SSA-LSTM neural network for hourly PM2.5 concentration prediction in Shenyang, China

Ye Zhang¹ and Weide Li¹,*

¹School of Mathematics and Statistics, Lanzhou University, China

*Email: weideli@lzu.edu.cn

Abstract. Atmospheric environment has become the focus of global attention. Fine particulate matter has posed a serious threat to human respiratory system. In order to effectively control atmospheric environment and protect human health, pollutant prediction has become a necessary work for human survival and development. In order to improve the accuracy of PM2.5 prediction, this paper developed a new combined prediction model — SSA-LSTM. Firstly, the observed time series are decomposed into periodic component and noise component by SSA (Singular spectrum analysis). Then, LSTM (Long short-term memory) neural network was used to forecast the decomposed components. Finally, the predicted results of different components are integrated to generate the final predicted results. The results show that the proposed model has a significant improvement in the accuracy of prediction.

1. Introduction

With the rapid development of economy, the atmospheric environment has become the focus of global attention, especially the pollutants represented by PM2.5 have caused serious harm to the atmospheric environment and human health ([1]). PM2.5, also known as fine particulate matter, refers to particulate matter with aerodynamic equivalent less than or equal to 2.5 microns in the air ([2, 3]). Fine particulate matter in the air is easy to enter the body through respiration and deposit in the alveolar, which is difficult to get out of the body. Long time deposition of fine particulate matter in the lung will cause serious damage to the respiratory system and lead various respiratory diseases ([4, 5]). According to China's latest environmental announcement in 2018, 64.2 percent of the 338 cities (Prefecture-level cities and above) in the country had excessive air quality. Among the days with heavy air pollution, PM2.5 accounted for 60 percent. Therefore, PM2.5 concentration prediction is particularly important to air pollution control and public health protection ([6]).

The existing research methods for air prediction are mainly divided into two methods: empirical model and deterministic model. Empirical models use various statistical and/or machine learning techniques to quantify the underlying complex relationships between air pollutants and potential predictors based on large numbers of data sets under various atmospheric conditions ([7]). Deterministic model refers to the use of physical and chemical knowledge to describe the formation of pollutants or transmission process. Although this kind of method does not require a large amount of observation data, it is usually based on a number of empirical assumptions and parameters so that that method might not be applicable to all urban environments.

At present, using statistics method or machine learning to establish empirical model for pollutants concentration prediction is increasing in recent years. The typical work can summary to three
categories: single model, decomposed model and combined model. The single model mainly includes some statistical and artificial intelligent models. Among statistical methods, MLR (multivariate linear regression) can be used to predict $O_3$ and NO$_2$ concentrations ([8]). But, in the aspect of prediction accuracy, NLR (nonlinear regression) can express a better prediction effect. Cobourn used the NLR model to predict PM$_{2.5}$ concentration, and added the variable PM24 (24-h backward track PM$_{2.5}$ concentration) into the prediction model ([9]). Then, this NLR model was used to predict the concentration of PM$_{2.5}$ and $O_3$ in three large Chinese cities, meantime, the above variable PM24 and the new variable TWS (trajectory wind speed) were added to the prediction model ([10]). Besides, the GM (1,1) (grey differential equation model) was used to predict PM$_{10}$ and PM$_{2.5}$ concentrations ([11]), and ARIMA (auto regressive integrated moving average) model was used to predict PM$_{10}$ concentrations ([12]). Among the single artificial intelligence methods, some classic models are widely used, such as random forest model ([13]) and multi-layer perceptron model ([7]). Both of these two methods improve the forecasting accuracy by adding the variable of AOD (aerosol optical depth). In addition, gradient boosted tree (GBT) is directly used as a peak PM$_{2.5}$ concentration predicting model ([14]). It can be seen from above that the selection of input variables is particularly important for the single model forecasting accuracy.

