

# Research on Air Quality Evaluation based on Principal Component Analysis

Xing Wang<sup>1,2</sup>, Zilin Wang<sup>2</sup>, Min Guo<sup>2,\*</sup> Wei Chen<sup>2</sup> and Huan Zhang<sup>2</sup>

<sup>1</sup>Hebei University of Technology, Tianjin, 300191, China

<sup>2</sup>Tianjin Academy of Environmental Sciences, Tianjin, 300191, China

\*Corresponding author e-mail: 251939587@qq.com

**Abstract.** Economic growth has led to environmental capacity decline and the deterioration of air quality. Air quality evaluation as a fundamental of environmental monitoring and air pollution control has become increasingly important. Based on the principal component analysis (PCA), this paper evaluates the air quality of a large city in Beijing-Tianjin-Hebei Area in recent 10 years and identifies influencing factors, in order to provide reference to air quality management and air pollution control.

## 1. Introduction

In recent years, the economic development and population growth have led to a deterioration of air quality. Therefore, it is very important to evaluate the quality of atmospheric environment. The evaluation methods concentrates on index evaluation, fuzzy comprehensive evaluation, gray clustering, and neural network. However, these analytical methods have some shortcomings. For example, index evaluation method fails to present a comprehensive evaluation of air quality, whereas fuzzy comprehensive evaluation method needs to give a weight for each index artificially, in which main indicators' role in air quality evaluation can hardly be reflected. Considering environment system is an open system, the change of environmental quality is the result of the combined effects of each variable. With the popularity of multivariate statistical methods, principal component analysis (PCA), which transforms multiple indicators into several unrelated comprehensive indicators on the basis of interrelationship among the multiple indicators, is applied widely in environmental quality assessment. In PCA, the correlation between the indicators is fully considered, in which can achieve the maximized retention of the original information. In addition, by means of integrated dimension reduction processing to high-dimensional data, PCA can objectively determine the weight of each indicator to avoid subjective randomness. In this research, a comprehensive evaluation on the air quality of one city in recent 10 years by PCA is carried out, so as to provide technical support for environmental quality monitoring and air pollution control.

## 2. The fundamental principles and analytical methods of PCA

### 2.1. Fundamental principles

The aim of PCA is to make a comprehensive simplification of multi-variable data and grasp main aspects of the problems in the condition of minimum information loss. By establishing a mapping relation, PCA transforms complex evaluation indicators into several simplified comprehensive



indicators, or in other words, converts data from high dimensional to low-dimensional<sup>8</sup>. Mathematically, a new comprehensive indicator is achieved by conducting linear combination of  $i$  ( $i=1,2,3,4,\dots,n$ ) indicators. The variance of the first linear combination,  $Z_1$ , indicates to what extent the information from original multi-variable data has been included. The larger the variance of  $Z_1$  is, the more original information  $Z_1$  contained. If the variance of  $Z_1$  is larger than any other linear combinations, then  $Z_1$  is defined as the first principal component of all the linear combinations. The ratio of  $Z_1$ 's variance in the total variance of all the linear combinations is defined as interpretation variance, which indicates  $Z_1$ 's ability of representing original data. The larger the interpretation variance is, the greater  $Z_1$ 's representativeness is in the linear combinations<sup>9</sup>. If the first principal component is insufficient to represent the information of the original  $i$  indicators, then the second linear combination  $Z_2$  is selected to reflect the original information along with  $Z_1$ . Generally, if the current  $j$  principal component variances occupy over 85% of the total variance, it shall be deemed that the  $j$  principal components can represent the most information<sup>10</sup>.

## 2.2. Analytical methods

(1) Data standardization. To exclude the impact of different orders of magnitude and dimensions, and make the original data comparable on the same scale, the data should be standardized.

(2) Calculation of covariance matrix.

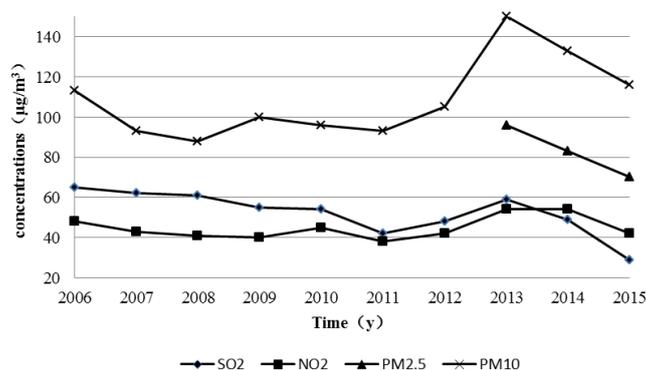
(3) Calculation of eigenvalues and eigenvectors of covariance matrices.

