Correlation Analysis between Meteorological Factors and Pollutants Based on Copula Theory

Qian Wang* and Xiuhua Li

Wenshang Ecological and Environmental Affairs Center of Jining City, Jining, China
Email: wsxhbjzljb@126.com.

Abstract. In recent years, the continuous acceleration of industrial and agricultural production has aggravated the deterioration of the environment and seriously affected people’s health. Gas pollutant index has become an important indicator to measure air quality. Taking a region of China as the study subject, this paper firstly uses the monitoring data of air pollutants in 5 years and meteorological data (rainfall, temperature and wind speed) of meteorological observation stations in the same period. Then, the autocorrelation function is selected to analyze the autocorrelation of pollutants, and on this basis, the significant correlation level between meteorological factors and pollutants is analyzed by Copula theory. The results show that the ACF values of pollutants are all higher than 0.6, which means that the autocorrelation of pollutants within 2 hours is relatively significant. PM$_{2.5}$ has a significant negative correlation with wind speed in four seasons, but a low correlation with rainfall. SO$_2$ is negatively correlated with most of the three meteorological factors. In addition, the study also finds that there are significant differences in the impact of meteorological factors on pollutants at different time scales. This paper proposes a correlation analysis method between meteorological factors and pollutants, which provides corresponding support and guarantee for China’s economic and environmental protection policies.

Keywords. Correlation analysis; meteorological factors; pollutants; Copula theory.

1. Introduction
Since entering the 21st century, the world economy has developed rapidly and the industrial structure has been constantly adjusted. However, at the same time, due to imperfect control measures and other reasons, the random discharge of various pollutants has aggravated the environmental pollution, especially the air pollution [1]. In the process of rapid social development in recent years, many countries must deal with the core issues of economic development and environmental protection when they are faced with economic planning and environmental policy formulation [2-3]. In this paper, air pollutants are taken as the research object. Because a variety of air pollutants exist at the same time with a high concentration, complex interactions occur, resulting in the formation of atmospheric complex pollution and haze and other phenomena, there may be a certain degree of autocorrelation [4]. At the same time, the environmental interaction of the earth system also makes pollutants and meteorological factors closely related, and the concentration of air pollutants in the same time period also shows a certain correlation with meteorological factors. With the continuous improvement of modern environmental monitoring system, real-time monitoring of pollutants in most key regions has been realized, which provides a lot of valuable data for this study [5-6]. This paper also combines with the meteorological monitoring data website corresponding to the meteorological factor data period, and then carry out correlation analysis.
The aggravation of environmental pollution has attracted extensive attention. Many scholars have conducted in-depth research on the relationship between air pollution and meteorological factors. These articles analyse the characteristics of air pollutants in 14 locations in Egypt, and makes a correlation between air pollutants and meteorology [7-8]. In order to further grasp the correlation model between meteorological factors and pollutants [9], this paper uses the observation database of daily meteorological elements and pollutants in the past 10 years [10]. Through detailed comparison, it is proved that their long-term cross-correlation behavior in rural areas is more obvious than that in urban areas. In addition, the influence of meteorological factors on the multifractal characteristics of pollutants is also studied [11]. Some scholars have studied the relationship between ozone concentration and meteorological factors, and find that there is an inverse correlation between atmospheric pressure tide and ozone concentration, and the long-distance transportation of favorable meteorological conditions (i.e., low relative humidity, high temperature and solar radiation, zero rainfall) will increase the ozone concentration [12-13]. Zhou et al. [14] showed the characteristics of PM$_{2.5}$ and O$_3$ atmospheric composition monitoring data and meteorological data, and used the empirical orthogonal function method to classify the types of PM$_{2.5}$ and O$_3$ pollution in winter and summer in China from 2015 to 2019. The results show that there is a significant positive correlation between the distribution of extreme pollution weather and the distribution of meteorological field [15].