Except for the single model mentioned above, decomposition prediction method is used occasionally. It is a prediction method based on signal decomposition and the prediction process can be divided into three steps. First, the original time series to be predicted is decomposed into different components by signal decomposition algorithm. Then, these components are respectively predicted using predictive model. Finally, the predicted results are re-integrated to generate the final predicted results. Among numerous signal decomposition methods, EEMD (ensemble empirical mode decomposition) is the most widely used decomposition method. Bai proposed an improved ensemble long short-term memory neural network for hourly PM$_{2.5}$ concentration forecasting ([15]). And, the EEMD decomposition is also used to predict the PM$_{2.5}$ concentration of the following day ([16]).

Among all the prediction models of pollutants, the most widely used model is the combined model. It combines different algorithms as predictor and most of these predictors are basing on the artificial intelligent method. Bai developed a W-BPNN model using wavelet technique and back propagation neural network (BPNN) to forecast daily air pollutants (PM$_{10}$, SO$_2$, and NO$_2$) concentrations ([17]). Xiao combined air mass trajectory analysis and wavelet transformation to improve the artificial neural network (ANN) forecast accuracy of daily average concentrations of PM$_{2.5}$ two days in advance ([18]). Basing on the classical model of SVR (support vector regression), GWO (grey wolf optimizer) was used to improve the prediction model of SVR and realize the prediction of PM$_{2.5}$ concentration ([19]). And, an optimal-combined SVR model based on CEEMD (complementary ensemble empirical mode decomposition), PSO-GSA (particle swarm optimization and gravitational search algorithm) and PSO (particle swarm optimization) was used to forecast AQI (air quality index) in five cities of China ([20]). Besides, a stacked ensemble model was developed for forecasting and analyzing the daily average concentrations of PM$_{2.5}$ in Beijing, China ([21]). Nowadays, the most widely used prediction models are the methods based on LSTM and CNN (convolutional neural network) ([22, 23]). These methods mainly construct different structures of CNN to extract features and use LSTM to predict pollutant concentrations.

Nowadays, some prediction methods connect multiple air quality monitoring stations in order to consider spatial and temporal information. For example, multiple LSTM neural network add full connection layer to forecast PM$_{2.5}$ concentration ([24, 25]) or multiple output model based on SVR to forecast PM$_{2.5}$ concentration ([26]). However, in some cases, this approach may not be practical when the air quality monitoring stations are not highly correlated or have few stations. Hence, in this paper, a new prediction method SSA-LSTM is proposed, which can improve the accuracy of prediction and increase the application range of the model without adding spatial information. SSA (singular spectrum analysis) is a method of signal decomposition, which can decompose the original time series into noise components and other components, while LSTM (Long short-term memory) neural network has the characteristics of memory and can deal with the long-term dependence of time series.
Therefore, this paper chooses this method and uses two representative air quality monitoring stations for prediction and obtains better prediction results.

2. Research area and data

Shenyang is the capital of Liaoning province, locating in the northeast of China with a total area of 12,948 square kilometers and an urban area of 3,495 square kilometers. Since 1951, the extreme maximum temperature in Shenyang was 38.3°C and the extreme minimum temperature was -32.9°C. It is of great significance to take Shenyang as the research area. On the one hand, as China's heavy industry base, Shenyang has serious industrial emissions. On the other hand, the huge temperature difference between four seasons determines the increase of fuel heating in winter, which also aggravates the environmental pollution. Therefore, the prediction of PM$_{2.5}$ concentration in Shenyang is of great significance to pollution forecasting and health protection.

The air quality data used in this paper comes from web: http://www.envicloud.cn, which provides various environmental data for scientific research. We select hourly data from nine monitoring stations in Shenyang and each data include seven pollutants: AQI, PM$_{10}$, PM$_{2.5}$, NO$_2$, SO$_2$, CO and O$_3$. Among them, two stations to be predicted are denoted as C1 and C2 respectively, while others are denoted as A1, A2, A3, A4, A5, A6, A7. The meteorological data used in the paper were provided by Shenyang meteorological monitoring center and recorded every hour. Specific variables included: temperature(T), relative humidity (RH), wind speed (WS) and wind direction (WD). The two stations (C1, C2) are selected as the sites to be predicted in this paper because they are obviously different from each other. Station C1 is far away from the city center, with few surrounding sites, and has little correlation with other sites. It can be regarded as an isolated site. Station C2 is located in the city center, close to other sites, and the correlation between sites is strong. So, the choice of these two sites dedicated to prediction is more illustrative.

