

Detecting Clustering Scales with the Incremental K-Function: Comparison Tests on Actual and Simulated Geospatial Datasets

Ran Tao, Jean-Claude Thill and Ikuho Yamada

Abstract The detection of so-called hot-spots in point datasets is important to generalize the spatial structures and properties in geospatial datasets. This is all the more important when spatial big data analytics is concerned. The K-function is regarded as one of the most effective methods to detect departures from randomness, high concentrations of point events and to examine the scale properties of a spatial point pattern. However, when applied to a pattern exhibiting local clusters, it can hardly determine the true scales of an observed pattern. We use a variant of the K-function that examines the number of events within a particular distance increment rather than the total number of events within a distance range. We compare the Incremental K-function to the standard K-function in terms of its fundamental properties and demonstrate the differences using several simulated point processes, which allow us to explore the range of conditions under which differences are obtained, as well as on a real-world geospatial dataset.

Keywords K-function · Point data · Hot spots · Spatial clustering · Scale · Spatial big data

R. Tao · J.-C. Thill (✉)

Department of Geography and Earth Sciences and Project Mosaic,
University of North Carolina at Charlotte, Charlotte, NC, USA
e-mail: Jean-Claude.Thill@uncc.edu

R. Tao

e-mail: rtao2@uncc.edu

I. Yamada

Department of Integrated Science and Engineering for Sustainable Society,
Chuo University, Tokyo, Japan
e-mail: iyamada.87@g.chuo-u.ac.jp

© Springer International Publishing Switzerland 2015

V. Popovich et al. (eds.), *Information Fusion and Geographic Information Systems (IF&GIS' 2015)*, Lecture Notes in Geoinformation and Cartography,
DOI 10.1007/978-3-319-16667-4_6

1 Introduction

The analysis of spatial point patterns has a long tradition in various scientific disciplines including geography, economics, demography, ecology, forestry, criminology, epidemiology, planning, and business [1, 2]. By detecting and analyzing the spatial point patterns, something interesting and informative about the underlying process that would have generated the events can be unveiled. For example, studies of animal behavior suggest that agglomerative or clustering point patterns are helpful to verify theories of territoriality and social organization; also, a diffusion point process in both spatial and temporal dimensions can provide evidence for various theories about information transmission or disease spreading [3]. A spatial concentration of cell phone users with certain attributes within a narrow time window may present a business opportunity or a risk to public safety.

Spatial point pattern analysis has recently regained popularity in spatial sciences and affiliated disciplines as an increasingly large volume of thematically diverse geospatial data is available as point data, with their own x-y coordinates, often with a time stamp. Developments in spatial technologies such as location-aware and remote sensing, advancements in information and communication technologies, along with data sharing inclination by public organizations and individuals in the form of social networking services and other public participation initiatives have created a data rich environment for social sciences that has no precedent in human history. Volunteered Geographic Information (VGI) has become a legitimate complement of existing data sets, which are still often published only at a spatially aggregated level for confidentiality reasons. The importance of this so-called “big data” defined with V criteria [4] has been well recognized by scientists across disciplines. In this “data avalanche” revolution [5], spatial sciences have a unique and vital role to play as most of the big data are georeferenced.

This paper is a contribution to the body of literature aimed at determining the existence of patterns in geospatial point datasets, particularly clumping or clustering structures. We propose a variant on the statistics based on the well-known K-function that would avoid overstating the spatial scope of clusters that may be detected in the empirical datasets. The so-called Incremental K-function is presented and evidence of its efficacy on real data and multiple simulated data sets are presented.

The rest of this paper is organized as follows. In the second part, we summarize the literature on spatial point pattern analysis. Next, we present the Incremental K-function method and its important properties. Then we conduct the comparison experiments with the conventional K-function on both real-world datasets and simulated ones and analyze the results. In the final part, we conclude on the characteristics and practical usefulness of the Incremental K-function.

