A Comparison of Air Pollution in Developed and Developing Cities: A Case Study of London and Beijing

S. X. Sun, Q. Wang

***Abstract*—**With the rapid development of industrialization, countries in different stages of development in the world have gradually begun to pay attention to the impact of air pollution on health and the environment. Air control in developed countries is an effective reference for air control in developing countries. Artificial intelligence and other technologies also play a positive role in the prediction of air pollution. By comparing the annual changes of pollution in London and Beijing, this paper concludes that the pollution in developed cities is relatively low and stable, while the pollution in Beijing is relatively heavy and unstable, but is clearly improving. In addition, by analyzing the changes of major pollutants in Beijing in the past eight years, it is concluded that all pollutants except O3 show a significant downward trend. In addition, all pollutants except O3 have certain correlation. For example, PM10 and PM2.5 have the greatest influence on air quality index (AQI). Python, which is commonly used by artificial intelligence, is used as the main software to establish two models, support vector machine (SVM) and linear regression. By comparing the two models under the same conditions, it is concluded that SVM has higher accuracy in pollution prediction. The results of this study provide valuable reference for pollution control and prediction in developing countries.

***Keywords*—**Air pollution, particulate matter, AQI, correlation coefficient, air pollution prediction.

I. Introduction

AIR pollution causes seven million deaths a year worldwide and is the biggest environmental risk to human health today. A third of respiratory illnesses and a quarter of heart disease deaths are caused by air pollutants [1]. In addition, air pollution is the fifth most attributable risk factor for the global burden of disease. Up to 91% of the population live in areas with air pollutant concentrations higher than those recommended by the WHO [2]

The main pollutant in the air is particulate matter. Particulate matter is all non-gas in the air and is composed of a variety of chemical compounds and materials, some of which may be toxic. Because many particles of travel particulate matter are small, some of these toxins may enter the bloodstream and travel around the body, settling in the heart, brain and other organs. Therefore, exposure to PM can have serious health effects, especially for the elderly, the children, and vulnerable groups with respiratory problems. And these particles can travel great distances in the air [3].

Xie designed a survey to explore the association between long-term $PM_{2.5}$ exposure and health outcomes of Chinese people of reproductive age between 20 and 49 years old. The investigation results show that long-term exposure to $PM_{2.5}$ in the environment of people of childbearing age increases the risk of hypertension, tachycardia, systolic blood pressure, diastolic blood pressure and resting heart rate. In addition, about 2.3% of hypertension cases in the study population could be attributed to PM2.5 exposure, and about 4% of tachycardia cases could be attributed to PM2.5 exposure [4].

With the continuous development of industrialization, the problem of air pollution has become more and more serious. Countries around the world have gradually realized the harm of air pollution and have introduced relevant policies and measures to deal with air pollution.

II. Air Pollution Situation and Control in Beijing and London

A. Air Pollution in London

Britain is the first industrialized country in the world, and also the first country to encounter environmental pollution events. Faced with such serious consequences, the British government introduced a series of strict measures and regulations to deal with the haze problem. Air control is divided into three stages:

1) The early stage of strict legislation to curb industrial pollution. A series of laws such as the Clean Air Act have been introduced [5].
2) The middle stage of all-round control of domestic emissions. The British government has taken measures to limit pollution, such as promoting unleaded gasoline, imposing strict controls on pollutants from car exhaust, and charging traffic congestion.
3) The later treatment as industrial upgrading and pollution output. Britain has begun to vigorously develop the service industry and high-tech industry [6].

B. Air Pollution in Beijing

Domestic and industrial pollution are the main sources of pollution in China [7]. Therefore, China has taken steps on several fronts. In law, the air pollution Prevention and Control action plan has been issued to set targets for emission reduction.

Shixian Sun is with Department of Engineering, Durham University, Durham, DH1 3LE UK (corresponding author, e-mail: shixian.sun@durham.ac.uk).

Qing Wang is with Department of Engineering, Durham University, Durham, DH1 3LE UK (e-mail: qing.wang@durham.ac.uk).

In terms of environmental taxes and fees, environmental protection taxes are levied on enterprises that discharge pollution to reduce pollution emissions and improve environmental quality [8]. In addition, in terms of motor vehicles, many cities have encouraged new energy vehicles, natural gas buses to reduce exhaust emissions [9].