3. Methodologies

In this paper, long short-term memory neural network is used to predict PM$_{2.5}$ hourly concentration. LSTM neural network is a variant of recurrent neural network (RNN), which overcome the inability of RNN to deal with long-distance dependence well and has the characteristics of long-term memory. Meanwhile, we also combine SSA method into the prediction model. SSA can decompose the original time series into sub-sequences, including trend, period, quasi-periodic and noise. Next, LSTM neural network and SSA are introduced in detail.

3.1. LSTM neural network

LSTM neural network is a variant of recurrent neural network, which is widely used in time series prediction. However, as the increasing for past information dependence, the gradient disappears or explodes will appear. Therefore, LSTM neural network adds special cell structure based on the original neural network, which effectively solves this problem and can achieve the effect of the long-term prediction. The three added cell structures called Forget gate, Input gate and Output gate can be described as follows ([27]):

1) Forget gate (FG): it is used to calculate which information needs to be forgotten.

\[ f(t) = \sigma(W_f[h_{t-1}, x_t] + b_f) \] \( \#(1) \)

where \( h_{t-1} \) represents the output of previous cell state, \( x_t \) represents the input of current cell state, \( W_f \) and \( b_f \) are the weights and bias of the forget gate. The degree of information retention depends on the value of \( f(t) \) with the range [0,1] (‘0’ full fail and ‘1’ full pass).

2) Input gate (IG): it is used to calculate which information is saved into the state unit, divided into two parts. The first part is:

\[ i(t) = \sigma(W_i[h_{t-1}, x_t] + b_i) \] \( \#(2) \)

This section can be viewed as how much of the current input needs to be saved to the cell state;
The second part is:

\[ \bar{c}_{(t)} = \tanh \left( W_c \cdot [h_t-1, x_t] + b_c \right), #(3) \]

This section can be viewed as the new information generated by the current input to be added to the unit state and combined to create a new memory. The unit state at the current moment is the product of the input gate and the state at the previous moment plus the product of the two parts of the input gate, namely:

\[ C_{(t)} = f_{(t)} \cdot C_{(t-1)} + i_{(t)} \cdot \bar{c}_{(t)}, #(4) \]

where \( W_i \) and \( b_i \) are the weights and bias of the input gate, \( W_c \) and \( b_c \) are the weights and bias of the cell state, and “\(*\)” means pointwise multiplication.

(3) Output gate (OG): it is used to calculate what information needs to be output through \textit{sigmoid} function, then multiply the value of the current cell state through \textit{tanh} function to get the output.

\[ O_{(t)} = \sigma (W_o \cdot [h_{t-1}, x_t] + b_o), #(5) \]

\[ h_{(t)} = O_{(t)} \cdot \tanh (C_{(t)}), #(6) \]

where \( W_o \) and \( b_o \) are the weights and bias of the output gate. \( O_{(t)} \) is used to evaluate which part of cell state to be exported, and \( h_{(t)} \) calculates the final outputs.

3.2. Singular spectrum analysis

Singular spectrum analysis is a non-parametric algorithm, which is widely used in nonlinear time series analysis. First, it constructs the trajectory matrix based on the observed time series. Then, decomposing and reconstructing the trajectory matrix is used to extract different components of the original time series. In details, the algorithm includes four steps: embedding, singular value decomposition, regrouping and reconstruction.

