Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jes

ELSEVIER

JES
JOURNAL OF
ENVIRONMENTAL
SCIENCES

www.jesc.ac.cn

Influencing factors and prediction of ambient Peroxyacetyl nitrate concentration in Beijing, China

Boya Zhang[1,**], Bu Zhao[2,**], Peng Zuo[1], Zhi Huang[1], Jianbo Zhang[1,*]

1. *State Key Joint Laboratory of Environmental Simulation and Pollution Control, College of Environmental Sciences and Engineering, Peking University, Beijing 100871, China*
2. *School of Environment, Tsinghua University, Beijing 100084, China.*

ARTICLE INFO

Article history:
Received 11 February 2018
Revised 18 July 2018
Accepted 19 July 2018
Available online 5 August 2018

Keywords:
Artificial neural network
Conventional atmospheric pollutants
Meteorological parameters
Concentration prediction
Multiple linear regression

ABSTRACT

Peroxyacyl nitrates (PANs) are important secondary pollutants in ground-level atmosphere. Accurate prediction of atmospheric pollutant concentrations is crucial to guide effective precautions for before and during specific pollution events. In this study, four models based on the back-propagation (BP) artificial neural network (ANN) and multiple linear regression (MLR) methods were used to predict the hourly average PAN concentrations at Peking University, Beijing, in 2014. The model inputs were atmospheric pollutant data and meteorological parameters. **Model 3** using a BP-ANN based on the original variables achieved the best prediction results among the four models, with a correlation coefficient (*R*) of 0.7089, mean bias error of −0.0043 ppb, mean absolute error of 0.4836 ppb, root mean squared error of 0.5320 ppb, and Willmott's index of agreement of 0.8214. Based on a comparison of the performance indices of the MLR and BP-ANN models, we concluded that the BP-ANN model was able to capture the highly non-linear relationships between PAN concentration and the conventional atmospheric pollutant and meteorological parameters, providing more accurate results than the traditional MLR models did, with a markedly higher goodness of *R*. The selected meteorological and atmospheric pollutant parameters described a sufficient amount of PAN variation, and thus provided satisfactory prediction results. More specifically, the BP-ANN model performed very well for capturing the variation pattern when PAN concentrations were low. The findings of this study address some of the existing knowledge gaps in this research field and provide a theoretical basis for future regional air pollution control.

Introduction

Peroxyacyl nitrates (PANs, $RC(O)OONO_2$) are important secondary pollutants in the ground-level atmosphere where there are no direct anthropogenic emissions. Among PANs, peroxyacetyl nitrate (PAN, $R = CH_3$) is the most important, with the highest atmospheric concentrations (Seinfeld and Pandis, 2012). Studies of PANs have gained widespread attention since the discovery of PAN in photochemical smog in the Los Angeles area in the 1950s (Stephens et al., 1956). Relevant monitoring performed around the world during recent decades has shown that PANs are ubiquitous in the

[*] *Corresponding author.*
E-mail address: jbzhang@pku.edu.cn. (J. Zhang).
[**] These two authors contributed equally to this work.

https://doi.org/10.1016/j.jes.2018.07.011

global atmosphere (Singh et al., 1986; Payne et al., 2013; Fischer et al., 2014). Related studies have also shown that daily exposure to high concentrations of PAN may have adverse effects on plant growth and human health (Vyskocil et al., 1998). Such impacts imply the importance of ground-level air pollution forecasting as an indispensable warning system to enable effective preparation in the case of severe episodes (Bishop, 1995). However, current studies of PANs are mainly focused on local monitoring data derived from scattered monitoring sites and relatively short monitoring periods (Zhang et al., 2009; Gao et al., 2014; Zhang et al., 2014; Zhang et al., 2015; Zhang et al., 2017). In previous studies, due to the limitation of human and financial resources, the monitoring sites are mainly located in megacities and the monitoring periods are generally less than 3 months. Thus, the results of these studies reflect only the local concentration levels at a specific time. There have been few studies to date on the patterns of variation and prediction of ambient PAN concentration in China.

The formation of PANs requires a series of complex photochemical reactions that not only involve NO_x and different VOCs but also are closely related to meteorological parameters such as temperature and wind speed (Kleindienst, 1994; Fischer et al., 2014). Accordingly, the numerical prediction of ant in the regression resultsons requires a complex systematic procedure due to the multivariate, strong coupling and nonlinear characteristics involved (Lei et al., 1998). Therefore, an appropriate modeling approach that can deal with the above-mentioned problems is necessary to obtain reliable forecasting outcomes. In current studies related to the prediction of atmospheric pollutant concentrations, the most commonly used methods include gray forecast models, multiple statistical analysis theory, the fuzzy recognition method, and artificial neural networks (ANNs). Among these methods, ANN models appear to offer an effective mathematical analysis tool because they have been found to perform remarkably well in capturing complex interactions and dealing with nonlinear problems (Baawain et al., 2007). Gardner and Dorling (1998) summed up the application of ANN models in atmospheric science, dividing it into three main categories: prediction, function approximation, and pattern classification. Several previous studies have used ANN models to successfully predict ground-level air pollution levels, assessing O_3 (Chattopadhyay and Chattopadhyay, 2012; Faris et al., 2014; Pires et al., 2012), PM_{10} (de Gennaro et al., 2013; Zhang et al., 2013), NO_2 (Russo et al., 2013), and SO_2 (Akkoyunlu et al., 2010; Ozkan et al., 2010; Sánchez et al., 2013). However, the prediction of PAN concentrations has not previously been studied.