III. Related Work

By comparing the difference of air pollution between dry and hot Makkah in Saudi Arabia and wet and cold Leeds in U.K., Habeebullah et al. [10] obtained the difference of correlation between various pollutants under different climate conditions and the influence of climate on pollution.

Hao [11] proposed a pollutant prediction model based on a BP(back propagation) neural network. Based on the variation rule of pollutant concentration, he proposed that the pollutant concentration of the previous day could be used to predict the pollutant concentration of the next day. By adding a meteorological factor to improve the prediction method, the prediction result has high accuracy. Magdin et al. [12] used linear regression and other methods to analyze vehicle exhaust noise and pollution and calculated that there is a strong correlation between them. Through Pearson correlation analysis, Oji et al. [13] analyzed the correlation between different meteorological factors and air pollutants, indicating that meteorological factors have a strong impact on air pollution. Karimian et al. [14] use the hybrid model based on long short-term memory to forecast PM2.5 at different times, and the results have high accuracy. Jeya et al. [15] used a bidirectional long short-term memory model to predict air pollution, producing highly accurate models.

IV. Research Objects and Methods

This paper mainly studies the pollution of five major pollutants, PM2.5, PM10, SO_2, CO and O_3, and the changes before and after treatment, and analyzes the pollution situation by comparing the data of Beijing and London, who are at two different stages of development. Software such as Python and Excel have been used for the data analysis.

A. The Main Sources

The sources of air pollution are extremely wide ranging. Pollution caused by natural disasters is mostly temporary and local, while pollution caused by human activities is regular and large-scale. The main types of pollution sources caused by human factors are industrial pollution source, domestic pollution source and transportation pollution [16]. Moreover, there are many types of air pollutants, and there are about 100 air pollutants that have been identified. Air pollutants in different periods and in different regions vary [17]. The forms of pollutants can be divided into aerosol pollutants and gaseous pollutants. The formation process can be divided into primary pollutants and secondary pollutants. If the air pollutant is the original material discharged directly from the pollution source, and its behavior does not change after entering the air, it is called the primary pollutant, for example, CO_2, NO_2. If a series of changes or photochemical reactions occur between the primary pollutants discharged from the pollution source and the original components in the air, or between several primary pollutants, forming new pollutants with different properties from the original pollutants, then the new pollutants formed are called secondary pollutants. Examples of these are sulfuric acid smoke, photochemical smoke and haze [18], [19]. In particular, the rapid increase of coal-based energy consumption and motor vehicle ownership has led to the emission of carbon monoxide (CO), carbon dioxide (CO_2), sulfur dioxide (SO_2), methane (CH_4), nitrogen oxide (NOx), ozone (O_3), and atmospheric particulate matter (PM2.5, PM10) into the atmosphere, seriously affecting city air quality [20]. Moreover, there are three main sources of air pollutants: industrial pollution, such as coal-fired power plants and various types of factories, and exhaust gases and soot emitted by burning fuels; traffic pollution sources, such as cars, planes, trains and other vehicles, emit emissions containing carbon monoxide, nitrogen oxides, hydrocarbons and so on; and domestic pollution sources, such as lampblack, cold air leakage refrigerant, etc. [21].

In this research, six major pollutants, PM2.5, PM10, CO, SO_2, O_3 and NO_2 will be selected as the research objects. AQI is also formulated with these six pollutants as a reference. Table I shows the indicators of AQI corresponding to different pollution levels [22].

TABLE I
The Level of Pollution Represented by the AQI

301-500	Hazardous
201-300	Very Unhealthy
151-200	Unhealthy
101-150	Unhealthy for Sensitive Groups
51-100	Moderate
0-50	Good

B. Research Method

The experimental data were provided by Beijing and London's official monitoring stations, which provide data on PM2.5, PM10, SO_2, CO, O_3 and NO_2, the main pollutants that pose a threat to human health. The study will use Python and Excel to analyze the data, such as comparing the pollution data of London and Beijing and creating a comparison map to compare the correlation between pollutants in Beijing and analyze the changes of Beijing's air pollutants before and after treatment, to provide valuable reference for future treatment direction.

