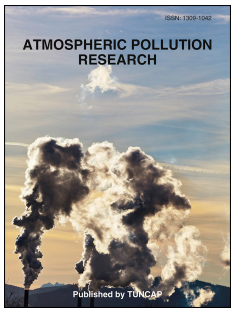
Journal Pre-proof



Prediction of ozone hourly concentrations by support vector machine and kernel extreme learning machine using wavelet transformation and partial least squares methods

Xiaoqian Su, Junlin An, Yuxin Zhang, Ping Zhu, Bin Zhu

PII: S1309-1042(20)30057-X

DOI: https://doi.org/10.1016/j.apr.2020.02.024

Reference: APR 761

To appear in: *Atmospheric Pollution Research*

Received Date: 22 November 2019

Revised Date: 24 February 2020

Accepted Date: 28 February 2020

Please cite this article as: Su, X., An, J., Zhang, Y., Zhu, P., Zhu, B., Prediction of ozone hourly concentrations by support vector machine and kernel extreme learning machine using wavelet transformation and partial least squares methods, *Atmospheric Pollution Research*, https://doi.org/10.1016/j.apr.2020.02.024.

Highlights

- SVR, KELM, BPNN and SR are applied to the hourly O_3 concentration prediction.
- Meteorological elements, precursors and O_3 concentrations 6 hours ago are used as model inputs.
- Improving models using WT and PLS methods.
- The KELM-WT-PLS model has better prediction effect with R^2 up to 0.78.

Prediction of ozone hourly concentrations by support vector machine and kernel extreme learning machine using wavelet transformation and partial least squares methods

Xiaoqian Su[a], Junlin An[a,*], Yuxin Zhang[b], Ping Zhu[a], Bin Zhu[a]

[a] Key Laboratory of Meteorological Disaster, Ministry of Education (KLME), Joint International Research Laboratory of Climate and Environment Change (ILCEC), Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD), Nanjing University of Information Science and Technology, Nanjing 210044, China

[b] Weather Modification Office of Qinghai Province, Xining 810001, China

[*] Corresponding author

ABSTRACT

In this paper, we develop a method for predicting ozone(O_3) concentration based on kernel extreme learning machine (KELM) and support vector machine regression (SVR) and pretreat it by wavelet transformation (WT) and partial least squares (PLS). To test the method's effectiveness, the observation (2014 – 2016 summer) of the precursors, meteorology and hourly O_3 concentrations in the Nanjing industrial zone were applied. The mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), normalized root mean square error (NRMSE) and coefficient of determination(R^2) were chosen to evaluate the model. Results demonstrate that the KELM and SVR perform better than stepwise regression (SR) methods and back propagation neural network (BPNN) for predicting O_3 concentration. WT decomposes the original time series of O_3 concentration into a few sub–series with less variability, and then improve the performance of SVR and KELM by 16.99%~30.91% and 16.00%~25.86%, respectively. The variable importance in projection (VIP) value was used to filter the influence factors of each sub–sequence, which can remove redundant information and reduce the calculation amount of the model. In addition, the WT and PLS methods enhance the predictive abilities of KELM and SVR for higher O_3 concentrations by 21% and 35%

respectively. The KELM-WT-PLS model shows the best fit of the O_3 hourly concentration, and the corresponding MAE, MAPE, RMSE, NRMSE and R^2 are 7.71 ppb, 0.37, 9.75 ppb, 11.83% and 0.78, while KELM predict the O_3 hourly concentration more accurately.

keywords: ozone concentration forecast; kernel extreme learning machine; support vector machine; wavelet transformation; partial least squares; variable importance in projection

1. Introduction

With the development of industry and Commerce and the increase of cars ownership, the anthropogenic emissions of nitrogen oxides (NO_x) (Duncan et al., 2016; Krotkov et al., 2016) and volatile organic compounds (VOCs) (Lu et al., 2013; Smedt et al., 2015) in the near–surface atmosphere are increasing. The complex air pollution problems such as photochemical smog (Hamer et al., 2015) are seriously affecting human health (Goodman et al., 2015; Karlsson et al., 2017) and ecological environment (Feng et al., 2015; Wang et al., 2012). Studying the forecasting method of air pollutants and establishing a timely early warning mechanism of air pollutants are of great application value for improving the air quality of cities and formulating control strategies.

Air pollution prediction methods can be generally divided into numerical and statistical prediction. The numerical prediction methods realize the simulation of pollutant transformation, migration and diffusion and reflect the changing law of pollutants, but they are based on a large number of meteorological data, pollutant emission source data and air monitoring data, need to grasp the mechanisms of pollution change and take a long time to calculate.O_3 concentration forecast involves nonlinear, strong coupling and multivariate problems, thus the numerical prediction will be a very complicated system engineering. Statistical forecasting methods such as regression model (An and Wang, 2010; Zhai et al., 2018) are widely used in operational forecasting which have the advantages of simple calculation, low data requirement and high accuracy. However, most of them are based on linear regression theory, assuming that the pollutant concentrations are not directly related to the source of pollution, it is difficult to apply them to non–linear

and strong coupling systems. In recent years, with the development of computer technology, artificial intelligence and machine learning theory have been widely studied and applied. Machine learning methods that based on statistical theory such as neural network (Gao et al., 2018; Mao et al., 2017), decision tree (Ding et al., 2018), SVR and ELM have shown excellent performances in dealing with non–linear problems.

SVR follows the principle of structural risk minimization. Unlike traditional machine learning methods which follow the principle of empirical risk minimization, the SVR avoids the problems of over–fitting, local optimization, difficulty in parameter adjustment and slow convergence (Lu and Wang, 2005; Nieto et al., 2013), and requires fewer parameters to be adjusted. In recent years, SVR was not only used to predict solar radiation (Quej et al., 2017), cloud cover (Zhao et al., 2016) and visibility (Wu et al., 2017), but also widely used to predict air pollutant concentrations such as O_3 (Lee et al., 2018; Luna et al., 2014; Mehdipour and Memarianfard, 2019; Ortiz-García et al., 2010; Salazar-Ruiz et al., 2008; Xu et al., 2016). Some researchers compared the SVR with commonly used statistical forecasting models such as multilayer perception (MLP) (Nieto et al., 2017), linear regression model (Canu and Rakotomamonjy, 2001), vector auto regressive model (VARMA) and auto–regressive integral moving average model (ARIMA) (Nieto et al., 2018), found that the SVR has better prediction effect on pollutants. Compared with many statistical forecasting methods, the SVR showed more feasibility and superiority.

In order to overcome the limitations of the artificial neural network based model, the extreme learning machine (ELM) was proposed, which is a new single hidden layer feed–forward neural network training method (Huang and Chen, 2007; Huang et al., 2011). After randomly determining the input hidden weights and hidden biases, the hidden output weights can be directly obtained by calculating the Moore–Penrose generalized inverse of the hidden output matrix. Compared to gradient–based methods, the ELM has better generalization capabilities and faster learning rates, and it has been initially applied in O_3 prediction and performed well (Feng et al., 2019; Peng et al., 2017; Zhang and Fu, 2017). However, the main drawback of ELM and its variants is that they rely on the choice of neurons in the hidden layer and the correct activation function, and the established prediction system is unstable. To solve this

problem, Huang et al. (2012) proposed the KELM by comparing the modelling and solving processes between ELM and vector machines. Compared with ELM, the KELM is more robust and performs better in solving regression prediction problems. In recent years, KELM has been widely used in engineering technology (Lin et al., 2018; Shi et al., 2019), however, its validation in air pollutants prediction has not been tested.

WT is a useful tool for obtaining time and frequency information from sequences (Chen and Zhao, 2013), and has been widely used in information science. In recent years, WT theory has also shown strong vitality in the atmospheric field (Li and Tao, 2018; Liu et al., 2018). O_3 has complex non–linear response relationship with precursors and meteorological conditions (Borrego et al., 2003; Liu et al., 2017a), whose time series have non–stationary and high variability characteristics, and it is difficult to accurately predict. Using WT theory to transform high–variability time series into multiple low variability sub–sequences has obvious advantages. For most models, the WT method was an effective way to improve the prediction accuracy (Dunea et al., 2015; Farajzadeh and Alizadeh, 2017). On the other hand, the ozone concentrations are affected by multiple factors such as meteorological conditions and precursors, it is not advisable to take all variables into account because of the interaction of factors. PLS is a supervised feature extraction method, which can extract the most comprehensive explanatory variables through the principal component analysis (PCA) and the synthesis of variable extraction, and effectively improved the explanatory and accuracy of the prediction models (Li and Tao, 2018; Yeganeh et al., 2012).

The work presented in this study aims to examine the feasibility of applying SVR and KELM models to predict O_3 hourly concentrations based on the measured database in Nanjing, try to optimize these two models with WT and PLS methods and compare their performances. These hybrid models provide novel alternative and optimized idea for air pollutants concentration forecasting.

2. Data and Method

2.1 Sampling site and modelling dataset

The modelling data including hourly O_3, NO, NO_2, CO and VOCs concentrations (1–h mean value) were collected at the observation site of Qixiang Building (32.21°N, 118.72°E) located on the Nanjing University of Information Science and Technology, China (Fig. 1). The data's observation period during 1/5/2014 to 25/8/2014, 1/6/2015 to 30/8/2015 and 1/5/2016 to 27/8/2016 were mostly in the summer with high O_3 concentrations.

