.0086	0.0136	0.0147	0.0106	0.0039	0.0263	0.0038	0.0038	0.0013	0.0182	0.0474	0.0024	0.0018
	max.E	4.2931	6.2686	8.4904	3.422	2.1569	9.2235	9.9212	1.5028	2.4958	5.0337	8.1801	7.6894	6.0805
	Var.E	3.7291	4.6545	4.282	1.4157	0.6563	7.9943	10.5457	0.2739	0.6459	3.8421	9.753	9.6812	2.0456
	Time	63.6136	65.0956	67.7701	71.8366	67.2112	70.9237	68.4242	68.8076	68.1498	69.0608	74.1707	71.644	68.8802
DEGWO-SVM (PM_{10})	best.*C*	3.5114	18.3746	4.8387	61.0409	56.5602	0.847	83.9649	51.5595	5.1477	19.6408	12.4613	1.2529	0.553
	best.*g*	4.7786	7.0919	0.01	0.0366	0.0187	0.3343	0.0103	0.01	0.0563	5.9592	3.3231	0.4991	4.3841
	SSE	$2.58*10^3$	$9.35*10^2$	$8.17*10^3$	$1.49*10^3$	$8.61*10^2$	$2.98*10^3$	$6.80*10^3$	$8.81*10^2$	$1.69*10^3$	$1.71*10^3$	$1.72*10^3$	$1.67*10^3$	$1.46*10^3$
	min.E	0.0176	0.0076	0.0036	0.0026	0.0005	0.0778	0.1005	0.0003	0.1025	0.0015	0.0039	0.03	0.0134
	max.E	10.9659	8.4861	14.0939	8.4406	5.9929	10.8543	15.9743	6.9388	7.4138	10.0266	10.3846	7.3109	8.9293
	Var.E	10.2272	5.1871	28.3304	8.9646	5.1568	17.5229	33.6806	3.9992	7.9292	5.211	14.8334	9.6962	8.9122
	Time	81.873	85.2247	85.6066	84.1662	79.4368	80.1318	80.0802	86.2987	82.036	83.7506	82.7005	81.4053	85.2475
GWO-SVM (PM_{10})	best.*C*	5.9521	18.4212	0.6008	56.4077	82.7556	0.8624	88.0444	51.7353	5.124	67.6243	5.4252	1.2881	9.1351
	best.*g*	4.7723	7.0158	0.0101	0.0367	0.0187	0.3284	0.0101	0.0101	0.0565	5.9744	0.01	0.4842	0.01
	SSE	$2.61*10^3$	$9.43*10^2$	$8.30*10^3$	$1.50*10^3$	$8.65*10^2$	$3.00*10^3$	$6.88*10^3$	$9.74*10^2$	$1.69*10^3$	$1.72*10^3$	$3.11*10^3$	$1.68*10^3$	$1.51*10^3$
	min.E	0.0249	0.0083	0.0207	0.0192	0.0528	0.081	0.1022	0.079	0.1026	0.0071	0.0256	0.066	0.0899
	max.E	10.9762	8.5423	14.7419	8.4423	6.0089	10.9071	16.0619	6.9653	7.416	10.0514	11.5914	7.4373	12.3556
	Var.E	10.2477	5.2718	38.3341	8.8981	5.1728	17.469	33.8092	4.1143	7.9363	5.2235	10.2618	9.5788	8.6932
	Time	73.6691	70.3273	73.1108	72.0992	69.9234	64.3441	65.0347	75.2683	71.9299	72.8245	69.424	65.4326	71.6009

