

# PM<sub>2.5</sub> Concentration Prediction Based on Markov Blanke Feature Selection and Hybrid Kernel Support Vector Regression Optimized by Particle Swarm Optimization

Lian-Hua Zhang<sup>1,2,3\*</sup>, Ze-Hong Deng<sup>1</sup>, Wen-Bo Wang<sup>2,3</sup>

<sup>1</sup> School of Literature, Law and Economics, Wuhan University of Science and Technology, Wuhan 430065, China

<sup>2</sup> Hubei Province Key Laboratory of Systems Science in Metallurgical Process (Wuhan University of Science and Technology), Wuhan 430081, China

<sup>3</sup> College of Science, Wuhan University of Science and Technology, Wuhan 430065, China

## ABSTRACT

This study employed air quality and meteorological data as research materials and extracted the optimal feature subset by using the approximate Markov blanket-based normal maximum relevance minimum redundancy (nMRMR) algorithm to serve as the input data of the prediction model. In addition, a hybrid kernel (HK) was created to improve upon the traditional support vector regression (SVR) model. Particle swarm optimization (PSO) was used to calculate the optimal parameters of hybrid kernel (HK) SVR, which were then used to establish the nMRMR-PSO-HK-SVR model for PM<sub>2.5</sub> concentration prediction. The 2016–2019 year air quality and weather data of Wuhan and Tianjin were employed to test the proposed method. The experimental results show that the mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE) and Theil's inequality coefficient (TIC) of nMRMR-PSO-HK-SVR model are lower than those of SVR, PSO-SVR, nMRMR-SVR and PSO-HK-SVR model. But also, the proposed model could more precisely track moments of sudden PM<sub>2.5</sub> concentration change. Thus, the nMRMR-PSO-HK-SVR model has more satisfactory generalizability and can predict PM<sub>2.5</sub> concentration more precisely.

**Keywords:** PM<sub>2.5</sub>, Maximum relevance minimum redundancy (MRMR), Hybrid kernel, Support vector regression, Prediction model

## 1 INTRODUCTION

Rapid development of economies worldwide has caused increasingly severe air pollution. A major pollutant, PM<sub>2.5</sub> remains in the air for a long time and can be transported over great distances because of its small size. These results in lowered visibility and severe deterioration of air quality and the atmospheric environment because of the copious toxic substances PM<sub>2.5</sub> carries, thus posing a health risk (Gu *et al.*, 2006). Estimation of the PM<sub>2.5</sub> concentration has critical value and significance to early warnings of severe pollution events. To date, PM<sub>2.5</sub> concentration has generally been estimated using validation or statistical models (Patricio *et al.*, 2020). Validation models are primarily constructed using historical weather information and chemical initial and boundary conditions to infer the complex process of pollutant formulation (Wang *et al.*, 2017a). Therefore, the estimation precision of these models is dependent on the accuracy of complex historical records, and accurate historical records are usually difficult to obtain. Because of the development in regression learning, artificial neural network and support vector regression (SVR) models have been successfully applied to estimate PM<sub>2.5</sub> concentration. Perez and Gramsch (2016) verified that when historical pollutant concentration data and weather data are available,

## OPEN ACCESS

Received: May 15, 2020

Revised: June 6, 2020

Accepted: December 25, 2020

\* Corresponding Author:

lyzhang200109@163.com

**Publisher:**

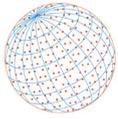
Taiwan Association for Aerosol  
Research

ISSN: 1680-8584 print

ISSN: 2071-1409 online

 **Copyright:** The Author(s).

This is an open access article distributed under the terms of the [Creative Commons Attribution License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.

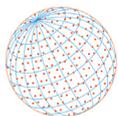


a feed forward neural network model can effectively predict the hourly concentration of PM<sub>2.5</sub>. Sun and Sun (2017) combined principal component analysis (PCA) with least-squares SVR to predict the daily PM<sub>2.5</sub> concentration; their experimental results revealed that the prediction precision was high. Statistical models describe the relationship between PM<sub>2.5</sub> concentration and the factors influencing it; therefore, the models have high prediction precision (Yin *et al.*, 2018). Although artificial neural network models can be employed in PM<sub>2.5</sub> concentration estimation, problems tend to occur, such as local optimal solutions being obtained and over fitting (Niu *et al.*, 2016). SVR models are based on statistical learning theories (Zhang, 2019); structural minimization is adopted as a principle, and the problem of over fitting does not exist. Hence, such models exhibit favorable generalizability (Yan *et al.*, 2020). As a major type of air pollutant, PM<sub>2.5</sub> has complex origins and forms through a complicated process under the influence of numerous factors (Ni *et al.*, 2017; Song *et al.*, 2018). It exhibits high complexity and nonlinearity (Wang *et al.*, 2017b). Most studies on PM<sub>2.5</sub> concentration estimation have been based on PM<sub>2.5</sub> time series (Qin *et al.*, 2016; Wang *et al.*, 2020), which strongly affect the prediction precision. In the present study, six air quality indices (PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, CO, and O<sub>3</sub> concentrations) and five weather factors (temperature, relative humidity, precipitation, wind speed, and air pressure) were employed to predict the PM<sub>2.5</sub> concentration on the next day. However, redundant data had to be eliminated, and this required dimension reduction of the aforementioned 11 factors. Some scholars (Wang *et al.*, 2017c; Jayakumar and Sangeetha, 2020) have used correlation coefficients or PCA for feature selection. Qiao *et al.* (2017) combined with principal component analysis (PCA) and fuzzy neural network to predict the concentration of PM<sub>2.5</sub>, and obtained better prediction results. Singh and Gupta (2012) used stepwise linear regression method to select the original features in the prediction of urban air quality, and used linear and nonlinear prediction models for experimental comparison. Kim *et al.* (2010) used the partial least squares (PLS) method to select the variables that have a greater impact on the output to predict PM<sub>2.5</sub> and PM<sub>10</sub> in the subway station, and compared with the prediction results obtained by taking all the measured variables as inputs, which proved the necessity of selecting characteristic variables.

The selection of PM<sub>2.5</sub> characteristic variables by the above common methods only reflects the linear relationship between variables, and does not consider the nonlinear relationship between variables. However, correlation coefficients are appropriate only when all features are independent, PCA can only process linear questions, and the 11 features considered in this study exhibit strong correlations and nonlinearity (Wang *et al.*, 2015). Therefore, this paper proposes a feature selection method based on the approximate Markov blanket-based normal maximum relevance minimum redundancy (nMRMR) algorithm (Zhang and Wang, 2018; Cai *et al.*, 2019), aiming at the shortcomings of the commonly used PM<sub>2.5</sub> feature variable selection method. The maximum correlation and minimum redundancy mutual information based on the approximate Markov blanket is used to calculate and sort out some features with small correlation, and the optimal feature collection is selected.

Support Vector Regression (SVR) is one of the most robust and accurate methods in data mining algorithms (Vapnik, 1998), which mainly includes SVM classification and support vector regression. Sun *et al.* (2016) presented a hybrid model based on PCA and least square support vector regression (LSSVR) optimized by cuckoo search algorithm to predict PM<sub>2.5</sub> concentrations. The prediction precision of an SVR model is dependent on the type of kernel functions (Cura, 2020). The radial basis function kernel (RBF kernel) has a high degree of local fitting, whereas the polynomial kernel (Poly) has strong generalizability (Dhamecha *et al.*, 2019). Poly and the RBF kernel can be linearly integrated to form a hybrid kernel (HK), which retains the advantages of the two original functions, enhances generalizability in the SVR model, and has been applied satisfactorily in numerous fields (Fei, 2016; Zhong and Carr, 2016).

In the Hybrid kernel SVR algorithm, the choice of SVR parameters (penalty factor and kernel function parameters) and kernel function combination parameters has an important impact on the accuracy of prediction, but at present, there is no optimal value method for kernel parameter, penalty factor and kernel function combination parameters. In traditional SVR, parameter selection is obtained by repeated experiments, which has great randomness of artificial selection. However, it takes a lot of time to select parameters by cross validation, although it overcomes human randomness to some extent. Particle swarm optimization (PSO), as an optimization method developed in recent years, has been widely used in function optimization, pattern recognition



and other fields due to its easy implementation and deep intelligent background. In order to achieve the optimal selection of parameters in the hybrid kernel SVR model, this paper combines PSO with SVR, and uses the global search ability of PSO to search the parameters in HK-SVR

In summary, the present study employed historical air quality and weather data. The nMRRM algorithm was used to first select features, and the optimal feature subset was chosen and input to the SVR model. A particle swarm optimization (PSO)-based HK was constructed to improve the conventional SVR model. Eventually, an nMRRM-PSO-HK-SVR model was established for estimating PM<sub>2.5</sub> concentration.