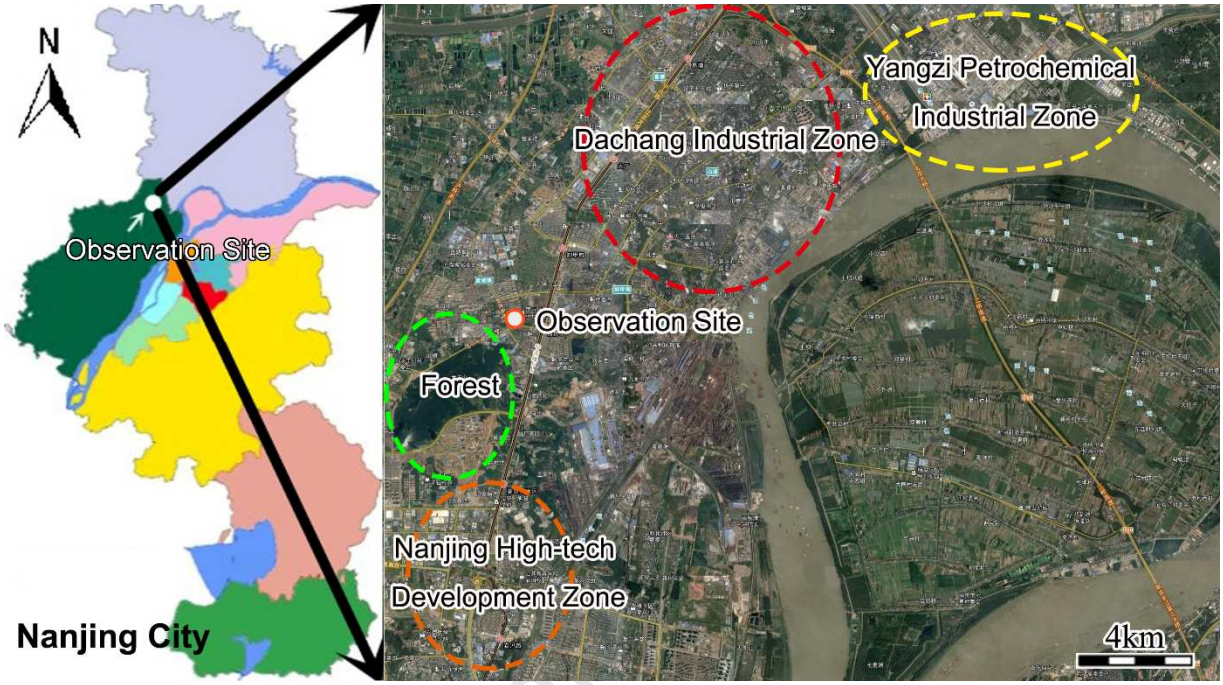


Fig. 1. Location of the observation site and the surrounding environment.

O_3, NO, NO_2, and CO concentrations were continuously measured using calibrated instruments from Thermo Environmental Instruments (TEI Inc., USA). The instruments' detailed parameters and calibration methods can be found in reference (An et al., 2015). VOCs concentrations were online monitored using the GC5000 analysis system from AMA Instruments GmbH (AMA, Germany), and they were classified into four classes in the study: alkane (*ALK*), olefin (*OLE*), acetylene (*ACE*) and aromatic hydrocarbon (*AH*), the specific species classified and detailed analytical method can be found in reference (An et al., 2014). In this paper, values were expressed by volume mixing ratios (ppb). The intuitive statistical information of pollutants data is shown in Figure S1.

The meteorological hourly data (1–h mean value) were collected by the automatic weather station on the campus which is approximately 800 m from the observation site, including temperature (*T*), atmospheric pressure (*P*), relative

humidity (*RH*), wind speed (*WS*), and wind direction (*WD*).

Table 1 shows the correlation analysis results of various variables. The correlation between O_3 and relative humidity (-0.74) is the best, followed by temperature (0.58). Among several O_3 precursors, NO, NO_x, olefin and aromatic hydrocarbon are relatively more important. In this paper, the single–site ozone prediction dominated by local photochemical reactions is mainly considered, thus days with rainfall more than 0 mm were removed. Further, figure 2 shows that the wind speed during observation period is mostly lower than 5 m/s, and it can be considered that there is no strong horizontal transport of pollutants, and the data used for modelling can represent the local photochemical reaction. A detailed statistical summary of modelling data including minimum, maximum, mean and standard deviation is shown in Table 2.

Table 1. Correlation analysis of all data used in prediction of O_3.

	O_3	*NO*	NO_2	NO_x	*CO*	*ALK*	*OLE*	*ACE*	*AH*	*WS*	*T*	*RH*	*WD*	*P*
O_3	1	**-0.50**	-0.36	**-0.45**	-0.31	-0.38	**-0.41**	-0.20	**-0.44**	0.35	**0.58**	**-0.74**	0.07	-0.12
NO	-0.50	1	0.52	0.67	0.32	0.37	0.47	0.16	0.33	-0.06	-0.26	0.32	-0.15	0.09
NO_2	-0.36	0.52	1	0.97	0.51	0.62	0.49	0.52	0.58	-0.20	-0.23	0.12	-0.32	0.16
NO_x	-0.45	0.67	0.97	1	0.53	0.63	0.53	0.50	0.59	-0.20	-0.27	0.21	-0.31	0.15
CO	-0.31	0.32	0.51	0.53	1	0.53	0.38	0.32	0.54	-0.10	–	0.26	-0.35	0.05
ALK	-0.38	0.37	0.62	0.63	0.53	1	0.65	0.56	0.77	-0.28	-0.19	0.29	-0.40	0.05
OLE	-0.41	0.47	0.49	0.53	0.38	0.65	1	0.25	0.51	-0.24	-0.15	0.33	-0.27	0.07
ACE	-0.20	0.16	0.52	0.50	0.32	0.56	0.25	1	0.56	-0.28	-0.15	0.17	-0.10	–
AH	-0.44	0.33	0.58	0.59	0.54	0.77	0.51	0.56	1	-0.24	-0.22	0.36	-0.39	0.07
WS	0.35	-0.06	-0.20	-0.20	-0.10	-0.28	-0.24	-0.28	-0.24	1	0.21	-0.51	-0.15	0.05
T	0.58	-0.26	-0.23	-0.27	–	-0.19	-0.15	-0.15	-0.22	0.21	1	-0.45	–	-0.47
RH	-0.74	0.32	0.12	0.21	0.26	0.29	0.33	0.17	0.36	-0.51	-0.45	1	–	–
WD	0.07	-0.15	-0.32	-0.31	-0.35	-0.40	-0.27	-0.10	-0.39	-0.15	–	–	1	–
P	-0.12	0.09	0.16	0.15	0.05	0.05	0.07	–	0.07	0.05	-0.47	–	–	1

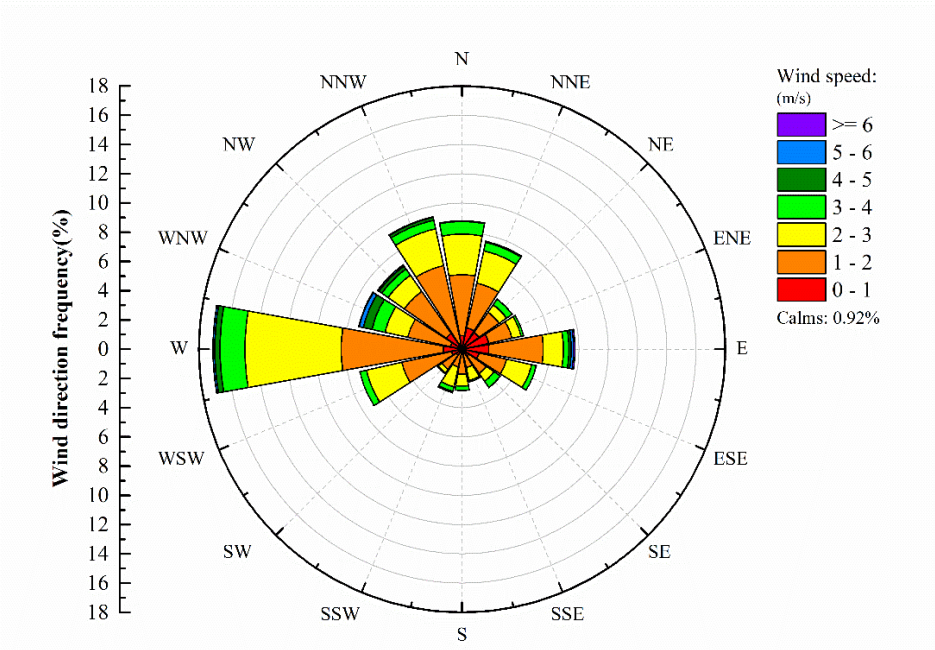


Fig. 2. Wind speed and wind direction during the observation period from 2014 to 2016.

Table 2. Statistics of measured values. Unit, minimum, maximum, mean, and standard deviation values during observation period.

Variable	Unit	Min	Max	Mean	Rstd (%)
O_3	ppb	0.50	116.20	32.94	74.86
NO	ppb	0.25	102.87	5.16	146.90
NO_2	ppb	0.25	81.00	20.11	57.28
NO_x	ppb	1.60	129.45	25.27	64.62
CO	ppb	20.00	4288.46	807.05	68.46
Alkane	ppb	2.76	73.98	16.45	61.28
Olefin	ppb	0.14	97.07	7.18	94.01
Acetylene	ppb	0.02	17.59	3.52	61.08
Aromatic hydrocarbon	ppb	0.80	71.85	8.39	79.14
Wind speed	$m \cdot s^{-1}$	0.00	6.40	1.82	50.55
Wind direction	°	0.00	359.00	165.91	52.08
Temperature	°C	9.80	40.30	25.96	16.87
Relative humidity	%	13.00	100.00	64.39	29.77
Atmospheric pressure	hPa	995.70	1021.40	1004.70	0.37

2.2 Wavelet transformation

Wavelet transformation (WT) is an effective time–frequency analysis method for the signal process, and it decomposes a signal directly according to the frequency. The WT can be divided into continuous wavelet

transformation (CWT) and discrete wavelet transformation (DWT). Among them, the DWT proposed by Mallat (1989) requires less time and is easy to implement, and has been widely used. For n–level decomposition and reconstruction, the original signal *s* can be expressed as:

$$s = a_n + \sum_{j=1}^{n} d_j \quad (1)$$

where a_n is the approximation series representing the low–frequency component, which contains trend information; d_j is the detail series on j level representing the high–frequency component, which contains periodic information. Essentially, this is a process in which the low frequency sequence is decomposed into low frequency subsequences and relatively high frequency subsequences with the increase of n (Fig. 3). The details and calculation procedures of the WT algorithm can be found in reference (Liu et al., 2017b).

