

УДК 004.032.26

Minglei Fu¹, Chen Wang¹, Zichun Le¹, Dmytro Manko²

¹College of Sciences, Zhejiang University of Technology
288 Liuhe Road, 310023 Hangzhou, China

²Institute for Information Recording, National Academy of Sciences of Ukraine
M. Shpak, 2 st., 03113 Kyiv, Ukraine

Double layer back propagation neural network based on restricted Boltzmann machines for forecasting daily particulate matter 2.5

Particulate matter 2.5 (PM_{2.5}) pollution is an actual problem in the modern world and forecasting of the daily concentration of PM_{2.5} is a challenging task for researchers. In this study, a novel neural network model that effectively forecasts daily PM_{2.5} in Hangzhou city was developed in the form of a restricted Boltzmann machines double layer back propagation neural network model (RBM-DL-BPNN). Air quality index, the air pollutants, e.g., particulate matter 10 (PM₁₀), PM_{2.5}, SO₂, CO, NO₂, O₃, and meteorological parameters (temperature, dew point, humidity, pressure, wind speed, and precipitation) of Hangzhou city were used in this study to train and test three models: RBM-DL-BPNN, double layer back propagation neural network (DL-BPNN), and back propagation neural network (BPNN). The results of experiments and analyses performed indicate that RBM-DL-BPNN has a smaller mean absolute percent error (MAPE), smaller overall daily absolute percentage errors, and more results in terms of absolute percentage error within the range 0–50 % than DL-BPNN and BPNN.

Key words: *Neural network, restricted Boltzmann machines, Particulate matter 2.5, Forecasting.*

1. Introduction

With sustainable growth of a social economy and rapid expansion of urban populations, urban air pollution problems have become increasingly serious [1, 2]. Among various kinds of air pollutants, particulate matter 2.5 (PM_{2.5}) is the main pollution in Hangzhou city [3]. Nowadays, merely measuring urban PM_{2.5} is not enough. Understanding the development trend of PM_{2.5} concentration to prevent air pollution in cities and guarantee the health of urban residents is a task of vital importance.

In recent years, artificial neural networks (ANNs) have been proven effective in forecasting trends in air pollution such as predicting CO ambient concentration [4] and

predicting hourly $PM_{2.5}$ concentration [5]. Furthermore, various models have successfully optimized ANNs [6–20]. In general, the following methods can be used to optimize ANNs: input data selection, algorithm optimization, and model combination. Some researchers have already conducted studies and analyzed the best choices for input data to optimize the effectiveness of ANNs. For example, Voukantsis *et al.* used principal component analysis (PCA) to reduce the dimension of the original input data and transform the original input dataset to a linear combination in order to optimize the artificial neural networks multilayer perceptron (ANN-MLP) model [6]. Gennaro *et al.* performed sensitivity analysis to understand the importance of different variables in developing an ANN [8]. Antanasijević *et al.* selected and optimized the input data to an ANN using a genetic algorithm [10]. In addition, some researchers have incorporated techniques such as fuzzy logic, k -means clustering, chaotic particle swarm optimization (CPSO), and wavelet transformation into neural networks to optimize ANNs. For instance, Mishra *et al.* combined a neural network and fuzzy logic to forecast $PM_{2.5}$ during haze conditions [11]. Elangasinghe *et al.* combined ANN with k -means clustering to analyze PM_{10} and $PM_{2.5}$ [13]. He *et al.* created a novel hybrid model combining ANN and CPSO to improve forecasting accuracy [14]. Siwek and Osowski combined wavelet transformation and neural network to forecast the daily average concentration of PM_{10} [15]. Researchers have also combined ANNs with other models to create new models that can be used for more accurate forecasting. For example, Perez *et al.* combined a nearest neighbor model (NNM) with ANN to improve the accuracy of PM_{10} concentration forecasting [16]. Di'z-Robles *et al.* created a novel hybrid model combining Box-Jenkins Time Series (ARIMA) and ANN and improved the forecast accuracy of particulate matter [18]. Al-Alawi *et al.* combined principal component regression (PCR) and ANN to predict ozone concentration levels in the lower atmosphere [20].