Through a regression equation, the relationship between different pollutants can be understood and analyzed, so as to evaluate pollution in the two cities. Pearson correlation analysis is a statistical method to analyze the relationship between variables. The correlation coefficient is a measure of the degree of linear correlation between the variables. Data can be predicted by the linear regression analysis of variables through the correlation coefficient. The closer the correlation coefficient is to 1, the higher the degree of linear correlation between variables. The closer the correlation coefficient is to -1, the lower the degree of linear correlation between variables. When the correlation coefficient is 0, it indicates that there is no linear correlation between variables. Therefore, the correlation

coefficient can be used to describe the correlation between PM2.5 and other pollutants [23]. The correlation coefficient calculation formula is:

$$r(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2(y_i-\bar{y})^2}} \quad (1)$$

where r: correlation coefficient; n: sample size; x: random variable (monitoring indicators of PM10, O_3, NO_2, SO_2, CO); y: random variable (monitoring indicators of PM2.5); $x_i y_i$: individual sample points indexed with i.

This study uses Python as a data processing tool because Python has powerful functions and can be used to write almost all types of applications. In addition, one of the features of Python programming is easy access to mature libraries, which greatly improves the efficiency of programming. Among them, the Numpy package is the data processing package that can quickly deal with the array. Pandas is a data analysis and modeling kit that provides intelligence for data alignment and flexibility for handling lost data. Matplotlib is a Python-based 2D graph library that can generate a variety of graphs, such as histograms, bar charts, scatter plots, pie charts, box plots, etc., with a few lines of code [24].

A multi-linear regression model and a SVM model will be used as air pollution prediction models, and the most accurate will be selected by comparing the accuracy data of the two models. Multiple linear regression models are usually used to describe random linear relationships between variables Y and X, the formula is as follows:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon \quad (2)$$

where $x_1, \ldots, x_k$: nonrandom variable; y: random dependent variable; $\beta_0, \ldots, \beta_k$: regression coefficient; ε: random error term.

SVM are learning algorithms that use classification and regression to analyze data. Given a set of training samples, each training sample is marked as belonging to one or the other of the two categories. It can be used to classify and regression data.

The mathematical expression of SVM is:

$$\begin{aligned} w^T x_i + b > 0,\ y_i = 1 \\ w^T x_i + b < 0,\ y_i = -1 \end{aligned} \quad (3)$$

Fig. 1 shows a schematic diagram of the SVM model, which shows the relationship between coefficients and variables.

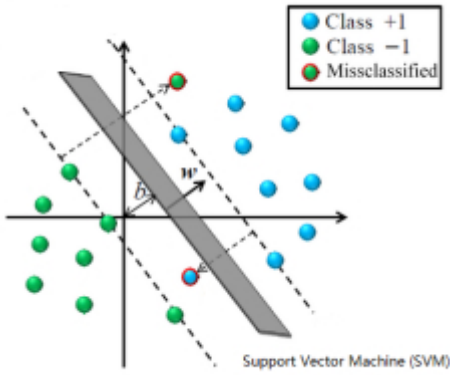

Fig. 1 SVM model

In this research, the regression equation will be used for data regression analysis and Python will be used as a tool to calculate correlations between pollutants, make heat maps and run the forecast models.

V. Results and Discussion

This paper compares the pollution indexes of Beijing and London, and changes in pollutant data in Beijing in the process of continuous control.

A. Variation of Pollutants in Beijing

Fig. 2 shows the variation of pollutants in Beijing from December 2013 to November 2021. It can be seen from these data that all data pollutants are constantly fluctuating and gradually decreasing. Among them, AQI, PM2.5, PM10, CO and NO_2 increased in 2016 and then continued to decline. SO_2 has been declining every year and has remained at an extremely low level below 10 since April 2017. O_3 fluctuates every year, which may be related to the season, showing the characteristics of more summer and less winter. To sum up the above data, the following conclusions can be drawn: China's air control effect is very significant, and all the data have shown a stable and continuous decline.