## 2 MATERIALS AND METHODS

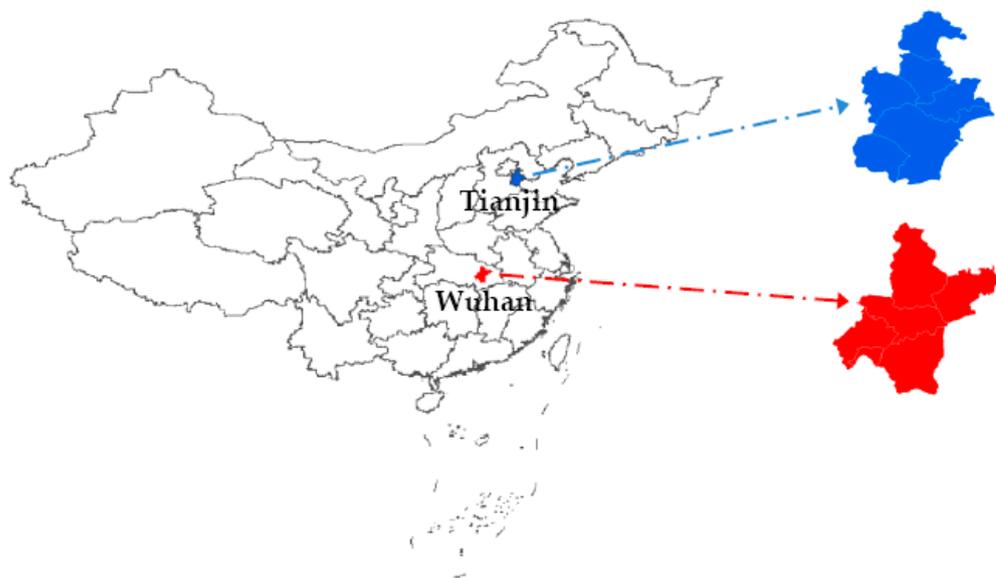
### 2.1 Research Site

Fig. 1 briefly shows the geographical location of Wuhan and Tianjin. Wuhan is located in the eastern part of the Jiangnan Plain in the middle and lower reaches of the Yangtze River (39°N, 114°E). Wuhan is the largest and most populous city in Hubei Province, with a total area of 8494.41 km<sup>2</sup> and a population of 11.08 million. Wuhan's terrain is mainly hilly, which is not conducive to the spread of PM<sub>2.5</sub>. Although Wuhan is currently making efforts to improve its environment, the air quality does not meet the national secondary standard, especially in winter, when the concentration of PM<sub>2.5</sub> rises sharply. In recent years, PM<sub>2.5</sub> pollution has become a thorny issue in Wuhan, and it is of great significance to study the pollution control of Wuhan, a mega-city in a period of economic growth.

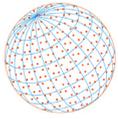
Tianjin is the largest coastal city in northern China, located on the west coast of Bohai Bay (39°N, 117°E), which has become China's new growth pole and center of advanced industrial and financial activities. Tianjin has a warm temperate sub-humid monsoon climate, with a significant monsoon and four distinct seasons. In the past decades, with the rapid urbanization, Tianjin has become a mega-city with a population of over 10 million. At the same time, due to the development of industrialization and the increase of motor vehicles, hazy weather is a common occurrence. In order to protect public health and atmospheric environment, there is an urgent need to simulate and predict the concentration of PM<sub>2.5</sub> in Tianjin.

### 2.2 Research Data

In this paper, data were collected from January 1, 2016 to June 30, 2019, including air quality data and meteorological data. The air quality data from <https://www.aqistudy.cn/historydata/>, including PM<sub>2.5</sub> and PM<sub>10</sub>, sulfur dioxide (SO<sub>2</sub>), nitrogen dioxide (NO<sub>2</sub>), carbon monoxide (CO),



**Fig. 1.** Geographical location of Wuhan City and Tianjin City.



and the daily average density of ozone ( $O_3$ ) eight hours. The "8 hours of ozone" is a sliding average value, which is calculated based on the average concentration of the 8 consecutive hours with the highest ozone level between 8:00 and 24:00. The meteorological data were collected from <http://www.wunderground.com/history/>, including wind speed, precipitation, atmospheric pressure, relative humidity and temperature. A total of 1065 data from January 1, 2016 to November 30, 2018 in Wuhan and Tianjin were selected as the training sample set, and 31 data from December 1, 2018 to December 31, 2018 in the two regions were selected as the test sample set for the short-term prediction. A total of 181 data from January 1, 2019 to June 30, 2019 in the Wuhan region were selected as the test sample set for the long-term forecasts.

For a small number of missing data (1 in 2016, 3 in 2017, and 3 in 2018), we fill in the data with the mean value of the adjacent two days. For the prediction training, we first normalize the data. The normalization formula is as follows:

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

In the formula,  $y$  denotes normalized data;  $x$  denotes pre-normalized data;  $x_{\min}$  denotes data minimum;  $x_{\max}$  denotes data maximum. In Section 1.4, four evaluation metrics are given in this paper to evaluate the prediction accuracy of the prediction model for  $PM_{2.5}$ , and all algorithmic models and experiments in this paper are implemented in Matlab 2018b.

## 2.3 Description of the Problem

The goal of  $PM_{2.5}$  concentration prediction is to predict  $PM_{2.5}$  concentration for a fixed period of time in the future (for example, the next 24, 48 or 72 hours) using the observed values for a fixed period of time in the past (for example, 24 hours). The aim of this manuscript is to make a short-term prediction of  $PM_{2.5}$  concentrations for the next 24 h (next day) using past observations.

For a given moment  $t$ , assume that the observed data in the past  $L$  ( $L \times 24$  h) day are

$$ST = \{X_{t-L}, X_{t-L+1}, \dots, X_t\}$$

Each observation  $X_{t-i}$  ( $0 \leq i \leq L$ ) in the sequence is a  $d$  dimensional vector consisting of pollutant concentration and some meteorological element observations. After the prediction model is constructed,  $PM_{2.5}$  concentration value in the next  $\{t + 1, t + 2, \dots, t + K\}$  day can be predicted by the data set  $ST$  as input.

In this manuscript, we use the maximum correlation minimum redundancy algorithm to extract the optimal features from the observed data  $ST$ , and use particle swarm optimization kernel support vector regression as a prediction model to make a short-term prediction of  $PM_{2.5}$  concentration in the next 24 hours (the next day).

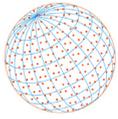
## 2.4 Research Methods

### 2.4.1 Approximate Markov blanket-based nMRMR algorithm

Because the factors influencing  $PM_{2.5}$  concentration are strongly correlated, data redundancy and specifically nonlinear relationships must be considered. Therefore, the approximate Markov blanket-based nMRMR algorithm was employed in this study for feature selection. The core of this algorithm is maximization of the correlation between a feature and the target feature at the same time as minimization of the correlations among the other features. The correlation at this time was expressed through mutual information (Ju and He, 2018). The mutual information between variables  $x$  and  $y$  is defined as

$$I(x, y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad (1)$$

where,  $p(x)$  and  $p(y)$  are respectively the probability density function of  $x$  and  $y$ , and  $p(x, y)$  is the joint probability density function of  $x$  and  $y$ . The measurement indicators of maximum relevance



and minimum redundancy are defined respectively as

$$\max D(S, p), D = \frac{1}{n} \sum_{i=1}^n I(x_i, p) \quad (2)$$

$$\min R(S), R = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n I(x_i, x_j) \quad (3)$$

where,  $S$  is the feature subset;  $n$  is the number of features;  $I(x_i, p)$  is the mutual information between the 11 features (i.e., six air quality features and five weather features) and the PM<sub>2.5</sub> data for the next day;  $p$  is the target feature and  $I(x_i, x_j)$  is the mutual information among the 11 features.

The criteria for feature selection are excellent classification performance and the smallest possible number of dimensions. These entail the maximum relevance within the feature set and categories as well as the minimum redundancy among the features. After comprehensive consideration of the aforementioned two measurement indicators, the following criterion for maximum relevance and minimum redundancy is obtained:

$$\max \Phi(D, R), \Phi(D, R) = D - R \quad (4)$$

**Markov blanket** (Yu *et al.*, 2019): Let  $F$  be the feature set and  $f_i$  be a feature within  $F$ . For a feature subset  $S$ , if  $S \subset F$ , then  $f_i \in S$  and the Markov blanket condition of  $f_i$  is  $f_i \perp \{F - S - \{f_i\}, C\} | S$ , where  $\perp$  indicates independence and  $|S$  means condition on  $S$ . In a given  $S$ ,  $f_i$  is independent of  $F - S - \{f_i\}$  and  $C$ , suggesting when  $S$  exists,  $f_i$  makes no contribution to the label  $C$  and should therefore be deleted. In addition, this shows that the smallest  $S$  that satisfies the aforementioned conditions is the Markov blanket of feature  $f_i$ .

According to the definition of the Markov blanket, the criteria for determining the approximate Markov blanket can be obtained (Hua *et al.*, 2020); that is, for features  $f_i$  and  $f_j$  ( $i \neq j$ ), the Markov blanket criteria for feature  $f_i$  to be feature  $f_j$  are

$$\begin{cases} I(f_i, C) > I(f_j, C) \\ I(f_j, C) > I(f_i, f_j) \end{cases} \quad (5)$$

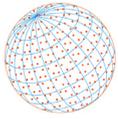
The specific steps of the approximate Markov blanket-based nMRMR algorithm used in this study are as follows:

- Step 1:** Initialize  $F$ , the set of factors influencing PM<sub>2.5</sub> concentration, and establish an empty set  $S$ .
- Step 2:** Calculate the mutual information between each feature in  $F$  and the PM<sub>2.5</sub> concentration of the next day. Arrange the features in  $F$  in descending order by the size of their mutual information.
- Step 3:** Deposit the first feature in  $F$  into  $S$ , and delete  $f_i$  from  $F$ .
- Step 4:** Arrange the features in accordance with the principle of maximum relevance and minimum redundancy. Deposit the arranged features into  $S$ .
- Step 5:** Delete irrelevant and redundant features in  $S$  according to the criteria of the approximate Markov blanket.
- Step 6:** Export the optimal set of factors affecting PM<sub>2.5</sub> concentration, which is  $S$ .

#### 2.4.2 PSO hybrid Kernel SVR (PSO-HK-SVR)

Based on statistical principles, particularly that of structural minimization, SVR can satisfactorily solve high-dimensional and over fitting problems. The core idea is to use the kernel function to map the imported data to a high-dimensional feature space rather than transforming the nonlinear problem into a linear problem. Through the use of the kernel function, dot product operations in high-dimensional space can be avoided, and an objective function can be formulated as follows:

$$f(x) = w \cdot \phi(x) + b \quad (6)$$



where,  $w$  the weight is vector and  $b$  is the offset constant. Substituting them into the kernel function yields the optimal hyper plane fitting function:

$$f(x) = \sum_{i=1}^m (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (7)$$

where,  $\alpha_i$  and  $\alpha_i^*$  are Lagrange multipliers and  $K(x, x_i)$  is the kernel function.

