CHEMICAL ENGINEERING TRANSACTIONS
VOL. 62, 2017
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors: Fei Song, Haibo Wang, Fang He

ISBN 978-88-95608- 60-0; **ISSN** 2283-9216
DOI: 10.3303/CET1762085

Analysis of the Relationship Between Economic Development and Environmental Pollution of Chemical Industry Based on Principal Component Analysis

Yixin Xing

School of Public Administration, Hohai University, Nanjing 211100, China
yixinhhu@163.com

After respectively analyzing and evaluating the economic development level and environmental pollution status of chemical industry in 30 regions in China by using the principal component analysis (PCA) of multivariate statistical analysis, we obtain the comprehensive scores of the economic development and environmental pollution of chemical industry in all regions. Through the cluster analysis of the two groups of comprehensive scores, the 30 regions in China are divided into four categories. The analysis of the comprehensive status of the four categories of regions reveals the relationship between regional economic development and environmental pollution of chemical industry.

1. Introduction

The social development of human beings depends on the ecological environment and natural resources, so environmental changes affect the direction of social development of human beings all the time. China's economy has developed rapidly since the reform and opening up. As one of the pillar industries in the industrialization process of China, the chemical industry has played a decisive role in promoting China's economic development. However, since the 1980s, the rapid development of the chemical industry has brought serious environmental problems while promoting economic development. The waste water, waste gas and waste residue discharged by the chemical industry are at the forefront of emissions of the industrial sector and chemical COD and inhalable particles harmful do great harm to human health (Mooibroek and Schaap, 2011; Ovadnevaitė et al., 2007).

The relationship between economic development and the environment has always been the research focus of academic world. The earliest research could date back to the one conducted by Grossman and the World Bank (World Bank, 1992) in the 1990s. According to the two researches, there is an inverted "U" curve between environmental pollution and per capita income. Lopez (1994), Jones and Manuelli (1994), Selden and Song (1995), Stocky (1998), Lopez and Mitra (2000) have respectively proved the possibility of an "inverted U" curve between environmental pollution and economic growth by constructing different models. Later, many scholars paid attention to the environmental pollution of chemical industry and analyzed the characteristics of pollution of chemical industry by using PCA (Jain et al., 2017; Aziz et al., 2017; Seklaoui et al., 2016). Domestic scholars have also paid increasing attesting to the relationship between economy and environment. Zhang and Gong Longyan analyzed the causes, status quo and characteristics of heavy metal pollution in cities by adopting PCA and 3D model.

Based on the existing researches, this paper uses PCA and cluster analysis to analyze the economic development and environmental pollution of chemical industry in 30 provinces and cities in China in 2015. And according to the actual situation, this paper puts forward the contradictions between economy and environment in each region, providing a realistic basis for the sustainable development of economy, society and environment in China.

Please cite this article as: Yixin Xing, 2017, Analysis of the relationship between economic development and environmental pollution of chemical industry based on principal component analysis, Chemical Engineering Transactions, 62, 505-510 DOI:10.3303/CET1762085

2. Research Methods and Indicator Selection

2.1 Research methods and models

(1) PCA

PCA is an analysis method that is to study how to explain multivariate variance—covariance structure by a few principal components, that is, to find a few principal components so that they retain the information of the original variables as much as possible. The use of PCA can simplify the analysis and reflect the essence of things.

Assume the original variables are x_1, x_2, ..., x_n. and the new variables obtained after PCA are y_1, y_2, ... y_m, which are linear combinations of x_1, x_2, ..., x_n($m<n$). Most information of the original variable is retained and unrelated. The steps of PCA are as follows:

First, in order to eliminate the impact of different dimensions and orders of magnitudes of data, the original data needs to be standardized:

$$x_{ij}^{*}=(x_{ij}-\bar{x}_i/e_i, \ i=1, 2, \ldots, n; j=1, 2, \ldots, p \tag{1}$$

Where, x_{ij} is the original data of the i_{th} indicator in the j_{th} partition, $\bar{x_i}$ and e_i are respectively sample average and standard deviation of the i_{th} indicator.

