Research on the forecasting of Air Quality Index (AQI) based on FS-GA-BPNN: A case study of Beijing, China

Binxu Zhai
Institute of Public Safety Research,
Department of Engineering Physics,
Tsinghua University, Beijing, China
dbx15@mails.tsinghua.edu.cn

Jianguo Chen
Institute of Public Safety Research,
Department of Engineering Physics,
Tsinghua University, Beijing, China
chenjianguo@mail.tsinghua.edu.cn

ABSTRACT

The analysis and forecasting of eminent air quality play a significant role in municipal regulatory planning and emergency preparedness. In this paper, a FS-GA-BPNN model forecasting the daily average Air Quality Index (AQI) is proposed. Special procedures for feature extraction to find more potential significant variables and feature selection to remove redundant information and avoid overfitting are conducted before modelling. Three different models – BPNN, GA-BPNN and FS-GA-BPNN are established to compare the prediction accuracy, generalization ability and reliability. 17 parameters involving pollutant concentration, meteorological elements and surrounding factors are found essential for the method effectiveness. The result shows that the FS-GA-BPNN model generally performs superior to ordinary BPNN, suggesting the necessity of extensive data mining and feature extraction for successful machine learning. The results of this paper can help to conduct air quality pre-warning system and improve the emergency planning process of extreme weather events.

Keywords

Feature Selection, Genetic Algorithm, Backpropagation Neural Network, Air Quality Index, forecasting

ABBREVIATIONS

ANN	Artificial Neural Network
AP1	AQI Value Of The Next Day
AQI	Air Quality Index
BPNN	Back Propagation Neural Network
FS	Feature Selection
GA	Genetic Algorithm
MAE	Mean Absolute Error
MSE	Mean Squared Error
OLS	Ordinary Least Squares
R	Pearson Correlation Coefficient
RMSE	Root Mean Squared Error
SVR	Support Vector Regression
WNN	Wavelet Neural Network

INTRODUCTION

There is robust scientific evidence confirming that the air pollution has been the largest single environmental risk and a leading cause of physiological diseases and respiratory mortality globally (Fischer et al., 2011; Mahiyuddin et al., 2013). Chinese cities experience particular high airborne particle concentrations. According to the latest China Environment Status Bulletin 2015, the ambient air quality of 265 cities exceed the national standard, accounting for 78.4% of the total 338 monitored cities, and typical annual maximal daily average $PM_{2.5}$ concentration was reported to be as high as 477.5 μg/m^3 in Beijing. To avoid these extreme weather events, a reliable forecasting technique is needed which plays an important role in the crisis response and emergency planning.

Existing air quality forecasting techniques and tools can be grouped into three categories (Hajek and Olej, 2015): simple empirical approaches, physically-based approaches and statistical approaches. The simple empirical approaches are usually too delicate to handle abrupt changes of weather, emissions and air quality (U.S. EPA, 2003). The physically-based forecasting systems generally demand sound knowledge of the pollution sources and are sensitive to the initial and boundary conditions. Besides, they require the assist of equipment with high performance computing capacity and weather forecast model with high spatio-temporal resolution (Zhang et al., 2012). The statistical approaches are usually confined to the area and conditions present during the measurements and lack a better understanding of the chemical and physical evolution processes. So the growing trend is to combine the physically-based models with statistical ones to provide more accurate prediction.

Since the first application for ambient pollutant concentrations modeling by Boznar et al. in 1993, ANN has been recognized as a cost-effective method for this task, superior to traditional statistical techniques and gained widely extension in this field (Challoner et al., 2015). However most of the inputs variables of established ANN models were based on experience, common sense or subjective inference from the existing scientific literature (Fu et al., 2015; Perez and Salini, 2008). Some researchers attempted to process feature selection before the training of artificial neural network. Grivas et al. (2006) applied a genetic algorithm optimization procedure for the selection of the input variables and compared the prediction results with multiple linear regression models. Mesin et al. (2010) introduced partial mutual information criterion to select the predictors and utilized Hampel test criterion to assess the significance of selected variables for the system output. Qin et al. (2014) conducted Gray Correlation Analysis (GCA) to search possible predictors for developed CS-EEMD-BPANN model to forecast following hourly particulate matter concentrations. But the interpretability and multicollinearity of most feature selection results still remain ambiguous and controversial.

This study introduces regularization-based feature selection for the filtering of model inputs to eliminate redundant information, alleviate multicollinearity and enhance generalization ability. Two ordinary BPNNs, an evolutionary model optimized by GA and a hybrid model with feature selection are established as contrast, respectively. Statistical indices like Pearson correlation coefficient (R), root mean square error (RMSE) and mean absolute error (MAE) are used to evaluate the reliability of different models.

STUDY AREA AND THE DATA

Study area

This study focuses on the area of Beijing municipality and surrounding cities including Baoding, Chengde, Tianjin, Zhangjiakou and Langfang whose suspended particulate matters and air pollutants can easily affect the air quality of Beijing (Figure 1).

Since 2012 an air pollution monitoring network including 35 monitors has been constructed gradually by Beijing Municipal Environment Monitoring Center. Among the 35 monitoring stations, 17 are located in the urban districts and 18 in suburban districts (Figure 1). The monitoring objects include major air pollutants and meteorological factors according to the new regulation issued by the Ministry of Environmental Protection of China (MEP).