In this study, the PAN concentrations at Peking University (PKU), Beijing, were continuously monitored from March to July 2014. Based on the monitoring results, the variations in PAN concentration were analyzed and summarized. ANN and multiple linear regression (MLR) methods were used to predict the hourly average PAN concentrations. The model inputs were atmospheric pollutant data (PAN, CO, SO_2, NO, NO_2, O_3, and $PM_{2.5}$) and meteorological parameters [temperature (TEMP), relative humidity (RH), wind speed (WS), and wind direction (WD)]. Different prediction models were evaluated based on the performance indices when compared with the actual monitoring data. This study successfully used conventional atmospheric pollutants and meteorological parameters to predict PAN concentrations, addressing existing gaps in this research field and providing a theoretical basis for future regional air pollution control.

1. Methods and materials

1.1. *Data collection and preprocessing*

The ambient air quality data used in this study were collected from 1 March to 10 July 2014. The monitoring site (39.99°N, 116.31°E) was located at Peking University (PKU) at a height of 25 m above ground level. PKU is located in Zhongguancun Street, which is one of the busiest areas in Beijing, with large crowds and heavy traffic. There is no major source of pollution near this site.

Monitoring data included concentrations of PAN and other conventional atmospheric pollutants (CO, SO_2, NO, NO_2, O_3, $PM_{2.5}$), as well as meteorological parameters (TEMP, RH, WS, and WD). The instrument detection limit was 5–10 pptv, and the detection time resolution was 5 min. The observation data were processed into an hourly format for further analysis, and the entire valid data set consisted of 2771 hr of observations. The final data format is shown in Table 1.

1.2. *Prediction of PAN concentrations*

Different categories of methods exist for predicting atmospheric pollutant (e.g., O_3, $PM_{2.5}$, $PM_{10,}$ etc.) concentrations. Among these methods, the statistical method has the

Table 1 – Final data format of monitoring data.

Variables	Variable symbols	Unit
Peroxyacetyl nitrate hourly average concentration	PAN	ppb
Carbon monoxide hourly average concentration	CO	ppb
Sulfur dioxide hourly average concentration	SO_2	ppb
Nitric oxide hourly average concentration	NO	ppb
Nitrogen dioxide hourly average concentration	NO_2	ppb
Ozone hourly average concentration	O_3	ppb
$PM_{2.5}$ hourly average concentration	$PM_{2.5}$	mg/m^3
Temperature	TEMP	°C
Relative humidity	RH	%
Wind speed	WS	°
Wind direction	WD	m/sec

advantage of high accuracy for short-term prediction. In this study, the MLR method and an ANN model were used to predict the PAN concentrations in Beijing. The multiple regression analysis, principal component analysis, and ANN modeling were performed using the Statistical Product and Service Solutions (SPSS) software program.

1.2.1. Multiple linear regression

Multiple statistical analysis theory considers the correlation between environmental factors and atmospheric pollutant concentrations. Thus, this method can determine the relationship between elements in the forecast model and achieve an effective quantitative prediction (Huang, 1991). Several previous studies have used ANN models to successfully make a prediction about ground-level air pollution levels, including O_3, $PM_{2.5}$, PM_{10} et al. Among these studies, several of them (Cai et al., 2009; Grivas and Chaloulakou, 2006; Sousa et al., 2006; Sousa et al., 2007) also use MLR as a comparison with the ANN methods in order to verify the outcome of ANN methods. As this is a preliminary study to apply ANN model on the prediction of the concentration of PANs, we believe that the MLR method can be an effective way to test the ANN results.

Therefore, in this study, we used the MLR method to depict the relationship between the input parameters and the prediction values. In contrast to the unitary linear regression method, MLR is used to reveal the interdependencies between one dependent variable Y and multiple independent variables x_1, x_2... x_m (independent of one another). The established regression model is represented by Eq. (1):

$$\begin{cases} y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_m x_{tm} + \varepsilon_t (t = 1, 2, \cdots, n) \\ E(\varepsilon_t) = 0, \mathrm{Var}(\varepsilon_t) = \sigma^2, \mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0 (if\ i \neq j) \\ or\ \varepsilon_t \sim N(0, \sigma^2), (t = 1, 2, \cdots, n) \end{cases} \quad (1)$$

1.2.2. Artificial neural network

An ANN is a computing system inspired by biological neural networks, and is composed of a number of simple and highly interconnected processing elements that process information by their dynamic state response to external inputs (Nelson and Illingworth, 1991).

A back-propagation (BP) ANN is a multi-layer feed-forward network using an error back propagation algorithm, which is one of the most widely used algorithms in ANN models (Baawain and Al-Serihi, 2014). A BP-ANN is usually composed of an input layer, one or more hidden layers, and an output layer. All layers are made up of a large number of simple non-connected 'neurons,' and every two 'neurons' of each end-to-end layer are connected. Different layers are connected by specific connection weights based on the training algorithm. The neurons of the input layer transmit the received data to the hidden layer, and the neurons of the hidden and output layers calculate their respective inputs via a non-linear transfer function. The ANN system 'learns' through these pairs of input and output data, and finally forms nonlinear mapping that describes the relationship between the input and output variables.

The learning process of the BP-ANN is shown in Fig. 1. BP learning starts with all weights initialized randomly. Weights are then modified as the algorithm progresses until steady-state values are reached. An input vector, given as a signal to the network, passes from the input layer to the hidden layer. After being processed by the hidden layer, the vector then passes to the output layer, where an output is produced. This is a feed-forward propagation process. Then, the error between the output and the actual value is calculated and propagated backward along the network to correct the connection weigh. This is a back-propagation process. Finally, another input is given, and the above learning processes are repeated. The ANN continues the learning process until the error minimization criteria are reached.

