ELSEVIER

Chemosphere 59 (2005) 693–701

CHEMOSPHERE

www.elsevier.com/locate/chemosphere

Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends

Wei-Zhen Lu [a,*], Wen-Jian Wang [b]

[a] Department of Building and Construction, City University of Hong Kong, 83, Tat Chee Avenue, Kowloon Tong, Hong Kong
[b] Department of Computer Science, Shanxi University, Taiyuan, Shanxi 030006, PR China

Received 26 January 2004; received in revised form 29 September 2004; accepted 12 October 2004

Abstract

Monitoring and forecasting of air quality parameters are popular and important topics of atmospheric and environmental research today due to the health impact caused by exposing to air pollutants existing in urban air. The accurate models for air pollutant prediction are needed because such models would allow forecasting and diagnosing potential compliance or non-compliance in both short- and long-term aspects. Artificial neural networks (ANN) are regarded as reliable and cost-effective method to achieve such tasks and have produced some promising results to date. Although ANN has addressed more attentions to environmental researchers, its inherent drawbacks, e.g., local minima, over-fitting training, poor generalization performance, determination of the appropriate network architecture, etc., impede the practical application of ANN. Support vector machine (SVM), a novel type of learning machine based on statistical learning theory, can be used for regression and time series prediction and have been reported to perform well by some promising results. The work presented in this paper aims to examine the feasibility of applying SVM to predict air pollutant levels in advancing time series based on the monitored air pollutant database in Hong Kong downtown area. At the same time, the functional characteristics of SVM are investigated in the study. The experimental comparisons between the SVM model and the classical radial basis function (RBF) network demonstrate that the SVM is superior to the conventional RBF network in predicting air quality parameters with different time series and of better generalization performance than the RBF model.

Keywords: Air pollution; Generalization performance; Neural networks; Radial basis function; Support vector machine; Time series

1. Introduction

With continuous economy development and population increase in Hong Kong, severe problems relating to environmental pollution have attracted much attention than ever before, e.g., air pollution, noise pollution, shortage of land resources, waste and sewage disposal, etc. Among these, air pollution has direct impact on humans by exposure to high pollutant levels existing in the atmosphere. Air pollution monitoring and prediction are needed for preventing the situation from worsening in the long run. Besides, short-term forecasting of air quality is also required in order to take preventive and evasive action during the episodes of airborne pollution.

* Corresponding author. Tel.: +852 27844316; fax: +852 27887612.
E-mail address: bcwzlu@cityu.edu.hk (W.-Z. Lu).

0045-6535/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chemosphere.2004.10.032

Monitoring and forecasting of air pollutant trends in ambient air involve using a variety of approaches, e.g., on-site measurement, computational fluid dynamics (CFD) simulation, computational intelligence, etc. Artificial neural network (ANN) methods are regarded as cost-effective method to achieve the prediction of air pollutants in time series and became popular in recent years. The ANN model, in particular, the multi-layer feed-forward neural network perceptron, can be trained to approximate virtually any smooth, measurable function, and has produced certain promising results to date (Boznar et al., 1993; Song and Hopke, 1996; Yi and Prybutok, 1996; Comrie, 1997; Roadknight et al., 1997; Gardner and Dorling, 1998; Lee et al., 1999; Reich et al., 1999; Fan et al., 2000, 2003; Perez et al., 2000; Lu et al., 2002a,b,c, 2003). These studies have shown that the ANN approach is effective in simulating and describing the dynamics of non-stationary time series due to its unique non-parametric, non-assumable, noise-tolerant and high-adaptive properties. ANN models are universal function approximators and can be used to map any non-linear function without prior assumptions on the original data.

Although some recognised ANN models were developed, e.g., Back-propagation (BP) algorithm, radial basis function (RBF) network, principal component analysis/radial basis function (PCA/RBF) network, etc., the inherent drawbacks, i.e., over-fitting training, local minima, difficult determination of network architecture, and poor generalizing performance, remain unsolved and impede the application of the ANN approach into practice. The support vector machine (SVM) method, developed by Vapnik (1995), can provide an effective novel approach to improve generalization performance of neural networks and achieve global solutions simultaneously. Originally, the SVM model was developed for pattern recognition problems. Recently, with the introduction of ε-insensitive loss function, SVM has been extended to solve non-linear regression estimation and time series prediction (Mukherjee et al., 1997; Müller et al., 1997; Vapnik et al., 1997; Broomhead and Lowe, 1998). Unlike traditional learning machines, which normally adopt the *Empirical Risk Minimization Principle* (ERMP) like feed-forward neural networks, SVM implements *Structural Risk Minimization Principle* (SRMP), which seeks to minimize an upper bound of generalization error rather than minimize training error. This process leads to better generalization than conventional methods. This paper presents a pioneer study of using the SVM model to investigate potential variations of air pollutants, which were measured at Causeway Bay Roadside Monitoring Station during 1999, one of fourteen air pollutant monitoring stations in Hong Kong territory established by Hong Kong Environment Protection Department (HKEPD). The SVM was firstly trained by selected data sets from the original database, and then used to predict the pollutant levels in different time series by selected test sets. The performances of SVM were evaluated by comparing with the results produced by conventional RBF network. The variability performance regarding to the free parameters of SVM was also examined.