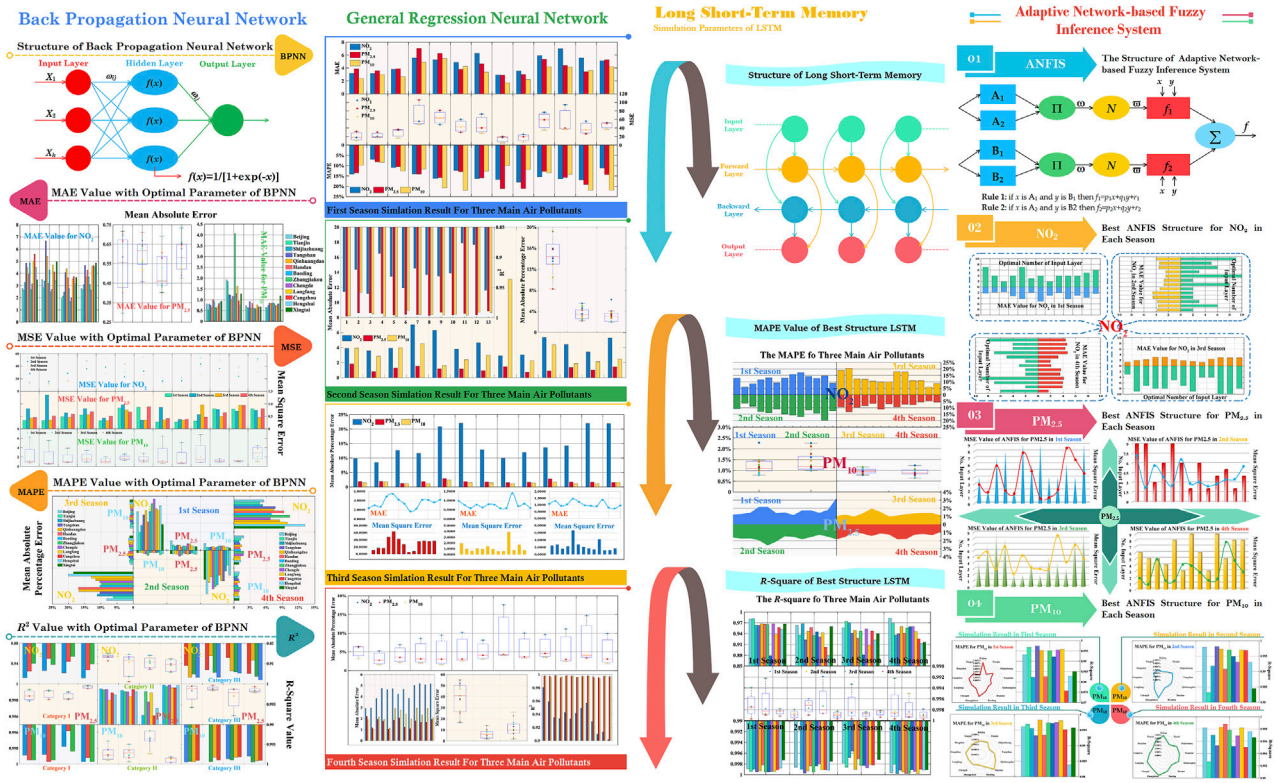

FIGURE 6 | The simulation result of each ANN model for three main air pollutants in different cities.

performance of each ANN (the best forecasting metric in **Table 9**) were selected to forecast three main air pollutants in each experiment. However, it can be found that there were no giant differences on forecasting performance among the networks trained with the same configuration, even if the neural network has the randomness and probability mechanism inside the training processing. A large sample of the times series is another reason that the training stability of the neural network can be ensured. For example, the best MAPE of BPNN for the forecasting of NO_2 in Beijing shown in **Figure 6** is 5.82%, and the worst one is 8.94%. Most of the MAPEs are between 6% and 7%. It is practical to use ANN in real air pollutants forecasting application where forecasting the changing air pollutant time series is suitable.

In summary, with the rapid development of ANN, it has become a powerful tool to solve prediction problems. Neural network is used in the field of air pollution to solve the problem of non-linear forecasting which cannot be solved by statistical models. Its non-linear mapping is especially suitable for the application of air pollutant forecasting. The main reason is that the ANN has the following advantages:

1) Non-linear mapping ability: ANN realizes a mapping function from input to output in essence. Mathematical theory proves that three-layer neural network can approximate any non-linear continuous function with arbitrary precision. This makes it especially suitable for solving complex internal mechanism problems, that is, ANN has strong non-linear mapping ability.
2) Self-learning and self-adapting ability: ANN can automatically extract the "reasonable rules" between output and output data by learning and self-adaptively memorizing the learning content in the weights of the network. ANN has a high ability of self-learning and self-adaptation.
3) Generalization ability: When designing pattern classifiers, the so-called generalization ability refers to whether the network can forecast the unknown time series correctly after training, while ensuring that the classified objects are correctly classified. ANN can apply learning results to new knowledge.
4) Fault-tolerant ability: ANN will not have a great impact on the global training results after its local or partial neurons are damaged; the system can work normally even when it is damaged locally. ANN has certain fault-tolerant ability.

