Analysis on the Influencing Factors of PM2.5Concentrations in Chinese Cities

Hao Zhou [1], Bingxiang Zhu [2 *]

[1]National Research Center of Cultural Industries, Central China Normal University, Wuhan, Hubei Province, China
[2]School of Electrical Engineering, Wuhan University, Wuhan, Hubei Province, China

[*]Corresponding author e-mail: 2321075389@qq.com

Abstract. Recently, haze weather has become a common phenomenon in large cities. PM2.5 usually refers to fine particulate matter with a diameter less than 2.5 micrometres, which is considered to be the "culprit" of haze weather. Modern society develops at a high speed, a series of economic and social activities are bound to discharge a large amount of PM2.5. Once it goes beyond air endurance, with unstable weather factors, it is easy to appear a wide range of haze. Based on data from 31 provincial capitals and municipalities directly under the central government, main reasons for the concentrations of PM2.5 in major cities were studied. The related factors were choosing to establish a multivariate regression model, MATLAB software was used as an instrument. The correlation coefficient method and stepwise regression method were used to eliminate redundant variables. By carrying out economic significance test and statistical significance test, the revised accurate model results were achieved. Finally, according to the results of the model, feasible ways to reduce PM2.5 concentrations are obtained, which provide policy suggestions for reducing haze and improving air conditions in China.

Keywords: PM2.5 Concentrations; Influencing Factors; Regression Model; Policy Suggestions; MATLAB.

1. Introduction

With the increase of national population and the process of urbanization and industrialization, air pollution has become increasingly serious [1]. Haze weather has caused the rising morbidity of various diseases and threatened people's health greatly [2]. The death rate, caused by lung cancer, was among the topest 10 in China. Masks have been widely used to avoid inhaling harmful particles. This has caused other kinds of diseases because of poor breathability of such masks [3]. It is imperative to understand the dominant reasons causing PM2.5 rather than using auxiliary method.

Current researches can be summarised as two main streams. Some researchers studied on the formation of PM2.5 from the perspective of Physics and Chemistry. For example, Mai J et al. used a WRF-CMAQ model to simulate and test weather factors on haze [4]. Zhong et al. collected OC/EC values in typical days to study haze weather in Wuhan, a large city in middle China [5]. Other researches focused on the cause of PM2.5 from the perspective of economics and statistics, using statistical methods

IOP Conf. Series: Earth and Environmental Science **170** (2018) 032064 doi:10.1088/1755-1315/170/3/032064

to analyse. Eeftens et al. chose land use, population and transportation as three main factors to study a regression model on PM2.5 [6]. Wu et al. divided transportations into two main modes and found that the on-road mode had a higher PM2.5 than the in-cabin mode [7].

This research focuses on 31 provincial capital cities and municipalities directly under the central government. Influencing factors of PM2.5 concentrations, including weather factors and human activities, were chose to study. The aim is to improve the current situation of haze weather so as to reduce air pollution.

2. Data collection and modelling.

This article studied 31 main cities from five aspects, GDP of the second industry, dust emission, the number of public cars, annual rainfall and built-up area. Rainfall was selected as a weather factor, others were selected to stand for typical human activities. These factors were used as independent variables, PM2.5 concentration is dependent variable. The cross-section data is listed in Table 1.