2. Mathematical background and original database

2.1. Theory of support vector machine

In support vector machine (SVM) analysis, the basic idea is to map original data x into a feature space F with higher dimensionality via a non-linear mapping function ϕ, which is usually unknown, and then carry on linear regression in the feature space (Vapnik, 1995; Vapnik et al., 1997). Hence, the regression approximation addresses a problem of estimating a function based on a given data set $G = \{(x_i, d_i)\}_i^l$ (x_i is input vectors, d_i is desired values), which is produced from ϕ. SVM method approximates the function in the following form:

$$y = \sum_{i=1}^{l} w_i \phi_i(x) + b \tag{1}$$

where $\{\phi_i(x)\}_{i=1}^l$ represent the features of inputs, $\{w_i\}_{i=1}^l$ and b are coefficients. These are estimated by minimizing the regularized risk function listed below:

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

where

$$L_\varepsilon(d, y) = \begin{cases} |d - y| - \varepsilon & |d - y| \geqslant \varepsilon \\ 0 & \text{others} \end{cases} \tag{3}$$

Here, ε is a prescribed parameter. The $L_\varepsilon(d, y)$ is ε-insensitive loss function, which does not penalize errors less than ε. The term $\frac{1}{2}\|w\|^2$ is used as a measurement of function flatness. C is a regularized constant determining the trade-off between the training error and the model flatness. Introducing slack variables ζ, ζ^* would lead equation (2) to the following constrained formation:

$$\text{Minimize:} \quad R(w, \zeta^*) = \frac{1}{2}\|w\|^2 + C^* \sum_{i=1}^{n}(\zeta_i + \zeta_i^*) \tag{4}$$

$$\text{Subjected to:} \quad \begin{aligned} & w\phi(x_i) + b - d_i \leqslant \varepsilon + \zeta_i \\ & d_i - w\phi(x_i) - b_i \leqslant \varepsilon + \zeta_i \\ & \zeta, \zeta^* \geqslant 0 \end{aligned} \tag{5}$$

Thus, Eq. (1) becomes the following explicit form:

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(x, x_i) + b \tag{6}$$

In Eq. (6), α_i and α_i^* are Lagrange multipliers, which satisfy the equality of $\alpha_i * \alpha_i^* = 0$, $\alpha_i \geqslant 0$, $\alpha_i^* \geqslant 0$,

$i = 1, \ldots, l$, and can be obtained by maximizing the dual form of Eq. (4):

$$\phi(\alpha, \alpha^*) = \sum_{i=1}^{l} d_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^{l}(\alpha_i + \alpha_i^*) - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\alpha_i, \alpha_j) \quad (7)$$

$$\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0$$

with constrains: (8)

$$0 \leqslant \alpha_i \leqslant C \quad i = 1, \ldots, l$$

$$0 \leqslant \alpha_i^* \leqslant C \quad i = 1, \ldots, l$$

According to the nature of quadratic programming, only those data corresponding to non-zero (α_i, α_i^*) pairs can be referred to support vectors. Here, $K(x_i, x_j)$ is called kernel function and obtained by $K(x_i, x_j) = \phi(x_i) * \phi(x_j)$ in the feature space. Thus, all computations related to ϕ will be carried on by kernel function in input space (Vapnik, 1995; Vapnik et al., 1997). The common Gaussian kernel function $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2/(2\sigma^2))$ is used in the study ($\sigma = 1.0$). Both C and ε are specified with 100 and 10^{-3} respectively.

2.2. Air quality in Hong Kong Urban Air

Hong Kong possesses the highest vehicle density in the world, i.e., approximately 271 vehicles/km [Transport Department, 2001]. Hence, vehicle exhaust is a major air pollutant source in Hong Kong. Owing to the diesel running on the majority of vehicles in Hong Kong, large amount of pollutants, e.g., respirable suspended particulate (RSP), nitrogen dioxides (NO_2), nitrogen oxides (NO_x), etc., are emitted to urban atmosphere from vehicle exhaust, particularly in congested downtown areas (Chan and Kwok, 2000; Chan and Liu, 2001; Wang et al., 2001). According to HKEPD annual reports, RSP is reckoned as the most severe pollutant (HKEPD Annual Report, 1998, 1999, 2000). Previous studies have shown that high RSP level may cause chronic and acute effects on human health, particularly the pulmonary function and the respiratory problems (Harrison et al., 1997; Spurny, 1998; Perez et al., 2000). Such effects are deteriorated if RSP pollution is associated with high level of other pollutants, e.g., NO_2, NO_x, etc. (Burnett et al., 1999). The RSP containing small sizes (e.g., less than 5 μm) may also influence visibility and images of historical, cultural treasures via surface deposition. The records of air quality monitoring database show quite a number of times that RSP, NO_2, and NO_x levels were reported exceeding the corresponding Air Quality Objectives (AQOs), stipulated by HKEPD, in certain congested downtown areas in Hong Kong during the past five years. The situations are getting worse as the road vehicles increase continuously. Long-term exposures to RSP, NO_2, and NO_x can lower human resistance to respiratory infections and aggravate existing chronic respiratory diseases at the same time. Therefore, these three pollutants, i.e., RSP, NO_x and NO_2, are selected as outputs of the proposed SVM model to be tested.

