# Comprehensive evaluation of environmental pollution in Shandong Province based on principal component and cluster analysis

Mingqing Zhao
Shandong University of Science and Technology College of Mathematics and Systems Science
Qingdao, China

Jinzhi Yue, Haimei Liu
Shandong University of Science and Technology College of Mathematics and Systems Science
Weifang, China
13053858693@163.com

**Abstract:** By principal component analysis and cluster analysis, overall analysis and evaluation on environmental pollution situation in Shandong province is conducted.Studies show that Weifang, Linyi and Zibo's pollution is serious, particularly in the sulfur dioxide emissions in the more serious aspects. Although Weihai, Rizhao and Dongying in sulfur dioxide and soot aspects slightly heavier than the other pollutants, the overall environmental situation is comparatively ideal in the province.

***Keywords: Principal component analysis; cluster analysis; environmental pollution assessment***

## I. Introduction

In recent years, with the development of our society, life has been greatly improved. But at the same time, the environment that people living also has a great change. In the process of sustainable development, we will face the dual pressures of environmental pollution and economic development. We need to reduce environmental pollution under the precondition of meeting people's living standards.

As the conflict between economic development and environmental protection, more and more scholars begin to focus on researches concerning economic and environment, especially researches on comprehensive evaluation of the environment pollution have caused widely public concern. The literature [1] has carried on the comprehensive evaluation to regional environmental situation in Shandong province by fuzzy cluster analysis. The literature [2] has carried on the overall evaluation to our country environmental pollution condition by principal component analysis. The literature [3] is discussed the application and characteristics of comprehensive environmental evaluation ideas in the health impact assessment of air pollution. The literature [4] has carried on the comprehensive evaluation to 1990-2009 China's provincial environmental pollution using the principal component analysis. The literature [5] has carried on the comprehensive evaluation of regional environmental quality to our country through the principal component analysis.[1]

The work is supported by the project of Shandong province graduate education innovation project (SDYY14086), Shandong University of Science and Technology graduate students in science and Technology Innovation Fund Project (YC150337)

As you can see, most of literatures above focus on the use of a single method to analyze the environmental quality of a country or some province, while literatures that using several methods less. In this paper, using principal component analysis with clustering analysis[1,3], it assesses environmental quality in major cities in Shandong province. Firstly, analyzing environment pollution in Shandong province by the principal component analysis. Then, clustering analysis method is used to analyze types of environmental pollution in Shandong province, looking forward to reflect the current state of the current environmental quality of Shandong province more fully.

## II. Data source

In this paper, with the help of Spss software, the relevant data of major pollutants' discharging in 17 cities in Shandong province in 2012 were analyzed, with a total of five measurements: wastewater emissions ($x_1$)、chemical oxygen demand emissions( $x_2$ ), ammonia nitrogen emissions ($x_3$), emissions of sulfur dioxide ($x_4$) and soot emissions( $x_5$). The data is found in Statistical Yearbook of Shandong in 2012, shown in table 1.

Table 1: Discharge of major pollutants in cities of Shandong province in 2012

| region | Quantiy of wasteaer (Million tons） | Chemical oxygen demand emissions（Ton） | Ammonia nitrogen emissions（Ton） | Sulfur dioxide emissios（Ton) | Soot emissions（Ton） |
|---|---|---|---|---|---|
| Ji nan | 33338.42529 | 115806.5925 | 9613.441 | 114520.2423 | 62824.84 |
| Qing dao | 51310.98922 | 149429.6717 | 12573.2789 | 99623 | 41193.026 |
| Zibo | 36248.10981 | 66717.9743 | 6185.3493 | 224245.83 | 62844.8052 |
| Zao zhuang | 20393.12123 | 55596.2217 | 5903.1164 | 87204.951 | 31061.8595 |
| Dong ying | 20540.43681 | 68534.2832 | 4037.9451 | 57300.26 | 7979.033 |
| Yan tai | 32069.98376 | 148806.8861 | 12818.138 | 100230.525 | 45914.492 |
| Wei | 52153.5 | 181290.9 | 17585.75 | 151821. | 52522.29 |