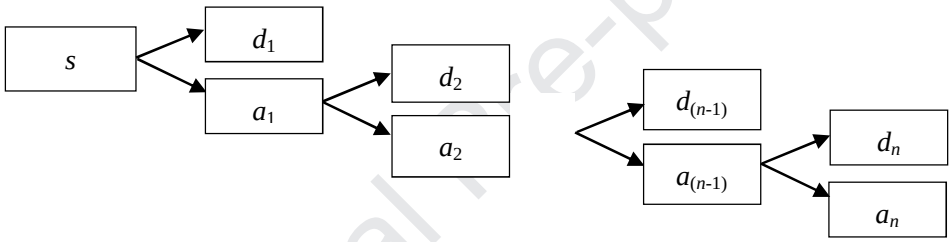


Fig. 3. Schematic diagram of *n*-level wavelet decomposition.

Fig. 4 shows the results of the 5–level wavelet decomposition of the original time series of O_3 concentrations by applying Daubechies Db5 wavelets implemented in the wavelet toolbox of MATLAB R2014b. Db5 is chosen as the wavelet function because it provides smaller variability of time series at the particular levels and its demonstrated good performance in related studies (He et al., 2017; Siwek and Osowski, 2012; Xiao et al., 2015). The optimal value of *n* was determined by the smoothness measure which can be written as:

$$\text{Smooth}(j) = \frac{\sum_{i=1}^{N-1}(a_j(i+1)-a_j(i))^2}{\sum_{i=1}^{N-1}(s(i+1)-s(i))^2} \quad (2)$$

where *N* represents the length of the series; j represents decomposition levels; *s* is the original series, and a_j is the approximation series on j level. Once Smooth(j) $\leq$ 0.005, j can be selected as the optimal decomposition level.

Calculate the Smooth value of *j* from 1 to 7, as shown in Table 3, Smooth (5) =0.003, so 5–level wavelet decomposition is used in this paper.

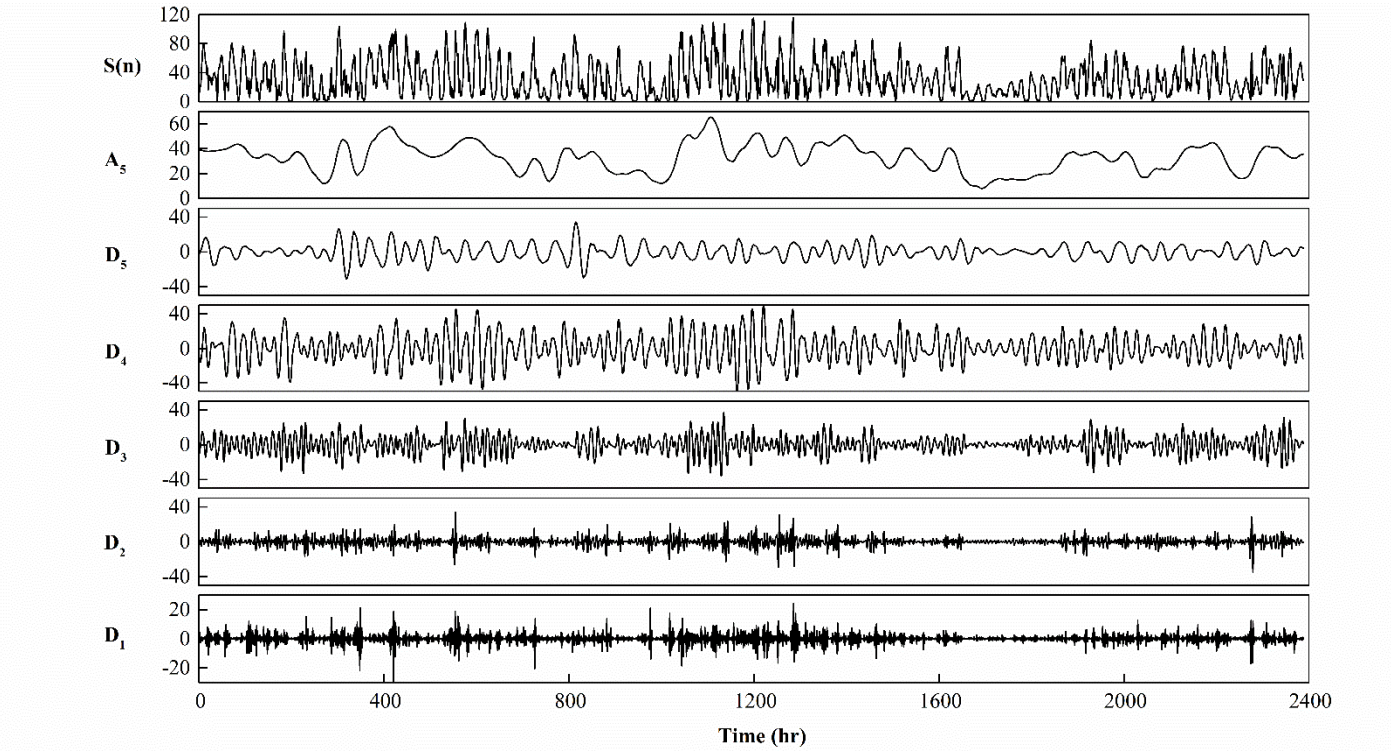


Fig. 4. The wavelet decomposition of the original time series *s* of O_3 concentrations: D_1~D_5 denote the wavelet coefficients at different levels and A_5 is the approximated signal of *s* on the fifth level.

Table 3. Value of Smooth from level 1 to level 7.

Level	1	2	3	4	5	6	7
Smooth	0.6481	0.4104	0.2501	0.0209	0.0021	0.0004	0.0001

2.3 Support vector machine regression

Support vector machine (SVM) developed by Vapnik is supervised machine learning algorithm that is widely employed for regression and forecasting. As compared to the ANN model that normally uses empirical risk minimization, the SVR uses the structural risk minimization principle which provides an upper bound on the generalization error (Vapnik, 2000). The approximated regression functions in the SVR algorithm is generated by

applying a set of high–dimension linear functions as follows:

$$y = w\varphi(x) + b \tag{3}$$

where $\varphi(x)$ is the higher–dimensional feature space; x is the input space; y is the output; w is the weights vector and b is a parameter of bias, that both can be estimated by minimizing the following regularized risk function:

$$R = \frac{\|w\|^2}{2} + C\sum_{i=1}^{N} L_\varepsilon\,(x_i, y_i, f) \tag{4}$$

where

$$L_\varepsilon(x_i, y_i, f) = \begin{cases} |y_i - f(x_i)| - \varepsilon, & |y_i - f(x_i)| \geq \varepsilon \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

Here, the $\frac{\|w\|^2}{2}$ regularization term is used as a measurement of function flatness; C is the cost parameter that determines the trade–off between the model flatness and the training error; ε is the tube size; The $L_\varepsilon(x_i, y_i, f)$ is ε–insensitive loss function, which penalizes the error greater than ε. Eq. (4) is expressed in the following constrained formation by introducing slack variables ξ_i,ξ_i^*:

$$\min\left(\frac{\|w\|^2}{2} + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)\right) \tag{6}$$

subject to

$$y_i - \big((w \times x_i) + b\big) \leq \varepsilon + \xi_i \tag{7}$$

$$\big((w \times x_i) + b\big) - y_i \leq \varepsilon + \xi_i^* \tag{8}$$

$$\xi_i^*, \xi_i \geq 0 \tag{9}$$

The method of using the Lagrangian multiplier α can be used to solve this quadratic programming problem. Thus, $f(x)$ can be finally expressed in an explicit form:

$$f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(x_i, x) + b \tag{10}$$

where $\alpha_i\alpha_i^* = 0,\ \alpha_i, \alpha_i^* \geq 0$. $K(x_i, x)$ is the kernel function and obtained by $K(x_i, x) = \varphi(x_i)^T\varphi(x)$ in the feature space. The common radial basis function $K(x_i, x) = \exp(-g\|x_i - x\|^2)$ where g is the spread of the RBF kernel is used in this study. The details and calculation procedures of the SVR algorithm can be found in Vapnik (2000).

The SVR algorithm in this study is applied using the LIBSVM3.22 software package

(http://www.csie.ntu.edu.tw/~cjlin/libsvm/) on the MATLAB platform, the cost parameter *C* and the parameter *g* in the kernel function are optimized by systematic grid search method using 5–fold cross validation on the training set.

2.4 Kernel extreme learning machine

The extreme learning machine (ELM) is a simple learning algorithm for single–hidden layer feed–forward network (SLFN), randomly assigning the input weights and hidden layer biases (Huang et al., 2006), and has the advantages of simple implementation, fast learning speed, less intervention conditions and strong generalization ability (Huang et al., 2012). For any given set of *N* training samples $\{(\boldsymbol{x}_j, \boldsymbol{y}_j), \boldsymbol{x}_j \in \boldsymbol{R}^m, \boldsymbol{y}_j \in \boldsymbol{R}^n, j = 1,2, \dots, N\}$, the output of a conventional SLFN with *L* hidden nodes, can be expressed as below:

$$\boldsymbol{o}_j = \sum_{i=1}^{L} \boldsymbol{\beta}_i g\left(\boldsymbol{w}_i \boldsymbol{x}_j + b_i\right), \quad j = 1,2, \dots, N \tag{11}$$

where $\boldsymbol{x}_j = [x_{j1}, x_{j2}, \dots, x_{jm}]^{\mathrm{T}}$ is the input vector with m nodes; $\boldsymbol{o}_j = [o_{j1}, o_{j2}, \dots, o_{jn}]^{\mathrm{T}}$ is the output vector with n nodes; $\boldsymbol{w}_i = [w_{j1}, w_{j2}, \dots, w_{jm}]^{\mathrm{T}}$ is the weight vector between the input nodes and the i–th hidden node; $\boldsymbol{\beta}_i = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jn}]^{\mathrm{T}}$ is the weight vector between the output nodes and the i–th hidden node; b_i represents the bias of the i–th hidden node and $g(\cdot)$ is the nonlinear activation function of the hidden layer.

