Study On Probability Estimation Of Haze Weather In Beijing Based On Cumulative Logistic Regression Model

Y H Zhu, S Z Li
2 Beinong Road, Huilongguan Town, Changping District, Beijing, China
zhuyonghua@ncepu.edu.cn   18010039283@163.com

Abstract. Based on the meteorological data and PM2.5 concentration data in Beijing every three hours from January 1 to December 31, 2016, this paper establishes a cumulative logistic regression model of haze pollution probability analysis, and then analyses the main meteorological factors affecting haze weather in different seasons, estimating the probability of occurrence of haze weather at various levels under different meteorological factors.

1. Introduction
Frequent haze weather in China has attracted the attention of the community. In many areas, haze weather has been used as a warning and forecast for severe weather. Fog is an aerosol system consisting of a large number of tiny water droplets or ice crystals suspended in the air near the ground. Haze refers to the deterioration of visibility caused by aerosol pollution in the near-surface atmosphere. The nature of haze weather is atmospheric PM2.5 pollution. PM2.5 refers to fine particles with an aerodynamic equivalent diameter less than or equal to 2.5 μm in the ambient atmosphere. Haze weather has a direct impact on many aspects, especially the people's lives, transportation and health. Domestic and foreign scholars have conducted extensive research on PM2.5. The research mainly focuses on the chemical composition of PM2.5, its impact on human health, visibility and climate, and source analysis. At present, domestic research on PM2.5 focuses on the analysis of the relationship between basic meteorological conditions, other atmospheric pollutants and PM2.5. Methods focus on simple correlation analysis, linear regression models, and principal component analysis methods.

2. Data collection and variable selection
The original data of Beijing PM2.5 concentration used in this study is from the official website of the US Embassy in China. Meteorological data comes from Beijing Station, a regular observational data set of the China Meteorological Data Network. This paper selects the PM2.5 concentration values and meteorological data for each three-hour period from January 1 to December 31, 2016. After processing the missing data, a total of 2,526 data is collected. According to the ambient air quality standards, the PM2.5 data is used to classify the haze. Currently, it is classified into 6 levels, namely excellent, good, light pollution, medium pollution, heavy pollution, and severe pollution. In this paper, we will study the ratio of haze weather with high pollution degree to haze weather with low pollution degree. Therefore, the cumulative response variables are sorted as follows:

| PM(μg/m³) | AQI | Grade of Air Quality | Response variable (Haze) | Occurrence probability |
|-----------|-----|----------------------|--------------------------|-----------------------|

Table 1: Haze pollution classification
Explanatory variables selected the meteorological factors such as temperature, press (The average atmospheric press in Beijing is 1012.9 Pa. Set press = 0 to indicate that the air press is less than the annual average, and press =1 to indicate that the air press is higher than the annual average), precipitation(Pre_1h), relative humidity (Rhu), wind direction(according to Beijing meteorological data, the wind direction is divided into 4 locations, namely north, east, south, west), and wind scale.

3. Establishment and Examination of Cumulative Logistic Regression Model

3.1 Establishment and Examination of Cumulative Logistic Regression Model

Because of the orderly relationship among the response variable levels, a cumulative logistic regression model was used for data analysis. The model contains five cumulative logistic functions and it is estimated simultaneously:

\[
\begin{align*}
\ln\left(\frac{p_1}{p_2 + p_3 + p_4 + p_5 + p_6}\right) &= \beta_01 - \sum_{k=1}^{K} \beta_k x_k, \\
\ln\left(\frac{p_1 + p_2}{p_3 + p_4 + p_5 + p_6}\right) &= \beta_02 - \sum_{k=1}^{K} \beta_k x_k, \\
\ln\left(\frac{p_1 + p_2 + p_3}{p_4 + p_5 + p_6}\right) &= \beta_03 - \sum_{k=1}^{K} \beta_k x_k, \\
\ln\left(\frac{p_1 + p_2 + p_3 + p_4}{p_5 + p_6}\right) &= \beta_04 - \sum_{k=1}^{K} \beta_k x_k, \\
\ln\left(\frac{p_1 + p_2 + p_3 + p_4 + p_5}{p_6}\right) &= \beta_05 - \sum_{k=1}^{K} \beta_k x_k,
\end{align*}
\] (1)

Among them, \(p_1, p_2, p_3, p_4, p_5, p_6\) respectively represent the probability of occurrence of severe pollution, heavy pollution, medium pollution, light pollution, good and excellent. And there are

\(p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1\) (2)

