An analysis and investigation of the factors affecting the air quality index in various cities in China

Xiaotian Jiang
Jilin University, Jilin, China
jiangxt1021@mails.jlu.edu.cn

Abstract. With the rapid progress of society, air pollution has gradually become the focus of peoples concern, the impact of air pollution is very many factors, including but not limited to natural factors, human factors, geographic factors, etc., this paper is mainly from the natural factors and human factors, to explore its impact on air quality in the past five years, this paper first from the data itself, summarizes the distribution characteristics of the air quality in different regions of China, and then from the temperature and humidity throughout the year, to quantify its impact, and then with the help of time-series methods, to explore the air pollution indicators in the case of de-seasonalization of the air pollution indicators. This paper firstly summarizes the distribution characteristics of air quality in different regions of China from the data itself, then quantifies the impact of temperature and humidity throughout the year, and then explores whether the air pollution indicators satisfy the smooth distribution under the situation of de-seasonalization with the help of time series correlation method. The paper then compares the effects of several anthropogenic factors on the air through the gray correlation method. Finally, the influence of each factor in different parts of the country is summarized to provide reference for future studies.

Keywords: air pollution index; Person correlation coefficient; Run test; Grey correlation degree.

1. Introduction

With the advancement of national industrialization and urbanization, the national economy is in a state of rapid development, but at the same time, the impact of industrialization on the environment has become a new focus of attention. The rapid development of industry not only provides convenience for peoples lives, but also has a significant impact on the surrounding environment. Published in Lancet Star Healthhealthy[1] A study on "Ambient Air Pollution Causes About 4.5 Million Deaths in 2019, Up from 4.2 Million in 2015 and 2.9 Million in 2000" reported that air pollution has become one of the major contributors to premature deaths around the worldone[2] At the same time, air pollution in China has become a national problem. Based on the air pollution situation, this paper provides a series of explanations on the air quality situation in different regions of China.

2. Sources of data collection and related concepts

Relevant definitions of air pollution index: Air Quality Index (AQI) is a single index that uses the numerical value of air pollution level to report the air quality status. The AQI calculation formula is as follows:

$$\mathrm{AQI} = ([\mathrm{C_i}/\mathrm{S_i}] \times \mathrm{L_i}) \times 100$$

Among them. $\mathrm{C_i}$ Indicates the concentration of the specified pollutant. $\mathrm{S_i}$ Indicates the primary criterion value for the concentration of the specified pollutant. $\mathrm{L_i}$ Air Quality Index (AQI) coefficients for specified pollutants. In particular, the urban air pollution index has a strict classification from one to six, as follows.

AQI	AQI level	AQI catagory	Representation color	Impact on health	Recommended actions
0~50	Level 1	Excellent	Green	The air is satisfactory and	All kinds of people can move normally

				basically free of air pollution	
51~100	Level 2	Good	Yellow	The air is acceptable, but some pollutants may have a weak impact on the health of a very small number of abnormal people	Very small number of people with unusual sensitivities should reduce outdoor activities
101~150	Level 3	Mild pollution	Orange	Symptoms of susceptible people are slightly aggravated, and symptoms of healthy people appear	Children, the elderly, and people with heart and respiratory diseases should reduce long-term, high-intensity outdoor exercise
151~200	Level 4	Moderate pollution	Red	Further aggravating the symptoms of susceptible people may have an impact on the heart and respiratory system of healthy people	Children, the elderly, and patients with heart disease and respiratory diseases should reduce long-term, high-intensity outdoor exercise, and the general population should reduce outdoor exercise moderately
201~300	Level 5	Heavy pollution	Purple	Symptoms of patients with heart and lung diseases are significantly aggravated. Exercise tolerance is common in healthy people	Children, the elderly, and people with heart and lung disease should stay indoors and stop outdoor exercise, and the general population should reduce outdoor exercise
>300	Level 6	Serious pollution	Maroon	Healthy people have obvious strong symptoms of exercise tolerance and some diseases appear in advance	Children, the elderly and patients should stay indoors to avoid physical exertion. The general population should avoid outdoor activities

Because of the large spatial span and long time span of collecting air quality data from all cities in the country, it is difficult to analyze each city, this paper mainly collects the more representative cities in each region of the country to be analyzed, and this paper roughly divides the country into seven regions and choose seven cities, which are from East China, South China, North China, Southwest China, Northwest China, Northeast China, and Central China,. The monthly AQIs for the past 5 years (2019-2023) are compared horizontally, and the factors affecting the air quality are further explored.