The methods cited above have been proved to effectively improve the performance of ANNs. However, the ANN models proposed in previous works were usually ANNs with only a single hidden layer. Hence, the prediction accuracy of the models might be restricted by the inherent shortcomings of the single layer ANNs. For example, it is usually difficult to optimize the weights of the neurons in single layer ANNs to obtain higher prediction accuracy. This study focused on optimization of a double layer back propagation neural network (DL-BPNN) for forecasting daily $PM_{2.5}$ using restricted Boltzmann machines (RBM). This is in contrast to the previous works cited above, which focused primarily on the input parameters and models to make the prediction of ANNs more accurate. RBM, which can learn input data features, is used to train the weights that initialize the DL-BPNN. The proposed RBM-DL-BPNN model was evaluated by comparing its results in predicting the $PM_{2.5}$ concentration in Hangzhou city to those obtained from the standard DL-BPNN and BPNN models. Further, its advantages and disadvantages were analyzed.

2. Material and methods

2.1. Data collection

In recent years, concern about the air quality of Chinese cities has been increasing. Hangzhou city covers an area of 16596 square kilometers and had an estimated population of approximately 9 million people in 2015 [21]. In 2014, there were 137 pollution

days in Hangzhou and 93 days with PM_{2.5} as the primary pollutant [3]. Therefore, the air quality daily data and the meteorological daily data of Hangzhou from December 2013 to August 2016 were both used in this study. The data used include the air quality index (AQI), concentration of PM_{2.5}, PM₁₀, SO₂, CO, NO₂, and O₃ in the air, temperature, dew point, humidity, pressure, wind speed, and precipitation. The data from December 2013 to May 2016 were used for training, whereas the data from June 2016 to August 2016 were used to test the models. The air quality daily data were collected from PM_{2.5} monitoring network websites [22] and the meteorological daily data were collected from the Weather Underground websites [23]. Before starting, the input data were normalized. Subsequently, the value of the input data was transformed into (0,1–0,9), which is helpful for training. The following equation was used:

$$x^* = 0,1 + \frac{0,8(x - x_{\min})}{x_{\max} - x_{\min}}, \quad (1)$$

where x is the original data and x_{\min} and x_{\max} are the minimum and maximum values, respectively. x^* is the normalized data.

2.2. Back propagation neural network (BPNN)

ANNs are mathematical structures consisting of a number of interconnected neurons. An ANN is able to emulate the process that people use to recognize patterns, acquire knowledge, and solve problems [1]. A BPNN is a classic neural network with three layers: an input layer, a hidden layer, and an output layer, as shown in Fig. 1. The working principle of a BPNN can be divided into two processes. In the first process, called the signal forward propagation process, the training data are introduced from the input layer, propagated through the hidden layer, and finally outputted from the output layer. The neurons in the hidden layer sum the weighted arriving signal:

$$h_{1j} = \sum_{i=1}^n w_{1ij} x_i^* + b, \quad (2)$$

where h_{1j} ($j = 1, 2, \dots, m$) is the output value of the hidden layer neuron, w_{1ij} ($i = 1, 2, \dots, n$) are the weights between the input layer and the hidden layer, x_i^* ($i = 1, 2, \dots, n$) are the normalized input data, and b is a bias value [24].

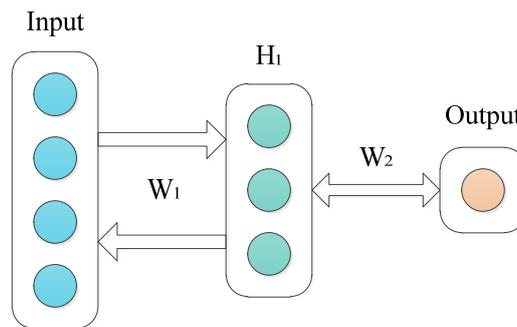


Fig. 1. Structure of a back propagation neural network (BPNN)

The second process is signal back propagation. In this process, errors in the output values and the actual values are propagated back to the hidden layer to adjust the weights between the hidden layer and the output layer. Similarly, the weights between the input layer and the hidden layer are adjusted when the error returns. BPNN is considered as being completely trained when the error has been reduced to a stable value. After training the BPNN, test data can be introduced to forecast daily PM_{2.5} concentration.

