



Air Pollutant Concentration Forecast Based on Support Vector Regression and Quantum-Behaved Particle Swarm Optimization

Xiaoli Li^{1,2,3} · Aorong Luo^{1,2} · Jiangeng Li^{1,2} · Yang Li⁴

Received: 7 April 2017 / Accepted: 16 August 2018 / Published online: 29 September 2018
© Springer Nature Switzerland AG 2018

Abstract

In order to improve the forecasting accuracy of atmospheric pollutant concentration, a prediction model of atmospheric PM_{2.5} and nitrogen dioxide (NO₂) concentration based on support vector regression (SVR) is established. Quantum-behaved particle swarm optimization (QPSO) algorithm is used to select the optimal parameters influencing the performance of SVR. And in order to improve the problem that the fixed SVR model is difficult to adapt to the highly nonlinear process, a simple online SVR based on re-modeling method is proposed instead of the fixed one. According to hourly PM_{2.5} and NO₂ concentrations and meteorological conditions from May 2014 to April 2015 in Wanliu Monitoring Station of Beijing in China, the experiment is carried out based on the data of 3 months. Meanwhile, PM_{2.5} concentration is predicted by three different prediction methods, including the recursive prediction method, direct prediction method, and online direct prediction method. The results show that the online direct prediction method is the most accurate in the three prediction methods. In addition, compared with original particle swarm optimization (PSO) algorithm, QPSO algorithm is tested more efficiently for the improvement of global search ability and robustness during the procedure of parameter selection. Moreover, the hybrid QPSO-SVR model proposed in this paper has higher prediction accuracy and less computational time compared with the PSO-SVR model, genetic algorithm (GA)-SVR model, and grid search (GS)-SVR model, which indicates reliability and effectiveness of the QPSO-SVR model in prediction of these two pollutant concentrations.

Keywords SVR · QPSO · Recursive prediction · Direct prediction

Abbreviations

AI	Artificial intelligence	PSO	Particle swarm optimization
ARMA	Autoregressive moving average	QPSO	Quantum-behaved particle swarm optimization
ANN	Artificial neural network	GA	Genetic algorithm
SVM	Support vector machine	GS	Grid search
SVR	Support vector regression	PM	Particulate matter
PLS	Partial least squares	NO ₂	Nitrogen dioxide
		CO	Carbon monoxide
		CO ₂	Carbon dioxide
		SO ₂	Sulfur dioxide
		CH ₄	Methane
		NO _x	Nitrogen oxides
		O ₃	Ozone

✉ Jiangeng Li
lijg@bjut.edu.cn

Xiaoli Li
lixiaolibjut@bjut.edu.cn

- ¹ Faculty of Information Technology, Beijing University of Technology, Beijing, 100124, China
- ² Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing, 100124, China
- ³ Engineering Research Center of Digital Community, Ministry of Education, Beijing, 100124, China
- ⁴ School of International Studies, Communication University of China (CUC), Beijing, 100024, China

1 Introduction

Recently, with the rapid development of urbanization and industrialization in China, a large amount of harmful substances have been released into the atmosphere, and more and more attention has been paid on the transformation of the air pollutants data, such as carbon monoxide (CO), carbon dioxide (CO₂), sulfur dioxide (SO₂), methane (CH₄),

nitrogen oxides (NO_x), ozone (O₃), and particulates (PM_{2.5} and PM₁₀). These harmful substances affect the urban air quality and pose a great threat to human health [24, 47]. Many regions have suffered from serious air pollution, especially Beijing, Tianjin, Hebei, and Shandong province in China [51, 56].

PM_{2.5} and nitrogen dioxide (NO₂), as dominant pollutants, have attracted wide attention [13, 15]. PM_{2.5} refers to the particulate matter whose aerodynamic diameter is 2.5 μm or less. It is made of toxic and hazardous substances with high activity; it has the character of long residence time and far transportation distance in the atmosphere [43]. The sources of PM_{2.5} include fuel combustion from automobiles, power plants, wood burning, industrial processes, and vehicles such as buses and trucks. It is also formed in the atmosphere when gases such as SO₂ and NO_x and volatile organic compounds are transformed in the air by chemical reactions. NO₂ is a poisonous gas with reddish brown and pungent odor at room temperature. Its participation in the photochemical reaction catalyzes ozone production, thus leads to photochemical smog pollution. NO₂ is mainly derived from the fossil fuel and biomass burning, soil emissions, and lightning. Meanwhile, the contribution of anthropogenic sources accounts for a larger proportion, including motor vehicle emissions, power plants, and other industrial sources [11, 27]. Numerous studies [8, 9, 40, 55] have shown that exposure to high levels of NO₂ and PM_{2.5} leads to breathing difficulty, lung and cardiovascular diseases, acid deposition, and eco-environmental system damages. To provide an early warning for air quality changes and protect human health and environment, an effective and accurate model for the short- and long-term forecasts of PM_{2.5} and NO₂ concentration is more necessary [48, 53, 54].

Forecasting methods can be divided into three main categories, i.e., numerical methods, statistical methods, and artificial intelligence (AI)-based methods [21]. A large number of numerical models [20, 44, 50], such as box model, Gaussian model, Lagrangian model, and Euler model, have been used for air pollutant concentration forecast. These models can simulate the physical and chemical process in the atmosphere and they are also called atmospheric dispersion models. However, such models are restricted in many operational conditions because they require accurate and detailed data, such as meteorology, terrain geomorphology, pollution sources, and other data [6, 37]. For the statistical methods, multiple regression model [13, 45], grey model [34], Kalman filter techniques [38], and autoregressive moving average (ARMA) model [25] have been widely used to forecast air pollutant levels; such models can be generalized and are consistent with actual observations. However, due to the existence of strong nonlinearity problem of air pollutant concentration, the predicting accuracy is difficult to improve by using the abovementioned methods [32].

In recent decades, the AI-based methods have aroused public interest in air pollutant concentration forecasting. Among them, artificial neural network (ANN) and support vector machine (SVM) are more popular. ANN is good at solving nonlinear problem and is considered as a promising forecasting tool [4, 16, 41]. Moustris et al. [29] presented an ANN to forecast the maximum daily value of pollutants index in Athens and Greece. The results indicated that ANN could give reliable forecast for the air quality. Gennaro et al. [10] proposed an ANN to forecast daily PM₁₀ concentration in regional site and urban site. The results showed that ANN could be a powerful tool to obtain real-time information on air quality status. Feng et al. [15] introduced a novel hybrid model combining air mass trajectory analysis and wavelet transformation to improve the ANN accuracy for PM_{2.5} concentration forecast. The mass trajectory was applied to recognize different corridors, and the wavelet transformation was used to deal with the fluctuation of PM_{2.5} concentration. Nevertheless, ANN suffers from a number of weakness, such as overfitting problem, local minimal problem, network construction problem, and the need of a large number of data for network training. So there are more difficulties when ANN is applied to some forecasting problems [2, 41].

SVM has been proposed on the basis of statistical learning approach and it overcomes the shortcomings of ANN model [39]. It employs the structural risk minimization principle to obtain the global optimum, instead of empirical risk minimization principle. Originally, SVM was applied for pattern classification. With the introduction of ϵ -insensitive loss function, SVM was gradually developed to solve the nonlinear regression estimation and time series prediction problems [17, 46], namely support vector regression (SVR). Ortiz-García et al. [32] established SVR model to forecast hourly O₃ concentration in Madrid urban area, and the model parameter was optimized by an improved grid search method. The findings showed that the SVR model is superior to multi-layer perceptron. Yeganeh et al. [49] used a hybrid model based on partial least squares (PLS) and SVM to forecast hourly and daily CO concentration. The results indicated that this hybrid model performed faster prediction and more accurate ability. Moazami et al. [28] applied SVR model to predict the carbon monoxide (CO) concentrations of the next day in Tehran metropolitan; the results showed that the SVR has less uncertainty in CO prediction than adaptive neuro-fuzzy inference system (ANFIS) and ANN models.

However, some shortcomings still exist in these studies. On one hand, the original time series of air pollutant concentration is highly nonlinear and time-varying. The fixed SVR model is difficult to adapt to this feature, while the online SVR model can update model dynamically; therefore, the online SVR model based on re-modeling

method is used to predict air pollutant concentration in this study. On the other hand, SVR model performance is greatly affected by three parameters (penalty factor C , kernel parameter σ , and insensitive coefficient ε). The traditional methods, such as grid search, cross validation, and gradient descend, exist some limitations due to low calculation efficiency and poor accuracy [2, 19]. Therefore, it is necessary to overcome these shortcomings. Heuristic algorithm is a kind of local optimization algorithm based on intuition or experience; it is applied in many fields, such as the optimization of neural network [30], the optimization of scheduling problem [35, 36], and so on. While for the parameter selection, the heuristic algorithms have also shown great superiority. Several heuristic algorithms have been applied to select parameters, such as genetic algorithm [14], immune algorithm [26], and simulated annealing algorithm [33]. However, compared with particle swarm optimization (PSO) algorithm, these methods perform slow search speed and poor accuracy in multi-dimensional optimization problems [5, 12]. PSO algorithm was introduced by Kennedy and Eberhart [22]; it is equipped with the mechanism of memory and has a simple structure. Therefore, it is more suitable to select the SVR parameters [12, 52].

In order to prevent premature convergence and local minimum of the standard PSO algorithm, a quantum-behaved PSO algorithm (QPSO) is applied, and a hybrid QPSO-SVR model is established to forecast $\text{PM}_{2.5}$ and NO_2 concentration. At the same time, in order to select the optimal prediction method, the recursive multi-step prediction, direct multi-step prediction, and online direct multi-step prediction methods are compared to predict $\text{PM}_{2.5}$ concentration in three selected months.

The rest of this paper is organized as follows: Section 2 describes the preliminary knowledge of mathematics, including SVR method, QPSO algorithm, and multi-step prediction method. Section 3 shows the data required for the experiment, and the experiment results for $\text{PM}_{2.5}$ and NO_2 concentration prediction. Finally, conclusions are given in Section 4.

2 Preliminary Knowledge of Mathematics Algorithm and Model

2.1 Support Vector Regression Model

Support vector machine (SVM) was developed on the basis of statistical learning [1]. In 1992, Boser, Guyon, and Vapnik proposed the optimal boundary learning theory in the conference paper about computational learning for the first time, which was also the initial form of SVM. In 1995, Vapnik proposed a SVM learning algorithm completely; it

had outstanding advantages in theory and it realized the nonlinear mapping of the high-dimensional space by kernel function, and it was used to solve nonlinear classification and regression estimation problems.