To sum up, most studies focus on the correlation analysis between meteorological factors and pollutants, but do not study the autocorrelation characteristics, and deeply explore the correlation relationship within the time period. First of all, this paper uses the pollutant data of meteorological monitoring stations, takes gaseous pollutants (PM$_{2.5}$ and SO$_2$) as the research object, and combines the meteorological factor data of the same period to carry out data filtering. Secondly, the autocorrelation analysis of air pollutants is carried out to grasp their own change characteristics. Finally, a correlation analysis method between air pollutants and meteorological factors based on Copula theory is proposed and verified by simulation.

2. Correlation Analysis Method for Meteorological Factors and Pollutant

2.1. Autocorrelation Analysis Method Based on Pearson Theory

Based on previous studies, Autocorrelation Function (ACF) is selected in this paper to analyze the autocorrelation of variables [16], and its formula is shown as follows:

$$\rho = \frac{E[(X_t - \mu)(X_{t+\lambda} - \mu)]}{\sigma^2}$$

where $\rho$ represents the autocorrelation coefficient; $X_t$ represents the sample data of variable X at time $t$; $X_{t+\lambda}$ represents the sample data lagging behind $X_t$ and with a lag time $\lambda$; $\mu$ and $\sigma$ represent the mean and standard deviation of the sample sequence, respectively.

2.2. Correlation Analysis Method Based on Copula Theory

Nowadays, with the advancement of Copulas theory research, it can more accurately describe the correlation between variables. Compared with the traditional linear correlation coefficient of Pearson it has more excellent results on the relationship between various nonlinear. The Copula function can construct the joint distribution function without obtaining the specific form of the edge distribution of each study variable [17].

Let the edge distribution function of multiple random variables be $F_1(x_1), F_2(x_2), F_3(x_3), \ldots, F_N(x_N)$ and its joint distribution function is $F(x_1, x_2, x_3, \ldots, x_N)$. When $F_1(x_1), F_2(x_2), F_3(x_3), \ldots, F_N(x_N)$ are continuous, there exists a single deterministic Copula function that satisfies:

$$F(x_1, x_2, x_3, \ldots, x_N) = C(F_1(x_1), F_2(x_2), F_3(x_3), \ldots, F_N(x_N))$$

where $C$ is a Copula function.
For the random variables of meteorological factors and pollutants in this paper, when $X-F(X)$, $Y-G(Y)$, and corresponding Copula function $C(u, v)$ exists, let $u = F(x)$, $v = G(y)$, $u, v \in [0, 1]$, Kendall rank correlation coefficient $\tau$ can be expressed as:

$$\tau = 4\int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 \quad (3)$$

The Kendall rank correlation coefficient $\tau$ is uniquely determined by the Copula function theory, and then the correlation between meteorological factor $X$ and pollutant $Y$ is calculated. However, there are many kinds of Copula functions, and the rank correlation coefficients obtained by different Copula functions are different. Therefore, the accuracy of correlation measure depends on the goodness of fit of selected Copula functions.

The tail correlation coefficient can be expressed by the Copula function:

$$\lambda = \lim_{\alpha \to 1} P\left\{ Y > G^{-1}(\alpha) \mid X > F^{-1}(\alpha) \right\} = \lim_{\alpha \to 1} \frac{C(1-\alpha, 1-\alpha)}{1-\alpha} \quad (4)$$

3. **Statistical Variation Law of Meteorological Factors and Pollutants Data**

In this paper, a county in Jining City, Shandong Province, China is selected as the research area, and a total of 5 environmental monitoring stations are selected. The monitoring frequency is once an hour, and the 24-hour continuous monitoring lasted from January 1, 2016 to December 31, 2020. The research object of air pollutants includes $\text{SO}_2$ and fine particulate matter $\text{PM}_{2.5}$, and the mean value represents the monthly concentration.

The monitoring data of meteorological factors come from the Website of China Meteorological Administration, and the time interval and resolution correspond to the environmental monitoring data one by one, mainly including rainfall, temperature and wind speed.