Second, the correlation coefficient matrix $R=\left(r_{ij}\right)_{n\times n}$ is calculated by using the standardized data table $(x_{ij}^{*})_{n\times p}$, where,

$$r_{ij}=\frac{1}{n}\sum_{k=1}^{n}\left(x_{ki}-\bar{x}_i\right)(x_{kj}-\bar{x}_j)/e_ie_j \tag{2}$$

Third, the eigenvalue and eigenvector of R are calculated. According to the characteristic equation $|R-\lambda I|=0$, the characteristic root λ_i is calculated and arranged in descending order: $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_n$ and the corresponding eigenvectors u_1, u_2, …, u_n are obtained.

Fourth, the variance contribution rate and cumulating contribution rate of each principal component are obtained.

$$e=\lambda_i/\sum_{i=1}^{n}\lambda_i \tag{3}$$

$$E_j=\sum_{j=1}^{m}\lambda_j/\sum_{i=1}^{n}\lambda_i \tag{4}$$

Fifth, the principal component is calculated and the principal component to be retained whose eigenvalue is greater than 1 or the cumulating variance contribution rate is greater than 85% is determined.

$$y_i=\sum_{j=1}^{p}\sum_{i=1}^{n}u_{ij}x_{ij}^{*} \tag{5}$$

(2) Cluster analysis

For the evaluation scores of economic development and environmental pollution of chemical industry obtained from PCA, we need to classify them so as to find the differences and regional characteristics of economic development and environmental pollution in various regions. There are many cluster analysis methods, of which K-means is the most famous and most widely used because of its good effect.

The calculation steps are as follows: Firstly, k objects are randomly selected from n data objects as the initial cluster centers. For the remaining objects, they are respectively assigned to the most similar clusters (represented by the cluster centers) according to their similarities (distances) with the cluster centers. Then we calculate the center of each new cluster (the average of all objects in the cluster) and repeat this process until the standard measure function begins to converge. Usually, the mean square error is taken as a standard measure function and the specific definition is as follows:

$$E=\sum_{i=1}^{k}\sum_{p\in C_i}|p-m_i|^2 \tag{6}$$

Where, E is the sum of the mean square error of all the objects in the cluster; p is a point in the cluster that can be multidimensional; and m_i is the mean value of cluster C_i which can be multidimensional. The cluster standard in the formula aims at making the k clusters have the characteristic that the clusters themselves are as compact as possible while the clusters are separated as much as possible.

2.2 Indicator selection and data source

(1) Selection of economic indicators

To better reflect the degree and level of economic development in various regions, this paper selects a total of 8 indicators reflecting economic position in 30 regions from "China Statistical Yearbook 2016" and "China City Statistical Yearbook 2016". The meanings of indicators are as follows:
x_1: gross regional domestic product (GDP) (100 million yuan)
x_2: GDP per capita (yuan / person)
x_3: gross industrial production (100 million yuan)
x_4: fixed investment (100 million yuan)
x_5: household consumption expenditure (yuan)
x_6: proportion of primary industry (%)
x_7: proportion of secondary industry (%)
x_8: proportion of tertiary industry (%)
Among the above eight indicators, x_1, x_3 and x_4 mainly reflect the overall economic scale and economic aggregate of the region; x_2 and x_5 mainly reflect the regional economic benefits and people's living standards; x_6, x_7 and x_8 mainly reflect the regional economic structure.
(2) Selection of environmental indicators
There are six indicators reflecting the level of environmental pollution of chemical industry in the regions. The original data come from "China Statistical Yearbook 2016" and "China Energy Statistical Yearbook 2016". The meanings of indicators are as follows:
y_1: total industrial wastewater discharge (10,000 tons)
y_2: domestic sewage discharge (10,000 tons)
y_3: total industrial waste gas discharge (100 million cubic meters)
y_4: industrial sulfur dioxide discharge (ton)
y_5: industrial dust discharge (ton)
y_6: industrial solid waste discharge (ton)
Among the above six indicators reflecting the level of environmental pollution of chemical industry, y_1 and y_2 mainly reflect the pollution level of water environment; y_3, y_4 and y_5 mainly reflect the pollution level of atmospheric environment; y_6 mainly reflects the impact of industrial solid waste on the total environment.