In conventional ε -support vector regression (ε -SVR) algorithm, the basic idea is to map the input vector into a high-dimensional feature space via a nonlinear mapping function. The structure risk minimization principle is applied to construct the optimal decision function in the feature space so that the relationship between the input and the output is approximated. Given the data set $\{(\mathbf{x}_i, y_i), i = 1, 2, \dots, l\}$ (\mathbf{x}_i is the input vector, y_i is the desired value, l is the number of samples), the regression estimation can be performed by the following formula:

$$f(\mathbf{x}) = \omega^T \phi(\mathbf{x}) + b \quad (1)$$

where ω and b are the coefficients to be adjusted, and $\phi(\mathbf{x})$ is a mapping function of the input vector in the high-dimensional space. These can be estimated by minimizing the structure risk function described as follows:

$$R(f) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l L_\varepsilon(y_i, f(\mathbf{x}_i)) \quad (2)$$

where $\frac{1}{2} \|\omega\|^2$ is used as a measurement of function smoothness, and C is a regularized constant determining the trade-off between the model complexity and promotion ability. The ε -insensitive loss function is denoted by L_ε and is described as the following:

$$L_\varepsilon(y_i, f(\mathbf{x}_i)) = \begin{cases} 0 & |y_i - f(\mathbf{x}_i)| < \varepsilon, \\ |y_i - f(\mathbf{x}_i)| - \varepsilon & |y_i - f(\mathbf{x}_i)| \geq \varepsilon \end{cases} \quad (3)$$

where y and $f(\mathbf{x})$ are the observation and predictive value respectively. This function is utilized to panelize the training error between y and $f(\mathbf{x})$. The above problem to find ω and b can be expressed in the form of convex quadratic programming, which can be described as follows:

$$\begin{cases} \min_{\omega, b} \left(\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \right) \\ s.t. \begin{cases} y_i - \omega \phi(\mathbf{x}_i) - b \leq \varepsilon + \xi_i & i = 1, 2, \dots, l \\ -y_i + \omega \phi(\mathbf{x}_i) + b \leq \varepsilon + \xi_i^* & i = 1, 2, \dots, l \\ \xi_i \geq 0, \xi_i^* \geq 0 & i = 1, 2, \dots, l \end{cases} \end{cases} \quad (4)$$

where ε defines the error requirement of regression function, which determines the number of support vectors and guarantees the sparseness of the solution. The slack variables ξ_i, ξ_i^* are used to control the upper and lower bounds of the output.

In order to solve the above quadratic programming problem, the Lagrange function is introduced. In this case, the dual form of optimization problem is described as follows:

$$\begin{cases} \max_{\alpha, \alpha^*} [-\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^l (\alpha_i + \alpha_i^*)\varepsilon + \sum_{i=1}^l (\alpha_i, \alpha_i^*)y_i] \\ s.t. \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i \leq C \\ 0 \leq \alpha_i^* \leq C \end{cases} \end{cases} \tag{5}$$

where α_i and α_i^* are the Lagrange multipliers. The function $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)\phi(\mathbf{x}_j)$ is the kernel matrix and can be replaced by any function satisfying the Mercer’s condition. A common election for this kernel function is the radial basis function (RBF):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right) \tag{6}$$

where σ is the width of RBF; it reflects the degree of correlation between support vectors. The impact of support vector is too strong to achieve sufficient accuracy if σ is too large; in contrast, the support vector is relatively loose if σ is too small, and the model is relatively complex.

By solving the optimization problem described above, the coefficients of Eq. (1) can be found as the following:

$$\omega^* = \sum_{i=1}^l (\alpha_i - \alpha_i^*)\phi(\mathbf{x}_i) \tag{7}$$

$$b^* = \frac{1}{N_{nsv}} \left\{ \sum_{0 < \alpha_i < C} [y_i - \sum_{\mathbf{x}_j \in SV} (\alpha_i - \alpha_i^*)K(\mathbf{x}_i, \mathbf{x}_j) - \varepsilon] + \sum_{0 < \alpha_i < C} [y_i - \sum_{\mathbf{x}_j \in SV} (\alpha_j - \alpha_j^*)K(\mathbf{x}_i, \mathbf{x}_j) + \varepsilon] \right\} \tag{8}$$

where N_{nsv} is the number of normal support vectors, and SV is the support vector. The following equation is the regression function:

$$\begin{aligned} f(\mathbf{x}) &= \omega^*\phi(\mathbf{x}) + b^* = \sum_{i=1}^l (\alpha_i - \alpha_i^*)\phi(\mathbf{x}_i)\phi(\mathbf{x}) + b^* \\ &= \sum_{i=1}^l (\alpha_i - \alpha_i^*)K(\mathbf{x}_i, \mathbf{x}) + b^* \end{aligned} \tag{9}$$

The fixed ε -SVR takes the existing sample data to build the model and then predicts the unknown value based on the established fixed model. While, for the highly nonlinear and time-varying data, it is difficult for fixed SVR model to adapt to such characteristics, and this leads to the decrease of prediction accuracy. Therefore, an online SVR model based on re-modeling method is proposed to overcome this shortcoming. The main idea of this approach is to

re-establish the SVR model based on the online updated time series. When a new sample arrives, it is added to the previous training set and then a new SVR model is obtained, and this new model is used for the next forecast. The single forecasting process of the fixed SVR model and the proposed online SVR model are shown in Fig. 1.

As mentioned above, the SVR parameters (C , σ , and ε) affect the performance of the model. Hence, it is essential to select appropriate parameter, and a quantum-behaved particle swarm optimization algorithm is utilized to find the proper SVR parameters.

2.2 Particle Swarm Optimization

2.2.1 The Original Particle Swarm Optimization

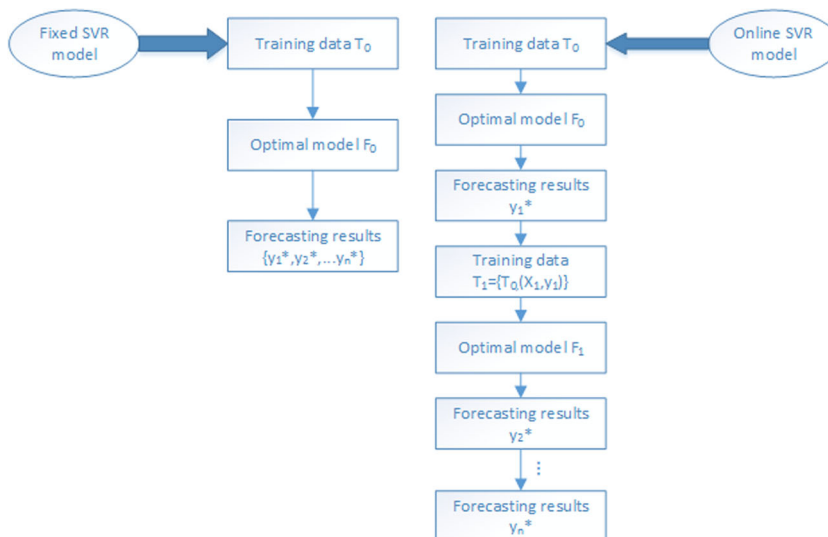
PSO [23] is a kind of stochastic optimization algorithm on the basis of population intelligence. It features a feasible and simple structure without gradient information. In continuous function optimization problems especially, it shows advantage in performance, such as the speed of convergence, computational time, and so on. Hence, it has become a hot research algorithm in the field of intelligent optimization. The basic principle is described below.

A swarm consists of m particle flies with a certain speed in a D -dimensional search space, and each particle represents a bird in the search space. For the problem to be solved, a potential solution is determined by a particle, and each particle has a velocity that determines the distance and direction of its flight. Moreover, all particles have a fitness value determined by the optimized function. In the process of flight, the particles will be adjusted dynamically by their own and group flight experience. After several iterations, the optimal solution is obtained. In each iteration, the particle updates itself by tracking two “extremes,” one is the optimal solution found by itself, called the individual extremum, another is the optimal solution found by the whole population, called the global extremum. Their velocity and position are updated according to the following equations.

$$\mathbf{V}_{(i+1)} = \omega \cdot \mathbf{V}_i + c_1 \cdot r_1 \cdot (\mathbf{p}_{Best_i} - \mathbf{X}_i) + c_2 \cdot r_2 \cdot (g_{Best} - \mathbf{X}_i) \tag{10}$$

$$\mathbf{X}_{i+1} = \mathbf{X}_i + \mathbf{V}_{(i+1)} \tag{11}$$

Fig. 1 The single forecasting process of the fixed SVR model and the proposed online SVR model



where ω is the inertia weight; c_1 and c_2 are the two positive constants, called cognitive learning rate and social learning rate respectively; r_1 and r_2 are random numbers in the range $[0,1]$; $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ represents the i th particle; $\mathbf{p}_{Best_i} = (p_{Best_{i1}}, p_{Best_{i2}}, \dots, p_{Best_{iD}})$ represents the best previous position of the i th particle; the g_{Best} represents the best particle among all the particles in the population; $\mathbf{V}_i = (v_{i1}, v_{i2}, \dots, v_{iD})$ represents the velocity for the i th particle, and the velocities are confined within $[\mathbf{V}_{min}, \mathbf{V}_{max}]^D$; if \mathbf{V}_i exceeds the threshold \mathbf{V}_{min} or \mathbf{V}_{max} , it is set equal to the corresponding threshold.

2.2.2 Quantum-Behaved Particle Swarm Optimization

The main disadvantage of PSO is that global convergence cannot be guaranteed [31]. To deal with this problem, QPSO was developed and reported by [42].

In traditional PSO algorithm, the dynamic behavior of the particle is widely divergent due to that the exact values of \mathbf{V} and \mathbf{X} cannot be determined simultaneously. While in QPSO algorithm, the state of a particle is determined by wave function $\psi(\mathbf{x}, t)$ instead of velocity and position. It is only necessary to learn the probability that the particles will appear at position \mathbf{X} with probability density function $\|\psi(\mathbf{x}, t)\|^2$, the form of which depends on the potential field that the particles lie in. Thus, the particles can appear at any point of space with a certain probability and the whole space can be searched without diverging to infinity. The particles move according to the following iteration equations:

$$\mathbf{X}_{t+1} = \begin{cases} \mathbf{P}_i - \beta(m_{Best} - \mathbf{X}_t) \ln(1/u) & \text{if } k \geq 0.5 \\ \mathbf{P}_i + \beta(m_{Best} - \mathbf{X}_t) \ln(1/u) & \text{if } k < 0.5 \end{cases} \quad (12)$$

where,

$$\mathbf{P}_i = \varphi \cdot \mathbf{p}_{Best_i} + (1 - \varphi) \cdot g_{Best_i} \quad (13)$$

$$m_{Best} = \frac{1}{N} \sum_{i=1}^N \mathbf{p}_{Best_i} \quad (14)$$

u , k , and φ are random numbers in the range of $[0,1]$ respectively; m_{Best} is the mean best position defined as the mean of all the best positions of the population; β , called contraction-expansion coefficient, can be tuned to control the convergence speed of the algorithm and it is only parameter in QPSO algorithm.

