Research of Spatial Distribution Rules of Mass Incidents Based on GIS

HU Jun, SHU Xue-ming, SHEN Shi-fei, TANG Shi-yang
Department of Engineering Physics, Institute of Public Safety Research, Tsinghua University,
Beijing 100084, China
008618810654098
hujun18810654098@126.com

ABSTRACT
The research of hot-spots is an important subject in public safety field as well as GIS system application, it reveals the spatial distribution law of crime and help develop targeted security preventive strategy, providing safeguards for the whole society.
The mass incident is a special kind of public safety incident, but there are little research on its spatial distribution rules from the perspective of geography. This article analyzes the spatial distribution rules of mass incidents in China from the macro point of view and based on the spatial autocorrelation and Kernel density estimation, then finds hot-spots and calculates the degree of spatial agglomeration to express the spatial distribution law of crime.

CCS Concepts
• Information systems – Database management system – Database design and models – Graph-based database models.

Keywords
Mass incidents; Spatial distribution; Spatial autocorrelation; Kernel density estimation.

1. INTRODUCTION
Now China is in the period of rapid economic development as well as urbanization, the social security risk becomes more and more serious. In recent years, mass incidents occur in many places of China, and most of them are caused by forced building demolition, land expropriation and environment problems, such as the PX incidents happened in Ningbo, Zhejiang. The current researches of mass incidents mainly focus on the cause analysis, prevention and response. For example, Yang analyzed the cause with Coser’s social conflict theory [1], and Hua gave some advice and countermeasures for mass incidents [2]. But there are little research on the spatial distribution rules from the perspective of geography. In fact, there is a relationship among all the incidents in geographical space, and there are some spatial distribution rules of mass incidents. Recognizing and understanding the rules and features of mass incidents in the space is very meaningful for incidents prevention and social security risk control.

The study of public order cases from the view of geography dates back to the 19th century. In 1833, Guerry analyzed the crime data of France from 1825 to 1830 by mapping and found out high levels of violent crime in southern France while high levels of property theft cases in northern France [3]. By the early 1900s, the Sociology Department at Chicago University explored the distribution of crime and other social problems in the city and explained with community difference. An famous example is Shaw and McKay’s research of distribution model of juvenile delinquency [4]. And today with the development of data processing technology and spatial information technologies based on GIS platform, the research of cases can be more specific. For example, Mencken and Bamett analyzed the spatial autocorrelation of average murder rates in 6 states of the south central United States from 1989 to 1991, and found out New Orleans is a hot-spot [5]. Yan built an model which took community as the unit and considered the spatial autocorrelation among cases and non-stability of the spatial structure to analyze the relationship among theft cases, population, roads and police force distribution [6]. Those researches are all case analysis on small scales, such as at a particular city or a certain community. And on the other hand, as a kind of incident happened under special conditions, the geography space analysis of mass incidents are less. This article researches the spatial distribution rules of mass incidents based on the spatial autocorrelation and Kernel density estimation in macro scale, in the purpose of providing reference for regional security assessment and targeted security preventive strategy to protect the social security.

2. Research Methods
This article is based on the spatial autocorrelation and Kernel density estimation to research the spatial distribution rules of mass incidents in China from the macro point of view on the GIS platform. This article’s software is ArcGis10.2.2, using the spatial analysis tool developed by ESRI (Environmental Systems Research Institute, Inc.). The following are the theoretical base of research methods in brief [7-10].

2.1 Spatial Autocorrelation
Spatial autocorrelation means the potential interdependency among the observational data in the same distribution. In the research of crime distribution, the spatial autocorrelation also refers to the interaction effect of crime rate among different zones. The tests for spatial autocorrelation includes global indexes and local indexes, the global indexes are used to detect the spatial distribution in the whole study area, reflecting the degree of autocorrelation by a single value, such as Moran’s Index and Getis-Ord general G, both of which are classical global clustering test index; while the local indexes are used to calculate the degree of correlation for a particular attribute among the spatial unit and its neighboring units, such as local Moran’s Index and Getis-Ord Gi.

The global Moran’s Index can test the neighboring districts are similar, different, or independent. The calculating formula is below:

\[ I = \frac{N \sum_{i} \sum_{j} w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\left( \sum_{i} \sum_{j} w_{ij} \right) \sum_{i} (X_i - \bar{X})^2} \]  \hspace{1cm} (1)

Where N is the number of face shape elements in the research area, \( X_i \), \( X_j \) are the attribute value of zone i and j, \( w_{ij} \) is the spatial weighted matrix, which expresses the connection of object between position i and j. The value of I ranges from -1 to 1, when I is closer to 1, it means that the similar attributes cluster (positive spatial
correlation), namely High-Clusters or Low-Clusters; when I is closer to -1, it means that the different attributes cluster (negative spatial correlation), namely High-Low-Clusters; when I is closer to 0, it means the attributes are random.