Computing Time for Each Model

In order to improve the computing efficiency and save the computing time, training and forecasting processing of all the models for the main air pollutants time series with parallel computing by central processing unit (CPU) and graphics processing unit (GPU). The computing times of every independent hybrid model in each experiment are shown in **Table 10**, from which we can further research the computational efficiency of the developed model selection forecasting system for the main air pollutants. Specifically, the

TABLE 9 | The simulation result of each ANN model.

Model	BPNN						ANFIS				
Metric	**NO. Input**	**NO. Hidden**	**MAE**	**MSE**	**MAPE**	R^2	**NO. Input**	**MAE**	**MSE**	**MAPE**	R^2
Beijing	7	29	2.7515	13.4095	12.82%	0.9825	7	2.2208	14.1852	11.61%	0.9816
Tianjin	4	21	2.6124	12.3784	6.07%	0.9830	4	1.8908	18.6934	5.00%	0.9747
Shijiazhuang	7	23	3.2795	28.5628	8.32%	0.9628	2	1.7853	9.7964	5.20%	0.9876
Tangshan	6	15	4.8308	49.0277	11.76%	0.9583	4	3.2171	25.9431	8.18%	0.9779
Qinhuangdao	8	24	4.4735	38.3830	13.34%	0.9672	7	4.2183	33.3884	14.23%	0.9722
Handan	10	21	3.9385	30.7437	9.47%	0.9602	3	1.9940	16.6067	5.23%	0.9788
Baoding	9	26	4.8845	54.5972	12.89%	0.9626	4	5.1107	52.7317	16.64%	0.9637
Zhangjiakou	2	21	2.6965	14.1237	15.52%	0.8956	2	3.0306	16.3035	19.68%	0.8805
Chengde	5	21	3.1457	21.7384	17.12%	0.9642	4	2.0878	12.3017	13.23%	0.9799
Langfang	4	21	5.0108	53.2976	13.89%	0.9393	4	4.2474	37.6082	13.43%	0.9602
Cangzhou	3	24	5.5917	58.3720	13.31%	0.9547	4	3.0234	31.2348	9.23%	0.9759
Hengshui	7	23	4.5165	37.0716	14.85%	0.9482	5	3.7816	33.4263	11.99%	0.9550
Xingtai	9	21	3.4928	30.4963	8.57%	0.9702	6	2.4056	28.1033	5.40%	0.9725
Model	**LSTM**						**GRNN**				
Metric	**NO. Input**	**NO. Hidden**	**MAE**	**MSE**	**MAPE**	R^2	**NO. Input**	**MAE**	**MSE**	**MAPE**	R^2
Beijing	8	27	2.8379	15.0298	14.37%	0.9805	2	3.1653	17.7058	14.09%	0.9772
Tianjin	9	9	2.6240	13.6968	6.72%	0.9812	8	3.1290	20.3228	6.97%	0.9720
Shijiazhuang	3	30	3.1771	28.2116	8.22%	0.9634	9	3.8065	34.9429	10.99%	0.9577
Tangshan	5	21	4.7107	43.2069	10.36%	0.9630	10	5.5704	55.6046	14.00%	0.9522
Qinhuangdao	4	30	4.5057	37.7889	13.13%	0.9673	7	5.2814	47.7487	16.22%	0.9590
Handan	6	29	3.6311	25.9993	8.01%	0.9663	2	4.8767	60.1796	12.32%	0.9271
Baoding	8	9	4.7872	50.7279	11.98%	0.9651	5	6.2932	72.5647	16.31%	0.9538
Zhangjiakou	7	29	2.7874	17.0202	18.70%	0.8783	8	2.9293	17.9413	16.74%	0.8711
Chengde	1	29	3.1558	20.0247	15.77%	0.9673	3	3.5448	25.0984	18.05%	0.9588
Langfang	10	30	5.0278	53.3276	13.90%	0.9383	10	5.9260	76.8659	15.51%	0.9170
Cangzhou	8	30	5.3683	56.0168	12.77%	0.9570	7	6.9785	94.2087	16.45%	0.9265
Hengshui	10	27	4.3802	36.2420	14.52%	0.9523	6	5.5657	56.0395	16.84%	0.9239
Xingtai	4	27	3.8483	33.5689	9.43%	0.9670	9	5.1267	52.2457	11.13%	0.9551

TABLE 10 | Computing time by each model.

