Geographic visualization and spatial analysis of COVID-19 based on GIS

Yu Wang¹

¹ Shandong Agricultural University, Taian, Shandong, 271018, China

Abstract. By using the spatial analysis function of ArcGIS and the statistical function of EXCEL, the cumulative confirmed cases, cumulative cured cases and cumulative death cases of new-type corona pneumonia from January 2020 to July 2020 were correlated with the vector map of China’s administrative region to obtain the spatial distribution of new-type corona pneumonia nationwide. The cumulative diagnosis, cure and death tolls of the new type of coronary pneumonia were represented by color classification, and the situation of the new type of coronary pneumonia nationwide was visualized. The analysis showed that the distribution of the new type of coronary pneumonia at this stage was clustered, mainly centered in Hubei Province, spreading to the surrounding Hunan Province, Henan Province, Jiangxi Province and Anhui Province, and the sudden occurrence of the new type of coronary pneumonia occurred in very few regions.

1. Introduction
ArcGIS is widely used in the field of public health, and its various functions can visualize the distribution and trend of diseases. At the end of 2019, with the outbreak of COVID-19 in Wuhan, Hubei Province, China entered the stage of epidemic prevention and control. In 2020, the national epidemic broke out in a wide range, and the new corona virus pneumonia is an infectious disease. Therefore, it is very important to analyze the spatial distribution of its incidence for disease research and epidemic monitoring and defense.

2. Data Sources and Research Methods

2.1. data sources
The data used in this study are the data related to the new type of coronary pneumonia from January 2020 to July 2020. The data types include cumulative confirmed cases, cumulative deaths, cumulative cured cases and number of close contacts. The above data are collected in the daily publication of the National Health Commission, and the data of all provinces in China are integrated through the news bulletin issued by the National Health Commission. The Chinese vector topographic map and the nine-segment line in this study are downloaded through the Geographical Conditions Monitoring Platform (Dsac.cn, ).

2.2. Research Method
This study mainly uses ArcGIS for data processing, and the Federal Geographic Data Commission (FGDC) defines it as a computer system for input, storage, maintenance, management, retrieval, analysis, synthesis and output of geographic information or location-based information[1]. ArcGIS is a computer technology system, which is based on geographic spatial database and supported by
computer hardware. It can collect, manage, operate, analyze, simulate and display spatial related data, and provide a variety of spatial and dynamic geographic information in real time by using geographic model analysis methods. The data of the new coronavirus disease can be correlated with the topographic map of China, and the spatial and temporal distribution of various data can be analyzed. At present, the research in China mainly focuses on the spatial heterogeneity of diseases, which is described by statistical charts and tables. Some studies study the spatial heterogeneity of diseases through probability distribution function [2–3]. Shan Ke et al. [4] used ArcGIS function for exploratory data analysis and spatial analysis to conduct spatial analysis and statistical research on tuberculosis data in China. By analyzing the spatial distribution of high and low areas of tuberculosis incidence and the regional distribution of positive and negative hotspots, it is concluded that the incidence of tuberculosis in China has obvious regional distribution, and there is moderate spatial aggregation. Yang et al. [5] used GIS to analyze the multi-factor spatial composite model to predict the distribution of malaria epidemic areas in China and concluded. Wen [6] used GIS to analyze the spatio-temporal trend of hand-to-hand mouth disease, and made thematic maps for spatio-temporal trend analysis. Lu Yijie [7] Based on the characteristics of real-time monitoring and precise analysis of big data, the global epidemic situation was monitored visually and interactively by ArcGIS, and the development trend of coronavirus epidemic in 2019. Zhang Yanyang, etc. [8] used ArcGIS function to carry out the spatial distribution of disease incidence areas, and combined with time data to process spatial and temporal distribution, so as to predict and prevent diseases. Zhang et al. [8] collected the reported data of new coronavirus pneumonia in Henan Province, established GIS database, and applied spatial correlation analysis software and spatio-temporal scanning software to conduct spatial autocorrelation analysis and spatio-temporal scanning statistical analysis of the data. In this study, we mainly processed and analyzed the data of various types of cases of new-type coronary pneumonia from January 2020 to July 2020, and studied the spatio-temporal distribution and incidence range of new-type coronary pneumonia.