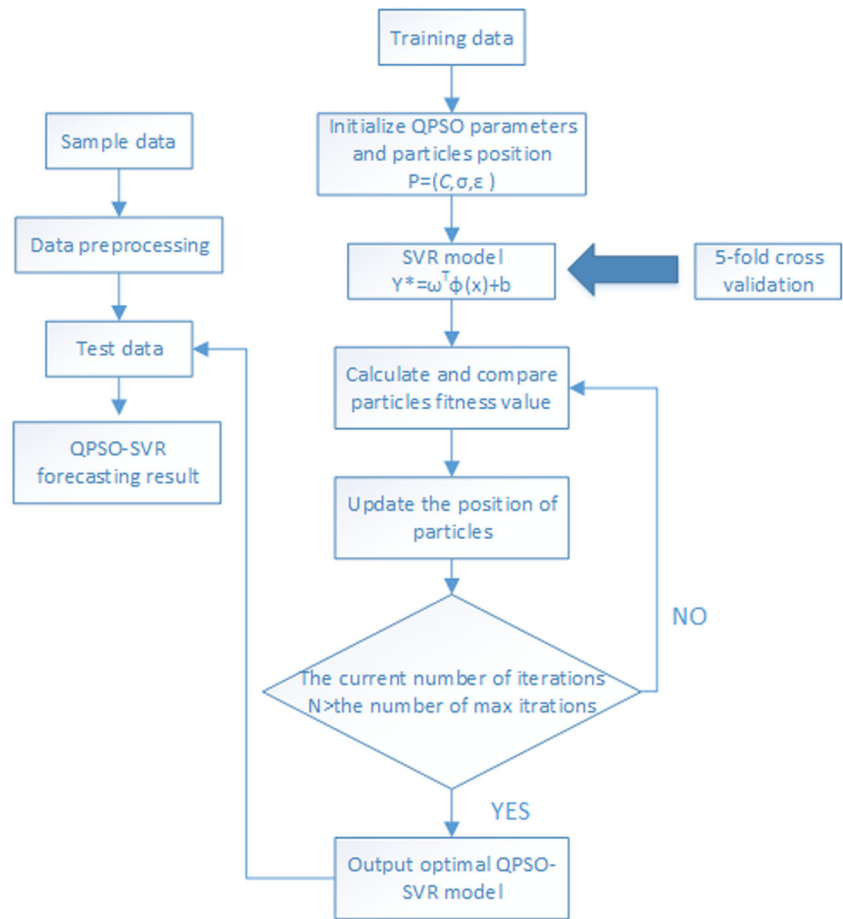
QPSO has already been applied in various optimization problems with excellent results [3, 7]. Therefore, QPSO is used to optimize the parameters of SVR, and the optimized SVR model is applied to predict air pollutant concentrations.

2.3 QPSO for Parameter Determination of the SVR Model

As it has been demonstrated above, QPSO algorithm is used to select the penalty factor C , kernel parameter σ , and insensitive coefficient ε in the SVR model, and then use the optimized SVR model to forecast the $PM_{2.5}$ and NO_2 concentrations. The flowchart of the QPSO algorithm for the three-parameter selection in the SVR model is shown in Fig. 2 and the procedures of the QPSO-SVR model are presented as follows:

Step 1: Initializing the QPSO parameter. The number of particles is 10, the maximum iteration number is 30, and the search ranges of C , σ , and ε are $[0.1,100]$, $[0.1,100]$, and $[0.01,10]$ respectively. Each particle's position is determined by the three-dimensional parameters, and the particle swarm position is randomly initialized according

Fig. 2 The flowchart of QPSO-SVR model



to the initial range of given variables. Contraction-expansion coefficient β is set to the following linear decreasing form:

$$\beta = (1.0 - 0.5)(T - t)/T + 0.5 \quad (15)$$

where T is the maximum iteration number and t is the current iteration number.

Step 2: Calculating the current fitness of all particles. The fitness value for each particle's position is determined by the fivefold cross validation error. In this study, mean square error (MSE) is utilized as cross-validation error, which is defined as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - Y_i^*)^2 \quad (16)$$

where Y_i is the measured value, Y_i^* is the predicted value, and n is the number of the data points.

Step 3: Choosing the individual history optimal position and the global optimal position. The current position of each particle is initialized to the

individual historical optimal position, and the position with the smallest fitness value among all the particles is chosen as the global optimal position.

Step 4: Updating the position of the particles. First, calculating the average position of the particles according to Eq. (14), then calculating random position for each particle according to Eq. (13); finally, the position of the particles is updated according to Eq. (12).

Step 5: The fitness value of the updated particle is recalculated and compared with the fitness value of the previous iteration. If it is better, the position of the particle is updated to the current position of the particle.

Step 6: The current global optimal position and fitness value of the population are calculated and compared with the fitness value of the global optimal position of the previous iteration. If it is better, the global optimal position of the population is updated to the current global optimal position.

Step 7: Checking the termination criterion. Optimal parameters are determined if the termination criterion is satisfied. Otherwise, return to step 2.

2.4 Multi-step Ahead Forecast Method

Multi-step ahead forecast can be described as an estimation of future values in the case of the given previous observations. There are several strategies for multi-period forecast, such as recursive strategy, direct strategy, MIMO strategy, and so on [18]. Therefore, the multi-step ahead forecast methods based on the recursive strategy and the direct strategy are compared to select optimal prediction method. The main idea of recursive strategy is that M

$$\begin{cases} \hat{y}(t+1) = f(y(t), y(t-1), \dots, y(t-n+1)) \\ \hat{y}(t+2) = f(\hat{y}(t+1), y(t), \dots, y(t-n+2)) \\ \vdots \\ \hat{y}(t+H) = f(\hat{y}(t+H-1), \hat{y}(t+H-2), \dots, \hat{y}(t), y(t-1), \dots, y(t-n+H)) \end{cases} \quad (17)$$

$$\{ \hat{y}(t+H) = f(y(t), y(t-1), y(t-2), \dots, y(t-n+1)) \quad (18)$$

where n is the maximum embedding order, y is the observed value, \hat{y} is the predicted value, and f represents the established model. And $H = 1, 2, \dots, M$, M is the maximum horizon of prediction.

By the above equations, the H step prediction results can be obtained. When the value of H is large, with the increase of the prediction step, it may appear that all the inputs are the predicted values, which may reduce the forecasting accuracy. In this study, in order to avoid error accumulation and computational complexity, the values of H and n are set to 4 and 1, respectively. Therefore, the following recursive equation and direct equation are obtained respectively:

$$\begin{cases} \hat{y}(t+1) = f(y(t), \mathbf{P}(t)) \\ \hat{y}(t+2) = f(\hat{y}(t+1), \mathbf{P}(t+1)) \\ \hat{y}(t+3) = f(\hat{y}(t+2), \mathbf{P}(t+2)) \\ \hat{y}(t+4) = f(\hat{y}(t+3), \mathbf{P}(t+3)) \end{cases} \quad (19)$$

$$\{ \hat{y}(t+4) = f(y(t), \mathbf{P}(t)) \quad (20)$$

where $y(t)$ is the concentration of air pollutant to be predicted at time t . $\mathbf{P}(t)$ represents the value of the auxiliary variables at time t . And in the experiment, it is assumed that the values of all the auxiliary variables, which will be described in Section 3.1, can be known 4 hours in advance.

3 Simulation Results and Discussions

3.1 Original Dataset

Beijing, as the capital of China, has built 35 air quality monitoring sites so far, among which Wanliu monitoring site located in Haidian District of Beijing is an environmental assessment point, and it is close to the

samples are trained to obtain regression model firstly. Secondly, a single-step forecast can be determined using the established regression models. Finally, the following forecasting steps are calculated iteratively using the single-step predicted values as a historical time series for the subsequent point. And the estimation of the H next values is defined as Eq. (17), while the direct strategy presents an easily understandable result when forecasting H steps ahead. And the estimation of the H next values can be obtained by Eq. (18).

city center. So the evaluation of its air quality has certain representation for the overall air quality of Beijing. This is why Wanliu monitoring site is selected as the object of this experiment. According to the dataset which was collected in the Urban Air project [57–59], the available air quality dataset measured at the Wanliu Monitoring Station in May 2014 to April 2015 is selected as the original dataset. The selected dataset includes five major air pollutants, i.e., PM_{2.5}, NO₂, CO, O₃, and SO₂, and six meteorological parameters, i.e., weather (W), temperature (T), pressure (P), relative humidity (RH), wind speed (WS), and wind direction (WD), which were hourly measured at the Wanliu Monitoring Station. The weather are described by 17 different values and the wind direction is represented by 10 different situations, as shown in Tables 1 and 2, respectively. All input variables in the models are shown in Table 3. And three kinds of prediction methods are used to forecast PM_{2.5} and NO₂ concentration for 4 hours ahead: (1) multi-step prediction based on recursive strategy (recursive forecast): the pollutant concentration are predicted according to Eq. (19). (2) Multi-step prediction based on direct strategy (direct forecast): the value of all input variables at time t is used to directly forecast the pollutant concentration at time $(t+4)$. (3) Online multi-step prediction based on direct strategy (online direct forecast): the regression model is updated dynamically in the process of direct multi-step prediction. While for the first two methods, they use a fixed model in the forecasting process.

Because meteorological conditions have great impact on atmospheric pollutant concentrations, and Beijing is a city with four distinct seasons, in order to assess the effect of seasonal variations on model performance, the recorded levels of PM_{2.5} and NO₂ in July 2014, November 2014, and January 2015 are selected as original samples. The number

Table 1 Different weather conditions are represented by 17 different values

0	1	2	3	4	5
Sunny	Cloudy	Overcast	Rainy	Sprinkle	Moderate rain
6	7	8	9	10	11
Heaver rain	Rain storm	Thunder storm	Freezing rain	Snowy	Light snow
12	13	14	15	16	
Moderate snow	Heavy snow	Foggy	Sand storm	Dusty	

of valid data in those months were 688 (July 3, 2014, 00:00–July 31, 2014, 15:00), 720 (November 1, 2014, 00:00–November 30, 2014, 23:00), and 720 (January 1, 2015, 00:00–January 30, 2015, 23:00). In the experiments, the first 70% of the data is selected as training set, and the remaining data is used as testing set. At the same time, the fivefold cross-validation method is adopted to obtain the optimal prediction model in the experiments. And its main idea is that the previous 70% training data are divided into five equally sized and mutually complementary subsets firstly, and then the data from the four subsets are trained to obtain a model and the remaining subsets are tested to evaluate the obtained model; this process is repeated for the five possible choices. Finally, the model with the smallest error in five experiments is selected as the optimal model, and then the previously divided 30% test data is used to evaluate this optimal model. In the experiments, all the algorithms were coded in matlab and C++ language and their code was run on an Intel(R) Core(TM) i5-4210U, 1.70GHZ PC with 4GB of RAM.

In order to eliminate the influence of different dimension and unit, the input and output data of samples are normalized respectively in the data process. The formula is as follows:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{21}$$

The formula normalizes the original data into the range of [0, 1], where X_{norm} is the normalized data, X is the original data, and X_{max} and X_{min} are the maximum and minimum values in the original data set respectively.

3.2 Evaluation of the Model Performance

The test results of the QPSO-SVR model are analyzed quantitatively based on mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination

(R^2) in this study. The three evaluation functions are defined as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - Y_i^*| \tag{22}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - Y_i^*)^2} \tag{23}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - Y_i^*)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2} \tag{24}$$

where Y_i is the measured concentration level, Y_i^* is the forecast value, \bar{Y}_i is the average of the measured value, and n is the number of the data points.

In MAE, the deviation is absolute; it can reflect the actual situation better for prediction error. RMSE is most useful for large error due to the existence of relatively high weight for large error. The better performance is always given by smaller MAE and RMSE and the better fitting result is always described by the value of R^2 which is close to 1.

3.3 PM_{2.5} Concentration Forecasting Results

In the experiments, the QPSO-SVR model proposed in this paper is used to select the optimal prediction method among three prediction methods, and then this optimal prediction method is applied to compare the performance of different optimization algorithms for SVR parameter selection.