 

| | | | | | |
|---|---|---|---|---|---|
| fang | 9175 | 241 | 14 | 9003 | 35 |
| Ji ning | 42425.92544 | 143001.3199 | 14443.8536 | 145224.0145 | 62312.1357 |
| Tai an | 23274.79089 | 121800.0879 | 10089.7824 | 90302.8424 | 29549.214 |
| Wei hai | 10960.99705 | 31685.4701 | 4498.2268 | 47054.73 | 17018.212 |
| Ri zhao | 15252.18771 | 48861.9409 | 4975.4474 | 66459.775 | 37208.036 |
| Lai wu | 5729.3517 | 18120.1 | 2064 | 88387 | 78500.608 |
| Lin yi | 36252.0225 | 159253.921 | 17836.1925 | 115697.088 | 52472.012 |
| De zhou | 22196.90844 | 170735.6205 | 13637.6947 | 93278.534 | 27851.4347 |
| Liao cheng | 22975.76333 | 150651.9288 | 9890.3806 | 85962.034 | 21625.378 |
| Bin zhou | 27848.71367 | 146204.6798 | 8505.2273 | 84555.7614 | 21896.349 |
| Heze | 26129.04424 | 144735.6404 | 13925.082 | 96932.833 | 42495.74 |

## III. Analysis of environmental pollution degree of Shandong Provence

Using principal component analysis, we analyzed environmental pollution of Shandong province by the indexes of wastewater contained converted to reflect the original index information of several principal components, specific steps are as follows:

*A. Calculating the sample mean and sample standard deviation of each index*

The sample means of five indicators respectively are:

$$u_1 = 2.8182E4, u_2 = 1.1301E5, u_3 = 9.9166E3,$$

$$u_4 = 1.0287E5, u_5 = 4.0898E4;$$

Sample standard deviations respectively are:

$$s_1 = 1.29001609E4, s_2 = 5.2678775E4,$$

$$s_3 = 4.80600991E3, s_4 = 4.1169836E4,$$

$$s_5 = 1.92477750E4.$$

*B. Data standardization process*

The results of data standardization process are shown in table 2.

Table 2 Standardization data of major pollutants emissions in cities of Shandong province in 2012

| Region | quantity of wastewater (Million tons) | Chemical oxygen demand emissions（Ton） | Ammonia nitrogen emissions（Ton） | Sulfur dioxide emisions（Ton） | Soot emissions（Ton） |
|---|---|---|---|---|---|
| Jinan | 0.39969 | 0.053017 | -0.06309 | 0.28296 | 1.13918 |
| Qing dao | 1.79289 | 0.691283 | 0.55277 | -0.07888 | 0.01532 |
| Zibo | 0.62524 | -0.87883 | -0.77638 | 2.94816 | 1.14021 |
| Zao zhuag | -0.6038 | -1.08996 | -0.83511 | -0.38051 | -0.51104 |
| Dong ying | -0.5923 | -0.84435 | -1.2232 | -1.10689 | -1.71028 |
| Yan tai | 0.30136 | 0.67946 | 0.60372 | -0.06413 | 0.26062 |
| Wei fang | 1.85821 | 1.29610 | 1.59573 | 1.18901 | 0.60392 |
| Ji ning | 1.10414 | 0.56925 | 0.94199 | 1.02875 | 1.11254 |
| Taian | -0.3804 | 0.16679 | 0.03603 | -0.30527 | -0.58963 |
| Wei hai | -1.3349 | -1.54385 | -1.12742 | -1.35575 | -1.24066 |
| Ri zhao | -1.0023 | -1.21779 | -1.02813 | -0.88441 | -0.19172 |
| Laiwu | -1.7405 | -1.80136 | -1.63392 | -0.35180 | 1.95359 |
| Linyi | 0.62554 | 0.87778 | 1.64784 | 0.31155 | 0.60131 |
| De zhou | -0.4639 | 1.09573 | 0.77425 | -0.23299 | -0.67783 |
| Liao cheng | -0.4036 | 0.71449 | -0.00546 | -0.41070 | -1.00130 |
| Bin zhou | -0.0258 | 0.63006 | -0.29368 | -0.44486 | -0.98722 |
| Heze | -0.1591 | 0.60218 | 0.83405 | -0.14423 | 0.082999 |

