Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 8, No. 1, 2023

Spatial Data Analysis of Sudden Epidemic Diseases Based on Case-Based Travel Data

-- Take Corona Virus Disease 2019 As An Example

Yukun Zhai[1, *]

[1]School of Mining Engineering, Taiyuan University of Technology, Taiyuan, Shanxi, 030002, China
[*]Corresponding author's e-mail: zhaiyukun3911@link.tyut.edu.cn

Abstract: In recent years, health geography has attracted the attention of more and more institutions and scholars in China, but as far as the results of a large number of studies are concerned, there are not many related types of research. Geographic Information System (GIS) and spatial data method can effectively deal with complex spatial data, time data and attribute data in diseases, analyze the relationship between the distribution and changes of diseases in different regions, and explore the causes of diseases and various environmental factors. According to the analysis results, the hierarchical model of spatial aggregation can find out the dynamic spatial distribution law of epidemic occurrence and prevalence more intuitively, and carry out epidemic detection and early warning in depth, thus laying the foundation for formulating appropriate prevention and control strategies. Corona Virus Disease 2019 is highly contagious, with diverse and complicated transmission methods, wide spread and high incidence. Therefore, taking Corona Virus Disease 2019 as an example, it is of great significance to study the application of spatial aggregation hierarchical model and spatial data analysis technology in epidemiology through GIS technology, especially in the spatial aggregation hierarchical model to visualize the travel data of epidemic cases and show the temporal and spatial distribution characteristics. In this work, Kriging interpolation will be tried to supplement the travel data of cases and then analyze the spatial data.

Keywords: Geographical Information System; Spatial data analysis; Kriging interpolation method; Health geography; Spatial aggregation model.

1. Introduction

Geographic Information System appeared and developed since the 1960s, while spatial data analysis technology has developed rapidly only in recent 20 years. There are still many imperfections in some models, methods and technologies of spatial data analysis. [1] In recent years, GIS has been successfully applied in the field of public health, showing a good development trend and application prospects. However, on the whole, the application of Geographic Information System in the field of public health is still in the initial stage, and its main research fields mainly focus on the distribution monitoring of endemic diseases and individual vector-borne diseases, and it is far from going deep into many fields of public health that urgently need spatial analysis. At the same time, the application of GIS in the field of public health is limited by data sources. The development of public health research began in industrial revolution in Europe in the 17th century, when a large number of farmers concentrated in cities, which worsened and led to the prevalence and spread of diseases such as cholera, dysentery and tuberculosis. The health status of residents has obviously declined, and the research on public health has been gradually established in the process of taking environmental health measures and fighting diseases. The lack of corresponding spatial attributes in the development of public health limits the development of spatial data analysis and public health. There are many researches in the current literature on establishing time propagation rate model and spatial statistical analysis to estimate the spread of epidemic diseases, such as the SARS virus outbreak in China in 2003 and Wang Jinfeng.[2]Using Geographic Information System, the epidemic data of SARS in Beijing are estimated by establishing a time propagation rate model. Fang Liqun used GIS technology, the data of highly pathogenic avian influenza in China in January and February, 2004 were collected and the database was established[3]. The relationship between the data and digital map was established in GIS software. At the same time, the meteorological data, vegetation remote sensing images and migratory bird migration data of China were collected to analyze the related environmental factors of avian influenza. In this paper, GIS technology will be used to analyze the spatial data of patients' travel data in order to assist local epidemic prevention and control.

2. Screening of Patient Travel Data

First, Download the travel data of COVID-19 cases from the local disease surveillance center, and then count the travel data of the case before, during and after the illness. And people are classified according to different ages, sexes, occupations, nationalities or races. Because family members live a common life and have the closest contact, the travel data of family members of cases are also of great reference value, and are classified according to different ages, genders, occupations, nationalities or races. For the data classified by people, it is classified according to the time characteristics of their travel data. Classification includes: short-term fluctuation, long-term variation, periodic epidemic and seasonal increase.

3. Transformation of Data

Because the patient's travel data collected from the disease monitoring center is a set of data that presents a linear shape in three-dimensional space. In order to perform Kriging interpolation on this linear data in three-dimensional space. In this work, will present a data conversion method. The specific step is to classify this kind of linear data in three-dimensional space according to time latitude. Set a certain time threshold, When the case stays in a certain location for more than the threshold, the data information of this location will be retained, and when the stay time in a certain location is less than the threshold, the data information of this location will be discarded. In this way, the transformed data can be used for exploratory spatial data analysis and kriging interpolation by geostatistics module in ArcGIS software.