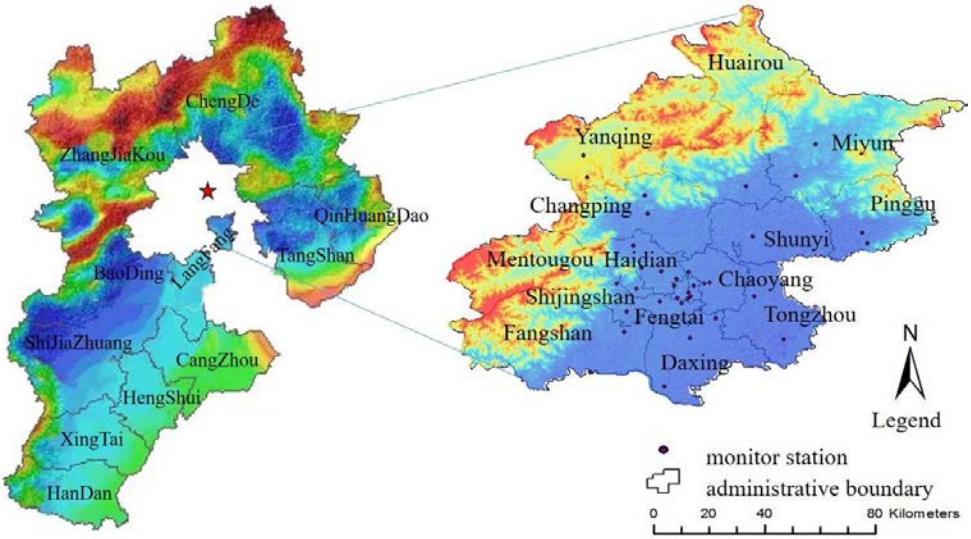

Figure 1. The surrounding cities and administrative division of Beijing municipality, with the monitoring stations labeled on the right part.

Data collection and pre-processing

The Air Quality Index proposed in the new regulation is defined as a function of several sub-indicators shown in Eq.(1) and Eq.(2), involving such pollutants as $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, O_3, which are classified by the mass concentration respectively (HJ 633-2012). So it is imperative to list pollutants, meteorological data and surrounding factors as alternate variables for neural network inputs.

$$AQI = max\{IAQI_1, IAQI_2, IAQI_3, \dots IAQI_n\} \tag{1}$$

$$IAQI_n = \frac{IAQI_{Hf} - IAQI_{Lo}}{BP_{Hf} - BP_{Lo}}(C_n - BP_{Lo}) + IAQI_{Lo} \tag{2}$$

In Eq.(1) and Eq.(2), the $IAQI_n$, $IAQI_{Hf}$, $IAQI_{Lo}$ refers to the corresponding sub-indicators of AQI respectively. The C_n refers to the mass concentration of pollutant n. BP_{Hf} and BP_{Lo} refers to the high-value and low-value of pollutant concentration n.

The analysis is based on historical daily averaged data from 1st February, 2013 to 31st July, 2016, spanning a total of 1277 days. Descriptive statistics for AQI and Particulate Matters during the observed period are presented in Table 1. Ground measured hourly pollute concentration data including PM_{10}, $PM_{2.5}$, CO, NO_2, SO_2, O_3 and AQI are collected from the Beijing Municipal Environment Monitoring Center (http://zx.bjmemc.com.cn/) and then calculated as daily mean values. The meteorological data including Barometric pressure, Air temperature, Relative humidity, Wind direction, Wind velocity, Precipitation and Duration of sunshine are acquired from National Meteorological Information Center (http://data.cma.cn/). The AQI values of surrounding 5 cities are obtained from the data center of MEP (http://www.mep.gov.cn/).

Table 1. Basic descriptive statistical characteristics for the main indicators

Statistical Characteristics	AQI	$PM_{2.5}$	PM_{10}
Arithmetic mean	113.4	80.1	106.1
Median	92.0	61.0	90.6
Standard Deviation	74.6	66.8	72.0
Range	15 - 475	5.2 – 477.5	1.8 – 480.8

All values in μg/m^3 except Standard Deviation and AQI (dimensionless).

Apart from the basic quality control (e.g. excluding missing data) of row data, additional attention is paid to feature extraction to integrate both quantitative and qualitative data and discover more potential input variables.

The AQI of the next day (AP1) is illustrated as a response of wind direction of the previous day (Figure 2). And the wind direction (in degree) are transformed to sine and cosine variables to help interpret the periodicity of prevailing wind. Besides, according to the wind roses, an additional variable is defined to fit the corresponding relationship between wind direction and AP1 as follow:

$$\mathrm{WD} = \left|\sin\left(\theta + \frac{4}{3}\pi\right)\right| \tag{3}$$

where θ refers to the wind direction (in radians), and the $4/3\pi$ represents the additional bias which makes the WD reach maximum when θ obtains the prevailing direction (i.e. south west and north east) and zero when the direction is south east or north west.

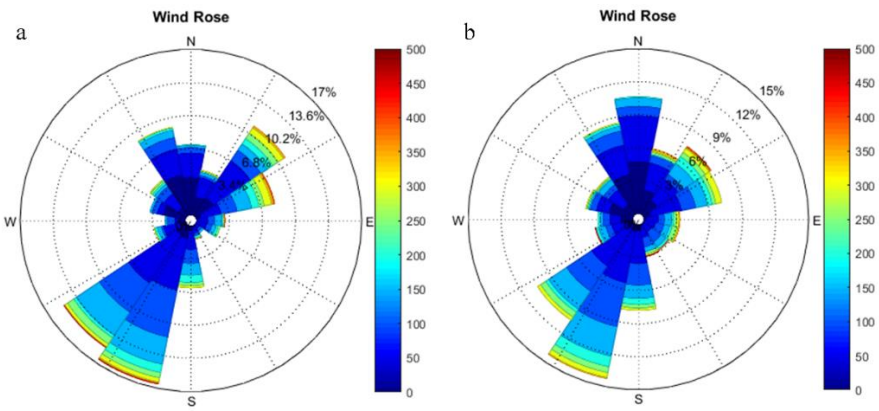

Figure 2. a. Wind Rose for direction of maximum wind speed VS AQI; b. Wind Rose for direction of extreme wind speed VS AQI. The radius refers to the frequency of specific wind direction and the intensity of color refers to the value of AQI.