2 Spatial Point Pattern Analysis

As one of the most common spatial patterns due to the general tendency of spatial phenomena (i.e. events) to co-occur spatially as encapsulated by Tobler's First Law of Geography [6], spatial clustering represents a general tendency of events occurring closer to each other than one might expect and it always draws great attentions in academics [7]. Clusters or clumps of events in the geographic space are commonly called hot spots. The detection of hot spots in one-dimensional spatial big data sets is of significance because it serves to generalize data and their spatial properties, which is critical for inferential purposes. A significant body of research has contributed to developing methodologies and tool sets for detecting spatial clustering. In the context of point pattern analysis, this family of methods is named second-order analysis of point processes [8, 9], or hot spots detection.

Early studies were primarily concerned with the overall pattern embedded in the spatial events. Therefore, a number of single-index spatial statistics, sometimes labeled as "global" statistics, were designed to depict the nature of events within the entire study area. Well-known examples include Moran's I, Geary's C, Quadrat Analysis, Nearest Neighbor Index, G statistic, and Ripley's K-function. Later on, scholars found that one of the fundamental assumptions of the global statistics, namely the spatial stationarity, is difficult to be held in many real situations. In addition, with only a single statistic to describe the entire study area, it is inadequate to further investigate more detailed aspects such as how the distribution of one variable would affect another in a localized fashion, or where departures from a random spatial distribution can be found [10]. To address these issues, along with the fast development of GIS in the 1990s, the study trend shifted to developing local statistics of detecting spatial clusters, i.e. 'hot spots'. Noticeable approaches include the geographical analysis machine (GAM) [11] and its derivative methods [1, 12], the local version of Ripley's K-function [9], local indicators of spatial association (LISA) especially the local Moran's I statistic, local Geary's C [13], and local G statistic [14, 15]. In contrast with their global counterparts, the local techniques are aimed at finding anomalies and interesting collections of spatial events within the study area that appear to be inconsistent with the background conceptual model of how events arise or at pinpointing the specific locations that serve as foci for clustering that repeats itself over the study area [7, 12]. Sometimes local methods with predetermined locations are given a special name "focused tests" to differentiate them with the ones based on randomly-chosen event locations [12]. More recently, related studies are aimed at handling emerging large point datasets with accurate locational information, sometimes with a time stamp. Techniques of Exploratory Spatial Data Analysis (ESDA), GeoComputation, and GeoVisualization are frequently incorporated for this purpose [16–18].

Among various methods of point pattern analysis, Ripley's K-function is regarded as one of the most effective methods at detecting whether a spatial process significantly departs from randomness, or whether it is more dispersed or more concentrated than random. The K-function is routinely used as a technique for hot spots detection,

that is for the discovery of high concentrations of point events. It has been enhanced through the decades since it was originally proposed by Ripley in 1976 [8, 19]. The fundamental idea of the K-function is to count the number of events within a certain distance to randomly selected event locations. The number is then used to calculate K-function value by dividing the density of events. To obtain statistical conclusions, the K-function value is compared with the expected value according to the null hypothesis, for example Complete Spatial Randomness (CSR). If the observed value is higher than expected, the study events have the tendency of clustering; or dispersed, if it is lower than expected. Monte Carlo simulation is frequently applied to assess the statistical significance of the results [11]. Several meaningful extensions and applications of the K-function have been conducted through the years. Of note is work by Getis and Franklin [9], where spatially local clusters are detected within the range of K-function, which is subsequently known as local K-function analysis. Boots and Okabe [20] discussed applying the Cross K-function as a focused test to identify clusters of events around specific locations for example crime cases surrounding rail stations. Yamada and Thill [2] adjusted both the global and local K-function to network-constrained space to study transportation-related cases. Tang et al. [18] incorporated Graphics Processing Units (GPU) technique to accelerate the computing process of Ripley's K Function for massive spatial point datasets.

Compared with other hot spot detection techniques, the K-function holds a unique advantage that spatial dependence is examined over a range of distances and spatial scales. For instance to analyze the spatial pattern of crime events within the city, with the K-function we can set the detecting radius from 0.1 to 10 miles so that clusters of the size within this range can all be discovered. This advantage enables the K-function to compare point patterns across scales and easily pick the most interesting ones without the painful process of choosing appropriate spatial weight matrix for LISA or deciding proper size and shape for quadrats. This is the reason the K-function is also called 'Multi-Distance Spatial Cluster Analysis' in the literature.