The above expression can be written as:

$$\mathbf{T} = \mathbf{H}\boldsymbol{\beta} \tag{12}$$

where **T** is the respected output of the SLFN; **H** is the output matrix of the hidden layer of the SLFN, which can be expressed as:

$$\mathbf{H} = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} g(w_1 x_1 + b_1) & \cdots & g(w_L x_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(w_1 x_N + b_1) & \cdots & g(w_L x_N + b_L) \end{bmatrix}_{N \times L} \tag{13}$$

The optimization objective of ELM is to find appropriate parameters making $\sum_{j=1}^{N} \left\| \boldsymbol{o}_j - \boldsymbol{y}_j \right\| = 0$ hold, when existing $\boldsymbol{w}_i$, $\boldsymbol{\beta}_i$ and b_i make Eq. (14) set up:

$$\boldsymbol{y}_j = \sum_{i=1}^{L} \boldsymbol{\beta}_i g\left(\boldsymbol{w}_i \boldsymbol{x}_j + b_i\right), \quad j = 1,2, \dots, N \tag{14}$$

The minimal norm least square solution of weight vector **β** is shown as below:

$$\boldsymbol{\beta} = \mathbf{H}^{\dagger}\mathbf{T} \tag{15}$$

where $\mathbf{H}^{\dagger}$ is the Moore–Penrose generalized inverse of matrix $\mathbf{H}$. Based on orthogonal projection method and ridge regression theory (Hoerl and Kennard, 1970), the regularization coefficient C was adopted in the optimization phase, then the solution of output weight $\boldsymbol{\beta}$ becomes:

$$\boldsymbol{\beta} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H} + \frac{\mathbf{I}}{C}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{T} \tag{16}$$

where $\mathbf{I}$ denotes the identity matrix. Hence, the output function of ELM can be written as follows:

$$f(x) = h(x)\boldsymbol{\beta} = h(x)\left(\mathbf{H}^{\mathrm{T}}\mathbf{H} + \frac{\mathbf{I}}{C}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{T} \tag{17}$$

In order to overcome the randomness of ELM, and improve its stability and generalization capability, kernel functions can be used for the optimized ELM, namely the kernel extreme learning machine (KELM) (Huang et al., 2012; Yao et al., 2014).In the KELM, the activation function $h(x)$ is replaced by kernel matrix $K(x, x_j)$, which can be shown as:

$$\Omega = \mathbf{H}\mathbf{H}^{\mathrm{T}}: \Omega_{\mathrm{ELM}i,j} = h(x_i)h(x_j) = K(x_i, x_j) \tag{18}$$

The output function of the KELM can be expressed as

$$f(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}\left(\Omega_{\mathrm{ELM}} + \frac{\mathbf{I}}{C}\right)^{-1}\mathbf{T} \tag{19}$$

Thus, only the type of the kernel function needs to be defined instead of the hidden layer's node number and activation function. The details and calculation procedures of the ELM algorithm can be found in Huang et al.(2006). In this study, the radial basis function is used as a kernel function similar to SVR.

2.5 Back Propagation neural network

Back Propagation neural network (BPNN) is one of the commonly used neural networks with strong nonlinear regression capability (Bai et al., 2016; Feng et al., 2011). The architecture of this network is consisted of input layer, one or more hidden layers and output layers. Each layer consists of multiple neurons. In this paper, a single hidden layer was selected. The tansig and purelin function are used as the transfer functions for the hidden layer and the

output layer, respectively. When modelling with BPNN, the number of hidden nodes is directly related to training efficiency. Normally, the trial and error method and empirical formula are applied to solve this problem (Shen et al., 2008), which can be expressed as:

$$\text{hidden nodes} = \sqrt{m + p} + a \tag{20}$$

where m is the number of input layer nodes; p is the number of output layer nodes; a represents a constant between 0 and 10. Through experiments, mean squared error between the networks outputs and the actual observations is minimal, when the number of hidden layer nodes is set to 6. Therefore, the structure of the BPNN is determined to be 14–6–1 in this study.

2.6 Variable importance in projection technique based on PLS

Partial Least Squares (Wold, 1966) is one of the features of extraction method, which constructs a linear model to describe the connection between dependent variables Y and predictor variables X. This linear model attempts to find the multidimensional direction in the X space that explains the maximum multidimensional co–variance direction in the Y space. The variable importance in projection (VIP) is an assistant technology based on PLS, which can be used to select important variables. The VIP can measure the explanatory power of each independent variable to dependent variable, for the j–th independent variable, its VIP formula is as follows:

$$\text{VIP}_j = \sqrt{\frac{k}{\sum_{h=1}^{m} r^2(y, c_h)} \sum_{h=1}^{m} r^2(y, c_h) {w_{hj}}^2} \tag{21}$$

where x_j is the independent variables; y is the dependent variables; k is the number of independent variables; c_h is the principal components extracted from the relevant variables; $r(y,c_h)$ is the correlation coefficient between dependent variables and principal components, indicating the explanatory ability of the principal components to y; w_{hj} is the weight of the independent variables on the principal components.

The explanatory effect of x_j on y is transmitted through the principal component c_h. If the explanatory effect of c_h on y is very strong, and the interpretation of x_j on c_h is also very large, then it can be considered that x_j has a greater

explanatory effect on *y*.

2.7 Performance evaluation

The models' performance on learning and testing are assessed using the following five indexes: mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), normalized root mean square error (NRMSE) and coefficient of determination (R^2). These parameters can be defined as below:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|O_i - P_i| \tag{22}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|P_i - O_i|}{O_i} \tag{23}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2} \tag{24}$$

$$\mathrm{NRMSE} = \frac{RMSE}{O_{max} - O_{min}} \cdot 100\% \tag{25}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - O_{\mathrm{m}})^2} \tag{26}$$

where O_i is the observed value at time *i*; O_{max}, O_{min} and O_m are the maximum, minimum and average values of the observed value, respectively; P_i is the predicted value at time *i* and *n* is the total number of samples.

3. Results and discussion

In this study, experiments were carried out in MATLAB 2014 environment running in an Intel i5, 1.6 GHZ CPU. Data for 2014 and 2015 were used for training and validation, accounting for about 80% of the total dataset, and the data for 2016 were used for testing. Before modelling, all variables were normalized to [0,1] by Equation (1):

$$y = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{27}$$

where x is the original variable; $x_{min}$is minimum value of the variable; $x_{max}$is maximum value of the variable and y is the variable transformed by normalization.

The previous ozone concentration characterizes the cumulative effect of pollutants, which could effectively improve the prediction accuracy when forecasting ozone concentration (Chelani, 2010). Taking into account the

timeliness of the forecast and the diurnal variation of ozone, the ozone concentration 6–h ago (O_3_6) was also taken as a variable in this study.

3.1 Comparing the prediction performance of SVR, KELM, BPNN and SR

The stepwise regression, as a traditional statistical method, is still widely used in business forecasting. By inputting all variables into the stepwise regression model (F probability ≤0.05 to enter and F probability ≥ 0.10 to exit), the ozone hourly concentration prediction equation could be constructed as follows:

$$O_3 = -315.693 - 0.245NO_2 - 0.197NO_x - 0.002CO + 0.249ALK - 0.730OLE + 1.634ACE$$

$$-0.209AH + 2.612T + 0.334P - 0.681RH - 1.722WS - 0.011WD + 0.040O_3\_6 \quad (28)$$

Fig. 5 shows the prediction effect of inputting all variables into the KELM, SVR, BPNN and SR separately, the scatter plots are used to describe the relation between the observed and predicted value, which illustrate that the higher the consistency between the two data sets, the more points tend to concentrate near the identity line marked as 'ideal fit' in the figure. As can be seen from Table 4, the KELM has the best performance among the four models, with its MAE, MAPE, RMSE, NRMSE and R^2 are 10.50 ppb, 0.60, 12.97 ppb, 15.73%, 0.58 respectively. The five indicator values of SVR and BPNN, which represent prediction accuracy, are similar in the study, and the former has a slight advantage. The SR method has the lowest performance and its MAE, MAPE, RMSE and NRMSE are about 16%, 40%, 9%, 9% worse than SVR or BPNN, respectively. The results indicate the relationships between ozone concentration, precursors and meteorological conditions are nonlinear, which are difficult to be accurately reflected with traditional linear statistical models. Moreover, for the three machine learning methods, the modelling time of KELM including cross validation was 22.35 s, which was about 1.5 times faster than SVR and 3 times faster than BPNN, KELM performs better in terms of the learning speed. In this study, KELM and SVR with better prediction effect were selected for further optimization and comparison.