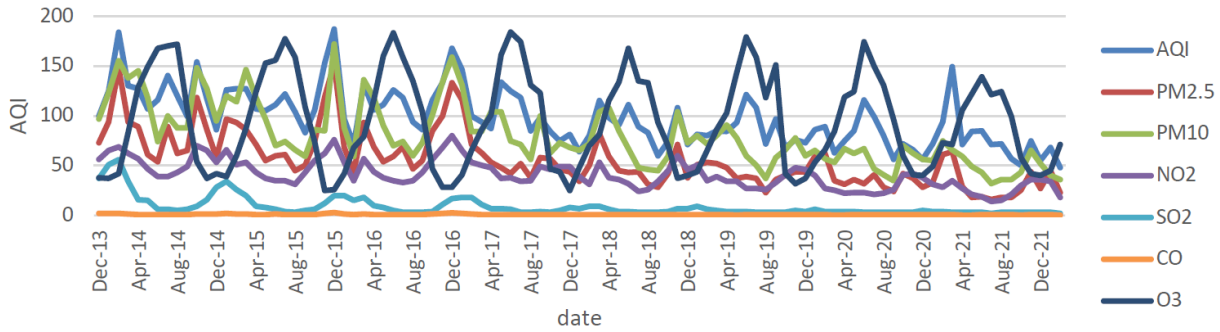

Fig. 2 Changes of pollutants in Beijing before and after treatment

B. Correlations between Different Pollutants

Table II and Fig. 3 show the correlation of various air pollution indexes in Beijing from 2013 to 2021 calculated using Python and the thermal map of their correlation. The correlation was calculated using the Pearson correlation analysis method. It can be clearly seen from Fig. 3 that AQI is highly correlated with PM10, CO and PM2.5, while AQI is relatively correlated with SO_2, reaching 0.47. O_3 is negatively correlated with all data. PM10 is highly correlated with CO, NO_2 and PM2.5. SO_2 is relatively correlated with other pollution data except O_3, which are all around 0.5. CO is highly correlated with PM10, AQI, NO_2 and PM2.5. NO_2 was highly correlated with AQI, PM10, CO and PM2.5. PM2.5 is highly correlated with other pollution data except O_3. Among them, PM2.5 and PM10 have the greatest influence on AQI, reaching 0.91 and 0.82 respectively. O_3 has the lowest correlation and influence on AQI and other pollution data.

TABLE II
CORRELATION BETWEEN DIFFERENT POLLUTANTS

	AQI	O_3_8h	PM10	SO_2	CO	NO_2	PM2.5
AQI	1	0.18	0.82	0.47	0.75	0.67	0.91
O_3_8h	0.18	1	-0.03	-0.27	-0.29	-0.32	-0.08
PM10	0.82	-0.03	1	0.53	0.71	0.72	0.84
SO_2	0.47	-0.27	0.53	1	0.63	0.62	0.54
CO	0.75	-0.29	0.71	0.63	1	0.80	0.85
NO_2	0.67	-0.32	0.72	0.62	0.80	1	0.77
PM2.5	0.91	-0.08	0.84	0.54	0.85	0.77	1

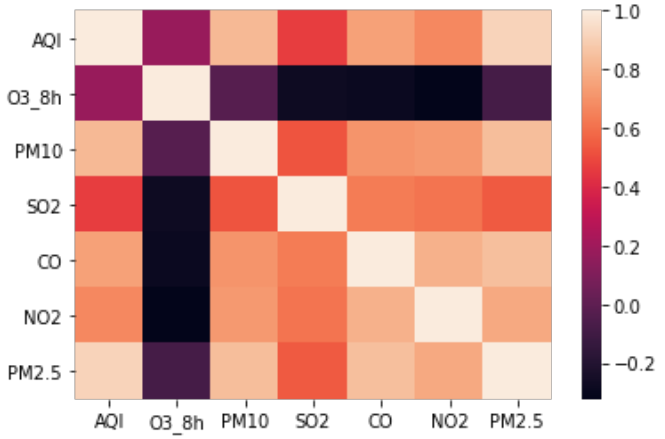

Fig. 3 Pollutant correlation thermal map

C. Comparison of PM2.5 between London and Beijing

Fig. 4 shows the comparison of PM2.5 between London and Beijing. The results show that the PM2.5 pollution index in LONDON IS RELATIVELY STABLE AND STAYS AT A LOW LEVEL FOR A LONG time, while the pollution index in Beijing fluctuates considerably. Although there may be consecutive good days, there are still many days when the pollution index exceeds 200 AQI.