However, determining the exact scale of an observed pattern from the multiple scales examined by K-function remains problematic. In fact, most of the existing application studies simply choose an "optimal" scale to show the results without further explanation. Moreover, when applied to a pattern exhibiting local clusters, the observed K-function tends to exceed the upper significance envelope even at scales beyond the process's true scale [2]. As a result, the point patterns detected by K-function would be untrue at certain scales and further inferential conclusions built on this would be misleading as well. In order to remedy this issue, we use a variant of the K-function namely the Incremental K-function in this paper. Instead of counting the total number of events within a distance range, the Incremental K-function examines the number of events within a particular increment of distance. We will further discuss the fundamental properties of the Incremental K-function in the next section, compare them to those of the standard K-function and demonstrate the differences on several real-world geospatial datasets as well as simulated point processes. The purpose is to explore the capability of the Incremental K-function that complements the K-function methodology family for detecting point patterns at the spatial process's true scale.

3 Incremental K-Function

The Incremental K-function is a variant of the standard K-function. To differentiate with their local versions, the standard K-function and the Incremental K-function is also referred to as the global K-function and the global Incremental K-function in this paper. Let us consider a point process including n events $P = \{p_1, p_2, \dots, p_n\}$ in a study region, the global K-function is defined as

$$K(r) = \frac{1}{\rho} E(\text{number of event in } P \text{ within distance } r \text{ of an arbitrary event of } P) \tag{1}$$

here ρ is the density of events within the study region. It can be further written as:

$$K(r) = \frac{A}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n I_{i,j} \tag{2}$$

where r is the detection window radius or scale; A represents the total area of the study region; n is the total number of events in the study region; $I_{i,j}$ equals to one if the distance between event i and event j is less than r , and zero otherwise.

By decomposing the global K-function down to an individual event location i we can obtain the local K-function [9] as:

$$LocK_i(r) = E(\text{number of events in } P \text{ within distance } r \text{ of event } i) \tag{3}$$

Or it can be formulized as:

$$LocK_i(r) = \sum_{j=1}^n I_{i,j} \tag{4}$$

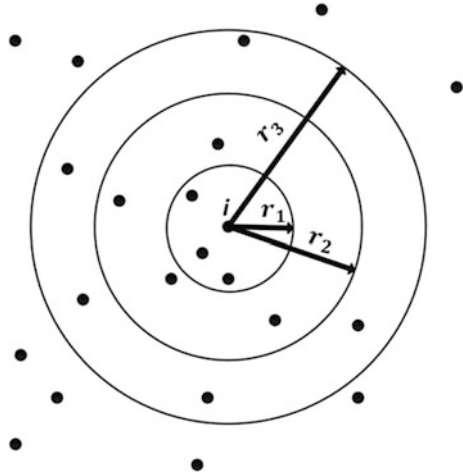
In contrast, the Incremental K-function counts the number of events within a particular increment of distance, i.e. the “donut” area from the smaller scale r_{t-1} to the current scale r_t . The only exception is when r_t is the smallest scale r_1 the Incremental K-function is then the same as the K-function. The formula of the global Incremental K-function is defined as:

$$IncK(r_t) = \begin{cases} K(r_t) - K(r_{t-1}), & t = 2, 3, \dots, \\ K(r_t), & t = 1 \end{cases} \tag{5}$$

And the local version of the Incremental K-function is defined by the same logic:

$$LocIncK_i(r_t) = \begin{cases} LocK_i(r_t) - LocK_i(r_{t-1}), & t = 2, 3, \dots, \\ LocK_i(r_t), & t = 1 \end{cases} \tag{6}$$

Fig. 1 Example of (incremental) K-function



Taking the example shown in Fig. 1, and assuming a unit density of events, the local K-function value for event location i equals 3 at the scale of r_1 ; 7 at the scale of r_2 ; and 12 at the scale of r_3 , while the local Incremental K-function value equals to 3, 4, and 5 at scale of r_1 , r_2 , and r_3 respectively. The global K-function and the global Incremental K-function will also result differently as they are comprised of their local counterparts. As illustrated by Yamada and Thill [2], a false positive error may be associated with the observed K-function of a pattern exhibiting a local clustering tendency at certain scales where the Incremental K-function will truthfully fail to detect such tendency. Because the K-function averages densities over the entire distance range, it may overshoot the true cluster size.

