ISSN 2616-5902 Vol. 4, Issue 6: 59-65, DOI: 10.25236/AJBM.2022.040610

Analysis of Influencing Factors of Haze in Zhengzhou Based on Linear Regression

Yifan Gao

Shanghai University, Shanghai, China

Abstract: *In recent years, various regions in China have been plagued by smog problems, and Zhengzhou, as the capital city of Henan, has a particularly prominent smog problem than other cities due to climate and economic development. In the past two years, smog pollution has improved slightly, but smog still plagues our lives. This paper analyses the changes in haze pollution in Zhengzhou in the past five years, draws the main reasons for the changes in haze pollution. Principal component analysis and regression analysis were used to establish a regression model of the influencing factors of haze in Zhengzhou. The results show that pollutant emissions and coal consumption have the most obvious impact on the degree of haze pollution in different years, and reasonable suggestions are put forward based on the current situation of haze pollution in Zhengzhou and the existing governance policies.*

Keywords: *Haze; Zhengzhou City; Principal Component Analysis; Linear Regression*

1. Introduction

With the industrialization of our country, the degree of urbanization continues to increase, the rapid development of the economy, people's living standards have been qualitatively improved. At the same time, environmental problems are becoming more and more prominent, of which the smog problem has become a particularly important aspect of environmental problems, and looking up at the blue sky and starry sky has gradually become a luxury. Since the word smog became the keyword of the year in 2013 and came into our field of vision, smog has been widely concerned and discussed.

Since 2013, all parts of China are facing serious smog pollution problems, and solving the smog problem is an inevitable requirement for fighting the battle against pollution and building a well-off society in an all-round way, so the state and local governments at all levels have attached great importance to the problem of smog control. In recent years, the state has carried out real-time monitoring of 337 cities at the prefecture level and above in the country, and publishes the pollution rankings of each city in the country every year, as well as the proportion of cities that meet the standards. According to the air quality status report released by the National Bureau of Statistics from Table 1, among the 337 cities at the prefecture level and above in 2019, 157 cities with air quality standards (better than the national secondary standards) accounted for 46.6%. Compared with the previous two years, the number of cities that have reached the standard has increased significantly, indicating that the situation of smog pollution in China has improved in recent years, but more than half of the cities in the country are still unqualified in air quality, the smog problem is still very serious nationwide, and the smog control still has a long way to go.

Table 1: The number of cities with air quality standards in the country and their proportion.

Year	Number of cities that meet the criteria	Proportion
2017	99	29.38%
2018	121	35.91%
2019	157	46.59%

Zhengzhou has always been among the cities with substandard air quality in the country, and in the air quality ranking of 168 key cities in the country, Zhengzhou city was in the bottom 20 in 2018 and 2019, and was in a state of serious pollution. In 2017, Zhengzhou was rated as a new national central city and has great economic development potential, but the smog problem has dragged the development of Zhengzhou. Smog will pose a threat to the health of citizens, bring inconvenience to citizens' lives, reduce people's sense of well-being, but also have an adverse impact on the sustainable development of the economy, the prosperity of tourism and the absorption of talents, damage the image of Zhengzhou and is not conducive to the future development of Zhengzhou, so we should analyze the current situation of haze pollution in Zhengzhou, increase the intensity of smog control in a targeted manner, and solve the smog problem as soon as possible.

2. Manuscript Preparation

In this paper, nine indicators of climate factors, pollutant factors and economic factors are selected in combination with the references, and the data provided by the official website of the weather post-report, the Henan Provincial Bureau of Statistics and the Zhengzhou Municipal Bureau of Statistics are used for statistical analysis of the data changes from 2014 to 2018, and the indicators that have a significant impact on the haze changes are selected as the independent variables in the regression analysis.

Due to the strong information correlation between the variables, the principal component analysis method is used to extract the common factor representing the original multiple explanatory variables, and the regression equation between the extracted principal components and the Zhengzhou smog pollution index is established, and finally the regression equation between the Zhengzhou haze pollution index and the variables is obtained according to the coefficient relationship between the principal components and the variables. Establish a regression model of smog in Zhengzhou. Finally, according to the above analysis results, corresponding governance measures and suggestions are proposed.