When performing PM<sub>2.5</sub> prediction using an SVR, kernel function selection has a decisive influence on the prediction result. Of the numerous types of kernel functions created for SVR applications, the following three types are the most commonly used:

(1) Poly Kernel

$$K(x, x_i) = [(x, x_i) + 1]^q$$

(2) Gaussian RBF kernel

$$K(x, x_i) = \exp\left(-\|x - x_i\|^2 / \sigma^2\right)$$

(3) Sigmoid kernel

$$K(x, x_i) = \tanh(v(x \cdot x_i) + c)$$

Poly kernel is a global kernel function that has a significant influence on dots that are far apart; it has extremely strong generalizability but weak learning capacity. By contrast, Gaussian RBF is a local kernel function that only influences dots that are relatively close; it has strong learning capacity but weak generalizability (Huanrui, 2016; Liu *et al.*, 2016).

To create a kernel that influences close-together dots as well as far-apart dots in fuzzy prediction, Poly and the Gaussian RBF kernel can be weighted and integrated to formulate a new linear HK (Huanrui, 2016):

$$K = r \cdot K_{poly} + (1 - r) \cdot K_{rbf} = r \cdot [(x, x_i) + 1]^q + (1 - r) \cdot \exp\left(-\|x - x_i\|^2 / \sigma^2\right) \quad (8)$$

where,  $K_{rbf}$  is the RBF kernel;  $K_{poly}$  is Poly;  $q$  is the polynomial order;  $\sigma$  is the bandwidth parameter of the RBF kernel;  $r$  is the weight coefficient for the HK, and  $r \in [0, 1]$ . When  $r = 1$ , the HK functions as Poly, whereas when  $r = 0$ , the HK is the RBF kernel. An analysis of Eq. (8) reveals that three variables in the equation—namely  $q$ ,  $\sigma$  and  $r$ —must be optimized. Therefore, obtaining the solution to the HK entails optimization of three variables, which can be expressed as  $x = [q, \sigma, r]$ . Smart algorithms such as the genetic algorithm, PSO, and artificial bee colony algorithm are one of the most effective methods of solving such problems. In this study, PSO, the most easily performed smart algorithm, was used to optimize the three parameters of HK-SVR. The core of the PSO algorithm is update of the speed and location of particles (Jiao and Liu, 2009)

$$\begin{cases} v_{id}^k = \lambda \cdot v_{id}^{k-1} + c_1 \cdot u_1 \cdot (p_{id}^k - x_{id}^k) + c_2 \cdot u_2 \cdot (g_{id}^k - x_{id}^k) \\ x_{id}^k = x_{id}^{k-1} + v_{id}^k \end{cases} \quad (9)$$

where,  $\lambda$  is the inertia weight, the value of which generally decreases from 0.9 to 0.4;  $c_1$  and  $c_2$  are the learning factors, the values of which are typically 2; and  $u_1$  and  $u_2$  are two random numbers. The PSO algorithm was employed to optimize  $x = [q, \sigma, r]$ , a parameter combination in HK-SVR. The parameter combination with the optimal fitness was used in the HK-KELM model. The fitness was obtained by

$$g(x) = \sqrt{\frac{1}{m} \sum_{i=1}^m [y_i - y_i^*(x)]^2} \quad (10)$$

where,  $y_i$  is the actual value;  $y_i^*$  is the predicted value, and  $m$  is the total number of trained samples. The optimized parameters were then substituted into Eq. (8) to derive the final equation of the HK-SVR model.

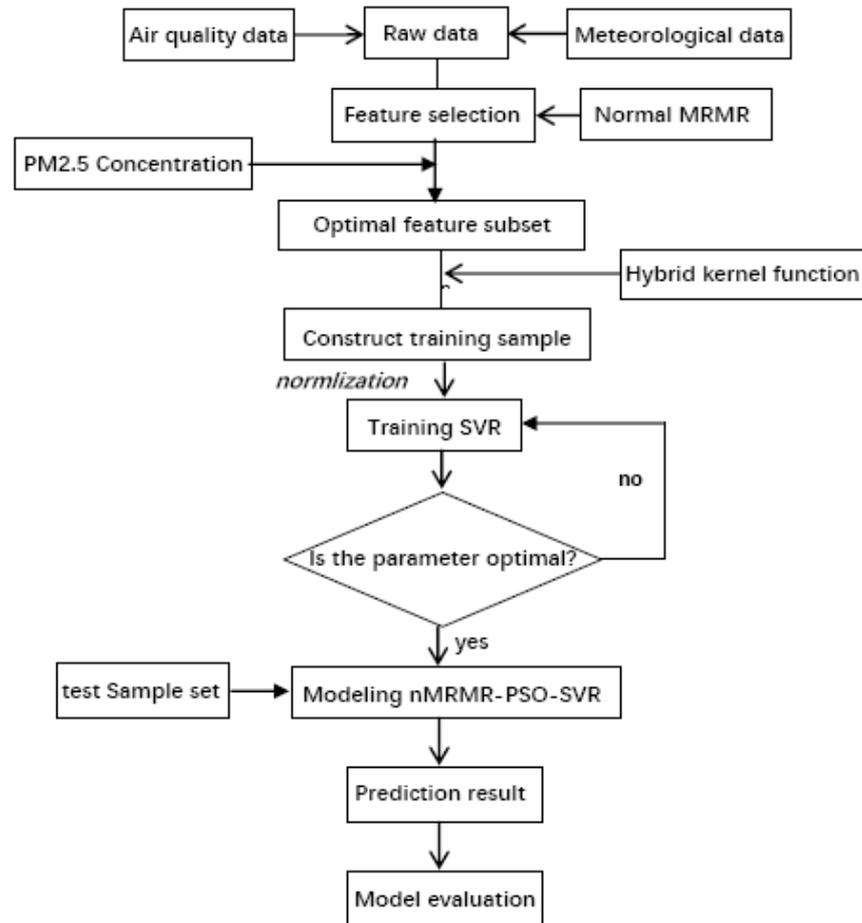
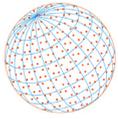


Fig. 2. Flow diagram of nMRMR-PSO-HK-SVR model.

### 2.4.3 Construction of nMRMR method-based PSO-HK-SVR model

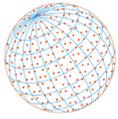
Because  $PM_{2.5}$  concentration is affected by numerous factors and strong correlations exist among these factors, this study first used the approximate Markov blanket-based nMRMR algorithm to select features. The optimal feature subset was then employed as the SVR model input. Because  $PM_{2.5}$  concentration variation is difficult to describe using a single kernel function, the HK was employed to modify the conventional SVR model; furthermore, the PSO algorithm was adopted to optimize the parameters of the HK. Finally, the nMRMR-PSO-HK-SVR model was constructed in MATLAB. Fig. 2 illustrates the process.

- (1) Collect the 2016–2019 year air quality statistics and weather information of Wuhan and produce the time series data of next-day  $PM_{2.5}$  concentration to be used as the model output.
- (2) Use the approximate Markov blanket-based nMRMR algorithm to select the optimal feature subset.
- (3) Create a training set based on the optimal feature subset selected in Step (2). Subject the training set to normalization using Eq. (10) to eliminate interference caused by unit differences among the features.

$$x' = 0.1 + 0.8 \cdot \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (11)$$

where,  $x$  and  $x'$  are the values before and after normalization respectively; and  $x_{\min}$  and  $x_{\max}$  are the minimum and maximum values of the original data series respectively.

- (1) Use the HK to improve the conventional SVR model.
- (2) Train the SVR using the normalized training set; use the optimized PSO algorithm to obtain the corresponding parameters and a prediction model.



(3) Input the testing sample to the prediction model to obtain the PM<sub>2.5</sub> concentration of the next day.

## 2.4 Prediction Model and Evaluation Indices

In order to verify the validity of the proposed model, the following error evaluation indexes were selected: mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), mean square error (RMSE) and Theil's inequality coefficient (TIC) (Chen and Li, 2019), which is calculated as follows.

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |y_i - y_i^*| \quad (12)$$

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^m \left| \frac{y_i - y_i^*}{y_i} \right| \times 100\% \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m |y_i - y_i^*|^2} \quad (14)$$

$$\text{TIC} = \frac{\sqrt{\frac{1}{m} \sum_{i=1}^m |y_i - y_i^*|^2}}{\sqrt{\frac{1}{m} \sum_{i=1}^m y_i^2} + \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i^*)^2}} \quad (15)$$