3. Data Processing

3.1. Treating Process

By pre-processing such as projection transformation of the downloaded vector image, the vector image that meets the requirements is obtained. Excel was used to collect and process the epidemic data. Link the epidemic data to the vector map of China’s administrative region, which contains daily cumulative case data. The data will be graded according to the size of the number by the function of color in GIS. In this study, the graded classes are selected according to the actual situation of each cumulative case data to form a thematic map, and the growth and trend of each data type of epidemic are distinguished by color difference and color depth. Spatial autocorrelation analysis was used to obtain the spatial autocorrelation distribution of cumulative confirmed cases in all provinces of China. The spatial autocorrelation in this study was analyzed by Moran index. Moran’s I > 0 means spatial positive correlation, the larger the value, the more obvious the spatial correlation. Otherwise, Moran’s I = 0, the space is random. The Global Moran’s I Spatial Statistics function in GIS was used for image processing to analyze the spatial distribution of new coronavirus pneumonia. Spatial autocorrelation Moran’s I statistics can be expressed as:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \omega_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2}$$

Where $z_i$ is the deviation between the attribute of element i and its average value (xi), is the spatial weight between element i and j, n is equal to the total number of elements, $S_0$ is the aggregation of all spatial weights:

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} \omega_{i,j}$$

The statistical z score is calculated as follows:
$z_1 = \frac{I - E[I]}{\sqrt{V[I]}}$

Among them:

$E[I] = -1/(n - 1)$

$V[I] = E[I^2] - E[I]^2$

For the pattern analysis tool, p value represents the probability that the observed spatial pattern is created by a random process. When p is small, it means that the observed spatial pattern is unlikely to be generated in random processes (small probability events). Hotspot analysis based on the new type of coronal pneumonia data is characterized by high values and surrounded by other factors with high values, which is called hot spots for statistical significance. Hotspot analysis tools can calculate Getis-Ord $G_i^*$ statistics for each element in the dataset. By looking at the z score and p value of each element in the adjacent element environment, the location of clustering of high or low value elements in space can be analyzed. Clustering analysis of epidemic data is also called group analysis or point group analysis. It is a quantitative method to study the classification of multi-factor things, a new multivariate statistical method, and a combination of contemporary taxonomy and multivariate analysis. The Getis-Ord General G statistics can be used to measure the clustering degree of high or low values.

3.2. Processing Result

Figure 1 shows the distribution and quantity changes of the four stages of the country’s cumulative diagnosed cases. From the figure, it can be concluded that Hubei Province has been the largest number of people infected with new coronavirus pneumonia. At the initial stage, Hunan Province and Hubei Province around Hubei Province are more serious areas. With the change of time, Jiangxi Province and Anhui Province around Hubei Province have become a serious disaster area of new coronavirus pneumonia. At the same time, the number of confirmed cases in Heilongjiang Province is small in the initial stage, but the number of cases in the later stage is increasing. The number of confirmed cases in Zhejiang Province has a continuous high incidence stage at this stage. The number of confirmed cases in Beijing is more in the early and late stages, but the number is less in the middle stage, and there is no high incidence. It can be clearly seen from the figure that the epidemic-prone areas are still centered in Hubei Province and spread to the surrounding areas, but there will also be a surge in the number of individuals in some provinces.