(4) Explain the calculation of variance, and determine the principal component. In theory, the cumulative variances of the former  $n$  principal components are required to be over 85% of the total variance.

(5) Comprehensive evaluation. Select  $n$  principal components, calculate the corresponding eigenvector, and give a reasonable explanation to the selected principal component.

## 3. An analysis on variation trends of typical pollutants in a city

According to the newest revision of Environmental Air Quality Standard (GB3095-2012) issued by Ministry of Environmental Protection (MEP) in 2012,  $PM_{2.5}$  was incorporated into the scope of routine air quality monitoring, which includes six air pollutant indicators, such as sulfur dioxide ( $SO_2$ ), nitrogen dioxide ( $NO_2$ ), carbon monoxide CO, ozone ( $O_3$ ),  $PM_{2.5}$ , and  $PM_{10}$ . In this paper, the variation trends of typical pollutants, such as  $SO_2$ ,  $NO_2$ , and  $PM_{10}$ , are analyzed in the time scope of 2005 to 2015, along with  $PM_{2.5}$  from 2013 to 2015 as it was first introduced in the GB3095-2012 in 2013. The annual average concentrations of major air pollutants are shown in Figure 1.



**Figure 1.** Trend chart of annual average concentration of main air pollutants

In general, annual average concentrations of  $SO_2$  and  $NO_2$ , have a decline from 2005 to 2015. During the execution of the Eleventh Five-Year Plan (2005-2010), two pollutants' concentrations decreased year by year, while during the period of the Twelfth Five-Year Plan (2011-2016) annual

average concentrations of SO<sub>2</sub> and NO<sub>2</sub> firstly increased, reached a peak around 2013-2014, then decreased in 2015. The annual average concentration of PM<sub>10</sub> showed a general decreasing trend from 2005 to 2015, increased slightly in 2013, and declined sharply during the later period of the Twelfth Five-Year Plan. Since the beginning of regular PM<sub>2.5</sub> monitoring in 2013, its concentration has been a continuously declining trend.

It can be seen that average concentrations of air pollutants have a slight fluctuation annually, particularly the annual average concentration of PM<sub>10</sub> varies significantly. In addition, the concentration fluctuation is relatively stronger during the Twelfth Five-Year Plan. On one hand, the fluctuation can be attributed to rigorous environmental protection policies from the Eleventh Five-Year, in which mandatory indexes on environmental protection, ecological construction, and energy conservation were determined. Therefore, annual concentrations of SO<sub>2</sub> and NO<sub>2</sub> gradually declined. On the other hand, the original parsing on air pollutants in cities shows that construction dust, which is produced by construction activities, is the main source of PM<sub>10</sub>. Considering construction activities one after another in city development, the average annual concentration of PM<sub>10</sub> has been showing a fluctuating trend.

#### 4. Analysis and evaluation based on principal components

##### 4.1. Data source

The air quality of a city from 2005 to 2015 is shown in Table 1. The paper takes the Environmental Air Quality Standard (GB3095-1996, 2012), as the evaluation criteria, which is shown in Table 2<sup>11</sup>.

**Table 1.** Environmental air quality mg/m<sup>3</sup>

Year	SO <sub>2</sub>	NO <sub>2</sub>	PM <sub>10</sub>
2005	0.077	0.047	0.106
2006	0.065	0.048	0.113
2007	0.062	0.043	0.093
2008	0.061	0.041	0.088
2009	0.055	0.04	0.1
2010	0.054	0.045	0.096
2011	0.042	0.038	0.093
2012	0.048	0.042	0.105
2013	0.059	0.054	0.15
2014	0.049	0.054	0.133
2015	0.029	0.042	0.116

**Table 2.** GB3095-1996, GB3095-2012 mg/m<sup>3</sup>

	SO <sub>2</sub>	NO <sub>2</sub>	PM <sub>10</sub>
Level 2 (2012)	0.06	0.04	0.07
Level 1	0.02	0.05	0.04
Level 2	0.06	0.05	0.10
Level 3	0.10	0.10	0.15

