Research Article
Prediction of PM2.5 Concentration in a Agricultural Park Based on BP Artificial Neural Network

[1]Bo Chen, Xiaoping Wang, [1]Lixuan Yu, [1]Huan Wang, [2]Yanfang Li, [2]Junqi Chen, [2]Jiangang Zhu,
[3]Hailong Nan and [3]Limin Hou
[1]College of Agricultural Parkry, Beijing Agricultural Parkry University, Beijing 100083, China
[2]College of Soil and Water Conservation, Beijing Agricultural Parkry Carbon Administration, Beijing
100013, China
[3]College of Landscape Architecture, Daxing Bureau of Landscape and Agricultural Parkry,
Beijing 102600, China

Abstract: The accurate prediction of PM2.5 concentration in a agricultural park is important to understand the role agricultural park plays in regulating PM2.5 pollution and guide public close to the nature healthily. An artificial neural network model was established, with meteorological data, atmospheric PM2.5 concentration outside the agricultural park and agricultural park structure as the input factors and PM2.5 hourly average concentration inside the agricultural park as the output factors. Its prediction accuracy was also evaluated in this study. The results show that it can be concluded that BP artificial neural network model is a promising approach in predicting PM2.5 concentration inside a agricultural park.

Keywords: Agricultural park structure, BP artificial neural network, multiple linear regression, PM2.5

INTRODUCTION

PM2.5 refers to the diameter of the atmosphere is less than or equal to 2.5 µm particles, often referred to as particulate matter into the lungs or fine particulate matter (Shao *et al*., 2000). Although PM2.5 is only a few components content of atmospheric composition, but its gravity sedimentation rate is very low in the air, floating in the air for a long time, attached a lot of organic pollutants and heavy metals (Shuping *et al*., 2004), have important effects on human health, environment and climate. agricultural park vegetation can have an impact on the concentration of particulate matter in the atmosphere through direct and indirect ways (Yu *et al*., 2011; Hailong *et al*., 2012; Beckett *et al*., 2000), however, the effect of agricultural park vegetation on PM2.5 is still in its infancy, PM2.5 concentrations within the agricultural park change rule and agricultural park compares concentration inside and outside is unclear. It has been reported that the mass concentration of PM2.5 in the agricultural park affected by many factors, such as sporadic contamination source, weather conditions and stand structure and so on. Namely: the existence of PM2.5 concentrations below the agricultural park within the agricultural park, but also the presence of agricultural park within the agricultural park PM2.5 concentration is higher than

outside (Langner, 2008; Cavanagh *et al*., 2009). In view of this, accurately predict PM2.5 concentrations in the agricultural park for the study of agricultural park regulate the effects of PM2.5 travel, close to the agricultural park to guide public health is of great significance. Studies show that it's presents a strong linear relationship between changes in PM2.5 concentrations and meteorological conditions, the traditional multiple linear regression model to predict PM2.5 mass concentration significant limitations (Guo *et al*., 2010; Yin *et al*., 2007). However, artificial neural networks can build very complex nonlinear models; well reflect the relationship between PM2.5 mass concentrations and parameters of Gardner and Dorling (2000), Chaloulakou *et al*. (2003) and Mckendry (2002), such prediction methods have been used successfully in a variety of pollutants in Kolehmainen *et al*. (2001) and Papanastasiou *et al*. (2007). In this study, Huangcun town landscape ecological agricultural park in Beijing's Daxing district as the research object, applications within the BP artificial neural network model to predict agricultural park-hour average PM2.5 concentration and with the multiple linear regressions to predict the results for comparison, shows the feasibility and accuracy within the artificial neural network model to predict agricultural park-based PM2.5 mass concentrations.

Corresponding Author: Bo Chen, College of Agricultural Parkry, Beijing Agricultural Parkry University, Beijing 100083, China
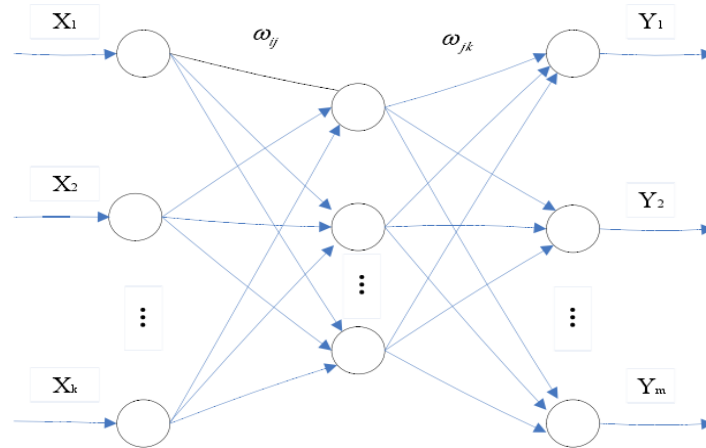
Fig. 1: BP artificial neural network with three layers

