Correlation Analysis of City Air Quality in Henan Province

Kai-guang ZHANG*, Ming-ting BA, Hong-ling MENG, Yan-min SUN and Huan-cheng SHI
Zhengzhou Normal University, Zhengzhou 450044, China
*Corresponding author

Keywords: Air quality index, Cluster analysis, Correlation coefficient, Spatial distribution, Henan

Abstract. Based on the daily average data of city air quality index in Henan province in 2017, utilizing geostatistical analysis and spatial analysis methods, this paper studies the correlations of air quality among cities, the provincial distribution characteristics of air quality, and the spatial relationship between air quality and topography, explores the evolution mechanism and internal mechanism of air quality in the province. The research shows that: City air quality in the province has obvious spatial agglomeration characteristics, the air quality index shows continuous block distribution. The local air quality index is inversely proportional to the topography elevation, and the overall difference of the air quality index is large. There are clear interactions between cities, the air quality indices between adjacent cities have strong spatial correlations, affected by the terrain and distance between cities. There are two independent air circulation systems and five air quality correlation belts, closely related to the topographic distribution in the province.

Introduction

In 21st century, China's economic development and urbanization have been promoting the rapid accumulation of material wealth and the substantial improvement of the people’s living standards, at the same time, large-scale infrastructure construction, over-consumption of energy, low efficiency of agricultural by-products and rapid expansion of the traffic scale make the air pollution is becoming more and more serious. Air quality has become one of the main environmental problems affecting population health. Air quality problem is attracting more and more scholars’ attention [1-2]. In recent years, the study of air quality based on geoscience vision has mainly focused on the variation characteristics analysis of air quality at different scales and in typical regions, the influencing factors analysis of air quality, and has achieved many valuable results [3-9]. In fact, although the amount of atmospheric pollutant emissions is the main factor determining regional air quality, the regional topographic conditions, wind direction, wind speed, precipitation and temperature have some impact on the spread of pollutants. Regional transmission and transport is one of the important features of air pollution, the air quality of a region is affected by the pollution conditions of adjacent regions, and there is a certain correlation between regions [2, 6].

Based on the impact of various pollutants on human health, ecology and environment, according to environmental air quality standard, the Air Quality Index (AQI) simplifies the monitored main air pollutant concentrations into a single conceptual index, classifies air quality as 0~500, the larger the index, the more serious the pollution situation and the greater harm to the human body [1, 7-10].

This paper utilizes the daily average data of city AQI in Henan province in 2017, and geostatistical analysis and spatial analysis methods to study the correlation of city air quality, the spatial distribution characteristics of air quality, the spatial relationship between air quality and terrain, explores their evolution mechanism and internal mechanism, in order to provide some scientific reference for the design of regional air pollution prevention and control measures.
Data and Research Methods

Research Area Overview and Data. Henan province (31°23'N-36°22'N, 110°21'E-116°39'E) is located in the transition zone of China's terrain from the second to the third ladder, made up of 18 cities. Its three sides as the north, west and south are semi-circular surrounded by Taihang, Funiu and Tabie mountains[11], its central and east regions are Huanghuaihai alluvial plain. the data used in the study mainly includes 30m×30m DEM provided by Computer Network Information Center (http://www.gscloud.cn), the air quality data, mainly includes the 155-day (from 2017.6.30-2017.12.30) daily average air quality index of 18 cities, derived from the key city air quality daily report data of the Ministry of Environmental Protection of the People's Republic of China (http://datacenter.mep.gov.cn).

Clustering Analysis

Clustering analysis is a spatial statistical analysis method that classifies the population into different groups according to the similarity degree of observed value and a certain criteria. a group is a collection of similar elements, the intra-group difference is much smaller than the difference between groups [12]. City air quality index K-mean clustering analysis method, first, defines the distance between elements as

$$d_{ij} = \sqrt{\sum_{t=1}^{n}(x_{it} - x_{jt})^2}$$

and the criterion function as

$$W = \sum_{t=1}^{k} \sum_{i=1}^{n} (X_{it} - \bar{X}_t)^T (X_{it} - \bar{X}_t),$$

according to the prior knowledge, determines the number of classifications $k$ and the initial clustering center $\bar{X}_0$; then the remaining samples are allocated to the corresponding group of each cluster center according to the principle of minimum distance, and the new clustering center and criterion function are calculated until the change of the criterion function is fitted in the allowable range. At the same time as the classification, the final cluster center and the distance from the element to the center are obtained.