By calculating, we get correlation coefficient matrix of the sample.

$$R = \begin{pmatrix} 1.000 & 0.694 & 0.738 & 0.622 & 0.295 \\ 0.694 & 1.000 & 0.911 & 0.245 & -0.19 \\ 0.738 & 0.911 & 1.000 & 0.344 & 0.185 \\ 0.622 & 0.245 & 0.344 & 1.000 & 0.634 \\ 0.295 & -0.19 & 0.185 & 0.634 & 1.000 \end{pmatrix}$$

*C. Determining the principal components*

The characteristic roots respectively are: $\lambda_1$=2.955, $\lambda_2$=1.399, $\lambda_3$=0.398, $\lambda_4$=0.182, $\lambda_5$=0.065.

Easy to know when the number of principal component is 2, the cumulative contribution rate can be 87.094%. It's suitable for principal component analysis, and the principal components are as follows:

$$\begin{cases} F_1 = 0.308x_1 + 0.277x_2 + 0.300x_3 + 0.233x_4 + 0.148x_5 \\ F_2 = -0.020x_1 - 0.373x_2 - 0.252x_3 + 0.431x_4 + 0.571x_5 \end{cases}$$

The first principal component $F_1$ shows obvious positive correlation with $x_1$, $x_2$, $x_3$, $x_4$, $x_5$, while these variables reflect pollution condition of waste water and exhaust gas. So we can think that the first principal component $F_1$ is the representative of the emission of waste water and exhaust gas.

The second principal component $F_2$ shows a strong positive correlation with $x_4$, $x_5$, shows a strong negative correlation with $x_1$, $x_2$, $x_3$, but $x_4$, $x_5$ comprehensively reflect the emission of industrial pollutants. So we can think

that the second principal component $F_2$ is the representative of the emission of industrial pollutants.

*D. Calculating the scores and sorting them*

By calculation, the results of scoring and sorting in areas are shown in table 3. And $F_0 = 0.59109F_1 + 0.27985F_2$

Table 3 the score and sorting of regions

| N | region | $F_1$ | $F_1$ (s) | $F_2$ | $F_2$ (s) | $F_0$ | $F_0$ (s) |
|---|---|---|---|---|---|---|---|
| 1 | Jinan | 0.3534 | 11 | 0.76073 | 15 | 0.2571 | 12 |
| 2 | Qingdao | 0.8941 | 14 | -0.4577 | 6 | 0.2993 | 13 |
| 3 | Zibo | 0.5729 | 13 | 2.43239 | 17 | 0.1127 | 9 |
| 4 | Zaozhuang | -0.90 | 5 | 0.17283 | 12 | -0.5520 | 4 |
| 5 | Dongying | -1.29 | 3 | -0.8190 | 4 | -0.5957 | 3 |
| 6 | Yantai | 0.4860 | 12 | -0.2901 | 9 | 0.1384 | 11 |
| 7 | Weifang | 1.7780 | 17 | -0.0647 | 11 | 1.5058 | 17 |
| 8 | Jining | 1.1856 | 16 | 0.60725 | 14 | 0.7686 | 16 |
| 9 | Taian | -0.218 | 6 | -0.5319 | 5 | 0.4533 | 15 |
| 10 | Weihai | -1.677 | 1 | -0.4067 | 8 | -1.0693 | 1 |
| 11 | Rizhao | -1.189 | 4 | 0.24224 | 13 | -1.0061 | 2 |
| 12 | Laiwu | -1.319 | 2 | 2.08174 | 16 | -0.09871 | 6 |
| 13 | Linyi | 1.0925 | 15 | -0.2771 | 10 | 0.3897 | 14 |
| 14 | Dezhou | 0.2384 | 9 | -1.0818 | 1 | 0.0161 | 8 |
| 15 | Liaocheng | -0.172 | 7 | -1.0057 | 2 | -0.0231 | 7 |
| 16 | Binzhou | -0.171 | 8 | -0.9158 | 3 | -0.3306 | 5 |
| 17 | Heze | 0.3468 | 10 | -0.4462 | 7. | 0.1238 | 10 |

