ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 12, No. 1, 2017, *pp*.014-019

Analysis of influencing factors of $PM_{2.5}$ based on regression equation

Jingrong Sun[1]
[1]*School of Information and Science Technology, Jinan University, Guangzhou, 510000, China, E-mail:1196972348@qq.com*
(Received July 21, 2016, accepted November 14, 2016)

Abstract. According to the AQI data and the meteorological data of Xi'an in the last years, the relationships and the influence principles between $PM_{2.5}$ and other five monitoring indicators of AQI, weather factors and heating time were analyzed, respectively, by the regression analysis and the ridge regression analysis. The main results include: (1) There were positive correlations between $PM_{2.5}$ and SO_2, NO_2 and CO, which shows that SO_2, NO_2 and CO may be the major gaseous components of forming $PM_{2.5}$. Therefore, the concentration of $PM_{2.5}$ can be reduced by considering how to efficiently decrease the concentrations of SO_2, NO_2, and CO. (2) The relationships between $PM_{2.5}$ and temperature, sea level press, visibility, wind speed and accumulated precipitation are significantly negatively correlated based on the multiple regression. (3) The concentration of $PM_{2.5}$ during the heating period was 1.868 times higher than that during non-heating period. Finally, the ridge regression between $PM_{2.5}$ and all the factors mentioned above shows that SO_2, NO_2, PM_{10}, CO and heating time were more significant than others.

Keywords: $PM_{2.5}$, Air Quality Index, meteorological factors, heating period, multiple regression, ridge regression.

1. Introduction

SQI(Air Quality Index) is a dimensionless index quantitatively describing air quality, which has six basic indexes: SO_2、NO_2、PM_{10}、CO、O_3 and $PM_{2.5}$[1-3].The $PM_{2.5}$ is the emission of particles less than 2.5μm, which is one of the main reasons for producing fog and haze weather[4-10]. All kind of toxic and hazardous substances adsorbed by $PM_{2.5}$, such as virus, can be inhaled into the respiratory system directly into the blood, which will the incidence of lung cancer, heart disease and chronic bronchitis[11-15]. Therefore, it is not only very necessary but also urgent to monitor mass concentration of $PM_{2.5}$ and analyze its pollution characteristics and influencing factors, which plays an important part to harness and predict $PM_{2.5}$ effectively [16].

But now when most of the papers come to discuss the influencing factors they involved nothing but AQI, meteorological factors and heating time. For example, Liu et al [17] draw the conclusions that the relationship between $PM_{2.5}$ and SO_2, NO_2, PM_{10} and CO are positive correlations while the relationship between $PM_{2.5}$ and O_3 are negative positive. Wang et al [18] draw the conclusion that the relationship between $PM_{2.5}$ and SO_2, NO_2 and OX(NO2+O3) are positive correlations; Zhang et al[19] draw the conclusion that the relationship between $PM_{2.5}$ and temperature is positive correlation, while the relationships between $PM_{2.5}$ and wind speed ,relative humidity and air pressure are negative correlation. Chen et al [20-22] draw the conclusion that the relationships between $PM_{2.5}$ and temperature and wind speed are negative correlation while the relationships between $PM_{2.5}$ and humidity and air pressure are positive correlation. Chen et al [23] draw the conclusion that the relationships between $PM_{2.5}$ and temperature, humidity, wind speed, wind direction and precipitation are significantly. When it goes to discuss $PM_{2.5}$, there is little paper take AQI, meteorological factors and heating time into consider at the same time, and there is also little paper considering heating time when they analyze $PM_{2.5}$ by establishing regression equation[24-25].

Basing to the studies having done, this paper selects the object, Xi'an, to analyze for its special location, Guangzhong Plain, surrounded by mountains in the south and Loess Plateau in the north, which limit the dispersion of air pollutants, and the observation of satellite indicates that the Xi'an is one of the most polluted cities [26-28]. This paper considers every factor fully when we study the relationship and regression between $PM_{2.5}$ and other factors, including other five basic indexes of AQI, seven meteorological factors and heating time.