3.1. **Statistical Variation Law of Meteorological Factors**

Meteorological factors are used to reflect the function of atmospheric system, which can reflect the trend of climate change and the impact scale of pollutants in this region to a certain extent. In this section, based on the monitoring data from environmental monitoring stations, the curve of average concentration change of meteorological factors in this area in five years is obtained, as shown in figures 1 and 2.
As can be seen from figures 1 and 2, in the statistical data in the past five years, the monthly average temperature presents a single-peak distribution characteristic, with the peak appearing in July and August and the highest reaching 25°C, and the lowest in January and December. However, the wind speed varies little from season to season. The annual change has a certain time series trend.

3.2. Statistical Variation Law of Pollutants

Through the monitoring of the monitoring sites for five consecutive years, the monitored mass concentration of PM$_{2.5}$ and SO$_2$ showed certain statistical rules between years, as shown in figures 3 and 4.

![Figure 3. Annual variation of average concentrations of PM$_{2.5}$.](image)

![Figure 4. Annual variation of average concentrations of SO$_2$.](image)

It is found that the annual variation of the average concentration of different pollutants is different. PM$_{2.5}$ and SO$_2$ showed a single-peak distribution in the year, with the peak value mostly appearing in January and February, and the trough appearing from June to August.

4. Case Study

4.1. Autocorrelation Analysis

The autocorrelation analysis of PM$_{2.5}$ and SO$_2$ is carried out by using equation (1). The ACF values of the two pollutants under different lag times can be obtained from figures 5 and 6. The ACF values of the two pollutants decreased slowly with the growth of lag time, showing typical ACF trailing characteristics, so the concentration of PM$_{2.5}$ and SO$_2$ has obvious autocorrelation. Meanwhile, when the lag time is less than 2 hours, the minimum ACF value is still greater than 0.6, and in the first half hour, the ACF value is greater than 0.7, indicating that there is a strong autocorrelation between air pollutants in each period in the first hour.
4.2. Correlation Analysis
After the previous analysis, the research area of atmospheric pollutants is obvious since the correlation, and closely related to pollutant concentration and the seasonal change, although also has certain relation to the intensity of industrial and agricultural production, but the influence of meteorological factor on air quality is much significant, and the pollution sources in the same season changes are also smaller. Therefore, when studying the correlation between pollutants and meteorological factors, this section chooses season as the scale to further significantly study the cross-correlation relationship.

Table 1. Correlation between PM$_{2.5}$ and meteorological factors in different seasons.

|       | Temperature | Rainfall | Wind speed |
|-------|-------------|----------|------------|
| Spring| 0.53        | 0.15     | -0.62      |
| Summer| 0.37        | 0.08     | -0.57      |
| Autumn| -0.12       | -0.05    | -0.38      |
| winter| 0.06        | -0.02    | -0.45      |

As can be seen from table 1, PM$_{2.5}$ concentration and meteorological factors show significant correlation differences in four seasons. PM$_{2.5}$ has a positive correlation with air temperature in spring and summer, but it is not obvious, and the correlation is weak in autumn and winter. The correlation between PM$_{2.5}$ concentration and rainfall is weak, and the correlation coefficient is generally lower than 0.15. PM$_{2.5}$ and wind speed show a strong negative correlation in a year. From the perspective of weather system, when wind speed is higher, there is less cloud cover and air flow accelerate, which is conducive to the diffusion of PM$_{2.5}$.

Table 2. Correlation between SO$_2$ and meteorological factors in different seasons.

|       | Temperature | Rainfall | Wind speed |
|-------|-------------|----------|------------|
| Spring| -0.12       | -0.11    | -0.24      |
| Summer| 0.37        | -0.12    | -0.19      |
| Autumn| -0.23       | -0.04    | -0.32      |
| winter| -0.16       | -0.07    | -0.28      |
As can be seen from table 2, the monthly average concentration of SO$_2$ is negatively correlated with air temperature, rainfall and wind speed in most cases, which are more obvious in summer and autumn. This is because when the area is controlled by low pressure, the higher the temperature, the faster the wind speed and the more rain, the pollutants are easy to diffuse under the control of such pressure field, which can quickly reduce the concentration of pollutants and improve the air quality.