2.1. Literature Review

Since the emergence of haze pollution abroad precedes China, foreign research on the influencing factors of haze is also earlier than that in China. As early as the 1970s, foreign countries began to analyze the smog problem. Since 2013, China has been plagued by smog weather throughout the country, and local governments at all levels have increased their vigilance against smog problems, so the domestic literature on the influencing factors of smog weather appeared after 2013. The research on smog at home and abroad is from different angles, through reading references, analyzing the research perspectives of different scholars, some scholars are introducing the basic concepts and main hazards of smog; some are analyzing the main influencing factors of haze formation; some use certain scientific methods to analyze the influencing factors of smog; some focus on proposing haze control schemes according to the specific conditions of the city.

Zhao (2015) used principal component analysis to analyze the causes of smog in five smog-prone provinces of Beijing, Tianjin, Hebei, Henan and Shandong, and used game theory to establish a game model to explore the regulatory role of government factors in smog governance. Xie et al. (2017) used the analytic hierarchy method to obtain the weight of the factors affecting smog, believing that pollutants are the main factors that produce smog, and on this basis, the fuzzy comprehensive evaluation method was used to analyze the pollutant emission levels of various industries, and the smog tax amount of each industry was sorted. Li et al. (2015) used the principal component analysis method based on meanness to analyze the smog environment in Beijing and put forward corresponding suggestions. Dong (2017) used partial least squares regression to establish a regression model of the influencing factors of haze in Zhengzhou. Meng et al. (2015) used multi-layer regression analysis to establish a regression model of the influencing factors of haze in Zhengzhou. Xiong (2020) used a multivariate gray model to predict the smog index of Nanjing. In this paper, with reference to the research methods of the above scholars, combined with the situation in Zhengzhou City, principal component analysis method and linear regression are selected as the data analysis methods in this paper

Because each person studies the problem from a different perspective, there are often different index systems for the same problem, and the establishment of the indicator system determines the scientific and persuasiveness of the research problem. In this paper, after a detailed understanding and analysis of the haze pollution situation in Zhengzhou, the indicators of the regression model of the influencing factors of haze are selected in strict accordance with the principle of index selection and the current situation of haze pollution in Zhengzhou.

2.2. Selection of haze influence factor indicators

Natural factors, pollutant emission factors, and urban development levels all have an impact on haze. The influence of natural factors on haze is seasonal, and the frequency of haze varies from season to season, which is caused by natural factors. The pollutant emission factors and urban development level are influencing the haze in terms of annual variation, and the different degree of haze pollution in different years is due to the different pollutant emission and urban development level in different years. This

ISSN 2616-5902 Vol. 4, Issue 6: 59-65, DOI: 10.25236/AJBM.2022.040610

provides a reference for the selection of indicators to study the annual variation of haze.

In summary, both pollutant emission factors as well as economic factors have a significant influence on the annual change of haze pollution status. In this paper, two primary indicators of pollutant emissions and economic development and seven secondary indicators, including dust emissions, sulfur dioxide emissions, nitrogen oxide emissions, coal consumption, motor vehicle ownership and construction site area, are selected to establish a progressive hierarchical model as shown in Table 2.

Table 2: Selected indicators.

Target layer	Guideline layer	Indicator layer
Haze influence factors	Pollutant emission factors	Dust emissions
		Sulphur dioxide emissions
		Nitrogen oxide emissions
	Level of urban development	GDP per capita
		Coal consumption per capita
		Motor vehicle ownership per capita
		Construction site area per capita

2.3. Empirical analysis of factors influencing haze in Zhengzhou City based on multiple regression

2.3.1. Identification of factors influencing haze based on principal component analysis

Because the correlation between these factors affecting haze is large, such as per capita coal consumption and sulfur dioxide and nitrogen oxide emissions, and the Zhengzhou Environmental Protection Bureau started to monitor and publish data on haze pollution in 2012, the sample size is limited and the number of variables is large, so it is not advisable to establish a regression model directly, so this paper first extracts the main information of independent variables through principal component analysis. Therefore, this paper firstly extracts the main information of independent variables through principal component analysis to achieve the purpose of condensing information. Then, the extracted principal components are used as new independent variables, and the standardized annual average pollution index of Zhengzhou City is used as the dependent variable to establish the regression model of haze pollution in Zhengzhou City.