MATERIALS AND METHODS

Methods and principles of the experiment: The BP neural network is a kind of multilayer feed forward neural networks, according to statistics; 80-90% of the neural network models adopt BP network or its variations (Hong, 2012). BP neural network consists of an input layer, one or more hidden layers and an output layer; the main feature is the signal to pass before the error back-propagation. BP algorithm principle is: the forward pass, the signal through the input layer, hidden layer and output layer, layer by layer processing and transmission, if the prediction error between the output with a given output not meet accuracy requirements, they turn back propagation and correct connection weights, which predict output looming expectation output. The structure of the three-layer BP neural network is shown in Fig. 1. The transfer function between the input layer and the hidden layer commonly used S-type transformation function, the transfer function between the hidden layer and output layer generally use pure linear transformation function (Min *et al.*, 2013). The BP artificial neural network model between input layer and hidden layer S transfer function expression is:

$$F(x) = \frac{1}{1+e^{-x}} \tag{1}$$

In the formula: x represents the input to the neurons, the neurons $(-\infty, +\infty)$ is mapped to the input range $(0+1)$ on the range, easy to train BP neural network algorithm.

Data acquisition: In this study, 283 sets of data used is obtained from the field in real time. In July 2013-in May 2014, each quarter, randomly selected nine days without precipitation, continuous monitoring of nine hours a day. In Beijing within HUANGCUN large landscape ecological agricultural park monitoring points Wacun Daxing District Site 1 (N 39°43.684', E 116 °18.772') and outside the agricultural park
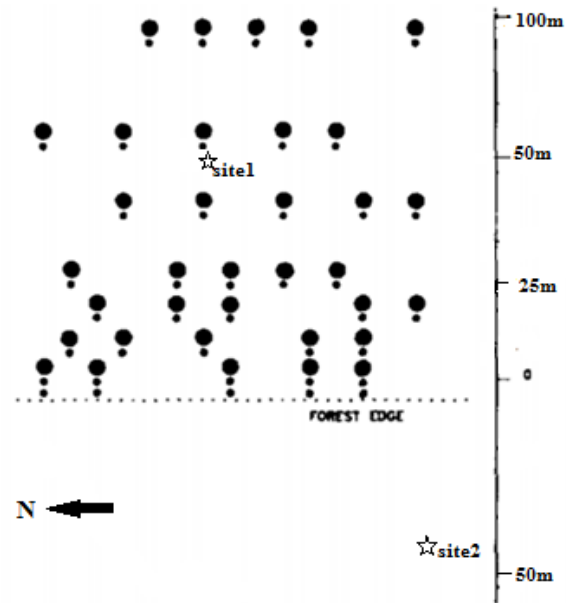

Fig. 2: Monitoring sites

monitoring sites Site 2 (N 39°43.642', E 116 °18.724') synchronous monitoring PM2.5 concentrations in the air is 1.5 m high and meteorological elements (temperature, relative humidity, wind speed, wind direction), straight-line distance between two points is set up to monitor the situation 100 m, monitoring samples are shown in Fig. 2, the main plant experimental area within the agricultural park for Salix (Salix matsudana). Concentration of PM2.5 h using DUST MATE DUST instrument measured, temperature and relative humidity is Taiwan az az8918 triad wind temperature humidity tester, wind speed and direction using the TN-F cups anemometer measured.

Data processing: Due to weather conditions directly affect the change of the atmospheric pollutants concentration, therefore, to predict the quality of agricultural park PM2.5 must consider reunite PM2.5

Table 1: Input variable for prediction models

Input variables	Variable name	Variable symbol	Unit
Background concentrations	Background concentrations of PM2.5	$PM^0_{2.5}$	µg/m³
Meteorological parameters	Temperature	T	°C
	Relative humidity	H	%
	Wind velocity	V_W	m/sec
	Wind sine	$\sin\theta$	
	Wind direction cosine	$\cos\theta$	
Season variable	Season	A_{season}	
Stand parameters	Canopy density	CD	%

background values and agricultural park meteorology conditions.

