Dean&Francis
ISSN 2959-6157

The Analysis of Nanjing Air Quality by Using Multiple-Linear Regression Model

Liuhao Huang

Reading Academy, Nanjing University of Information Science and Technology, Nanjing, Jiangsu province, 211800, China

vm808696@student.reading.ac.uk

Abstract:

For the sake of industrial evolution, booming population, increased vehicles, and so forth, the air quality index (AQI) needs to be emphasized. This research paper takes Nanjing's AQI as the survey background. This essay has deployed the multiple linear regression model to help analyze the basic circumstances of the AQI by using monthly data on six basic factors. The results show that the most significant factor of Nanjing's AQI is the density of O3. What's more, NO2 plays the role of the negative correlation and there exists a linear relationship between the variables. To improve or maintain the "generally good" circumstances of the AQI, it's better to take measures like enlarging the vegetation, reducing the activities of incomplete incendiary, and increasing the usage of public transport... Additionally, this essay's goal is to raise the awareness of the public and point out the mainly significant factors, that can offer evidence for the proceeding survey and measures.

Keywords: AQI, Analyse, multiple-linear regression model.

1. Introduction

Due to the rapid development of urban and industrialization, booming population and economy, and increased number of vehicles, air quality in China has decreased. Nowadays, worse condition of air quality can lead to reduced visibility, harm human health, have significant effects on climate change, and damage the gorgeous image of the cities [1, 2]. Therefore, analysis of the air quality has a deeply significant meaning. Additionally, it can propel urban development in various dimensions.

In recent years, more and more humanity begun to pay attention to air quality problems. For example, the China air quality monitoring stations in main cities have notified hourly average mass concentration statistics of air pollutants and published a vast number of research literature on air quality. Beijing-Tianjin-Hebei region, Yangtze River Delta, Pearl River Delta, and Chengdu-Chongqing region have become the main research areas [2]. Nanjing is a typical city in the Yangtze River Delta, whose population has reached 9.547 million and the number of vehicles has become 0.308 million until 2023. There also gathered at least 2 heavy industrial firms [2]. Additionally, Nanjing is situated in the northern subtropical area. Subtropical locality is humid in summer, however, aridity in winter. Then inversion phenomenon can be

ISSN 2959-6157

easily formed. As a result, the air quality issues become severe for the sake of climate features [3]. What's more, someone has found that the digital economy can make giant contributions to air pollution migration in China via industrial optimization and green innovation [4]. Wang has also made research on the application of the Nanjing air quality health index based on functional time series analysis. As a result, Wang has found that the main pollutant is PM10 and the survey based on the functional time series is superior to others [5]. Additionally, China activates the Air Pollution Prevention and Control Action (APPCA) in the year 2013 [6, 7]. Chen did the estimated PM2.5 concentration in China by using machine learning technology [8]. Wu et.al has noted that the large area of vegetation can alleviate air pollution to some extent, which is proved by doing experiments [9]. However, the survey on Nanjing's specific air quality changes still has demerits.

This essay will make the corresponding diagrams by using the data from China's air quality online monitoring and analysis platform to help analyze the air quality of Nanjing. This research paper tends to propel various aspects of the AQI's improvements and provide specific suggestions for protecting air quality in the long term.

2. Multiple-Linear Regression Model's Set and Test

2.1 Main Components of the Pollutants

First of all, PM2.5 is a pollutant whose diameter is less than or equal to 2.5 microns. PM2.5 is one of the significant standards to evaluate air pollution, it can penetrate deep into people's lungs and have a giant effect on the populace's health. Additionally, PM2.5's main sources include incendiary activities, industrial emissions, and raised dust. Secondly, PM10 is a pollutant whose diameter is less than or equal to 10 microns, also called Inhaler particulate matter. PM10 can enter the body's respiratory tract, and harm human health.PM10 has similar sources to PM2.5, including various incendiary processes, industrial production, raised dust, and so forth. Thirdly, NO2 is a kind of normal nitrogen oxide, one of the main pollutants. NO2 possesses stimulating smells, harmful to an individual's health, and mainly influences the respiratory system. NO2 is mainly from motor vehicle exhaust, industrial emissions, and other high-temperature incendiary processes.