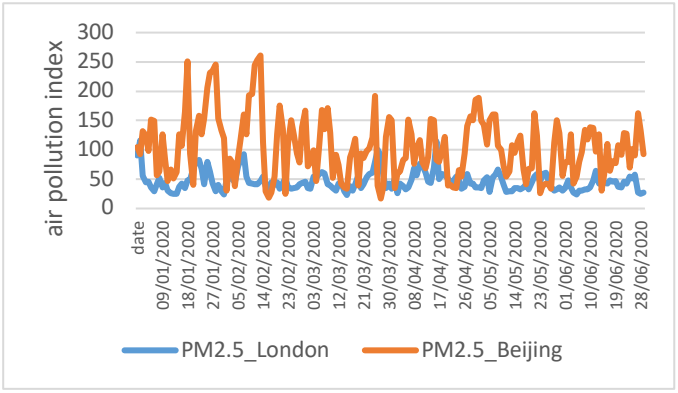

Fig. 4 comparison of PM2.5 between London and Beijing

D. Comparison of the monthly number of excellent weather days in Beijing and London

Figure 5 shows a comparison of the monthly number of excellent weather days in Beijing and London from 2014 to 2021. The abscissa is the date, the ordinate is the number of days. As can be seen from the chart, the number of good days in London each month is at a very high level and remains stable. The number of good days in Beijing fluctuates greatly, and there are fewer good days per month. However, the number of good days in Beijing has been increasing gradually, even by nearly half in the last two years. In July 2021, it reached an excellent level for the whole month, on par with London.

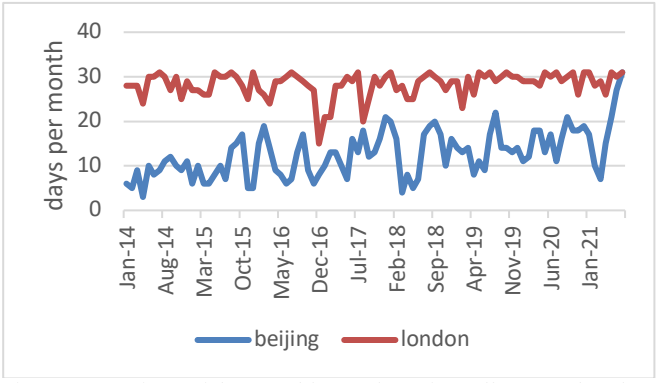

Fig. 5 Comparison of the monthly number of excellent weather days

E. Comparison between Multiple Linear regression results and SVM results

The prediction model uses Python to predict the air quality level of the day based on the historical weather and air quality data of Beijing and the weather and air quality conditions of the previous two days. The model uses two methods of multiple linear regression prediction and SVM classification prediction as a comparison and chooses the better algorithm. Experimental process is as follows, firstly, air pollution and weather data of Beijing in the past ten years are collected, and then the weather conditions such as sunny, cloudy, cloudy, rain, snow, fog, sand are coded. Then air quality such as excellent, good, light pollution and moderate pollution, and high levels of pollution, serious pollution were coded. Then, the training data and testing data were divided into 7:3. The linear regression model and SVM model were each used for training and the respective results were obtained. Table 3 shows the results of the multiple

linear regression prediction model and SVM model. By comparison, the SVM results are more accurate.

TABLE III THE RESULTS OF THE MULTIPLE LINEAR REGRESSION PREDICTION MODEL AND SVM MODEL

	Multiple linear regression results	SVM results
Accurate rate	0.868	0.94
Precision ratio	0.874	0.94
Recall rate	0.868	0.94

VI. CONCLUSIONS

Different countries have different development periods, but they have experienced rapid economic development, and at the same time, they are also facing a series of air pollution problems. As the earliest industrialized country, Britain has left valuable reference experience on air control. China has gradually begun to pay attention to the issue of air pollution. Comparison with the UK's air pollution can provide valuable reference and help for China's pollution control. The comparison shows that there is still a gap between Beijing's air pollution and that of London, and pollution levels are not stable. In addition, by analyzing the changes of various pollutants before and after the control, it can be found that the pollution level in China, represented by Beijing, is decreasing year by year. Through continuous control, the air quality may reach the level of developed countries in the future. In addition, the prediction model of air pollution based on a large amount of historical data achieves ideal accuracy.

REFERENCES

[1] UN, “Air pollution, which kills an estimated 7 million people every year, is the biggest environmental health risk of our time.,” 2 2022. [Online]. Available: https://www.unep.org/explore-topics/air/why-does-air-matter.