##### 4.2. Data standardization

To eliminate the dimensional effects, the specific calculation method is conducted as follows:

$$X_{ij}^* = \frac{x_{ij} - \bar{x}_i}{\sigma_i} \quad (1)$$

$$\bar{X}_i = \frac{\sum_{j=1}^n X_{ij}}{n} \tag{2}$$

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^n (X_{ij} - \bar{X}_i)^2}{n-1}} \tag{3}$$

$X^*$ ,  $\bar{X}_i$ , and  $\sigma_i$  are standardized data, and sample mean and standard deviation of indicator  $i$  respectively. The result of standardization is shown as follows.

$$X = \begin{Bmatrix} 0.06 & 0.04 & 0.07 \\ 0.02 & 0.05 & 0.04 \\ 0.06 & 0.05 & 0.10 \\ 0.10 & 0.10 & 0.15 \\ 0.077 & 0.047 & 0.106 \\ 0.065 & 0.048 & 0.113 \\ 0.062 & 0.043 & 0.093 \\ 0.061 & 0.041 & 0.088 \\ 0.055 & 0.04 & 0.1 \\ 0.054 & 0.045 & 0.096 \\ 0.042 & 0.038 & 0.093 \\ 0.048 & 0.042 & 0.105 \\ 0.059 & 0.054 & 0.15 \\ 0.049 & 0.054 & 0.133 \\ 0.029 & 0.042 & 0.116 \end{Bmatrix}$$

$$X^* = \begin{Bmatrix} 0.21027 & -0.59509 & -1.18847 \\ -1.92805 & 0.07106 & -2.2517 \\ 0.21027 & 0.07106 & -0.12523 \\ 2.34859 & 3.40178 & 1.64684 \\ 1.11906 & -0.12879 & 0.08742 \\ 0.47756 & -0.06217 & 0.33551 \\ 0.31718 & -0.39525 & -0.37332 \\ 0.26373 & -0.52848 & -0.55052 \\ -0.05702 & -0.59509 & -0.12523 \\ -0.11048 & -0.26202 & -0.26699 \\ -0.75198 & -0.72832 & -0.37332 \\ -0.43123 & -0.46186 & 0.05198 \\ 0.15681 & 0.33751 & 1.64684 \\ -0.37777 & 0.33751 & 1.04434 \\ -1.44693 & -0.46186 & 0.44183 \end{Bmatrix}$$

**4.3. Solve the covariance matrix of standardized data**

Calculate the correlation coefficient by using the normalized sample data to get the covariance matrix  $R$ .

$$R = \begin{Bmatrix} 1 & 0.622 & 0.532 \\ 0.622 & 1 & 0.546 \\ 0.532 & 0.546 & 1 \end{Bmatrix}$$

#### 4.4. Calculate eigenvalues and eigenvectors.

The eigenvalues, eigenvectors, interpretation variance and cumulative interpretation variance can be obtained by solving the covariance matrix. The concrete results are shown in Table 3, among them,  $\mu_1$ ,  $\mu_2$  and  $\mu_3$  represent the corresponding eigenvector of the principal component respectively<sup>12-13</sup>.

**Table 3.** Characteristic value and characteristic vector

Principal Component	Eigenvalues	interpretation variance %	cumulative interpretation variance %	eigenvectors			
				SO <sub>2</sub>	NO <sub>2</sub>	PM <sub>10</sub>	
Z <sub>1</sub>	2.134	71.139	71.139	0.5847	0.5891	0.5578	$\mu_1$
Z <sub>2</sub>	0.488	16.280	87.419	-0.4462	-0.3408	0.8275	$\mu_2$
Z <sub>3</sub>	0.377	12.581	100.000	0.6776	-0.7327	0.0636	$\mu_3$

#### 4.5. Determine the principle components and comprehensive evaluation

According to the analytical principles of principal component, the cumulative interpretation variance can be regarded as principal component when it is over 85%. It can be seen from Table 3 that the cumulative contribution rate of the first two eigenvalues is 87.419%, hence the first two eigenvalues shall be chosen. Based on the comprehensive analysis of main components of Z<sub>1</sub> and Z<sub>2</sub>, the comprehensive principal component Z<sub>1-2</sub> is obtained, as shown in Table 4.