As can be seen from table 3, Qingdao, Linyi, Weifang and Jining get higher scores on $F_1$, this interprets that waste water, chemical oxygen demand (cod), ammonia nitrogen, sulfur dioxide and fuel dust are serious in these cities. That's because there are heavy industries in these cities,which make the dust, sulfur dioxide and others more. Furthermore, because of these developed cities, large population, high automobile exhaust, so waste water caused by life could also be more. At the same time, we find that air environment is better in most of these cities, pollution is relatively low. Jinan, Laiwu, Zibo and Jining get higher scores on $F_2$, explaining that pollution of sulfur dioxide and fuel dust in these cities are more serious. Pollution by, Jining, Weifang, Taian and Linyi are at the top of the population list, Weihai, Dongying, Rizhao ,Laiwu and other regions have better environmental protection.

## IV. The type analysis of environmental pollution in Shandong province

Using principal component to do cluster analysis can reflect the characteristics of the class better. We take the two principal components $F_1$ and $F_2$ as variables to cluster training samples, so as to analyze the types of environmental pollution in Shandong province.

### A. Clustering scatter pattern

Firstly, making a scatter diagram on $F_1$ and $F_2$, as shown in figure 1.

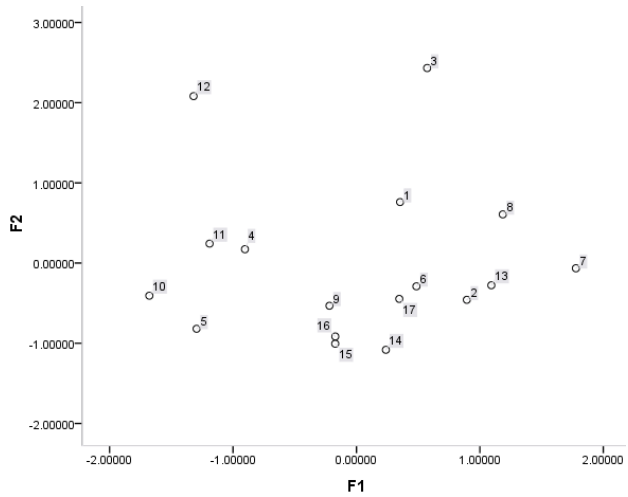


Figure 1 Cluster scatter diagram

As can seen from figure 1 intuitively, 17 cities in Shandong province can be divided into 5 groups, which more appropriate. This intuitive judgment is good for hierarchical clustering method to determine the clustering number.

### B. Hierarchical clustering figure

Clustering with system clustering method and using the sum of squared residuals method calculate the distance between classes, we get the hierarchical clustering diagram shown in figure 2. Combined with intuitive analysis above, the distance between the class is 7, and is divided into 5 types.

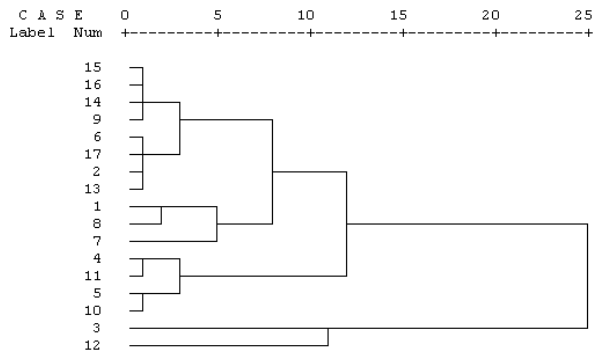


Figure 2 Pedigree cluster diagram

The first category: Liaocheng, Binzhou, Dezhou, Taian, Yantai, Heze, Qingdao and Linyi. Liaocheng, Dezhou, Taian

and Binzhou ranked in the middle on $F_1$, but ranking the top five on $F_2$. That is to say, Liaocheng, Dezhou, Taian and Binzhou, have a heavy pollution on waste water, chemical oxygen demand (cod) and ammonia nitrogen emissions; Yantai, Qingdao, Heze and Linyi ranked in the middle on $F_1$ and $F_2$, that is, Yantai, Qingdao, Heze and Linyi have a heavy pollution on sulfur dioxide and soot emissions.