CO is a kind of gas with toxic, with no colour and no smell, mainly produced from the incomplete incendiary procession, for example, motor vehicle exhaust, forest fires, domestic gas equipment et al. Additionally, CO will harm human health. Additionally, SO2 is a kind of normal sulfur oxide, one of the main pollutants. SO2 also has a stimulating smell, harmful to civilization's health, and mainly has an impact on the respiratory system. It is mainly produced from burning processes (burning the fuels like coal and petroleum). O3 takes a bit of proportion of the atmosphere, however, it's crucial to human survival. The O3 is divided into 'good types' and 'bad types'.

2.2 Data Collection and Release Channels

This essay adapted the data from China's air quality online monitoring and analysis platform in the year 2014-2024. Additionally, six pollutants were selected as the interpretation variables (PM2.5, PM10, CO, NO2, SO2, O3). About the data collection and posting channels, firstly, the relevant data is collected primarily by air quality monitoring stations located in various locations. For example, China has set up a large number of air quality monitoring stations in main cities. Especially in the Beijing-Tianjin-Hebei region, Yangtze River Delta, Pearl River Delta, and Chengdu-Chongqing region [2]. Secondly, the data posting channel can be the official website, apps, social media, and the third-party API interface.

2.3 Methods and Resources

The research paper uses the linear regression method to do the analysis and prediction. First of all, this essay will deploy the basic statistical strategy to check the layout of the data. This research paper adopts the layout of data by using three kinds of figures. The frequency bar chart, Boxplot chart, and density estimation line graph represent the data layout clearly and precisely in three ways.

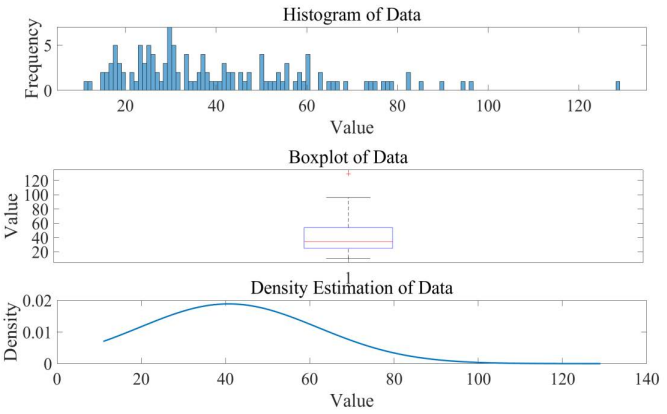

Fig. 1 The data distribution visualization of PM2.5 (cited from: original)

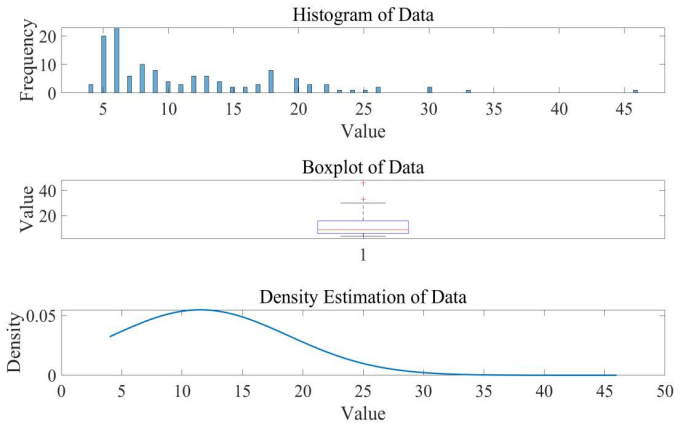

Fig. 2 The data distribution visualization of SO2 (cited from: original)