To incorporate the effect of the weekly variation of pollutants emission, another index is introduced by adding a weight to every week day expressed as Eq.(4). The $\bar{x_i}$ refers to the mean AQI of specific day in a week and $\bar{x}$ represents the mean value of the whole year.

$$\mathrm{DoW} = \frac{\bar{x_i}}{\bar{x}} \times 100\% \tag{4}$$

$$\{i = \ Monday, Tuesday - Friday, Saturday, Sunday \ and \ other \ official \ hoidays\}$$

The effect of precipitation is also considered as a binary variable (0/1) represented by Pre_b. And based on the statistical analysis of the row data, the seasonal effect is taken into account as illustrated in Eq.(5). The $\bar{x_i}$ refers

Table.2. Definitions of raw data parameters

Parameter	Abbreviation	Unit	Description
Index	AQI		Air quality index of Beijing
	Dow		The effect of the weekly variation
	SE		Seasonal effect
	SCI		Shanghai Composite Index
Pollutants	$PM_{2.5}$	μg/m^3	Daily averaged concentration of $PM_{2.5}$
	PM_{10}	μg/m^3	Daily averaged concentration of PM_{10}
	SO_2	μg/m^3	Daily averaged concentration of SO_2
	CO	mg/m^3	Daily averaged concentration of CO
	NO_2	μg/m^3	Daily averaged concentration of NO_2
	O_3	μg/m^3	Daily averaged concentration of O_3
Meteorological Factors	P_mean	0.1hPa	Mean barometric pressure
	P_max	0.1hPa	Maximum barometric pressure
	P_min	0.1hPa	Minimum barometric pressure
	ΔP	0.1hPa	Daily maximum pressure difference
	T_mean	0.1℃	Mean air temperature
	T_max	0.1℃	Maximum air temperature
	T_min	0.1℃	Minimum air temperature
	ΔT	0.1℃	Daily maximum temperature difference
	Vap	0.1hPa	Average vapor pressure
	Eva_s	0.1mm	Small evaporation
	Eva_l	0.1mm	Large evaporation
	H_mean	1%	Mean relative humidity
	H_min	1%	Minimum relative humidity
	Pre	0.1mm	Precipitation
	Pre_b	binary	Precipitation in binary code
	v_mean	0.1m/s	Mean wind velocity
	v_max	0.1m/s	Maximum wind velocity
	v_max_n	1-16	Maximum wind velocity direction in number
	v_max_d	degree	Maximum wind velocity direction in degree
	Sin_max		Sine of maximum wind velocity direction
	Cos_max		Cosine of maximum wind velocity direction
	Sin(x+4/3π)_max		Transferred maximum wind velocity direction
	v_ext	0.1m/s	Extreme wind velocity
	v_ext_n	1-16	Extreme wind velocity direction in number
	v_ext_d	degree	Extreme wind velocity direction in degree
	Sin_ext		Sine of extreme wind velocity direction
	Cos_ext		Cos of extreme wind velocity direction
	Sin(x+4/3π)_ext		Transferred extreme wind velocity direction
	Sun	0.1hour	Duration of sunshine

Table 2. Definitions of raw data parameters

Parameter		Unit	Description
Background factors	BD		Air quality index (AQI) of Baoding
	CD		Air quality index (AQI) of Chengde
	LF		Air quality index of (AQI) Langfang
	TJ		Air quality index of (AQI) Tianjin
	ZJK		Air quality index of (AQI) Zhangjiakou

to the mean AQI of specific season and $\bar{x}$ represents the mean value of the whole year, with *i* ranging from spring to winter.

$$\mathrm{SE} = \frac{\overline{x_i}}{\bar{x}} \times 100\% \qquad \{i = Spring, Summer, Autumn, Winter\} \tag{5}$$

To implement the estimation of the emission intensity of pollution source and the economical background conditions, the daily volume of Shanghai Composite Index (reflecting the daily variability in the trading of main listed shares in China, abbreviated as SCI) is taken as a supplementary variable. The total variable considered in this paper is listed in Table 2.

Besides, all of the raw data are normalized to fall in the range of [−1, 1] via the following formula:

$$y = \frac{(x - x_{min})}{(x_{max} - x_{min})} * (y_{max} - y_{min}) + y_{min} \tag{6}$$

where x is the element of each observation vectors, y is the normalized value at observation i, x_{min} and x_{max} are the minimum and maximum of each observation vectors respectively, and y_{min} , y_{max} are the minimum and maximum of each normalized vectors (-1/1).

METHODOLOGY

Back propagation neural network

Back Propagation Neural Network (BPNN), a multilayer feed forward neural network trained by error back propagation algorithm to minimize the sum squared error of network is adopted in this paper, with the learning algorithm ranging from Levenberg-Marquardt backpropagation to Bayesian Regularization backpropagation and Scaled Conjugate Gradient backpropagation, seeking for the best performance of networks. It has been proved that a single hidden layer network can approximate any continuous function to any desired accuracy, given sufficient neurons in the hidden layer (Guliyev et al., 2016). A typical neural network architecture is briefly illustrated in Figure 3. Each input is weighted with an appropriate weight W and the sum of the weighted inputs and bias ($IW+b^1$) forms the input to the hidden layer transfer function f_H. The output layer works similarly with the hidden layer with few differences on the transfer function f_O to generate their output. The structures adopted in this paper are all based on this architecture with the specific input preprocessing method, neuron count and optimization algorithm vary and complicate for different models. The early-stopping technique and regularization method are applied respectively as contrast to avoid overfitting. For each method, the original data are divided into two subsets evenly with one set reserved as training set (85%) and the other cross validation set (15%).