Here, $x_{it}$ is the observed value of the city $i$ in day $t$, and $n_t$ and $\bar{X}_t$ respectively are the number of cities in group $t$ and the cluster center of group $t$.

Correlation Coefficient

The correlation coefficient is a statistical indicator used to describe the close relationship between observed values, by the sum of multiplied the dispersions of two variables [13].

Suppose $(x_i, y_i), \ i = 1, 2, 3, \ldots, n$ is the observations of two-dimensional random variable $(X, Y)$, and $\bar{x}, \bar{y}, D(X)$ and $D(Y)$ respectively are the mean and variance of $X$ and $Y$, the correlation coefficient of $X$ and $Y$ is defined as [14]

$$C = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{D(X)D(Y)}}$$

$C$ belongs to $[0, 1]$, the greater $C$ means strong relevance, as $C = 1$ means $X$ and $Y$ is linearly dependence, $C = 0$ means $X$ and $Y$ is mutual independence.
Distribution Characteristics Analysis of Air Quality Index in Henan Province

Topographic Distribution Characteristics Analysis in Henan Province

Fig.1a and Fig.1b respectively are the elevation DEM of Henan province and its 250m elevation contour interpolation results. The topography of the whole province shows west high and east and middle low, three-ladder features from the east to the west. The first ladder is a plain region with the elevation of less than 250m, distributed in the east of 113°30'E including eastern Anyang, Puyang, eastern Xinxiang, Kaifeng, Shangqiu, Zhoukou, eastern Xuchang, Luohe, eastern Zhumadian and northern Xinyang; The second ladder is a half mountain region[11], with the elevation of 250-500m, located between 112°30'E and 113°30'E, including Jiaozuo, northeastern Luoyang, western Xinxiang, Zhengzhou, Pingdingshan, western Xuchang, western Zhumadian, southeastern Nanyang and western Xinyang; The third ladder is a mountain region, with the elevation above 500m (a standard mountain height), located in the west of 112°30'E, including Jiyuan and Sanmenxia. Southwestern Luoyang and northeastern Nanyang.

Figure 1. The topographic distribution characteristics in Henan province.

Clustering Analysis of Air Quality Index in Henan Province

Carrying out K-mean clustering analysis on the AQI of the 155-day, using Euclidean distance (Eq.1) and the squares sum of the inter-class deviations (Eq.2) as the distance between samples and the criterion function, the results are shown in Fig. 2a. The first group is a heavily polluted region represented by Xinxiang, with the weighted average AQI 121.64, and the air quality pollution days account for 43.14% (Fig.2b, air quality level days distribution), including Anyang, Hebi, Puyang, Jiaozuo, Zhengzhou, Kaifeng and Xuchang. The second group is a lightly polluted region represented by Luoyang, with the weighted average AQI 103.59, and the air quality pollution days account for 37.91%, including Jiyuan, Sanmenxia and Pingdingshan. The third group is a good air quality region represented by Zhumadian, with the weighted average AQI 95.92, and the air quality pollution days account for 20.91%, including Shangqiu, Zhoukou, Luohe, Nanyang and Xinyang.

Figure 2. The clustering analysis of air quality index in Henan province.
In the view of spatial distribution, air quality has obvious regional characteristics, the 3 groups are distributed in continuous blocks. The first group is mainly concentrated in the middle and northern of the province, the AQI in the group is significantly different with the variance of 6.71. The AQIs of Anyang and Puyang, the northernmost city are the largest, followed by Zhengzhou and Kaifeng in the central part. The geometric center of the group is located at (114°11'03.98997''E, 35°11'17.76282''N), the AQI weighted center is (114°07'55.73142''E, 35°09'03.93208''N), indicating that the pollutants in the southern part of the group are higher than that in other regions. The second group is mainly located in the western region of the province. The difference of AQI within the group is small, its variance is only 1.96. The AQI in the eastern region is slightly higher than that in the western region. The third group is mainly distributed in the eastern and southern regions of the province, the AQI in the group is different with the variance of 5.25, the AQI of Shangqiu and Luohe are significantly higher than that of the other cities.

Considering the topographical distribution characteristics (Fig.1b), and the weights of AQI in calculating, it can be seen that the second group is mainly distributed on the third elevation ladder, the first group is mainly distributed in the north of the first and second elevation ladders, the third group is mainly distributed in the south of the first and second elevation ladders.