2. Materials and methods

2.1.Data and sources

Data of AQI is downloaded in http://www.xianemc.gov.cn/ including six basic indicators data. Data of meteorology is downloaded in https://www.wunderground.com/ including temperature, the dew point, humidity, sea level pressure, visibility, wind speed and precipitation. Data of longitude and altitude of thirteen areas in Xi'an is downloaded in http://haiba.qhdi.com/.

2.2.research method

When the relationship between $PM_{2.5}$ and heating time is analyzed, we consider 2015-3-1 to 2015-3-15 and 2015-11-15 to 2016-2-29 as heating period, and 2015-3-16 to 2015-11-14 to as non-heating period according to Xi'an's polices. In this paper, we discuss the relationship between $PM_{2.5}$ and its influencing factors, mainly by using the method of multiple regression equation. However, in order to avoid multicollinearity, ridge regression equation between $PM_{2.5}$ and other indexes of AQI, meteorological factors and heating time is built.

3. Results and discussion

Following, we are going to analyze the influencing factors of $PM_{2.5}$ concentration. The relationships between $PM_{2.5}$ and AQI, meteorological factors and heating time are analyzed in section 2.1, 2.2 and 2.3, respectively.

3.1. Ridge regression analysis of $PM_{2.5}$ and other monitoring indexes in AQI

The multiple regression equation is established with the variable of $PM_{2.5}$ as the dependent variable, and $SO_2(x_1)$, $NO_2(x_2)$, $PM_{10}(x_3)$, $CO(x_4)$ and $O_3(x_5)$ as independent variables. As is shown in Fig 1, the relationship is significantly correlated with each monitoring indicator of AQI, therefore, we adopt the ridge regression to avoid multicollinearity. After analyzing, the ridge regression equation of $PM_{2.5}$ to x_1、x_2、x_3、x_4 and x_5 is

$$\hat{y}_{PM_{2.5}} = -16.386 + 0.622x_1 + 0.375x_2 + 0.404x_3 + 0.619x_4 - 0.011x_5. \tag{1}$$

The standard ridge regression equation is

$$\hat{y}_{PM_{2.5}} = 0.150x_1 + 0.135x_2 + 0.315x_3 + 0.212x_4 - 0.006x_5 \tag{2}$$

