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

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

The standard ridge regression equation is

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