3. Specific analysis of factors

3.1 Data analysis methods

First, calculate some basic statistics of the sample, including mean, variance, IQR, kurtosis, and biasdegree[3].

Here the skewness$\beta_S$with the skewness$\beta_k$is calculated as follows.

$$\beta_S = \frac{E(X-E(X))^3}{[Var(X)]^{\frac{3}{2}}} \qquad \beta_k = \frac{E(X-E(X))^4}{[Var(X)]^2} - 3$$

All the results are obtained in the following table.

Name	mean ± standard deviation	Variance	IQR	Kurtosis	Skewness	Coefficient of variation (CV)

Beijing	79.800±19.312	372.942	22.5	1.765	1.049	24.20%
Harbin	71.450±29.039	843.269	34.75	4.914	1.741	40.64%
Lanzhou	83.717±18.003	324.105	16.75	12.547	2.593	21.51%
Wuhan	78.600±15.263	232.956	20	0.316	0.834	19.42%
Chengdu	76.533±15.914	253.27	19.75	-0.551	0.163	20.79%
Hangzhou	71.717±11.539	133.156	17.5	-0.473	0.097	16.09%
Guangzhou	65.217±13.625	185.63	16	0.944	0.816	20.89%

By comparing the above data, it can be concluded that under the criterion of average value, there is not much difference between the data of the northern and southern cities, but comparing the variance, kurtosis and skewness, it can be found that the air pollution index of the northern cities (e.g., Beijing, Harbin, Lanzhou in the figure) has a larger variation, compared to the north, the air pollution index of the southern places (e.g., Wuhan, Chengdu, Hangzhou, Guangzhou in the figure) is more stable.

3.2 Influence of natural factors on air pollution indices

The previous article analyzed the overall situation of air quality in various regions of the country, and the following consideration was given to the impact of natural factors on air quality, first of all the temperature factor. In order to describe their corresponding correlation, this paper introduces the concept of Pearson correlation coefficient to conduct a quantitative analysis of their correlation degree. The calculation formula is as follows:

$$r = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}}$$

The value range of r is - 1 to 1. The closer to 1, the stronger positive linear correlation exists between the two variables, the closer to - 1, the stronger negative linear correlation exists between the two variables, and the closer to 0, the weaker or no obvious linear correlation exists between the two variables. The calculation results are shown in the following table:

	Beijing	Harbin	Lanzhou	Wuhan	Chengdu	Hangzhou	Guangzhou
Average high temperature for the month	0.3707	-0.6309	-0.1554	-0.2834	-0.1286	0.2265	0.2819
average low temperature for the month	0.3455	-0.6688	-0.2068	-0.3333	-0.2132	0.2081	0.1446
average monthly humidity	-0.1361	-0.5073	-0.5358	-0.1033	-0.0540	0.0233	-0.0120

By observing the results of the final data, it can be found that the cities in the north (such as Harbin and Lanzhou) have a greater impact on the air pollution index, temperature and humidity. This effect shows the law that the more south, the less impact. But in the south, where the temperature is high all year round, people consume more electricity and other resources in summer, which will lead to a change trend of air pollution index that is positively related to temperature. In addition, the south is humid all year round, so the impact of humidity on the air is relatively insignificant. In contrast, for example, in Beijing, the impact of its temperature on it is more similar to the trend in the south, because Beijing has a large population and a large population density, so the impact of human factors diluted the impact of natural factors, resulting in some deviation in the results.

The following time-series partitioning of theanalyse[4] (taking Beijing as an example), to see whether the AQI satisfies the smooth distribution after removing the seasonal factors.

For time series decomposition models.

$$X_t = T_t + S_t + R_t$$

where $\{T_t\}$ is the trend term.$\{S_t\}$Its a seasonal term, $\{R_t\}$Is a random term. When the known seasonal cycle is m, the jth serial value of the ith cycle is recorded as$S_{i,j}$Then there is