The second category: Jinan, Jining and Weifang. These three cities ranked at the bottom of seven both on $F_1$ and $F_2$. That is to say, Jinan, Jining and Weifang suffer more serious pollution on wastewater, chemical oxygen demand (cod), ammonia nitrogen, sulfur dioxide and soot emissions. These cities' comprehensive environment quality is poorer, rank at the bottom of the province.

The third category: Zaozhuang, Rizhao, Dongying and Weihai. Zaozhuang and Rizhao rank six on $F_2$, but ranking the top five on $F_1$,that is to say, Zaozhuang and Rizhao suffer serious pollution on sulfur dioxide and soot emissions;Dongying ranked the top three on $F_1$, and the top four on $F_2$ ,that is, Dongying's pollution on dioxide and carbon emissions is heavier than other pollution ;Weihai ranked middle on $F_2$, but ranked first on $F_1$, that is slightly heavier on sulfur dioxide and soot emissions.

The forth category: Zibo. Zibo ranked at the bottom fifth on $F_1$, and ranked at the last on $F_2$, that is heavier in terms of sulfur dioxide and soot emissions.

The fifth category: Laiwu. Laiwu ranked at last second on $F_2$,but ranked second on $F_1$, that is heavier in terms of sulfur dioxide and soot emissions.

By calculating, we get the mean values of the $F_0$ respectively: $u_{10}$ =0.1333625, $u_{20}$ =0.843833, $u_{30}$ = -0.805775, $u_{40}$ = 0.1127, $u_{50}$ =-0.09871. Accordingly, we can get $u_{30} < u_{50} < u_{40} < u_{10} < u_{20}$ . All kinds of urban environmental quality from good to bad arrangement for the third, the fifth, the fourth, the first and the second.

## V. Conclusion

This article selects five indicators to make the comprehensive evaluation on environmental quality in regions of Shandong province. First, analyzing regional environmental pollution degree by principal component analysis; through clustering analysis, we analyzed the environment pollution types, and found out the polluting emissions of regions which are serious. The results show that the degree of pollution is heavier in Linyi, Weifang and Qingdao, and there are more pollution types;And Jinan, Jining, Zibo and Laiwu are heavier in terms of sulfur dioxide and soot emissions. But in Dongying, Weihai and Rizhao, pollution emissions is lighter, environmental protection is ideal. Through this analysis, we suggest that environment department can take the following measures for all kinds of pollution, in order to make the province's environmental quality better.

a) To change waste material into things of value, reducing wastewater emissions

Quantity of wastewater in cities like Weifang, Qingdao and Jining are high. Therefore, these cities should focus on wastewater treatment technology, improve the utilization rate of wasteing water and turn "waste" into wealth.

b) To optimize of industrial structure, reducing emissions of three-waste

In cities like Weihai, Rizhao, Laiwu, industrial emissions of three-waste is relatively high. So, with the principle of "less pollution, we should develop the tertiary industry like high and new technology industry, tourism and service, so as to reduce emissions of three-waste, to protect the environment.

c) To increase the environmental investment

At present, economic development model is still relatively backward in Heze and Linyi. We not only develop the economy but also increase capital investment of environmental protection, to cocreate a win-win situation both the economic development and environmental protection.

## References

[1] Zhang wenlan, Zheng zhaopei. Fuzzy clustering analysis of environmental pollution in various regions of Shandong Province[J]. 2009, 11(3):34-36

[2]Qi Yongfang. Application of principai component analysis in environmental pollution[J]. Pingxiang advanced specialist.2 012,29(3):40-44

[3]Wan yue, Yang hongwei. The application and characteristics of comprehensive environmental assessment in the assessment of the health effects of air pollution [J].Journal of environment and health,2006,23(4):376-379.

[4]Qu xiaoe.1990—2009 year, Comprehensive assessment of environmental pollution in china [J].China population. Resources and environment, 2012, 22(5):158-163.

[5]Yuan xiaohe, Li xiaoqing. Comprehensive assessment of environmental quality in various regions of China[J]. 2011, 11(5):89-90.