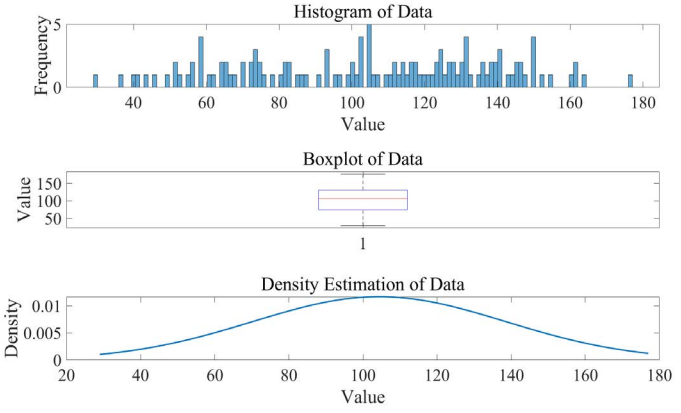

Fig. 3 The data distribution visualization of O3 (cited from: original)

As Figure 1 indicates, it can be found that the data of PM2.5 is mainly distributed between 20 to 60, the same as the NO2. However, PM10 mainly spread from 40 to 100. The data layout of the three is similar. What's more, as Figure 3 indicates, O3 has a wider range of layout than PM2.5, which is spread from 40 to 140. Additionally, figure 2 shows that CO is mainly located from 0.5 to 1 and SO2 is from 5 to 15. The data layout of these two seems to be similar. This implies that there may exist some relationship between emission sources and the environment

ISSN 2959-6157

under particular requisitions. In addition, the density estimation reaches the peak of about 40, 40, 72, 108, 0.8, and 12 respectively.

NO2, CO, and O3's distributions seem more like a normal distribution, however, the images of CO and O3 should be partly moved as a whole to the left. The left-moved curves are named 'Positive Skewness', which is due to the existence of extreme maximum value among data, data truncation, or the nature of natural phenomena. The distribution images of PM2.5, PM10, and SO2 seem to be similar, however, the exact data are different.

2.3 Model's Set

By using the scatter plot analysis of each air pollutant and AQI, which is shown in Figure 4 (the linear relationship between PM10 and AQI), it's simple to find that there exists a linear relationship between each air pollutant and AQI index. Then, the multiple regression equation is set up as follows:

$$Y\beta + \text{¨»}1X1 + \text{¨»}2X2 + \text{¨»}3X3 + \text{¨»}4X4 + \text{¨»}5X5 + \text{¨»}6X6 \quad (1)$$

For the equation (1), the X1, X2, X3, X4, X5, X6 and Y denote PM2.5, PM10, CO, NO2, SO2, O3 and AQI respectively. While α and β represent unknown coefficients. Next, it deploys the least squares to do the parameter estimation for equation (1), through MATLAB's calculation, the multiple regression equation is as follows:

$$Y = 11.951 + 0.70759X1 + 0.073317X2 + 15.827X3 - 0.22188X4 + 0.14947X5 + 0.29897X6 \quad (2)$$

The results are demonstrated in Table 1. Additionally, the R2 is 0.757, the modified R2 is 0.745. F Statistics (constant model): 62.7 and p-value equal to 8.31e-35. The number of samples is 128. All of these results are indicated in Table 2.

Table 1. The equation statistics and the test statistics

	Estimate	SE	T-Stat	P-value
(Intercept)	11.9510	7.0895	1.6857	0.0944
X1(PM2.5)	0.7076	0.1469	4.8185	4.2388e-06
X2(PM10)	0.0733	0.1158	0.6331	0.5278
X3(CO)	15.8270	7.1586	2.2109	0.0289
X4(NO2)	-0.2219	0.1457	-1.5225	0.1305
X5(SO2)	0.1495	0.2717	0.5500	0.5833
X6(O3)	0.2990	0.0353	8.4701	6.8527e-14

Table 2. The general statistics of the multiple-linear regression model

R2	0.757
Modified R2	0.745
F statics	62.7
P-value	8.31e-35

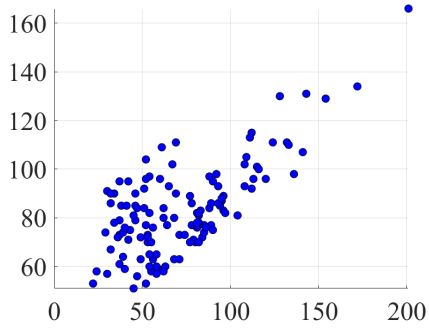

Fig. 4 The image illustrates the liner relationship between PM10 and AQI (cited from: original)