2.1. Data collection

Table 1. Variable Data

City	Dust Emission (ton)	Annual Rainfall (mm)	GDP of the Second Industry, (billion yuan)	Public Cars(Unit)	built-up area (square km)	PM2.5 Concentrations (micrograms per cubic meter)
Beijing	7874	669.1	494.44	22688	1401.01	83.2
Tianjin	57314	608.6	757.14	12699	885.43	85.8
Shijiazhuang	52705	712.6	269.39	4882	278.05	122.6
Taiyuan	41174	528.4	106.75	2671	340	67.7
Huhehaote	79103	531.3	88.44	2128	260	44
Shenyang	30130	968	213.56	5444	465	70.9
Changchun	24451	890.8	291.57	4342	506.33	64.6
Haerbin	21781	537.8	189.67	7408	402.9	72.5
Shanghai	72782	1596.1	840.63	16693	1380	52.2
Nanjing	48591	1807.7	411.73	9208	755.27	73.7
Hangzhou	20414	1797.3	412.09	8770	506.09	60.9
Hefei	11483	1502	318.12	4916	438.2	80
Fuzhou	67548	2263.4	259.04	4386	260.05	31.4
Nanchang	33926	1869	230.72	3423	307.3	49.8
Jinan	54678	1008.2	236.89	5476	392.96	91
Zhengzhou	21097	833	372.87	6230	437.6	87.6
Wuhan	54089	1827.1	522.71	8970	566.13	79.5
Changsha	6983	1704.8	452.1	7187	312.3	75
Guangzhou	8951	2939.7	575.16	14074	1237.25	47.4
Nanning	9694	1546.4	142.72	3327	287.4	47.6
Haikou	156	1913.7	23.36	1597	152.4	22.4
Chongqing	83787	1348	789.89	8753	1329.45	62.8
Chengdu	12534	983.9	523.2	10781	615.71	72.8
Guiyang	8475	1045.8	121.88	3265	299	45.5
Kunming	13853	1150.2	166.01	6262	420.5	32.2
Lasa	131	551.6	16.28	522	90.72	23.6
Xian	2853	456	219.78	7829	500.59	75.7
Lanzhou	15892	310	79.01	2800	305.28	58.8
Xining	28348	444.1	59.56	2449	90	62.1
Yinchuan	11220	264.9	82.56	1818	166.82	47.4
Wulumuqi	38411	387.1	70.41	4669	429.96	62.9

Source: China Statistical Yearbook (2017) and China Construction Yearbook (2017).

2.2. Scatter diagram

For the convenience of further study, take logarithm of the explanatory variables and the explained variable. The scatter diagram is shown in Figure 1:

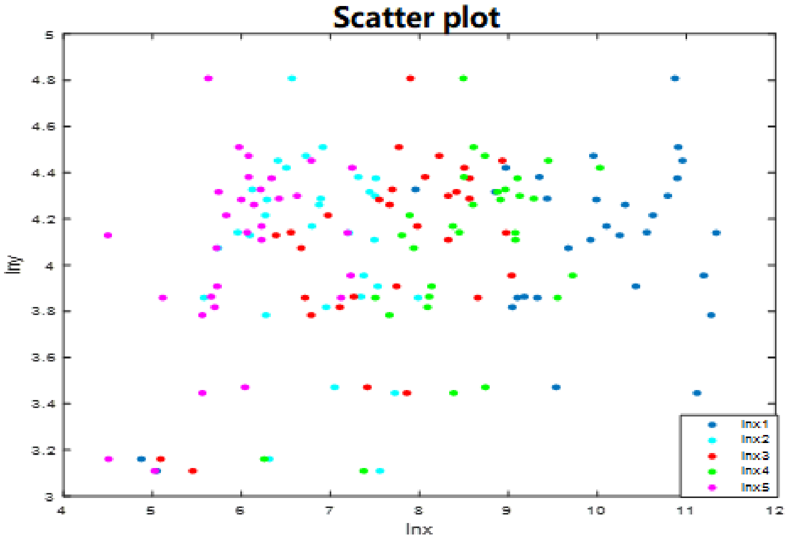

Figure 1. Scatter diagram

As can be seen from above, the explanatory variables and the explained variable have a linear trend. Therefore, the following model uses multiple linear regression model, and analyse whether there are insignificant variables. If there are insignificant variables, methods should be used to eliminate them to get more accurate model.

2.3. Establish multiple regression model

Adopted model is as follows:

$$\text{lny} = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \beta_3 \ln x_3 + \beta_4 \ln x_4 + \beta_5 x_5 + \varepsilon_i \tag{1}$$

Where: β_0 represents intercept, β_{i} represents coefficient, ε_i represents he random disturbances. Through the regression analysis of the model, the influence degree and direction of each variable are obtained.