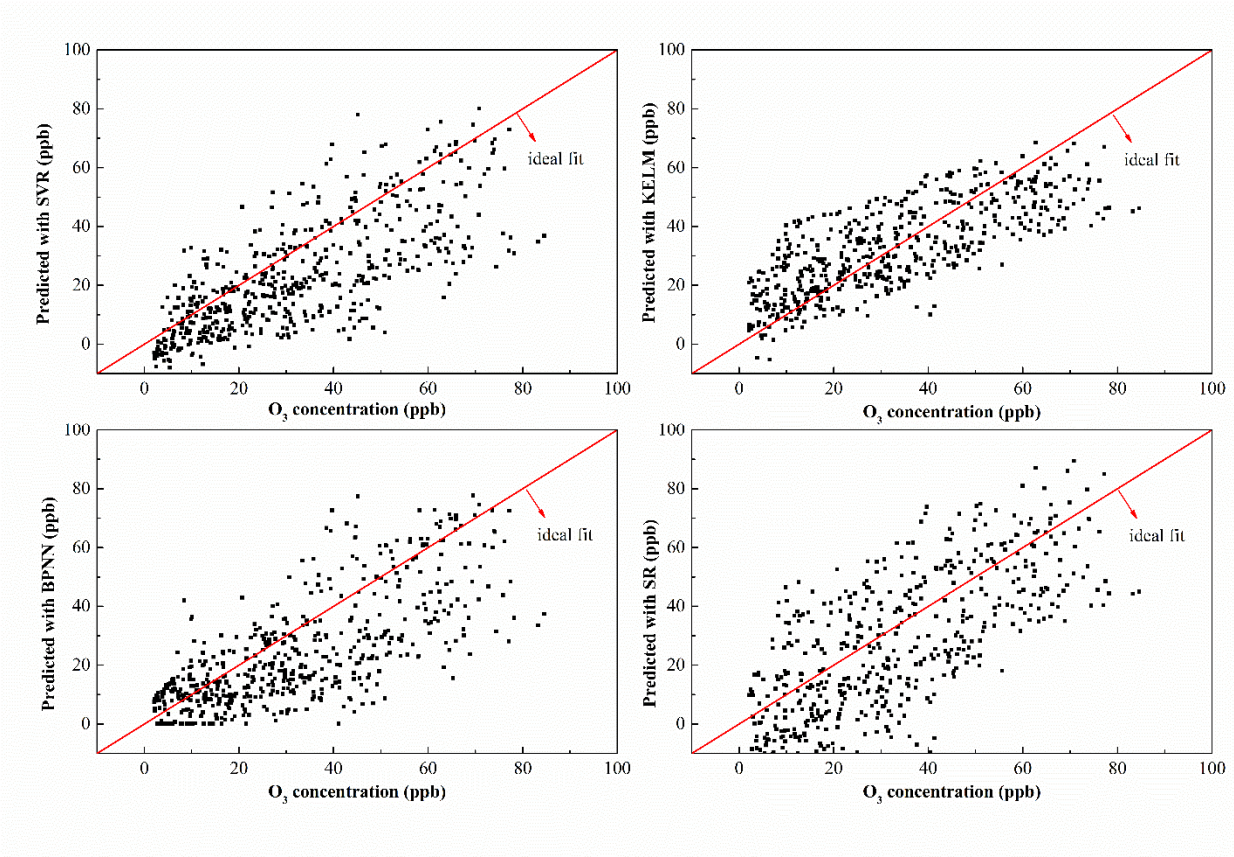


Fig. 5. Comparison of prediction effect of the SVR, KELM, BPNN and SR.

Table 4. Comparison of five evaluation indexes among SR, BPNN, SVR and KELM.

	MAE (ppb)	MAPE	RMSE (ppb)	NRMSE (%)	R^2
SR	15.58	0.89	18.61	22.57	0.54
BPNN	12.98	0.51	16.90	20.50	0.53
SVR	12.95	0.53	16.82	20.40	0.55
KELM	10.50	0.60	12.97	15.73	0.58

3.2 The result of WT and PLS methods on prediction

The 14 variables were decomposed into six sub–sequences respectively, including five high frequency sequences (D_1~D_5) which describe random characteristics, and one low frequency sequence (A_5) which describes trend change. The sub–sequences of O_3 concentration were used as the prediction object, and the sub–sequences of the remaining 13 variables at corresponding levels were used as input factors. Modelling and prediction were carried out at six levels respectively, and the final prediction results are obtained by adding the prediction results of the O_3 sub–sequences of

the six models. Applying the above method to KELM and SVR, it can be seen from Table 3 that KELM–WT fits better than SVR–WT at most levels, and its R^2 even reaches 0.91 in D_4.

The VIP was used in variable screening in this paper. If each variable has the same explanatory effect on dependent variables, the VIP values of each variable are all 1. For the independent variable with a relatively large VIP value, its contribution to dependent variables is particularly important. Generally, a variable with the VIP ≥ 0.8 can be considered to have a large contribution to the dependent variable (Wold, 1995). In this study, simplified calculations were used at all levels, as long as the VIP of an independent variable for a principal component is greater than 1, it is retained. The influencing factors of O_3 sub–sequences are shown in Table 5. Temperature and relative humidity are important at most levels. Precursors are the main influencing factor for detailed sub–sequences that characterize the sudden change of O_3. The traffic trunk lines and integrated industrial areas near the observation point were the main anthropogenic sources of these precursors (Lin et al., 2015). In addition, aromatic hydrocarbons are more important among the four VOCs. In Nanjing, the contribution of aromatic hydrocarbons to OFP (O_3 formation potential) and RIR (relative incremental reactivity) were very large, and it was one of the two most effective species to control the concentration of O_3 (Yang et al., 2016; Zhang et al., 2018). With the increase of decomposition levels, the main influencing factors of A_5 become *T*, *RH* and *O3_6*, indicating that the trend of O_3 concentration changes is dominated by meteorological conditions and the cumulative effect of pollutants.

As shown in Table 5, after using the PLS method to select variables, the accuracy of most O_3 sub–sequences prediction models including SVR and KELM is improved, and R^2 is increased by 1.10%~16.18%, which shows the effectiveness of the input variables selected using the PLS method.

Table 5. Input data used in prediction of six O_3 scales respectively and effect of using PLS to select variables.

Scales	WT+PLS		WT +all variables
	Input variable	R^2	R^2
		SVR/KEL	SVR/KELM

		M	
D_1	*NO*, NO_2, NO_x, *CO*, *T*	0.63/0.63	0.55/0.60
D_2	*NO*, NO_2, NO_x, *CO*, *T*, *RH*	0.60/0.58	0.59/0.57
D_3	NO_2, NO_x, *AH*, *T*, *RH*	0.71/0.79	0.71/0.81
D_4	NO_2, NO_x, *ALK*, *ACE*, *AH*, *WS*, *T*, *RH*	0.87/0.92	0.87/0.91
D_5	*NO*, NO_x, *AH*, *T*, *RH*	0.58/0.61	0.55/0.59
A_5	*ACE*, *T*, *RH*, O_3_*6*	0.68/0.79	0.62/0.68

3.3 Discussion on SVR–WT–PLS and KELM–WT–PLS

Table 6 shows the statistics of the final prediction results of each model. It can be seen that the prediction accuracy of the model is improved by using WT based on the original prediction method. Among them, MAE, MAPE, RMSE, NRMSE and R^2 of SVR–WT are 10.75 ppb, 0.54, 13.46 ppb, 16.33% and 0.72, respectively. The prediction accuracy is 16.99%~30.91% higher than that of SVR. MAE, MAPE, RMSE, NRMSE and R^2 of KELM–WT are 8.82 ppb, 0.47, 10.79 ppb, 13.08% and 0.73, which are 16.00%, 21.67%, 16.81%, 16.85% and 25.86% better than KELM, respectively. The results show that the accuracy of model prediction can be effectively improved by decomposing the original O_3 time series with high variability into several sub–series with lower variability, and then performing modelling and prediction separately.

Comparing the final prediction results of SVR–WT and SVR–WT–PLS, it is found that the MAE, MAPE, RMSE, NRMSE and R^2 of SVR–WT–PLS are 18.60%, 24.07%, 15.97%, 15.98% and 1.39% better than that of SVR–WT respectively after using PLS to select variables. In addition, KELM–WT–PLS shows the best prediction effect, with MAE, MAPE, RMSE, NRMSE and R^2 are 7.71 ppb, 0.37, 9.75 ppb, 11.83% and 0.78, which were 12.59%, 21.28%, 9.64%, 9.56% and 6.85% better than KELM–WT, respectively. The excessive number of input variables will cause information redundancy and slow down the calculation speed of the model. The PLS selection method can retain more important variables related to O_3 prediction, shorten the running time of the model and greatly improve the accuracy of the prediction.

Fig. 6 shows the model predictions and observations more intuitively. Both the KELM and the SVR have an ideal

fit to the O_3 hourly concentration, which can capture the peak and valley values of O_3 concentration more accurately. With the addition of optimization methods such as WT and PLS, the predictions of the high and low values of the two models are closer to the observations to varying degrees. According to the actual situation of O_3 observation data, the O_3 concentration were divided into three classes: low($[O_3] < 40$ ppb), medium (40 ppb $\leq [O_3] < 60$ ppb) and high($[O_3] \geq 60$ ppb) to further evaluate the prediction results of the model, and the results are shown in Table 4. With the increase of O_3 concentration, the prediction work is more difficult, and the prediction accuracy of each model for O_3 high–level is lower than that of the mid–level and low–level. However, after optimization by WT and PLS methods, the accuracy of the model for O_3 concentration prediction in three classes is greatly improved. For high–level O_3 that people are concerned about, the MAE, MAPE, RMSE and NRMSE of SVR–WT–PLS are decreased by about 35% compared with SVR, and the prediction accuracy of KELM–WT–PLS is also increased by about 21%, indicating that the prediction abilities of the two methods for high ozone concentration are improved.