Figure 3 shows the original time series of hourly PM_{2.5} concentrations in July 2014, November 2014, and January 2015. It can be observed that the highest concentrations of PM_{2.5} are 262 $\mu\text{g}/\text{m}^3$, 435 $\mu\text{g}/\text{m}^3$, and 482 $\mu\text{g}/\text{m}^3$ in the 3 months respectively, and PM_{2.5} concentration increased significantly in late November; this is mainly due to the combustion of coal which produces a large amount of pollutants. In addition, meteorological conditions affect

Table 2 Different wind directions are represented by 10 different values

0	1	2	3	4	9	13	14	23	24
No	East	West	South	North	Unstable	Southeast	Northeast	Southwest	Northwest

Table 3 All input variables for the models

Input variables	Name	Symbol	Unit
Basic value	PM _{2.5}	$\rho(PM_{2.5}^0)$	$\mu\text{g}/\text{m}^3$
	CO	$\rho(CO^0)$	mg/m^3
	O ₃	$\rho(O_3^0)$	$\mu\text{g}/\text{m}^3$
	SO ₂	$\rho(SO_2^0)$	$\mu\text{g}/\text{m}^3$
	NO ₂	$\rho(NO_2^0)$	$\mu\text{g}/\text{m}^3$
Meteorological parameters	Weather	W	–
	Temperature	T	°C
	Pressure	P	<i>hPa</i>
	Relative humidity	H	%
	Wind speed	WS	<i>m/s</i>
	Wind direction	WD	–

the diffusion of atmospheric pollutants. Therefore, it is necessary to analyze and predict the PM_{2.5} concentration.

In order to select the best prediction method, three prediction methods were tested based on the QPSO-SVR model. Figure 4 shows results of the three prediction methods based on QPSO-SVR model for the prediction of

PM_{2.5} concentration in July 2014, November 2014, and January 2015. It is observed that, for the recursive prediction method, the prediction result is the worst one in the selected months. The reason is the cumulative effect of errors in the recursive strategy, while the other two methods have little difference in the prediction result. However, it still can be seen that the online direct forecast method is slightly better than the result of the direct prediction method. This can be seen from Table 4; both MAE and RMSE produced by the online direct prediction method are smaller than those created by the recursive prediction and direct prediction methods in the selected months. Hence, it can be concluded that the online direct prediction method is superior to the other two methods. Therefore, the online direct prediction method is selected to test the prediction performance of several models.

Figures 5, 6 and 7 present the prediction results of PM_{2.5} concentration based on the QPSO-SVR model and the PSO-SVR model in the selected months, respectively, which include the fitting curve of the two models in the test phase and the corresponding absolute error. It can be seen from Fig. 5 that there are many deviation points in the prediction

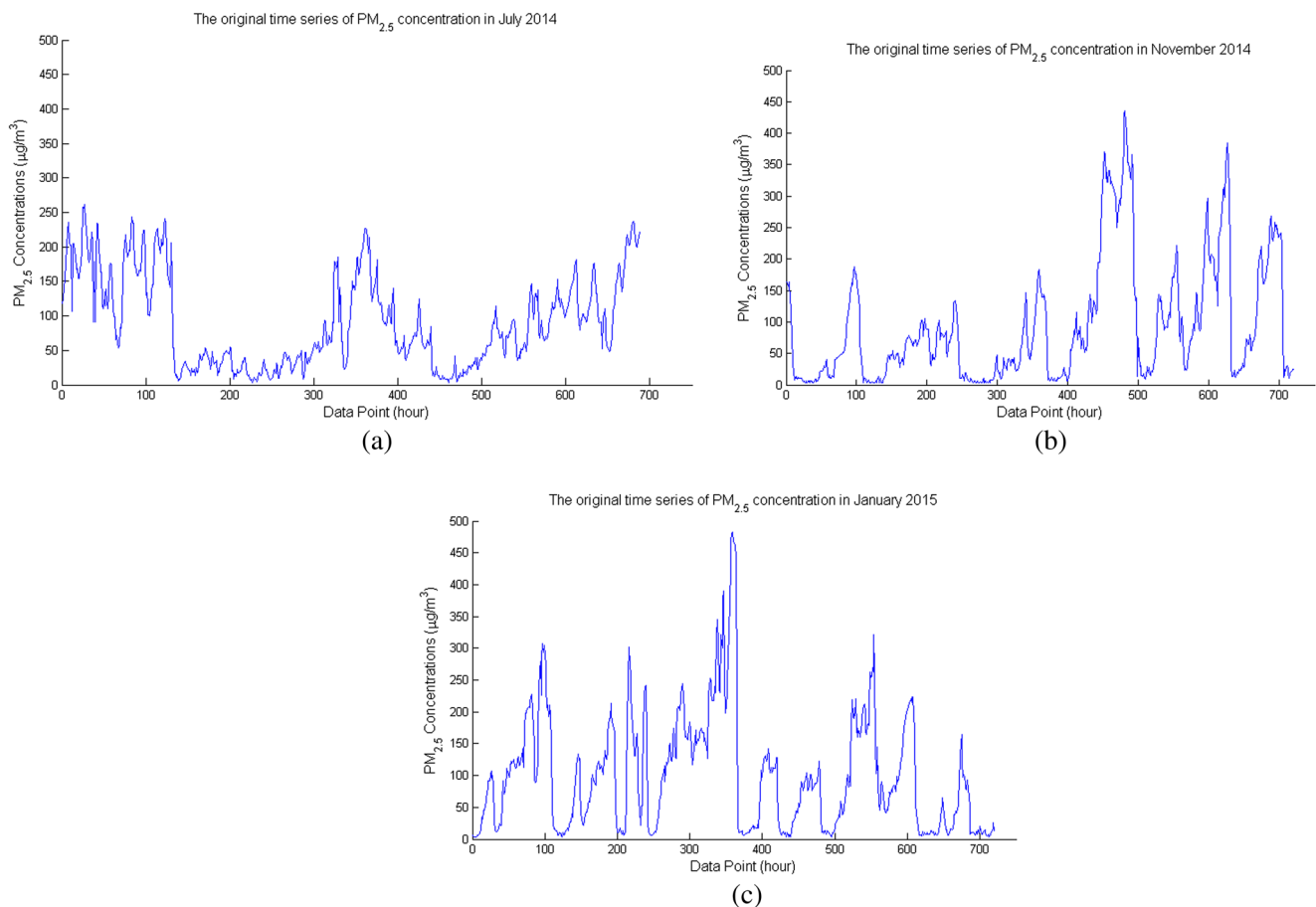
**Fig. 3** The original time series of hourly PM_{2.5} concentrations **a** July 2014, **b** November 2014, **c** January 2015

Fig. 4 The results of the three prediction methods based on QPSO-SVR model for the prediction of $PM_{2.5}$ concentrations. **a** July 2014. **b** November 2014. **c** January 2015

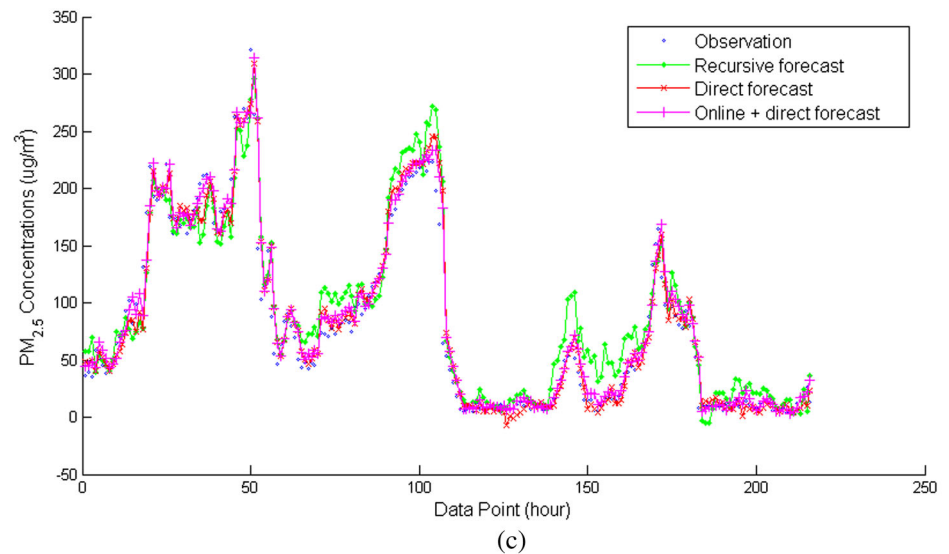
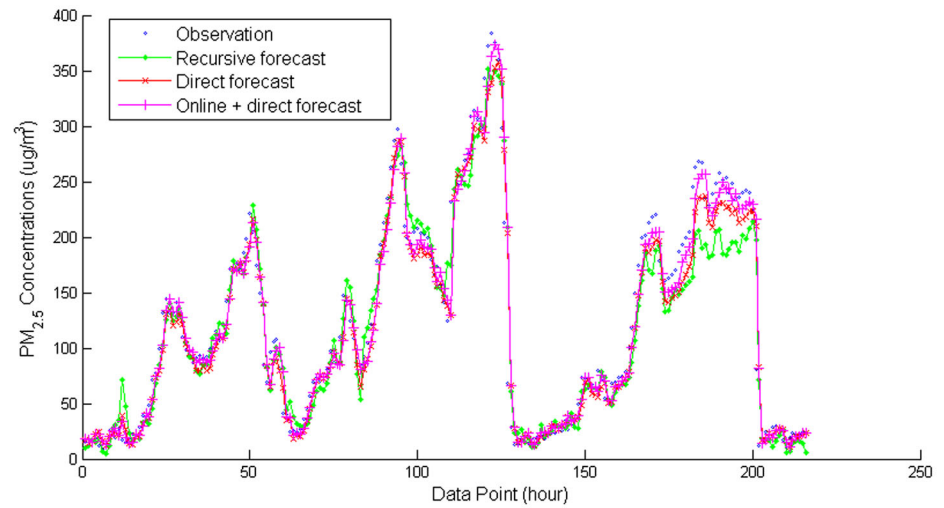
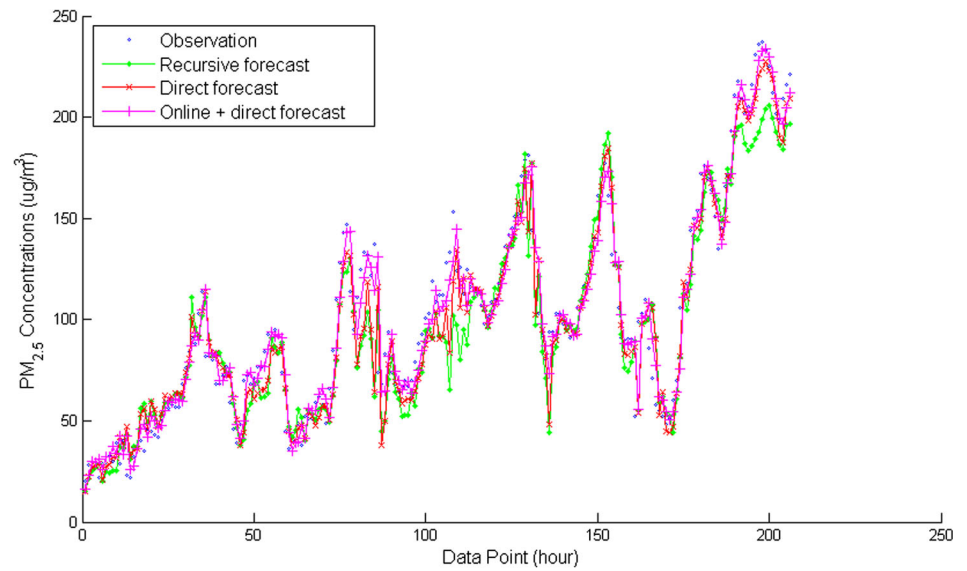


Table 4 The comparison of three prediction methods for PM_{2.5} concentration prediction based on the QPSO-SVR model

Prediction methods	July 2014		November 2014		January 2015	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
Recursive forecast	14.89	19.69	20.33	28.73	20.25	27.16
Direct forecast	12.20	16.43	15.82	24.00	12.68	20.41
Online direct forecast	10.42	14.05	13.57	23.29	10.90	18.66

of the PSO-SVR model, while the QPSO-SVR model only appears with few deviation points. Either for the individual case or for the average case, the QPSO-SVR model shows better prediction performance than the PSO-SVR model. Figures 6 and 7 can also prove the same conclusion. In addition, the robustness of both QPSO-SVR model and PSO-SVR model is also inspected under the impact of meteorological factors such as weather, temperature, pressure, humidity, wind speed, and wind direction in the three different seasons. Hence, it can be concluded that the QPSO-SVR model possesses advantages to the PSO-SVR

model although the impact of meteorological factors exists.