Based on the selected indicators, the data of the selected indicators from 2014-2018 were found from Henan Provincial Bureau of Statistics, Zhengzhou City Bureau of Statistics.

Table 3: Raw data.

Year	2014	2015	2016	2017	2018
Dust emissions (tons)	6.76	9.43	3.56	2.61	2.48
Sulfur dioxide emissions (tons)	10.46	12.53	3.88	3.67	3.35
Nitrogen oxide emissions (tons)	17.83	16.77	8.54	3.18	2.98
GDP per capita (million yuan)	7.30	7,.72	8.41	9.31	10.13
Coal consumption per capita (tons)	3.80	3.17	2.90	2.53	2.23
Motor vehicle ownership per capita (units)	0.24	0.25	0.24	0.28	0.31
Construction site construction area per capita (square meters)	18.98	24.39	25.24	26.06	32.01

Analysis of results

Table 4: Common factor variance.

	Initial	Extraction
Dust emissions	1.000	.763
Sulfur dioxide emissions	1.000	.799
Nitrogen oxide emissions	1.000	.963
GDP per capita	1.000	.980
Coal consumption per capita	1.000	.915
Motor vehicle ownership per capita	1.000	.752
Construction site construction area per capita	1.000	.806

ISSN 2616-5902 Vol. 4, Issue 6: 59-65, DOI: 10.25236/AJBM.2022.040610

As can be seen from the metric variance table in Table 4, most of the information for each original variable was extracted, with more than 75% of the information extracted, with NOx emissions, GDP per capita and coal consumption extracting the most information, at more than 90%.

Table 5: Total Variance Explained.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative%	Total	% of Variance	Cumulative%
1	5.978	85.404	85.404	5.978	85.404	85.404
2	.769	10.986	96.390			
3	.189	2.707	99.096			
4	.063	.904	100.000			
5	1.761E-16	2.516E-15	100.000			
6	-2.819E-17	-4.027E-16	100.000			
7	-3.746E-16	-5.352E-15	100.000			

From the table of eigenroot and variance contribution rate in Table 5, it can be seen that 7 principal components can be extracted, and if all of them are proposed, the significance of the method is lost. Therefore, the eigenvalue is set to 1, and according to the principle of principal component extraction, only the principal components with eigenvalue greater than 1 are extracted, and the cumulative contribution rate should reach 85%. From the table, it can be seen that only the first principal component has an eigenvalue greater than 1, carrying 85.404% of the information, and it can be considered that most of the information is retained, so one principal component can be extracted, which not only solves the problem of multicollinearity between variables, but also fully reflects the information carried by the original variables and achieves the purpose of information concentration.

Table 6: Component Matrix

	Component
	1
Dust emissions	.874
Sulfur dioxide emissions	.894
Nitrogen oxide emissions	.981
GDP per capita	-.990
Coal consumption per capita	.956
Motor vehicle ownership per capita	-.867
Construction site construction area per capita	-.898

The loading matrix of the principal components is obtained in Table 6. After obtaining the loading matrix, the principal component eigenvectors are obtained by dividing the loading vector of the principal components by the arithmetic square root of the eigenvalues of the principal components. As shown in Table 7

Table 7: Eigenvector matrix

	Component
	1
Dust emissions	.358
Sulfur dioxide emissions	.366
Nitrogen oxide emissions	.402
GDP per capita	-.406
Coal consumption per capita	.392
Motor vehicle ownership per capita	-.355
Construction site construction area per capita	-.368

Expression of principal components $Y = Z_1 X_1^* + Z_2 X_2^* + \cdots + Z_7 X_7^*$, X_1^*, X_2^*, ..., X_7^* is the value of the variable after normalization. Z_1, Z_2, ... Z_7 is the variable eigenvector of the principal component. Bringing the specific values of the eigenvectors in Table 7 into the above expression gives.

$$Y = 0.358X_1 + 0.366X_2 + 0.402X_3 - 0.406X_4 + 0.392X_5 - 0.355X_6 - 0.368X_7$$

The principal component expressions show that the principal components are strongly correlated with each variable, and are strongly positively correlated with dust emissions, sulfur dioxide emissions,

ISSN 2616-5902 Vol. 4, Issue 6: 59-65, DOI: 10.25236/AJBM.2022.040610

nitrogen oxide emissions, and per capita coal consumption, and negatively correlated with per capita GDP, per capita motor vehicle ownership, and per capita construction site area.