3. Results and Analysis

All the indicators need to be standardized because they have different dimensions and numerical value so as to eliminate the impact of differences on results.

3.1 PCA of economic level

The KMO test and Bartlett spherical test are conducted on each variable of economic development level and the results show that the KMO test is 0.686, indicating that the statistical effect is still good and suitable for factor analysis. The Sig of Bartlett spherical test is 0.000, showing that there is a significant correlation between each variable. The characteristic root and contribution rate of principal components of regional economic development are obtained as shown in Table 1.

Table 1: Characteristic root and contribution rate of principal components of economic development

Components	Initial eigenvalue			Extraction of quadratic sum of load	
	Total	Variance percentage	Accumulation %	Total	Variance percentage
1	3.730	46.624	46.624	3.730	46.624
2	2.974	37.170	83.794	2.974	37.170
3	.952	11.904	95.698		
4	.196	2.450	98.147		
5	.088	1.104	99.251		
6	.057	.707	99.958		
7	.002	.027	99.986		
8	.001	.014	100.00		

It can be seen from Table 1 that the cumulating contribution rate of the first two components has reached 83.794%. When these two principal components replace the original indicators, most of the information of the original data can be well reflected. The two principal components are recorded as Y1 and Y2.The coefficients of each principal component are shown in Table 2.

Table 2: Coefficient matrix of principal components score of economic development level

	Components	
	1	2
Zscore (GDP)	.059	.273
Zscore (per capita GDP)	.252	.044
Zscore (industrial added value)	.023	.291
Zscore (fixed investment)	-.048	.295
Zscore (household consumption level)	.279	-.003
Zscore (proportion of primary industry)	-.202	-.087
Zscore (proportion of secondary industry)	-.170	.201
Zscore (proportion of tertiary industry)	.272	-.129

From Table 2, we can get the decomposition expressions of the two principal components.

Y1=0.059*Zscore (GDP)+0.252*Zscore (per capita GDP)+0.023*Zscore (industrial added value) -0.048*Zscore (fixed investment)+0.279*Zscore (household consumption level)-0.202*Zscore (proportion of primary industry)-0.170*Zscore (proportion of secondary industry)+0.272* Zscore (proportion of tertiary industry

Y2=0.273*Zscore (GDP)+0.044*Zscore (per capita GDP)+0.291* Zscore (industrial added value)+0.295* Zscore (fixed investment)-0.003* Zscore (household consumption level) -0.087* Zscore (proportion of primary industry) +0.021* Zscore (proportion of secondary industry)-0.129* Zscore (proportion of tertiary industry)

3.2 PCA of environmental pollution of chemical industry

KMO test and Bartlett spherical test are carried out on all the variables of environmental pollution of chemical industry. The results show that the KMO test result is 0.712, indicating that the statistical effect is still good and suitable for factor analysis. The Sig of Bartlett spherical test is 0.000, showing that there is a significant correlation between each variable. The characteristic root and contribution rate of the principal components of environmental pollution in each region are obtained as shown in Table 3.