4 Comparison Experiments

In this section we implement a series of comparison experiments with both data simulated to represent known spatial processes and real-world geospatial data. The global and local K-function as well as the global and local Incremental K-function values are calculated according to the equations given in the previous section. Statistical inference is determined by 1,000-time Monte-Carlo simulation based on the assumption of Complete Spatial Randomness (CSR) of the point process. We adopt the significance level of 5 % for global results and 0.1 % for local results to account for test simultaneity. Edge effects are corrected through shrinking the size of the analysis area by a distance equal to the largest distance band to be used in the analysis. Only the points within the shrunk area are used to calculate K-function and Incremental K-function values, while the background point process and the simulated points remain within the original area. The implementation program is

coded in C/C++ and computed with a 64 bit PC with 4 CPUs and 16 GB RAM. The parallel computing technique OpenMP is applied to accelerate the computation process, which is particularly beneficial for the Monte Carlo simulation task.

We first conduct tests with simulated datasets. In a square region of 10,000 by 10,000 units, we simulate a series of point patterns generated according to different known processes. The total number of points is 2,500 for every point pattern. We set our detection scale from 100 units to as many as 2,500 units in order to scan the full extent of point patterns. In order to verify our approach, we start by simulating a random point pattern based on the null hypothesis namely Complete Spatial Randomness, which is used as a benchmark (as shown in Fig. 2). On the right-hand side of Fig. 2, we show the global K-function results on top and the global Incremental K-function results at the bottom. The blue line represents the value of either K-function or Incremental K-function, while the red and green lines stand for statistical significance envelopes numerically simulated through Monte Carlo process. Not surprisingly, in both charts of Fig. 2 the three lines closely overlap with each other. It indicates that neither the K-function nor the Incremental K-function of this random point pattern has escaped the expected range generated by CSR. In other words the two functions both have successfully verified the random point pattern.

We are more interested in the capabilities of these two functions to deal with clustered point patterns. Figure 3 shows a simulated clustered point patterns in which 25 independent point clusters are generated. Each of the point clusters consists of 100 points distributed according to a Bivariate Normal Distribution with standard deviation of 100, while the centers of these clusters are located on a square

