Advances in Computer Science Research, volume 86
International Symposium on Communication Engineering & Computer Science (CECS 2018)

Analysis of Correlation Between $PM_{2.5}$ and Major Pollutants by the Method of Path Analysis

Zhenghua Wang [1, a], Zhihui Tian [2]

[1.] Zhengzhou University, Zhengzhou, China,450000, China;

[2.] Zhengzhou University, Zhengzhou, China,450000, China.

[a] wangzhenghua99199@163.com

Abstract. The air quality is poor in Xuchang city in winter. It often happens that severe and moderate pollution phenomenon. And the primary pollutant is $PM_{2.5}$. The synergistic effects of SO_2, NO_2, PM_{10}, CO, $O_{3/1h}$, and $O_{3/8h}$ air pollutants on $PM_{2.5}$ in winter were analyzed in this paper from the perspective of mathematical statistics using correlation analysis, principal component regression analysis and the path analysis method. The results showed that $PM_{2.5}$ was positively correlated with SO_2, NO_2, PM_{l0} and CO, and negatively correlated with $O_{3/1h}$ and $O_{3/8h}$. PM_{10} has the greatest effect on $PM_{2.5}$, and is the main factor affecting the concentration of $PM_{2.5}$. The direct effect of CO on $PM_{2.5}$ is secondary, and it also has an important influence on the formation of $PM_{2.5}$. The direct effects of $O_{3/1h}$ and $O_{3/8h}$ on $PM_{2.5}$ were small, but the indirect effects of PM_{10} and CO on $PM_{2.5}$ were greater.

Keywords: path analysis; $PM_{2.5}$; direct influence; indirect influence; interaction.

1. Introduction

In February 29, 2012, the State Council issued the revised "ambient air quality standards", the new standard increased the fine particulate matter $PM_{2.5}$ and ozone 8 hours concentration limit monitoring indicators. In recent years, $PM_{2.5}$ has become the primary pollutant that endangers people's health and air quality, and $PM_{2.5}$ has become the hotspot and frontier of atmospheric aerosol research at present [1,2]. $PM_{2.5}$, also known as fine particles, refers to aerosol particles smaller than or equal to 2.5μm in diameter in the atmosphere, which can be suspended in the air for a long time. Because of its small particle size, large area and strong activity, it is easy to carry toxic and harmful substances, which has a great impact on the quality of the atmospheric environment and human health. $PM_{2.5}$ is an important reason for haze formation [3]. Studies have shown that the harm of atmospheric fine particulate matter to human health is higher in children and women [4,5]. Less than 2.5 μm can be directly into the blood [6]. At the same time, it is easy to cause respiratory, cardiovascular and nervous system diseases. The chemical composition of the two particles in the fine particles is more complex and more toxic among them, [7]. The two major particulate matters include sulfate, nitrate, ammonium salt and two organic aerosols. They are produced by photochemical reactions such as SO_2, NO_2 and other gaseous substances and photochemical oxidants in the atmosphere, such as ozone. [8]. The concentrations of SO_2, NO_2 and CO in the atmosphere are bound to have a certain connection with the change of $PM_{2.5}$ concentration. The correlation analysis between $PM_{2.5}$ and SO_2, NO_2, CO and PM_{10} was carried out in References 9. The principal component regression model of $PM_{2.5}$ and SO_2, NO_2, CO and PM_{10} was established. The path analysis showed that CO had the greatest direct impact on $PM_{2.5}$. This paper studies the correlation between $PM_{2.5}$ and SO_2, NO_2, CO, PM_{10}, $O_{3/1h}$ and $O_{3/8h}$, and establishes the principal component regression model of $PM_{2.5}$ and SO_2, NO_2, CO, PM_{10}, $O3_{/1h}$ and $O3_{/8h}$. Path analysis showed that PM_{10} had the greatest direct impact on $PM_{2.5}$. Tianjin were studied by correlation analysis, PLS1 and path analysis in the literature 10. The results show that the effects of several major pollutants on $PM_{2.5}$ are different in different seasons. The principal component regression model of $PM_{2.5}$, SO_2, NO_2, CO, PM_{10}, $O_{3/1h}$ and PM_{10} is established by principal component analysis. this paper analyzes the indirect effect, direct effect and total effect of air pollutants on $PM_{2.5}$ and the internal factors of winter air quality change in Xuchang. It provides the theoretical support for Atmospheric Research and management in Xuchang.