Table 5 lists the comparison of prediction performance among the QPSO-SVR model, PSO-SVR model, GA-SVR model, and GS-SVR model for PM_{2.5} concentration on the test stage. It can be seen from the table that the QPSO-SVR model has the lowest prediction error and its calculation time is less than that of the GA-SVR model and GS-SVR model. Although the QPSO-SVR model and the PSO-SVR model have little difference in the running time, it can still be seen that the QPSO-SVR runs faster than the PSO-SVR model. Therefore, it can be concluded that QPSO is superior

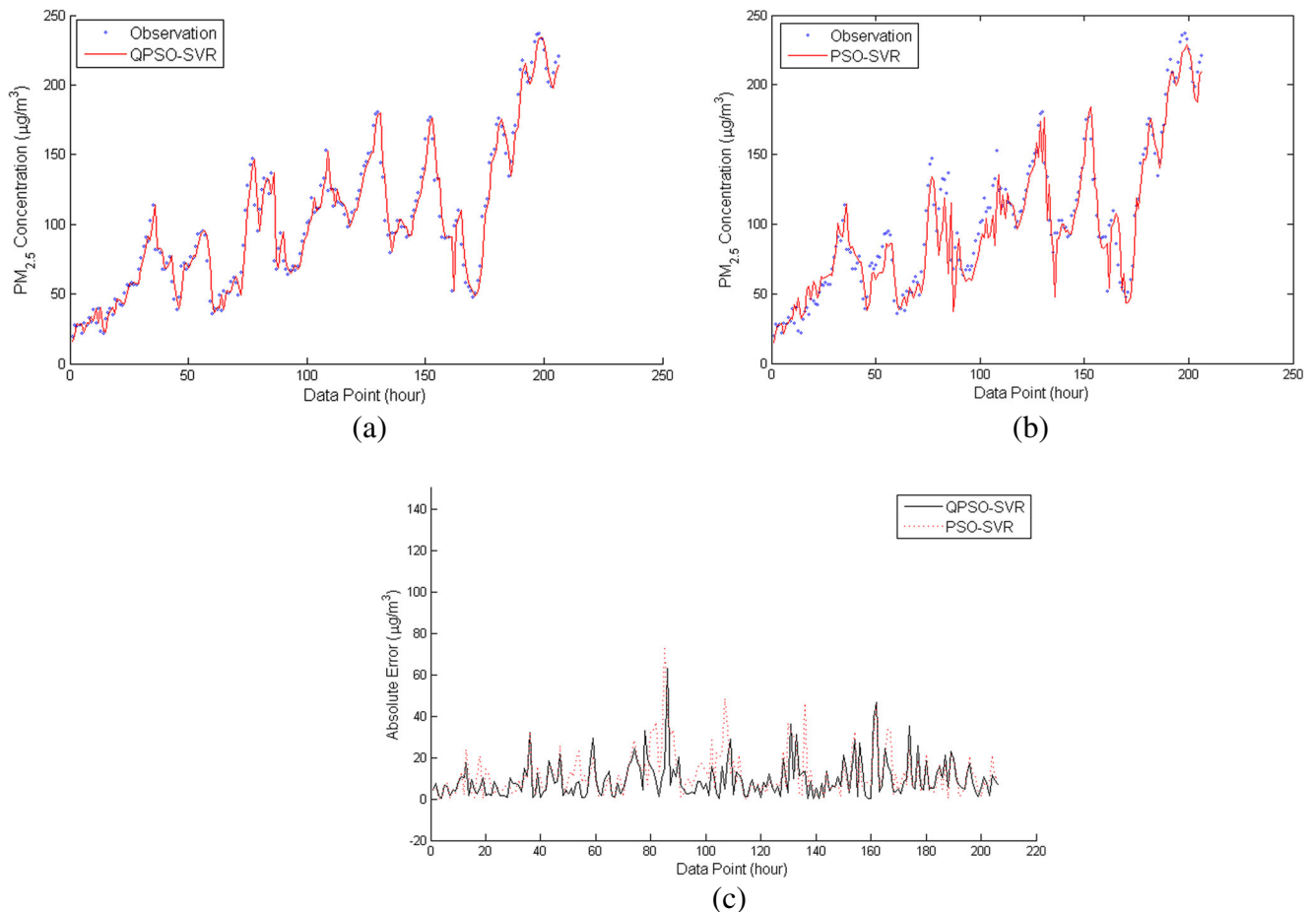


Fig. 5 The online multi-step prediction results for PM_{2.5} concentration and the absolute error of two models in July 2014. **a** QPSO-SVR model. **b** PSO-SVR model. **c** Absolute error of two models

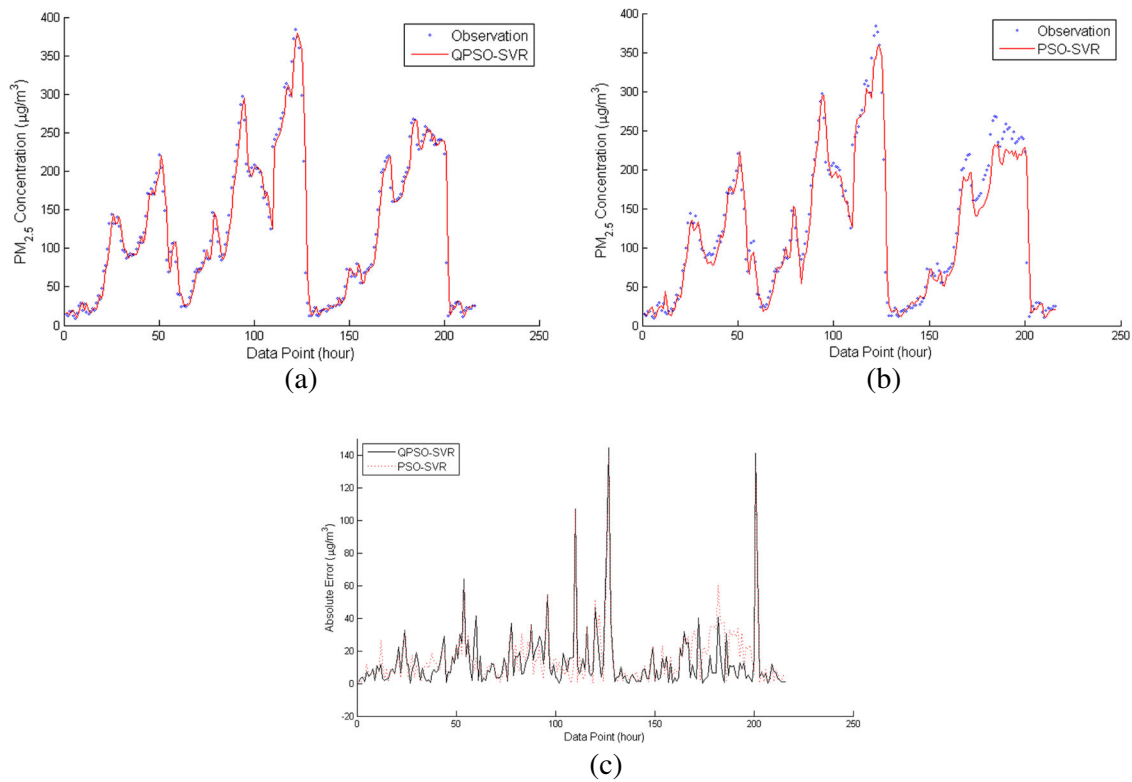


Fig. 6 The online multi-step prediction results for PM_{2.5} concentration and the absolute error of two models in November 2014. **a** QPSO-SVR model. **b** PSO-SVR model. **c** Absolute error of two models

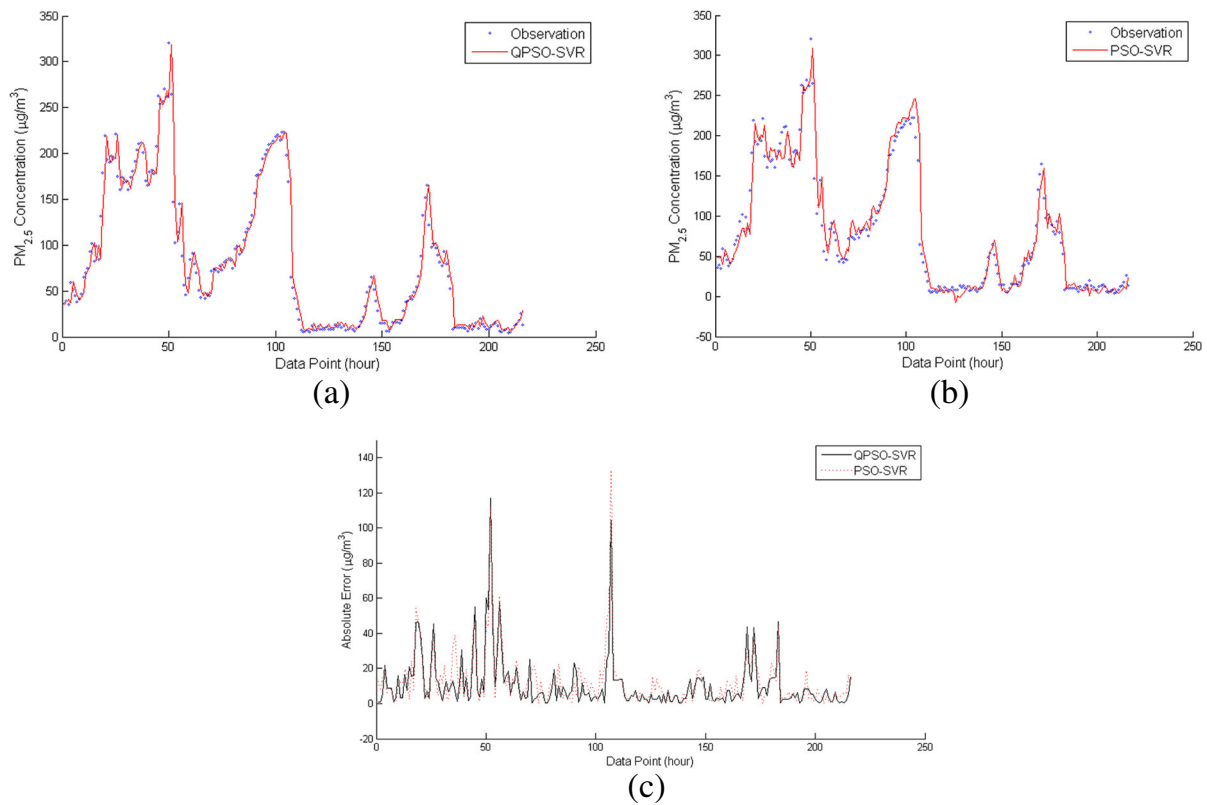


Fig. 7 The online multi-step prediction results for PM_{2.5} concentration and the absolute error of two models in January 2015. **a** QPSO-SVR model. **b** PSO-SVR model. **c** Absolute error of two models

Table 5 The comparison of model performance for PM_{2.5} concentration prediction based on the online direct prediction method

Models	July 2014				November 2014				January 2015			
	MAE	RMSE	R^2	Time(s)	MAE	RMSE	R^2	Time(s)	MAE	RMSE	R^2	Time(s)
QPSO-SVR	10.42	14.05	0.93	26	13.57	23.29	0.90	28	10.90	18.66	0.92	31
PSO-SVR	12.12	16.30	0.91	28	16.01	24.41	0.88	31	12.68	20.41	0.90	33
GA-SVR	13.56	17.80	0.88	34	18.12	26.34	0.85	37	13.23	20.04	0.88	37
GS-SVR	19.68	23.52	0.78	141	22.05	30.11	0.76	145	23.83	30.43	0.76	156

to other optimization algorithms in parameter selection of SVR model; it is proved that the proposed hybrid QPSO-SVR model is effective in the prediction of atmospheric PM_{2.5} concentration.