4. Exploratory Analysis of Travel Data

This paper will make an exploratory analysis of the data converted above. The exploratory analysis methods mainly involved are histogram, QQplot, Voronoi diagram, semivariogram cloud map.

4.1. histogram

Histogram is to classify the existing data, and then express this classification in the form of histogram. The statistical result is the proportion of sampling points in the total number of each level. According to Figure 1-1, the incidence data of COVID-19 in Handan City from 2019 to 2023 are very uneven in frequency distribution, and its main distribution pattern is skewness.

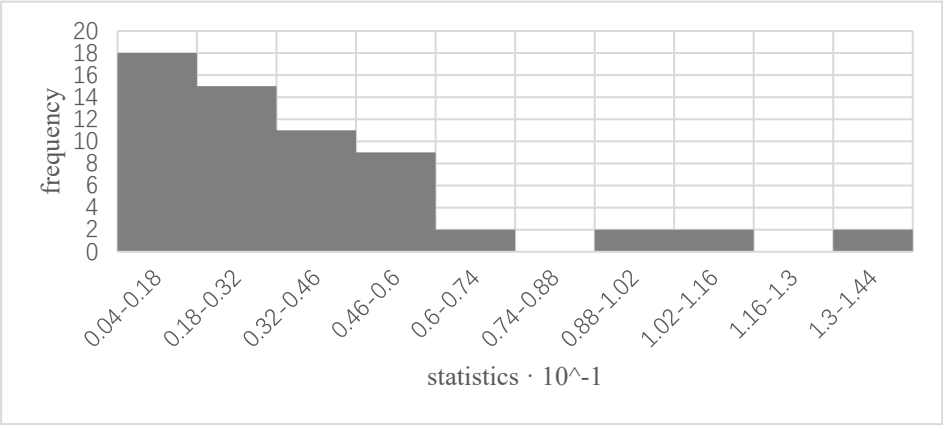

Figure 1. Incidence histogram

4.2. QQplot

QQplot is a normal quantile graph, which is mainly used to visually show whether a group of data is normal distribution, and its image shows the comparison between normal distribution and existing data. If the data is closer to a straight line, it is close to normal distribution. The following figure shows the distribution of the incidence data of COVID-19 in Handan from 2019 to 2023 after logarithmic changes. It is basically a normal distribution.

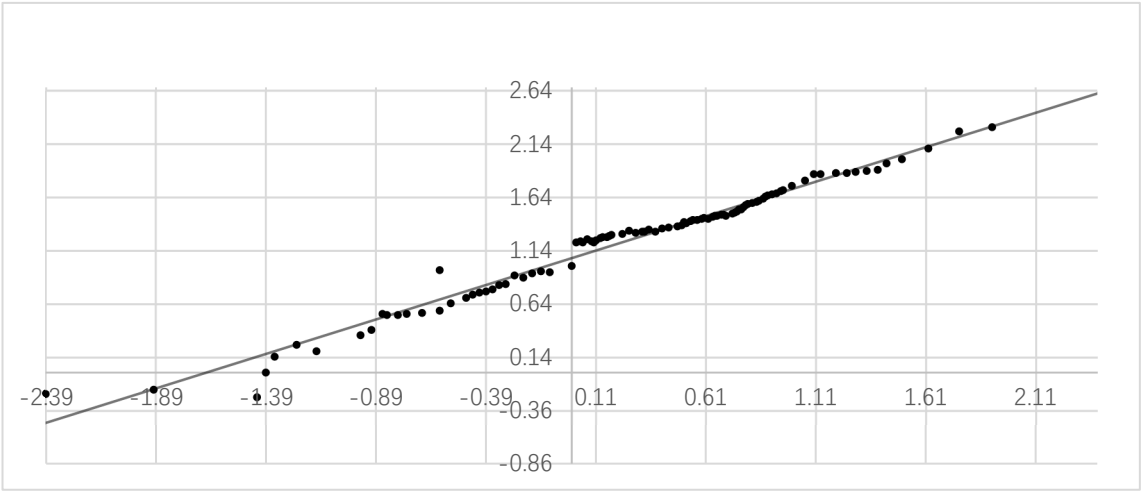

Figure 2. Logarithmic incidence QQplot diagram

4.3. Voronoi

Voronoi diagram, also called Tai Sen polygon, is composed of a group of perpendicular bisector of line segments connecting two adjacent points, and the distance between the points of the polygon and the corresponding discrete points is the shortest. The algorithm is generated by using entropy and clustering method, and it is mainly used to find outliers of spatial points in exploratory data analysis, because entropy shows the dissimilarity of adjacent units, so local outliers are identified by the level of entropy.