2.4 Model's Test and Modification

Table 2 has implies that X1, X2, X3, X5, and X6 are all greater than 0, however, X4 is less than 0. The former denotes the positive correlation, while the latter denotes the negative correlation. All of the statistics above imply that when each unit increased in NO2 concentration, the corresponding index decreased in AQI concentration. Vice versa, each unit increased in PM2.5, PM10, SO2, O3, and CO concentrations, and the corresponding index increased in AQI concentration. Then, the X1, X2, X3, X5, and X6 fit the environmental theory and possess economic meaning. It find that X4 betrays the environmental theory and lacks economic significance, which needs to be modified.

What's more, X2, X4, and X5 are non-significant since the P-value of these is less than 0.05. Because the P-values of X1, X6 are more than 0.05, then both of the two are significant. In addition, as Figure 5 has shown there is a linear relationship between the variables X2 and X4, which seems to be the reason for the non-significant of X2, X4, and X5.

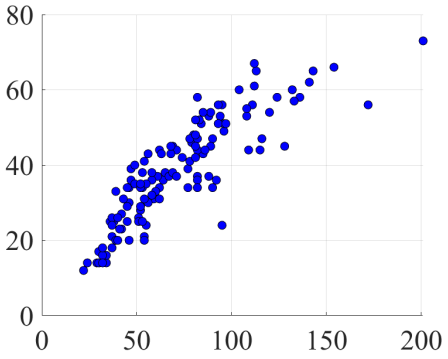

Fig. 5 The dependency between non-significant variable PM10 and NO2 (cited from: original)

2.5 Goodness of Fit Test

As this essay referred to before, the R2 is 0.757 and the modified R2 is 0.745, which implies that the goodness of fit has reached 0.75. This data is generally good enough since the degree of explanation of the variables to the whole population reached 75%.

3. Results and Discussion

From 2014 to 2024, as the essay referred to before, it has the results as follows.

PM2.5 and NO2 are mainly distributed between 20 to 60, PM10 is mainly located from 40 to 100, O3 is mainly spread from 40 to 140, CO is mainly distributed from 0.5 to 1 and SO2 is spread from 5 to 15, which indicates Nanjing's corresponding measures on controlling air pollution exceed success in some extent during the past few years. The above statistics demonstrate the general situation of the AQI.

The AQI index mainly remains in the range of "generally good". The result indicates the effectiveness of recent years' measures. For instance, the government of Nanjing has notified the <Implementation Plan of Nanjing air quality continuous improvement action plan>, spur the green transformation of industry and energy structure, and so forth.

According to equation (2), the PM2.5, PM10, SO2, O3, and CO display a positive correlation, while the NO2 behaves a negative correlation. This implies that if PM2.5, PM10, SO2, O3, and CO's density increased, the AQI decreased, and vice versa.

X2, X4, X5 are non-significant since the P-value of these is less than 0.05. The non-significant of X2, X4, X5 seems due to the linear relationship among themselves. However, X1, X6 are significant. As a result, it's paramount to pay attention to the density of PM2.5 and O3. Nanjing has deployed various measures to help reduce PM2.5 and O3. For example, control industrial emissions, Energy structure adjustment, control dust pollution, and strengthen the controls of volatile organic compounds.

The distribution images of NO2, CO, and O3 seem more like a normal distribution with different mean values, additionally, the distribution images of PM2.5, PM10, and SO2 seem to be similarly positive skewness distribution graphs.