2.3. Double layer BPNN (DL-BPNN)

BPNNs can be used to solve nonlinear problems. The structure of a double layer back propagation neural network is similar to that of a BPNN. However, DL-BPNN has two hidden layers, as shown in Fig. 2. The equations utilized by the first and the second hidden layers are:

$$h'_{1j} = \sum_{i=1}^n w'_{1ij} x_i^* + b, \quad (3)$$

$$h'_{2l} = \sum_{j=1}^m w'_{2jl} h'_{1j} + b, \quad (4)$$

where h'_{1j} ($j = 1, 2, \dots, m$) is the output value of the first hidden layer; h'_{2l} ($l = 1, 2, \dots, k$) is the output value of the second hidden layer; w'_{1ij} ($i = 1, 2, \dots, n$) are the weights between the input layer and the first hidden layer; w'_{2jl} ($j = 1, 2, \dots, m$) are the weights between the first hidden layer and the second hidden layer; x_i^* ($i = 1, 2, \dots, n$) are the normalized input data, and b is the bias value.

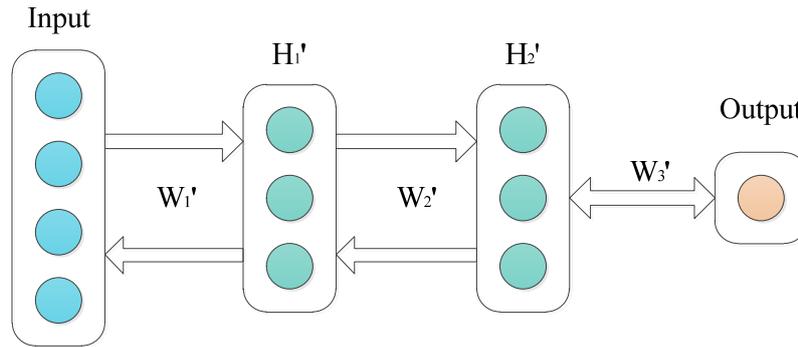


Fig. 2. Structure of a double layer back propagation neural network (DL-BPNN)

Theoretically, DL-BPNN is more useful than BPNN for solving nonlinear problems. However, DL-BPNN may in fact not be able to show its advantages because of the error attenuation during the error back propagation process. Consequently, weights adjustment between the input layer and the first hidden layer is small. However, the signal forward propagation process begins with the input layer and the first hidden layer, which means that regardless of how the second hidden layer is trained, the output values are confused by the first two layers and the entire network will operate poorly. As a result, DL-BPNNs are not widely used as BPNNs.

2.4. Double layer back propagation neural network based on restricted Boltzmann machines (RBM-DL-BPNN)

Deep neural networks (DNNs) were introduced in 2006 and have been successfully applied in speech recognition and image recognition [25, 26, 27]. RBM is an important component of any DNN, as it can learn the features of input data through unsupervised training [27]. Random weight initialization of a DL-BPNN fails with very high probability in the basin of attraction of a poor local minimum [28]. Thus, an RBM is used to pre-train the weights of DL-BPNN [29]. The training of the RBM-DL-BPNN can be divided into two parts. In the first part, RBM is used to learn the features of the input data. An RBM has a visible layer and several hidden layers but no visible-visible or hidden-hidden connections. In a binary RBM, the weights on connections and the biases of individual units define a probability distribution over the joint states of the visible and hidden units via an energy function. The energy of a joint configuration is given as follows [26]:

$$E(v, h | \theta) = - \sum_{i=1}^V \sum_{j=1}^H w_{ij} v_i h_j - \sum_{i=1}^V b_i v_i - \sum_{j=1}^H a_j h_j, \tag{5}$$

where $\theta = (w, b, a)$ and w_{ij} represents the symmetric interaction term between visible unit i and hidden unit j while b_i and a_j are their bias terms. V and H are the numbers of visible and hidden units.

The aim of RBM training is to learn the parameter $\theta = (w, b, a)$, whereas the value of parameter w is desired. The structure of RBM-DL-BPNN is shown in Fig. 3. In the figure, the unsupervised structure RBM 1 trains the weights between the input layer and the first hidden layer (W_1'') of RBM-DL-BPNN. Then, W_1'' is used as the initial weight to train the weights between the first hidden layer and the second hidden layer (W_2'') of RBM-DL-BPNN through a double layer unsupervised structure, RBM 2.