3.4 NO₂ Concentration Forecasting Results

Considering the characteristic of each pollutant, such as the accumulation of PM_{2.5}, and chemical and physical

complexity of NO₂, prediction performance and the robustness of the QPSO-SVR model can be further verified by forecasting NO₂.

Figure 8 shows the original time series of hourly NO₂ concentrations in July 2014, November 2014, and January 2015 respectively. It can be observed that NO₂ also shows same change regulation with PM_{2.5} concentration, and the frequent fluctuation of NO₂ concentration may have an impact on the prediction model.

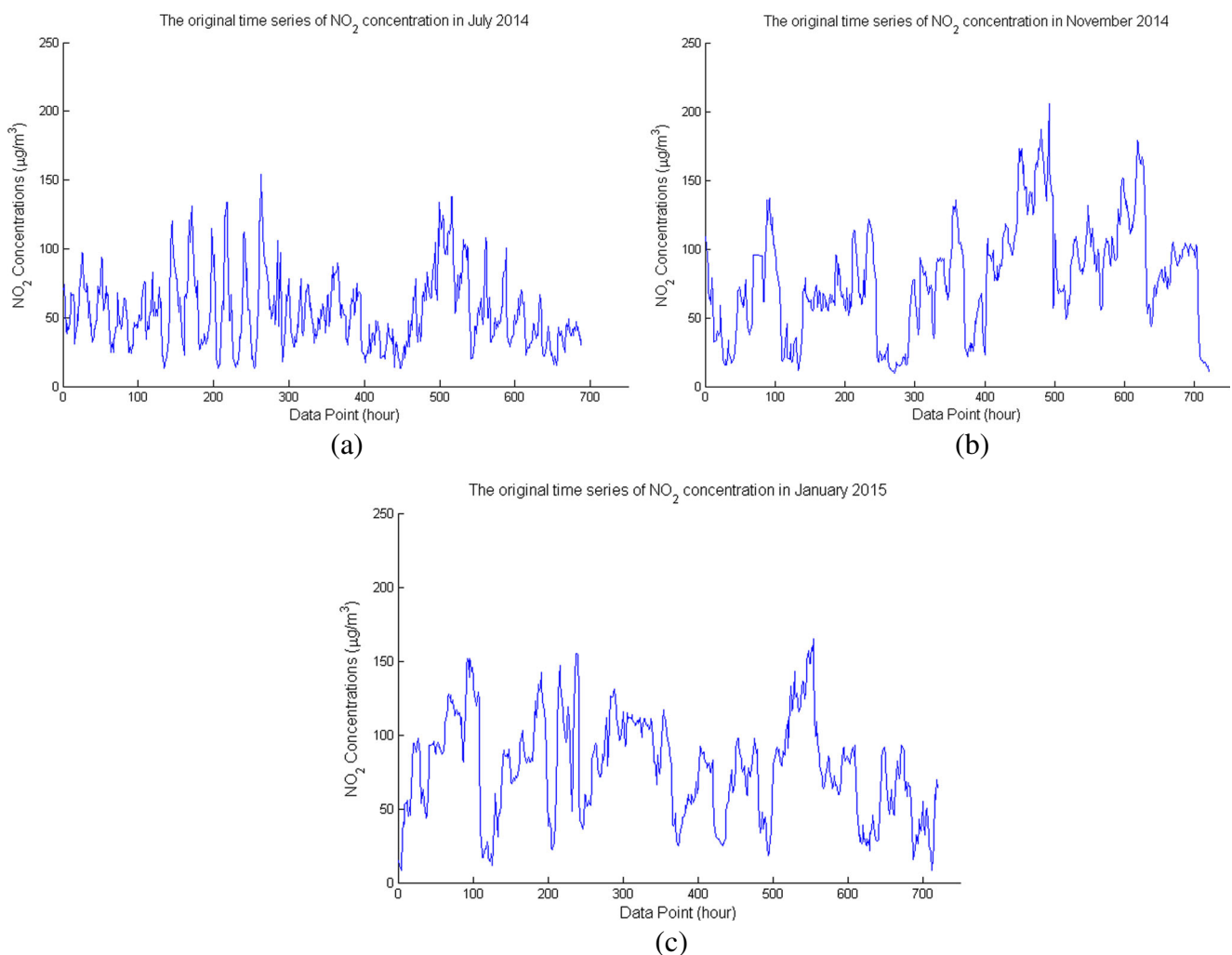


Fig. 8 The original time series of hourly NO₂ concentrations. **a** July 2014. **b** November 2014. **c** January 2015

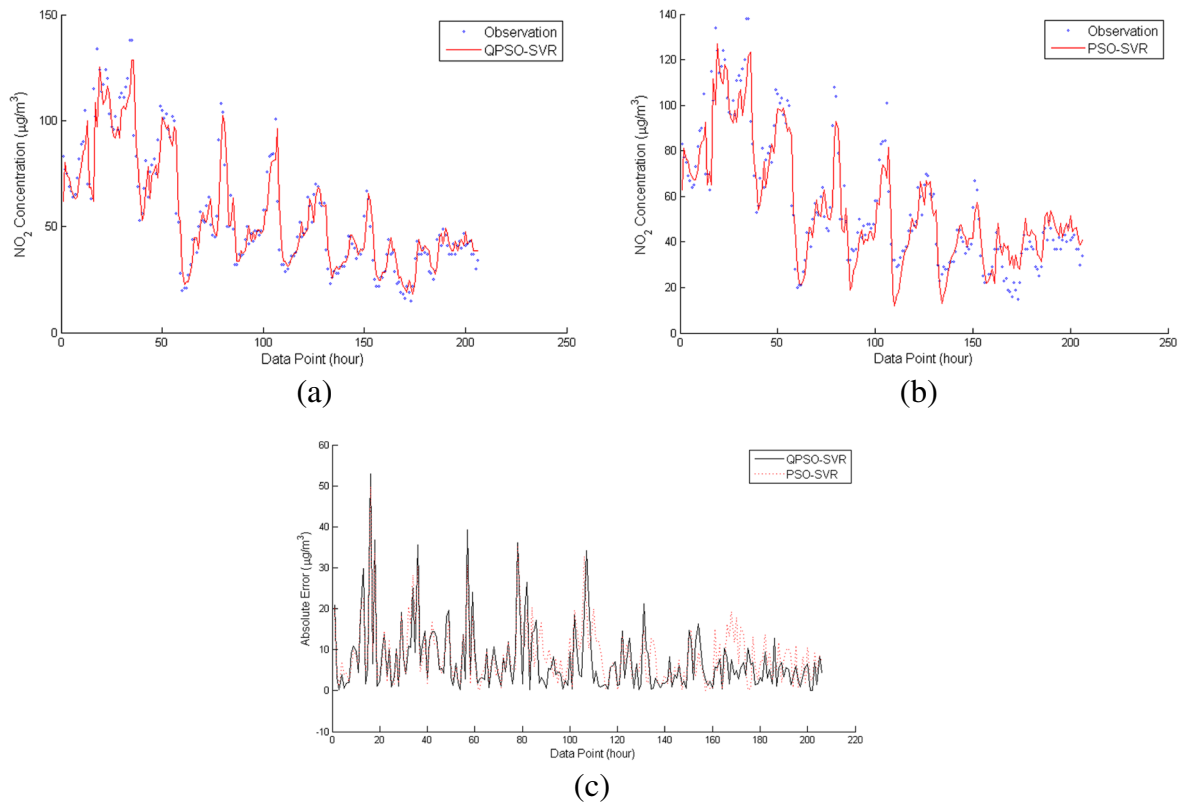


Fig. 9 The online multi-step prediction results for NO₂ concentration and the absolute error of two models in July 2014. **a** QPSO-SVR model. **b** PSO-SVR model. **c** Absolute error of two models

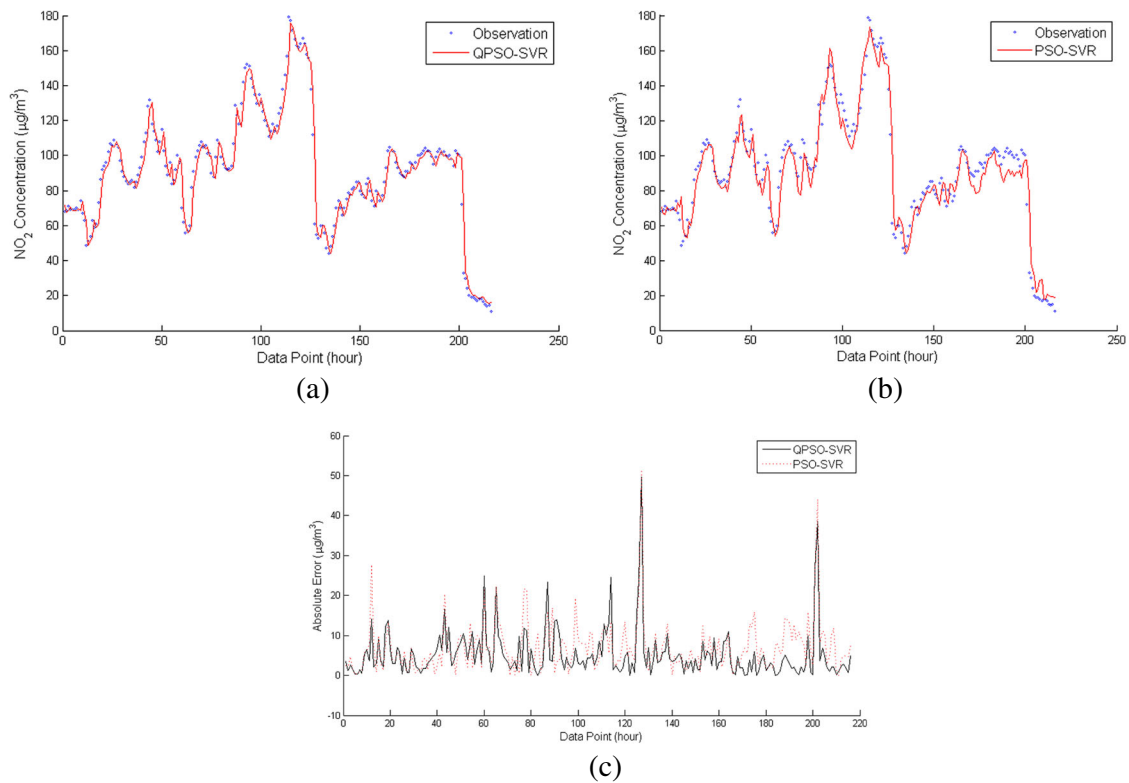


Fig. 10 The online multi-step prediction results for NO₂ concentration and the absolute error of two models in November 2014. **a** QPSO-SVR model. **b** PSO-SVR model. **c** Absolute error of two models