The number of neurons in different layers were decided due to various principles. For the neurons in the input layer, the number of neurons comprising in the input layer should be completely and uniquely determined by the training dataset we have, which means the number of neurons in the input layer should be equal to the number of features (columns) in the data. In our case, as we generally understand the generate mechanism of PANs, 10 parameters (CO, SO_2, NO, NO_2, O_3, $PM_{2.5}$, TEMP, RH, WS, and WD), which were thought to be related to the generation of PANs, are included in our data set. That is why there were 10 neurons in the input layer of model 3. For the neurons in output layer, the number of neurons would vary with the problems need to be solved. While for the neurons in hidden layer, there is no hard-and-fast rule for it. The number is often determined based on empirical formulas or trying & testing. Too few nodes will lead to high error for the system as the predictive factors might be too complex for a small number of nodes to capture. And too many nodes will overfit to the training data and not generalize well. Based on some previous research, we thought that the number of neurons in hidden layer should be somewhere between the size of the input and output layer and the exact number was decided by the empirical formula from Huang's Paper (Huang et al., 2010):

$$S = \sqrt{0.43mn + 0.12n^2 + 2.54m + 0.77n + 0.35} + 0.51 \quad (2)$$

The sample is subdivided into 'training,' 'validation,' and 'test' sets. According to the definitions of Ripley (1996), a training set is a set of examples used for learning to fit the parameters (i.e., weights) of the classifier; a validation set is a set of examples used to tune the parameters (i.e., architecture, not weights) of a classifier, e.g., to choose the number of hidden units in a neural network; and a test set is a set of examples used only to assess the performance (generalization) of a fully specified classifier. However, in general, when it comes to practical applications, the data set is subdivided into only a training set and a test set. In this study, a feed-forward BP multi-layer preceptor (MLP) neural network architecture was selected for modeling the concentrations, which meant that there was no need for the validation set. Thus, we chose to subdivide the prepared data set into two subsets with a training set size of 2699 and a test set size of 72.

1.3. Model evaluation criteria

To evaluate and compare the forecast models, several commonly used performance indices were chosen to describe the performance of the models (Chaloulakou et al., 2003a; Chaloulakou et al., 2003b; Lu et al., 2004), which are shown in Table 2.

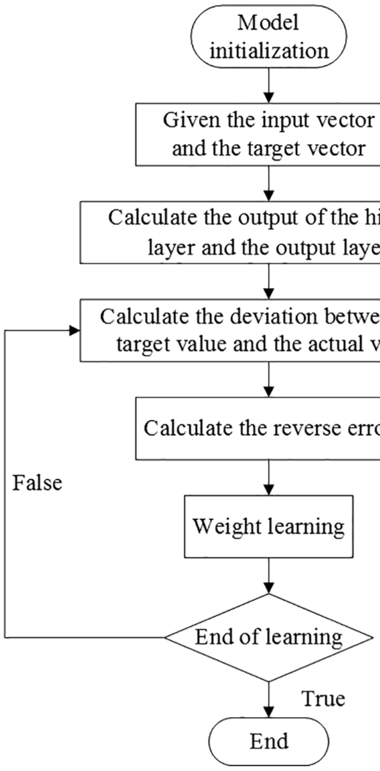
Model initialization
Given the input vector and the target vector
Calculate the output of the hidden layer and the output layer
Calculate the deviation between the target value and the actual value
Calculate the reverse error
False
Weight learning
End of learning
True
End

Fig. 1 – Learning process of the BP-ANN.

In this study, the correlation coefficient (R), mean bias error (MBE), mean absolute error (MAE), root mean squared error (RMSE), and Willmott's index of agreement ($d_{2)}$ were selected to comprehensively evaluate the model performance.

2. Results and discussion

2.1. Variation pattern of PAN concentrations based on the monitoring data

Fig. 2 shows the continuous monitoring data from March to July 2014. The concentrations of CO, SO_2, NO, and $PM_{2.5}$ in spring were higher than those in summer, whereas the concentrations of O_3 and PAN were higher in summer than in spring. However, high PAN concentrations were also observed during the Spring Festival (25 March).

2.2. PAN concentration predictions

2.2.1. MLR prediction model

After stepwise regression of all 10 input variables, the PAN concentration prediction model obtained using MLR based on the original variables, denoted as **Model 1,** was represented by:

$$\begin{aligned}PAN = 0.147 + 0.011PM_{2.5} + 0.037O_3 + 0.042NO_2 – 0.051TEMP\\ –0.010NO – 0.0491CO + 0.19SO_2 – 0.057WS – 0.001WD\end{aligned} \tag{3}$$

Model 1 preserved nine input variables via the stepwise regression method, and the variable RH was negligible because it was not significant in the regression results. **Model 1** shows that the PAN concentration was influenced by various pollutants and meteorological parameters. Among these, the closest correlation was between PAN concentration and SO_2 concentration. SO_2, NO_2, O_3, and $PM_{2.5}$ concentration were positively correlated with PAN concentration, whereas CO, NO, TEMP, WS and WD were negatively correlated with PAN concentration.