**Correlation Analysis of Air Quality Index in Henan Province**

Based on the 155 days air quality index, using Eq.3 to calculate the correlation coefficient between cities, the results are showed in Fig.3. There are great correlations between the inter-city in the province, the mean is of 0.807, the maximum value is of 0.985 (between Luoyang and Jiyuan), and the minimum value is of 0.599 (between Hebi and Xinyang). For the convenience of analysis, this paper divides them into 4 levels, namely weak correlation \(0.70 \leq C < 0.90\), basic correlation \(0.90 \leq C < 1.00\), correlation \(0.80 \leq C < 0.90\) and strong correlation \(0.90 \leq C < 1.00\), the correlation coefficients between cities are shown in Fig. 3.

**Anyang (Fig.3a):** Its strongly related cities are adjacent Hebi, Puyang and Xinxiang, the strong correlation region shows west high and east low, mainly located in the north of the first ladder. Its overall air quality is the worst in the region, with the weighted average of 121.64, is the main pollution source in the region, impacting on the reduction of air quality in other cities. Its correlation region coincides with the first group of clustering analysis, is basic correlation with other cities except for Xinyang.

**Hebi (Fig.3b):** Beside of adjacent Anyang and Xinxiang, its strongly related cities also includes Zhengzhou and Puyang. The strong correlation region shows west high and east low, mainly located in the north of the first and second elevation ladder. Its overall air quality is the best in the region, with the weighted average of 100.36. The correlation region includes Jiaozuo, Jiyuan, Pingdingshan, Xuchang, Luohe and Kaifeng, is weak correlation with Xinyang and Sanmenxia.

**Puyang (Fig.3c):** Located in the northeastern part of the province, bordering with Hebei province and Shandong province. The cities with strong correlations include its adjacent Anyang and Xinxiang, also include Kaifeng which is located on the first elevation ladder as it, its air quality index weighted average is 112.15, its air quality reduction is affected by Anyang. The correlation region includes Jiaozuo, Jiyuan, Zhengzhou, Xuchang, Zhoukou and Shangqiu, and is weak correlation with Xinyang and Sanmenxia.

**Xinxiang (Fig.3d):** Beside of its adjacent Anyang, Hebi, Jiaozuo and Puyang, its strongly related cities also include Zhengzhou and Kaifeng. The strong correlation region shows west high and east low, mainly located in the north of the first and elevation ladders, its air quality weighted average is 103.05. Its air quality reduction is affected by Zhengzhou. The correlation region includes Jiyuan, Luoyang, Pingdingshan, Xuchang and Luohe, and is weak correlation with Xinyang and Sanmenxia.

**Jiyuan (Fig.3e):** its strongly related cities are Jiaozuo, Luoyang, Xinxiang and Zhengzhou, mainly located in the north of the second elevation ladder and the transition zone from the second to the third elevation ladder, the air quality index is opposite to the terrain elevation in its strong correlation region, shows east high and west low. Its correlation region includes Hebi, Xinxiang, Kaifeng,
Xuchang, Luohe, Pingdingshan and Sanmenxia, is basic correlation with other cities except for Xinyang.

Jiaozuo (Fig.3f): its strongly related cities include adjacent Xinxiang, Jiyuan, Zhengzhou and Luoyang, its strong correlation region spans three elevation ladders, where the air quality index is opposite to the terrain elevation, shows east high and west low. Its correlation region includes Anyang, Puyang, Hebi, Kaifeng, Luohe and Pingdingshan, is basic correlation with other cities except for Nanyang and Xinyang.

Zhengzhou (Fig.3g), Its strongly related cities are all located in the north, including Xinxiang, Jiaozuo, Jiyuan and Hebi, the region spans three elevation ladders, its air quality weighted average value is of 111.56, is an important pollution source in the region, impacts on the air quality reduction in other cities, as the distance increases, the impact gradually decreases. Its correlation region includes Anyang, Puyang, Kaifeng, Xuchang, Luohe, Pingdingshan and Luoyang, is basic correlation with other cities except for Xinyang.

Kaifeng (Fig.3h): Its strong correlation region shows a band distribution, including Puyang, Xinxiang and Xuchang, mainly located in the first elevation ladder, its air quality weighted average value is of 110.53. Its correlation region includes Anyang, Hebi, Jiyuan, Jiaozuo, Zhengzhou, Pingdingshan, Zhumadian and Zhoukou, is basic correlation with other cities except for Xinyang and Sanmenxia.
Figure 3. The correlation analysis of air quality index in Henan province.