Background concentration: In this study, we will reunite monitoring of PM2.5 concentrations is set to the background concentration prediction model input parameters, namely PM to characterize the effect of the pollution sources is strong to the agricultural park to PM2.5 concentrations.

Meteorological parameters: Meteorological parameters, including Temperature (T), relative Humidity (H) and Wind speed (VW) data directly into the analysis. Wind direction (θ) by the sine, cosine transform into two variables sinθ and cosθ (Min *et al.*, 2013), were converted to what is about to wind direction and the north-south direction of two variables.

Seasonal parameters: Consider a year of plant growth at different rates in different seasons have different effects on changes in PM2.5 concentrations in the agricultural park, the introduction of seasonal variables Aseason, variable Aseason season as a dual variables. According to the plants growth habit, vigorous spring and summer, make Aseason = 0; autumn and winter, slow growth makes Aseason = 1.

Plant parameters: Canopy density is an important indicator of stand density, has an important impact on the agricultural park particulate matter concentrations and relatively easily obtained, so it is necessary to introduce variable Canopy Density (CD) characterization of plant parameters.

In this study, all the input variables of the prediction model used is shown in Table 1.

BP artificial neural network model: This study was based on internal BP artificial neural network simulation agricultural park PM2.5 concentration process, mainly for programming and establishes appropriate model implemented by MATLAB R2013a Neural Network Toolbox provides functions. Specific experimental procedure is as follows.

Step 1: Network initialization: Set the input layer to the hidden layer, hidden layer to the output layer is any ownership small random number and sets the initial threshold.

Step 2: Reading input variables and output variables (PM2.5 mass concentration within the agricultural park): All sample data were randomly divided into 2 groups: training group 70% of samples, used to train the network and in accordance with the error debugging network, which uses Levenberg-Marquardt training rule algorithms; test group 30% samples for training by independent testing network performance, no effect the training process.

Step 3: Select the training function, the preset number of neurons in the hidden layer begins to train the network. If the network does not converge, replace the training function; if the result of the error is large, then adjust the number of neurons in the hidden layer to train the network again, repeated training network, until you get the best results so far.

Step 4: To determine the final structure of the BP neural network: each layer neuron number is 8-6-1, training function identified as trainlm, the agricultural park and use of the trained network PM2.5 concentration are simulated.

Multiple linear regression models: In this study, MATLAB R2013a multiple linear regression (Multiple linear regression, MLR) model and artificial neural network model for comparison. Multiple linear regression mathematical models are as follows:

$$Y = b0 + \sum_{i=1}^{n} b_i X_i + \varepsilon_i \tag{2}$$

where, X_i is the value of the input variable i; Y is the measured PM2.5 mass concentration within the agricultural park; constant term b0 b_i and regression coefficient calculation is obtained by the least squares method; ε_i is the regression error, return solving process is the average minimum error process.

This study all input variables by gradually, stepwise regression to get agricultural park PM2.5 mass concentration rho (PM2.5) in the multivariate linear regression forecast model is as follows:

ρ (PM2.5) = -19.2179 + 0.9689 ρ (PM) + 0.4959 (CD) - 0.4583T + 0.0946H + 0.3425 VW

Forecast model shows that: when predicting PM2.5 concentrations in the agricultural park, the effect of

background concentration is very important, the second is the relative humidity, the forecasting model by stepwise regression method retains the five input variables, wind variable and seasonal variables are not entered into the regression.

RESULTS AND DISCUSSION

Comparison of predicted results: Figure 2 is based on internal BP artificial neural network simulation and MLR method for predicting agricultural park study area generated PM2.5 concentrations predicted and observed values of the comparison. In the diagram, you can see very clearly the BP artificial neural network and the differences between the MLR model prediction results. The BP artificial neural network model to predict very well the agricultural park PM2.5 mass concentration change under different weather conditions, despite the dramatic changes in the data (sudden increase or decrease) the point, prediction error relative change in network smooth point slightly larger, but the overall

effect is very good prediction. MLR model predicted effect compared with BP artificial neural network model is a lot of difference. When the atmospheric concentration of PM2.5 higher (>100 $\mu g/m^3$), the agricultural park was significantly higher than the actual PM2.5 predictions observations; when atmospheric PM2.5 concentration is low (<10 $\mu g/m^3$), the predicted results were significantly lower than the actual observed values and even negative, this shows that there are many limitations and uncertainties when using PM2.5 concentration MLR model predictions agricultural park. In order to better describe the predictive ability of the model to predict, this study will be evaluated in accordance with the model prediction accuracy.