2.3. Original data set

The available air quality database measured at Causeway Bay roadside monitoring station in 1999 was selected as the original data set. The database includes seven major air pollutants, i.e., carbon monoxide (CO), nitric oxide (NO), NO_2, sulphur dioxide (SO_2), NO_x, ozone (O_3), and RSP, and five meteorological parameters, i.e., indoor and outdoor temperature (IT and OT), solar radiation (SD), wind direction (WD) and wind speed (WS), which were hourly measured at the said location.

In prediction experiments, the recorded levels of RSP, NO_x and NO_2 in June and December are selected as original samples. The reason to choose the data in these two months is because December and June represent two different seasons in Hong Kong, i.e., December corresponds to dry, cold weather and is normally accompanied by the prevailing north-eastern wind and the highest pollutant levels (i.e., local vehicle exhausts combining with migration of industrial pollutants from Mainland China), while June corresponds to hot, wet season and often undergoes the dominant south-eastern wind and the lowest pollutant concentrations (i.e., local vehicle pollution dominates) (HKEPD Annual Report, 1998). Hence, the robustness of the SVM model can be verified by seasonal variation. In simulations, the data of the first ten days (240 data points) in each month are used as training sets. The trained SVM model is then used to predict the pollutant levels in different time series, i.e., one day and one week predictions for coming periods. Thus the simulation results have either 24 test points corresponding to the hourly measurements on the 11th day of selected month, or 168 test points representing the hourly data for the week of the 11th–17th day of each month.

3. Results and discussion

The simulation programs are constructed using Matlab with sequential minimal optimization (SMO) algorithm for solving time series prediction. A conventional, adaptive RBF network is used as a benchmark. The performance assessments of both models are carried out by comparing the simulation results produced by both models. In both models, the input variables include

six pollutants: SO_2, NO_x, NO, NO_2, CO and RSP, while Outputs are RSP, NO_x and NO_2 in corresponding simulations.

3.1. Recovery performances of SVM and RBF models

Taking the RSP levels as training examples, we compare the recovery performance between the SVM and the RBF models on the training set (i.e., 240 data of 1st–10th January 1999) and the testing set (i.e., 168 data of 11th–17th January 1999). Fig. 1 illustrates the recovery performances on the training data using both methods. For the SVM method, the computed results are almost identical to the original data and the maximum deviation is 65.7 μg/m^3. While, for the RBF network, more deviating points are observed than the SVM model does, and the maximum deviation is 69.2 μg/m^3. In general, both models show good recovery performance on the training data except some individual deviating points observed, more in RBF model than that in SVM model.

The comparison of prediction errors on the testing data between the SVM and the RBF models is shown in Fig. 2. It can be seen that the results of RBF network fluctuate for the first 60 data, while the ones produced by SVM method are very close to the original data points (Fig. 2b). Both models present good performance on simulating the data points between 60 and 150. However, the prediction errors of RBF network increase sharply at the later testing stages with the maximum error at 2601.3 μg/m^3 (Fig. 2a). While, for SVM method, the computing error keeps within a small range, the predictions still remain close to the original ones, and the maximum error is only 92.0 μg/m^3 (Fig. 2b). From Fig. 2, it can be observed that the SVM model has better generalization performance than the RBF network on the testing process, which is in consistency with statistical learning theory.

3.2. Predictions of pollutant levels in different time series

The robustness and tolerance of both SVM and RBF models are inspected and discussed under the impact of meteorological factors such as temperature, humid-

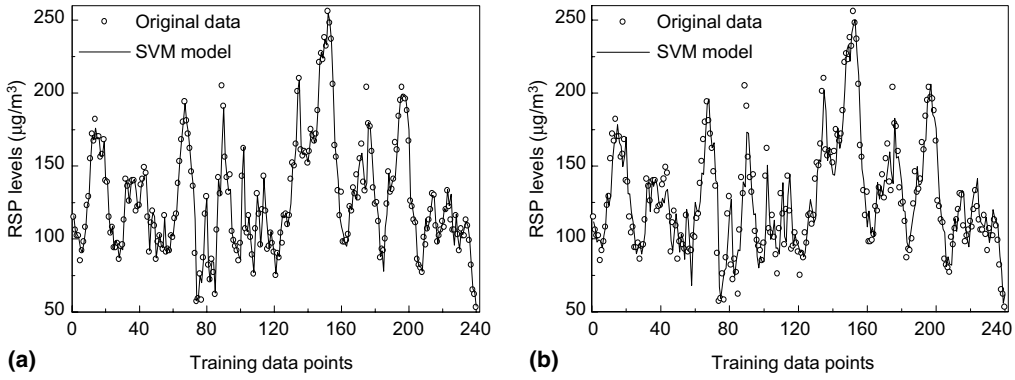

Fig. 1. Recovery performances of SVM and RBF models on training data: (a) SVM model and (b) RBF model.