Taking weather factors and seasonal factors as independent variables, among them: using temperature, relative humidity (Rhu), and wind scale as continuous independent variables; as categorical variables, seasons are divided into spring, summer, autumn, winter, and winter is used as a reference variable; as categorical variables, the wind direction is divided into north, east, south, west, and west is used as a reference variable. According to whether or not the press is higher than the average value at this moment, it is divided into two categories, so the press is introduced as a categorical variable into the model. Based on the degree of pollution of haze at this moment as a dependent variable, the maximum likelihood method is used to solve the problem, and the annual haze pollution estimation model for 2016 is obtained.

| Haze_1 | Haze_2 | Haze_3 | Haze_4 | Haze_5 |
|--------|--------|--------|--------|--------|
| 0-35   | 0-50   | excellent | 6      | P6     |
| 36-75  | 51-100 | good    | 5      | P5     |
| 76-115 | 101-150| light   | 4      | P4     |
| 116150 | 151-200| medium  | 3      | P3     |
| 151-250| 201-300| heavy   | 2      | P2     |
| 251 or more | 301-500| severe  | 1      | P1     |
According to the estimation results of the model, for the whole year, when the temperature, wind direction, wind scale, season, and press of the variable are the same, the relative humidity in the air increases by 10%, and the haze weather occurrence ratio is 1.05 times as much as the original. In the same situation of relative humidity, wind direction, wind scale, season and press, when the temperature is increased by 1 degree, the occurrence rate of haze weather is 1.11 times as much as the original. In the seasonal occurrence of haze, winter is the season with high haze. The occurrence of haze in winter is 15.1 times as much as spring, which is 148.6 times of summer and 11.8 times of autumn. Looking from the wind direction, when the wind direction is east, the occurrence of haze is the largest. The occurrence of haze in the east wind is 1.56 times as much as the westerly winds, which is 2.39 times of the south winds and 1.55 times of the north winds.

3.2 Seasonal Cumulative Logistic Regression Model

Next the cumulative logistic regression model is established by season, and the impact of various meteorological factors on haze weather is analysed.

After sorting the data by season, the degree of pollution (severe pollution, heavy pollution, medium pollution, light pollution, good, and excellent) of haze weather is used as a reaction variable, and atmospheric press, wind force, temperature, relative humidity, and wind direction as independent variables are introduced in the model. The result of the calculation is as follows:

Table 3: Spring Haze Pollution Model Estimation Results in Beijing

| Haze_1 | Haze_2 | Haze_3 | Haze_4 | Haze_5 |
|--------|--------|--------|--------|--------|
| estimated value | -9.257 | -8.444 | -7.299 | -5.034 |
| Standard error | 0.895 | 0.859 | 0.828 | 0.772 |

| Press | Wind_Scale | Tem | Rhu | North | East | South | West |
|-------|------------|-----|-----|-------|------|-------|------|
| estimated value | 1.065 | 0.094 | -0.174 | -4.33 | -0.96 | -1.2 | -0.322 | 0 |
| Standard error | 0.382 | 0.165 | 0.028 | 0.646 | 0.375 | 0.369 | 0.361 |
Table 4: Summer Haze Pollution Model Estimation Results in Beijing

| Haze_1   | Haze_2   | Haze_3   | Haze_4   | Haze_5   |
|----------|----------|----------|----------|----------|
| estimated value | -11.79   | -10.691  | -9.479   | -7.373   | -5.154   |
| Standard error     | 1.262    | 0.966    | 0.837    | 0.78     | 0.758    |

| Press  | Wind_Scale | Tem        | Rhu      | North    | East     | South    | West |
|--------|------------|------------|----------|----------|----------|----------|-------|
| estimated value | 0.052     | -0.388    | -0.087   | -3.103   | 0.37     | -0.117   | -0.33 |
| Standard error  | 0.319     | 0.112     | 0.022    | 0.431    | 0.244    | 0.217    | 0.207 |

Table 5: Autumn Haze Pollution Model Estimation Results in Beijing

| Haze_1   | Haze_2   | Haze_3   | Haze_4   | Haze_5   |
|----------|----------|----------|----------|----------|
| estimated value | -5.063   | -4.145   | -3.08    | -1.986   |
| Standard error     | 0.748    | 0.736    | 0.726    | 0.715    |

| Press  | Wind_Scale | Tem        | Rhu      | North    | East     | South    | West |
|--------|------------|------------|----------|----------|----------|----------|-------|
| estimated value | 0.896     | -0.044    | -0.017   | -3.454   | -0.123   | -0.48    | -0.957 |
| Standard error  | 0.234     | 0.162     | 0.018    | 0.6      | 0.347    | 0.323    | 0.324 |