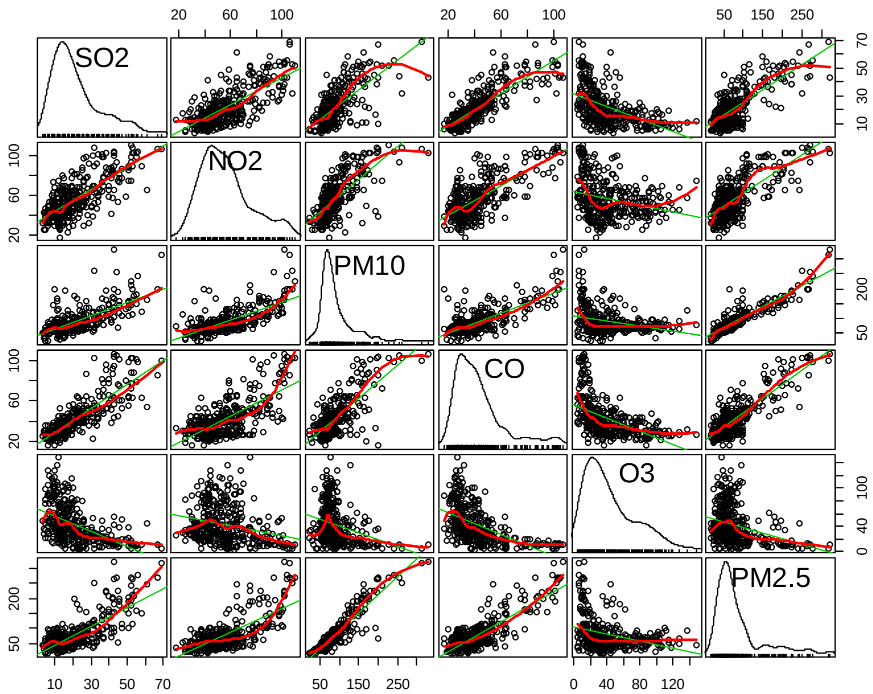


Fig.1 Matrix scatter diagram of AQI monitoring indicators

The standard ridge regression equation indicates that there are high positive correlation between $PM_{2.5}$ and $SO_2(x_1)$, $NO_2(x_2)$, $PM_{10}(x_3)$ and $CO(x_4)$, so we can speculate that $PM_{2.5}$ may be formed by those polluted gases. The relationship between $PM_{2.5}$ and $O_3(x_5)$ is negative correlation. The standard ridge regression equation shows that, when other factors are fixed, $PM_{2.5}$ concentration adds 0.15 units when SO_2 concentration rose for every one unit and so on.

The influencing principles between $PM_{2.5}$ and other polluted gases are as followed: the incomplete combustion of fossil fuels produce not only CO but also $PM_{2.5}$; NO_2 can generating nitrogen oxides and ammonia, and SO_2, nitrogen oxides and ammonia can converted into $PM_{2.5}$ under certain circumstances, thus, $PM_{2.5}$ should have a positive correlation with CO, SO_2 and NO_2; but O_3 can oxide CO, SO_2 and NO_2, further to resist the rising of $PM_{2.5}$ concentration, thus $PM_{2.5}$ should have a negative correlation with O_3.

3.2. Multiple regression analysis of $PM_{2.5}$ and meteorological factors

The multiple regression equation is established with the variable of $PM_{2.5}$ as the dependent variable, and temperature(x_6), the dew point(x_7), humidity(x_8), sea level pressure(x_9), visibility(x_{10}), wind speed(x_{11}) and precipitation(x_{12}) as independent variables. Regression diagnosis of regression equation found residual does not satisfy the Gauss Markov assumptions, so the model should be built after having a logarithmic transformation to explained variable. After eliminating not significant explanatory and collinearity, multiple regression equations is as follows:

$$\log(\hat{y}_{PM2.5}) = 24.456 - 0.017x_6 - 0.019x_9 - 0.129x_{10} - 0.011x_{11} - 0.019x_{12} \tag{3}$$

The p value of t test of regression equation is 4.68e–65, and both the constant and linear terms of regression equation are significant under the significant level of 0.05. The regression equation indicates that the relationships between $PM_{2.5}$ and temperature, sea level pressure, visibility, wind speed and the cumulative precipitation are significantly negatively correlated. The equation shows that when other factors are fixed, $PM_{2.5}$ concentration adds 0.15 units when temperature rose for every one unit and so on.

The regression diagnostic plots showed that residual values were satisfied with the equal variance, independent variance and normal distribution, therefore, the model of regression equation pass the test.