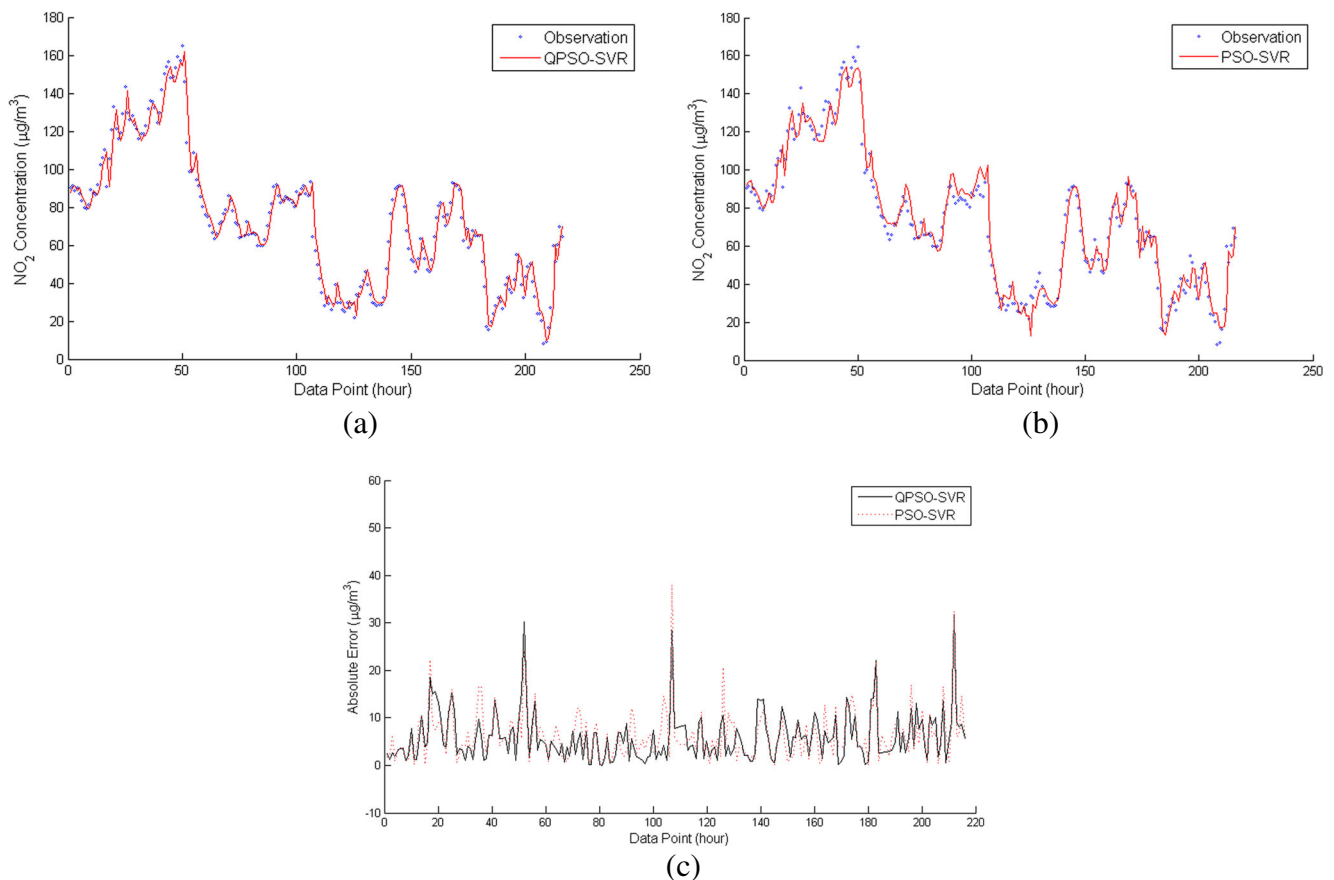


Fig. 11 The online multi-step prediction results for NO₂ concentration and the absolute error of two models in January 2015. **a** QPSO-SVR model. **b** PSO-SVR model. **c** Absolute error of two models

Figures 9, 10 and 11 present the prediction results of NO₂ concentration based on the QPSO-SVR model and the PSO-SVR model in the selected months respectively. By comparison, it can be seen that the prediction results generated by the QPSO-SVR model are much better than those produced by the PSO-SVR model in the 3 months. Especially in July, more prediction points by the PSO-SVR model are deviated from the measured points, but only several prediction points by the QPSO-SVR model are away from the measured ones. And the MAE produced by the QPSO-SVR model is smaller than that obtained by the

PSO-SVR model. Therefore, the same conclusion that the QPSO-SVR model possesses better prediction performance than the PSO-SVR model can be obtained.

Table 6 shows predicting error and computational time comparison among QPSO-SVR model, PSO-SVR model, GA-SVR model, and GS-SVR model for NO₂ concentration on the test stage. It can be seen that, both MAE and RMSE produced by the QPSO-SVR model are smaller than those created by the other models in the three selected months, while the values of R^2 generated by the QPSO-SVR model are greater than those produced by the other models for all

Table 6 The comparison of models performance for NO₂ concentration prediction based on the online direct prediction method

Models	July 2014				November 2014				January 2015			
	MAE	RMSE	R^2	Time(s)	MAE	RMSE	R^2	Time(s)	MAE	RMSE	R^2	Time(s)
QPSO-SVR	7.42	10.98	0.85	25	5.26	8.07	0.94	27	5.92	7.78	0.95	28
PSO-SVR	8.68	11.35	0.84	28	6.94	9.46	0.92	27	6.38	8.32	0.94	29
GA-SVR	9.92	12.79	0.80	37	6.97	9.52	0.92	46	8.29	10.16	0.83	43
GS-SVR	12.25	14.27	0.77	142	11.41	14.31	0.78	148	12.26	14.92	0.77	167

Table 7 The comparison of different methods in this study

Experimental materials	Variables (fixed variables)	Variables value	Results
PM _{2.5} concentration in 3 selected months	Prediction method (fixed, QPSO-SVR)	Recursive	It has the cumulative effect of error, which resulting in reduced prediction accuracy.
		<i>Direct</i>	The prediction error is smaller than the error of recursive prediction, but the prediction accuracy of these two methods is affected by the values of H and n .
	Model (direct, QPSO-SVR)	Fixed	It is difficult to adapt to the nonlinear and time-varying characteristics of pollutant concentration.
		<i>Online</i>	The prediction model is continuously updated with arriving of the new sample; the prediction accuracy is higher than that of the fixed model, but the running time becomes longer.
PM _{2.5} and NO ₂ concentrations in 3 selected months	Optimization algorithms (online, direct)	GS	It is very time-consuming compared with the heuristic optimization algorithm, and the optimization result is poor.
		GA	It has more control parameters and more complicated calculation.
		PSO	It has simpler principle and fewer parameters, and it is easier to implement.
		<i>QPSO</i>	It possesses less control parameters, shorter calculation time, higher optimization accuracy, and stronger robustness. Compared to PSO and GA, it is more suitable for the optimization of SVR parameters in the PM _{2.5} and NO ₂ concentration predictions.

the selected months. In addition, the heuristic optimization algorithm is more efficient than the grid search method in selecting SVR parameters. Moreover, the PSO-SVR model possesses much less computational time than the GA-SVR model, and the QPSO-SVR model also reduces the calculation time compared with the PSO-SVR model.

Based on the above experiments, it can be concluded that the QPSO-SVR model is more excellent compared with the PSO-SVR model, GA-SVR model, and GS-SVR model. It can always possess good, robust prediction performance for air pollutants.

3.5 Experiments summary

In order to summarize, visualize, and compare the studies and to emphasize the contribution of the algorithm proposed in this study, the comparison of different methods in this study is presented in the Table 7. In order to improve the PM_{2.5} and NO₂ concentration prediction accuracy, three aspects are considered, including prediction methods, models, and optimization algorithms. According to the above experimental results, it is concluded that the QPSO-SVR model based on the online direct prediction method is more suitable for the prediction of atmospheric PM_{2.5} and NO₂ concentration than other methods.

4 Conclusions

This paper mainly develops a hybrid QPSO-SVR model to predict atmospheric PM_{2.5} and NO₂ concentrations in the short term, and the QPSO algorithm is mainly used to select the optimal parameters (C , σ , and ε) influencing the performance of SVR. Firstly, the three prediction methods are proposed, including multi-step prediction method based on recursive strategy, multi-step prediction method based on direct strategy, and online multi-step prediction method based on direct strategy. PM_{2.5} concentration was predicted by these three methods based on the QPSO-SVR model; the results show that the online multi-step prediction method based on direct strategy has best prediction results. Secondly, the prediction performances of the QPSO-SVR model, PSO-SVR model, GA-SVR model, and GS-SVR model were compared by using the online direct prediction methods. And the atmospheric PM_{2.5} and NO₂ concentrations in the three different seasons were predicted. The results demonstrate that the QPSO-SVR model possesses better prediction performance in terms of prediction accuracy and computational time. Moreover, the QPSO-SVR model is more robust because it is less affected by the meteorological factors. Finally, the model proposed in this paper can be used for the prediction of

other pollutant concentration, and our team have installed device in our campus to collect pollutant concentration and meteorological data in order to analyze and evaluate the environment of our campus. Additionally, the value of H and n in multi-step prediction will have an effect on the prediction results, and the problem of large computation and poor real-time performance will appear in the online SVR model when the sample is too large. How to solve these problems will be our future research work.

Funding Information This work is supported by the Fund of National Natural Science Foundation of China (61873006, 61473034 and 61673053), Beijing Nova Programme Interdisciplinary Cooperation Project (Z161100004916041) and Project of Beijing science and technology commission: Research on key technologies of intelligent factory interconnection and industrial Internet of things equipment.