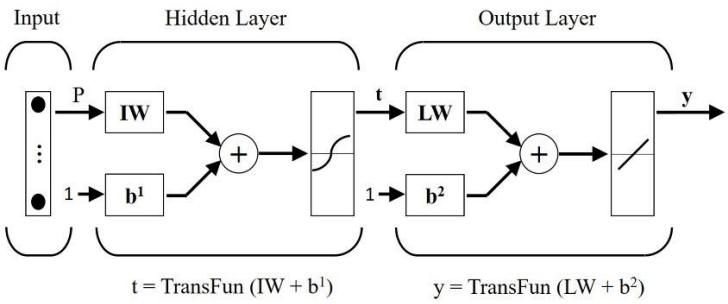

Figure 3. Typical structure of neural networks.

Regularization-based feature selection

In supervised machine learning settings with many input variables, overfitting is usually a potential problem unless there is ample training data (Srivastava et al., 2014). Efficient and robust feature selection methods are usually needed to extract meaningful features and eliminate noisy ones. In this paper, two standard regularization methods – the L_1 regularization, which uses a penalty term to force the sum of the absolute values of the parameters to be small, and L_2 regularization, which minimizes the sum of the squares of the parameters - are applied to identify important predictors, reduce input dimension and enhance the generalization ability.

Lasso regression

Lasso (least absolute shrinkage and selection operator), a regression analysis method that performs both variable selection and regularization, was originally formulated for ordinary least squares models (OLS) and then extended to a wide variety of statistical models. Different with OLS, it adds a positive penalty term (L_1 norm of coefficient vector) to the error function that forces the estimated coefficients of minority features to be zero (Tibshirani, 1996). The sparse parameter vector makes it a natural candidate for feature selection, and meanwhile, can present smaller mean squared error than an OLS estimator when applied to new data. The objective of Lasso can be written compactly as

$$min_{\beta_0,\beta}\left\{\frac{1}{N}\sum_{i=1}^{N}(y_i-\beta_0-x_i^T\beta)^2+\lambda\sum_{J=1}^{P}\left|\beta_j\right|\right\} \tag{7}$$

where N is the number of observations, each of which consists of p covariates and a single outcome. x_i and y_i are the values and response at observation i. λ is a positive regularization parameter.

Ridge regression

Ridge regression is the most popular method of regularization for ill-posed problems in statistics. Similar with Lasso, Ridge improves generalization ability by shrinking regression coefficients to avoid overfitting, but it does not set any of them to zero (Ng, A. Y., 2004). So it can be an implement for Lasso to avert the missing of important variables. The difference between Lasso and Ridge lies in that Ridge introduces L_2 norm of coefficient vector into the error function, which gains the Ridge more stable performance for coefficients estimation. The Ridge regression model can be interpreted as follow:

$$min_{\beta_0,\beta}\left\{\frac{1}{N}\sum_{i=1}^{N}(y_i-\beta_0-x_i^T\beta)^2+k\sum_{J=1}^{P}(\beta_j)^2\right\} \tag{8}$$

where the meanings of variables are the same as those in Lasso, except that the $||\beta||_1$ is replaced as $||\beta||_2$ and the k represent the Ridge parameter.

Genetic Algorithm

Genetic algorithm (GA) is a highly parallel, auto-adaptive computation metaheuristic inspired by the process of natural selection to search the global optimal solution by bio-inspired operation rules such as selection, crossover and mutation. The selection rules choose the parent individuals for the next generation based on their scaled values at the fitness function. The crossover rules combine two parents to form children for the next generation. The mutation rules implement random changes to parent individuals to form children with a fixed probability. The algorithm is frequently used for optimization of various established models (Gan et al., 2016; Wang et al., 2015).

In this paper, GA is implemented as an optimization for the weight and bias of ANN to avoid the local minimum and improve the prediction accuracy. The RMSE of the predicted daily mean AQI is taken as the fitness function of GA, and the schematic representation of the hybrid algorithm is shown in Figure 4.

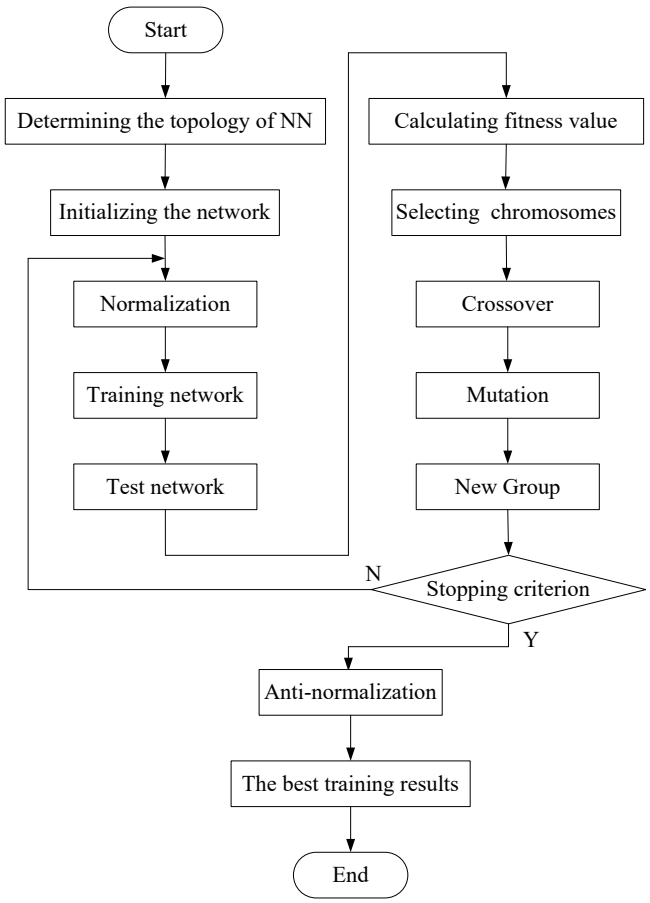

Figure 4. The algorithm flow of the optimized model. The left part of the diagram is the process of ANN while the right part the procedure of GA. Note that the input variables has been optimized by feature selection.