2. Research Areas and Data Sources

2.1 Research Area Overview

Xuchang City is located in the middle of Henan Province which is 80 kilometers away from Zhengzhou city. It is between 113 degrees 3 minutes east longitude and 114 degrees 190 minutes east longitude, 33 degrees 16 minutes and 34 degrees 24 minutes north latitude. Xuchang includes Changge City, Yuzhou City, Xuchang City, Xiangcheng City and Yanling City.

2.2 Data Source

The air quality data is derived from the main official medias. There is a total of ninety days of data.

3. Research Method

3.1 Correlation Analysis

In most cases, one thing is usually affected by many factors, the correlation analysis is a common method to study whether there is a dependency relation between different phenomenon's and discuss the degree of correlation. It is a method to study the relationship between the random variables.

3.2 Principal Component Regression Analysis

Principal component regression analysis is based on the principal component as independent variable, which is a method to analyze the multicollinearity problem. To eliminate the use of principal component analysis and multiple regression model of the linear, principal component variables as independent variables in regression analysis, and then according to the new model score coefficient matrix of the original variables back to get the total elimination by principal component analysis, principal component regression model [11].

3.3 Path Analysis

Path analysis is a statistical methods which studies the relationship between the study variables. It divides the relationship between independent variables and dependent variables into direct and indirect effect. It not only reflects the direct effect of an independent variable on dependent variable, but also reflects the indirect effect of an independent variable on dependent variable through other independent variables, thus obtaining the total effect of dependent variable on dependent variable.In essence, it is a standardized multiple linear regression analysis method. We can get the internal influence of X on Y.

4. Results and Analysis

4.1 Correlation Analysis

Correlation analysis and of $PM_{2.5}(y)$, $SO_2(x_1)$, $NO_2(x_2)$, $PM_{10}(x_3)$, $CO(x_4)$, $O_{3/1h}(x_5)$ and $O_{3/8h}(x_6)$ were conducted by using SPSS software based on the daily air quality data from January 2015 to February 2015.The results show in Table I and Table II.

Table I shows that $PM_{2.5}$ is significantly positively correlated with NO_2, PM_{10} and CO, positively correlated with SO_2, and negatively correlated with $O_{3/1h}$ and $O_{3/8h}$. SO_2 is significantly positively correlated with NO_2 and PM_{10}, and PM_{10} is positively correlated with CO, and $O_{3/1h}$ was positively correlated with $O_{3/8h}$.

Table 1. Correlation Analysis Between Y and X1, X2, X3, X3, X4, X5 AND

	y	x1	x2	x3	x4	x5	x6
y	1	0.356	0.734	0.891	0.854	0.348	0.476
x1		1	0.718	0.570	0.490	0.157	0.041
x2			1	0.851	0.744	0.037	0.185
x3				1	0.769	0.123	0.280
x4					1	0.416	0.552
x5						1	0.937
x6							1

Table 2. Collinearity statistics

constant	tolerance	VIF
SO_2	0.443	2.258
NO_2	0.178	5.618
PM_{10}	0.227	4.411
CO	0.238	4.193
$O_{3/1h}$	0.100	9.986
$O_{3/8h}$	0.089	11.252

Seen from table II, the expansion factor VIF of the variance of $O_{3/8h}$ is more than 10. It indicates the existence of multicollinearity between variables. Aiming at the multicollinearity between variables, we use the analysis method to eliminate the collinearity between constants, establishing the principal components analysis of the model between $PM_{2.5}$ and SO_2, NO_2, PM_{10}, CO, $O_{3/1h}$, $O_{3/8h}$.

4.2 Establishment and Test of Principal Component Regression Model Of $PM_{2.5}$

Taking $PM_{2.5}$ (y) as dependent variable, SO_2 (x1), NO_2 (x2), PM_{10} (x3), CO (x4), $O_{3/1h}$ (x5) and $O_{3/8h}$ (x6) as independent variables, establish principal component regression model. The results show in Table III and table IV.