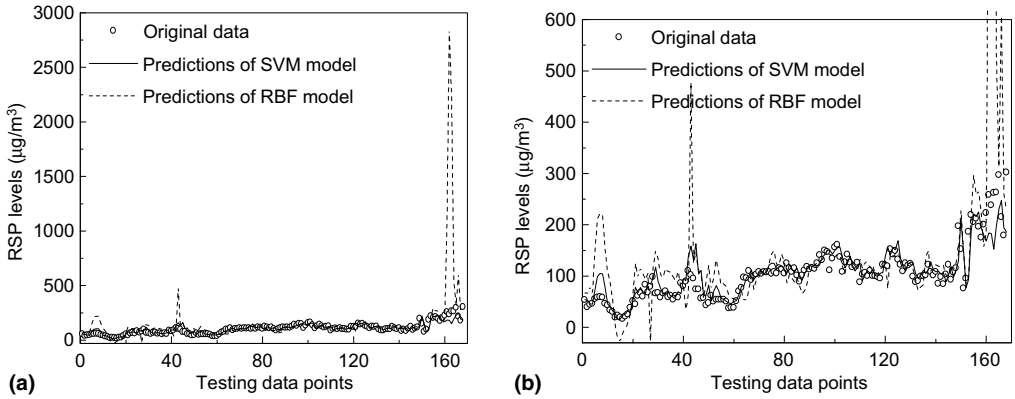

Fig. 2. Predictions by SVM and RBF models on testing data: (a) comparison in full scale and (b) comparison in resolved scale.

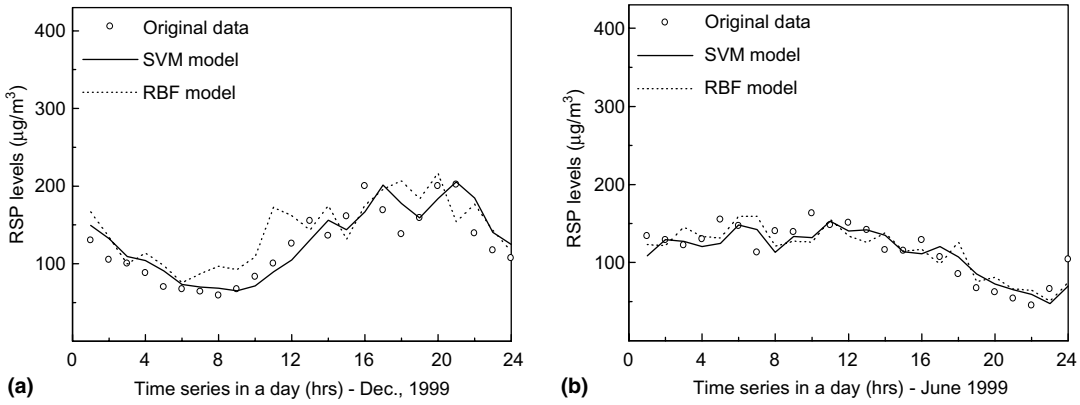

Fig. 3. Prediction comparisons between SVM and RBF models for 24-h period.

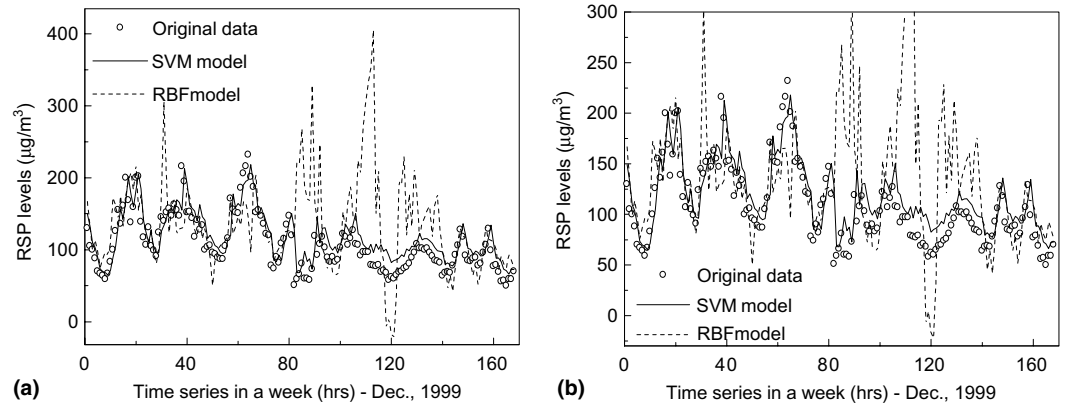

Fig. 4. Prediction comparisons between SVM and RBF models for one-week in December, 1999: (a) comparison in full scale and (b) comparison in resolved scale.

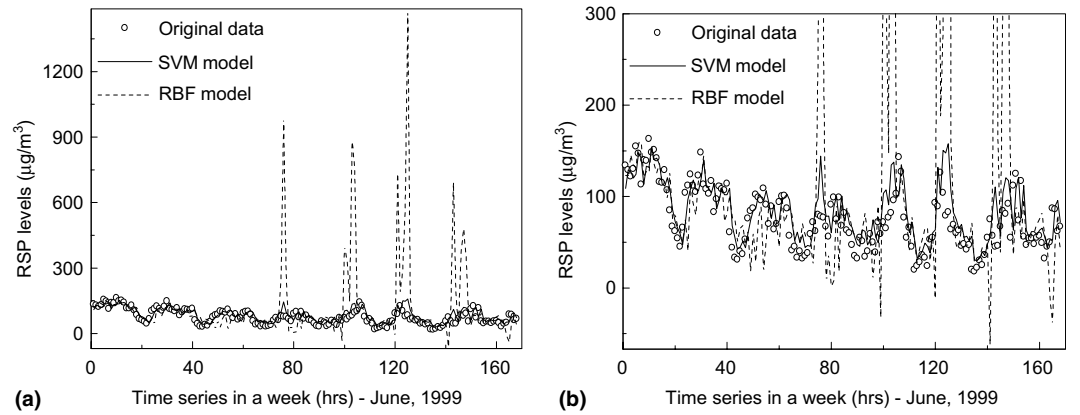

Fig. 5. Prediction comparisons between SVM and RBF models for one-week in June, 1999: (a) comparison in full scale and (b) comparison in resolved scale.