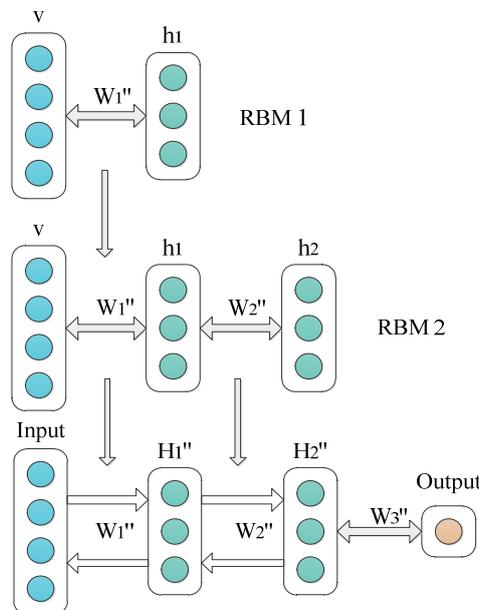


Fig. 3. Structure of the double layer back propagation neural network model based on restricted Boltzmann machines (RBM-DL-BPNN)

Then, the weights between the output layer and the second hidden layer (W_3'') are trained and W_1'' and W_2'' are fine-tuned. The W_1'' and W_2'' trained by RBM 1 and RBM 2 are used as the initial weights of the first two layer weights of the RBM-DL-BPNN model. Then, we use supervised training to finally train W_3'' and adjust W_1'' and W_2'' .

3. Results and discussion

3.1. Results

The efficiency of all three models presented above was evaluated using efficiency indexes: root mean square error (RMSE) (6), mean absolute error (MAE) (7), and mean absolute percent error (MAPE) (8). RMSE and MAE were used to evaluate the reliability, whereas MAPE was used to evaluate the accuracy of the models [1, 18]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |t_i - y_i|^2}, \quad (6)$$

$$MAE = \frac{1}{n} \left(\sum_{i=1}^n |t_i - y_i| \right), \quad (7)$$

$$MAPE = \frac{1}{n} \left(\sum_{i=1}^n \frac{|t_i - y_i|}{t_i} \right) \cdot 100 \%, \quad (8)$$

where n is the number of data points, y_i is the predicted value, and t_i is the actual observed value.

There are 33 months from December 2013 to August 2016. The data for the first 30 months were used to train models and the data for the last three months were used to test them. Thus, 912 pieces of data were used for training and 90 pieces of data for testing. Figs. 4–6 show curves of the daily $PM_{2.5}$ concentration forecasted by three models and the actual data in June, July, and August 2016. The black, red, blue, and green curves represent actual data, RBM-DL-BPNN forecasted data, DL-BPNN forecasted data, and BPNN forecasted data, respectively. On the whole, the trends of the three models are all similar to the actual data. However, the trend of the red curve is closer to the trend of the black curve than the trends of the blue and green curves. Further, the blue curve is closer to the black curve than to the green curve, which means that, to some extent, the DL-BPNN model can forecast more accurately than the BPNN model. Further, the RBM-DL-BPNN model can forecast more accurately than both the DL-BPNN model and the BPNN model.

Table 1 shows the efficiency indexes of BPNN, DL-BPNN, and RBM-DL-BPNN in forecasting $PM_{2.5}$ from June 2 to August 31, 2016 in Hangzhou city. From Table 1 it is clear that the MAPE of RBM-DL-BPNN is lower than that of both DL-BPNN and BPNN. Further, the MAPE of DL-BPNN is lower than that of BPNN. RBM-DL-BPNN's RMSE and MAE are sometimes higher than those of the other two models; however, the difference is minimal. From Figs. 4–6 and Table 1, it is clear that the DL-BPNN model actually has several advantages over the BPNN model and that RBM is useful for the DL-BPNN model.

Table 1. Efficiency indexes of BPNN, DL-BPNN, and RBM-DL-BPNN for PM2.5 from June 2016 to August 2016 in Hangzhou city

	BPNN			DL-BPNN			RBM-DL-BPNN		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
June	12,941	10,348	36,353 %	11,498	9,179	31,941 %	12,710	9,335	28,054 %
July	9,666	7,413	31,376 %	9,753	7,613	31,295 %	10,612	8,253	26,468 %
August	9,620	7,949	40,148 %	10,981	8,930	38,912 %	7,835	6,602	29,484 %