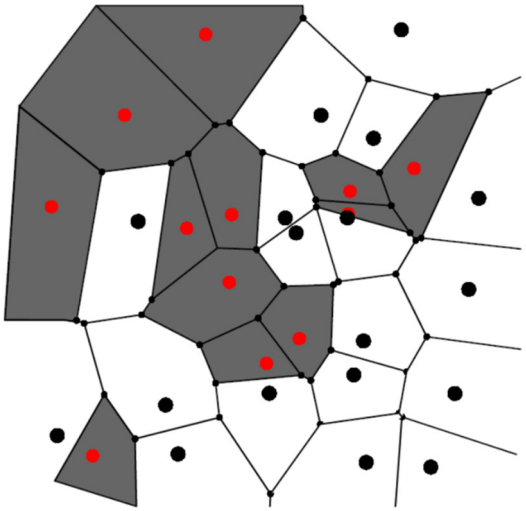

Figure 3. Voronoi diagram of logarithmic COVID-19 incidence

4.4. Semi-variogram

The semivariogram represents the theoretical semivariogram of all the sample points in the data center, and represents them as a function of the distance between two points, and this function becomes a semivariogram. Because the characteristic points are discrete, the calculation formula of the distributor, that is, the formula of the semivariogram, refers to Qu Rong.[4]

5. Kriging Interpolation and Disease Mapping Expression

The essence of Kriging method is to take the regionalized variables as the research object and use the variation function as the research tool to make unbiased optimal estimation of the values of the regionalized variables of the non-sampled points. This paper mainly uses geostatistics function in ArcGIS for Kriging interpolation and exploratory spatial data analysis. ArcGIS provides users with a statistical tool. First, we understand the characteristics and properties of the overall data before further spatial analysis, and then establish a new prediction surface through deterministic methods and geostatistics methods. According to the previous exponential models, this paper uses Kriging method to interpolate the incidence rate of COVID-19 in Handan from 2019 to 2023, and then we can get the isoline map of COVID-19 incidence rate in Handan over the years. According to the spatial representation of the data converted from the travel data of cases, the spatial location of the incidence rate in a certain area can be directly expressed, so that the relevant departments can more clearly understand and study the overall pattern and local laws of the incidence rate of diseases here.

6. Conclusion

The variation function used in geostatistics studies the spatial distribution and spatial correlation of regionalized variables with randomness and structure. This method has been widely used in the study of spatial distribution pattern of various natural and social phenomena. Based on the analysis of the travel data of COVID-19 cases, this paper puts forward a method of transforming the travel data of cases into point data by using time classification, and then realizes the spatialization of the travel data of cases by combining exploratory spatial data analysis method with geostatistics method, and obtains good cartographic expression effect. Spatial data analysis technology can intuitively describe the spatial distribution and pattern of diseases, and such results usually affect the health status and related decisions of an area. Studying the spatial distribution of diseases and their changing laws and characteristics can provide basis for relevant departments and scholars to further study the causes of diseases, promote the health of public health to a certain extent, and play a role in preventing diseases and providing strategies and measures for public health services. Based on the spatial data analysis method, this paper discusses the problems related to disease mapping and spatial distribution in the field of sudden epidemic diseases. The research in this paper is still very shallow, and there are still a lot of related problems that need further study.

References

[1] Wu, J.L., Wang, B.F. (2003) application of air data analysis technology in public health leading cities. Progress in Geographical Science, 39: 219-228.

[2] Jin, F., Wang Anthony, Bin, M., et al. (2006) Spatial dynamics if an epidemic of severe acute respriratory syndrome in an urhan area. Bulletin of the World Health Organization, 12: 965-969.

[3] Fang, L.Q., Cao, C.X., Chen, G.S., et al. (2005) Analysis of spatial distribution and environmental factors of highly pathogenic avian influenza in Chinese mainland by GIS. Chinese Journal of Epidemiology, 11: 839-842.

[4] Qu, R. (2013) Study on spatial pattern of epidemic diseases based on spatial data analysis technology. Thesis of Wuhan University, 1: 38-40.