ity, wind speed and direction, and solar condition in different seasons. Figs. 3–5 compare the RSP levels predicted by both SVM and RBF networks with two time periods, i.e., 24-h and one-month advancing, in December and June of 1999. It can be seen that, for the 24-h cases, both models produce generally good results for the selected months, but the results produced by SVM method are slightly closer to the original data than those by RBF network (shown in Fig. 3a and b). While for one-week cases, the SVM method expresses great advantages over the RBF network (Figs. 4 and 5). The predictions produced by SVM are generally close to the original data in both months (Figs. 4a and 5a). The maximum absolute errors of the SVM model for selected months are 77.87 μg/m^3 (Fig. 4b) and 74.67 μg/m^3 (Fig. 5b) respectively. The results created by RBF fluctuate and, at certain points, deviate from the measured points in great range. The maximum absolute errors by RBF network for case of December is nearly 327.44 μg/m^3 (Fig. 4a), and even reaches 1389.16 μg/m^3 for case in June (Fig. 5a). The SVM method performs better than the conventional RBF one does. Hence, it can be concluded that, although the impact of meteorological variables exists, the SVM method still possesses superior advantages to the conventional RBF network and can produce good prediction performance due to its features of noise-tolerance, high stability, adaptive properties and better generalization performance.

Considering the characteristics of each pollutant, e.g., accumulation of RSP matter, physical and chemical complexity of NO_x and NO_2, etc., the generalization performance of the SVM model can be further verified by predicting the other two pollutants, i.e., NO_x and NO_2. Figs. 6 and 7 describe the predictions of hourly NO_x and NO_2 levels in one week advancing time series in June and December of 1999. It is noticed that the RBF network demonstrates very poor performance to predict one-week NO_x levels both in June and December (Fig. 6). At some points, the deviations (i.e., maximum absolute errors) between original data and predictions reach 80054.91 μg/m^3 in June (Fig. 6a) and 38758.5 μg/m^3 in December (Fig. 6c) respectively; While the SVM method presents satisfactory predicting performance

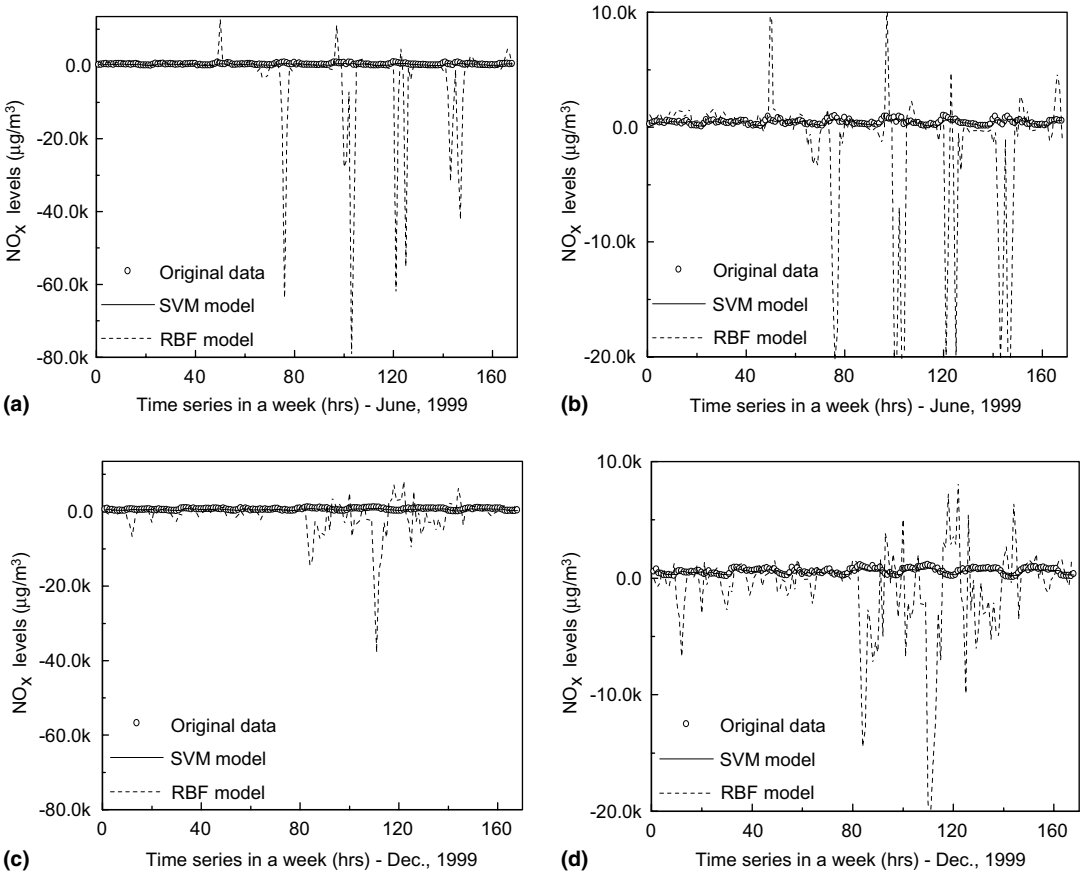

Fig. 6. Prediction comparisons between SVM and RBF models for NO_x levels in 1999: (a,c) comparison in full scale and (b,d) comparison in resolved scale.