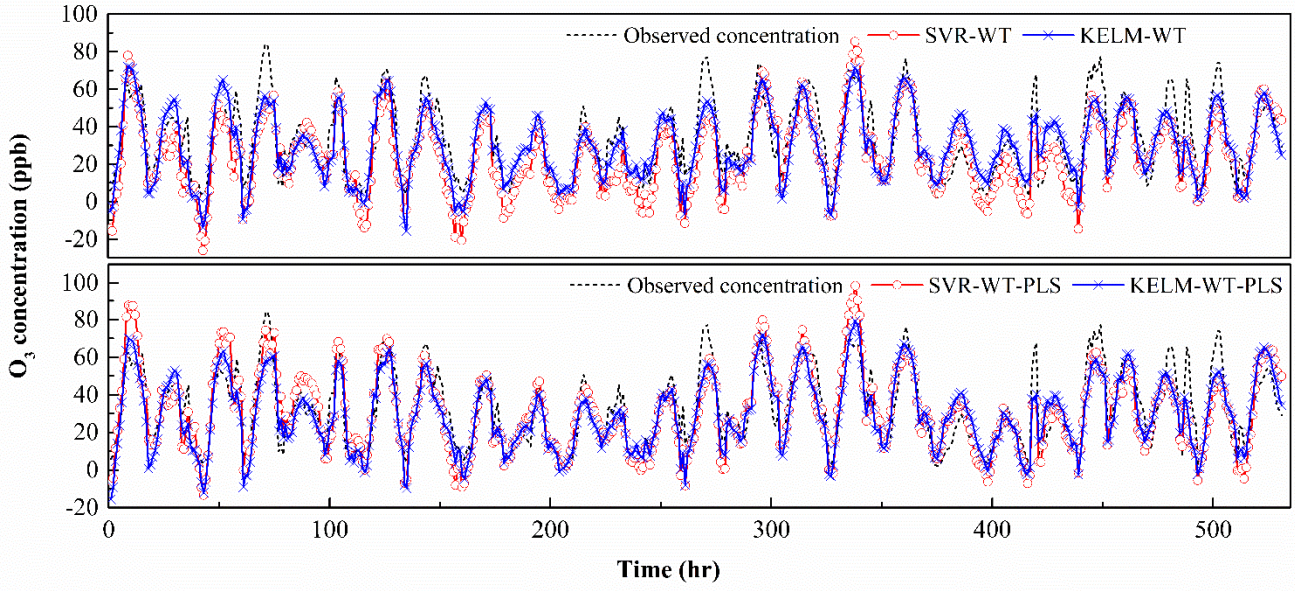


Fig. 6. Predictions of the SVR–WT, KELM–WT, SVR–WT–PLS and KELM–WT–PLS.

Table 6. Performance measures of six models at different ozone concentration levels.

Estimator	Classification Intervals (ppb)

		$[O_3]$ <40	40≤ $[O_3]$ <60	$[O_3]$ ≥60	all test database
SVR/KELM	MAE (ppb)	9.61/9.32	18.02/10.23	21.78/17.27	12.95/10.50
	MAPE	0.62/0.80	0.37/0.21	0.32/0.25	0.53/0.60
	RMSE (ppb)	11.97/11.62	21.68/12.13	26.43/19.53	16.82/12.97
	NRMSE (%)	31.74/30.81	108.67/60.81	108.26/80.01	20.40/15.73
	R^2	0.39/0.34	0.15/0.19	0.01/0.00	0.55/0.58
SVR–WT/KELM–WT	MAE (ppb)	9.14/7.72	11.31/8.87	18.28/14.53	10.75/8.82
	MAPE	0.70/0.61	0.23/0.18	0.27/0.21	0.54/0.47
	RMSE (ppb)	11.43/9.29	13.75/10.48	20.87/16.94	13.46/10.79
	NRMSE (%)	30.32/24.64	68.92/52.53	85.51/69.38	16.33/13.08
	R^2	0.54/0.51	0.20/0.23	0.00/0.03	0.72/0.73
SVR–WT–PLS/KELM–WT–PLS	MAE (ppb)	7.20/6.37	10.39/8.55	14.12/13.37	8.75/7.71
	MAPE	0.52/0.47	0.21/0.18	0.21/0.20	0.41/0.37
	RMSE (ppb)	9.33/8.04	12.54/10.05	17.12/15.70	11.31/9.75
	NRMSE (%)	24.74/21.32	62.87/50.39	70.15/64.31	13.72/11.83
	R^2	0.58/0.62	0.20/0.24	0.01/0.03	0.73/0.78
Number of samples		350	115	66	531

Whether WT and PLS are used or not, the KELM exhibits superior performance over the SVR in the prediction process. Compared with SVR–WT–PLS, the MAE, MAPE, RMSE and NRMSE of KELM–WT–PLS are 11.89%, 9.76%, 13.79%, 13.78% lower respectively, and R^2 is 6.85% higher. KELM–WT–PLS shows certain advantages in the prediction of O_3 hourly concentration.

4. Conclusions

Based on the KELM and SVR methods, the summer hourly ozone concentration is predicted and improved by WT and PLS methods. The KELM and SVR methods were selected as the optimization objects because of the relatively higher prediction accuracy when compared with the result of linear regression and the BPNN method. The WT method was used to decompose the high variability time series into several sub-series with lower variability, and then prediction strategy was applied to each sub-series of different scales. Meanwhile the prediction accuracy of SVR and KELM is improved by 16.99%~30.91% and 16.00%~25.86%, respectively. In addition, the application of VIP value based on PLS method for variable selection keep more important information for prediction and shorten the

model running time. Temperature and relative humidity are input parameters for prediction at most levels. The main influencing factors of the low-level detail series are precursors, while the approximation series is affected by meteorological conditions and the accumulated O_3. The WT and PLS methods improve the predictive performance of both KELM and SVR significantly. In particular, the prediction accuracy of high ozone concentration which is the focus of air pollution forecasting increase by 21% and 35% respectively. Overall, the KELM has the better prediction ability than SVR. The MAE, MAPE, RMSE, NRMSE and R^2 of KELM-WT-PLS method are 7.71 ppb, 0.37, 9.75 ppb, 11.83% and 0.78, respectively, implying the advantage of predicting the summer O_3 hourly concentration in Nanjing.

Acknowledgements

This work was supported by grants from the National Key Research and Development Program of China (Nos. 2016YFA0602003, 2017YFC0210003), the National Natural Science Foundation of China (No. 91544229), and the Qing Lan Project. The authors also thank the reviewers for their constructive and valuable comments.

References

An, J.L., Wang, Y.S., 2010. Principal component and multiple regression analysis predicting ozone concentrations: case study in summer in Beijing. Acta Scientiae Circumstantiae 30, 1286–1294.(in Chinese)

An, J.L., Zhu, B., Wang, H.L., Li, Y.Y., Lin, X., Yang, H., 2014. Characteristics and source apportionment of VOCs measured in an industrial area of Nanjing, Yangtze River Delta, China. Atmospheric Environment 97, 206–214.

An, J.L., Zou, J.N., Wang, J.X., Lin, X., Zhu, B., 2015. Differences in ozone photochemical characteristics between the megacity Nanjing and its suburban surroundings, Yangtze River Delta, China. Environmental Science & Pollution Research 22, 19607–19617.

Bai, Y., Li, Y., Wang, X.X., Xie, J.J., Li, C., 2016. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. Atmospheric Pollution Research 7,

557–566.

Borrego, C., Tchepel, O., Costa, A.M., Amorim, J.H., Miranda, A.I., 2003. Emission and dispersion modelling of Lisbon air quality at local scale. Atmospheric Environment 37, 5197–5205.

Canu, S., Rakotomamonjy, A., 2001. Ozone peak and pollution forecasting using support vectors, IFAC Workshop on Environmental Modeling, Yokohama, pp. 22–23.

Chelani, A.B., 2010. Prediction of daily maximum ground ozone concentration using support vector machine. Environmental Monitoring & Assessment 162, 169–176.

Chen, Y.L., Zhao, Z.J., 2013. A multi–step–ahead prediction of ozone concentration using wavelet transform and traditional time series model. Acta Scientiae Circumstantiae 33, 339–345. (in Chinese)

Ding, S., Chen, B.Z., Wang, J., Chen, L., Zhang, C.L., Sun, S.B., Huang, C.W., 2018. An applied research of decision–tree based statistical model in forecasting the spatial–temporal distribution of O_3. Acta Scientiae Circumstantiae 38, 3229–3242. (in Chinese)

Duncan, B.N., Lamsal, L.N., Thompson, A.M., Yoshida, Y., Lu, Z.F., Streets, D.G., Hurwitz, M.M., Pickering, K.E., 2016. A space–based, high–resolution view of notable changes in urban NO_x pollution around the world (2005–2014). Journal of Geophysical Research Atmospheres 121, 976–996.

Dunea, D., Pohoata, A., Iordache, S., 2015. Using wavelet–feedforward neural networks to improve air pollution forecasting in urban environments. Environmental Monitoring & Assessment 187, 477.

Farajzadeh, J., Alizadeh, F., 2017. A hybrid linear–nonlinear approach to predict the monthly rainfall over the Urmia Lake watershed using wavelet–SARIMAX–LSSVM conjugated model. Journal of Hydroinformatics 20, 246–262.

Feng, R., Zheng, H.J., Zhang, A.R., Huang, C., Gao, H., Ma, Y.C., 2019. Unveiling tropospheric ozone by the traditional atmospheric model and machine learning, and their comparison:a case study in hangzhou, China. Environmental Pollution 252, 366–378.

Feng, Y., Zhang, W.F., Sun, D.Z., Zhang, L.Q., 2011. Ozone concentration forecast method based on genetic algorithm

optimized back propagation neural networks and support vector machine data classification. Atmospheric Environment 45, 1979–1985.

Feng, Z.Z., Hu, E.Z., Wang, X.K., Jiang, L.J., Liu, X.J., 2015. Ground–level O_3 pollution and its impacts on food crops in China: a review. Environmental Pollution 199, 42–48.