The calculating formula of Getis-Ord general G is:

\[ G = \frac{\sum \sum \omega_{ij} X_i X_j}{\sum \sum X_i X_j} \quad \ldots \quad (2) \]

The meaning of symbols are as before. The difference of global Moran’s Index and Getis-Ord general G is that Moran’s Index can judge the spatial distribution pattern is clustered or random, but it cannot distinguish the cluster types; while Getis-Ord general G can distinguish the cluster types are High-Clusters, Low-Clusters or just random. So we can use the indexes both to make the cluster research more general and accurate.

The global spatial autocorrelation assumes spatial homogeneity, there exists only a global trend of the element attribute. However, the spatial heterogeneity exists almost in all geographical research, especially in the mass incidents research, the mass incidents in different place are different in scale, cause and other aspects. So the local features should be considered to measure the local correlation of each elements. Local Moran Index is used to test whether the similar or different observed values gather or not in the local areas. The Local Moran Index of element i is given by:

\[ I_i = \frac{(X_i - \overline{X})}{S^2} \sum \omega_{ij} \left( X_j - \overline{X} \right) \quad \ldots \quad (3) \]

The meaning of symbols are as before. When \( I_i \) is positive, it means the attributes around the zone are similar, namely a high value is surrounded by high values, or a low value is surrounded by low values; while \( I_i \) is negative, it means the attributes around the zone are different, namely a high value is surrounded by low values, or a low value is surrounded by high values.

Getis-Ord \( G_i \) is also used to test local cluster, and it is defined as:

\[ G_i = \frac{\sum \omega_{ij} X_j}{\sum X_j} \quad \ldots \quad (5) \]

The meaning of symbols are as before. High \( G_i \) means samples with high value cluster, while low \( G_i \) means samples with low value cluster. Generally, the \( G_i \) need standardize to get Z score, and there are difference between \( G_i \) and \( L \). The difference will be analyzed in the following parts.

2.2 Kernel Density Estimation

Kernel density estimation is a kind of nonparametric testing method and is used to estimate the density function of random variable. It is applicable to generate a smooth estimation surface to describe the cluster degree of observational data. Compared with the discrete point, which can’t show the spatial trend clearly, Kernel density estimation can transform the set of points into a surface whose density changes continuously and enhance the display of spatial distribution pattern. Kernel density estimation doesn’t use any prior knowledge of data distribution or make any assumptions, it is a method just based on the data itself to research the data distribution characters.

The substance of Kernel density estimation can be summed up as: choosing appropriate kernel function \( K \), and calculating the probability density estimation \( f(x) \) of every grid center point \( X \) after meshing the research areas. \( f(x) \) is depended on the relationship between k subjects (observation point \( X_i \)) contained within the window width \( h \) (a threshold chosen artificially) and grid center point \( X \).

3. Research Areas and Data

As the particularity of mass incident, the research areas of this article are just the mainland of China, except for Hong Kong, Macao and Formosa. And in order to ensure the statistical regularity, the number of mass incident cases can’t be too little. The data in this article is the mass incidents nationwide from 2005 to 2015. With focusing crawler technology, we grabbed the spatial location and time information of each cases from the websites, and totally over 11800 cases. Turning the data into spatial coordinates, the cases are shown in Figure 1. The cases numbers of each provinces, autonomous regions and municipalities are shown in Figure 2. It’s important to note that the geolocations of some cases are not so precise, so the locations are fuzzified. And the requirement of scale is more than ten people.

- **space points of mass incidents**

![Figure 1 the space points of mass incidents in China](image1)

- **The number of mass incidents number of each provinces, autonomous regions and municipalities in China**

![Figure 2 the mass incidents number of each provinces, autonomous regions and municipalities in China](image2)

Look at the purely number, Shandong, Fujian and Guangdong province are focus areas. Of course, considering the management of public reporting of mass incidents, the result may not mean the significance of public protest in those provinces, but to some degree, it reflects the high social management risk during the process of economic construction and urbanization.

Certainly, the analysis of hot-spots is not just quantitative analysis. The following is the analysis of Spatial Autocorrelation.
4. Result

4.1 Global Spatial Autocorrelation Test

The result of global spatial autocorrelation test is shown in Table 1. With the help of ArcGIS 10.2.2, the value of global Moran’s Index and Getis-Ord general G as well as their Z-score can be calculated. From the statistical test method, Z test, we can know that when the confidence value is 99%, the corresponding Z score is 2.58. As the Z-score of global Moran’s Index is 2.86, which is greater than 2.58, we can conclude that the mass incidents has spatial connection, whose probability is greater than 99%.

| Table 1 The result of global spatial autocorrelation test |
|-------------------------------|-------------------|
| global indexes                | Value             | Z-score |
| global Moran’s Index          | 0.025             | 2.86   |
| Getis-Ord general G           | 0.000031          | 3.37   |

Similarly, the Z-score of Getis-Ord general G is 3.37, greater than 2.58, indicating that the cluster type of mass incidents are High-Clusters, and the probability is greater than 99%.