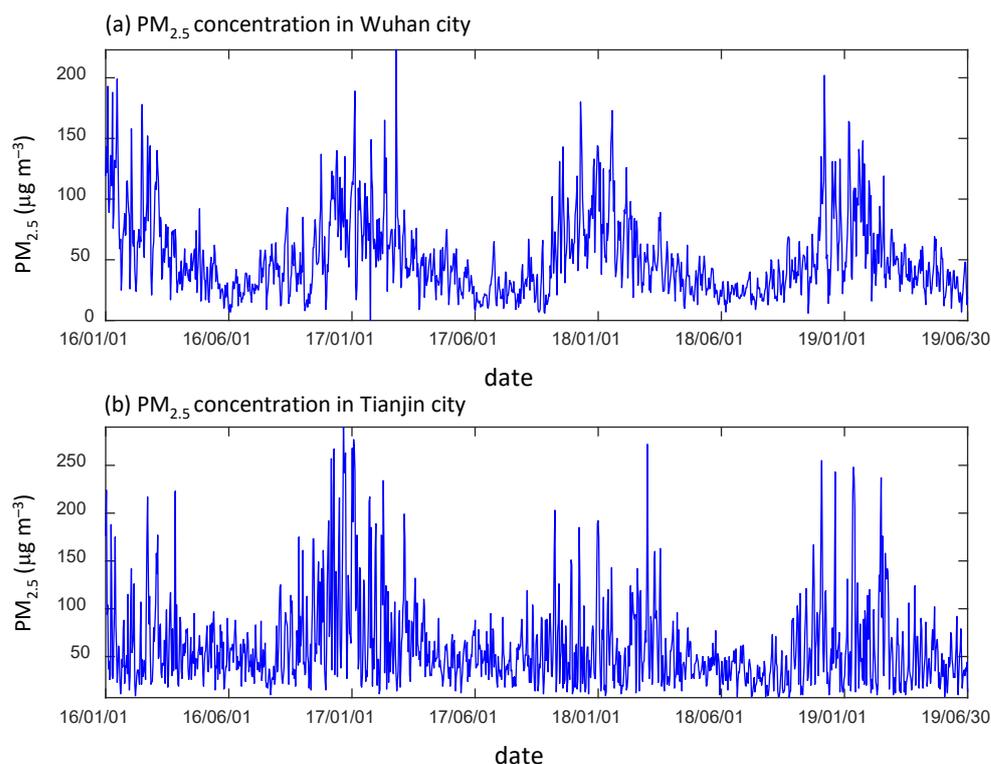
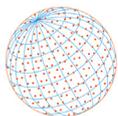
where,  $y_i$  and  $y_i^*$  represent the measured and predicted values of PM<sub>2.5</sub> concentration, and  $m$  is the number of samples in the test set. In the evaluation index, MAE, RMSE and MAPE are used to quantify the error of prediction results. The smaller MAE, RMSE and MAPE indicate that the prediction accuracy is higher. TIC is used to evaluate the prediction ability of different prediction models. The smaller TIC is, the better the prediction ability of the model is.

To compare the validity of the nMRMR-PSO-HK-SVR method proposed in this paper, we also use back propagation (BP), support vector regression (SVR), normal maximum relevance minimum redundancy SVR (nMRMR-SVR) and Particle Swarm Optimization-hybrid kernel-SVR (PSO-HK-SVR) to predict the concentration of PM<sub>2.5</sub>. Finally the results of the five prediction methods are compared and analyzed. In the BP neural network prediction model, two hidden layers are set up; the first hidden layer has 30 neuron nodes and the second hidden has 20 neurons; the iteration step size is chosen to be 0.001; the number of training sessions is 1,000; the training target is taken to be 0.0001. In the SVR and nMRMR-SVR models, the radial basis function is chosen as the nuclear function, and the nuclear function parameters are selected by ten-count cross-validation. In the PSO-HK-SVR and nMRMR-PSO-HK-SVR models, radially based and polynomial kernel functions are selected to form hybrid kernel functions, the parameters of which are determined by a particle swarm optimization algorithm. The parameters of the particle swarm algorithm are set as follows: population size is 25, the number of evolutions is 200 times, and the learning factor is taken to be  $c_1 = c_2 = 2$ .

## 3 RESULTS AND ANALYSIS

### 3.1 Analysis of Wuhan's and Tianjin's Pollution

Fig. 3 shows that the concentration of PM<sub>2.5</sub> is U-shaped and has obvious seasonal characteristics, with a low concentration in summer and a sharp increase in winter. It can be seen from Table 1 that the current pollution situation in Wuhan and Tianjin has been relieved after the treatment in previous years. And the annual average value is about 45.71  $\mu\text{g m}^{-3}$ . But the concentration of PM<sub>2.5</sub> also often reaches 150  $\mu\text{g m}^{-3}$  in winter.



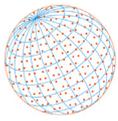
**Fig. 3.** PM<sub>2.5</sub> concentration of Wuhan city and Tianjing city.

**Table 1.** Statistical results of air quality data and meteorological data in Wuhan.

| Index                                   | Average value | Maximum value | Minimum value | Standard deviation |
|---|---------------|---------------|---------------|--------------------|
| PM <sub>2.5</sub> (µg m <sup>-3</sup> ) | 55.6          | 290           | 7             | 39.9               |
| PM <sub>10</sub> (µg m <sup>-3</sup> )  | 69.8          | 220           | 15.2          | 34.8               |
| SO <sub>2</sub> (µg m <sup>-3</sup> )   | 22.3          | 71            | 3.4           | 11.7               |
| NO <sub>2</sub> (µg m <sup>-3</sup> )   | 23.0          | 72.3          | 9.1           | 11.6               |
| CO (µg m <sup>-3</sup> )                | 1.2           | 2.7           | 0.9           | 0.6                |
| O <sub>3</sub> (µg m <sup>-3</sup> )    | 105           | 253           | 21.6          | 41.4               |
| Wind speed (m s <sup>-1</sup> )         | 1.3           | 3.7           | 0.5           | 0.7                |
| Rainfall (mm)                           | 3.6           | 138           | 0.4           | 11.0               |
| Pressure (hpa)                          | 998           | 1018          | 980           | 7.3                |
| Relative humidity (%)                   | 76            | 99            | 42.8          | 12.5               |
| Temperature (°C)                        | 20.5          | 33            | 5             | 7.8                |

### 3.2 nMRMR-method Based Feature Selections

In this study, the 2016–2019 year data for 11 features in Wuhan were obtained to produce the times series of next-day PM<sub>2.5</sub> concentration, which is expressed as PM<sub>2.5</sub>. Table 2 presents the between-feature mutual information values of the 12 features. A greater value indicates stronger correlation. A few features were indeed strongly correlated; therefore, feature selection could not be conducted as if the features were mutually independent. When employed as the output of the prediction model, PM<sub>2.5p</sub> was most strongly correlated with air pressure, temperature, and the concentrations of PM<sub>10</sub>, CO, and O<sub>3</sub>. When the between-feature degree of redundancy was also considered, the optimal subset selected using the nMRMR algorithm was air pressure, temperature, PM<sub>10</sub> concentration, 46 concentration and precipitation. These five features were used as the input of the prediction model. Despite being extremely strongly correlated with PM<sub>2.5p</sub>, CO was not included in the optimal input subset because CO had extremely strong correlation with PM<sub>10</sub> and O<sub>3</sub>. This proves that the nMRMR algorithm indeed takes the redundancy in data into consideration when selecting the optimal input subset.



**Table 2.** Mutual information values between the 12 feature data (Air quality data and meteorological data).

| Parameter          | PM <sub>2.5</sub> | PM <sub>10</sub> | SO <sub>2</sub> | NO <sub>2</sub> | CO  | O <sub>2</sub> | WS  | RF  | P   | RT  | T   | PM <sub>2.5p</sub> |
|--------------------|-------------------|------------------|-----------------|-----------------|-----|----------------|-----|-----|-----|-----|-----|--------------------|
| PM <sub>2.5</sub>  | -                 | 4.5              | 3.3             | 3.0             | 4.5 | 5.0            | 2.1 | 1.6 | 4.9 | 3.2 | 4.9 | 3.8                |
| PM <sub>10</sub>   |                   | -                | 3.6             | 3.3             | 5.1 | 5.5            | 2.4 | 1.8 | 5.5 | 3.7 | 5.5 | 4.5                |
| SO <sub>2</sub>    |                   |                  | -               | 2.5             | 3.6 | 4.2            | 1.5 | 1.2 | 4.3 | 2.5 | 4.1 | 3.1                |
| NO <sub>2</sub>    |                   |                  |                 | -               | 3.5 | 4.0            | 1.4 | 1.2 | 4.0 | 2.3 | 4.0 | 2.9                |
| CO                 |                   |                  |                 |                 | -   | 5.7            | 2.6 | 1.8 | 5.8 | 3.9 | 5.7 | 4.3                |
| O <sub>2</sub>     |                   |                  |                 |                 |     | -              | 3.0 | 2.0 | 6.3 | 4.3 | 6.2 | 4.9                |
| WS                 |                   |                  |                 |                 |     |                | -   | 1.0 | 3.1 | 1.5 | 3.0 | 2.0                |
| RF                 |                   |                  |                 |                 |     |                |     | -   | 2.1 | 1.5 | 2.1 | 1.6                |
| P                  |                   |                  |                 |                 |     |                |     |     | -   | 4.4 | 6.2 | 4.9                |
| RH                 |                   |                  |                 |                 |     |                |     |     |     | -   | 4.4 | 3.3                |
| T                  |                   |                  |                 |                 |     |                |     |     |     |     | -   | 4.9                |
| PM <sub>2.5p</sub> |                   |                  |                 |                 |     |                |     |     |     |     |     | -                  |

Note: PM<sub>2.5p</sub> represents the concentration of PM<sub>2.5</sub> on the next day; the other characteristics represent the observed values on the same day. WS, RF, P, RH, T represents wind speed, rainfall, pressure, relative humidity and temperature, respectively.