Evaluation of prediction accuracy: Forecast for the same object, you can choose more than one prediction method, but there is an optimal method. Prediction accuracy is the basis of measuring prediction method is reasonable, is to evaluate the merits of the standard
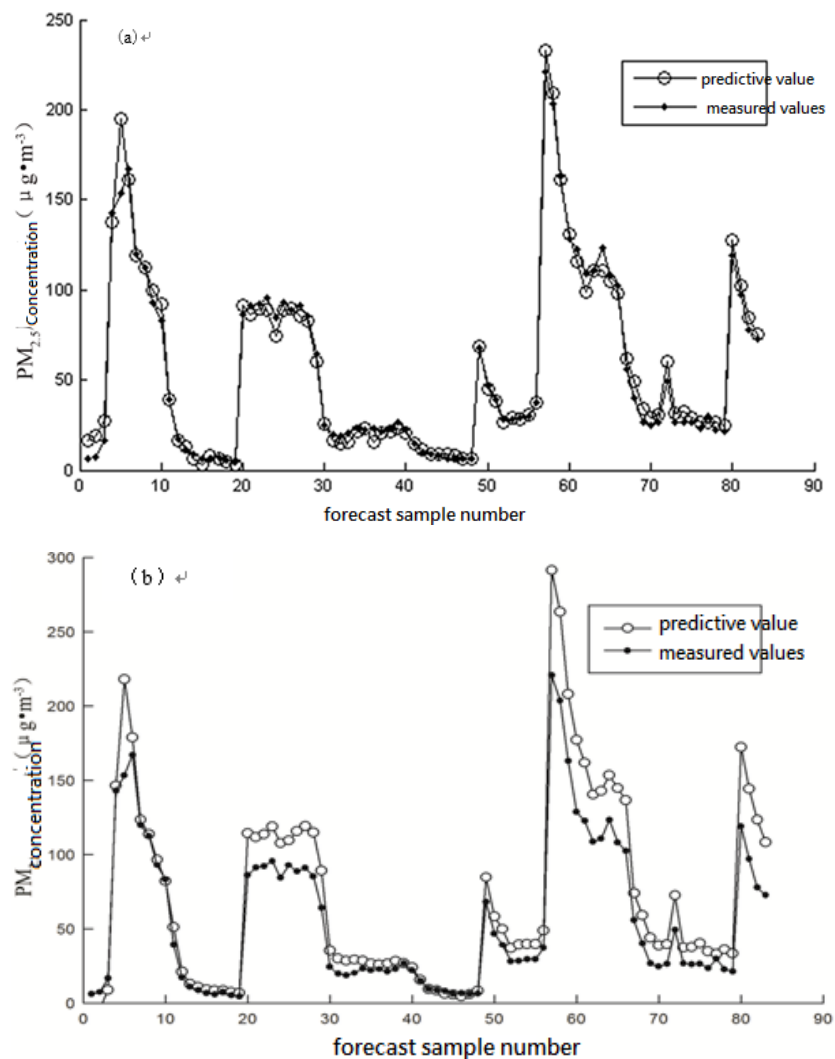

Fig. 3: Line plots of predicted results

Table 2: Performance indicators for different predictive models

Method	R^2	E_{MR}	E_{RMS}
BP neural network	0.99	1.71×10^{-3}	6.77
MLR	0.97	0.27	22.92

model prediction. Performance evaluation model of the commonly used inspection index (Lingzhi *et al.*, 2012; Songmei *et al.*, 2011) including: R^2 (goodness of fit), R^2 closer to 1, represents more relevant, the closer Found predictive value; EMR (average relative error), EMR smaller, more accurate predictions; ERMS is used to measure the deviation of the observed values with the measured values, the smaller the ERMS prediction

result is more ideal. The above two prediction methods of R^2 are shown in Fig. 3, prediction accuracy evaluation indexes are shown in Table 2.