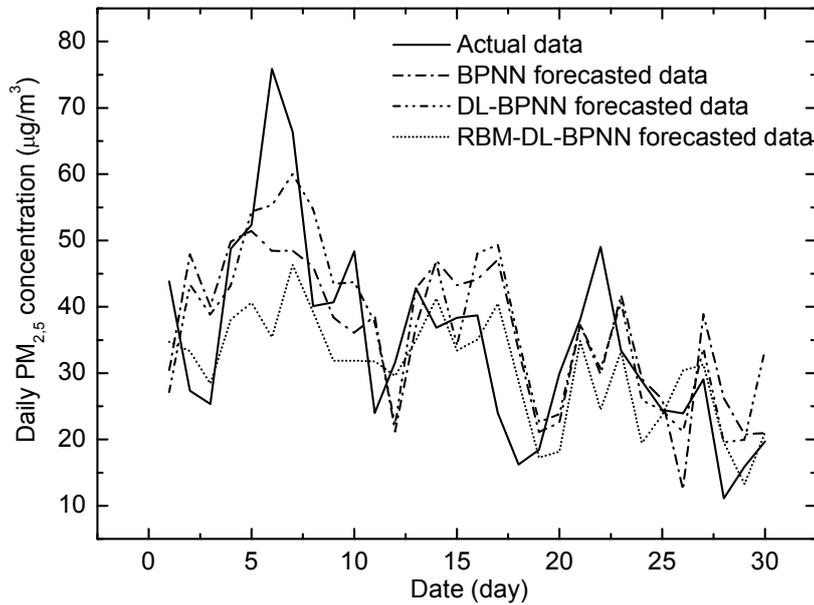


Fig. 4. Actual data for June 2016 versus the daily PM_{2.5} concentration forecasted by the three models

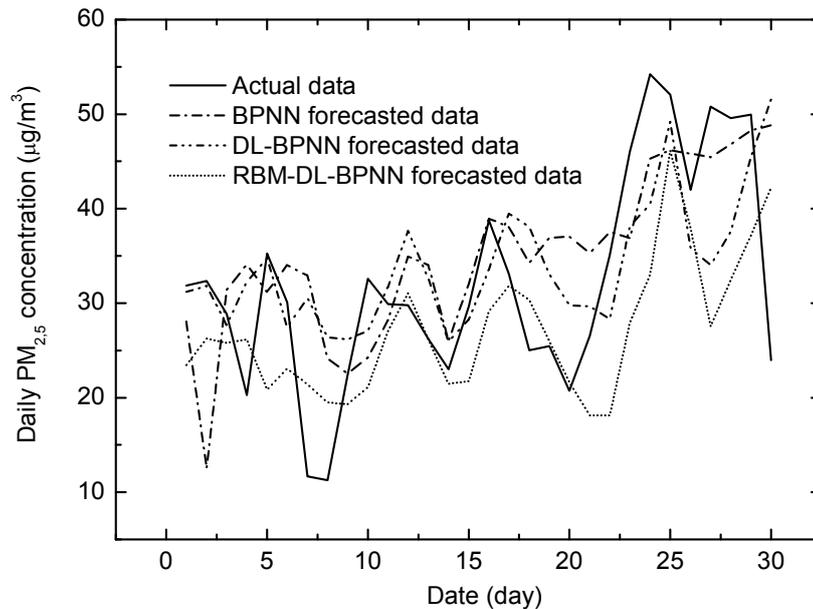


Fig. 5. Actual data for July 2016 versus the daily PM_{2.5} concentration forecasted by the three models

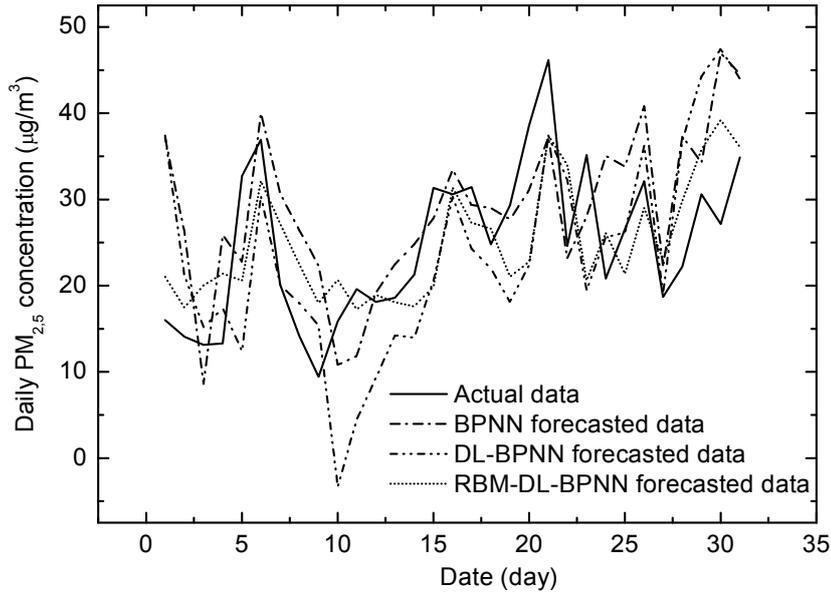


Fig. 6. Actual data for August 2016 versus the daily $PM_{2.5}$ concentration forecasted by the three models