$$S_t = \bar{x} \cdot I_j$$

Among them. I_j is called the seasonal index, which is calculated as

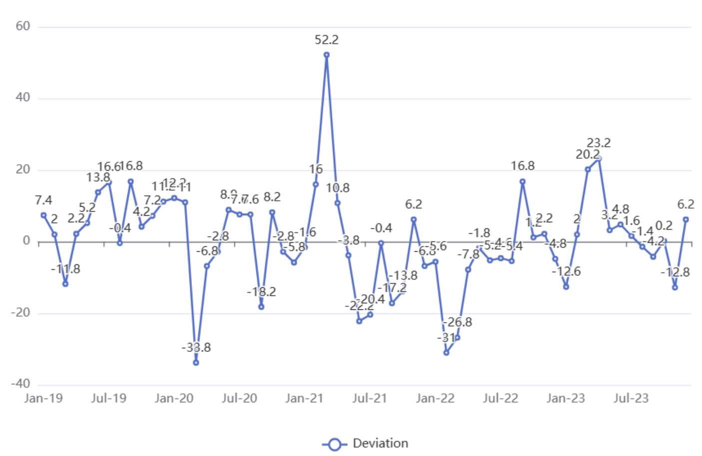
$$I_j = \frac{\overline{X_J}}{\bar{X}} \qquad j=1,2,...m,$$

Among them.

$\overline{X_J} = \frac{1}{n}\sum_{i=1}^{n} X_{i,j}$, j=1,2,...m, ; $\bar{X} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} X_{i,j}$.

Calculations were made by substituting the corresponding data, and the results, as shown in the figure below, show that there is no longer a clear seasonal effect.

Time sequence chart of Beijing air index after de-seasoning



The number of occurrences of the positive sign in the table is then counted by means of the travel testN_1, the number of times the negative sign appearsN_2, the number of its tripsN_r,the mean and variance of the sequence are, respectively, the

$$E(N_r) = \frac{2N_1N_2}{N} + 1\,; \qquad \mathrm{Var}(N_r) = \frac{2N_1N_2(2N_1N_2 - N)}{N^2(N-1)}$$

and the sample statistic, the

$$Z = \frac{N_r - E(N_r)}{\sqrt{\mathrm{Var}(N_r)}} \sim \mathrm{N}(0,1)$$

Taking the statistics from this paper, theN_1=31， $N_2 = 29$， N_r=21 is substituted into the formula, and the calculation results show that at the significance level of 0.005, the assumption that the series is stationary is accepted. That is, after de seasoning, the air quality index series has no obvious trend.

3.3 Anthropogenic influences on AQI

This paper considers the impact of natural factors on air pollution index, while human factors also have a greater impact on air pollution. The change of air pollution index may also be related to human behavior. However, there are many human factors that cause air pollution. This paper starts from three parts: per capita GDP, urban car ownership and annual per capita carbon emissions, To analyze which factor has the most impact on the air pollution index big[5] . This article will use grey correlationlaw[6] to calculate the results of this three-factor comparison. The calculations are as follows.

Firstly, the relevant data of the given factors were found through the China Statistical Yearbook.

	Beijing	Harbin	Lanzhou	Wuhan	Hangzhou	Chengdu	Guangzhou
Per capita GDP (10000 yuan/year)	19.03	5.38	7.6	13.78	15.26	9.81	15.36
Vehicle ownership (10,000)	622.4	232	119.8	403.9	360.1	589.6	331
annual per capita carbon emissions	0.44	0.32	0.24	0.21	0.24	0.27	0.3

air quality index	79.8	71.45	83.72	78.6	71.72	76.53	65.22

Among them, per capita GDP, car ownership, and per capita carbon emissions are independent variable series, and the corresponding air quality index is dependent variable series. In order to make the different factors comparable, the dependent variable series and independent variable series are first dimensionally eliminated. The dimensionality elimination method used in this paper is to use each item of data to be in its mean value to calculate the results, that is, the following formula is used:

$$x^*(k) = \frac{x(k)}{\bar{x}} \qquad \bar{x} = \sum_{k=1}^{n} x(k)$$

The results are as follows.

	Beijing	Harbin	Lanzhou	Wuhan	Hangzhou	Chengdu	Guangzhou
GDP per capita	1.545	0.4368	0.617	1.1188	1.2389	0.7965	1.247
automobile ownership	1.6386	0.6108	0.3154	1.0634	0.948	1.5523	0.8714
annual per capita carbon emissions	1.5246	1.1088	0.8316	0.7277	0.8316	0.9356	1.0395
air quality	1.0599	0.949	1.112	1.0439	0.9526	1.0165	0.8662

Further, substituting this result into Eq

$$\alpha_i(k) = \frac{min_k|x_0^*(k) - x_i^*(k)| + \rho.max_k|x_0^*(k) - x_i^*(k)|}{|x_0^*(k) - x_i^*(k)| + \rho.max_k|x_0^*(k) - x_i^*(k)|}$$