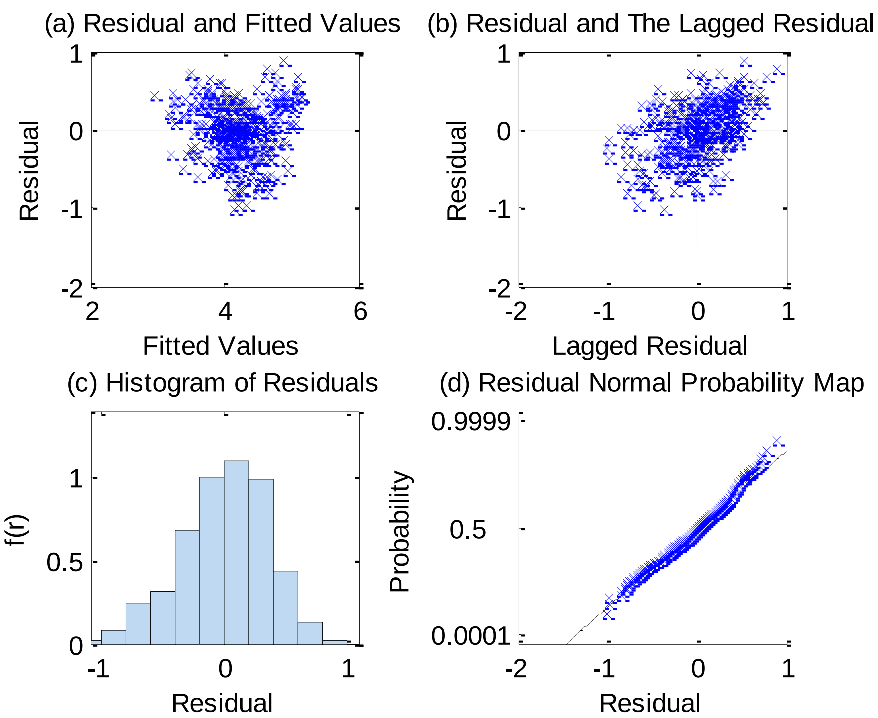
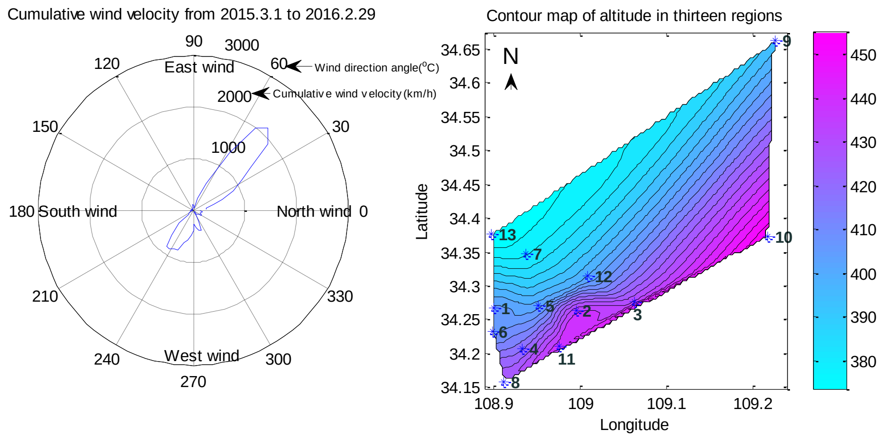


Fig.2 Regression diagnostic residuals

The influencing principles between $PM_{2.5}$ and meteorological factors are as followed: the relationship between $PM_{2.5}$ concentration and temperature is negative correlation. The reason why their relationship is negative correlation is that the temperature increases, the solar radiation and the atmospheric turbulence increase, making the air flow and the mixing layer become larger, which can promote the diffusion and dilution of $PM_{2.5}$ [20]. The relationship between $PM_{2.5}$ and sea level pressure is negative correlation. The reason why their relationship is negative correlation is that when the pressure decrease, which will make the wind speed decrease thus to resist the diffusion of $PM_{2.5}$, the $PM_{2.5}$ concentration will increase [19]. The reason why the relationship between $PM_{2.5}$ and visibility is negative correlation is that the high $PM_{2.5}$ concentration can not only diffuse but also absorb light to decrease the visibility [23,28]. The reason why the relationship between $PM_{2.5}$ and accumulated precipitation is negative correlation is that precipitation can scour the $PM_{2.5}$ in the air. It is easy to understand that why the relationship between $PM_{2.5}$ and wind speed is negative correlation is that the higher wind speed the more beneficial to the diffusion of $PM_{2.5}$. In addition, there is little papers analyzing the $PM_{2.5}$ concentration by combining with the cumulative wind speed, wind direction and altitude of thirteen monitoring stations in Xi'an at the same time. Therefore, the further analysis draw the conclusion showing in Fig.3 that the northeast is the main wind direction and the altitude is gradually increasing from the northwest to the southeast in Xi'an. The $PM_{2.5}$ concentration in the districts of Yanliang and Lintong may be lower, because those districts are located in the northeast direction, direction of air inlet. The $PM_{2.5}$ concentration in the districts of Guangyuntan and Fangzhicheng may be higher, for not only they are located in the direction of air outlet of Lintong and lower altitude than Lintong, but also the altitude in the direction of air outlet of Guangyuntan and Fangzhicheng is higher than Guangyuntan and Fangzhicheng's. Contrary to Guangyuntan and Fangzhicheng, Xingqinxiaoqu's $PM_{2.5}$ concentration may be lower, because not only Xingqingxiaoqu is not be polluted by the direction of air inlet, but also the altitude in the areas of the direction of air outlet is lower than Xingqingxiaoqu.