Residual lever diagram can be obtained meantime. For residual lever diagram, it is better to distribute the data evenly around zero and have no regularity, which means that regression analysis is ideal to some extent. In general, stast returns four values: R^2, F, f, p associated with significant test (if the p value does not exist, only the first three items are output).

(1) Generally speaking, the larger the R^2 is, the better the result is.

(2) The larger the F test value is, the better it is. Especially, F should be greater than f.

(3)p value should be smaller than alpha value. If p is bigger than the latter, there is insignificant variables in the equation.

After calculation, β_{i} β_0 are :3.5394, 0.0246, -0.3268, 0.3435, 0.1741, -0.2626. Stast returns four values:0.6557, 9.5220, 3.5120e-05,0. 0624.That is to say, the confidence level is about 95%, R^2 is 0.6557, F(=3.5120e-05) is larger than f. However, p (=0.0624) is larger than alpha (=0.05), which indicates there are extra independent variables. The correlation coefficient method and stepwise regression method were used later to eliminate redundant variables.

2.4. The correlation coefficient method

Table 2. correlation coefficient matrix

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1	0.0257	0.5896	0.4709	0.4320	0.5543
X_2	0.0257	1	0.4522	0.3479	0.3341	-0.1738
X_3	0.5896	0.4522	1	0.9000	0.8447	0.5985
X_4	0.4709	0.4522	0.9000	1	0.8958	0.5669
X_5	0.4320	0.4522	0.9000	0.8958	1	0.4376
X_6	0.5543	0.4522	0.9000	0.8958	0.4376	1

2.5. As can been seen, there is a high correlation between explanatory variables according to the correlation coefficient method. The variance inflation factor (VIF) test was then used to determine the extent of collinearity between variables. By calculation, VIF=2.9044<<10. It indicates that there is no high degree of collinearity between variables, but there is a high correlation between variables.

From the correlation coefficient matrix, it can be found that there is a high correlation between X3, X4 and X5, which is considered to exclude X4 and X5 variables. The influences of public car ownership and built-up area can be excluded. The corresponding regression model is as follows:

$$\text{lny} = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \beta_3 \ln x_3 + \varepsilon_i \tag{2}$$

By calculation, β_i β_0 are: 3.6635, 0.0293, -0.3219, 0.3050

Stast returns four values: 0.6856, 14.3290, 8.8789e-06, 0.0347

$$\text{lny} = 3.6635 + 0.0293 \ln x_1 - 0.3219 \ln x_2 + 0.3050 \ln x_3 \tag{3}$$

2.5.1. Economic significance tests the estimated results of the model show that, assuming other variables remain unchanged, the concentration of PM2.5 increases by 0.305% when GDP of the second industry increases by 1%. PM2.5 concentration increases by 0.0293% when industrial dust emission (tons) increases by 1%. Rainfall increased by 1% will cause a 0.3219% reduction in PM2.5 concentration. This is also in line with the expected results.

2.5.2. Statistical test .1) Goodness of fit: $R^2 = 0.6856$,which indicates that the model fits well with the sample. 2)F test: given significance level $\alpha = 0.05$,when k=5 and n-k-1=51-5-1=45, $F_\alpha(5,45) = 2.45$ F=14.3290>>2.45, which indicates the equation is significant.3) p test: given significance level $\alpha = 0.05$,p=0.0347<0.05, which indicates the equation is significant.