3.2. Discussion

In order to deeply analyze the improvements attributable to the RBM-DL-BPNN model, we collected the daily absolute percentage error of the three models and calculated the days from different error ranges.

Figs. 7 shows the distribution of MAPE of the three models, in which the dense, sparse, and none filling of a column represents a distribution of the daily MAPE of the RBM-DL-BPNN model, DL-BPNN model, and BPNN model, respectively — compares the absolute percentage error in June-August 2016 of the three models. The distribution evidence, that the BPNN model is characterized by bigger number of relative errors less than 10 %. At the same time RBM-DL-BPNN model are characterized by slightly lower amount of relative errors less than 10 %, but the amount of relative errors less than 30 % higher than in the case of another two models.

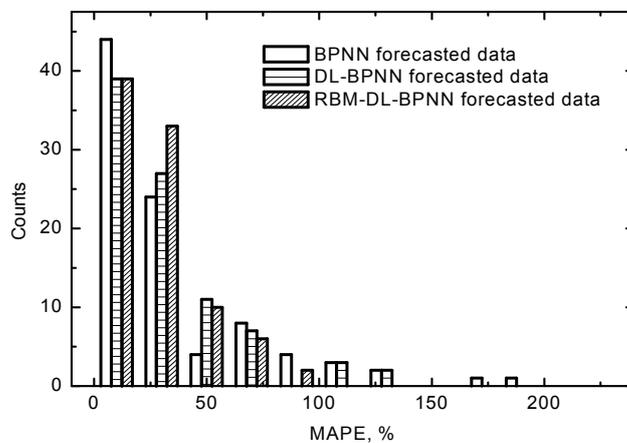


Fig. 7. Absolute percentage error of the three models in June 2016

The biggest relative errors are 88,6 % (dense filled column), 120,9 % (sparse filled column), and 154,8 % (none filled column), respectively. Considering the results presented in Figs. 7, the largest errors of the RBM-DL-BPNN model are all lower than those of the DL-BPNN model and the BPNN model. Furthermore, intuitively, the overall daily absolute percentage errors of the RBM-DL-BPNN model are lower than those of the DL-BPNN model and the BPNN model, which makes the RBM-DL-BPNN model more credible than the others.

Table 2. shows the days of the three models' results in different error ranges. From Table 2, it is clear that the results of the three models are primarily distributed within the range 0–50 %. In the 0–50 % range, the RBM-DL-BPNN model has more data points than the other two models, whereas they are all close in the 50–100 % range. Conspicuously, in the >100 % range, the RBM-DL-BPNN model has zero data points, whereas the DL-BPNN model and the BPNN model both have seven data points. Thus, from the above analysis, we can conclude that the RBM-DL-BPNN model is able to forecast more accurately than the other models.

Table 2. Number of days spent in the three different absolute percentage error scales by each of the three models

	BPNN			DL-BPNN			RBM-DL-BPNN		
	0–50 %	50–100 %	>100 %	0–50 %	50–100 %	>100 %	0–50 %	50–100 %	>100 %
June	23	3	3	23	4	2	25	4	0
July	25	4	2	25	3	3	28	3	0
August	22	7	2	24	5	2	26	5	0
Sum	70	14	7	72	12	7	79	12	0

However, the forecasting accuracy of the RBM-DL-BPNN model is not sufficiently satisfying. Two main factors could account for this result. Firstly, the data used in this work are data taken from websites; they represent the mean values of the meteorological parameters of the entire Hangzhou city, which makes the reliability of the relationships between variables weak. Secondly, we could not collect all the necessary data concerning PM_{2.5} concentration in this work. The pollution sources of PM_{2.5} and their proportions in Hangzhou city are shown in Fig. 10 [3].