RESULTS AND DISCUSSION

Feature selection result

In this paper, a 10-fold cross validation is performed to estimate the MSE during the Lasso regularization, where the original sample is randomly partitioned into 10 subsamples, with one single subsample remained as validation set and the other 9 subsamples as training set. The optimal regularization parameter (λ) with minimal MSE plus one standard deviation is then used to calculate the coefficient of corresponding variables. The trace plot of coefficient fit and cross validation MSE are shown as Figure 5.

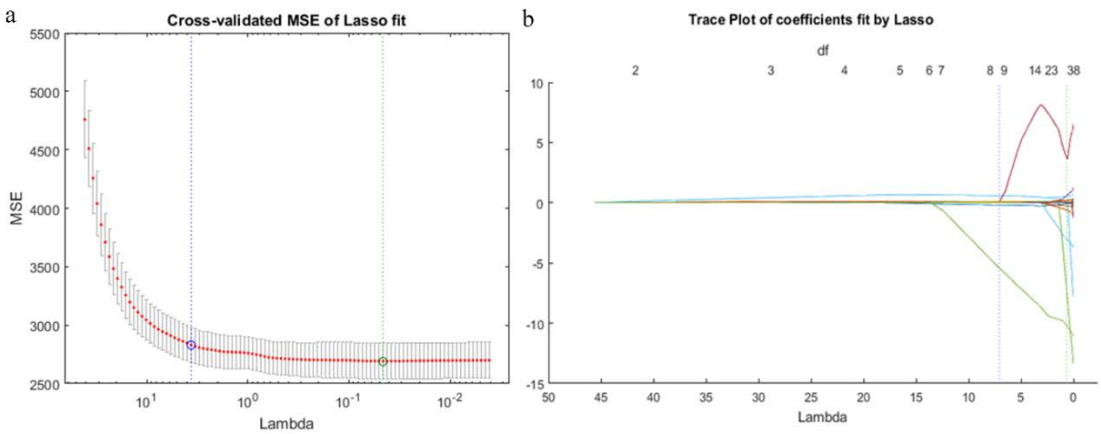

Figure 5. a. The cross validated MSE of Lasso fit. b. The trace plot of coefficient fit by Lasso.

The selection result of Lasso regularization is listed in Table 3, indicating that the most important variables for the prediction of AQI are cos_ext (negative correlation) and sin(x+4/3π)_max (positive correlation), followed by meteorological factors such as Pmin (positive correlation), Vext (negative correlation) and Sun (positive correlation) and main pollutant concentration such as $PM_{2.5}$, PM_{10} and NO_2.

Table.3. Coefficients estimated by Lasso

Factor	**AQI**	**$PM_{2.5}$**	**PM_{10}**	**NO_2**	**Pmin**	**sin(x+4/3π)_max**	
Coefficient	0.0826	0.0853	0.1116	0.517	0.0114	6.7743	
Factor	**Vext**	**cos_ext**	**Sun**	**LF**	**ZJK**	**BD**	**others**
Coefficient	-0.2419	-7.9065	0.0514	0.072	0.0765	0.0138	0

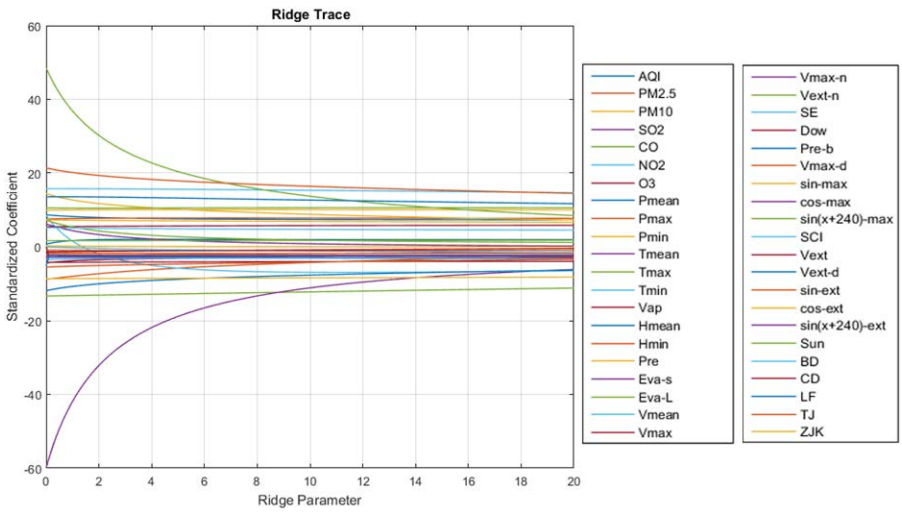

Figure 6. The trace plot of Ridge regression.