References

- Andrew, A.M. (2001). An introduction to support vector machines and other kernel-based learning methods. *Robotica*, 18(6), 687–689.
- Arabgol, R., Sartaj, M., Asghari, K. (2016). Predicting nitrate concentration and its spatial distribution in groundwater resources using support vector machines (svms) model. *Environmental Modeling & Assessment*, 21(1), 71–82.
- Bagheri, A., Peyhani, H.M., Akbari, M. (2014). Financial forecasting using ANFIS networks with quantum-behaved particle swarm optimization. *Expert Systems with Applications*, 41(14), 6235–6250.
- Bai, Y., Li, Y., Wang, X., Xie, J., Li, C. (2016). Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmospheric Pollution Research*, 7(3), 557–566.
- Bamakan, S.M.H., Wang, H., Ravasan, A.Z. (2016). Parameters optimization for nonparallel support vector machine by particle swarm optimization. *Procedia Computer Science*, 91, 482–491.
- Beckerman, B.S., Jerrett, M., Serre, M., Martin, R.V., Lee, S.J., Van, D.A., Ross, Z., Su, J., Burnett, R.T. (2013). A hybrid approach to estimating national scale spatiotemporal variability of PM_{2.5} in the contiguous united states. *Environmental Science & Technology*, 47(13), 7233–41.
- Ch, S., Anand, N., Panigrahi, B.K., Mathur, S. (2013). Streamflow forecasting by SVM with quantum behaved particle swarm optimization. *Neurocomputing*, 101(3), 18–23.
- Chen, R., Samoli, E., Wong, C.M., Huang, W., Wang, Z., Chen, B., Kan, H., Group, C.C. (2012). Associations between short-term exposure to nitrogen dioxide and mortality in 17 chinese cities: the China air pollution and health effects study (capes). *Environmental Research*, 45(14), 32–38.
- Chiusolo, M., Cadum, E., Galassi, C., Stafoggia, M., Berti, G. (2009). Short term effects of nitrogen dioxide exposure on mortality and susceptibility factors. *Epidemiology*, 20(6), S67.
- De, G.G., Trizio, L., Di, G.A., Pey, J., Pérez, N., Cusack, M., Alastuey, A., Querol, X. (2013). Neural network model for the prediction of PM₁₀ daily concentrations in two sites in the western mediterranean. *Science of the Total Environment*, 463-464(5), 875.
- Dijkema, M.B., van Strien, R.T., Sc, V.D.Z., Mallant, S.F., Fischer, P., Hoek, G., Brunekreef, B., Gehring, U. (2016). Spatial variation in nitrogen dioxide concentrations and cardiopulmonary hospital admissions. *Environmental Research*, 151, 721–727.
- Dong, Z., Yang, D., Reindl, T., Walsh, W.M. (2015). A novel hybrid approach based on self-organizing maps, support vector regression and particle swarm optimization to forecast solar irradiance. *Energy*, 82, 570–577.
- Donnelly, A., Misstear, B., Broderick, B. (2015). Real time air quality forecasting using integrated parametric and non-parametric regression techniques. *Atmospheric Environment*, 103(103), 53–65.
- Fang, S.F., Wang, M.P., Qi, W.H., Zheng, F. (2008). Hybrid genetic algorithms and support vector regression in forecasting atmospheric corrosion of metallic materials. *Computational Materials Science*, 44(2), 647–655.
- Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., Wang, J. (2015). Artificial neural networks forecasting of PM_{2.5} pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment*, 107, 118–128.
- Gorai, A.K., & Mitra, G. (2016). A comparative study of the feed forward back propagation (FFBP) and layer recurrent (LR) neural network model for forecasting ground level ozone concentration. *Air Quality Atmosphere & Health*, pp. 1–11.
- Ishak, A.B., Moslah, Z., Trabelsi, A. (2016). Analysis and prediction of PM₁₀ concentration levels in Tunisia using statistical learning approaches. *Environmental and Ecological Statistics*, 23(3), 1–22.
- Ji, Y., Hao, J., Reyhani, N., Lendasse, A. (2005). *Direct and recursive prediction of time series using mutual information selection*. Berlin: Springer.
- Juhos, I., Makra, L., Tóth, B. (2008). Forecasting of traffic origin NO and NO₂ concentrations by support vector machines and neural networks using principal component analysis. *Simulation Modelling Practice and Theory*, 16(9), 1488–1502.
- Juodis, L., Filistovič, V., Maceika, E., Remeikis, V. (2016). Analytical dispersion model for the chain of primary and secondary air pollutants released from point source. *Atmospheric Environment*, 128, 216–226.
- Kavousi-Fard, A., Samet, H., Marzbani, F. (2014). A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. *Expert Systems with Applications*, 41(13), 6047–6056.
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *IEEE International conference on neural networks, 1995. proceedings*, (Vol. 4 pp. 1942–1948).
- Kennedy, J., & Eberhart, R. (2011). *Particle swarm optimization* Vol. 4. USA: Springer.
- Krewski, D., & Rainham, D. (2007). Ambient air pollution and population health: overview. *Journal of Toxicology and Environmental Health, Part A*, 70(3-4), 275–283.
- Kumar, U., & Jain, V.K. (2010). Arima forecasting of ambient air pollutants (O₃, NO, NO₂ and CO). *Stochastic Environmental Research and Risk Assessment*, 24(5), 751–760.
- Lin, K.P., Pai, P.F., Yang, S.L. (2011). Forecasting concentrations of air pollutants by logarithm support vector regression with immune algorithms. *Applied Mathematics and Computation*, 217(12), 5318–5327.
- Malik, M.A., Jiang, C., Heller, R., Lane, J., Hughes, D., Schoenbach, K.H. (2016). Ozone-free nitric oxide production using an atmospheric pressure surface discharge – a way to minimize nitrogen dioxide co-production. *Chemical Engineering Journal*, 283, 631–638.
- Moazami, S., Noori, R., Amiri, B.J., Yeganeh, B., Partani, S., Safavi, S. (2016). Reliable prediction of carbon monoxide using developed support vector machine. *Atmospheric Pollution Research*, 7(3), 412–418.

29. Moustris, K.P., & Ziomas, I.C. (2010). Paliatsos, A.G.: 3-day-ahead forecasting of regional pollution index for the pollutants NO₂, CO, SO₂, and O₃ using artificial neural networks in athens, greece. *Water, Air, & Soil Pollution*, 209(1), 29–43.
30. 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_{2.5} concentration forecasting. *Atmospheric Environment*, 134, 168–180.
31. Omkar, S.N., Khandelwal, R., Ananth, T.V.S., Narayana Naik, G., Gopalakrishnan, S. (2009). Quantum behaved particle swarm optimization (QPSO) for multi-objective design optimization of composite structures. *Expert Systems with Applications*, 36(8), 11,312–11,322.
32. Ortiz-García, E.G., Salcedo-Sanz, S., Pérez-Bellido, M., Portilla-Figueroas, J.A., Prieto, L. (2010). Prediction of hourly O₃ concentrations using support vector regression algorithms. *Atmospheric Environment*, 44(35), 4481–4488.
33. Pai, P.F., & Hong, W.C. (2005). Support vector machines with simulated annealing algorithms in electricity load forecasting. *Energy Conversion and Management*, 46(17), 2669–2688.
34. Pan, L., Sun, B., Wang, W. (2011). City air quality forecasting and impact factors analysis based on grey model. *Procedia Engineering*, 12, 74–79.
35. Pei, J., Liu, X., Pardalos, P.M., Fan, W., Yang, S. (2017). Scheduling deteriorating jobs on a single serial-batching machine with multiple job types and sequence-dependent setup times. *Annals of Operations Research*, 249(1–2), 175–195.
36. Pei, J., Pardalos, P.M., Liu, X., Fan, W., Yang, S. (2015). Serial batching scheduling of deteriorating jobs in a two-stage supply chain to minimize the makespan. *European Journal of Operational Research*, 244(1), 13–25.
37. Reyes, J.M., & Serre, M.L. (2014). An LUR/BME framework to estimate PM_{2.5} explained by on road mobile and stationary sources. *Environmental Science & Technology*, 48(3), 1736–44.
38. Ridder, K.D., Kumar, U., Lauwaet, D., Blyth, L., Lefebvre, W. (2012). Kalman filter-based air quality forecast adjustment. *Atmospheric Environment*, 50(4), 381–384.
39. Schölkopf, B. (2008). *The nature of statistical learning theory* springer.
40. Song, X., Liu, Y., Hu, Y., Zhao, X., Tian, J., Ding, G., Wang, S. (2016). Short-term exposure to air pollution and cardiac arrhythmia: a meta-analysis and systematic review. *International Journal of Environmental Research and Public Health*, 13(7), 642.
41. Song, Y., Qin, S., Qu, J., Liu, F. (2015). The forecasting research of early warning systems for atmospheric pollutants: a case in yangtze river delta region. *Atmospheric Environment*, 118(118), 58–69.
42. Sun, J., Feng, B., Xu, W. (2004). Particle swarm optimization with particles having quantum behavior. In *2004. CEC2004. Congress on evolutionary computation*, (Vol. 1 pp. 325–331).
43. Sun, W., & Sun, J. (2016). Daily PM_{2.5} concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm. *Journal of environmental management*, 188, 144–152.
44. Suresha, C.M., Lakshminarayanachari, K., Prasad, M.S., Pandurangappa, C. (2012). Advection - diffusion numerical model of an air pollutant emitted from an area source of primary pollutant with chemical reaction and dry deposition. *International Journal of Engineering Science and Technology*, 4(1), 82–97.
45. Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S., Kukkonen, J. (2011). Evaluation of a multiple regression model for the forecasting of the concentrations of NO_x and PM₁₀ in athens and helsinki. *Science of the Total Environment*, 409(8), 1559–1571.
46. Wang, J., Hou, R., Wang, C., Shen, L. (2016). Improved μ -support vector regression model based on variable selection and brain storm optimization for stock price forecasting. *Applied Soft Computing*, 49, 164–178.
47. Wang, S. (2012). Air quality management in china: issues, challenges, and options. *Journal of Environmental Sciences*, 24(1), 2–13.
48. Xu, Y., Yang, W., Wang, J. (2016). Air quality early-warning system for cities in China. *Atmospheric Environment*.
49. Yeganeh, B., Motlagh, M.S.P., Rashidi, Y., Kamalan, H. (2012). Prediction of CO concentrations based on a hybrid partial least square and support vector machine model. *Atmospheric Environment*, 55(3), 357–365.
50. Miao, Y., Liu, S., Zheng, Y., Wang, S., Liu, Z., Zhang, B. (2015). Numerical study of the effects of planetary boundary layer structure on the pollutant dispersion within built-up areas. *Journal of Environmental Sciences*, 32(6), 168–179.
51. Zhang, H., Wang, S., Hao, J., Wang, X., Wang, S., Chai, F., Li, M. (2016). Air pollution and control action in Beijing. *Journal of Cleaner Production*, 112, 1519–1527.
52. Zhang, J., Tittel, F.K., Gong, L., Lewicki, R., Griffin, R.J., Jiang, W., Jiang, B., Li, M. (2016). Support vector machine modeling using particle swarm optimization approach for the retrieval of atmospheric ammonia concentrations. *Environmental Modeling & Assessment*, 21(4), 531–546.
53. Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., Baklanov, A. (2012). Real-time air quality forecasting, part i: history, techniques, and current status. *Atmospheric Environment*, 60(32), 632–655.
54. Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., Baklanov, A. (2012). Real-time air quality forecasting, part ii: State of the science, current research needs, and future prospects. *Atmospheric Environment*, 60(6), 656–676.
55. Zhao, R., Chen, S., Wang, W., Huang, J., Wang, K., Liu, L., Wei, S. (2017). The impact of short-term exposure to air pollutants on the onset of out-of-hospital cardiac arrest: a systematic review and meta-analysis. *International Journal of Cardiology*, 226, 110.
56. Zheng, S., Yi, H., Li, H. (2015). The impacts of provincial energy and environmental policies on air pollution control in china. *Renewable & Sustainable Energy Reviews*, 49, 386–394.
57. Zheng, Y., Capra, L., Wolfson, O., Yang, H. (2014). Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology*, 5(3), 38.
58. Zheng, Y., Liu, F., Hsieh, H.P. (2013). U-air: when urban air quality inference meets big data. In *ACM SIGKDD International conference on knowledge discovery and data mining* (pp. 1436–1444).
59. Zheng, Y., Yi, X., Li, M., Li, R., Shan, Z., Chang, E., Li, T. (2015). Forecasting fine-grained air quality based on big data. In *The ACM SIGKDD international conference* (pp. 2267–2276).