Shangqiu (Fig.3i): Located on the first elevation ladder in the eastern of the province, adjacent to Anhui province and Shandong province, there is no strong correlation with other cities in the province, its air quality index weighted average is of is 108.84. Its correlation region includes Puyang, Kaifeng, Xuchang, Luohe, Zhoukou and Nanyang, is weak correlated with Sanmenxia, Luoyang and Xinyang, and basic correlation with the remaining cities.

Sanmenxia (Fig.3j): Located on the third elevation ladder in the western of the province, adjacent to Shaanxi and Shanxi province. There is no strong correlation with other cities in the province. Its correlation region includes Luoyang and Jiyuan, its basic correlation region includes Anyang,
Jiaozuo, Zhengzhou, Xuchang, Pingdingshan, Luohe and Zhumadian, there are weak correlation with the remaining 8 cities.

Luoyang (Fig.3k): Its strongly related cities are all located in the north, including Jiyuan and Jiaozuo, mainly located in the second elevation ladder and the transition zone from the second to the third elevation ladder, its air quality index weighted average is of is 106.16. Its correlation region includes Xinxiang, Zhengzhou, Xuchang, Luohe, Pingdingshan, Zhumadian and Sanmenxia, is basic correlation with other cities except for Xinyang and Shangqiu.

Pingdingshan (Fig.3l): Its strong correlation region includes Xuchang and Luohe, mainly located in the second elevation ladder with south high and north low, its air quality index weighted average is of is 104.46. Its correlation region includes Sanmenxia, Xinyang, Shangqiu, Puyang and Anyang, and is basic correlation with the remaining 10 cities.

Xuchang (Fig.3m): Located on the geometric center of the province, the strong correlation region includes Pingdingshan, Luohe, Xuchang and Kaifeng, mainly located in the first and second elevation ladder with west high and east low, its overall air quality is the best in the region with the weighted average 103.83, is basic correlation with Sanmenxia and Xinyang, and correlation with the remaining 12 cities.

Zhoukou (Fig.3n): Its strongly related cities are all located in the south, including Luohe and Zhumadian, mainly located on the first elevation ladder. Its air quality index weighted average of the is 105.36, and the air quality is greatly affected by its adjacent Luohe. Its correlation region includes Shangqiu, Puyang, Kaifeng, Xuchang, Pingdingshan, Nanyang and Xinyang, is weak correlation with Sanmenxia, and basic correlation with the remaining 6 cities.

Luohe (Fig.3o): Its strong correlation region includes Xuchang, Pingdingshan, Zhumadian and Zhoukou, mainly located in the first elevation ladder and the transition zone from the first to the second elevation ladder, its overall air quality is the worst in the region with the weighted average 110.04, is the main sources of pollution in the region, impacting on the reduction of air quality in other cities. Its correlation region includes Shangqiu, Zhoukou, Xuchang, Pingdingshan, Nanyang and Xinyang, is weak correlation with Sanmenxia, and basic correlation with the remaining 6 cities.

Nanyang (Fig.3p): Located in the southwestern part of the province, on the second and third elevation ladder, its air quality weighted average is of 106.44, has no strong correlation with other cities in the province. Its correlation cities all are located in the east, including Pingdingshan, Xuchang, Luohe, Zhumadian, Zhoukou, Xinyang and Shangqiu, is weak correlation with Sanmenxia and Jiaozuo, and basic correlations with Luoyang, Zhengzhou, Kaifeng, Xinxiang, Hebi, Puyang and Anyang.

Zhumadian (Fig.3q): Its strong correlation region includes Luohe, Zhoukou and Xinyang, mainly located in the first elevation ladder. The weighted average of the air quality index is of 101.92, and its air quality index is greatly affected by the adjacent city Luohe. Its correlation region includes Kaifeng, Xuchang, Pingdingshan, Luoyang and Nanyang, is basic correlation with Sanmenxia, Jiyuan, Jiaozuo, Zhengzhou, Xinxiang, Shangqiu, Hebi, Puyang and Anyang.

Xinyang (Fig.3r): Located in the southern part of the province at the first elevation ladder. The overall air quality is the best with the weighted average of 95.92. Its strong correlation region only includes Zhumadian, the correlation region includes Nanyang, Luohe and Zhoukou, the basic correlation region includes Pingdingshan and Xuchang, is weak correlation with the other 12 cities.

Overall, there are clear regional correlation and interaction between city air quality in the province, there are large differences in indices among the cities, the three cities, as Anyang, Zhengzhou and Luohe, have large impacts on the rise of air quality index of surrounding cities(Fig.3s).