2.6. stepwise regression method

The result can be seen from the following figure:

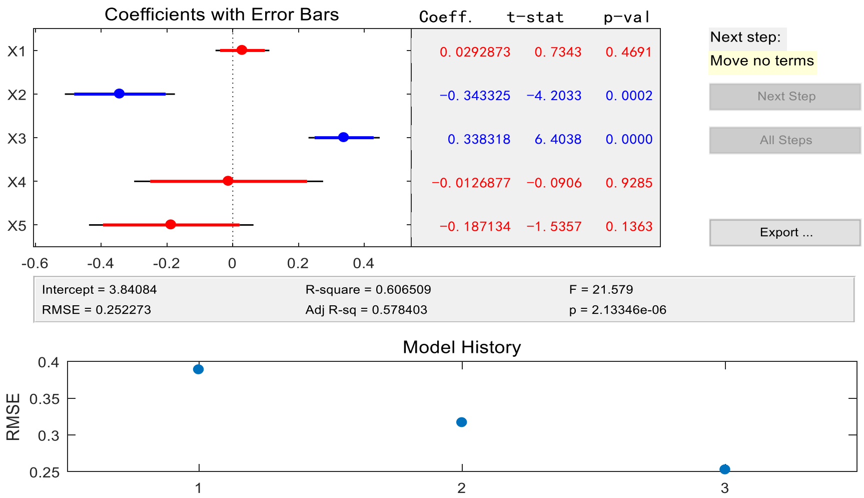

Figure.2. stepwise regression

As can been seen, X_2, X_3 are finally chose, *RSME*, R^2 and F have both improved. RSME becomes smaller, R^2 and F becomes bigger. As a result, significant level improves a lot.

Here is the mew equation:

$$\ln y = 3.8408 - 0.3433 * \ln x_2 + 0.3383 * \ln x_3 \tag{4}$$

F test: As can been seen in the Figure, $F \rangle F_{\alpha}(2{,}28)$

T test: Given $\alpha = 0.05$, when n-k-1=28, $t_{\alpha/2}(n-k-1) = 2.763$.Abviously,t-stat2, t-stat3 are both bigger than2.763.

p=3.133e-6<<0.05, which indicates that all variances are significant.

GDP of the secondary industry and rainfall both have significant influences on PM2.5 concentrations in major cities. However, their influences are reverse.

To further study whether X_1 should be eliminated, X_1 is added again to test whether the result goes bad or not. The result is as follows:

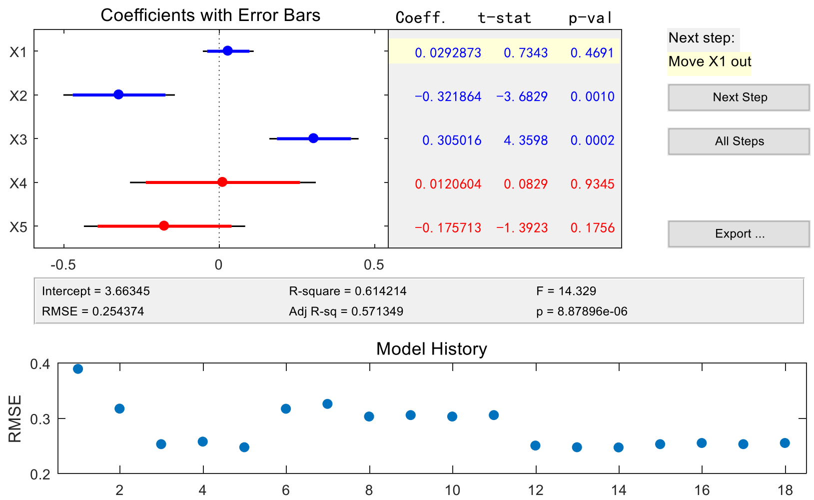

Figure.3. Stepwise regression when X1 is added

IOP Conf. Series: Earth and Environmental Science **170** (2018) 032064 doi:10.1088/1755-1315/170/3/032064

The equation when adding X_1 is:

$$\ln y = 3.6635 + 0.02929 \ln x_1 - 0.3219 * \ln x_2 + 0.3051 * \ln x_3 \quad (5)$$

When X_1 is added, it is obvious that the result is not so good as that of the former calculation. What is more, the coefficient of X_1 is 0.02929, which is much smaller than that of the other two. After these detailed test, the small effect of X_1 can be neglected so finally annual rainfall and GDP of secondary industry are selected as two main factors. The equation is:

$$\ln y = 3.8408 - 0.3433 * \ln x_2 + 0.3383 * \ln x_3 \quad (6)$$

2.7. Conclusion
Assuming other variables remain unchanged, GDP of the secondary industry increases by 1%, the average PM2.5 concentration in major cities increases by 0.3383%. Annual rainfall increases by 1%, the average PM2.5 concentration in major cities increases by 0.3433%.