Table 3.Contribution rate of variance

	Initial characteristic root		
principal component	All characteristic roots	Variance contribution rate%	Cumulative contribution rate %
1	3.304	55.062	55.062
2	1.949	32.483	87.545
3	0.399	6.644	94.189
4	0.179	2.980	97.169
5	0.121	2.020	99.189
6	0.049	0.811	100.000

It can be seen from table III and table IV. that the principal component of the eigenvalue 1 is 2, and the variance contribution rate is 87.5%, which contains most of the information. Therefore, the two principal components Y1 and Y2 can be extracted, and the principal component regression model is established.

Table 4. Contribution rate of variance

principal component	Extract sum of squares load		
	Total	Variance contribution rate %	Cumulative contribution rate%
1	3.304	55.062	55.062
2	1.949	32.483	87.545
3			
4			
5			
6			

Table 5. principal component score coefficient

Variable	principal component 1	principal component 2
x_1	0.4836	0.2643
x_2	0.4858	0.1698
x_3	0.5083	-0.0652
x_4	0.3581	0.3983
x_5	-0.225	0.6347
x_6	-0.3048	0.5802

According to the principal component score coefficient of table V, the equation is obtained:

$$y1=0.4836x1+0.4858x2+0.5083x3+0.3581x4-0.225x5-0.3048x6 \quad (1)$$

$$y2=0.2643x1+0.1698x-20.0652x3+0.3983x4+0.6337x5+0.5802x6 \quad (2)$$

Principal component regression analysis was used to get the coefficient of principal component regression model. See table VI.

Table 6. coefficient of principal component regression model

constant coefficient	constant coefficient
y1	0.858
y2	-0.074

According to the coefficient of principal component regression model, the linear regression equation between $PM_{2.5}$ and principal components Y1 and Y2 was established:

$$Y=0.858y1-0.074y2+4.879 \ldots \quad \ldots(3)$$

The main components of the expression (1) (2) into the principal component regression model (3), the standard variable y1, y2 replacement for x1, x2, the original variables x3, x4, x5, x6, available for variable principal component regression model of original variables:

$$Y=0.219x1+0.291x2+0.484x3+0.0126x4-0.663x5-0.691x6+4.879 \ldots \quad \ldots.(4)$$

Table 7. Significance Test Results

t	sig
0.407	0.000
22.562	0.000
-0.711	0.000

Significant tests were carried out on $PM_{2.5}$, and F=255.082. Significant linear relationships were found between Y and principal component Y1 and Y2. The P values of T were less than 0.05, indicating that there was a significant linear relationship between Y1 and Y2 at the significant level of 0.05. Principal component regression model goodness of fit test R value are 0.974 which are close to 1. Goodness of fit is very good. The principal component regression equation of variance expansion factor VIF=1.000 and $PM_{2.5}$ has no multicollinearity, and multicollinearity among independent variables has been eliminated. According to the test results, the $PM_{2.5}$ principal component regression equation was tested.

4.3 Path Analysis

The path analysis of $PM_{2.5}$ principal component regression model was carried out by SPSS software, R_1^2=5.48%, R_2^2=0.001%, R_3^2=51.8%, R_4^2=16.9%, R_5^2=1.59%, R_6^2=0.6%, R_{12}=0.24%, R_{13}=-19%, R_{14}=-9.4%, R_{15}=0.9%, R_{16}=-0.2%, R_{23}=-0.9%, R_{24}=-0.4%, R_{25}=-0.01%, R_{26}=0.02%,

R_{34}=45.5%, R_{35}=2.2%, R_{36}=-3.2%, R_{45}=4.3%, R_{46}=-3.6%, R_{56}=-1.9%. Coefficient of total determination is 90.9%. The decision factor of residual factor Y is 9.1%, which directly acts on =30%. The path results of the $PM_{2.5}$ principal component regression model show that:

(1) The direct effect of SO_2 on $PM_{2.5}$ was -0.234, and the total effect on $PM_{2.5}$ was 0.332.

(2) The direct effect of NO_2 on $PM_{2.5}$ was -0.007, the indirect effect on $PM_{2.5}$ was 0.741, and the total effect on $PM_{2.5}$ was 0.773.