(1) Embedding

Transform the original time series \( X = (X_1, X_2, \ldots, X_N) \) into the \( L \)-dimensional trajectory matrix \( Y = (Y_1, Y_2, \ldots, Y_L)^T \), \( 1 < L < N + 1 \). Hence, each element in the matrix \( Y \) can be defined as \( Y_i = (X_i, X_{i+1}, \ldots, X_{i+N-L})^T \), \( i = 1, 2, \ldots, L \) and the trajectory matrix \( Y \) can be defined as follows:

\[
Y = \begin{pmatrix}
Y_1^T \\
Y_2^T \\
\vdots \\
Y_L^T
\end{pmatrix} = \begin{pmatrix}
X_1 & X_2 & \cdots & X_{N-L+1} \\
X_2 & X_3 & \cdots & X_{N-L+2} \\
\vdots & \vdots & \ddots & \vdots \\
X_L & X_{L+1} & \cdots & X_N
\end{pmatrix} #(7)
\]

(2) Singular value decomposition

The trajectory matrix \( Y \) can be decomposed into \( d \) components, where \( d = \text{rank}(Y) \) is the number of non-zero singular values of the original matrix \( Y \). Through EAM (eigenvalue analysis method), the eigenvalues and eigenvectors of matrix \( YY^T \) can be obtained in descending order of eigenvalue \( \lambda_i \), \( \lambda_1 > \lambda_2 > \cdots > \lambda_d > 0 = \lambda_{d+1} = \cdots = \lambda_L \). At the same time, \( u_i \) is eigenvector corresponding to \( \lambda_i \), \( i = 1, 2, \cdots, L \). So, the relationship between \( YY^T \) and \( \lambda_i, u_i \) can be written as:

\[
YY^T u_i = \lambda_i u_i, \quad i = 1, 2, \ldots, L #(8)
\]

Similarly, we can calculate the eigenvalues and eigenvectors of matrix \( Y^T Y \) and write the relationship between them:

\[
Y^T Y v_j = \lambda_j v_j, \quad j = 1, 2, \ldots, n - L + 1 #(9)
\]

Among them, \( \lambda_1 > \lambda_2 > \cdots > \lambda_d > 0 = \lambda_{d+1} = \cdots = \lambda_{n-L+1} \).

So, the matrix \( Y \) can be expressed as follows:
\[
Y = U_{L \times L} \begin{pmatrix}
\sqrt{\lambda_1} & 0 & \cdots & 0 & \cdots & 0 \\
0 & \sqrt{\lambda_2} & \cdots & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\
0 & 0 & \cdots & \sqrt{\lambda_d} & \cdots & 0 \\
\vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & \cdots & 0
\end{pmatrix} V_{(N-L+1) \times (N-L+1)}^T, \tag{10}
\]

where \(U = (u_1, u_2, \cdots, u_L), \|u_i\| = 1 \) for \(1 \leq i \leq L\); \(V = (v_1, v_2, \cdots, v_{N-L+1}), \|v_j\| = 1 \) for \(1 \leq j \leq N - L + 1\).

Therefore, we can further rewrite the trajectory matrix:

\[
Y = \sum_{i=1}^{d} \sqrt{\lambda_i} u_i v_i^T = Y_1 + Y_2 + \cdots + Y_d. \tag{11}
\]

(3) Regrouping

The first \(m\) of these \(d\) components in \(Y = Y_1 + Y_2 + \cdots + Y_m\) are determined as trend information. In particular, \(I = \{I_1, I_2, \cdots, I_m\}\), \(Y_I = Y_{I_1} + Y_{I_2} + \cdots + Y_{I_m}\), where \(Y_I\) means the trend component extracted from the original sequence, and the remaining \(d - m\) components constitute the noise sequence.

(4) Reconstruction

Through matrix diagonalization, \(Y_I\) is converted to corresponding time series data. That is, \(\{Y_{I_1}, Y_{I_2}, \cdots, Y_{I_m}\}\) is converted to \(\{X_{I_1}, X_{I_2}, \cdots, X_{I_m}\}\) separately. So, the original time series can be expressed as:

\[
X = X_{\text{trend}} + X_{\text{noise}} = X_{I_1} + X_{I_2} + \cdots + X_{I_m} + X_{\text{noise}}, \tag{12}
\]