5. Conclusion
Based on the concentration of PM$_{2.5}$ and SO$_2$ pollutants and the meteorological factors (temperature, rainfall and wind speed) of environmental monitoring stations, and the autocorrelation analysis of pollutants by the Principle of Person coefficient, this paper uses Copula theory to conduct correlation analysis with meteorological factors of the same period, and draws the following conclusions:

1) By calculating the values of ACF, it is found that the ACF curves of PM$_{2.5}$ and SO$_2$ have obvious trailing characteristics, and the values are generally greater than 0.6, highlighting the strong autocorrelation within the time period.

2) The correlation coefficients of two kinds of pollutants and three kinds of meteorological factors in different seasons are calculated and analyzed. PM$_{2.5}$ has a significant negative correlation with wind speed within a year, but has a low correlation with precipitation and a significant seasonal difference with temperature. SO$_2$ showed negative correlation with most of the three meteorological factors, and the correlation is strong.

References
[1] Chan C K and Yao X 2008 Air pollution in mega cities in China Atmospheric Environment 42 (1) 1-42
[2] Richman M B and Lamb P J 1985 Climatic pattern analysis of three and seven-day summer rainfall in the Central United States: Some methodological considerations and a regionalization Journal of Applied Meteorology 24 (12) 1325-1343
[3] Bolton R 1989 Integrating economic and environmental models: Some preliminary considerations Socio-Economic Planning Sciences 23 (1-2) 25-37
[4] Qian L, Song J, Wang E, Hu H, Zhang J and Wang Y 2014 Economic growth and pollutant emissions in China: A spatial econometric analysis Stochastic Environmental Research and Risk Assessment 28 (2) 429-442
[5] Tang Y P, Chen B and Zhang D N 2008 Establishment of Nanjing modern environmental monitoring system The Administration and Technique of Environmental Monitoring 20 (3) 6-8
[6] Wittkowski F, Schöpfer C and Batternmann G 1997 Environmental Monitoring Supported by a Modern Data Management System Combining Standard Software with High Performance Databases (Netherlands: Springer)
[7] Elminir H K 2007 Relative influence of weather conditions and air pollutants on solar radiation—Part 2: Modification of solar radiation over urban and rural sites Meteorology & Atmospheric Physics 96 (3-4) 257-264
[8] El-Nouby M A 2013 Suspended particulates concentration (PM$_{10}$) under unstable atmospheric conditions over subtropical urban area (Qena, Egypt) Advances in Meteorology 2013 (7) 1903-1912
[9] Chithra V S and Nagendra S M S 2014 Impact of outdoor meteorology on indoor PM$_{10}$, PM$_{2.5}$ and PM$_1$ concentrations in a naturally ventilated classroom Urban Climate 10 77-91
[10] Aguilar-Velázquez D and Reyes-Ramírez I 2018 A wavelet analysis of multiday extreme ozone and its precursors in Mexico city during 2015-2016 Atmospheric Environment 188 112-119
[11] He H 2017 Multifractal analysis of interactive patterns between meteorological factors and pollutants in urban and rural areas Atmospheric Environment 149 47-54
[12] Toh Y Y, Lim S F and Von Glasow R 2013 The influence of meteorological factors and biomass burning on surface ozone concentrations at Tanah Rata, Malaysia *Atmospheric Environment* **70** 435-446

[13] Voulgarakis A and Field R D 2015 Fire influences on atmospheric composition, air quality and climate *Current Pollution Reports* **1** (2) 70-81

[14] Zhou L, Jing Z, Lu T, Bao M, Deng X and Hu X 2021 Pollution patterns and their meteorological analysis all over China *Atmospheric Environment* **246** 118108

[15] Guo H, Lin H, Zhang W, Deng C, Wang H, Zhang Q, Shen Y and Wang X 2014 Influence of meteorological factors on the atmospheric mercury measurement by a novel passive sampler *Atmospheric Environment* **97** 310-315

[16] Passarelli R E and Siggia A D 1983 The autocorrelation function and doppler spectral moments: Geometric and asymptotic interpretations *Journal of Applied Meteorology* **22** (10) 1776-1787

[17] Nelsen R B 2006 *An Introduction to Copulas* (New York: Springer)