[2] WHO, "New WHO Global Air Quality Guidelines aim to save millions of lives from air pollution," 22 9 2021. [Online]. Available: https://www.who.int/news/item/22-09-2021-new-who-global-air-quality-guidelines-aim-to-save-millions-of-lives-from-air-pollution. [Accessed 10 1 2022].

[3] D. f. E. F. a. R. Affairs, "Emissions of air pollutants in the UK – Particulate matter (PM10 and PM2.5)," 18 2 2022. [Online]. Available: https://www.gov.uk/government/statistics/emissions-of-air-pollutants/emissions-of-air-pollutants-in-the-uk-particulate-matter-pm10-and-pm25. [Accessed 2 1 2022].

[4] X. Xie, Long-term effects of ambient fine particulate matter on blood pressure and heart rate and attributable risk among adults aged 20-49 years in China, 2019.

[5] J. Martinez, “Great Smog of London,” 2 1 2022. [Online]. Available: https://www.britannica.com/topic/rock-London-1960s-overview-1371200. [Accessed 2 2 2022].

[6] L. J. Zhang X, From London to Beijing: Comparison and reflection on smog management between China and UK., 2014.

[7] S. Y. Z. L. Zhang H, Pollution control in urban China: A multi-level analysis on household and industrial pollution, Science of the Total Environment, 2020.

[8] W. Y. Z. B. Cai S, "The impact of the air pollution prevention and control action plan on PM2.5 concentrations in jing jin ji region during 2012-2020," Science of the total environment, 2017, pp. 197-209.

[9] T. F. Y. G. Qin Y, "Challenges of using natural gas as a carbon mitigation option in China," Energy Police, 2018, pp. 457-462.

[10] T. M. Habeebullah, “Acomparison of air quality in Arid and temperate climatic conditions- A case study of Leeds and Makkah,” *World academy of Science, engineering and technology international journal of enbironmental and ecological engineering,* pp. 750-758, 6 9 2015.

[11] J. Hao, research on air pollution prediction technology based on neural networks, Nanjing University, 2014.

[12] V. M. Kirill Magdin, "Correlation between Noise and air pollution from car sources," *international conference on developments in eSystems engineering,* no. 2019.00151, pp. 812-816, 2019.

[13] H. A. Sunday Oji, "Correlation between air pollutants concentration and meteorological factors on seasonal air quality variation," *Journal of air pollution and health,* vol. 5, no. 1, pp. 11-32, 2020.

[14] Q. L. Hamed Karimian, "evaluation of different machine learning approaches to fourecasting PM2.5 mass concentrations," *aerosol and air quality research,* vol. 19, pp. 1400-1410, 2019.

[15] L. S. S. Jeya, "air pollution prediction by deep learning model," *International conference on intelligent computing and control systems,* no. 2020.9120932, pp. 736-741, 2020.

[16] M. D. L. Liu J, "Air pollution emissions from Chinese households: Amajor and underappreciated ambient pollution source," Academy of Sciences, 2016, pp. 7756-7761.

[17] F. X, "Emission inventory of primary pollutants and chemical speciation in 2010 for the yangze river delta region," Atmospheric Environment, 2013, pp. 39-50.

[18] C. L, "Difference between primary pollutants and secondary pollutants," 12 4 2018. [Online]. Available: http://www.differencebetween.net/science/difference-between-primary-pollutants-and-secondary-pollutants/. [Accessed 10 2 2022].

[19] X. F, "Emission inventory of primary pollutants and chemical speciation," Atmospheric Environment, 2013, pp. 39-50.

[20] Z. G, "Exploring the inconsistent variations in atmospheric promary and secondary pollutants during the 2016 G20 summit in Hangzhou , China: implications from observations and models," Atmospheric Chemistry and Physics, 2020, pp. 5391-5403.

[21] S. O. P, "Effect of industrial pollution on residential neighbourhood: Amuwo Odofin Industrial layout Lagos as case study," Covenant journal of research in the built environment , 2017, p. 5.

[22] R. Z, "Let's discuss AQI," 9 7 2020. [Online]. Available: https://csu-ceams.com/blog/lets-discuss-aqi/. [Accessed 15 1 2022].

[23] Y. Sun, "Correlation analysis and temporal and spatial changes in a dispersion model of PM2.5 in UK cities," Durham University, 2020.

[24] W. Wentao, "Data analysis of PM2.5 and study on green economic efficiency of cities in China based on Python," China University of Petroleum, 2017.