(3) The direct effect of PM_{10} on $PM_{2.5}$ was 0.72, and the direct effect of CO on $PM_{2.5}$ was 0.411.

(4) The direct effect of $O_{3/1h}$ on $PM_{2.5}$ was -0.126, the direct effect of $O_{3/8h}$ on $PM_{2.5}$ was 0.079, and the decision coefficients were -8.1%.

(5) The direct effect of residual factor epsilon on $PM_{2.5}$ is 0.3, and the coefficient of decision making for $PM_{2.5}$ is 9.1%.

5. Conclusion

$PM_{2.5}$ was positively correlated with SO_2, NO_2, PM_{10} and CO, and negatively correlated with $O_{3/1h}$ and $O_{3/8h}$, among which PM_{10} and $PM_{2.5}$ had the strongest correlation, the correlation coefficient was 0.891, and the CO correlation intensity was the second, and the correlation coefficient was 0. 854.PM_{10} is the main factors affecting the change of $PM_{2.5}$ concentration. The direct effect of CO is next. The direct effect of NO_2 on $PM_{2.5}$ is minimal.The direct effect of NO_2 on $PM_{2.5}$ was the smallest, but the indirect effect was the greatest. When the concentration of SO_2 and NO_2 increased, the concentration of $PM_{2.5}$ increased with the indirect action of PM_{10} and CO. the total effect of CO on $PM_{2.5}$ was the largest, followed by the total effect of PM_{10}, and the total effect of $O_{3/8h}$ on $PM_{2.5}$ was the smallest. the decision factor of residual factors to $PM_{2.5}$ is 9.1%, and other factors play a role in $PM_{2.5}$.

Xuchang mainly relies on coal for heating in winter. The incomplete combustion of coal will lead to the increase of CO emissions in the air which leads to the increase of $PM_{2.5}$ concentration and the poor air quality. At the same time, industrial production, automobile exhaust, dry climate and other reasons will also have an important impact on the rise of $PM_{2.5}$n, it is useful to reduce the $PM_{2.5}$ concentration and improve the air quality that we reduce the amount of CO emissions in the process of coal combustion and improve the utilization rate of coal combustion in winter in Xuchang.

References

[1]. Wang Wei, Tang Dagang, Liu Hongjie, et al. Research on current pollution status and pollution characteristics of $PM_{2.5}$ in China[J]. Research of Environmental Science,2000,13 (1).

[2]. Huang D, Xu J, Zhang S. Valuing the health risks of particulate air pollution in the Pearl River Delta, China[J]. Environmental Science & Policy, 2012, 15 (1): 38-47.

[3]. YANG Xinxing, FENG Lihua, WEI Peng.Air particulate matter $PM_{2.5}$ in Beijing and its harm. Frontier Science, 2012, 6 (1): 22-31.

[4]. BAGHAPOUR M A. Air pollution impacts on women, s health[J]. Women Health Bulletin, 2015.2 (2): e26864.

[5]. BARBOSA S M, FARHAT S C, MARTINS L C, et al.Air pollution and children, s health:sickle cell disease[J].Cad Saude Publica, 2015.31 (2): 265-275.

[6]. Liu Yanli.$PM_{2.5}$ the ins and outs of [j]. Chinese Green Pictorial.2012 (5): 20-21.

[7]. Yang Xinxing, Wei Peng. $PM_{2.5}$ technology of atmospheric particulate matter and its source apportionment [j]. 2013, 7:12-19.

[8]. FENG J L, GUO Z Get al. Source and formation of secondary particulate matter in $PM_{2.5}$ in Asian continental outflow [J]. Journal Geophysical Research Atmospheres, 2012117 (D3).

[9]. Zheng Yu, Deng Lan. Correlation analysis of $PM_{2.5}$ and air pollutants in Harbin based on PLS1[11]. Journal of ecology and environment.2014.12.

[10]. Wang Yunge, et al. Correlation analysis of $PM_{2.5}$ and air pollutants in Tianjin based on PLS1. Journal of environmental engineering 2017.7 seedlings (1): 39-45.

[11]. Chinese public health, 2001, volume eighth, seventeenth. Principal component regression analysis using SPSS. Jinan University medical school. Liu Runxing.