Category	Model	Min	Max	Average	Model	Min	Max	Average
Category I (NO_2)	MODEGWO-SVM	39.4238	48.1795	43.9663	**First season for Category III (PM_{10})**			
	MODEGWO-GRNN	65.7006	80.2975	72.8432	DEGWO-SVM	32.49657	39.71566	36.03983
	MODEGWO-BPNN	65.7133	80.2837	72.7285	DEGWO-BPNN	54.17679	66.18566	60.13886
	MODEGWO-ANFIS	39.4207	48.1762	43.7417	Model selection	0.108317	0.132383	0.120239
	Model select	0.1314	0.1605	0.1460	Final time	87.15716	105.6679	96.29893
	Final time	214.8663	251.5786	233.4256	**Second season for Category III (PM_{10})**			
Category II (NO_2)	MODEGWO-SVM	44.5487	54.4188	49.4898	DEGWO-SVM	43.09741	52.62296	47.78303
	MODEGWO-GRNN	74.2109	90.6824	82.7233	DEGWO-GRNN	71.80499	87.73968	79.55603
	MODEGWO-BPNN	74.2103	90.7002	82.6234	DEGWO-BPNN	43.08113	52.63649	48.01184
	Adam-LSTM	133.5967	163.2337	148.2152	DEGWO-ANFIS	71.80864	87.74184	79.77636
	Model select	0.1484	0.1813	0.1650	Model selection	0.143618	0.175492	0.159656
	Final time	330.6037	393.0909	363.2167	Final time	235.9748	276.3926	255.2869
Category III (NO_2)	MODEGWO-SVM	41.5062	50.7174	46.0609	**Third season for Category III (PM_{10})**			
	MODEGWO-GRNN	69.1727	84.5312	76.9191	DEGWO-SVM	37.79408	46.17638	42.03216
	MODEGWO-BPNN	69.1691	84.5052	76.7223	DEGWO-BPNN	62.98502	76.9619	70.11255
	Model select	0.1383	0.1689	0.1537	DEGWO-ANFIS	75.59966	92.35553	83.84003
	Final time	183.1891	218.0664	199.8560	Model selection	0.125959	0.15388	0.139823

(Continued on following page)

TABLE 10 | (*Continued*) Computing time by each model.

Category	Model	Min	Max	Average	Model	Min	Max	Average
Category	Model	Min	Max	Average	Final time	178.6782	212.5414	196.1246
Category II ($PM_{2.5}$)	MODEGWO-SVM	42.00122	51.30858	46.52758	**Fourth season for Category III (PM_{10})**			
	MODEGWO-BPNN	69.98488	85.5151	77.81983	DEGWO-SVM	37.68701	46.0525	41.96226
	MODEGWO-ANFIS	83.97059	102.6216	93.34656	DEGWO-BPNN	62.81456	76.76278	69.93992
	Adam-LSTM	125.9622	153.9263	139.7002	DEGWO-ANFIS	75.37328	92.07837	83.83009
	Model select	0.139947	0.171026	0.155573	Model selection	0.125629	0.153373	0.139332
	Final time	326.5897	387.7865	357.5498	Final time	176.9772	213.5771	195.8716

average computation time of the model selection forecasting system ranges from 330.6037 to 363.2167 s for NO_2 forecasting in Category II, with the longest computing time appearing in the different categories. Notably, Adam-LSTM with complex model structure has longer computing time than the other hybrid models, taking more time in the iterative optimization process. Moreover, this paper establishes multiple hybrid models and uses the model selection method to find the best forecasting value, in which the final forecasting accuracy is improved but needs more computing time.

CONCLUSION

In this study, a novel model selection forecast system was proposed that overcomes the shortcomings of the single hybrid model, which cannot give the optimal results for the forecasting process. First, the FSE theory is employed to analyze the major pollutant for each city in BJ-TJ-HE, and the fuzzy c-means algorithm is used to analyze the feature of the 13 cities. Then, to further improve modeling accuracy and rationality of modeling, a modified optimization algorithm (DEGWO) was used to optimize the premasters of different models. Finally, the model selection forecasting system obtains forecasting results at each time point from different hybrid models.