### 3.3 Short-term Prediction of PM<sub>2.5</sub> in Wuhan City

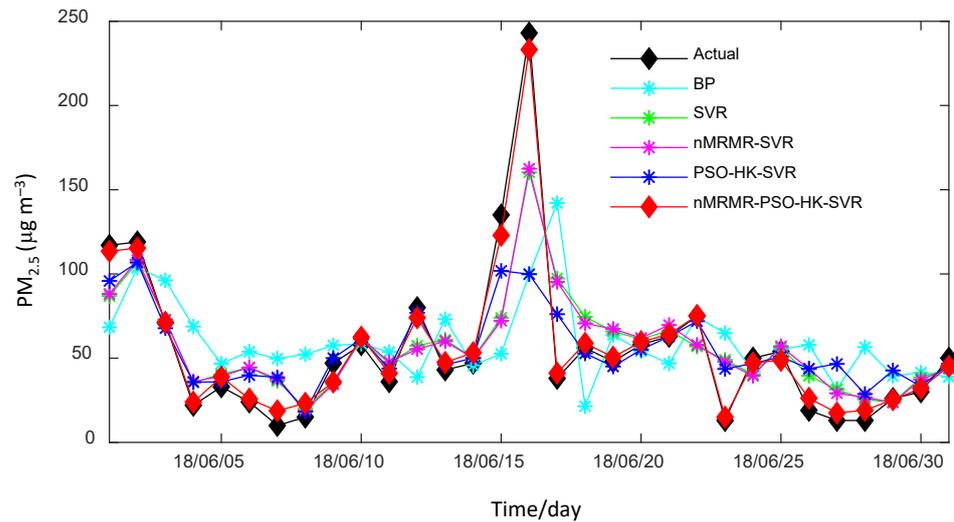
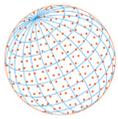
In order to analyze the validity of the nMRMR-PSO-HK-SVM model proposed in this paper, a total of 1065 data from 2016/01/01 to 2018/11/30 is selected as the training sample set, and a total of 31 data from 2018/12/01 to 2018/12/31 is selected as the test sample set for the short-term prediction experiment of PM<sub>2.5</sub>. Meanwhile, the prediction results of the proposed method are compared with BP method, SVR method, nMRMR-SVR method, and PSO-HK-SVR method. The prediction performance of all prediction models are evaluated using four error indices including MAE, MAPE, RMSE and TIC.

The prediction results of the five models are shown in Fig. 4, and the four error performance indicators (MAE, MAPE, RMSE, TIC) for all models are shown in Table 3. In Table 3, the minimum values for each error performance indicator are marked in bold black. From Table 3, it can be seen that the four error performance metrics (MAE, MAPE, RMSE, TIC) for the prediction results of the nMRMR-PSO-HK-SVR model are minimal compared to the other four methods. This indicates that the hybrid nMRMR-PSO-HK-SVR model proposed in this paper has the best prediction performance. To compare the prediction errors of different models more visually, in Fig. 5 we present histograms of MAE, MAPE, RMSE, and TIC for different methods.

In order to further analyze the effect of nMRMR feature selection method and PSO-HK technique on the prediction accuracy, the following three types of comparison tests are conducted in this paper. The first type of comparison test (Comparison Test I) is used to analyze the effect of nMRMR method on the prediction accuracy of PM<sub>2.5</sub>, in which BP model, SVR model and nMRMR-SVR model are compared and analyzed. The second type of comparison test (Comparison Test II) is used to analyze the effect of Particle Swarm Optimization (PSO) algorithm and hybrid kernel function on the prediction of SVR, in which we compare the prediction results of the SVR model, PSO-HK-SVR. The third type of comparison test (Comparison Test III) is used to analyze the effect of the hybrid model nMRMR-PSO-HK on the prediction accuracy of SVR, in which the PM<sub>2.5</sub> prediction results of the nMRMR-SVR model, the PSO-HK-SVR model, and the nMRMR-PSO-HK-SVR model are Contrast analysis. The results of the analysis of the three types of comparison experiments are shown in Table 4. From the results of Table 4, we can obtain the following conclusions.

#### 3.3.1 Comparison results between SVR and BP, nMRMR-SVR and SVR

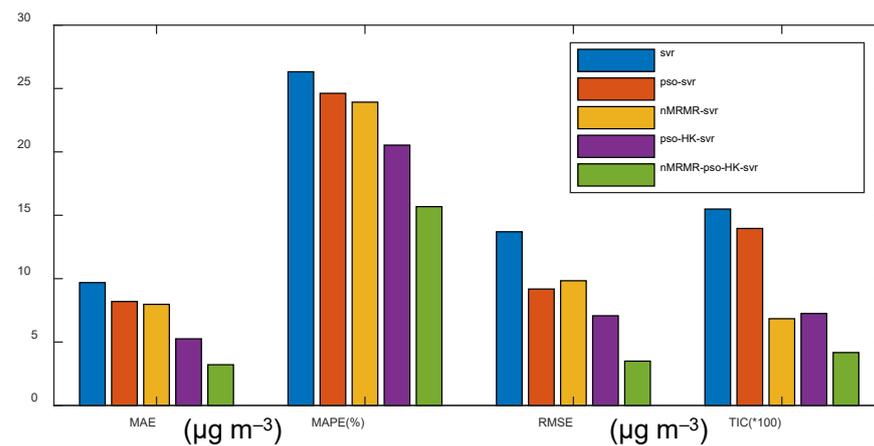
From Table 3, it can be seen that both SVR and BP models are able to complete the PM<sub>2.5</sub> concentration prediction to a certain extent, but the prediction accuracy of SVR model is somewhat improved compared to BP model. The final decision function of SVR is determined by only a few support vectors, and the complexity of the calculation depends on the number of support vectors rather than the dimension of the sample space, which in a sense avoids the "Dimensional catastrophe". Since only a few support vectors are needed to determine the final



**Fig. 4.** Comparison of short-term forecast results of five methods with measured values (Wuhan).

**Table 3.** Comparison of short-term prediction accuracy of five methods (Wuhan).

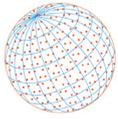
| Model            | MAE ( $\mu\text{g m}^{-3}$ ) | MAPE (%)    | RMSE ( $\mu\text{g m}^{-3}$ ) | TIC         |
|------------------|------------------------------|-------------|-------------------------------|-------------|
| BP               | 9.7                          | 26.3        | 13.7                          | 0.15        |
| SVR              | 8.2                          | 24.6        | 9.3                           | 0.11        |
| nMRMR-SVR        | 7.1                          | 21.9        | 7.8                           | 0.07        |
| PSO-HK-SVR       | 5.3                          | 20.5        | 7.2                           | 0.06        |
| nMRMR-PSO-HK-SVR | <b>4.2</b>                   | <b>15.7</b> | <b>3.5</b>                    | <b>0.04</b> |



**Fig. 5.** Comparison chart of error evaluation indexes in short-term prediction of the five methods (Wuhan).

**Table 4.** Comparison results of Comparison Test I, II and III in short-term prediction (Wuhan).

| Index    | Comparison of error performance indexes |         |            |                  |                  |
|----------|---|---------|------------|------------------|------------------|
|          | SVR                                     | nMRMR   | PSO-HK-SVR | nMRMR-PSO-HK-SVR | nMRMR-PSO-HK-SVR |
|          | vs. BP                                  | vs. SVR | vs. SVR    | vs. nMRMR-SVR    | vs. PSO-HK-SVR   |
| MAE (%)  | 15.4                                    | 13.7    | 35.9       | 40.5             | 19.9             |
| MAPE (%) | 6.5                                     | 11.0    | 16.6       | 28.5             | 23.6             |
| RMSE (%) | 33.0                                    | 14.7    | 23.0       | 55.4             | 50.6             |
| TIC (%)  | 29.2                                    | 37.6    | 45.9       | 39.7             | 30.5             |



result, it helps us to catch the key samples and "eliminate" a large number of redundant samples, which makes the algorithm more "robust". The experimental results also show that the SVR method outperforms the BP network method in terms of stability and generalization

Comparing the nMRMR-SVR model with the SVR model shows that the nMRMR-SVR model after feature selection is superior to the traditional SVR model in terms of MAE, MAPE, RMSE, and TIC evaluation indexes. From Table 4, it can be seen that compared with the SVR model the nMRMR-SVR model reduced the values of MAE, RMSE, MAPE, and TIC by 14%, 11%, 15%, and 38%

### 3.3.2 Comparison results of PSO-HK-SVR and SVR

From Table 4, it can be seen that the optimization of SVR by using the hybrid kernel function (HK) and particle swarm algorithm (PSO) significantly improves the prediction accuracy of SVR for PM<sub>2.5</sub>. Compared with SVR model, the values of MAE, RMSE, MAPE, and TIC of PSO-HK-SVR model are reduced by 15%, 6%, 32%, and 32%, 10%.

Therefore, optimizing the parameters of the kernel functions by mixing the kernel functions and PSO can improve the predictive ability of SVR. Furthermore, it can be seen that SVR cannot reduce the prediction error more effectively by relying only on a single radial basis function. This is because the radial basis kernel function, as a local kernel function, does not have a strong generalization ability, and cannot accurately track the sudden changes in PM<sub>2.5</sub> concentration, which limits the prediction accuracy of the model to a certain extent and makes it difficult for the traditional SVR model to warn of sudden air pollution events. Therefore, the use of mixed kernel functions has a significant impact on the improvement of the prediction accuracy.

### 3.3.3 Comparison results of nMRMR-PSO-HK-SVM and nMRMR-SVR, PSO-HK-SVR

From Table 3, it can be seen that the nMRMR-PSO-HK-SVR model has a higher prediction accuracy compared to the nMRMR-SVR model, PSO-HK-SVR model, and its prediction results are closer to the measured values. The four error performance indicators of the nMRMR-PSO-HK-SVR model are also minimal. From Table 4, it can be seen that compared with the nMRMR-SVR model the values of MAE, MAPE, RMSE and TIC of the nMRMR-PSO-HK-SVM model are reduced by 47%, 34%, 64% and 39%; compared with the PSO-HK-SVR model the values of MAE, MAPE, RMSE and TIC of the nMRMR-PSO-HK-SVM model are reduced by 20%, 24%, 51% and 24%.