Table4. Coefficients estimated by Ridge

factor	AQI	Dow	SE	SCI	$PM_{2.5}$	PM_{10}
k=0	6.81	1.41	-1.84	0.99	9.30	9.00
k=20	7.08	1.49	-1.83	1.28	8.67	9.29
change /%	3.99%	5.67%	0.54%	29.29%	6.79%	3.26%
factor	SO_2	CO	NO_2	O_3	P_mean	P_max
k=0	-4.49	-14.94	21.39	7.00	-2.31	-8.25
k=20	-2.67	-12.04	18.02	7.12	1.52	-2.69
change /%	40.49%	19.40%	15.77%	1.72%	165.96%	67.33%
factor	P_min	Vap	T_mean	T_max	T_min	Eva_s
k=0	16.26	4.93	-65.01	40.53	0.77	-0.23
k=20	8.11	1.63	-10.24	3.96	-10.89	-0.24
change /%	50.11%	66.93%	84.25%	90.22%	1520.99%	4.34%
factor	Eva_l	H_mean	H_min	Pre	Pre_b	v_mean
k=0	17.59	-7.71	20.15	4.87	0.42	-1.26
k=20	13.03	-3.65	14.17	5.26	0.42	-2.49
change /%	25.95%	52.73%	29.69%	8.01%	0%	98.42%
factor	v_max	v_max_n	v_max_d	Sin_max	Cos_max	Sin(Φ)_max*
k=0	-6.82	0.00	-3.52	-1.14	-2.28	1.75
k=20	-5.64	-1.82	-1.82	-1.25	-2.50	1.87
change /%	17.34%	-	48.22%	9.81%	9.89%	7.03%
factor	v_ext	v_ext_n	v_ext_d	Sin_ext	Cos_ext	Sin(Φ)_ext*
k=0	-3.18	0.00	-5.19	-0.42	-9.07	-1.99
k=20	-3.85	-2.42	-2.42	-0.53	-8.59	-2.01
change /%	21.23%	-	53.48%	25.23%	5.35%	1.21%
factor	Sun	BD	CD	LF	TJ	ZJK
k=0	8.69	6.65	-1.42	13.63	-6.86	7.73
k=20	9.54	5.86	-0.71	11.71	-4.41	6.91
change /%	9.67%	11.89%	50.35%	14.10%	35.60%	10.65%

*$\Phi=x+4/3\pi$

The Ridge regression is performed with the parameter *k* ranging from 0 to 20 (Figure 6). With the increase of *k*, most of the coefficients decrease towards zero but the traces are apparently steadier than Lasso. Except the same variables selected by Lasso, CO, Hmin and Eva-L gains similar estimation as NO_2 in Ridge, which may be omitted by Lasso and need to be taken into consideration. Besides, the Tmean shows distinctive performance in Ridge, with the coefficient ranging from around -65 to -10 and the same as T_{min}, with the coefficient multiplying by more than15 times from 0.76 to 10.88. Part of coefficients estimated by Ridge are listed in Table 4.

Based on the analysis above, this study totally finds 17 significant variables for the prediction of AQI and takes them as the input of NN models – *AQI, $PM_{2.5}$, PM_{10}, NO_2, CO, Pmin, Hmin, Tmean, Tmin, Eva-L, sin(x+4/3π)_max, Vext, cos_ext, Sun, LF, ZJK and BD* (see Table 2 for complete definition).

Statistical studies using meteorological data and air pollution monitoring data have confirmed that the meteorological condition affects the dispersion, transformation and removal of pollutants in the atmosphere and finally influences the spatial-temporal distribution characteristics and air pollution levels in numerous ways (Tian et al., 2014; Yin et al., 2016; Zhang et al., 2015). The selection result in this study shows partial consistency with

the previous researches, confirming the prominent relationship between air quality with main air pollutants and meteorological parameters. Besides, the contribution of such meteorological factors as daily minimum of relative humidity, barometric pressure and temperature (instead of the daily averaged values used in most relevant researches before) to the prediction of AQI is of notable significance but rarely reported. The air quality of surrounding cities also takes an indispensable role in the pollution level of observed subject, verifying the necessity of sufficient relevant variable mining. The selection of such variables as *cos_ext* and *sin(x+4/3π)_max* proves the validity of feature extraction in this research which are also rarely reported in previous work.

Comparative evaluation of the models

During the training process, the original data is divided into two subsets – one for supervised machine learning and the other for cross validation. Statistical indices like Pearson correlation coefficient (R), root mean square error (RMSE) and mean absolute error (MAE) are used to evaluate the reliability of different models. R is used to describe the degree the prediction approximating the observation. The model is more reliable when R is closer to one. RMSE and MAE are applied to quantify the forecasting errors which indicates a higher prediction accuracy when these indices are closer to zero. R, RMSE and MAE are calculated as follow:

- The correlation coefficient (R) of the observed and predicted data :

$$R = \frac{\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}\sqrt{\sum_{i=1}^{n}((\hat{y}_i-\bar{\hat{y}})^2}} \tag{9}$$

- The root mean square error (RMSE) :

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{10}$$

- The mean absolute error (MAE) :

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{11}$$

In Eq. (9-11), n is the number of data point. $\hat{y}_i$ is the predicted value. y_i is the real observed value. $\bar{\hat{y}}$ is the average predicted value and $\bar{y}$ is the average observed data.

Three different models are established to compare the reliability and prediction accuracy – BPNN, GA-BPNN and FS-GA-BPNN. GA-BPNN is the evolutional model of BPNN optimized by Genetic Algorithm, while the FS-GA-BPNN is the hybrid model of Feature Selection, Genetic Algorithm and Backpropagation Neural Network. In contrast of the old regulation, a BPNN model utilizing corresponding predictors included in the old national standard is constructed on parallel. The statistical indices for performance evaluation of these models are presented in Table5.