Table 3: Characteristic root and contribution rate of the principal components of environmental pollution of chemical industry

Components	Initial eigenvalue			Extraction of quadratic sum of load	
	Total	Variance percentage	Accumulation %	Total	Variance percentage
1	3.796	63.267	63.267	3.796	63.267
2	1.392	23.194	86.462	1.392	23.194
3	.434	7.241	93.703		
4	.174	2.899	96.602		
5	.141	2.357	98.959		
6	.062	1.041	100.000		

It can be seen from Table 3 that the cumulating contribution rate of the first two components has reached 86.462%. Therefore, when these two principal components replace the original indicators, most of the information of the original data can be well reflected. The two principal components are recorded as W1 and W2.

The coefficients of each principal component are shown in Table 4.

Table 4: Characteristic root and contribution rate of the principal components of environmental pollution of chemical industry

	Components	
	1	2
Zscore (industrial wastewater discharge)	-.091	.449
Zscore (domestic sewage discharge)	-.188	.506
Zscore (industrial waste gas discharge)	.239	.099
Zscore (industrial sulfur dioxide discharge)	.218	.108
Zscore (industrial dust discharge)	.338	-.112
Zscore (industrial solid waste discharge)	.420	-.226

W1=-0.091*Zscore (industrial wastewater discharge) -0.188*Zscore (domestic sewage discharge) +0.239*Zscore (industrial waste gas discharge) +0.218*Zscore (industrial sulfur dioxide discharge) +0.338*Zscore (industrial dust discharge)+0.42*Zscore (industrial solid waste discharge
W2=0.449*Zscore (industrial wastewater discharge) +0.506*Zscore (domestic sewage discharge) +0.099*Zscore (industrial waste gas discharge) +0.108*Zscore (industrial sulfur dioxide discharge) -0.112*Zscore (industrial dust discharge) -0.226*Zscore (industrial solid waste discharge)

3.3 Cluster analysis and evaluation of comprehensive scores

From formula (1) to formula (10), we can calculate the comprehensive scores and rankings of economic development and environmental pollution of chemical industry in each region, and use K-means cluster analysis to classify the economic development and environmental pollution into four categories as follows:
Category I: Beijing, Tianjin, Shanghai
Category II: Jilin, Heilongjiang, Anhui, Jiangxi, Hubei, Hainan, Chongqing, Sichuan, Guizhou, Yunnan, Shaanxi, Gansu, Qinghai, Ningxia, Sinkiang
Category III: Hebei, Shanxi, Inner Mongolia, Hunan
Category IV: Jiangsu, Zhejiang, Shandong, Guangdong

4. Conclusions and Discussion

This paper grasps the overall situation of economic development and environmental pollution of chemical industry in 30 regions in 2015 and divides them into four categories by using PCA and cluster analysis. Combined with geographical location, following conclusions have been drawn.
(1) Eight indicators are selected and combined into two principal components. Then PCA is conducted on the level of economic development in various regions, finding that economic development of cities ranking the forefront is better while that of cities at the bottom is more backward.
(2) Six indicators are selected and combined into two principal components. Then PCA is conducted on the environmental pollution of chemical industry in various regions, finding that environmental pollution of cities ranking the forefront is more serious while that of cities at the bottom is relatively light.
(3) Based on the comprehensive cluster analysis of the economic development and environmental pollution of chemical industry in various provinces and municipalities, the results can be grouped into four categories.
Category I belongs to the regions with high level of economic development and good environment represented by Beijing, Tianjin and Shanghai. For such regions, they should continue to readjust their economic structure, build an environment-friendly society and maintain a win-win economic and environmental development.
Category II belongs to the regions with low level of economic development but good environment represented by Hainan, Qinghai, Ningxia, and Sinkiang. Such regions should develop characteristic economy based on local conditions while raising awareness of environmental protection and governance to prevent embarking on the road of “pollution first, treatment later”.
Category III belongs to the regions with middle level of economic development and serious environmental pollution represented by Hebei and Shanxi. Such regions should carry out reform of the economic system under the premise of promoting environmental protection and governance, and resolutely eradicate the phenomenon of “sacrificing the environment in exchange for economic growth” while enhancing environmental governance to realize the social transformation and development.
Category IV belongs to regions with high level of economic development but serious environmental pollution represented by Jiangsu, Zhejiang, and Guangdong. It is urgent for these regions to strengthen environmental governance and reduce pollutant discharge.