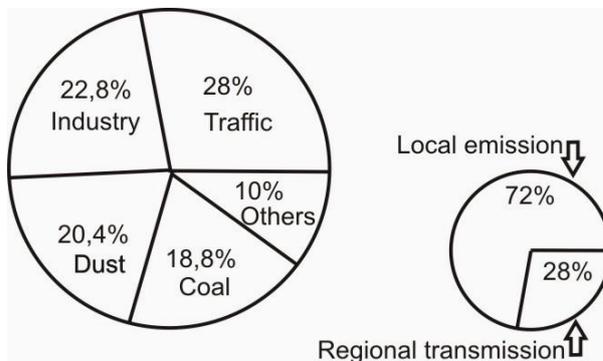


Fig. 8. Pollution sources of PM_{2.5} in Hangzhou [3]

Traffic pollution, industrial pollution, and dust and coal pollution are the top four pollution sources of PM_{2.5} in Hangzhou city. However, obtaining these data directly from official statistics is difficult. However, we will try to collect the related data in future work.

4. Conclusions

ANNs are widely used to process air quality and meteorological records, and many optimized neural networks have proved effective. In this study, we proposed and developed the RBM-DL-BPNN model for forecasting daily PM_{2.5} concentration. The model uses RBM to learn features of the input data and saves the information in weights to initialize the weights of the DL-BPNN model. RBM is firstly applied to optimize the prediction of the ANN, which makes the DL-BPNN more effective. The meteorological parameters of Hangzhou city for the period December 2013 to May 2016 were used to train three models: RBM-DL-BPNN, DL-BPNN and BPNN and the remainder from June 2016 to August 2016 used for testing. Experimental results and analysis show that the RBM-DL-BPNN model has a smaller MAPE, smaller daily absolute percentage errors on the whole, and no errors above 100 %. Thus, it can be concluded that the RBM-DL-BPNN model can relatively accurately and reliably forecast daily PM_{2.5} concentration for Hangzhou city. Although the RBM-DL-BPNN model is better than the DL-BPNN and BPNN models, the RBM-DL-BPNN model sometimes could not forecast accurately because of uncertain anthropogenic factors and cases of extreme weather conditions. In the future, we will study the relationship between daily PM_{2.5} concentration and anthropogenic factors so that human activity and extreme weather conditions can be added as parameters in an appropriate manner for more accurate forecasting.

Acknowledgments

This work was financially supported by the Special Funding of «the Belt and Road» International Cooperation of Zhejiang Province (2015C04005) and National Natural Science Foundation of China (61571399).

1. Minglei Fu., Weiwen Wang, Zichun Le, Mahdi Safaei Khorram. Prediction of particular matter concentration by developed feed-forward neural network with rolling mechanism and gray model. *Neural. Comput. & Applic.* 2015. Vol. 26. Iss. 8. P. 1789–1797.
2. Yifeng Xue, Hezhong Tian, Jing Yan [et al.]. Present and future emissions of HAPs from crematoriums in China. *Atmos. Environ.* 2016. Vol. 124. P. 28–36
3. Hangzhou released PM_{2.5} source analysis results: motor vehicle exhaust emissions first. URL: <http://zjnews.zjol.com.cn/system/2015/06/06/020686281.shtml>
4. Perez-Roa R., Castro J., Jorquera H., Perez-Correa J.R., Vesovic V. Air-pollution modelling in an urban area: Correlating turbulent diffusion coefficients by means of an artificial. *Atmos. Environ.* 2006. Vol. 400. Iss. 1. P. 109–125.
5. Perez P., Gramsch E. Forecasting hourly PM_{2.5} in Santiago de Chile with emphasis on night episodes. *Atmos. Environ.* 2016. Vol. 124. P. 22–27.
6. Voukantsis D., Karatzas K., Kukkonen J., Räsänen T., Karppinen A., Kolehmainen K. Inter-comparison of air quality data using principal component analysis, and forecasting of PM₁₀ and PM_{2.5} concentrations using artificial neural networks, in Thessaloniki and Helsinki. *Sci. Total Environ.* 2011. Vol. 409. P. 1266–1276.
7. Ul-Saufie A.Z., Yahaya A.S., Ramli N.A., Rosaida N., Hamid H.A. Future daily PM₁₀ concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA). *Atmos. Environ.* 2013. Vol. 77. P. 621–630.