Figure 2 shows the distribution of the total number of cured cases in the country. The analysis results show that the distribution of the number of cured cases is basically the same as that of the diagnosed cases, and the spread is mainly centered around Hubei Province. It can be shown that the allocation of relief resources in China is in line with the incidence of the epidemic. Through this analysis, we can monitor and prevent the epidemic in the later period.
Figure 3 shows the distribution of the cumulative number of deaths in the country. Hubei Province is the main high-incidence area of death. At the same time, the mortality rate of Henan Province is also high, which is in a growth trend. The initial situation is better, and the number of people increases in the later period. The mortality rate in Heilongjiang Province is also high in the whole country, and Beijing is also the same area. The number of deaths in Hebei, Tianjin and Shandong around Beijing also showed a small increase in the later period. The mortality rate in Anhui Province is in a trend of first increase and then decrease. The number of deaths in Xinjiang Uygur Autonomous Region increased in the later period.

Spatial autocorrelation test is divided into global spatial autocorrelation test and local spatial autocorrelation test. Global spatial autocorrelation represents the spatial dependence of the new coronavirus in the national space, and Morans’ I index is commonly used to measure and represent it. Through the spatial autocorrelation analysis of the cumulative confirmed cases in each province from January 2020 to July 2020. According to the official data of GIS, when z value is greater than 2.58 and p value is less than 0.01, the confidence is 99 %. According to the analysis results, it can be concluded that the distribution of confirmed cases of new coronavirus pneumonia is not random distribution, but clustering distribution. The spatial distribution of high and low values of cumulatively diagnosed cases in each province can be obtained through hotspot analysis. For the positive z score with significant statistical significance, the higher the z score is, the closer the clustering of high value (hotspot) is. From the analysis of the results, it can be concluded that the agglomeration of hot spots is mainly centered on Hubei Province, and there are also hot spots agglomeration in surrounding Henan Province, Anhui Province, Jiangxi Province and Hunan Province. Most of these provinces have a 95 % confidence level in most of the time, and the confidence level in the initial stage of Henan Province is 90 %. The cumulative diagnosed cases in China have high value agglomeration from January 2020 to July 2020. Except for the above five provinces, the cumulative diagnosed cases in other provinces have no high value agglomeration. Hot spot analysis results are shown in Figure 4.
By using the high / low clustering analysis of the national cumulative confirmed case data, \( z = 3.42, p = 0.000628 \). The distribution of the epidemic has a certain spatial agglomeration, which also proves that the high incidence area of new coronavirus pneumonia is close to the high incidence area, which conforms to the high infectious characteristics of new coronavirus pneumonia.

4. Conclusions and Discussion
Based on the data of cumulative diagnosed cases, cumulative cured cases and cumulative death cases of new coronavirus pneumonia from January 2020 to July 2020 and the vector map of China’s administrative region, the spatial distribution analysis was carried out by ArcGIS. The results showed that:

(1) The incidence of new-type corona pneumonia has obvious regional differences in time, and there is also a certain leap. Before January 30, it can be regarded as the initial stage of the epidemic. The new cases of the epidemic in China are mainly concentrated in Hubei Province, especially Wuhan City. The number of new cases accounts for about 60 % – 70 % of the new cases in China. Other cities are still in a low incidence, but the epidemic has spread outward. By mid-February, the epidemic had entered the outbreak stage. Driven by the return of the Spring Festival, the epidemic was widely spread throughout the country, and the impact range was rapidly expanded. This stage is characterized by an increasing number of confirmed cases throughout the country, peaking and fluctuating at high levels. Entering the mid-April is the epidemic control stage, and the number of new cases per day in China gradually decreases. In general, the number of newly diagnosed people gradually decreased at this stage, and the epidemic prevention and control measures in China were effective. In July 2020, when the epidemic ended, the cure rate in China reached 93.94 %, which was the highest in the world.

(2) The spatial dimension of the epidemic distribution has agglomeration, which has certain reasons for the population density of each province and the early onset of the disease. At the same time, there are also reasons for the timely treatment of national medical deployment, which makes the situation of new coronavirus pneumonia in China stable and alleviate, and there is no large area of diffusion. In general, the cumulative diagnosis, cumulative cure and cumulative death cases in Hubei Province are
the highest in the country, and the surrounding Henan Province, Hunan Province, Jiangxi Province and Anhui Province are relatively high, showing a high value of agglomeration. At the same time, there are also periodic population growth and decline in other regions, such as Heilongjiang, Beijing and Shandong.