4.2 Local Spatial Autocorrelation Test

Except for the global trend of the mass incidents, we also research the local correlation considering the spatial heterogeneity. By calculating the local Moran Index and significance testing, the cluster type of each provinces, autonomous regions and municipalities can be obtained as Not Significant, High-High Cluster, High-Low Outlier, Low-High Outlier and Low-Low Cluster. The result is shown in Figure 3 (the significant level is 5%).

![Figure 3](image-url)  
**Figure 3** The cluster type based on the local Moran Index and significance testing

According to Figure 3, it can be seen that there are three main cluster type in China, namely Not Significant, Low-Low Cluster and Low-High Outlier. In the western regions and most of central regions, the mass incidents is Low-Low Cluster, as the economic level in those regions are relatively lower, limiting the various contradictions in the urbanization. In the eastern regions, the number of mass incidents is Low-High Outlier, which means that the mass incident in eastern regions is low, but it is surrounded by high value.

If we calculate the Getis-Ord Gi and then standardize to get Z score, the result is shown in Figure 4. High positive Z score means high value cluster, low negative Z score means low value cluster, while Z score is close to 0, it means no significant cluster.

![Figure 4](image-url)  
**Figure 4** Z score of Getis-Ord Gi

According to Figure 4, in the western regions and most of central regions, the mass incidents is low value cluster while in some eastern regions, the mass incidents is high value cluster. Compare with the result of local Moran Index, we can conclude that in the western regions and most of central regions, the mass incidents is significantly less, but the eastern regions is high value cluster. This result is agree with actual states, as the economic level in eastern regions are higher, many mass incidents occur as the result of urbanization, such as forced building demolition, land expropriation and resist of chemical enterprise.

The difference result of Getis-Ord Gi and local Moran Index can be partly explained by the distinction in computing. There are difference in the capability of exploring cluster zones between Getis-Ord Gi and local Moran Index [11]. In general, local Moran Index is weaker than Getis-Ord Gi. Getis-Ord Gi can explore the cluster zones more accurately, while the local Moran Index can just explore the center of cluster zones, and the deviation of cluster range is large, the explored range is smaller than actual range.

4.3 Kernel Density Estimation Test

The above spatial autocorrelation is based on the surface factors, the analysis is on the scale of province, then get the results with limited precision. However, Kernel Density Estimation will not be limited by natural boundary of districts, it can reflect the gradation of gather degree. The cluster hot spots are in a smaller scale. The Kernel Density Estimation of mass incidents in China is shown in Figure 5.

![Figure 5](image-url)  
**Figure 5** Kernel Density Estimation of mass incidents in China

From Figure 5, there are two focusing points of mass incidents. One is in Beijing, and the other is in Chongqing. Many mass incidents are petitions, since Beijing is the administrative center of China, organizers of petition almost choose Beijing as appeal for help. Beijing is a hotspot; as for Chongqing, the most Economic development areas in western part of China is Sichuan Province and
Chongqing Municipality, where happen many mass incidents as the social contradictions are more complicated, so the another hotspot is the center of the two areas. In the southeast of China, each provinces (autonomous regions or municipalities) happens mass incidents, and the numbers are almost large and the place are relatively scattered, so there is no focus point. It also indicates the comprehensiveness of mass incidents in those regions on secondary side.

5. Conclusion
This study mainly analyzes the cluster degree and cluster type of mass incidents by spatial autocorrelation and kernel density estimation. By calculating the global Moran index, it can be proved that the mass incidents in China has spatial connection. The conclusion of local spatial autocorrelation is that in the western regions and most of central regions, the mass incidents is low value cluster and in some eastern regions, the mass incidents is high value cluster. With kernel density estimation, we can conclude that Beijing and Chongqing Municipality are two focusing points of mass incidents, and the mass incidents in the southeast of China are widely distributed.

The following research can discuss the cause of mass incident in different areas, and analyzing the correlation among population, economic and mass incidents based on the big attribute data in GIS.

6. ACKNOWLEDGMENTS
Supported by the National Science & Technology Pillar Program during the 12th Five-year Plan Period (No. 2015BAK12B03), Fire Department of Ministry of Public Security Research Programs (No.2014XFCX10), and the Collaborative Innovation Center of Public Safety.

7. REFERENCES