4. Results and discussion

In this section, we use LSTM neural network to predict \(F0\) and \(F1\) after SSA decomposition of original time series, where \(F0\) represents trend component plus periodic component, \(F1\) represents the noise component. Then, recombine the predicted results as the final prediction results. In the paper, the time span of the data used by all models is 2016-01-01 00:00 - 2016-12-31 23:00 including 366 days or 8,784 hours in total. The ratio of sample training set and test set is 7:3, namely: 8784*0.7=6150 samples are used as training set, and other samples are used as test set. In order to evaluate the prediction performance of the model, the following three criteria are used to evaluate the results, including RMSE, MAE and CC. The specific formulas are shown in Table 1. Meanwhile, in order to compare the prediction performance of the model, we selected three models as comparative models, LSTM (Long short-term memory), MLP (Multi-layer perceptron) and LSTM-FC (Long short-term memory-fully connection) respectively.

| Criteria                  | Formula                                      |
|---------------------------|----------------------------------------------|
| Root absolute error (RMSE)| \(RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y_i^*)^2}\) |
| Mean absolute error (MAE) | \(MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - y_i^*|\) |
| Correlation coefficient (CC) | \(R = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(y_i^* - \bar{y}^*)}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} (y_i^* - \bar{y}^*)^2}}\) |

In this paper, these three models are adopted as comparison models because LSTM model, as a
variant of RNN, is widely used in time series prediction. When considering spatial information, adding full connection layer is a widely used method ([24, 25]). Therefore, LSTM-FC is also applied in this paper as a comparison model. MLP model is a classic method in PM$_{2.5}$ prediction field. Many models using neural network for prediction are improved on the basis of MLP and have achieved good prediction effect.

4.1. Results of SSA - LSTM model
Figure 1 shows the prediction results of PM$_{2.5}$ concentration in the next 3 hours using SSA - LSTM model. Sub-graphs (a), (b) and (c) represent the prediction results of station C1 in three hours. The graph above shows the line graph of true value and predicted value change, while the graph below shows the scatter graph of true value and predicted value. Sub-graphs (d), (e) and (f) are the prediction results of station C2 in three hours, and the upper and lower figures have the same meaning as C1. As can be seen from the figure 1, the prediction result of the next hour is very good according to SSA-LSTM model. As can be seen from the line graph, except for a very small part (2 hours), PM$_{2.5}$ concentration suddenly rises and the prediction error is large, the prediction result of other time is very good. As can be seen from the scatter diagram, the predicted value is highly correlated with the real value.

Meanwhile, table 2 shows the statistical indicators of the predicted results for the next three hours. It can be seen from table 2 that, for the predicted results of station C1 for next 1h, RMSE is 6.632, MAE is 2.779, and CC is 0.991. The prediction results of station C2 for next 1h, RMSE is 6.365, MAE is 2.636 and CC is 0.994.

![Figure 1. Forecasting results and scatter plots with SSA - LSTM model. (a), (b), (c) next 3 hours for C1 and (d), (e), (f) next 3 hours for C2.](image)

| Station | Time | RMSE  | MAE   | CC    |
|---------|------|-------|-------|-------|
| C1      | 1h   | 6.632 | 2.779 | 0.991 |
|         | 2h   | 17.756| 9.033 | 0.936 |
|         | 3h   | 28.101| 15.384| 0.831 |
| C2      | 1h   | 6.365 | 2.636 | 0.994 |
|         | 2h   | 20.354| 11.001| 0.938 |
|         | 3h   | 33.002| 18.456| 0.828 |

4.2. Comparison study
To verify the efficiency and accuracy of proposed approach, three compared models LSTM, MLP and LSTM-FC are employed for hourly PM$_{2.5}$ concentration forecasting in 3 hours. The three comparison
models use the same training set and test set, so the partition ratio is 7:3. For LSTM model, the input of the model is 4 meteorological variables and 7 air quality variables and selects the best lag time. For MLP model, the input is the same as the input to LSTM. However, for LSTM - FC, first, LSTM is used to predict multiple stations. Then, a fully connected neural network is used to connect the predicted results to generate the final predicted results for every station. Therefore, the correlation between PM$_{2.5}$ concentrations among stations should be considered, and the stations with the strongest correlation should be selected as the prediction input. Table 3 shows the correlation coefficients for the two stations C1, C2 with other stations.