The region geographical conditions largely determine the interaction of air quality between cities. the city air qualities in the province presents two independent air circulation systems and five air quality correlation belts(Fig.3t). Sanmenxia and Shangqiu are not obvious correlation with other cities, where form two independent air circulation systems in the province. The first three air quality correlation belts shows north-south trend, the first belt is located on the first elevation ladder, running through Zhumadian-Zhoukou-Kaifeng-Xinxiang-Puyang from south to north; The second is located
Conclusion

This paper, based on the daily average data of AQI in Henan province in 2017, utilizing geostatistical analysis and spatial analysis methods, studies the air quality correlation among cities, the provincial distribution characteristics of air quality, and the spatial relationship between air quality and topography, explore the evolution mechanism and internal mechanism of air quality. The results show that:

(1) City air quality in the province has obvious spatial agglomeration characteristics, the air quality index shows the heavily polluted region, the lightly polluted region, the good air quality region, etc. 3 continuous block distributions by Euclidean distance and the squares sum of the inter-class deviations as the distance between samples and the criterion function, located respectively in the central and northern part of the province, the western part of the province, the eastern and southern part of the province.

(2) The local air quality index is inversely proportional to the topography elevation, and the overall difference is large. The three heavy pollution cities, as Anyang, Zhengzhou and Luohe, have been impacting on the rise of the air quality index of surrounding cities.

(3) There are clear interactions between cities, the air quality indices between adjacent cities have strong spatial correlations, these correlations are affected by the terrain and distance between cities. There are two independent air circulation systems and five air quality correlation belts, closely related to the topographic distribution in the province.

Reference

[1] Lin Xueqin, Wang Dai. Spatio-temporal variations and socio-economic driving forces of air quality in Chinese cities. Acta Geographica Sinica, 2016, 71(8): 1357-1371.

[2] Jiang Lei, Zhou Haifeng, Bai Ling, et al. The analysis of socio-economic factors of air quality index (AQI) based on the perspective of the exponential decay effects. Acta Scientiae Circumstantiae, 2018, 38(1):390-398.

[3] Zhang Nan Nan, Wu Sun Ze, Li Yi Fan. Evaluation of the ambient air quality improvement in China based on percentile method. Acta Scientiae Circumstantiae, 2018, 38(5): 2082-2089.

[4] Wang Zhenbo, Fang Chuanglin, Xu Guang, et al. Spatial-temporal characteristics of the PM 2.5 in China in 2014. Acta Geographica Sinica, 2015, 70(11): 1720-1734.

[5] Zhang Shengling, Wang Yao, Li Yue, et al. Does regional haze opinion affect air quality? Journal of Arid Land Resources and Environment, 2018, 32 (8):100-106.

[6] Qiao X, Jaffe D, Tang Y, et al. Evaluation of air quality in Chengdu, Sichuan Basin, China: Are China’s air quality standards sufficient yet? Environ. Monit. Assess, 2015, 187(5): 1-11.

[7] Peng Jian, Chen Sha, Lü Huiling, et al. Spatiotemporal patterns of remotely sensed PM 2.5 concentration in China from 1999 to 2011. Remote Sensing of Environment, 2016, 174: 109-121.

[8] Xue An, Geng Enze. Region Division Study of PM2.5 Pollution in Cities of China Based on Complex Networks. Journal of Basic Science and Engineering, 2018, 23(Supplement):68-78.

[9] Zhang Yinjun, Chen Xi, Xie Gaodi, et al. Pollution status and spatial distribution of PM 2.5 in China. Resources Science, 2015, 37 (7):1339-1346.
[10] Ning Miao Sun Yamei Lli Wei. Study on the Method of City Air Quality Attainment Judgment. Environment and Sustainable Development, 2017, 42 (1):154-157.

[11] Zhang Kaiguang, Ba Mingting, Meng Hongling. Characteristics Analysis of Land Use Spatial Distribution on Terrain Gradient in Henan Province. 2017 3rd International Conference on Artificial Intelligence and Industrial Engineering (AIIE 2017): 96-102.

[12] Zhang Kaiguang, Ba Mingting, Meng Hongling. Spatial pattern and forecast analysis of urbanization growth rates in Henan (2011-2015). AER-advance in engineering research, 2016, 57:45-49.

[13] Chen Yanguang. Mathematical Methods for Geography. Beijing: Science Press, 2017.

[14] Guo Renzhong. Spatial Analysis. Beijing: Higher education Press, 2011.