The number,1 to 13,on the subgraphs of right,represent the Gaoyakaiguanchang,Xingqingxiaoqu,Fangzhicheng,Xiaozhai,Shirenmintiyuch Gaoxinxiqu,Jingkaiqu,Changanqu,Yanliangqu,Lintongqu,Qujiangwenhuajituan,Guangyuntan and Caotan,respectively.

Fig.3 The comparison between wind and ρ(PM2.5) contour map

3.3. Analysis on the relationship between PM2.5 and heating period

After a series of calculations, the mean value of $PM_{2.5}$ concentration is 116.059 during heating period, while 62.135 during non-heating period. In other word, it means that mean value of $PM_{2.5}$ concentration during heating period is 1.868 times as much as the mean value during non-heating period, which indicates that the combustion of coal during heating period has a great impact on the rise of $PM_{2.5}$ concentration. Therefore, on the one hand, it is necessary for government to take emergency measure to improve the technology to produce coal of high quality; on the other hand, we should advocate the use of low emission energy, such as natural gas.

3.3. Comprehensive analysis of relationship between PM2.5 and other indexes of AQI, meteorological factors and heating time

The multiple regression equation is established with the variable of $PM_{2.5}$ as the dependent variable, and $SO_2(x_1)$, $NO_2(x_2)$, $PM_{10}(x_3)$, $CO(x_4)$, $O_3(x_5)$, temperature(x_6), the dew point(x_7), humidity(x_8), sea level pressure(x_9), visibility(x_{10}), wind speed(x_{11}) , precipitation(x_{12}) and heating time(x_{12}) as independent variables. In the equation, heating time is a dummy variable, in which the value 1 represents heating period, while value 0 represents not-heating period. Taking into account the explanatory variables too many, we adopt the ridge regression to avoid multicollinearity. After analyzing, the ridge regression equation is

$$\hat{y}_{PM_{2.5}} = -13.407 + 0.733x_1 + 0.221x_2 + 0.565x_3 + 0.641x_4 + 0.543x_7 - 3.354x_{10} + 3.235x_{13} \quad (4)$$

The standard ridge regression equation is

$$\hat{y}_{PM_{2.5}} = 0.176x_1 + 0.080x_2 + 0.440x_3 + 0.220x_4 + 0.105x_7 - 0.138x_{10} + 0.028x_{13} \quad (5)$$

The goodness of fit of multivariate linear regression giving above is 0.952. The standard ridge regression equation indicates that $SO_2(x_1)$, $NO_2(x_2)$, $PM_{10}(x_3)$, $CO(x_4)$, the dew point(x_7), visibility(x_{10}) and heating time(x_{12}) , compared to other factors, have a significant impact on $PM_{2.5}$. The standard ridge regression equation shows that, when other factors are fixed, $PM_{2.5}$ concentration rise 0.176 units when SO_2 concentration rose for every one unit; and $PM_{2.5}$ concentration during heating time is 0.028 units higher than $PM_{2.5}$ concentration during non-heating time and so on.