Table 3. Correlation coefficient of PM$_{2.5}$ concentration between stations.

| station | A1  | A2  | A3  | A4  | A5  | A6  | A7  | C1  | C2  |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| C1      | 0.754 | 0.783 | 0.881 | 0.771 | 0.783 | 0.792 | 0.748 | 1.000 | 0.793 |
| C2      | 0.801 | 0.868 | 0.812 | 0.900 | 0.947 | 0.877 | 0.860 | 0.793 | 1.000 |

Table 4 shows the comparison results of evaluation indexes predicted by different models. Compared with other models, SSA-LSTM model has obvious advantages. However, through the comparison between LSTM model and LSTM-FC, it can be seen that site C1's prediction result by using LSTM model is better than that of LSTM-FC model, which is the same as previously considered. When there are few sites or the correlation of sites is not strong, there will be a large error in using spatial information for prediction, and this method will be limited. Because C1 has a weak correlation with other sites, the addition of full connection layer cannot correct the prediction error and it will increase the prediction error. For C2 site, it has a strong correlation with other sites, adding full connection will improve the prediction results of LSTM model.

By comparing MLP model with LSTM model, it can be found that, with the increase of prediction time, the prediction effect of MLP model is better than that of LSTM model, which also indicates that MLP can show better effect for prediction, which also provides basis for future model improvement and further research.

Table 4. Comparison of forecasting performances using different models.

| site | Time | 1h | 2h | 3h |
|------|------|----|----|----|
|      | Methods | RMSE | MAE | CC | RMS E | MAE | CC | RMSE | MAE | CC |
| C1   | LSTM   | 15.837 | 8.285 | 0.949 | 26.732 | 14.855 | 0.848 | 33.443 | 18.95 | 0.749 |
|      | LSTM - FC | 16.257 | 8.336 | 0.951 | 27.924 | 16.600 | 0.848 | 34.324 | 19.29 | 0.758 |
|      | MLP     | 17.973 | 9.565 | 0.934 | 25.871 | 14.342 | 0.859 | 31.583 | 18.23 | 0.780 |
|      | SSA - LSTM | 6.632 | 2.779 | 0.991 | 17.756 | 9.033 | 0.936 | 28.101 | 15.38 | 0.831 |
| C2   | LSTM   | 17.123 | 9.505 | 0.957 | 30.883 | 17.700 | 0.851 | 40.844 | 24.88 | 0.719 |
|      | LSTM - FC | 16.982 | 9.124 | 0.956 | 28.709 | 15.955 | 0.868 | 37.129 | 21.99 | 0.766 |
|      | MLP     | 19.864 | 10.303 | 0.941 | 29.432 | 16.967 | 0.866 | 37.009 | 22.23 | 0.777 |
|      | SSA - LSTM | 6.365 | 2.636 | 0.994 | 20.354 | 11.001 | 0.938 | 33.002 | 18.45 | 0.828 |

Therefore, it can be concluded that the SSA-LSTM model proposed in this paper is effective for the prediction of PM$_{2.5}$ concentration hourly. Under the condition that spatial information cannot be used, the model can also be used to achieve a good prediction effect.

5. Conclusions
In this paper, the proposed SSA-LSTM model is effectively proved to improve the accuracy of prediction. Meanwhile, by comparing LSTM model with LSTM-FC model, the problem of adding full
connection for spatial sites is not applicable in some cases, and further points out that proposed model in the paper is more widely applicable.

**Acknowledgements**

This study was supported by the National Key Research and Development Program of China (NO. 2019YFA0607104) and National Nature Science Foundation of China (NO.41775069,41975076).