**Table 4.** Principal component, comprehensive principal component and ranking

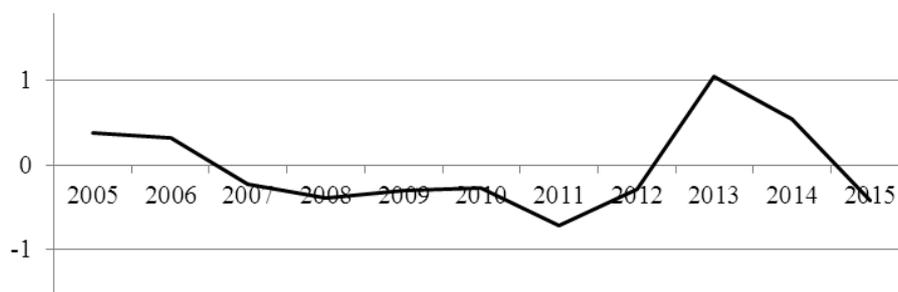
Serial number	Year	Z <sub>1</sub>	Z <sub>2</sub>	Z <sub>1-2</sub>	Ranking	Grade
1	2005	0.6272	-0.3831	0.3838	9	Level 3
2	2006	0.4298	0.0857	0.3197	8	Level 3
3	2007	-0.2556	-0.3157	-0.2333	7	Level 2 (2012)
4	2008	-0.4642	-0.3931	-0.3942	3	Level 2 (2012)
5	2009	-0.4538	0.1246	-0.3025	4	Level 2 (2012)
6	2010	-0.3679	-0.0823	-0.2751	6	Level 2 (2012)
7	2011	-1.0770	0.2748	-0.7214	1	Level 2 (2012)
8	2012	-0.4952	0.3928	-0.2883	5	Level 2 (2012)
9	2013	1.2091	1.1778	1.0519	11	Level 3
10	2014	0.5605	0.9177	0.5481	10	Level 3
11	2015	-0.8716	1.1686	-0.4298	2	Level 2 (2012)

The main components can be ranked as follows: Level 3<2013<2014<2005<2006<the standard in 2012<Level 2<2007<2010<2012< 2009<2008<2015 <2010<Level 1. Therefore, it can be concluded that the air quality from 2005 to 2015 is shown as follows according to order from superior to inferior: 2011, 2015, 2008, 2009, 2012, 2010, 2007, 2006, 2005, 2014, and 2013. The air quality in 2007, 2008, 2009, 2010, 2012, and 2015, are between Level 1 and Level 2, which should be treated as Level 2, meanwhile the air quality in 2005, 2006, 2013 and 2014 are between Level 3 (in 2012) and Level 3 (in 1996), which should be seen as Level 3.

From constitution characteristics of the eigenvector of the first principal component, the proportion of SO<sub>2</sub>, NO<sub>2</sub> and PM<sub>10</sub> accounted in the whole air quality assessment are almost equal, indicating that these three kinds of pollutants have similar roles in air pollution. Therefore, equal attention should be paid on the control of the three pollutants.

## 5. Results

PCA shows that air quality of the city from 2005 to 2015 was generally improved, as shown in Figure 2.



**Figure 2.** Principal component analysis results of atmospheric environmental quality

(1) During the period of the Eleventh Five-Year Plan, the air quality of the city tends to improve, but the degree of improvement is relatively gentle, which mainly depends on the implementation of environmental protection measures during the period of the Eleventh Five-Year Plan.

Comprehensive eco-city construction is the most important environment protection policy in this city, highlighting regional environmental protection, ecological construction, and the mandatory index of energy conservation. Although environmental management have been much more strengthened than before, most environmental protection activities focus on industrial pollution control, such as desulfurization projects of coal-fired power plants and boilers, large-scale central heating boiler room construction. Few measures have been conducted on air pollution by motor vehicle, dust, and extensive use of coal. As a result, the air quality of the city generally improved during the period of the Eleventh Five-Year Plan, but also gradually.

(2) During the period of the Twelfth Five-Year Plan, air quality of the city fluctuated significantly. Compared with the later period since 2013, air pollution in the beginning years of this period was much server, which can be attributed to the rapid development of regional economy with increasing atmospheric pollutant emission. The regional economy, particularly in the field of aerospace, petrochemical industry and metro construction, development rapidly, meanwhile the growth rate of motor vehicle increased significantly. Compared with 2006, the total consumption of the coal of the city in 2012 increased by 38.2%, construction site area increased by 155%, and motor vehicle ownership increased by 101.7%, resulting in sulfur dioxide and nitrogen oxides increased by 38.3% and 192.1% respectively. Unfortunately, environmental protection measures were constantly strengthened but fell behind regional economic development, therefore air quality of the city worsen in the early stage of the Twelfth Five-Year Plan period.