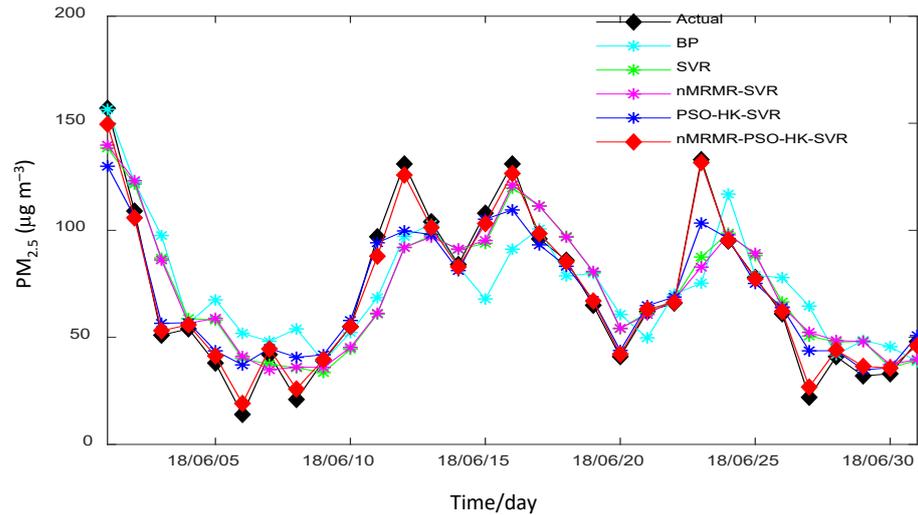
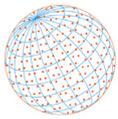
In the nMRMR-PSO-HK-SVM model, the input vector is still the optimal subset elected by the nMRMR algorithm. But in order to improve the shortcomings of the nMRMR-SVM model and enhance the generalization ability of the model, the hybrid kernel function (HK) is constructed by Eq. (8) and used as the kernel function to improve the model. Compared with the radial kernel function (Chen and Pai, 2015), the hybrid kernel function in this paper is more suitable to describe the complex changes of PM<sub>2.5</sub>.

The weight coefficients of the mixed kernel function are determined by the PSO search algorithm (Zhou *et al.*, 2017). From the prediction results in Fig. 4, the model prediction results are very close to the measured values. And the prediction accuracy is also high in some positions with large fluctuations, which further indicates that the optimized mixed kernel function can further enhance the generalization ability of the model.

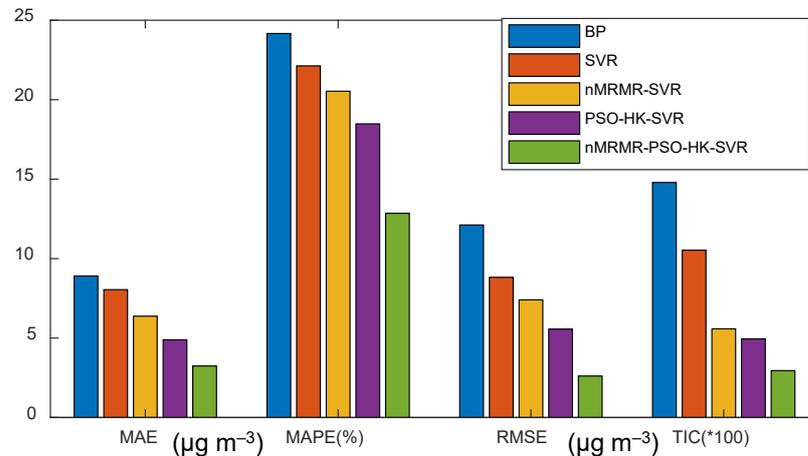
## 3.4 Short-term Prediction of PM<sub>2.5</sub> in Tianjin City

In order to further systematically and comprehensively analyze the validity and applicability of the proposed nMRMR-PSO-HK-SVR method, we use the Tianjin PM<sub>2.5</sub> data for short-term prediction experimental analysis. In order to analyze the validity of the nMRMR-PSO-HK-SVM model proposed in this paper, a total of 1065 data from 2016/01/01 to 2018/11/30 is selected as the training sample set, and a total of 31 data from 2018/12/01 to 2018/12/31 is selected as the test sample set for the short-term prediction experiment of PM<sub>2.5</sub> in Tianjin. The prediction results of PM<sub>2.5</sub> concentration for different models are shown in Fig. 6, and the prediction performance indexes of MAE, MAPE, RMSE and TIC for each prediction model are calculated and the results are shown in Fig. 7, Fig. 8, Table 5 and Table 6.

From the experimental results, it can be seen that different models can obtain similar results as those of Wuhan in predicting PM<sub>2.5</sub> in Tianjin. Compared with the BP model, SVR model,



**Fig. 6.** Comparison of short-term forecast results of five methods with measured values (Tianjin).



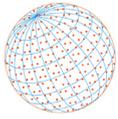
**Fig. 7.** Comparison chart of error evaluation indexes in short-term prediction of the five methods (Tianjin).

**Table 5.** Comparison of short-term prediction accuracy of five methods (Tianjin).

| Model            | MAE ( $\mu\text{g m}^{-3}$ ) | MAPE (%)    | RMSE ( $\mu\text{g m}^{-3}$ ) | TIC         |
|------------------|------------------------------|-------------|-------------------------------|-------------|
| BP               | 8.9                          | 24.2        | 12.1                          | 0.15        |
| SVR              | 8.0                          | 22.1        | 8.8                           | 0.12        |
| nMRMR-SVR        | 6.4                          | 20.5        | 7.4                           | 0.06        |
| PSO-HK-SVR       | 4.9                          | 18.5        | 5.6                           | 0.05        |
| nMRMR-PSO-HK-SVR | <b>3.2</b>                   | <b>12.9</b> | <b>2.6</b>                    | <b>0.03</b> |

**Table 6.** Comparison results of Comparison Test I, II and III in short-term prediction (Tianjin).

| Index    | Comparison of error performance indexes |           |            |                  |                  |
|----------|---|-----------|------------|------------------|------------------|
|          | SVR                                     | nMRMR-SVR | PSO-HK-SVR | nMRMR-PSO-HK-SVR | nMRMR-PSO-HK-SVR |
|          | vs. BP                                  | vs. SVR   | vs. SVR    | vs. nMRMR-SVR    | vs. PSO-HK-SVR   |
| MAE (%)  | 9.6                                     | 20.7      | 39.2       | 49.1             | 33.7             |
| MAPE (%) | 8.4                                     | 7.2       | 16.5       | 37.4             | 30.5             |
| RMSE (%) | 27.3                                    | 16.2      | 36.9       | 64.7             | 53.2             |
| TIC (%)  | 28.8                                    | 47.0      | 53.0       | 47.1             | 39.4             |



nMRMR-SVR model and PSO-HK-SVR model, the hybrid nMRMR-PSO-HK-SVR model proposed in this paper has the highest prediction accuracy. Compared with the BP prediction model, the MAE, MAPE, RMSE, and TIC of the nMRMR-PSO-HK-SVR model are reduced by 53.582%, 46.811%, 78.445%, and 80.054%. Compared with the SVR model, the MAE, MAPE, RMSE, and TIC of the nMRMR-PSO-HK-SVR model are reduced by 60%, 42%, 70%, and 72%.

Compared with the nMRMR-SVR model, the MAE, MAPE, RMSE, and TIC of the nMRMR-PSO-HK-SVR model are reduced by 49%, 37%, 65%, and 47%. Compared with the PSO-HK-SVR model, the MAE, MAPE, RMSE, and TIC of the nMRMR-PSO-HK-SVR model are reduced by 34%, 30%, 53%, and 39%. The experimental results for the prediction of PM<sub>2.5</sub> in Tianjin further confirmed that the proposed model is very suitable for the prediction of PM<sub>2.5</sub> concentration and has a strong adaptive capability. It also demonstrates that the hybrid model (nMRMR-PSO-HK-SVR) has better prediction performance than the single input feature selection model (nMRMR-SVR) and the single kernel function optimization model (PSO-HK-SVR). From the experimental results we can also see that the input feature selection (nMRMR) and particle swarm optimization hybrid kernel function (PSO-HK) can indeed improve the prediction ability of SVR.

### 3.5 Long-term Prediction of PM<sub>2.5</sub> in Wuhan City

In order to evaluate the performance of the model more comprehensively, a total of 1096 data from 2016/01/01 to 2018/12/31 in Wuhan is selected as the training sample set, and a total of 181 data from 2019/01/01 to 2019/06/30 is selected as the test sample set for the long-term prediction experiment of PM<sub>2.5</sub>.