**References**

[1] J.S. Apte, J.D. Marshall, A.J. Cohen, M. Brauer, Environ. Sci. Technol. 49 (13), 8057-8066 (2015)

[2] A.J. Cohen, H. Ross Anderson, B. Ostro, K.D. Pandey, M. Krzyzanowski, N. Kunzli, K. Gutschmidt, A. Pope, I. Romieu, J.M. Samet, K. Smith, J. Toxicol. Environ. Health 68 (13-14), 1301-1307 (2005)

[3] S.W. Wu, D. Yang, H.Y. Wei, B. Wang, J. Huang, H.Y. Li, M. Shima, F.R. Deng, X.B. Guo, Chemosphere 135, 347-353 (2015)

[4] Y.Z. Xu, W.D. Yang, J.Z. Wang, Atmos. Environ. 148, 239-257 (2016)

[5] H.F. Yu, A.L. Stuart, Sci. Total Environ. 576, 148-158 (2017)

[6] Y. Huang, H.Z. Shen, H. Chen, R. Wang, Y.Y. Zhang, S. Su, et al. Environ. Sci. Technol. 48 (23), 13834-13843 (2014)

[7] L. Hrust, Z.B. Klaic, J. Krizan, O. Antonic, P. Hercog, Atmos. Environ. 43 (35), 5588–5596 (2009)

[8] E. Agirre-Basurko, G. Ibarra-Berastegi, I. Madariaga, Environ. Model. Softw. 21 (4), 430-446 (2006)

[9] W.G. Cobourn, Atmos. Environ. 44 (25), 3015-3023 (2010)

[10] B.L. Lv, Y.Q. Bai, W.G. Cobourn, Atmos. Environ. 147, 209-223 (2016)

[11] L. Chen, T.Y. Pai, Atmos. Pollut. Res. 6 (4), 572-580 (2015)

[12] L. Jian, Y. Zhao, Y.P. Zhu, M.B. Zhang, D. Bertolatti, Sci. Total Environ. 426 (2), 336-345 (2012)

[13] X.T. Li, X.D. Zhang, Environ. Pollut. 249, 735-749 (2019)

[14] G. Miskell, W. Pattinson, L. Weissert, D. Williams, J. Environ. Manage. 242, 56-64 (2019)

[15] Y. Bai, B. Zeng, C. Li, J. Zhang, Chemosphere 222, 286-294 (2019)

[16] M.F. Niu, K. Gan, S.L. Sun, F.Y. Li, J. Environ. Manage. 196, 110-118 (2017)

[17] Y. Bai, Y. Li, X.X. Wang, J.J. Xie, C. Li, Atmos. Pollut. Res. 7, 557-566 (2016)

[18] X. Feng, Q. Li, Y.J. Zhu, J.X. Hou, L.Y. Jin, J.J. Wang, Atmos. Environ. 107, 118-128 (2015)

[19] M.F. Niu, Y.F. Wang, S.L. Sun, Y.W. Li, Atmos. Environ. 134, 168-180 (2016)

[20] S.L. Zhu, L. Yang, W.N. Wang, X.R. Liu, M.M. Lu, X.P. Shen, Environ. Pollut. 243, 842-850 (2018)

[21] B.X. Zhai, J.G. Chen, Sci. Total Environ. 635, 644-658 (2018)

[22] Y.L. Qi, Q. Li, H. Karimian, D. Liu, Sci. Total Environ. 664, 1-10 (2019)

[23] C.C. Wen, S.F. Liu, X.J. Yao, L. Peng, X. Li, Y. Hu, T.H. Chi, Sci. Total Environ. 654, 1091-1099 (2019)

[24] J.C. Zhao, F. Deng, Y.Y. Cai, J. Chen, Chemosphere 220, 486-492 (2019)

[25] X. Li, L. Peng, X.J. Yao, S.L. Cui, Y. Hu, C.Z. You, T.H. Chi, Environ. Pollut. 231, 997-1004 (2017)

[26] Y.L. Zhou, F.J. Chang, L.C. Chang, I.F. Kao, Y.S. Wang, C.C. Kang, Sci. Total Environ. 651, 230-240 (2019)

[27] F.A. Gers, J. Schmidhuber, F. Cummins, Neural Comput. 12 (10), 2451-2471 (2000)