4. Conclusion

a) In the analysis of $PM_{2.5}$ and other five monitoring indicators of AQI, there are high positive correlation between $PM_{2.5}$ and $SO_2(x_1)$, $NO_2(x_2)$, $PM_{10}(x_3)$ and $CO(x_4)$, so we can speculate that $PM_{2.5}$ may be formed by those polluted gases. Therefore, we can decrease $PM_{2.5}$ concentration by thinking about how to effectively reduce those polluted gases.

b) In the analysis of $PM_{2.5}$ and meteorological factors, not only the multiple regression of $PM_{2.5}$ and meteorological factors is giving, but also the influencing principles between $PM_{2.5}$ and meteorological factors are giving, which has an important guiding significance to predict the trend of $PM_{2.5}$.

c) In the analysis of $PM_{2.5}$ and heating time, mean value of $PM_{2.5}$ concentration during heating period is 1.868 times as much as the mean value during non-heating period, which indicates that the combustion of coal during heating period has a great impact on the rise of $PM_{2.5}$ concentration.

d) Finally, in the Comprehensive analysis of relationship between $PM_{2.5}$ and other indexes of AQI, meteorological factors and heating time, ridge regression equation indicates that $SO_2(x_1)$, $NO_2(x_2)$, $PM_{10}(x_3)$, $CO(x_4)$, the dew point(x_7), visibility(x_{10}) and heating time(x_{12}) , compared to other factors, have a significant impact on $PM_{2.5}$, which have a theoretical guidance for the effective governance and prediction of $PM_{2.5}$.

5. References

[1] Zhang Yanjie, Lin Zeyi, Zheng Yu. Correlation Analysis between Air Qualityu Index and Air Pollotants Harbin in Winter.Journal of Northeast Forestry University. 43(11):122–126(2015).

[2] LANZAFAME R,MONFORTE P,PATANE G, et al. Trend analysis of Air Quality Index in Catania from 2010 to 2014. Energy Procedia. 82:708-715(2015).

[3] CHEN Bo, LU Shaowei, LI Shaoning, et al. Impact of fine particulate fluctuation and other variables on Beijing's air quality index.Environmental Science and Pollution Research. 22(7):5139-5151(2015).

[4] Cheng Yali， Wang Bo. Research on Related Factors of PM2.5 in Shanghai. Mathematical Theory and Applications. 34(3) :96–103(2014).

[5] PEDERSEN D U, DURANT J L,TAGHIZADEH K, et al. Human Cell Mutagens in Respirable Airborne Particles from the Northeastern United States. 2. Quantification of Mutagens and Other Organic Compounds. Environmental Science and Technology. 39(24): 9547-9560(2005).

[6] AMIN N. The Association of PM2.5 and Surface Ozone with Asthma Prevalence among School Children in Japan 2006-2009. Health. 5(10B):1-7(2013).

[7] NINOMIYA Y, WANG Qunying, XU Shuyin, et al. Evaluation of a Mg-Based Additive for Particulate Matter (PM)2.5 Reduction during Pulverized Coal Combustion. Energy and Fuels. 24(1):199-204(2010).

[8] Kong Shaofei, Li Xuxu, Li Li, et al. Variation of polycyclic aromatic hydrocarbons in atmospheric PM2.5 during winter haze period around 2014 Chinese Spring Festival at Nanjing: Insights of source changes, air mass direction and firework particle injection. Science of The Total Environment. 520:59-72(2015).

[9] Qiao Ting, Zhao Mengfei, Xiu Guangli, et al. Simultaneous monitoring and compositions analysis of PM1 and PM2.5 in Shanghai: Implications for characterization of haze pollution and source apportionment. Science of The Total Environment. 557-558:386-394(2016).

[10] HO K F, HO S SH, Huang Rujin, et al. Chemical composition and bioreactivity of PM2.5 during 2013 haze events in China. Atmospheric Environment. 126:162-170(2016).

[11] TANDAY S. India launches air quality index to tackle pollution problems. Lancet Oncology. 16(5):e203(2015).