Gao, M., Yin, L.T., Ning, J.C., 2018. Artificial neural network model for ozone concentration estimation and Monte Carlo analysis. Atmospheric Environment 184, 129–139.

Goodman, J.E., Prueitt, R.L., Sax, S.N., Pizzurro, D.M., Lynch, H.N., Zu, K., Venditti, F.J., 2015. Ozone exposure and systemic biomarkers: evaluation of evidence for adverse cardiovascular health impacts. Critical Reviews in Toxicology 45, 412–452.

Hamer, P.D., Bowman, K.W., Henze, D.K., Attié, J.L., Marécal, V., 2015. The impact of observing characteristics on the ability to predict ozone under varying polluted photochemical regimes. Atmospheric Chemistry and Physics 15, 4909–4971.

He, J.J., Gong, S.L., Yu, Y., Yu, L.J., Wu, L., Mao, H.J., Song, C.B., Zhao, S.P., Liu, H.L., Li, X.Y., 2017. Air pollution characteristics and their relation to meteorological conditions during 2014–2015 in major Chinese cities. Environmental Pollution 223, 484–496.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12, 55–67.

Huang, G.B., Chen, L., 2007. Convex incremental extreme learning machine. Neurocomputing 70, 3056–3062.

Huang, G.B., Wang, D.H., Lan, Y., 2011. Extreme learning machines: a survey. International Journal of Machine Learning & Cybernetics 2, 107–122.

Huang, G.B., Zhou, H., Ding, X., Zhang, R., 2012. Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems Man & Cybernetics Part B 42, 513–529.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing 70,

489–501.

Karlsson, P.E., Klingberg, J., Engardt, M., Andersson, C., Langner, J., Karlsson, G.P., Pleijel, H., 2017. Past, present and future concentrations of ground–level ozone and potential impacts on ecosystems and human health in northern Europe. Science of the Total Environment 576, 22–35.

Krotkov, N.A., Mclinden, C.A., Li, C., Lamsal, L.N., Celarier, E.A., Marchenko, S.V., Swartz, W.H., Bucsela, E.J., Joiner, J., Duncan, B.N., Boersma, K.F., Veefkind, J.P., Levelt, P.F., Fioletov, V.E., Dickerson, R.R., He, H., Lu, Z.F., Streets, D.G., 2016. Aura OMI observations of regional SO_2 and NO_2 pollution changes from 2005 to 2015. Atmospheric Chemistry & Physics 16, 4605–4629.

Lee, N.U., Shim, J.S., Ju, Y.W., Park, S.C., 2018. Design and implementation of the SARIMA–SVM time series analysis algorithm for the improvement of atmospheric environment forecast accuracy. Soft Computing 22, 4275–4281.

Li, Y., Tao, Y., 2018. Daily PM_{10} concentration forecasting based on multiscale fusion support vector regression. Journal of Intelligent & Fuzzy Systems 34, 3833–3844.

Lin, X., Zhu, B., An, J.L., 2015. Potential contribution of secondary organic aerosols and ozone of VOCs in the northern suburb of Nanjing. China Environmental Science 35, 976–986. (in Chinese)

Lin, Y., Yu, J., Cai, J.Q., Sneeuw, N., Li, F.T., 2018. Spatio–temporal analysis of wetland changes using a kernel extreme learning machine approach. Remote Sensing 10, 1129.

Liu, H., Mi, X.W., Li, Y.F., 2018. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. Energy Conversion and Management 156, 498–514.

Liu, J., Wu, D., Fan, S.J., Liao, Z.H., Deng, T., 2017a. Impacts of precursors and meteorological factors on ozone pollution in Pearl River Delta. China Environmental Science 37, 813–820. (in Chinese)

Liu, S.J., Hu, Y.W., Chao, L., Lu, H.T., Zhang, H.C., 2017b. Machinery condition prediction based on wavelet and

support vector machine. Journal of Intelligent Manufacturing 28, 1045–1055.

Lu, Q., Zheng, J.Y., Ye, S.Q., Shen, X.L., Yuan, Z.B., Yin, S.S., 2013. Emission trends and source characteristics of SO_2, NO_x, PM_{10} and VOCs in the Pearl River Delta region from 2000 to 2009. Atmospheric Environment 76, 11–20.

Lu, W.Z., Wang, W.J., 2005. Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends. Chemosphere 59, 693–701.

Luna, A.S., Paredes, M.L.L., de Oliveira, G.C.G., Corrêa, S.M., 2014. Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil. Atmospheric Environment 98, 98–104.

Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 11, 674–693.

Mao, X., Shen, T., Feng, X., 2017. Prediction of hourly ground–level $PM_{2.5}$ concentrations 3 days in advance using neural networks with satellite data in eastern China. Atmospheric Pollution Research 8, 1005–1015.

Mehdipour, V., Memarianfard, M., 2019. Ground–level O_3 sensitivity analysis using support vector machine with radial basis function. International Journal of Environmental Science and Technology 16, 2745–2754.

Nieto, P.J.G., Combarro, E.F., Del Coz Díaz, J.J., Montañés, E., 2013. A SVM–based regression model to study the air quality at local scale in Oviedo urban area (Northern Spain): a case study. Applied Mathematics & Computation 219, 8923–8937.

Nieto, P.J.G., García–Gonzalo, E., Sánchez, A.B., Miranda, A.A.R., 2017. Air quality modeling using the PSO–SVM–based approach, MLP neural network, and M5 model tree in the metropolitan area of Oviedo (Northern Spain). Environmental Modeling & Assessment 23, 229–247.

Nieto, P.J.G., Lasheras, F.S., García–Gonzalo, E., Juez, F.J.D.C., 2018. PM_{10} concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM, MLP, VARMA and ARIMA: a case study. Science of the Total Environment 621, 753–761.

Ortiz–García, E.G., Salcedo-Sanz, S., Pérez-Bellido, á.M., Portilla-Figueras, J.A., Prieto, L., 2010. Prediction of hourly O_3 concentrations using support vector regression algorithms. Atmospheric Environment 44, 4481–4488.

Peng, H.P., Lima, A.R., Teakles, A., Jin, J., Cannon, A.J., Hsieh, W.W., 2017. Evaluating hourly air quality forecasting in Canada with nonlinear updatable machine learning methods. Air Quality Atmosphere & Health 10, 195–211.

Quej, V.H., Almorox, J., Arnaldo, J.A., Saito, L., 2017. ANFIS, SVM and ANN soft–computing techniques to estimate daily global solar radiation in a warm sub–humid environment. Journal of Atmospheric and Solar–Terrestrial Physics 155, 62–70.

Salazar–Ruiz, E., Ordieres, J.B., Vergara, E.P., Capuz–Rizo, S.F., 2008. Development and comparative analysis of tropospheric ozone prediction models using linear and artificial intelligence–based models in Mexicali, Baja California (Mexico) and Calexico, California (US). Environmental Modelling & Software 23, 1056–1069.

Shen, H.Y., Wang, Z.X., Gao, C.Y., Qin, J., Yao, F.B., Xu, W., 2008. Determining the number of BP neural network hidden layer units. Journal of Tianjin University of Technology 24, 13–15.

Shi, Y., Li, P.J., Yuan, H., Miao, J.Y., Niu, L.F., 2019. Fast kernel extreme learning machine for ordinal regression. Knowledge–Based Systems 177, 44–54.

Siwek, K., Osowski, S., 2012. Improving the accuracy of prediction of PM_{10} pollution by the wavelet transformation and an ensemble of neural predictors. Engineering Applications of Artificial Intelligence 25, 1246–1258.

Smedt, I.D., Stavrakou, T., Hendrick, F., Danckaert, T., Vlemmix, T., Pinardi, G., Theys, N., Lerot, C., Gielen, C., Vigouroux, C., Hermans, C., Fayt, C., Veefkind, P., Müller, J.F., Roozendael, M.V., 2015. Diurnal, seasonal and long–term variations of global formaldehyde columns inferred from combined OMI and GOME–2 observations. Atmospheric Chemistry and Physics 15, 12241–12300.

Vapnik, V.N., 2000. The nature of statistical learning theory. Springer, New York.

Wang, X.K., Zhang, Q.Q., Zheng, F.X., Zheng, Q.W., Yao, F.F., Chen, Z., Zhang, W.W., Hou, P.Q., Feng, Z.Z., Song, W.Z., 2012. Effects of elevated O_3 concentration on winter wheat and rice yields in the Yangtze River Delta, China.

Environmental Pollution 171, 118–125.

Wold, H., 1966. Estimation of principal components and related models by iterative least squares. Multivariate Analysis 1, 391–420.

Wold, S., 1995. PLS for multivariate linear modeling In chemometric methods in molecular design. Verlag–Chemie, Weinheim, Germany.

Wu, B., Hu, B.H., Wang, X.Z., Huang, H., Wang, J., 2017. Visibility forecast based on proximal support vector machine. Journal of Tropical Meteorology 23, 104–110. (in Chinese)

Xiao, F., Qi, L., Zhu, Y.J., Hou, J.X., Jin, L.Y., Wang, J.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.

Xu, Y.Z., Yang, W.D., Wang, J.Z., 2016. Air quality early–warning system for cities in China. Atmospheric Environment 148, 239–257.

Yang, X.X., Tang, L.L., Zhang, Y.J., Mu, Y.F., Wang, M., Chen, W.T., Zhou, H.C., Hua, Y., Jiang, R.X., 2016. Correlation analysis between characteristics of VOCs and ozone formation potential in summer in Nanjing urban district. Environmental Science 37, 443–451. (in Chinese)

Yao, W., Zeng, Z.G., Lian, C., Tang, H.M., 2014. A kernel ELM classifier for high–resolution remotely sensed imagery based on multiple features, Advances in Neural Networks–ISNN 2014. Springer,Cham, pp. 270–277.

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, 357–365.