Based on the 10 input variables, to rule out significant relationships among the original variables, the corresponding 10 principal components were calculated, as shown in Table 3. After stepwise regression of all 10 principal components, the PAN concentration prediction model obtained using MLR based on the principal components, denoted as **Model 2,** was represented by:

$$\begin{aligned}PAN = 1.701 + 0.704PC_2 + 0.422PC_3 + 0.368PC_{10} – 0.265PC_8\\ –0.235PC_6 + 0.213PC_1 + 0.191PC_4 – 0.153PC_9 + 0.107PC_5 – 0.044PC_7\end{aligned} \tag{4}$$

2.2.2. BP-ANN prediction model

To predict the PAN concentrations, a 10–6-1 BP-ANN model, denoted as **Model 3,** was constructed with an input layer with 10 'neurons,' a hidden layer with 6 'neurons,' and an output layer with the PAN concentration as output. The hyperbolic transfer function [shown in eq.(5)], was selected in this model. The structure of **Model 3** and the comparison between the

Table 2 – Performance indices of different models.

Performance indices	Calculation method	Evaluation criterion
Correlation coefficient (R)	$R = \sqrt{\frac{\sum_{i=1}^{n} (Y_i – \overline{Y}_i)^2 - \sum_{i=1}^{n} (Y_i – \widehat{Y}_i)^2}{\sum_{i=1}^{n} (Y_i – \overline{Y}_i)^2}}$	[0,1], 1 is optimal
Mean bias error (MBE)	$MBE = \frac{1}{n}\sum_{i=1}^{n}(\widehat{Y}_i – Y_i)$	(–∞,+∞), 0 is optimal
Mean absolute error (MAE)	$MAE = \frac{1}{n}\sum_{i=1}^{n} \lvert\widehat{Y}_i – Y_i\rvert$	[0,+∞], 0 is optimal
Root mean squared error (RMSE)	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (Y_i – \widehat{Y}_i)^2}$	[0,+∞], 0 is optimal
Willmott's Index of Agreement (d_2)	$d_2 = 1 – \frac{[\sum_{i=1}^{n} \lvert\hat{Y}_i – Y_i\rvert^2]}{[\sum_{i=1}^{n} (\lvert\hat{Y}_i – \overline{Y}_i\rvert + \lvert Y_i – \overline{Y}_i\rvert)^2]}$	[0,1], 1 is optimal

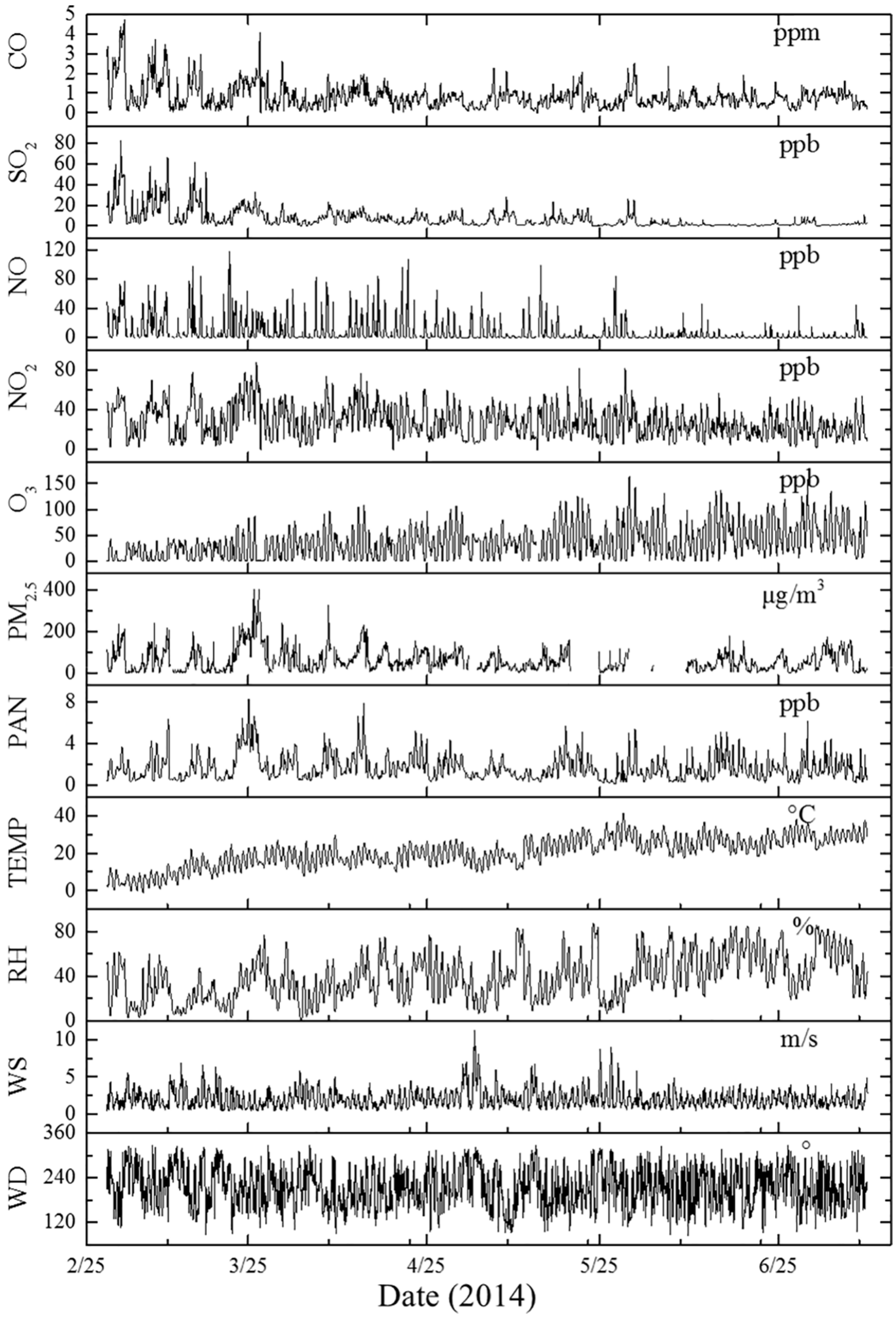

Fig. 2 – Continuous monitoring data from March to July 2014.