The developed model selection forecasting system was evaluated on hourly NO_2, $PM_{2.5}$, and PM_{10} from 13 cities, and several performance metrics were calculated, with experimental results indicating that the model selection forecasting system is superior to single hybrid models with the smallest MAPE in the different cities pollutant forecasting, indicating its strong forecasting performance. Overall, the proposed model selection forecast system exhibits outstanding performance in data analysis and time series forecasting for air pollutants. Specifically, it can not only deeply analyze major pollutants of AQI for BJ-TJ-HE but also approximate the actual values with high accuracy and stability.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

The design experiment, data analysis, and paper writing were conducted by YH; the forecasting experiment and data analysis were completed by YH and CW; supervision, paper writing, and editing were conducted by CW and YD; validation, methodology, paper editing, and supervision were handled by QL and GZ. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by Western Project of the National Social Science Foundation of China (Grant No. 18XTJ003).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2021.761287/full#supplementary-material

REFERENCES

Akyüz, M., and Çabuk, H. (2009). Meteorological Variations of PM2.5/PM10 Concentrations and Particle-Associated Polycyclic Aromatic Hydrocarbons in the Atmospheric Environment of Zonguldak, Turkey. *J. Hazard. Mater.* 170, 13–21. doi:10.1016/j.jhazmat.2009.05.029

Bessagnet, B., Couvidat, F., and Lemaire, V. (2019). A Statistical Physics Approach to Perform Fast Highly-Resolved Air Quality Simulations – A New Step towards the Meta-Modelling of Chemistry Transport Models. *Environ. Model. Softw.* 116, 100–109. doi:10.1016/j.envsoft.2019.02.017

Brereton, R. G., and Lloyd, G. R. (2010). Support Vector Machines for Classification and Regression. *Analyst* 135, 230–267. doi:10.1039/b918972f

D'Allura, A., Kulkarni, S., Carmichael, G., Finardi, S., Adhikary, Wei, et al. (2011). Meteorological and Air Quality Forecasting Using the WRF–STEM Model during the 2008 ARCTAS Field Campaign. *Atmos. Environ.* 45. doi:10.1016/j.atmosenv.2011.02.073

Dass, A., Srivastava, S., and Chaudhary, G. (2021). Air Pollution: A Review and Analysis Using Fuzzy Techniques in Indian Scenario. *Environ. Technology Innovation* 22, 101441. doi:10.1016/j.eti.2021.101441

Davood, N.-K., Goudarzi, G., Taghizadeh, R., Asumadu-Sakyi, A., and Fehresti-Sani, M. (2021). Long-term Effects of Outdoor Air Pollution on Mortality and

Morbidity–Prediction Using Nonlinear Autoregressive and Artificial Neural Networks Models. *Atmos. Pollut. Res.* 12, 46–56. doi:10.1016/j.apr.2020.10.007

Dhiman, G., and Kumar, V. (2018). Multi-objective Spotted Hyena Optimizer: A Multi-Objective Optimization Algorithm for Engineering Problems. *Knowledge-Based Syst.* 150, 175–197. doi:10.1016/j.knosys.2018.03.011

Díaz-Robles, L. A., Ortega, J. C., Fu, J. S., Reed, G. D., Chow, J. C., Watson, J. G., et al. (2008). A Hybrid ARIMA and Artificial Neural Networks Model to Forecast Particulate Matter in Urban Areas: The Case of Temuco, Chile. *Atmos. Environ.* 42, 8331–8340. doi:10.1016/j.atmosenv.2008.07.020

Feng, X., Li, Q., Zhu, Y., Hou, J., Lingyan, J., and Wang, J. (2015). Artificial Neural Network Forecasting of PM2.5 Pollution Using Air Mass Trajectory Based Geographic Model and Wavelet Transformation. *Atmos. Environ.* 107, 118–128. doi:10.1016/j.atmosenv.2015.02.030

Gayen, S., and Biswas, A., (2021). Pythagorean Fuzzy C-Means Clustering Algorithm. In *Computational Intelligence in Communications and Business Analytics Cham*. Editors J. K. Mandal and S. Mukhopadhyay.