GDP of the secondary industry reflects the development of industry to some extent. There is no deny that the development would cause pollution at the present. Adequate rainfall will improve the air condition. The result is consistent with economic significance.

2.8. Policy suggestions
At present, although the government has established various prevention and control policies on air pollution, the actual effect is not obvious. The problem of air pollution, especially the high concentration of PM2.5, has not been solved continuously and efficiently. There are many factors contributing to the high concentration of PM2.5 in major cities, involving economic, political and social aspects. The solution of the PM2.5 problem should be coordinated with the society and economy so that it can be more effective.

According to the result, increasing the basic research on the source of PM2.5 concentration and carrying out atmospheric environmental assessment are needed. Local governments should conduct special tracking studies on the causes and diffusion characteristics of PM2.5 concentrations in their respective regions. In this paper, the overall influencing factors of 31 major cities are studied but there is no detailed analysis of a single case, only common characteristics are obtained. Different regions must have their own priorities. Later, governments at all levels should strengthen cooperation with research institutions, using scientific evaluation system.

Moreover, a multi-participation governance system should be established. At present, the main participant of environmental governance is the government, which weakens the functions of other departments and multiple participation is insufficient. In the current situation, different advantages of various roles such as enterprises and residents should be fully used. The task of the government is to build the overall governance framework, guide and centralize power, and provide financial support in time. In this paper, the influences of the secondary industry and dust emission have both been proved. Therefore, for enterprises, especially the secondary industry sector, their task is to follow the rules and regulations, and use advanced scientific technology to reduce waste discharge. For the waste that has been produced, it is necessary to guarantee safety discharge.

References

[1] Wu J, Zhang P, Yi H, et al. What Causes Haze Pollution? An Empirical Study of PM2.5 Concentrations in Chinese Cities[J]. Sustainability, 2016, 8(2):132.

[2] Degobbi C, Lopes F D T Q S, Carvalho-Oliveira R, et al. Correlation of fungi and endotoxin with PM2.5 and meteorological parameters in atmosphere of Sao Paulo, Brazil[J]. Atmospheric Environment, 2011, 45(13):2277-2283.

[3] Li X, Gong Y. Design of Polymeric Nanofiber Gauze Mask to Prevent Inhaling PM2.5 Particles

IOP Conf. Series: Earth and Environmental Science **170** (2018) 032064 doi:10.1088/1755-1315/170/3/032064

from Haze Pollution[J]. Journal of Chemistry, 2015, 2015:1-5.

[4] Mai J, Deng T, Yu L, et al. A modeling study of impact of emission control strategies on PM2.5 Reductions in Zhongshan, China, Using WRF-CMAQ[J]. Advances in Meteorology,2016, (2016-4-14), 2016, 2016(1):1-11.

[5] Zhong, Zhangxiong, Huang, et al. Characteristic Analysis of OC and EC in PM2.5 of Typical Haze Weather in Wuhan City[J]. Meteorological and Environmental Research, 2014(4):19-22.

[6] Eeftens M, Beelen R, Hoogh K D, et al. Eeftens M, Beelen R, de Hoogh K, et al. Development of Land Use Regression models for PM (2.5), PM (2.5) absorbance, PM (10) and PM(coarse) in 20 European study areas; results of the ESCAPE project[J]. Environmental Science & Technology, 2012, 46(20):11195-11205.

[7] Wu D L, Lin M, Chan C Y, et al. Influences of Commuting Mode, Air Conditioning Mode and Meteorological Parameters on Fine Particle (PM2.5) Exposure Levels in Traffic Microenvironments[J]. Aerosol & Air Quality Research, 2013, 13(2):709-720.