8. Gennaro G.D., Trizio L., Gilio A.D. Neural network model for the prediction of PM10 daily concentrations in two sites in the Western Mediterranean. *Sci. Total Environ.* 2013. Vol. 463–464. P. 875–883.
9. Shan S.Q., Feng L., Jian Z.W., Beibei S. Analysis and forecasting of the particulate matter (PM) concentration levels over four major cities of China using hybrid models. *Atmos. Environ.* 2014. Vol. 98. P. 665–675.
10. Antanasijević D.Z., Pocajt V.V., Povrenović D.S., Ristić M.D., Perić-Grujić A.A. PM₁₀ emission forecasting using artificial neural networks and genetic algorithm input variable optimization. *Sci. Total Environ.* 2014. Vol. 443. P. 511–519.
11. Mishra D., Goyal P., Upadhyay A. Artificial intelligence based approach to forecast PM2.5 during haze episodes: A case study of Delhi, India. *Atmos. Environ.* 2015. Vol. 102. P.239–248.
12. Yildirim Y., Bayramoglu M. Adaptive neuro-fuzzy based modelling for prediction of air pollution daily levels in city of Zonguldak. *Chemosphere.* 2006. Vol. 63. P. 1575–1582.
13. Elangasinghe M.A., Singhal N., Dirks K.N., Salmond J.A., Samarasinghe S. Complex time series analysis of PM₁₀ and PM_{2.5} for a coastal site using artificial neural network modelling and *k*-means clustering. *Atmos. Environ.* 2014. Vol. 94. P. 106–116.
14. He H.D., Lu W.Z., Yu X. Prediction of particulate matter at street level using artificial neural networks coupling with chaotic particle swarm optimization algorithm. *Build. Environ.* 2014. Vol. 78. P. 111–117.
15. Siwek K., Osowski S. Improving the accuracy of prediction of PM₁₀ pollution by the wavelet transformation and an ensemble of neural predictors. *Eng Appl Artif Intell.* 2012. Vol. 25. P. 1246–1258.
16. Perez P. Combined model for PM10 forecasting in a large city. *Atmos. Environ.* 2012. Vol. 60. P. 271–276.
17. Arhami M., Kamali N., Rajabi M.M. Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations. *Environ Sci.Pollut. Res.* 2013. Vol. 20. P. 4777–4789.
18. Di'az-Robles L.A., Ortega J.C., Fu J.S. [et al.] A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmos. Environ.* 2008. Vol.42. P. 8331–8340.
19. Zhou Q.P., Jiang H.Y., Wang J.Z., Zhou J.L. A hybridmodel for PM2.5 forecasting based on ensemble empirical mode decomposition and a general regression neural network. *Sci. Total Environ.* 2014. Vol. 496. P. 264–274.
20. Al-Alawi S.M., Abdul-Wahab S.A., Bakheit C.S. Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone. *Environ. Model. Softw.* 2008. Vol. 23. P. 396–403.
21. The main data bulletin of 1 % population sample survey in Zhejiang Province. URL: http://www.zj.stats.gov.cn/tjgb/rkcydcgb/201601/t20160128_168706.html
22. Air Quality Index Monthly Statistical Data. URL: <http://www.aqistudy.cn/historydata/month-data.php?city=>
23. История погоды в Ханчжоу, Китай, Декабрь, 2013. URL: <https://www.wunderground.com/history/airport/ZSHC/2013/12/2/DailyHistory.html>
24. Liu D.J., Li L. Application Study of Comprehensive Forecasting Model Based on Entropy Weighting Method on Trend of PM2.5 Concentration in Guangzhou, China. *Int. J. Environ. Res. Public Health.* 2015. Vol. 12. P. 7085–7099.
25. Hinton G.E., Osindero S., Teh Y.W. A fast learning algorithm for deep belief nets. *Neural Comput.* 2006. Vol. 18. P. 527–1554.

26. Mohamed A., Dahl G.E., Hinton G. Acoustic Modeling using Deep Belief Networks. *IEEE T Audio Speech*. 2012. Vol. 20. P. 14–22.
27. Hinton G., Li D., Yu D. [et al.] Deep Neural Networks for Acoustic Modeling in Speech Recognition. *IEEE Signal Proc. Mag.* 2012. Vol. 29. P. 82–97.
28. Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent Erhan. Why Does Unsupervised Pre-training Help Deep Learning? *Mach. Learn Res.* 2010. Vol. 11. P. 625.
29. Dahl G.E., Yu D., Member S., Li D., Acero A. Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. *IEEE T Audio Speech*. 2012. Vol. 20. P. 30–42.

Received 26.07.2017