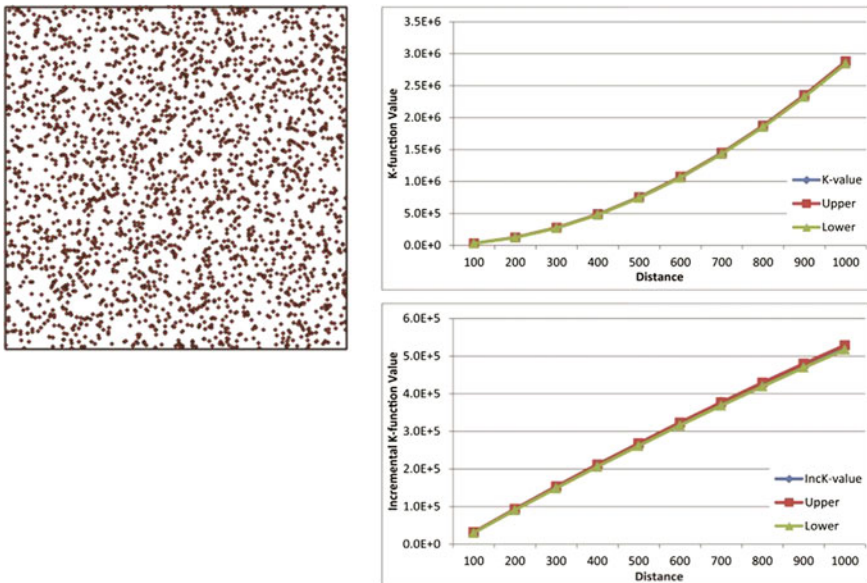


Fig. 2 Random point pattern results

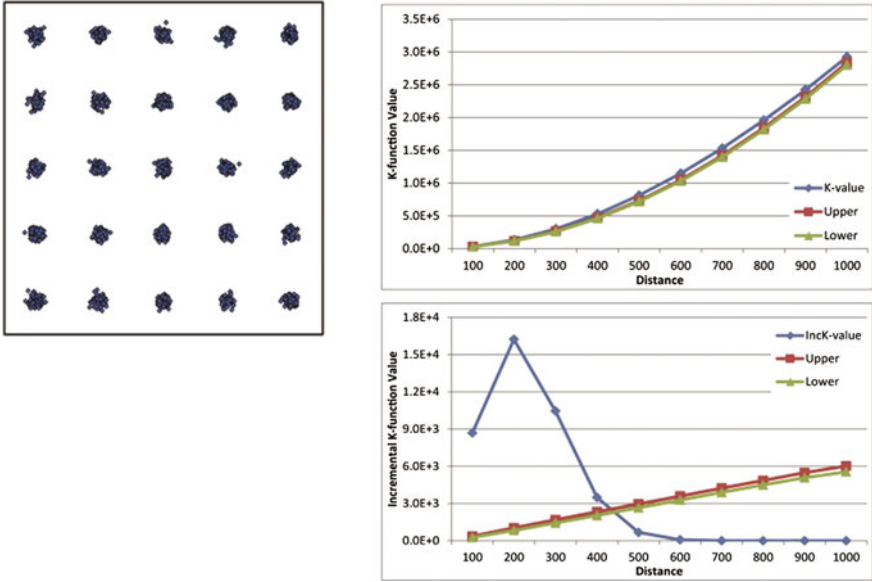


Fig. 3 Clustered point pattern (standard deviation = 100)

grid within the square study area with a distance of 2,000 units. This means that there is a 68 % probability that a point is within 100 units of its local cluster center; this probability becomes 95, 99, and 100 % for distances of 200, 300, and 400 units respectively. The results of the K-function and the Incremental K-function are quite different this time. As we can see from the two charts in Fig. 3, the observed global K-function is above the upper envelope, and thus detects a significant clustering pattern, at most scales. On the contrary, the global Incremental K-function detects a significant clustering pattern only at the first four scales and a significantly regular (or dispersed) pattern at all larger scales. The global incremental K-function has captured the properties of the point process accurately. In the bottom chart of Fig. 3, the two largest increments of ‘IncK-value’ are at the zero to 100 scale and the 100–200 scale, which correspond to the probability increasing from 0 to 68 % and then to 95 %. Beyond the 200-unit scale, the increment of the function value starts dropping as there only remains less than 5 % probability to include more points into a local cluster. Beyond the 400-unit scale, the probability drops down to zero. Correspondingly, we observe the ‘IncK-value’ in the figure drops below the lower significance envelope. In short, the incremental K-function varies consistently with the underlying point process. At the small scales the ‘IncK-value’ stays beyond the upper significance envelope indicating the local clusters dominate the global point pattern of the whole study area. With the scale increasing beyond a threshold between 400 and 500 units, the local clusters become more like single points and their grid-like centers have swayed the global point pattern to a regular one. Evidence can also be found in the results of the Incremental K-function at larger

scales. Overall, the K-function has failed to detect this changing point process. The clustered point pattern at almost all the scales indicated by the K-function is definitely inconsistent with our simulated point process. We also conduct a similar test in which only the size of local clusters has changed (to a standard deviation of 300). Figure 4 illustrates this point pattern and its two comparison results. Again the Incremental K-function has captured the nature of the point process across scales. It indicates a clustered pattern at small scales and a regular one at larger scales. The only difference with the results in Fig. 3 is the peak is shifted toward larger scales, which corresponds well to the enlarged local clusters. In contrast, the K-function result shows a clustered pattern across all scales, which would clearly be misleading at large scales.