monitored and predicted values of PAN concentrations are shown in Fig. 3.

$$f(x) = \frac{e^x + e^{-x}}{1 + e^{-x}} \tag{5}$$

The comparison of the predicted values from the model and the actual values is shown in the right-hand panel of Fig. 3. It can be seen that **Model 3** performed relatively well, and the predicted values were close to the actual values. The predicted concentration fitting line was close to the expected 1:1 line, and the residuals were small and relatively concentrated.

For any prediction model, including ANN modeling, the number and selection of appropriate input variables are very important (Karacan, 2007; Karacan, 2008). The aim of principal component analysis is to reduce the dimensionality of data sets that contain a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. In this study, principal components were used to simplify the BP-ANN model, and **Model 4** was constructed. We

Table 3 – Principle components constructed from the original variables.

Original variable	Principle component									
	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7	PC_8	PC_9	PC_{10}
CO	0.838	0.247	0.273	0.185	−0.074	0.175	−0.105	0.101	−0.196	−0.178
SO_2	0.707	0.056	0.596	0.011	−0.107	0.109	−0.242	0.086	0.211	0.085
NO	0.706	−0.245	−0.104	0.001	0.442	0.424	0.226	−0.050	0.040	0.027
NO_2	0.853	0.046	−0.113	−0.068	0.252	−0.334	−0.008	0.200	−0.118	0.157
$PM_{2.5}$	0.596	0.542	0.173	0.364	0.071	−0.285	0.191	−0.245	0.073	−0.033
O_3	−0.682	0.543	0.275	0.141	0.124	0.216	−0.128	−0.128	−0.144	0.170
TEMP	−0.611	0.568	−0.186	0.082	0.430	−0.024	−0.128	0.190	0.116	−0.106
RH	0.281	0.507	−0.626	0.237	−0.375	0.195	0.115	0.130	0.047	0.078
WS	−0.628	−0.158	0.556	0.309	−0.043	−0.012	0.348	0.229	−0.006	0.021
WD	−0.067	−0.604	−0.243	0.719	0.076	−0.033	−0.215	−0.022	0.001	0.020

chose PC_1, PC_2, PC_3, and PC_4, which provided 80% of the accumulated variance contribution, to constitute the four 'neurons' of the input layer. The hidden layer comprised three 'neurons,' and the output of the output layer was the PAN concentration. The hyperbolic transfer function was also selected in this model. The structure of **Model 4** and the comparison between the monitored and predicted values of PAN concentrations are shown in Fig. 4.

A comparison of the predicted values from **Model 4** and the actual values is shown in the right-hand panel of Fig. 4. For the low values, the predicted values were close to the actual values, and the residual values were comparatively small and concentrated. However, for the high values, the difference between the actual and predicted values was very large, and the residuals were also large and scattered. Compared with **Model 3**, the goodness of the **Model 4** declined. This was mainly due to the selection of the four principal components as the model input, which resulted in a lack of information and reduced the prediction accuracy.

The performance indices of the four forecast models are shown in Table 4 to compare the prediction results of the different models. The R-values of the four models were all <0.8, indicating that none of the models was able to completely capture the variation pattern of PAN concentration. However, comparing all four models, the performance indices of **Model 3** were superior to those of the other three