(3) In the current domestic epidemic prevention and control situation, it has been basically stable and controlled. However, the epidemic situation outside China is complex and severe, and there are still increasing cases of infection in some countries. Therefore, against the background of great pressure of “external defense input and internal defense rebound” [9], we should use the existing technology to monitor and control the new type of coronary pneumonia, and use the current technology to reasonably allocate and effectively use materials, so as to cope with the sudden outbreak of new type of coronary pneumonia and do a good job in safety and health monitoring.

References
[1] Fan Wei. GIS application and development trend in the field of public health [J]. Wireless Internet technology, 2012 (07): 166-168.
[2] Kulldorff M. A spatial scan statistic[J]. Communications in Statistics: Theory and Methods, 1997, 26(6): 1481-1496.
[3] Martin Kulldorff, Lan Huang, Linda Pickle, Luiz Duczmal. An elliptic spatial scan statistic[J]. Statistics in Medicine, 2006, 25(22).
[4] Shan Ke, Xu Lingzhong, Gai Ruoyan, Wang Xingzhou, Yu Huaijin, Cui Yimeng, Lu Yifan, Xu Rongfei. GIS spatial analysis of tuberculosis prevalence in China from 2002 to 2011 [J]. Public health in China, 2014, 30(04): 388 – 391.
[5] Epidemiological group of new coronavirus pneumonia emergency response mechanism in China Center for Disease Control and Prevention. Epidemiological characteristics of new coronavirus pneumonia [J]. Chinese Journal of Epidemiology, 2020, 41(2): 145 – 151.
[6] Wen Xiangzhi. Epidemiological characteristics of hand, foot and mouth disease in Handan City from 2009 to 2018 based on GIS technology [D]. North China University of Technology, 2020.
[7] Lu Yijie, Zhang Zhen. Design of a new real-time monitoring system for coronavirus pneumonia based on ArcGIS Dashboard [J]. Computer knowledge and technology, 2020, 16(14): 105-107.
[8] Zhang Yanyang, Xiao Zhanpei, Yang Kaichao, Zhao Dongyang, Li Jun, Zhang Mingyu, Zhang Xiaoxiao, Ma Yating, Wang Changshuang, Wang Yan, Chen Shuaiyin. Temporal and spatial clustering of new coronavirus pneumonia in Henan Province [J]. Chinese Journal of Disease Control, 2020, 24(05): 534-538 + 572.
[9] Lv Zhenhua, Cheng Shaowen. Spatial and temporal characteristics of new coronavirus in Hubei Province based on Crystal Ball and GIS [J]. Journal of Central China Normal University (Natural Science Edition), 2020, 54(06): 1059 – 1071.
[10] Wu Weishen, Li Yonggang, Wei Zhaofei, Zhou Penghui, Lv Likun, Zhang Guoping, Zhao Ying, He Haiyan, Li Xiaoyan, Gao Lu, Zhang Xiumei, Liu Hui, Zhou Ning, Guo Yan, Zhang Xiaomeng, Zhang Dan, Liu Jing, Zhang Ying. Investigation and analysis of new coronavirus pneumonia aggregation epidemic in a department store in Tianjin [J]. Chinese Journal of Epidemiology, 2020, 4(04): 489-490-491-492-493.
[11] Wen Xiangzhi. Epidemiological characteristics of hand, foot and mouth disease in Handan City from 2009 to 2018 based on GIS technology [D]. North China University of Technology, 2020.
[12] Review Xiujun, Li Xinlou, Liu Kun, Zhao Xiaobo, Ma Meng, Sun Bo. Review on the application progress of geographic information system in the prevention and control of new coronavirus pneumonia [J]. Journal of Shandong University (Medical Edition), 2020, 58(10): 13 – 19.