Acknowledgments

I would like to express my sincere gratitude to all those who helped her during the writing of this paper. This paper was supported by the “Research Innovation Program for College Graduates of Jiangsu Province” (KYLX_0410) and the “Fundamental Research Funds for the Central Universities” (2014B29514).

Reference

Aziz A., Steve P., Marie-Noëlle P., 2017, Assessment of organic pollution of an industrial river by synchronous fluorescence and UV-vis spectroscopy: the Fensch River (NE France), Environmental Monitoring & Assessment, 189(5), 1-12, DOI: 10.1007/S10661-017-5933-3.

Gong L., Zhou J., Ye Z., Qin T., 2012, The use of principal component analysis and three-dimensional diffusion model of urban heavy metal pollution, Proceedings - 2012 3rd Global Congress on Intelligent Systems, 50-53, DOI: 10.1109/GCIS.2012.65.

Grossman G., Krueger A., 1991, Environmental impacts of a North American Free Trade Agreement, NBER Working Paper, 3914, 13-56, DOI: 10.3386/w3914.

Jones L., Manueli R., 2001, Endogenous policy choice: the case of polution and growth, Review of Economic Dynamics, 4 (2), 369-405, DOI: 10.1006/redy.2000.0118.

Jain S., Sharma SK., Choudhary N., Masiwal R., Saxena M., Sharma A., Mandal TK., Gupta A., Gupta NC., Sharma C., 2017, Chemical characteristics and source apportionment of PM2.5 using PCA/APCS, UNMIX, and PMF at an urban site of Delhi, India, Environmental Science And Pollution Research International, 24(17), 1047-3203, DOI: 10.1007/S11356-017-8925-5.

Lopez R., 1994, The environment as a factor of production: the effects of economic growth and trade liberalization, Journal of Environmental Economics and Management, 27(2), 163-184, DOI: 10.1006/jeem.1994.1032

Lopez R., Mitra S., 2000, Coruption, pollution, and the Kuznets Environment Curve, Journal of Environmental Economics and Management, 40 (2), 137-150, DOI: 10.1006/jeem.1999.1107.

Mooibroek D., schaap M., 2011, Source apportionment and spatial variability of $PM_{2.5}$ using measurements at five sites in the Netherlands, Atmospheric Environment, 25(45), 4180-4191, DOI: 10.1016/j.atmosenv.2011.05.017.

Ovadnevaitė J., Kvietkus K., Šakalys J., 2007, Evaluation of the impact of long range transport and aerosol concentration temporal variations at the Eastern Coast of the Baltic Sea, Environment Monitoring Assess, 132(3), 365-375, DOI: 10.1007/S10661-006-9540-y.

Selden T., Song D., 1994, Environmental quality and development: is there a kuznets curve for air polution emisions? Journal of Environmental Economics and management, 27(2), 369-405, DOI: 10.1006/jeem.1994.1031.

Seklaoui M., Boutaleb A., Benali H., Alligui F., Prochaska W., 2016, Environmental assessment of mining industry solid pollution in the mercurial district of Azzaba, northeast Algeria, Environmental Monitoring & Assessment, 188(11), 1-25, DOI: 10.1007/s10661-016-5619-2.

Stokey N., 1998, Are there limits to growth? International Economic Review, 39(1), 1-31, DOI: 10.2307/2527228.

World Bank, 1992, World development report 1992: development and the environment, 1145-1161, Oxford University Pres, Oxford, United Kingdom.

Zhang L., Li H., 2016, Evaluation system of urban environmental pollution based on the principal component analysis, Proceedings-2016 International Conference on Smart City and Systems Engineering, ICSCSE, 47-49, DOI: 10.1109/ICSCSE.2016.0023.