Güçlü, Y. S., Dabanlı, İ., Şişman, E., and Şen, Z. (2019). Air Quality (AQ) Identification by Innovative Trend Diagram and AQ index Combinations in Istanbul Megacity. *Atmos. Pollut. Res.* 10, 88–96. doi:10.1016/j.apr.2018.06.011

Han, Y., Lam, J. C., Li, V. O., and Reiner, D. (2021). A Bayesian LSTM Model to Evaluate the Effects of Air Pollution Control Regulations in Beijing, China. *Environ. Sci. Pol.* 115, 26–34. doi:10.1016/j.envsci.2020.10.004

Hao, Y., Niu, X., and Wang, J. (2021). Impacts of Haze Pollution on China's Tourism Industry: A System of Economic Loss Analysis. *J. Environ. Manag.* 295, 113051. doi:10.1016/j.jenvman.2021.113051

Hao, Y., and Tian, C. (2018). The Study and Application of a Novel Hybrid System for Air Quality Early-Warning. *Appl. Soft Comput.*, 74.

Hao, Y., Tian, C., and Wu, C. (2019). Modelling of Carbon price in Two Real Carbon Trading Markets. *J. Clean. Prod.* 244, 118556.

Jang, J.-S. R. (1993). Anfis: Adaptive-Network-Based Fuzzy Inference System. *IEEE Trans. Syst. Man. Cybern.* 23, 665–685. doi:10.1109/21.256541

Jiang, P., Dong, Q., and Li, P. (2017). A Novel Hybrid Strategy for PM 2.5 Concentration Analysis and Prediction. *J. Environ. Manage.* 196, 443–457. doi:10.1016/j.jenvman.2017.03.046

Kim, J., Kim, D., Choi, M., and Choi, M. (2018). Impact of Air Pollution on Cause-specific Mortality in Korea: Results from Bayesian Model Averaging and Principle Component Regression Approaches. *Sci. total Environ.* 636, 1020–1031.

Kumar, K., Zindani, D., and Davim, J. P. (2019). Particle Swarm Optimization Algorithm. *Optimizing Engineering Problems through Heuristic Techniques*. 23–32. doi:10.1201/9781351049580-3

Land, W. H., and Schaffer, J. D. (2020). The Generalized Regression Neural Network Oracle. *The Art and Science of Machine Intelligence*, 77–105. doi:10.1007/978-3-030-18496-4_3

Lanzafame, R., Monforte, P., Patanè, G., and Strano, S. (2015). Trend Analysis of Air Quality Index in Catania from 2010 to 2014. *Energ. Proced.* 82, 708–715. doi:10.1016/j.egypro.2015.11.796

Latif, M. T., Ahamad, F., Abdul Halim, N. D., Khan, M., Juneng, L., Chung, J. X., et al. (2018). The Long-Term Assessment of Air Quality on an Island in Malaysia. *Heliyon* 4, 1–33.

Lauriks, T., Longo, R., Baetens, D., Derudi, M., Parente, A., Bellemans, A., et al. (2020). Application of Improved CFD Modeling for Prediction and Mitigation of Traffic-Related Air Pollution Hotspots in a Realistic Urban Street. *Atmos. Environ.* 246, 118127. doi:10.1016/j.atmosenv.2020.118127

Leong, W. C., Kelani, R. O., and Ahmad, Z. (2019). Prediction of Air Pollution index (API) Using Support Vector Machine (SVM). *J. Environ. Chem. Eng.* 8, 103208.

Li, W. k., Wang, W., Wang, Z., and Li, L. (2019). Opposition-based Multi-Objective Whale Optimization Algorithm with Global Grid Ranking. *Neurocomputing* 5. doi:10.1016/j.neucom.2019.02.054

Li, Z., Fung, J. C. H., and Lau, A. K. H. (2018). High Spatiotemporal Characterization of On-Road PM2.5 Concentrations in High-Density Urban Areas Using mobile Monitoring. *Building Environ.* 143, 196–205. doi:10.1016/j.buildenv.2018.07.014

Liu, H., Wu, H., Lv, X., Ren, Z., Liu, M., Li, Y., et al. (2019). An Intelligent Hybrid Model for Air Pollutant Concentrations Forecasting: Case of Beijing in China. *Sustainable Cities Soc.* 47, 101471. doi:10.1016/j.scs.2019.101471

Liu, Q., Wu, L., Xiao, W., Wang, F., and Zhang, L. (2018). A Novel Hybrid Bat Algorithm for Solving Continuous Optimization Problems. *Appl. Soft Comput.* 73. doi:10.1016/j.asoc.2018.08.012