Figure 3 shows the simulation accuracy based on artificial neural network model R^2 is 0.99, however, based on the simulation accuracy multivariate linear regression model was 0.97, lower than the former. From Fig. 3a and b fitting lines and expectations 1:1 line also can be seen that the relationship between artificial neural network models of prediction result is more close to the expected value.
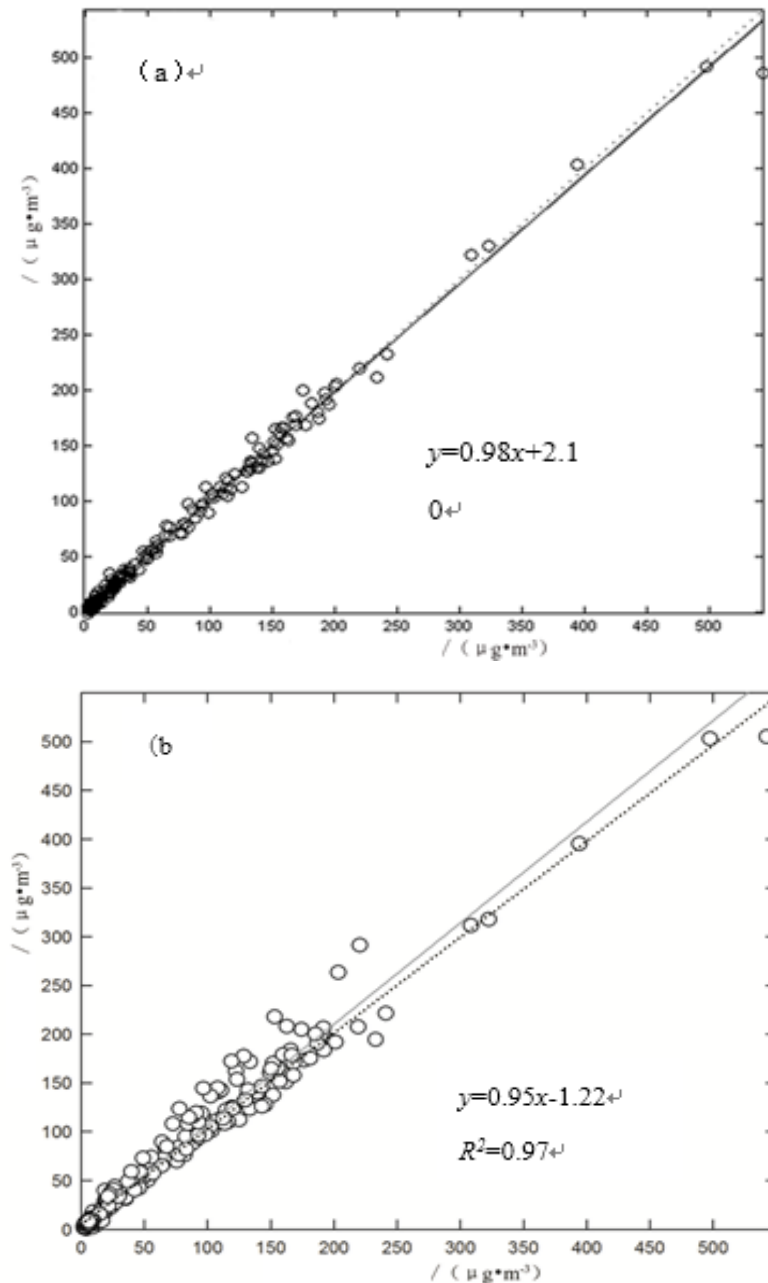

Fig. 4: Scatter plots of predicted versus observed PM2.5 concentrations; (a): Artificial neural network; (b): Multiple linear regression

Table 2 shows that the neural network to predict the average relative error of EMR is 1.71×10-3, the root mean square error of ERMS is 6.77; And MLR model to predict the average relative error of EMR is 0.27, Root Mean Square Error (ERMS) is 22.92. Although the two methods get the correlation between the predicted and the measured values are high (Fig. 4), but can be seen from the error size, BP neural network model to control the error of the stronger. Table 2 contrast accuracy results show that the neural network has strong robustness and fault tolerance, has a strong information comprehensive ability, can well handle complex nonlinear relationship. The above analysis shows that two kinds of prediction models could predict the hour average PM2.5 mass concentration within the agricultural park, in contrast, within the artificial neural network model to predict changes in agricultural park PM2.5 mass concentrations greater linear regression model results closer to the measured value changes, forecasting more accurate.

CONCLUSION

- BP artificial neural network and the traditional multiple linear regression model two kinds of methods can be used within the agricultural park PM2.5 air quality concentrations predicted, although two were dramatic changes in PM2.5 concentrations predicted some defects, but still be able to sparse monitoring data input conditions reflect the time variation of the basic agricultural park PM2.5 pollution.