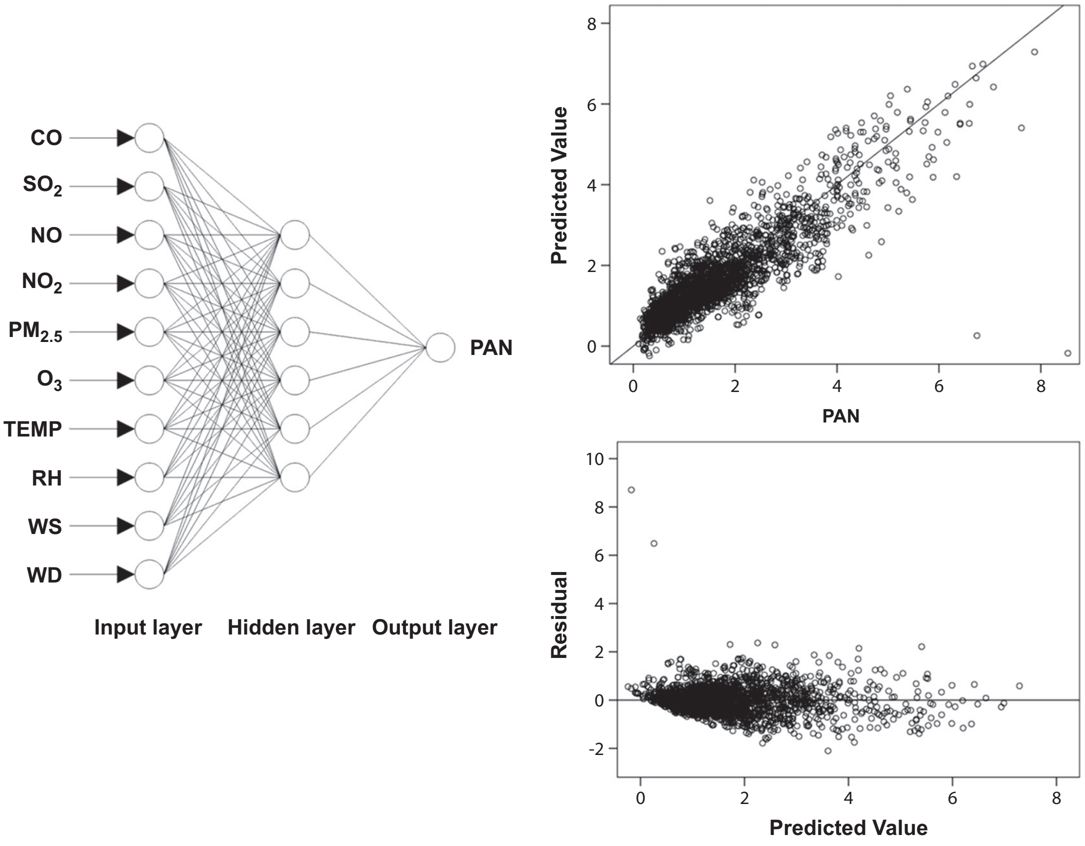

Fig. 3 – Structure of Model 3 (left) and comparison between monitored and predicted values of PAN concentrations (right).

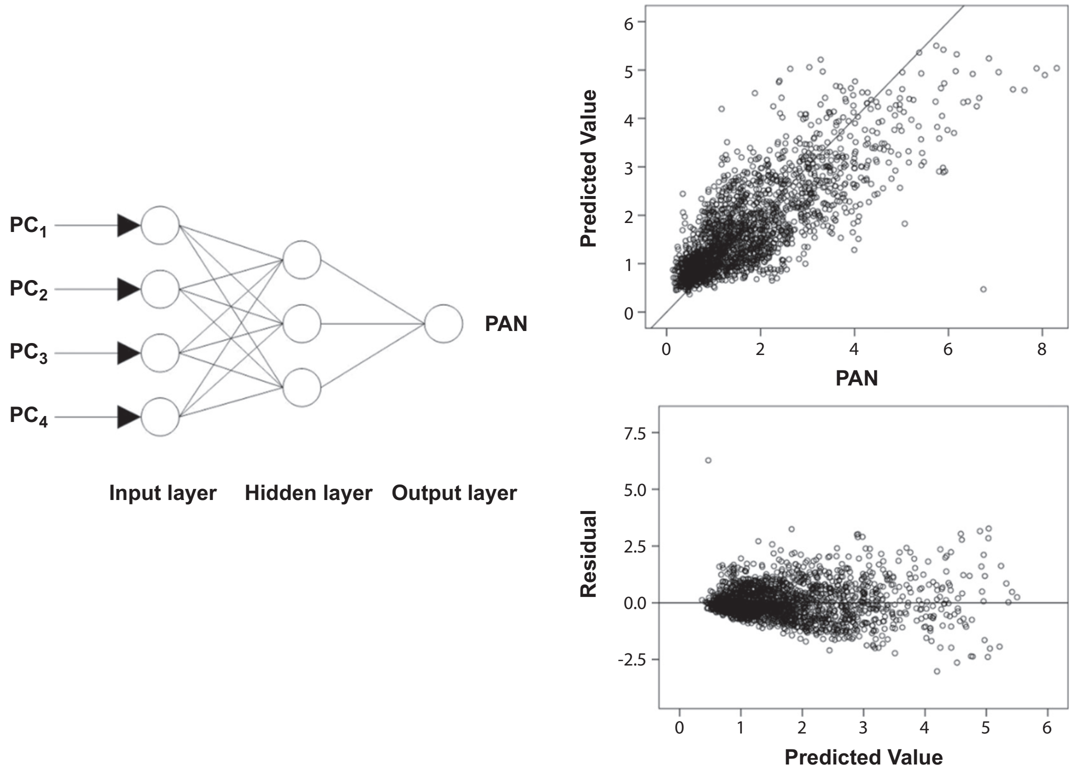

Fig. 4 – Structure of Model 4 **(left) and comparison between monitored and predicted values of PAN concentrations (right).**

models, and the prediction results from this model were closer to the actual values. Although **Model 4** was also based on an ANN, the model input was chosen as the four principal components, which missed important information, rendering the model's performance even worse than those of **Models 1** and **2**, which were based on the MLR method.

To better elucidate the prediction performance of the different models, we performed a time-series analysis of the prediction results. The chosen time period was 8–10 July 2014 (3 d, 72 h), a period of high PAN concentrations. The time-series analysis between the monitored and predicted values of PAN concentration is shown in Fig. 5. **Model 3**, obtained using a BP-ANN based on the original variables, achieved the best prediction results among the four models; the prediction accuracies of **Models 1** to **4** were 31, 33, 54, and 25%, respectively. **Model 3** performed relatively poorly during high-pollution periods, and it was difficult for the model to keep up with the changes in the actual values; however, the prediction results were accurate for most of the period. The reasons for the misfits in the model may be caused by the reasons below: First, the training dataset may be still not large enough for the ANN to capture the variation rule of PAN concentration which may affect the prediction accuracy. Second, the test set in our study is relatively small which may increase rate of misfits. To solve this problem, larger database of PAN concentration and related parameters is needed in the future. However, it should be noticed that no matter how large the dataset is, the difference between modeled and monitoring results will still be exist due to the unavoidable errors. MLR **Models 1** and **2**, the models only predicted the average variation trend of the PAN concentration, and the predicted values also did not keep up with the changes in PAN concentration. For extreme values, both models performed poorly, and both produced negative values in the prediction of low values, whereas the BP-ANN model was able to adequately capture this variation pattern and achieved better prediction results.