According to the results, this essay notified the mainly significant variable, data distribution, general circumstance of the model equation, and so forth. However, this essay still has some drawbacks. For example, this essay just completed the general analysis of the six variables and found out the main significant factors. However, this essay lacks an analysis of the spatial and temporal distribution of the pollutants due to the monthly data without specific regions. That leads to this essay's lack of pertinence in this realm. For instance, a group of researchers drew graphs that are related to the spatial and temporal distribution [10]. That made the content more persuasive. By analyzing the more specific data may yield concise results and more specific measures.

4. Conclusion

Concluding from the results, Table 1 and equation (2), O3 plays a vital role in influencing the AQI, and PM10, NO2, and SO2 are non-significant. According to Nanjing's air pollution circumstances, this essay urges us to offer suggestions as follows. Firstly, enlarge the realm of the vegetation. As this essay has referred to before the large area of vegetation can alleviate air pollution to some extent. For example, the flora can absorb part of VOCs and then help alleviate the O3. Secondly, reinforce the prevention of incomplete incendiary activities. Since CO has become the largest factor that leads to air pollution, then according to the CO producing reason, it better reinforces the prevention of incomplete incendiary activities. For example, limit the vehicle license plate to reduce the emission of vehicle exhaust, make precise policies to prevent the occurrence of forest fires, improve the technical of domestic gas equipment, and so forth. Thirdly, using the air filter to help reduce the air pollution during the high-polluted weather. Fourthly, increase the use of public transport.

ISSN 2959-6157

Relevant councils can encourage the populace to use more public transport. Raise citizens' awareness of protecting the environment. Only if the majority of the populace learns deeply about the significance of protection, all of the measures can be propelled faster and more effectively.

5. References

[1] Wang Z, Huang X, Ding A. Dome effect of black carbon and its key influencing factors: a one-dimensional modelling study. Atmos. Chem. Phys., 2018, 18: 2821-2834.

[2] Guo Q H, Chen K. Analysis of ambient air quality in Nanjing. Journal of Nanjing University of Information Science and Technology (Natural Science Edition), 2022, 14(03): 294-303.

[3] Xie M, Zhu K G, Wang T J, et al. Temporal characterization and regional contribution to O3 and NOx at an urban and a suburban site in Nanjing, China. Science of the Total Environment, 2016, 551/552: 533-545.

[4] Wei G, Yang Y, Li R, Liu Y, He B-J. Digital economy exhibits varying degrees of mitigation of air pollution in China: Total cities-economic subdivisions-urban agglomerations. iScience, 2024, 27(6): 110091.

[5] Wang W. Application study of Air Quality Health Index in Nanjing based on functional time series analysis. D, Nanjing University of Finance and Economics, 2023.

[6] Huang J, Pan X, Guo X, Li G. Health impact of China's Air Pollution Prevention and Control Action Plan: an analysis of national air quality monitoring and mortality data. Lancet Planet. Health, 2018, 2: e313-e323.

[7] Zhang Q, Zheng Y, Tong D, Shao M, Wang S, Zhang Y, Xu X, Wang J, He H, Liu W, Ding Y, Lei Y, Li J, Wang Z, Zhang X, Wang Y, Cheng J, Liu Y, Shi Q, Yan L, Geng G, Hong C, Li M, Liu F, Zheng B, Cao J, Ding A, Gao J, Fu Q, Huo J, Liu B, Liu Z, Yang F, He K, Hao J. Drivers of improved PM2.5 air quality in China from 2013 to 2017. Proc. Natl. Acad. Sci. USA, 2019, 116: 24463-24469.

[8] Chen G, Li S, Knibbs L D, Hamm N A S, Cao W, Li T, Guo J, Ren H, Abramson M J, Guo Y. A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information. Sci. Total Environ., 2018, 636: 52-60.

[9] Wu H-W, Kumar P, Cao S-J. The role of roadside green infrastructure in improving air quality in and around elderly care centres in Nanjing, China. Atmospheric Environment, 2024, 332: 120607. ISSN 1352-2310.

[10] Huang Z, Kong S, Seo J, Yan Y, Cheng Y, Yao L, Wang Y, Zhao T, Harrison R M. Achievements and challenges in improving air quality in China: Analysis of the long-term trends from 2014 to 2022. Environment International, 2024, 183: 108361.