As is seen in Table 5, there are not tremendous performance gap between the training sets and test sets in all models, suggesting a reasonable structure of the inner architecture. The BPNN* performs slightly better than BPNN. This may be attributed to the redundant information contained in the input of the latter, which makes the neural network learn too many meaningless details and erodes the generalization ability. The GA-BPNN shows a better performance than BPNN and BPNN* but a little poor prediction accuracy, with the R of testing set reaches

Table.5.The prediction results of different models

Model	BPNN*	BPNN	GA-BPNN	FS-GA-BPNN
Structure	4-4-1	28-24-1	28-24-1	17-17-1
Training Set				
R	0.89	0.80	0.89	0.90
MAE(μg/m^3)	37.62	40.27	34.99	30.02
RMSE(μg/m^3)	51.64	53.23	47.23	27.11
Testing Set				
R	0.73	0.69	0.75	0.82
MAE(μg/m^3)	41.52	41.33	37.73	33.91
RMSE(μg/m^3)	54.78	57.37	51.55	40.26

The model BPNN* represents the neural network using observed data according to the old regulation as input.

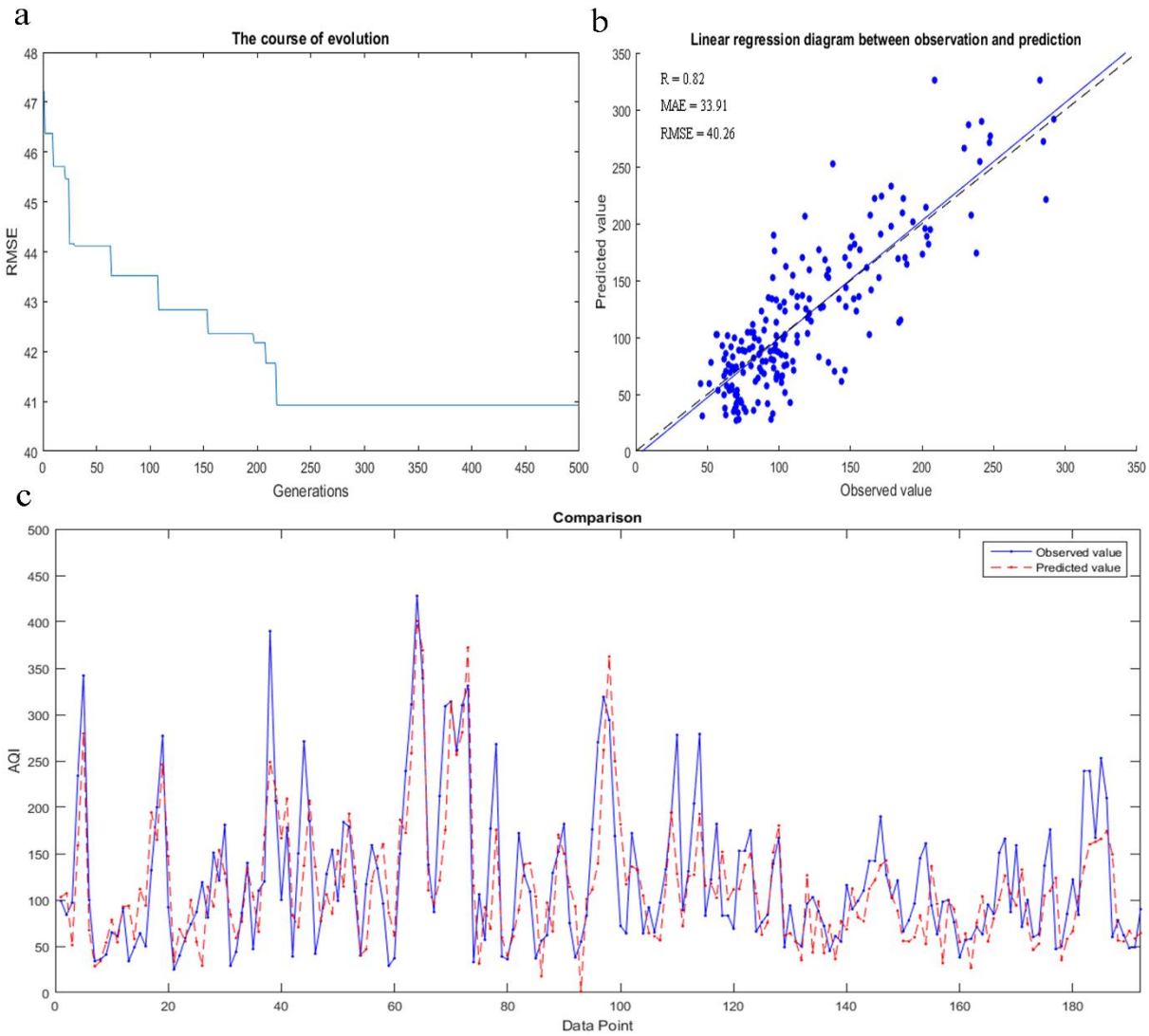

Figure 7. a. The GA optimization course, with the best RMSE=40.26 at generation 221; b. The regression line between the prediction and observation; c. Comparison of predicted value and observed value in testing set.

only 0.75 when the training set obtains 0.89. The GA still makes a remarkable contribution to the optimization of BPNN, minimizing the RMSE by around 10.14%. Both RMSE and MAE of FS-GA-BPNN model are lower than the other three models, and the correlation coefficient is the highest in all models, with the RMSE and MAE decreasing by 21.9% and 10.12% and R increasing by 9.33% compared with GA-BPNN, certifying the effectiveness of feature selection. The procedure of feature extraction and feature selection not only improves the generalization ability and reduce the complexity of NN models but also increases the prediction accuracy. The GA optimization course, regression line and the comparison between observed and predicted value in testing set of FS-GA-BPNN model are shown in Figure 7.

CONCLUSIONS

Driven by deteriorated air quality and increased societal demands, the emergency planning for extreme weather crises have raise extensive concerns from communities, organizations and government worldwide. Accurate air quality forecasting can be used to issue early air quality alerts that allow government and people to take precautionary measures such as temporarily shutting off major emission sources and avoid direct exposures to polluted outdoor environment. These actions can significantly alleviate the government pressure of municipal regulatory management and relieve the adverse health impacts on individuals, which therefore can bring tremendous societal and economic benefits in the long run.