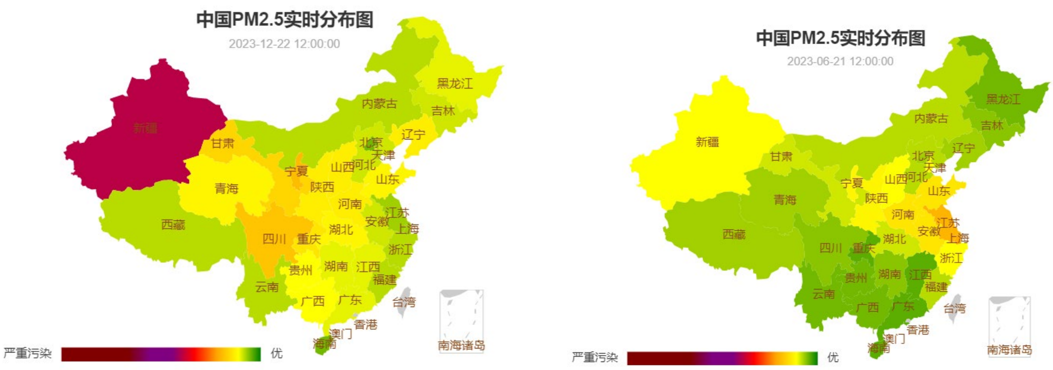
where, in Eq,ρIs a coefficient to control the coefficient of differentiation, the general calculation of the value of 0.5 is more appropriate, this paper will also be taken as 0.5, the results of each item are as follows.

	GDP per capita	automobile ownership	annual per capita carbon emissions
Calculations	0.604685714	0.656385714	0.6929

It can be seen from this that among the above three anthropogenic factors, the annual per capita carbon emissions have the greatest impact on the air quality index, followed by the car ownership, and finally the per capita GDP, which is consistent with the expected results at the beginning.

4. Conclusions

There are many factors affecting the AQI, which can be categorized as anthropogenic and natural factors, and these two factors have different impacts depending on the topography, climate, and level of economic development of each region.



(The left figure shows the real-time distribution of P.M2.5 at the winter solstice in 2023, and the right figure shows the real-time distribution of P.M2.5 at the summer solstice in 2023)

In conjunction with the conclusions reached in the previous section, air quality varies in different areas of the country at different times of the yearbe the same[7]The air quality index in summer is generally better than that in winter, and the air quality index in regions with high population density is generally lower than that in regions with low population density. Moreover, the impact is more significant in the north, and the air quality index caused by different geographical locations also has corresponding differences. Generally speaking, the air quality index in the north has a large range of change, and because the north is dry compared with the south, especially the impact of the natural environment in the northwest, the average air quality index is high, and the range of change is large, However, the air quality index of southwest China is relatively low due to its good natural environment and relatively small population. The natural environment in central China is good, but the population density is high, and the air pollution is relatively serious, leading to the general air quality. In South China and West China, due to the coastal areas in most areas, the temperature is at a relatively average level throughout the year, and the climate is humid, which is conducive to the dispersion of pollutants. Therefore, even though these areas have a large population, their air quality index remains at a low standard due to the rapid volatilization of pollutants.

References

[1] Richard F,JPL,KalpanaB ,etal.Pollution and health: a progress update.[J]. The Lancet.Planetary health,2022,6(6):e535-e547.

[2] Jitkajornwanich K ,Vijaranakul N ,Jaiyen S , et al. Enhancing risk communication and environmental crisis management through satellite imagery and AI for air quality index estimation [J]. MethodsX, 2024, 12 102611-.

[3] Mao Shisong, Cheng Yiming. Course of Probability Theory and Mathematical Statistics [M]. Beijing: Higher Education Press, 2011

[4] Zhou Yongdao, Wang Huiqi, Lv Wangyong. Time series analysis and application [M]. Beijing: Higher Education Press, September 2015

[5] Xu Ying. Spatial statistical analysis of air pollution in China and research on influencing factors [D]. Jiangxi University of Finance and Economics, 2016

[6] Wang Juan Research on calibration of air quality data based on grey correlation degree method [J] Microcomputer Applications, 2021, 37 (03): 44-47

[7] Bao Zhenhu, Liu Tao, Luo Jihua, etc Analysis on spatial-temporal distribution characteristics of ambient air quality in China [J] Geographic Information World, 2014, 21 (06): 17-21