From 2013 to 2015, air quality of the city was back along the track of improvement, indicating new promulgated environmental protection policies played an important role in regional air quality improvement. On the national level, in 2013, Chinese national government promulgated renewed Environmental Air Quality Standards (GB3095-2012), implemented the Action Plan on Air Pollution Control. In 2015, renewed the Law of Environmental Protection of the People's Republic of China was enforced.

On the regional level, the city revised the Air Pollution Prevention Ordinance. Particularly, the Fresh Air Action Program was issued in 2013, in which 66 measures on 10 aspects were raised for regional air quality improvement. Special attention was focused on strict measures on five main air pollution sources, including coal consumption, dust, vehicle, industrial activities and new construction projects, while rigorous requirement was set on  $PM_{2.5}$  control. The Fresh Air Action Program stipulated the goal that till 2017, the air quality should be significantly improved with sharp reduction of heavily polluted days and gradual increase of days with excellent/good air quality. Additionally, the annual average concentration of  $PM_{2.5}$  should decreased by 25% compared with 2012. Through the implementation of a series of new policies, environmental protection have been strengthened in an unprecedented way, and abovementioned goal was completed in advance in 2015, resulting in the decreasing concentration of atmospheric pollutants from 2013 to 2015.

(3)With increasingly intensive environmental protection in China from the period of the Eleventh-Five-Year-Plan and Twelfth Five-Year Plan, the air quality of the city has improved. In addition, the concentration of main atmospheric pollutants still has potential for further decreasing.

### Acknowledgments

This work was financially supported by National Science and Technology Support Program (2014BAC3B03) and Environmental protection commonweal research program (201409004) fund.

### References

- [1] Ministry of Environmental Protection of the People 's Republic of China. GB3095-2012 Ambient Air Quality Standard[S].Beijing: China Environmental Science Press, 2012
- [2] Zhang Fanyong. Multi - attribute comprehensive evaluation method and demonstration [J]. Statistics and Decision, 2015,(21):65-68.
- [3] Ma Guangwen, He Lihuan, Wang Xiaofei, et al. Comprehensive Evaluation Method of Rural Environmental Quality and Its Application in Typical Areas [J]. Chinese Journal of Environmental, 2014, (05): 10-15.
- [4] Han Peng, Li Tao. Study on Comprehensive Evaluation Method of Resources and Environment Carrying Capacity - Taking Zhongyuan Economic Zone as an Example [J]. Journal of Basic Science and Engineering, 2015, (S1): 88-101.
- [5] Zhang Yunhai, Ma Yanjun, Sun Caitao, et al. Application of Grey Clustering Method for Atmospheric Environmental Quality Assessment in Shenyang Economic Zone [J]. Environmental Science & Technology, 2010, (S1): 426-428.
- [6] Zhong Zhiming. Discussion on Method to Determine Grade of Atmosphere Environment Evaluation Work [J]. Coal Engineering, 2011, (06): 24-25.
- [7] Qiao Han. A Review on the Comprehensive Evaluation Method of Principal Components and Its Applicability [J]. Journal of Tianjin University of Commerce, 2015, 35(3): 9-14.
- [8] Pang Xiaoning,Wang Liu. Construction of Index System for Comprehensive Evaluation of Circular Economy Based on Principal Component Analysis [J]. Statistics and Decision, 2012, (14): 56-58.
- [9] Li Fenhong, Liu Xinping. Improvement and Application of Comprehensive Evaluation Model of Principal Component Analysis [J]. Statistics and Decision, 2010, (02): 162-163.
- [10] Liu Tao. Study on the Influencing Factors of Industrial Structure Change in Shandong Province Based on Principal Component [J]. Analysis Journal of Shandong University: Philosophy and Social Sciences, 2011, 3: 107-112.
- [11] Ministry of Environmental Protection of the People 's Republic of China. GB3095-1996 Ambient Air Quality Standard [S]. Beijing: China Environmental Science Press, 1996
- [12] Zheng Chengde, Yin Desong. Application of Principal Component Analysis Method in Urban Atmospheric Environmental Quality Assessment [J]. Journal of Dalian Railway Institute, 2002, (03): 39-41.
- [13] Tong Qihui. Application of Principal Component Analysis Method in Index Comprehensive Evaluation [J]. Journal of Beijing Institute of Technology:Social Sciences Edition, 2002, (01): 59-61.