Table 4 – Performance indices of different models.

	R	MBE	MAE	RMSE	d_2
Model 1	0.5091	−0.0030	0.4809	0.6492	0.7844
Model 2	0.5085	−0.0043	0.4836	0.6495	0.7837
Model 3	0.7089	0.0958	0.4058	0.5320	0.8214
Model 4	0.3488	0.3361	0.5757	0.7069	0.7334

3. Conclusions

This study investigated the potential use of a systematic approach to develop ANN models for predicting ground-level PAN concentrations at a specific receptor area in Beijing, in 2014. We determined the PAN concentrations according to their relationships with conventional atmospheric pollutants and meteorological conditions. Four models based on the BP-ANN and MLR methods were used to predict the hourly average PAN concentrations. Based on a comparison of the performance indices of the MLR and BP-ANN models, we concluded that the BP-ANN model adequately captured the

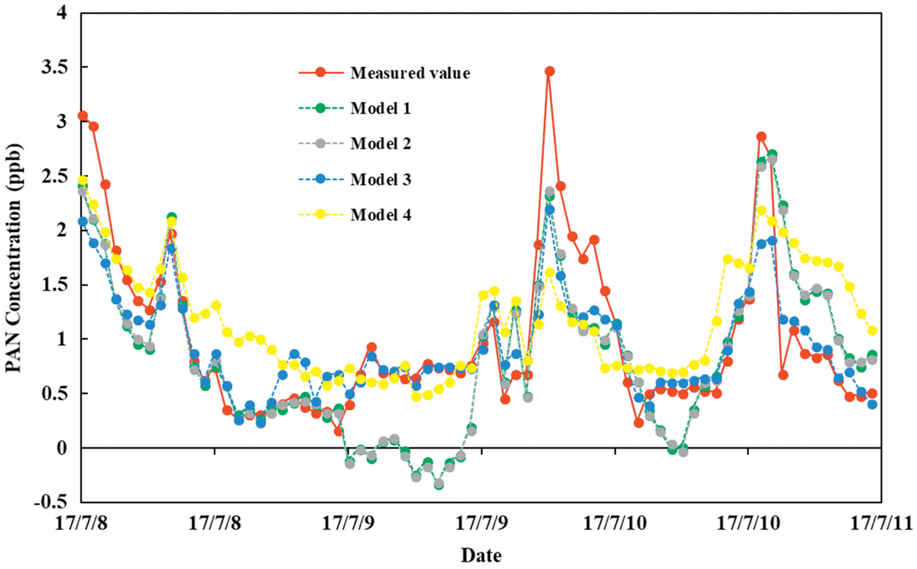

Fig. 5 – Time-series analysis of different models.

highly non-linear relationships between the PAN concentration and the conventional atmospheric pollutant and meteorological parameters, and provided superior results compared to the traditional MLR models, with much higher goodness of R. The selected meteorological and atmospheric pollutant parameters captured a sufficient amount of the PAN variation, and thus provided comparatively satisfactory prediction results. More specifically, the BP-ANN model performed very well for capturing the variation pattern when PAN concentrations were low. The findings of this study address existing gaps in this research field and provide a theoretical basis for future regional air pollution control.

Acknowledgment

This work was supported by the "State Key R & D Program" of China. (Nos. 2017YFC0212400, 2016YFC0202200).

REFERENCES

Akkoyunlu, A., Yetilmezsoy, K., Erturk, F., Oztemel, E., 2010. A neural network-based approach for the prediction of urban SO_2 concentrations in the Istanbul metropolitan area. Int. J. Environ. Pollut. 40, 301–321.

Baawain, M.S., Al-Serihi, A.S., 2014. Systematic approach for the prediction of ground-level air pollution (around an industrial port) using an artificial neural network. Aerosol air Qual. Res. 14 (1), 124–134.

Baawain, M.S., El-Din, M.G., Smith, D.W., 2007. Artificial neural networks modeling of ozone bubble columns: mass transfer coefficient, gas hold-up, and bubble size. Ozone Sci. Eng. 29 (5), 343–352.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon Press.

Cai, M., Yin, Y., Xie, M., 2009. Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. Transp Res D Trans Environ. 14 (1), 32–41.

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

Chaloulakou, A., Grivas, G., Spyrellis, N., 2003b. Neural network and multiple regression models for PM_{10} prediction in Athens: a comparative assessment. Air Waste management Assoc. 53 (10), 1183–1190.

Chattopadhyay, S., Chattopadhyay, G., 2012. Modeling and prediction of monthly total oozone concentrations by use of an artificial neural network based on principal component analysis. Pure Appl. Geophys. 169, 1891–1908.

Faris, H., Alkasassbeh, M., Rodan, A., 2014. Artificial neural networks for surface ozone prediction: models and analysis. Pol. J. Environ. Stud. 23, 341–348.

Fischer, E.V., Jacob, D.J., Yantosca, R.M., Sulprizio, M.P., Millet, D.B., Mao, J., et al., 2014. Atmospheric peroxyacetyl nitrate (PAN): a global budget and source attribution. Atmos. Chem. Phys. 14 (5), 2679–2698.

Gao, T., Li, H., Wang, B., Yang, G., Xu, Z., Zeng, L., et al., 2014. Peroxyacetyl nitrate observed in Beijing in august from 2005 to 2009. J. Environ. Sci. 26 (10), 2007–2017.

Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perception)—a review of applications in the atmospheric sciences. Atmos. Environ. 32, 2627–2636.

de Gennaro, G., Trizio, L., Di Gilio, A., Pey, J., Perez, N., Cusack, M., et al., 2013. Neural network model for the prediction of PM10 daily concentrations at two sites in the western Mediterranean. Sci. Total Environ. 463, 875–883.

Grivas, G., Chaloulakou, A., 2006. Artificial neural network models for prediction of PM hourly concentrations, in the greater area of Athens. Greece. Atmos. Environ. 40 (7), 1216–1229.

Huang, G., 1991. Application of the theory of multivariate statistical analysis to prediction of air pollution in the urban environment. Environ. Sci. 6 (4), 69–85.

Huang, M.L., Hung, Y.H., Chen, W.Y., 2010. Neural network classifier with entropy based feature selection on breast cancer diagnosis. J. Med. Syst. 34 (5), 865–873.

Karacan, O.C., 2007. Development and application of reservoir models and artificial neural networks for optimizing ventilation air requirements in development mining of coal seams. Int. J. Coal Geol. 72, 221–239.

Karacan, O.C., 2008. Modeling and prediction of ventilation methane emissions of U.S. longwall mines using supervised artificial neural networks, Int. J. Coal Geol. 73, 371–387.

Kleindienst, T.E., 1994. Recent developments in the chemistry and biology of peroxyacetyl nitrate. Res. Chem. Intermed. 20 (3–5), 335–384.

Lei, X.N., 1998. Air Pollution Numerical Forecasting Foundations and Patterns. China Meteorological Press.

Lu, W.Z., Wang, W.J., Wang, X.K., Yan, S.H., Lam, J.C., 2004. Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air. Hong Kong. Environ. Res. 96 (1), 79–87.

Nelson, M.M., Illingworth, W.T., 1991. A Practical Guide to Neural Nets (Addison-Wesely).

Ozkan, G., Ucan, L., Ozkan, G., 2010. The prediction of SO_2 removal using statistical methods and artificial neural network. Neural Comput. Appl. 19, 67–75.

Payne, V., Alvarado, M., Cadypereira, K.E., Worden, J., Kulawik, S. S., Fischer, E.V., 2013. Global satellite retrievals of Peroxy acetyl nitrate (PAN) in the troposphere// proceedings of the AGU fall meeting. San Francisco 2013, 0182.

Pires, J.C.M., Goncalves, B., Azevedo, F.G., Carneiro, A.P., Rego, N., Assembleia, A.J.B., et al., 2012. Optimization of artificial neural network models through genetic algorithms for surface ozone concentration forecasting. Environ. Sci. Pollut. Res. 19, 3228–3234.

Ripley, B.D., 1996. Pattern Recognition and Neural Nets. Cambridge University Press.

Russo, A., Raischel, F., Lind, P.G., 2013. Air quality prediction using optimal neural networks with stochastic variables. Atmos. Environ. 79, 822–830.

Sánchez, A.B., Ordóñez, C., Lasheras, F.S., Juez, F.J.D.C., Roca-Pardiñas, J., 2013. Forecasting SO_2 pollution incidents by means of Elman artificial neural networks and ARIMA models. Abstr. Appl. Anal. 2013 (3), 1728–1749.

Seinfeld, J.H., Pandis, S.N., 2012. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. John Wiley.

Singh, H.B., Salas, L.J., Viezee, W., 1986. Global distribution of peroxyacetyl nitrate. Nature 321 (6070), 588–591.

Sousa, S.I.V., Martins, F.G., Pereira, M.C., Alvim-Ferraz, M.C.M., 2006. Prediction of ozone concentrations in oporto city with statistical approaches. Chemosphere 64 (7), 1141–1149.

Sousa, S.I.V., Martins, F.G., Alvim-Ferraz, M.C.M., Pereira, M.C., 2007. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. Environ. Modelling and Software 22 (1), 97–103.

Stephens, E.R., Hanst, P.L., Doerr, R.C., Scott, W.E., 1956. Reactions of nitrogen dioxide and organic compounds in air. Ind. Eng. Chem. 48 (9), 1498–1504.

Vyskocil, A., Viau, C., Lamy, S., 1998. Peroxyacetyl nitrate: review of toxicity. Hum. Exp. Toxicol. 17 (4), 212–220.

Zhang, J.M., Wang, T., Ding, A.J., Zhou, X.H., Xue, L.K., Poon, C.N., et al., 2009. Continuous measurement of peroxyacetyl nitrate (pan) in suburban and remote areas of western China. Atmos. Environ. 43 (2), 228–237.

Zhang, H., Liu, Y., Shi, R., Yao, Q.C., 2013. Evaluation of PM_{10} forecasting based on an artificial neural network model and intake fraction in an urban area: a case study in Taiyuan City. China. J. Air Waste Manage. Assoc. 63, 755–763.

Zhang, H., Xu, X., Lin, W., Wang, Y., 2014. Wintertime peroxyacetyl nitrate (pan) in the megacity Beijing: role of photochemical and meteorological processes. J. Environ. Sci. 26 (1), 83–96.

Zhang, G., Mu, Y., Zhou, L., Zhang, C., Zhang, Y., Liu, J., et al., 2015. Summertime distributions of peroxyacetyl nitrate (pan) and peroxypropionyl nitrate (ppn) in Beijing: understanding the sources and major sink of pan. Atmos. Environ. 103 (1), 289–296.

Zhang, B., Zhao, B., Zuo, P., Huang, Z., Zhang, J., 2017. Ambient peroxyacyl nitrate concentration and regional transportation in Beijing. Atmos. Environ. 166, 543–550.