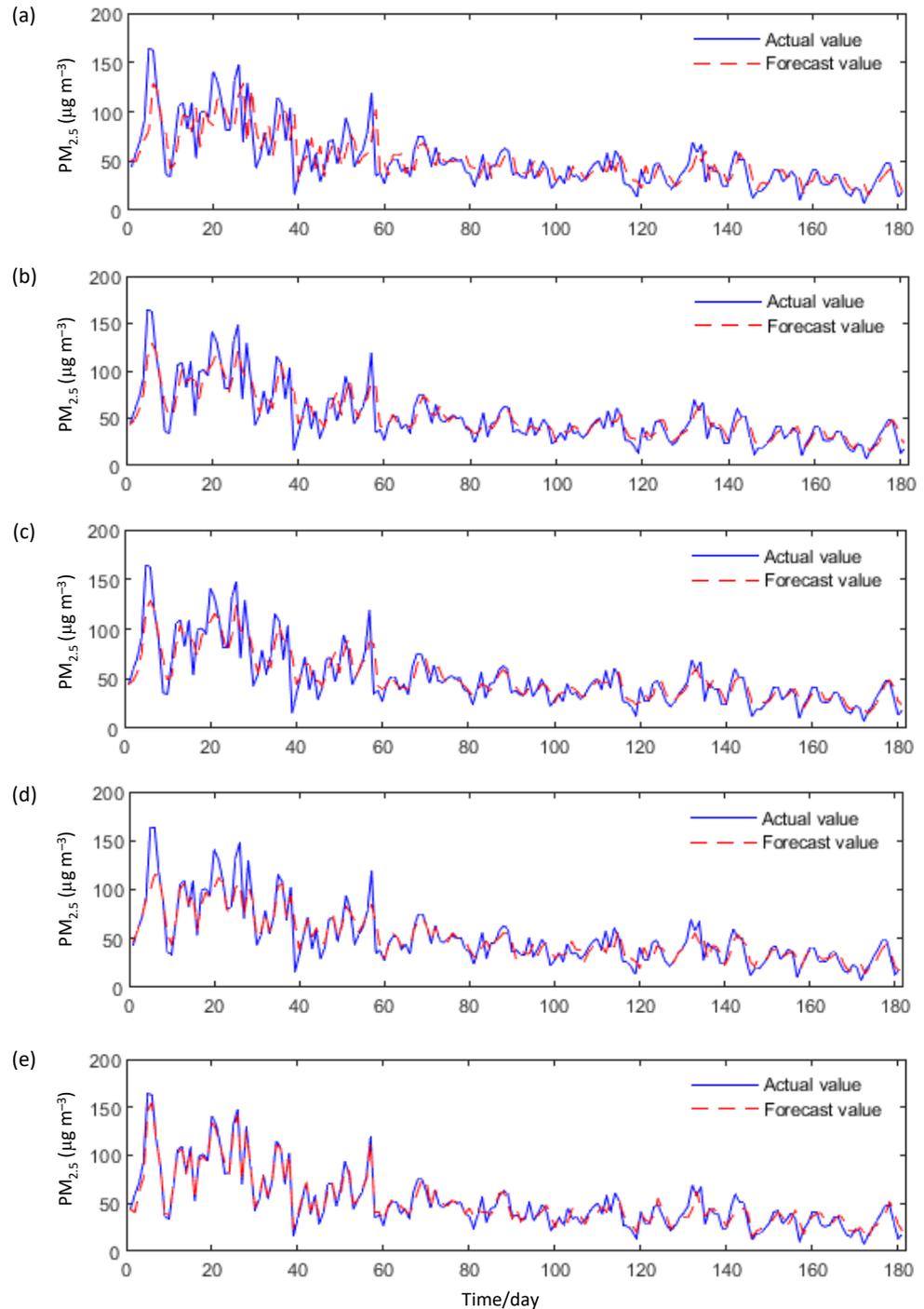
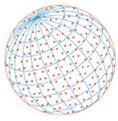
Fig. 8 presents the results of the comparison between predicted and measured values for the long-term predictions of the different models. It can be seen that the trend of the predicted and measured values of the nMRMR-PSO-HK-SVR model is almost the same, especially for the location of the "PM<sub>2.5</sub> concentration peak", which also has a good dependence effect. The long-term prediction results of PM<sub>2.5</sub> concentration in Wuhan show that the nMRMR-PSO-HK-SVR model proposed in this paper can be adapted to the long-term prediction needs, and can be used not only for the prediction of PM<sub>2.5</sub> concentration in the weather with good air quality, but also for the prediction of PM<sub>2.5</sub> concentration in the heavily polluted weather. Table 7 gives the values of the error performance indicators for the five models when predicted in the long term. As can be seen in Table 7, the four error performance metrics of the nMRMR-PSO-HK-SVR model are also minimal in the long-term predictions. The MAE, MAPE, RMSE, and TIC decreased by 6, 11%, 10, and 0.13 compared to the SVR model. The MAE, MAPE, RMSE, and TIC decreased by 6, 11%, 10 and 0.13 compared to the nMRMR-SVR model. The MAE, MAPE, RMSE, and TIC decreased by 4, 8%, 6, and 0.05 compared to the PSO-SVR model. The MAE, MAPE, RMSE, and TIC decreased by 1, 5%, 4, and 0.04 compared to the nMRMR-PSO-HK-SVR model.

The scatter plot of the predicted and measured values of the five methods on the test set is shown in Fig. 9. In Fig. 9, the horizontal coordinates indicate the measured values, the vertical coordinates indicate the predicted values, the trend line indicates the regression line between the predicted and measured values, and R<sup>2</sup> indicates the coefficient of determination, with the closer R<sup>2</sup> to 1 indicating a stronger linear relationship between the predicted and measured values. From the scatter plots, it can be seen that the R<sup>2</sup> for the predicted and measured values of the SVR model, nMRMR-SVR model, PSO-SVR model, PSO-HK-SVR model and nMRMR-PSO-HK-SVR model are 0.74, 0.88, 0.89, 0.92, and 0.97. nMRMR-PSO-HK-SVR model has the largest R<sup>2</sup> which is close to 1, indicating that the nMRMR-PSO-HK-SVR model is also close to the measured value in the long-term prediction of PM<sub>2.5</sub>.

Based on the above experimental results, the combined nMRMR and PSO-HK-SVR can achieve accurate prediction of PM<sub>2.5</sub> concentrations over longer periods and under different weather quality conditions, indicating that the nMRMR-PSO-HK-SVR model is feasible and reliable for application.

## 4 CONCLUSIONS

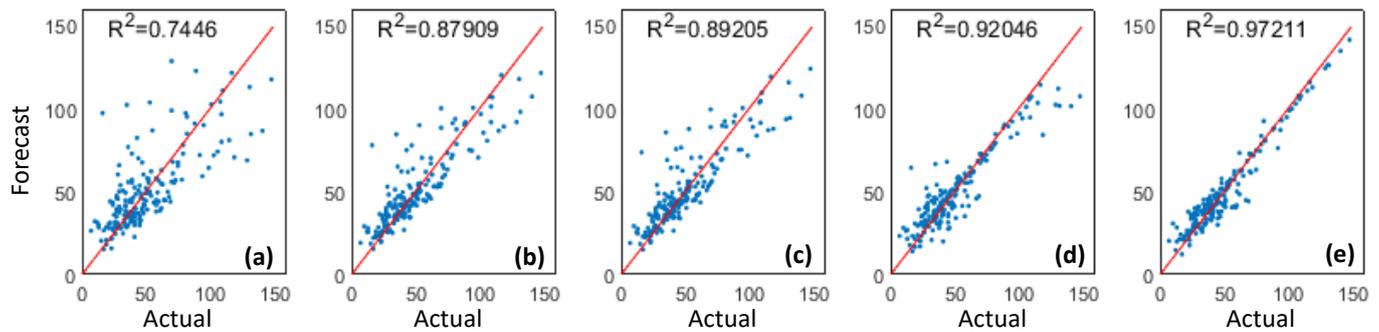
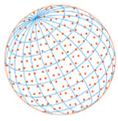
Fine particulate matter PM<sub>2.5</sub> is an important air pollution measurement data, and the prediction of PM<sub>2.5</sub> is of great significance for environmental protection. Considering the efficiency, practicability and accuracy of the prediction, this manuscript firstly extracts the optimal feature subset from



**Fig. 8.** The  $PM_{2.5}$  long term prediction result of five models (Wuhan) (a) prediction result of SVR (b) prediction result of PSO-SVR (c) prediction result of nMRMR-SVR (d) prediction result of PSO-HK-SVR (e) prediction result of nMRMR-PSO-HK-SVR.

**Table 7.** Comparison of long-term prediction accuracy of five methods (Wuhan).

| Model            | MAE ( $\mu\text{g m}^{-3}$ ) | MAPE (%)    | RMSE ( $\mu\text{g m}^{-3}$ ) | TIC         |
|------------------|------------------------------|-------------|-------------------------------|-------------|
| SVR              | 9.8                          | 26.4        | 13.8                          | 0.18        |
| PSO-SVR          | 8.2                          | 25.0        | 9.3                           | 0.15        |
| nMRMR-SVR        | 8.1                          | 24.3        | 9.9                           | 0.10        |
| PSO-HK-SVR       | 5.4                          | 20.8        | 7.1                           | 0.09        |
| nMRMR-PSO-HK-SVR | <b>4.2</b>                   | <b>15.8</b> | <b>3.5</b>                    | <b>0.05</b> |



**Fig. 9.** Long-term predicted scatter plot of PM<sub>2.5</sub> in the Wuhan region. (a) BP model, (b) SVR model, (c) nMRMR-SVR model, (d) PSO-HK-SVR model, (e) nMRMR-PSO-HK-SVR model.

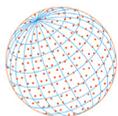
the air quality data and meteorological data by the approximate maximum correlation minimum redundancy algorithm. Then, the optimal feature subset was used as the input, and the mixed kernel function support vector regression model was used to predict the PM<sub>2.5</sub> concentration in the next 24 h (the next day). The optimal parameters of the hybrid kernel function can be adaptively determined by the particle swarm optimization algorithm.

- (1) Air quality elements and weather elements are both strongly correlated with PM<sub>2.5</sub> concentration. However, data redundancy between features exists because some features are strongly intercorrelated. This hinders the precision of the SVR model in predicting PM<sub>2.5</sub> concentration.
- (2) The approximate Markov blanket-based nMRMR algorithm considers the correlations among features while considering the correlations between ordinary features. The optimal feature subset selected by the algorithm can retain the majority of data even though its dimensionality is reduced.
- (3) When using the SVR model for PM<sub>2.5</sub> concentration prediction, the prediction precision is strongly affected by the kernel function employed. Because of the high complexity of PM<sub>2.5</sub> concentration variation, a single kernel function has difficulty describing all the varying features. By comparison, a PSO-based HK is more appropriate for describing complex variation.
- (4) The prediction experiments of PM<sub>2.5</sub> concentrations in Wuhan 2018 and 2019 show that the nMRMR-PSO-HK-SVM model has higher prediction accuracy. Compared with the traditional SVR model, the MAE, MAPE, RMSE and TIC of PM<sub>2.5</sub> short-term prediction decreased by 5, 11%, 10, 0.11, respectively. The MAE, MAPE, RMSE, and TIC of PM<sub>2.5</sub> long-term prediction decreased by 6, 11%, 10, and 0.13 respectively.