To fully unveil the capability and characteristics of the Incremental K-function, we also experiment on more complex datasets. Figure 5 shows a group test on a two-stage clustered point dataset. The basic features of this data remain the same as the one in Fig. 4 in terms of the number of clusters as well as the size of each cluster (standard deviation of bi-normal distribution is still 300). However the center of several clusters is relocated so as to have four larger clusters in the four corners, each one being former of smaller clusters (there is still a total of 25 small clusters). Comparing the results of the two global functions, we find extreme differences again. The K-function once again results in clustered patterns regardless of the scale that is varied from 100 to 2,500 units, from which we can hardly extract any more useful information. By contrast, the results of Incremental K-function coincide with the underlying two-stage clustered point process. It indicates two and only two

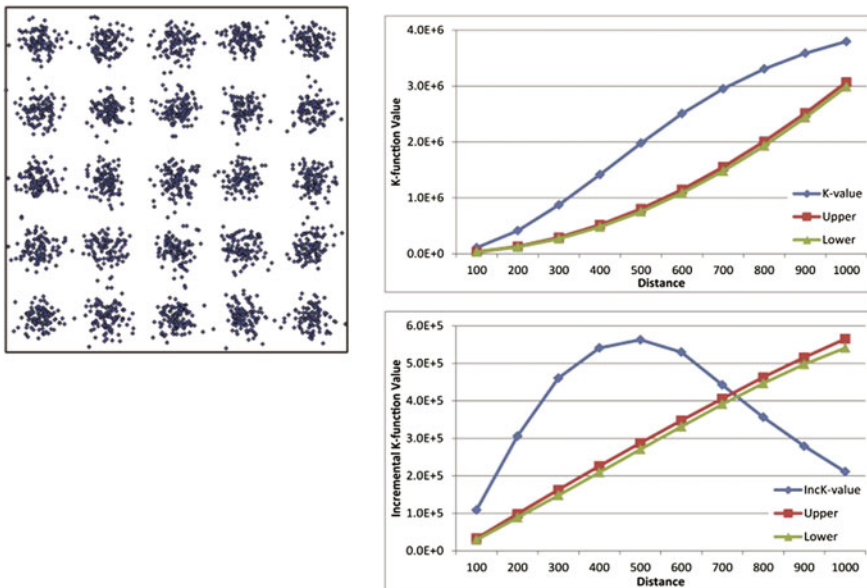


Fig. 4 Clustered point pattern (standard deviation = 300)

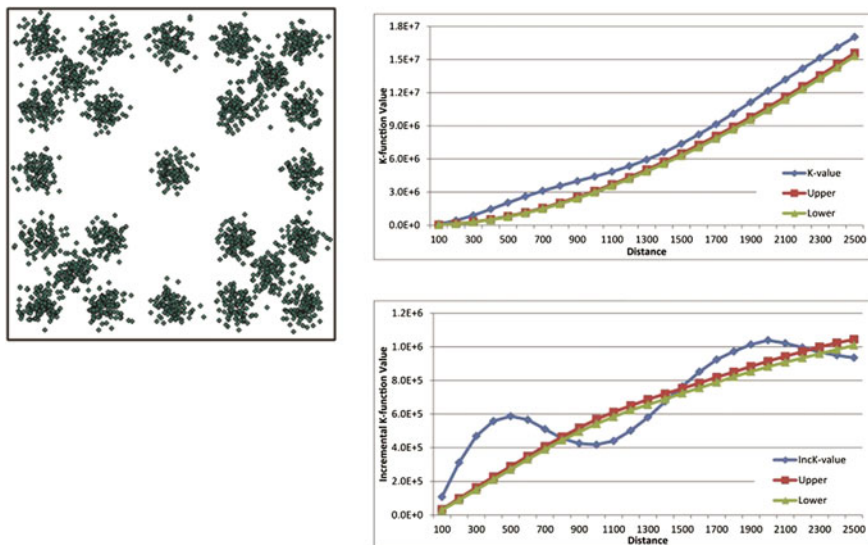


Fig. 5 Two-stage clustered point pattern (standard deviation = 300)

separate peaks of significant clustering. The first peak shows up at the same scale (500 units) as Fig. 4 because the size of local clusters remains the same. The second peak manifests itself at the scale of 2000 and demonstrates that the Incremental K-function is capable to detect those larger clusters formed of local clusters. Therefore it does neither overestimate nor underestimate the point patterns. The conclusion from this group of experiments is that the Incremental K-function can not only detect out at which scale it shows clustering, but is also able to capture the variation of cluster size across scales.