[12] Zhang Ying, LI Zhengqiang. Remote sensing of atmospheric fine particulate matter (PM2.5) mass concentration near the ground from satellite observation. Remote Sensing of Environment. 160:252-262(2015).

[13] BARI M A, KINDZIERSKI W B. Eight-year (2007–2014) trends in ambient fine particulate matter (PM2.5) and

its chemical components in the Capital Region of Alberta, Canada. Environment International. 91:122-132(2016).

[14] MISHRA D, GOYAL P, UPADHYAY A. Artificial intelligence based approach to forecast PM2.5 during haze episodes: A case study of Delhi, India. Atmospheric Environment. 102:239-248(2015).

[15] Sun Xiaoli, Luo Xiping, Zhao Chunmei, et al. The associations between birth weight and exposure to fine particulate matter (PM2.5) and its chemical constituents during pregnancy: A meta-analysis. 211:38-47(2016).

[16] Song Ci, Pei Tao, Yao Ling. Analysis of the Characteristics and Evolution Modes of PM2.5 Pollution Episodes in Beijing, China During 2013.International Journal of Environmental Research and Public Health. 12(2):1099-1111(2015).

[17] Liu Wenjin, Zheng Guoyi, Tian Xue, et al.PM2.5 Factors Associated with Multivariate Regression Analysis Model in Xi’an.Journal of Quantitative Economics. 32(1) :85–88(2015).

[18] Wang Zhanshan, Li Yunting, Chen Tian, et al. Spatial-temporal characteristics of PM2.5 in Beijing in 2013.ActaGeographica Sinica. 70(1):110–120(2015).

[19] Zhang Siyuan, Wu Kaiya. Analysis of Pollution Characteristics and Influencing Factors of PM2.5 in Hefei.Environmental Science And Management. 40(1) :51–56(2015).

[20] Chen Yang, Zeng Yu, Zhang Qin, et al. Study on the Influence of Meteorological Factors on PM2.5 Periodic Variation in Changsha City.Sichuan Environment. 33(6) :81–87(2014).

[21] Huang Fangfang, Li Xia, Wang Chao, et al.PM2.5 Spatiotemporal Variations and the Relationship with Meteorological Factors during 2013-2014 in Beijing, China. PLoS One. 10(11):e0141642(2015).

[22] Huang Ping, Zhang Jingyuan, Tang Yuxiang, et al. Spatial and Temporal Distribution of PM2.5 Pollution in Xi’an City, China. International Journal of Environmental Research and Public Health. 12(6):6608-6625(2015).

[23] Chen Guanyi, Zhang Wen, Hou Li'an, et al. Pollution Characteristics and Influence Factors of PM2.5 in Summer in Jixian County of Tianjin. Journal of Tianjin University. 48(2):95–102(2015).

[24] Zhang Chen, Ni Zhiwei, Ni Liping. Multifractal detrended cross-correlation analysis between PM2.5 and meteorological factors. Physica A: Statistical Mechanics and its Applications. 8:114-123(2015)

[25] Pu Weiwei, Zhao Xiujuan, Zhang Xiaoling, et al. Effect of Meteorological Factors on PM2.5 during July to September of Beijing. Procedia Earth and Planetary Science. 2: 272-277(2014).

[26] Zhang Qian, ShenZhenxing, Cao Junji, et al. Variations in PM2.5, TSP, BC, and trace gases (NO2, SO2, and O3) between haze and non-haze episodes in winter over Xi'an, China. Atmospheric Environment.12:64-71(2015).

[27] Wang Ping, Cao Junji, Shen Zhenxing, et al. Spatial and seasonal variations of PM2.5 mass and species during 2010 in Xi'an, China. Science of the Total Environment.508:477-487(2015).

[28] Hua Yang, Chen Zhen, Wang Shuxiao, et al. Characteristics and source apportionment of PM2.5 during a fall heavy haze episode in the Yangtze River Delta of China.Atmospheric Environment.123(Part B):380-391(2015).