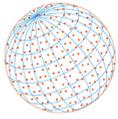
In the case of known air quality and weather data, the proposed model in the manuscript can effectively predict the PM<sub>2.5</sub> concentration value in the next day (24 h). After adjusting the predicted data structure, the proposed model can also be used to predict the PM<sub>2.5</sub> concentration in the next 2 days (48 h) or 3 days (72 h), however, the accuracy of long-term forecasting is affected by the accumulation of forecasting errors. How to make a more accurate long-term prediction of PM<sub>2.5</sub> concentration is what we will continue to study in the future

## REFERENCES

- Cai, Z., Ma, H., Zhang, L. (2019). Feature selection for airborne LiDAR data filtering: A mutual information method with Parzon window optimization. *GISci. Remote Sens.* 57, 323–337. <https://doi.org/10.1080/15481603.2019.1695406>
- Chen, J.F., Li, Y. (2019). Forecasting of PM<sub>2.5</sub> concentration based on multimodal support vector regression. *Environ. Eng.* 37, 122–126. (in Chinese)
- Chen, L., Pai, T.Y. (2015). Comparisons of GM (1, 1), and BPNN for predicting hourly particulate matter in Dali area of Taichung City, Taiwan. *Atmos. Pollut. Res.* 6, 572–580. <https://doi.org/10.5094/apr.2015.064>
- Cura, T. (2020). Use of support vector machines with a parallel local search algorithm for data classification and feature selection. *Expert Syst. Appl.* 145, 113133. <https://doi.org/10.1016/j.eswa.2019.113133>



- Dhamecha, T.I., Noore, A., Singh, R., Vatsa, M. (2019). Between-subclass piece-wise linear solutions in large scale kernel SVM learning. *Pattern Recognit.* 95, 173–190. <https://doi.org/10.1016/j.patcog.2019.04.012>
- Fei, S. (2016). A hybrid model of EMD and multiple-kernel RVR algorithm for wind speed prediction. *Int. J. Electr. Power Energy Syst.* 78, 910–915. <https://doi.org/10.1016/j.ijepes.2015.11.116>
- Gu, F., Hu, M., Wang, Y., Li, M., Guo, Q., Wu, Z. (2016). Characteristics of PM<sub>2.5</sub> pollution in winter and spring of Beijing during 2009–2010. *China Environ. Sci.* 36, 2578–2584. (in Chinese with English Abstract)
- Hua, Z., Zhou, J., Hua, Y., Zhang, W. (2020). Strong approximate Markov blanket and its application on filter-based feature selection. *Appl. Soft Comput.* 87, 105957. <https://doi.org/10.1016/j.asoc.2019.105957>
- Huanrui, H. (2016). New mixed kernel functions of SVM used in pattern recognition. *Cybern. Inf. Technol.* 16, 5–14. <https://doi.org/10.1515/cait-2016-0047>
- Jayakumar, C., Sangeetha, J. (2020). Kernellized support vector regressive machine based variational mode decomposition for time frequency analysis of Mirnov coil. *Microprocess. Microsyst.* 75, 103036. <https://doi.org/10.1016/j.micpro.2020.103036>
- Jiao, W., Liu, G.B. (2009). An Improved Particle Swarm Optimization Algorithm with Immunity. 2009 Second International Conference on Intelligent Computation Technology and Automation, Presented at the 2009 Second International Conference on Intelligent Computation Technology and Automation, pp. 241–244. <https://doi.org/10.1109/ICICTA.2009.66>
- Ju, Z., He, J.J. (2018). Prediction of lysine glutarylation sites by maximum relevance minimum redundancy feature selection. *Anal. Biochem.* 550, 1–7. <https://doi.org/10.1016/j.ab.2018.04.005>
- Kim, M.H., Kim, Y.S., Lim, J., Kim, J.T., Sung, S.W., Yoo, C. (2010). Data-driven prediction model of indoor air quality in an underground space. *Korean J. Chem. Eng.* 27, 1675–1680. <https://doi.org/10.1007/s11814-010-0313-5>
- Liu, P., Choo, K.K.R., Wang, L., Huang, F. (2016). SVM or deep learning? A comparative study on remote sensing image classification. *Soft Comput.* 21, 7053–7065. <https://doi.org/10.1007/s00500-016-2247-2>
- Ni, X.Y., Huang, H., Du, W.P. (2017). Relevance analysis and short-term prediction of PM<sub>2.5</sub> concentrations in Beijing based on multi-source data. *Atmos. Environ.* 150, 146–161. <https://doi.org/10.1016/j.atmosenv.2016.11.054>
- Niu, M., Wang, Y., Sun, S., Li, Y. (2016). A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM<sub>2.5</sub> concentration forecasting. *Atmos. Environ.* 134, 168–180. <https://doi.org/10.1016/j.atmosenv.2016.03.056>
- Patricio, P., Camilo, M., Camilo, R. (2020). PM<sub>2.5</sub> forecasting in Coyhaique, the most polluted city in the Americas. *Urban Clim.* 32, 100608. <https://doi.org/10.1016/j.uclim.2020.100608>
- Perez, P., Gramsch, E. (2016). Forecasting hourly PM<sub>2.5</sub> in Santiago de Chile with emphasis on night episodes. *Atmos. Environ.* 124, 22–27. <https://doi.org/10.1016/j.atmosenv.2015.11.016>
- Qiao, J., Cai, J., Han, H. (2017). Predicting PM<sub>2.5</sub> concentrations at a regional background station using second order self-organizing fuzzy neural network. *Atmosphere* 8, 10. <https://doi.org/10.3390/atmos8010010>
- Qin, X.W., Liu, Y.Y., Wang, X.M., Dong, X.G., Zhang, Y., Zhou, H.M. (2016). PM<sub>2.5</sub> prediction of Beijing city based on ensemble empirical mode decomposition and support vector regression. *J. Jilin Univ. (Earth Science Edition)* 46, 563–568.
- Singh, K.P., Gupta, S., Kumar, A., Shukla, S.P. (2012). Linear and nonlinear modeling approaches for urban air quality prediction. *Sci. Total Environ.* 426, 244–255. <https://doi.org/10.1016/j.scitotenv.2012.03.076>
- Song, G., Guo, X., Yang, X., Liu, S. (2018). ARIMA-SVM combination prediction of PM<sub>2.5</sub> concentration in Shenyang. *China Environ. Sci.* 38, 4031–4039. (in Chinese with English Abstract)
- Sun, W., Sun, J. (2016). Daily PM<sub>2.5</sub> concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm. *J. Environ. Manage.* 188, 144–152. <https://doi.org/10.1016/j.jenvman.2016.12.011>
- Vapnik, V. (1998). *Statistical learning Theory*. New York.
- Wang, H., Ling, Z., Yu, K., Wu, X. (2020). Towards efficient and effective discovery of Markov blankets for feature selection. *Inf. Sci.* 509, 227–242. <https://doi.org/10.1016/j.ins.2019.09.010>



- Wang, L.M., Wu, X.H., Zhao, T.L., Cheng, G.S., Zhang, X.Z., Tang, L.L., Jia, M.W., Chen, Y.S. (2017a). A scheme for rolling statistical forecasting of PM<sub>2.5</sub> concentrations based on distance correlation coefficient and support vector regression. *Acta Sci. Circumst.* 37, 1268–1276. (in Chinese with English Abstract)
- Wang, P., Zhang, H., Qin, Z., Zhang, G. (2017b). A novel hybrid-Garch model based on ARIMA and SVM for PM<sub>2.5</sub> concentrations forecasting. *Atmos. Pollut. Res.* 8, 850–860. <https://doi.org/10.1016/j.apr.2017.01.003>
- Wang, P., Zhang, H., Qin, Z.Z., Yao, Q.C., Geng, H. (2017c). PM<sub>10</sub> concentration forecasting model based on Wavelet-SVM. *Environ. Sci.* 38, 3153–3161 <https://doi.org/10.13227/j.hjkk.201612194>
- Wang, Z., Li, Y., Chen, T., Zhang, D., Sun, F., Pan, L. (2015). Spatial-temporal characteristics of PM<sub>2.5</sub> in Beijing in 2013. *Acta Geogr. Sin.* 70, 110–120. (in Chinese)
- Yan, H., Zhang, J., Rahman, S. S., Zhou, N., Suo, Y. (2020). Predicting permeability changes with injecting CO<sub>2</sub> in coal seams during CO<sub>2</sub> geological sequestration: A comparative study among six SVM-based hybrid models. *Sci. Total Environ.* 705, 135941. <https://doi.org/10.1016/j.scitotenv.2019.135941>
- Yin, J.G, Peng, F, Xie, L.K, Xu, Y., Liu, H., Gong, Q.Q., Wang, K. (2018). The study on the prediction of the PM<sub>2.5</sub> concentration based on model of the least squares support vector regression under wavelet decomposition and adaptive multiple layer residuals correction. *Acta Sci. Circumst.* 38, 2090–2098. (in Chinese with English Abstract)
- Yu, K., Liu, L., Li, J. (2020). Learning markov blankets from multiple interventional data sets. *IEEE Trans. Neural Networks Learn. Syst.* 31, 2005–2019. <https://doi.org/10.1109/TNNLS.2019.2927636>
- Zhang, K. (2019). Forecasting regional economic growth using support vector machine model. *Ecol. Econ.* 15, 186–192.
- Zhang, L., Wang, Z. (2018). Multi-label feature selection algorithm based on joint mutual information of max-relevance and min-redundancy. *J. Comm.* 39, 111–122. (in Chinese with English Abstract) <https://doi.org/10.11959/j.issn.1000-436x.2018082>
- Zhong, Z., Carr, T.R. (2016). Application of mixed kernels function (MKF) based support vector regression model (SVR) for CO<sub>2</sub>-Reservoir oil minimum miscibility pressure prediction. *Fuel* 184, 590–603. <https://doi.org/10.1016/j.fuel.2016.07.030>
- Zhou, G.Q., Gao, W., Gu, Y.X., Qu, Y (2017). Impact of precipitation on Shanghai PM<sub>2.5</sub> forecast using WRF-Chem. *Acta Sci. Circumst.* 37, 4476–4482. (in Chinese with English Abstract)