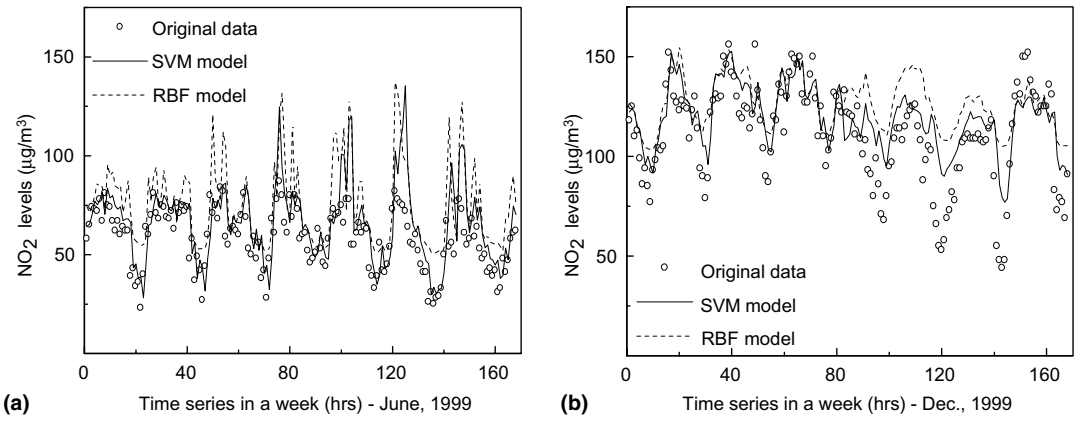

Fig. 7. Prediction comparisons between SVM and RBF models for NO_2 levels in 1999.

with the maximum absolute errors at 530.6888 μg/m^3 in June (Fig. 6b) and 415.99 μg/m^3 in December (Fig. 6d) respectively. For predicting NO_2 levels, although both methods are close to original data (Fig. 7a and b), the SVM method still possesses better generalization performance than the RBF network does. Either the mean absolute error (MAE) or the maximum absolute error produced by SVM model is smaller than that obtained by RBF one. Table 1 shows comparisons of predicting errors of three pollutants for one-week time series between SVM and RBF models. From Table 1, it can be seen that, for three pollutants, both the MAE and the maximum absolute errors produced by the SVM method are smaller than the ones created by the conventional RBF network in both months. Additionally, the errors in December are less than the ones in June by both methods. Based on the above experiments, it can be concluded that SVM method is superior to RBF model, and possesses good, robust predicting performance.

3.3. Sensitivity on free parameters in the SVM method

In the study, the Gaussian kernel function of SVM contains three free parameters: σ, C and ε. Since there are no general rules to determine these free parameters, it is necessary to investigate impacts of selecting these parameters on the resultant generalization errors. Here, the mean absolute error (MAE) is used as an assessment of deviation between original data and predictions. Generally, the smaller the values of MAE, the better results one can achieve. The MAE is defined as

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|a_i - p_i| \tag{9}$$

Fig. 8 demonstrates the MAE variation via C-values. The figure indicates that the prediction error is scarcely influenced by C-values and is only sensitive if C is small enough, e.g., $C \leqslant 0.001$. As C increases, the MAE value

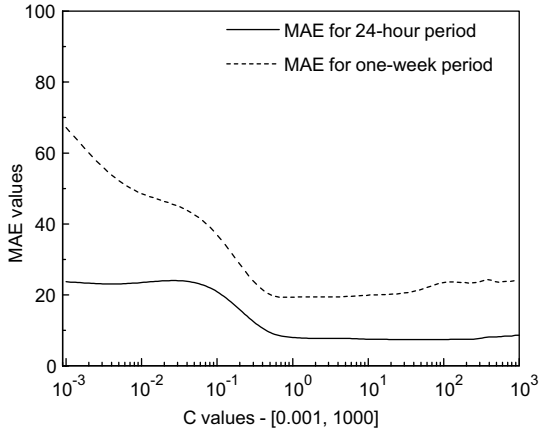

Fig. 8. SVM prediction errors versus C-value.