Liu, Y., Guo, H., Mao, G., and Yang, P. (2008). A Bayesian Hierarchical Model for Urban Air Quality Prediction under Uncertainty. *Atmos. Environ.* 42, 8464–8469. doi:10.1016/j.atmosenv.2008.08.018

Lu, D.-X., Weng, W.-Y., Su, J., Wang, Z.-B., and Yang, X.-J. (2011). An Improved Fuzzy Synthetically Evaluation 163, 214–222. doi:10.1007/978-3-642-25002-6_30

Muzaffar, S., and Afshari, A. (2019). Short-Term Load Forecasts Using LSTM Networks. *Energ. Proced.* 158, 2922–2927. doi:10.1016/j.egypro.2019.01.952

Pan, L., Sun, B., and Wang, W. (2011). City Air Quality Forecasting and Impact Factors Analysis Based on Grey Model. *Proced. Eng.* 12, 74–79. doi:10.1016/j.proeng.2011.05.013

Pereira, A., Patrício, B., Fonte, F., Marques, S., Reis, C. I., and Maximiano, M. (2018). Collecting Information about Air Quality Using Smartphones. *Proced. Computer Sci.* 138, 33–40. doi:10.1016/j.procs.2018.10.006

Rivas, E., Santiago, J., Lechon, Y., Martin, F., Ariño, A., Pons Izquierdo, J., et al. (2018). CFD Modelling of Air Quality in Pamplona City (Spain): Assessment, Stations Spatial Representativeness and Health Impacts Valuation. *Sci. Total Environ.* 649, 1362–1380. doi:10.1016/j.scitotenv.2018.08.315

Shenfield, A., and Rostami, S. (2015). *A Multi Objective Approach to Evolving Artificial Neural Networks for Coronary Heart Disease Classification*. 2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Auguat 12–15, 2015, 1–8. doi:10.1109/CIBCB.2015.7300294

Shanshan, Q., Liu, F., Wang, J., and Sun, B. (2014). Analysis and Forecasting of the Particulate Matter (PM) Concentration Levels over Four Major Cities of China Using Hybrid Models. *Atmos. Environ.* 98, 665–675. doi:10.1016/j.atmosenv.2014.09.046

Steffens, J., Kimbrough, S., Baldauf, R., Isakov, V., Brown, R., Powell, A., et al. (2017). Near-port Air Quality Assessment Utilizing a mobile Measurement Approach. *Atmos. Pollut. Res.* 8, 1023–1030. doi:10.1016/j.apr.2017.04.003

Urbancok, D., Payne, A., and Webster, R. (2017). Regional Transport, Source Apportionment and Health Impact of PM 10 Bound Polycyclic Aromatic Hydrocarbons in Singapore's Atmosphere. *Environ. Pollut.* 229, 984–993.

Xu, Y., Yang, W., and Wang, J. (2016). Air Quality Early-Warning System for Cities in China. *Atmos. Environ.* 148, 239–257. doi:10.1016/j.atmosenv.2016.10.046

Yang, Z., and Wang, J. (2017). A New Air Quality Monitoring and Early Warning System: Air Quality Assessment and Air Pollutant Concentration Prediction. *Environ. Res.* 158, 105–117. doi:10.1016/j.envres.2017.06.002

You, H., Ma, Z., Tang, Y., Wang, Y., Yan, J., Ni, M., et al. (2017). Comparison of ANN (MLP), ANFIS, SVM, and RF Models for the Online Classification of Heating Value of Burning Municipal Solid Waste in Circulating Fluidized Bed Incinerators. *Waste Management* 68, 186–197. doi:10.1016/j.wasman.2017.03.044

Zhang, L., Lin, J., Qiu, R., Hu, X., Zhang, H., Chen, Q., et al. (2018). Trend Analysis and Forecast of PM2.5 in Fuzhou, China Using the ARIMA Model. *Ecol. Indicators* 95, 702–710. doi:10.1016/j.ecolind.2018.08.032

Zhou, Q., Jiang, H., Wang, J., and Zhou, J. (2014). A Hybrid Model for PM 2.5 Forecasting Based on Ensemble Empirical Mode Decomposition and a General Regression Neural Network. *Sci. Total Environ.* 496, 264–274. doi:10.1016/j.scitotenv.2014.07.051

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.