In this study, a FS-GA-BPNN model forecasting the daily average Air Quality Index (AQI) based on Feature Selection (FS), Genetic Algorithm (GA) and Backpropagation Neural Network (BPNN) is proposed. Different from traditional practice of empirically choosing input for supervised machine learning, a special procedure for feature extraction to find more potential significant variables and feature selection to remove redundant information and avoid overfitting is conducted before training the neural network. Three different models are established to compare the prediction accuracy, generalization ability and reliability of different methods. The result shows that the FS-GA-BPNN generally performs 10% - 20% better than an ordinary BPNN and that optimized by Genetic Algorithm alone. The model can help to recognize the causes of weather crisis, improve the emergency preparedness and planning process and sequentially help to conduct air quality pre-warning system and enhance crisis management ability.

The training of the developed ANNs are based on the limited observation duration and restricted to the sampling location of Beijing. However, the benefit of ANN models can be improved by incorporating the following aspects in the future research: (1) Periodic expansion of training sets, exhaustive recognition and quantification on the pollution source and appropriate refinement for data preprocessing method (2) Comparison of other prevailing machine learning algorithms such as Support Vector Regression (SVR), Wavelet Neural Network (WNN) etc. ; (3) Investigation of the validity of ANN models for different cities around the country; (4) Exploration of the capability of ANN models in predicting AQI in two or three days advance.

REFERENCES

Boznar, M., Lesjak, M., & Mlakar, P. (1993). A neural network-based method for short-term predictions of ambient SO2 concentrations in highly polluted industrial areas of complex terrain. Atmospheric Environment. Part B. Urban Atmosphere, 27(2), 221-230.

Challoner, A., Pilla, F., & Gill, L. (2015). Prediction of Indoor Air Exposure from Outdoor Air Quality Using an Artificial Neural Network Model for Inner City Commercial Buildings. International journal of environmental

research and public health, 12(12), 15233-15253.

Fischer, P. H., Marra, M., Ameling, C. B., Janssen, N., & Cassee, F. R. (2011). Trends in relative risk estimates for the association between air pollution and mortality in The Netherlands, 1992–2006. Environmental research, 111(1), 94-100.

Fu, M., Wang, W., Le, Z., & Khorram, M. S. (2015). Prediction of particular matter concentrations by developed feed-forward neural network with rolling mechanism and gray model. Neural Computing and Applications, 26(8), 1789-1797.

Gan, R., Chen, N., & Huang, D. (2016). Comparisons of forecasting for hepatitis in Guangxi Province, China by using three neural networks models. PeerJ, 4, e2684.

Grivas, G., & Chaloulakou, A. (2006). Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece. Atmospheric Environment, 40(7), 1216-1229.

Guliyev, N. J., & Ismailov, V. E. (2016). A single hidden layer feedforward network with only one neuron in the hidden layer can approximate any univariate function. Neural computation.

Hajek, P., & Olej, V. (2015). Predicting Common Air Quality Index–The Case of Czech Microregions. Aerosol and Air Quality Research, 15(2), 544-555.

Mahiyuddin, W. R. W., Sahani, M., Aripin, R., Latif, M. T., Thach, T. Q., & Wong, C. M. (2013). Short-term effects of daily air pollution on mortality. Atmospheric environment, 65, 69-79.

MEP. (2012). Ambient air quality standards (GB 3095 - 2012).

MEP. (2012). Technical regulation on ambient air quality index (on trial) (HJ 633-2012).

MEP. (2016). China Environment Status Bulletin 2015.

Mesin, L., Orione, F., Taormina, R., & Pasero, E. (2010, September). A feature selection method for air quality forecasting. In International Conference on Artificial Neural Networks (pp. 489-494). Springer Berlin Heidelberg.

Ng, A. Y. (2004). Feature selection, L 1 vs. L 2 regularization, and rotational invariance. In Proceedings of the twenty-first international conference on Machine learning (p. 78). ACM.

Perez, P., & Salini, G. (2008). PM2.5 forecasting in a large city: Comparison of three methods. Atmospheric Environment, 42(35), 8219-8224.

Qin, S., Liu, F., Wang, J., & Sun, B. (2014). Analysis and forecasting of the particulate matter (PM) concentration levels over four major cities of China using hybrid models. Atmospheric Environment, 98, 665-675.

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929-1958.

Tian, G., Qiao, Z., & Xu, X. (2014). Characteristics of particulate matter (PM 10) and its relationship with meteorological factors during 2001–2012 in Beijing. Environmental Pollution, 192, 266-274.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. J. R. Statist. Soc. B, 58, 267–288.

U.S. EPA, 2003. Guideline for Developing an Air Quality (Ozone and PM2.5) Forecasting Program. U.S. Environmental Protection Agency Rep, EPA-456/R-03-002,126 pp.

Wang, Z., Lu, F., Lu, Q. C., Wang, D., & Peng, Z. R. (2015). Fine-scale estimation of carbon monoxide and fine particulate matter concentrations in proximity to a road intersection by using wavelet neural network with

genetic algorithm. Atmospheric Environment, 104, 264-272.

Yin, Q., Wang, J., Hu, M., & Wong, H. (2016). Estimation of daily PM 2.5 concentration and its relationship with meteorological conditions in Beijing. Journal of Environmental Sciences.

Zhang, H., Wang, Y., Hu, J., Ying, Q., & Hu, X. M. (2015). Relationships between meteorological parameters and criteria air pollutants in three megacities in China. Environmental research, 140, 242-254.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., & Baklanov, A. (2012). Real-time air quality forecasting, part I: History, techniques, and current status. Atmospheric Environment, 60, 632-655.