Table 1
Prediction errors of three pollutants with one-week in advance (μg/m^3)

Prediction errors		RSP		NO_x		NO_2	
		December	June	December	June	December	June
MAE	SVM method	17.657	19.016	131.645	119.667	13.128	11.548
	RBF network	46.157	70.745	2556.391	4058.23	18.707	17.604
Maximum absolute error	SVM method	77.87	74.67	415.99	530.689	48.974	64.817
	RBF network	327.44	1389.16	38758.5	80054.91	61.010	65.456

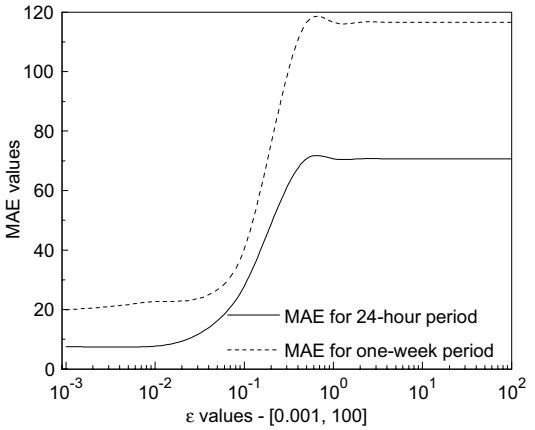

Fig. 9. SVM prediction errors versus ε-values.

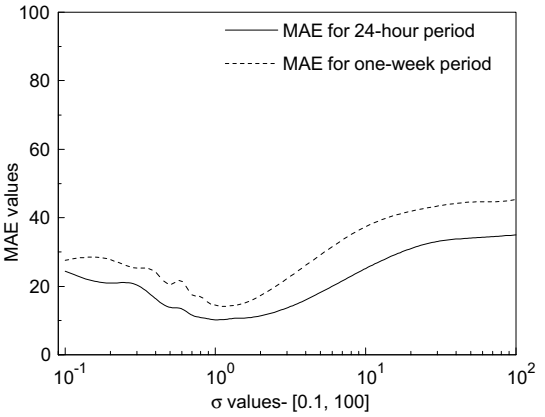

Fig. 10. SVM prediction errors versus σ-values.

decreases rapidly, keeps at small value, and changes slightly when $C \geqslant 0.5$. In general, to guarantee the stable learning process, C should be set with a large value, e.g., $C = 100$ in this study. Fig. 9 illustrates the prediction errors varying versus ε-values. It is also found that ε has little impact on the predicting performance. The MAE values remain almost constant once $\varepsilon < 10^{-2}$ and $\varepsilon > 0.5$. In SVM simulation, it is suggested that ε takes a small value. Hence, we set $\varepsilon = 0.001$ in the simulation.

Theoretically, the spread parameter σ greatly affects the prediction performance. Both too large (e.g., $\sigma \to \infty$) and too small (e.g., $\sigma \to 0$) values of σ may lead to poor predictions. If $\sigma \to 0$, all training data points would become support vectors. Then, for unseen data, the SVM model cannot provide valuable computing guidance and achieve good predicting performance. Otherwise, if $\sigma \to \infty$, all training data points would be regarded as one point. Hence, the SVM model may produce same calculating results for any new data points. Therefore, these two extreme situations should be avoided. It should be noticed that both $\sigma \to 0$ and $\sigma \to \infty$ represent two approximating processes. In actual cases, if $\sigma \ll \|x_i - x_j\|$ and $\sigma \gg \|x_i - x_j\|$, the extreme situations mentioned above will appear. Fig. 10 presents the varying trend of MAE profile with σ-values. It can be seen that MAE is large (i.e., around 24.7) when σ is small (e.g., $\sigma = 0.001$), then decreases to small values with increasing σ, and reach the minimum point (i.e., 10.1) at point of $\sigma = 1.00$. Based on Fig. 10, the MAE curve slightly fluctuates within a range of [10.1, 10.8] when σ falls into [0.9, 1.1], gradually increases with the increase of σ, and finally tends to maintain constant after $\sigma \geqslant 30$. Therefore, in practical applications, only the spread parameter σ of Gaussian kernel function needs to be determined during the simulations, while the other two parameters, i.e., C and ε, can be set in advance by experiences. In our study, the σ-value is finally specified as $\sigma = 1.0$ after trial-and-error process.

4. Conclusion

The potential of applying SVM model in ambient air pollutant prediction is studied and presented in this paper. It can be concluded that SVM model provides a promising alternative and advantage in time series forecast. It offers several advantages over the conventional feed-forward RBF neural networks. Firstly, it contains fewer free parameters than the RBF model does. In this study, the spread parameter σ is the only factor need to be considered in the SVM model once the kernel function is determined. While for conventional RBF model, the network size, the learning parameter estimation, and the network training greatly affect prediction performance and need extra care during the simulation. Secondly, due to the adoption of the *Structure Risk Minimization Principle*, the SVM model provides better predictions than the conventional RBF model. As shown above, the SVM model produces smaller MAE values either for the 24-h or for the one-week time series predictions than that of RBF network. Finally, the SVM model can eliminate the typical drawbacks of conventional neural network models, e.g., "over-fitting" training and local minima, and proves to be more expandable and robust than the conventional RBF network. The application of the SVM method in environmental aspect is a good, interesting attempt; it may be worthy to test its value in more areas.

Acknowledgments

The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (HKSAR), China [Project No. CityU 1013/02E], and a Strategic Research Grant #7001371(BC) from City Uni-

versity of Hong Kong, HKSAR. The provision of original data from HKEPD is also appreciated.