Table 6: Winter Haze Pollution Model Estimation Results in Beijing

| Haze_1   | Haze_2   | Haze_3   | Haze_4   | Haze_5   |
|----------|----------|----------|----------|----------|
| estimated value | -6.754   | -5.487   | -4.755   | -3.864   | -2.753 |
| Standard error     | 0.465    | 0.445    | 0.435    | 0.423    | 0.412  |

| Press  | Wind_Scale | Tem        | Rhu      | North    | East     | South    | West |
|--------|------------|------------|----------|----------|----------|----------|-------|
| estimated value | 0.492     | -0.008    | -0.133   | -6.497   | -0.326   | -0.542   | -0.311 |
| Standard error  | 0.241     | 0.099     | 0.014    | 0.389    | 0.215    | 0.207    | 0.224 |

Tables 3 to 6 show the estimated results of the seasonal haze pollution model in Beijing. Through calculating the estimated results from the model: in the four seasons, the most significant effect on haze is relative humidity (Rhu), under the control of other meteorological factors, when relative humidity increases by 1%, the occurrence of haze weather with high levels of pollution in spring increases by 6% compared with the original situation, increases by 3% in summer, increases by 4% in autumn, and increases by 7% in weather. Atmospheric press has the most significant influence on haze.

...
weather in spring and autumn, when atmospheric press is higher than average in the spring, the occurrence of haze weather with high pollution degree is 2.9 times than when press is lower than the average. 2.4 times in autumn. It is significant that the influence of wind scale on the occurrence of haze weather in summer, for every additional level of wind scale, the occurrence of haze weather with a high level of pollution is increased by 1.47 times compared with the original situation. In the seasonal regression model, the influence of temperature on haze weather in autumn is not obvious, under the condition of controlling other meteorological factors, in spring, when the temperature rises by 1 degree, the occurrence ratio of haze weather is 1.19 times as much as the original, 1.09 times in summer, and 1.14 times in winter. The regression of relative humidity (Rhu) in the four-season regression model performed well, for each 1% increase in relative humidity, the occurrence of haze in spring, summer, and winter increase to 1.04 times, 1.03 times, 1.03 times, and 1.07 times. The influence of the wind direction on the haze weather in spring is significant, under the same conditions of other meteorological factors, when the north wind is blowing, the haze weather with more serious pollution is 2.61 times as much as the westerly wind, and the easterly wind is 3.32 times as much as the westerly wind.

4. Summary and outlook
In this paper, according to different seasons, with meteorological factors as independent variables, the cumulative logistic regression model of Beijing haze pollution in 2016 was established, analysed the main meteorological factors that affect haze weather. At the same time, the probability of occurrence of haze weather at all levels under various meteorological factors is estimated. Of course, this paper only analyses the probability of occurrence of haze weather in Beijing in 2016 and its main influencing factors through cumulative logistic regression model. The sample data is not sufficient and extensive, which has a great influence on the accuracy and credibility of the analysis results. We hope that the country will have long-term and comprehensive monitoring of haze pollution. Only by obtaining a large amount of sample data can we establish a more complete mathematical model to obtain the more comprehensive, more accurate and credible research results of haze disciplinarian.

References
[1] Wu D. 2012. Haze weather research in China in the last decade: A review [J]. Acta Scientiae Circumstantiae, 32 (02):257-269.
[2] Gu W D. Study on the Special Formation Mechanism of Haze in China [J]. Macroeconomic research, 2014, (06):3-7+123(in Chinese).
[3] Wang J C, Guo Z G. Logistic Regression Model - Methods and Applications [M]. Beijing: Higher Education Press.2001
[4] Zhang Z H. Study on the Characteristics of Temporal and Spatial Variation of PM2.5 Concentration, Influencing Factors and Source Analysis [D]. Zhejiang University,2014.
[5] Han D W. Effects of temperature and relative humidity on the vertical distribution of aerosol mass concentration [A]. 2007:8.
[6] Wu D. Differences between haze and fog and discussion of early warning of haze [J]. Meteorology, 2005, (04):3-7.
[7] Wang Q, Sun W and Zhang X Y. Analysis of the Relationship between PM2.5 Mass Concentration Distribution and Meteorological Conditions in Beijing [J]. Computer and Applied Chemistry, 2014, (10):1193-1196.
[8] Cao C, Jiang W, Wang B, Fang J, Lang J, Tian G, Jiang J and Zhu T F, Inhalable microorganisms in Beijing’s PM2.5 and PM10 pollutants during a severe smog event. Environmental Science And Technology,Vol.48,No.3,2014.
[9] Zhang Y, Li G, et al. The spatial characteristics of ambient particulate matter and daily mortality in the urban area of Beijing, China. Science of the Total Environment, 2012, 435:14-20.
[10] Dong X L, Liu D M, Yuan Y S, et al. Pollution characteristics and influencing factors of atmospheric particulates in Beijing during the summer of 2005. Chinese Journal of Environmental Engineering, 2007, 1(9): 100-104