- With respect to the multiple linear regression model, artificial neural networks can achieve more accurate prediction of PM2.5 in the agricultural park and its predictions to better capture the meteorological factors and stand structure of the pollutant concentrations in the agricultural park influence of nonlinear. BP artificial neural network can be used to predict the agricultural park within the preferred method of PM2.5 mass concentration.

- This study selects the meteorological parameters, strong pollution characterization of variables and the agricultural park canopy density can accurately describe the climate change impact on the mass concentration of PM2.5 and the stand structure and the data is easier to get used to predict agricultural park PM2.5 mass concentration within the ideal.

- The use of artificial neural network to establish agricultural park PM2.5 mass concentration prediction model applies not only to the general pollution concentrations, for the period of the agricultural park in the high pollution PM2.5 mass concentration forecast is more accurate and wide applicability.

REFERENCES

Beckett, K.P., P.H. Freer-Smith and G. Taylor, 2000. Particulate pollution capture by urban trees: Effect of species and windspeed. Global Change Biol., 6: 995-1003.

Cavanagh, J.A.E., P. Zawar-Reza and J.G. Wilson, 2009. Spatial attenuation of ambient particulate matter air pollution within an urbanised native forest patch. Urban For. Urban Gree., 8: 21-30.

Chaloulakou, A., M. Saisana and N. Spyrellis, 2003. Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens. Sci. Total Environ., 313: 1-13.

Gardner, M.W. and S.R. Dorling, 2000. Statistical surface ozone models: An improved methodology to account for non-linear behaviour. Atmos. Environ., 34(1): 21-34.

Guo, E.G., C. Wang, G.F. Qie, C. Fang, Z.W. Sun and Z.H. Zhou, 2010. Seasonal variations of airborne particulate matter in typical recreation forests in the West Mountain of Beijing. J. Northeast Forestry Univ., 38(10): 55-57.

Hailong, W., Y. Xinxiao, S. Shen *et al.*, 2012. Advances in the study of PM2.5 characteristic and the regulation of agricultural parks to PM2.5. Sci. Soil Water Conserv., 10(6): 116-122.

Hong, C., 2012. Traffic Flow Prediction Based on BP Neural Network. Chang'an University, Xi'an, pp: 30.

Kolehmainen, M., H. Martikainen and J. Ruuskanen, 2001. Neural networks and periodic components used in air quality forecasting. Atmos. Environ., 35(5): 815-825.

Langner, M., 2008. Reduction of airborne particulates by urban green. Bundesamt für Naturschutz, 179: 129-137.

Lingzhi, S., D. Qihong, L. Chan and L. Wei-wei, 2012. Prediction of PM10 mass concentrations based on BP artificial neural network. J. Central South Univ. Sci. Technol., 43(5): 1969-1974.

Mckendry, I.G., 2002. Evaluation of artificial neural networks for fine particulate pollution (PM10 and PM2.5) forecasting. J. Air Waste Manage., 52(9): 1096-1101.

Min, W., Z. Bin, G. Yu and H. Jinqiang, 2013. BP artificial neural network-based analysis of spatial variability of urban PM2.5 concentration. Environ. Pollut. Control, 35(9): 63-70.

Papanastasiou, D.K., D. Melas and I. Kioutsioukis, 2007. Development and assessment of application of neural network and multiple regression models in order to predict PM10 levels in a medium-sized Mediterranean city. Water Air Soil Poll., 182: 325-334.

Shao, L., Z. Shi and Q. Huang, 2000. A study of inhalable particulates in urban ambient air. Environ. Protect., 1: 24-26, 29.

Shuping, S., G. Runze and Z. Jing, 2004. Inhalable particulate matter (PM10) related to different green cover percentage and greenbelt types in Beijing. Chinese Landscape Archit., 20(3): 334-346.

Songmei, Z., G. Dongdong, H. Qibiao and W. Ji, 2011. Soil moisture prediction based on artificial neural network model. Chinese Agric. Sci. Bull., 27(8): 280-283.

Yin, S., J.P. Cai, L.P. Chen, Z.M. Shen, X.D. Zou *et al.*, 2007. Effects of vegetation status in urban green spaces on particles removal in a canyon street atmosphere. Acta Ecol. Sinica, 27(11): 4590-4595.

Yu, Y., N. Schleicher, S. Norra, M. Fricker, V. Dietze, U. Kaminski, K. Cen and D. Stüben, 2011. Dynamics and origin of PM2.5 during a three-year sampling period in Beijing, China. J. Environ. Monitor., 13: 334-346.