References

Boznar, M., Lesjak, M., Mlakar, P., 1993. A neural network-based method for short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain. Atmospheric Environment 27 (2), 221–230.

Broomhead, D., Lowe, D., 1998. Multivariable functional interpolation and adaptive networks. Complex Systems 2, 321–355.

Burnett, R.T., Smith-Doiron, M., Stieb, D., Cakmak, S., Brook, J.R., 1999. Effects of particulate and gaseous air pollution on cardiorespiratory hospitalizations. Archives of Environment Health 54 (2), 130–139.

Chan, L.Y., Kwok, W.S., 2000. Vertical dispersion of suspended particulates in urban area of Hong Kong. Atmospheric Environment 34, 4403–4412.

Chan, L.Y., Liu, Y.M., 2001. Carbon monoxide levels in popular passenger commuting modes traversing major commuting routes in Hong Kong. Atmospheric Environment 35, 2637–2646.

Comrie, A.C., 1997. Comparing neural networks and regression models for ozone forecasting. Journal of Air and Waste Management 47, 653–663.

Fan, H.Y., Lu, W.Z., Xu, Z.B., 2000. An empirical comparison of three novel genetic algorithms. Engineering Computations 17 (8), 981–1001.

Fan, H.Y., Lu, W.Z., Xi, G., Wang, S.J., 2003. A new neural-network-based calibration method for aerodynamic pressure probes. ASME Journal of Fluids Engineering 125 (1), 113–120.

Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer feed-forward neural networks)—a review of applications in the atmospheric science. Atmospheric Environment 30 (14/15), 2627–2636.

Harrison, R.M., Deacon, A.R., Jones, M.R., 1997. Sources and processes affecting concentrations of PM10 and PM2.5 particulate matter in Birmingham (UK). Atmospheric Environment 31 (24), 4103–4117.

Hong Kong Environment Protection Department, 1998, 1999, 2000, Annual Report.

Lee, E., Chan, C.K., Paatero, P., 1999. Application of positive matrix factorization in source apportionment of particulate pollutants. Atmospheric Environment 33, 3201–3212.

Lu, W.Z., Fan, H.Y., Leung, A.Y.T., Wong, J.C.K., 2002a. Analysis of pollutant levels in Central Hong Kong applying neural network method with particle swarm optimisation. Environmental Monitoring and Assessment 79 (3), 217–230.

Lu, W.Z., Wang, X.K., Fan, H.Y., Wang, W.J., Leung, A.Y.T., Yuen, K.K., 2002b. A preliminary study of ozone trend and its impact on environment in Hong Kong. Environment International 28 (6), 503–512.

Lu, W.Z., Wang, W.J., Fan, H.Y., Leung, A.Y.T., Xu, Z.B., Lo, S.M., Wong, J.C.K., 2002c. Prediction of pollutant levels in Causeway Bay area in Hong Kong using an improved neural network model. ASCE Journal of Environmental Engineering 128 (12), 1146–1157.

Lu, W.Z., Wang, W.J., Wang, X.K., Xu, Z.B., Leung, A.Y.T., 2003. Using improved neural network model to analyze RSP, NO_x and NO_2 levels in urban air in Mong Kok, Hong Kong. Environmental Monitoring & Assessment 87 (3), 235–254.

Mukherjee, S., Osuna, E., Girosi, F., 1997. Nonlinear prediction of chaotic time series using a support vector machine. In: Proc. NNSP, pp. 511–520.

Müller, K.R., Smola, A.J., Rätsch, G., Schölkopf, B., Kohlmorgen, J., Vapnik, V., 1997. Predicting time series with support vector machines. In: Proc. ICANN, pp. 999–1004.

Perez, P., Trier, A., Reyes, J., 2000. Prediction of PM2.5 concentrations several hours in advance using neural networks in Santiago Chile. Atmospheric Environment 34, 1189–1196.

Reich, S.L., Gomez, D.R., Dawidowski, L.E., 1999. Artificial neural network for the identification of unknown air pollution sources. Atmospheric Environment 33, 3045–3052.

Roadknight, M., Balls, G.R., Mills, G.E., Palmer_Brown, B.D., 1997. Modelling complex environmental data. IEEE Transactions on Neural Networks 8 (4), 852–861.

Song, X.H., Hopke, P.K., 1996. Solving the chemical mass balance problem using an artificial neural network. Environmental Science & Technology 30 (2), 531–535.

Spurny, K.R., 1998. On the physics, chemistry and toxicology of ultrafine anthropogenic atmospheric aerosols: (UAAA) new advances. Toxicological Letters 96, 253–261.

Transport Department, 2001. Annual Transport Digest, Hong Kong Printing Department, Hong Kong.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.

Vapnik, V., Golowich, S., Smola, A., 1997. Support method for function approximation regression estimation, and signal processing. In: Mozer, M., Jordan, M., Petsche, T. (Eds.), Advance in Neural Information Processing System 9. MIT Press, Cambridge, MA.

Wang, T., Cheung, V.T.F., Lam, K.S., Kok, G.L., Harris, J.M., 2001. The characteristics of ozone and related compounds in the boundary layer of the South China coast: temporal and vertical variations during autumn season. Atmospheric Environment 35 (15), 2735–2746.

Yi, J., Prybutok, R., 1996. A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. Environmental Pollution 92 (3), 349–357.