Zhai, L., Li, S., Zou, B., Sang, H.Y., Fang, X., Xu, S., 2018. An improved geographically weighted regression model for $PM_{2.5}$ concentration estimation in large areas. Atmospheric Environment 181, 145–154.

Zhang, J.S., Fu, D.W., 2017. Prediction of air pollutants concentration based on an extreme learning machine: the case of Hong Kong. International Journal of Environmental Research & Public Health 14, 114.

Zhang, Y.X., An, J.L., Wang, J.X., Shi, Y.Z., Liu, J.D., Liang, J.S., 2018. Source analysis of volatile organic compounds in the Nanjing industrial area and evaluation of their contribution to ozone. Environmental Science 39, 502–510. (in Chinese)

Zhao, W.J., Zhao, Z.J., Wang, J.H., Shang, K.Z., Wang, S.G., Liu, Z.H., Kong, D.B., Su, J.L., 2016. A study on refined forecast of cloud cover based on support vector machine. Journal of Arid Meteorology 34, 568–574. (in Chinese)

Supplement

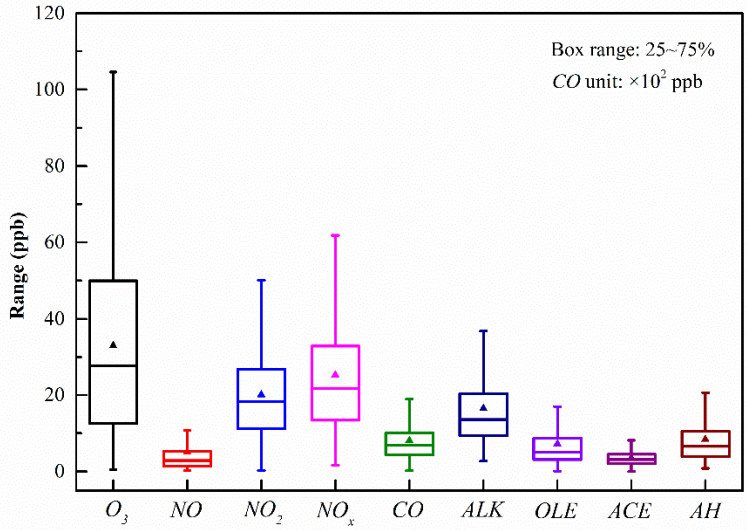


Fig. S1. Distribution characteristics of pollutants data.

Figures

Fig. 1. Location of the observation site and the surrounding environment.

Fig. 2. Wind speed and wind direction during the observation period from 2014 to 2016.

Fig. 3. Schematic diagram of *n*-level wavelet decomposition.

Fig. 4. The wavelet decomposition of the original time series *s* of O_3 concentrations: D_1~D_5 denote the wavelet coefficients at different levels and A_5 is the approximated signal of *s* on the fifth level.

Fig. 5. Comparison of prediction effect of the SVR, KELM, BPNN and SR.

Fig. 6. Predictions of the SVR–WT, KELM–WT, SVR–WT–PLS and KELM–WT–PLS.

Tables

Supplement

Table 1. Correlation analysis of all data used in prediction of O_3.

	O_3	NO	NO_2	NO_x	CO	ALK	OLE	ACE	AH	WS	T	RH	WD	P
O_3	1	**-0.50**	-0.36	**-0.45**	-0.31	-0.38	**-0.41**	-0.20	**-0.44**	0.35	**0.58**	**-0.74**	0.07	-0.12
NO	-0.50	1	0.52	0.67	0.32	0.37	0.47	0.16	0.33	-0.06	-0.26	0.32	-0.15	0.09
NO_2	-0.36	0.52	1	0.97	0.51	0.62	0.49	0.52	0.58	-0.20	-0.23	0.12	-0.32	0.16
NO_x	-0.45	0.67	0.97	1	0.53	0.63	0.53	0.50	0.59	-0.20	-0.27	0.21	-0.31	0.15
CO	-0.31	0.32	0.51	0.53	1	0.53	0.38	0.32	0.54	-0.10	–	0.26	-0.35	0.05
ALK	-0.38	0.37	0.62	0.63	0.53	1	0.65	0.56	0.77	-0.28	-0.19	0.29	-0.40	0.05
OLE	-0.41	0.47	0.49	0.53	0.38	0.65	1	0.25	0.51	-0.24	-0.15	0.33	-0.27	0.07
ACE	-0.20	0.16	0.52	0.50	0.32	0.56	0.25	1	0.56	-0.28	-0.15	0.17	-0.10	–
AH	-0.44	0.33	0.58	0.59	0.54	0.77	0.51	0.56	1	-0.24	-0.22	0.36	-0.39	0.07
WS	0.35	-0.06	-0.20	-0.20	-0.10	-0.28	-0.24	-0.28	-0.24	1	0.21	-0.51	-0.15	0.05
T	0.58	-0.26	-0.23	-0.27	–	-0.19	-0.15	-0.15	-0.22	0.21	1	-0.45	–	-0.47
RH	-0.74	0.32	0.12	0.21	0.26	0.29	0.33	0.17	0.36	-0.51	-0.45	1	–	–
WD	0.07	-0.15	-0.32	-0.31	-0.35	-0.40	-0.27	-0.10	-0.39	-0.15	–	–	1	–
P	-0.12	0.09	0.16	0.15	0.05	0.05	0.07	–	0.07	0.05	-0.47	–	–	1

Table 2. Statistics of measured values. Unit, minimum, maximum, mean, and standard deviation values during observation period.

Variable	Unit	Min	Max	Mean	Rstd (%)
O_3	ppb	0.50	116.20	32.94	74.86
NO	ppb	0.25	102.87	5.16	146.90
NO_2	ppb	0.25	81.00	20.11	57.28
NO_x	ppb	1.60	129.45	25.27	64.62
CO	ppb	20.00	4288.46	807.05	68.46
Alkane	ppb	2.76	73.98	16.45	61.28
Olefin	ppb	0.14	97.07	7.18	94.01
Acetylene	ppb	0.02	17.59	3.52	61.08
Aromatic hydrocarbon	ppb	0.80	71.85	8.39	79.14
Wind speed	$m \cdot s^{-1}$	0.00	6.40	1.82	50.55
Wind direction	°	0.00	359.00	165.91	52.08
Temperature	°C	9.80	40.30	25.96	16.87
Relative humidity	%	13.00	100.00	64.39	29.77
Atmospheric pressure	hPa	995.70	1021.40	1004.70	0.37

Table 3. Value of Smooth from level 1 to level 7.

Level	1	2	3	4	5	6	7

Smooth	0.6481	0.4104	0.2501	0.0209	0.0021	0.0004	0.0001

Table 4. Comparison of five evaluation indexes among SR, BPNN, SVR and KELM.

	MAE (ppb)	MAPE	RMSE (ppb)	NRMSE (%)	R^2
SR	15.58	0.89	18.61	22.57	0.54
BPNN	12.98	0.51	16.90	20.50	0.53
SVR	12.95	0.53	16.82	20.40	0.55
KELM	10.50	0.60	12.97	15.73	0.58

Table 5. Input data used in prediction of six O_3 scales respectively and effect of using PLS to select variables.

Scales	WT+PLS		WT +all variables
	Input variable	R^2 SVR/KELM	R^2 SVR/KELM
D_1	*NO*, NO_2, NO_x, *CO*, *T*	0.63/0.63	0.55/0.60
D_2	*NO*, NO_2, NO_x, *CO*, *T*, *RH*	0.60/0.58	0.59/0.57
D_3	NO_2, NO_x, *AH*, *T*, *RH*	0.71/0.79	0.71/0.81
D_4	NO_2, NO_x, *ALK*, *ACE*, *AH*, *WS*, *T*, *RH*	0.87/0.92	0.87/0.91
D_5	*NO*, NO_x, *AH*, *T*, *RH*	0.58/0.61	0.55/0.59
A_5	*ACE*, *T*, *RH*, O_3_*6*	0.68/0.79	0.62/0.68

Table 6. Performance measures of six models at different ozone concentration levels.

	Estimator	Classification Intervals (ppb)			
		$[O_3]$ <40	40≤ $[O_3]$ <60	$[O_3]$ ≥60	all test database
SVR/KELM	MAE (ppb)	9.61/9.32	18.02/10.23	21.78/17.27	12.95/10.50
	MAPE	0.62/0.80	0.37/0.21	0.32/0.25	0.53/0.60
	RMSE (ppb)	11.97/11.62	21.68/12.13	26.43/19.53	16.82/12.97
	NRMSE (%)	31.74/30.81	108.67/60.81	108.26/80.01	20.40/15.73
	R^2	0.39/0.34	0.15/0.19	0.01/0.00	0.55/0.58
SVR–WT/KELM–WT	MAE (ppb)	9.14/7.72	11.31/8.87	18.28/14.53	10.75/8.82
	MAPE	0.70/0.61	0.23/0.18	0.27/0.21	0.54/0.47
	RMSE (ppb)	11.43/9.29	13.75/10.48	20.87/16.94	13.46/10.79
	NRMSE (%)	30.32/24.64	68.92/52.53	85.51/69.38	16.33/13.08

SVR–WT–PLS/KELM–WT–PLS	R^2	0.54/0.51	0.20/0.23	0.00/0.03	0.72/0.73
	MAE (ppb)	7.20/6.37	10.39/8.55	14.12/13.37	8.75/7.71
	MAPE	0.52/0.47	0.21/0.18	0.21/0.20	0.41/0.37
	RMSE (ppb)	9.33/8.04	12.54/10.05	17.12/15.70	11.31/9.75
	NRMSE (%)	24.74/21.32	62.87/50.39	70.15/64.31	13.72/11.83
	R^2	0.58/0.62	0.20/0.24	0.01/0.03	0.73/0.78
Number of samples		350	115	66	531

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

None