Based on the above principal component expressions, the corresponding principal component scores can be calculated for each year in order to build a regression model.

Table 8: Principal component score.

Year	Principal Component(Y)
2014	1.14
2015	0.85
2016	-0.07
2017	-0.70
2018	-1.22

2.3.2. Identification of factors influencing haze based on principal component analysis

The standardized annual average air quality index Z for each year was used as the explanatory variable, and the principal components obtained in the previous chapter were used as explanatory variables for regression analysis.

The average air quality index for each year was derived from the air quality index of the weather hindcast network, the air quality index for each year was normalized according to the formula $Z_i^* = \frac{Z_i - \bar{Z}}{S}$ (where $\bar{Z}$ is the mean of the haze pollution index and S is the standard deviation), and the results are shown in Table 9.

Table 9: Air quality index.

Year	Average Air Quality Index	Standardized Air Quality Index
2014	133.19	1.49
2015	134.77	1.64
2016	119.01	0.18
2017	111.84	-0.49
2018	106.82	-0.96

The results of the regression analysis performed are presented in the following table:

Table 10: Model summary.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.987a	.973	.964	.21927

The adjusted complex correlation coefficient in the model summary in Table 4-8 is 0.964, indicating that the regression equation fits the sample data well and can explain the variation in the dependent variable to the extent of 96.4%.

Table 11: Analysis of variance table.

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	5.251	1	5.251	109.213	.002a
Residual	.144	3	.048		
Total	5.395	4			

The significance level of the model was determined to be 0.05, and the probability of the significance value of the statistic F was found to be $0.002 < 0.05$ based on the regression sum of squares and the residual sum of squares in the ANOVA table in Table 4-8, indicating that the linear relationship between the variables is significant. That is, the linear relationship between the extracted principal components and the standardized annual average air quality index is significant.

Table 12: Coefficient analysis table.

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	.371	.098		3.784	.032
Principal Components	1.146	.110	.987	10.450	.002

From the regression coefficient table in Table 12, it can be seen that the significance probability of the statistic t $0.002 < 0.05$, the regression coefficient passed the significance test, the regression

ISSN 2616-5902 Vol. 4, Issue 6: 59-65, DOI: 10.25236/AJBM.2022.040610

coefficient is significantly different from 0, and the regression model is obtained:

$$Z=0.371+1.146Y$$

The regression between the dependent variable and the principal component is transformed into a regression between the dependent variable and the independent variable by bringing the linear expression between the principal component and the independent variable obtained earlier into the above equation.

$$Z = 0.371 + 1.146Y$$

$$= 0.371 + 1.146(0.358X_1 + 0.366X_2 + 0.402X_3 + 0.406X_4 + 0.392X_5 - 0.355X_6 - 0.368X_7)$$

$$= 0.41X_1 + 0.42X_2 + 0.46X_3 - 0.46X_4 + 0.45X_5 - 0.41X_6 - 0.42X_7$$

The above equation is the regression equation of standardized dust emissions, sulfur dioxide emissions, nitrogen oxide emissions, per capita GDP, per capita coal consumption, per capita motor vehicle ownership, and per capita construction site area on the standardized annual average air quality index, which reflects the degree of influence of the selected indicators on the haze pollution situation.

3. Conclusions

It can be seen that each of these factors has a significant impact on the number of haze days, with NOx and coal consumption per capita having the most significant impact on haze weather and being positively correlated with the air quality index, i.e., a reduction in these indicators leads to a reduction in haze pollution. Among them, nitrogen oxides mainly come from the combustion process of fossil fuels in cars, airplanes and industrial machines, as well as from the burning of vegetative bodies. In recent years, the Zhengzhou government has introduced a series of policies to reduce NOx emissions, such as vehicle traffic restrictions, vigorous development and use of clean energy instead of fossil fuels, and a ban on straw burning. It can be seen that these have played an important role in reducing haze. Coal consumption has been the key control link of haze management in Zhengzhou, coal combustion will produce a variety of pollutants, such as sulfur dioxide, hydrogen sulfide, nitrous oxide, dust, these gases will aggravate air pollution. The share of coal combustion in energy consumption has been high in Zhengzhou, which has been working to reduce the share of coal consumption since 2013, and the proportion of coal consumption in total energy consumption in Zhengzhou has dropped to within 60% by 2019. As can be seen from the regression results, this effort has had a significant effect on the reduction of haze.

In addition, sulfur dioxide and dust emissions also play an important role in the haze pollution situation in Zhengzhou, and the government has also been strictly controlling the pollutant emissions from factories in recent years by strictly checking whether the pollutant emissions from factories are qualified, and closing factories with high pollution, high emissions, and seriously outdated equipment with substandard pollutant emissions, which will directly lead to the reduction of sulfur dioxide and dust emissions, and thus the air quality index decreases.

The negative correlation between GDP per capita and AQI shows that in recent years, Zhengzhou City has broken the "spell" of economic development bringing environmental pollution, and has not only not aggravated but also controlled the haze pollution while continuously developing its economy, which brings us great confidence to continue to reduce haze pollution while enhancing economic development in the future. This change has given us great confidence to continue to reduce haze pollution while enhancing economic development.

The correlation between the number of motor vehicles per capita and the area of construction sites per capita and the air quality index appears to be negative, which may seem contradictory to our perceptions, but it reflects the real changes in Zhengzhou in recent years, as the city government has made great efforts to solve the haze problem and make initiatives to reduce the haze without affecting the development of the city. The number of motor vehicles is increasing, but the government promptly eliminates unqualified yellow-label vehicles and sets up restriction policies, so although the number of motor vehicles has increased, the pollution caused by motor vehicles to the environment has been controlled. The construction site construction area is also increasing, but because of the strict regulations on the construction site dust emission standards, the amount of dust emitted from construction sites is strictly limited, and the haze pollution has decreased with the increase in construction site construction area.

References

[1] Cai F. (2017) Study on the impact of industrial structure on haze in Beijing-Tianjin-Hebei region [D]. Northwestern University.
[2] Huang B.B. (2019) An analysis of the causes, hazards and management measures of haze in Luohe City[J]. Agricultural disaster research, 9(04): 113-114.
[3] Li L., Sun Y.X. (2015) Analysis and research on haze environment based on mean-valued principal component analysis [J]. Computer Application Research, 32(05): 1373-1375.
[4] Li Y.Y., Wang L.H. (2018) Grey correlation variance analysis of PM2.5 drivers in Beijing-Tianjin-Hebei region [J]. China Development, 18(04): 13-25.
[5] Lu N., Zhou J.B., Li Z.G., Wang Y.T., Jin W. (2015) Causes of haze in China and countermeasures for its management [J]. Hebei Industrial Science and Technology, 32(04): 371-376.
[6] Meng Z.G., Yue X.N., Wang D., Yuan Z.H. (2015) Analysis of urban haze causation model based on multilayer regression [J]. Journal of Shenyang University (Natural Science Edition), 27(02): 139-142.
[7] Xie Q.J., Lu Y.L., Zhu J.M., Zhou J.B., (2017).Research on the quantitative levy of haze tax based on fuzzy comprehensive evaluation [J]. Journal of Huaiyin Normal College (Natural Science Edition), 16(02): 113-118.
[8] Yao, X.J., Li, Y.W., Liu, M.Z. (2014) Characteristics and formation mechanism of a winter haze–fog episode in Tianjin, China [J]. Atmospheric Environment, 98.
[9] Yan, S.Q., Zhu B., Kang H.Q. (2019) Long-Term Fog Variation and Its Impact Factors Over Polluted Regions of East China [J]. Journal of Geophysical Research: Atmospheres, 124(3).