Furthermore, we carry out experiments on even more complex datasets with various sizes of clusters at random locations. The geospatial dataset depicted in Fig. 6 includes 25 local clusters, of which 10, 8, and 7 clusters are generated based on a bi-normal distribution with standard deviation of 100, 300, and 500, respectively. The results are consistent with the previous experiment. On the one hand, the K-function detects significant clustering at all scales ranging from 100 to 2000, thus committing patent false positive error. On the other hand, three separate peaks of clustering pattern corresponding to the three scales used to generate the point distributions are picked out by the Incremental K-function in spite of the random location of cluster centers. It provides further evidence of the Incremental K-function's sensitivity and accuracy with respect to point patterns across scales.

As a matter of fact, real-world situations are usually more complex than simulated ones. Therefore we have also implemented comparison studies with real-world geospatial data to backup and supplement the conclusions obtained from simulated data. The data we use in this study are the records of vehicle theft and recovery locations in Charlotte, North Carolina. Given the extremely heterogeneous

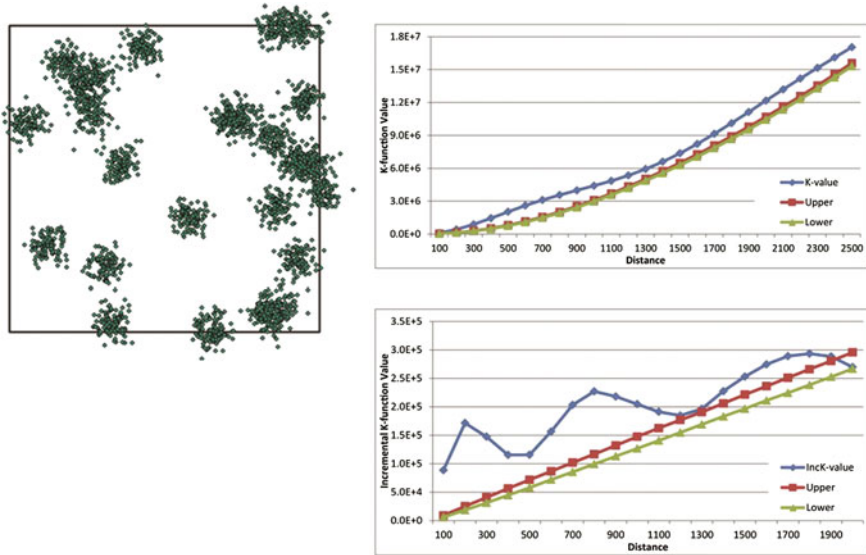


Fig. 6 Clustered point pattern with random seeds and various cluster scales

distribution of these records, the study area is narrowed down to an eight-mile by eight-mile square region surrounding the city center. According to the crime report database maintained by the Charlotte-Mecklenburg Police Department (CMPD), there were 1,832 vehicles stolen within this region from January 1st, 2012 to December 31st, 2013, of which 995 have been recovered somewhere else in the city. Recovery locations show a more varied point pattern across scales than the theft locations. Therefore we use these 995 vehicle recovery locations to illustrate the properties of the Incremental K-function and explore its practical usefulness as well.

Figure 7 shows the study region and the auto recovery locations for the stated period of analysis. The right-hand side panels present the global K-function and the global Incremental K-function charts. The smallest detection scale is set at 0.05 mile, i.e. 80 m, and 20 times that much as the largest possible scale. The K-function shows significant clustering over the entire range of scales as the ‘K-value’ stays well above the upper significance envelope and smoothly departs from it to indicate an even more significantly clustered pattern for larger scales. On the other hand, the ‘Inck-value’ in the bottom chart is quite uneven across scales, although it stays above the upper envelope at all scales. Its several up-and-down periods are believed to reflect the heterogeneous nature of this real-world dataset. A first peak shows up at the smallest possible scale and also exhibits the greatest departure from the significance envelopes. From this, we can conclude that there exist a number of auto recovery location clusters across the study region and that these clusters have very compact sizes as their radii are 80 m or less. Through further investigation we found these small clusters match specific geographical features such as car dismantlers or deserted parking lots where cars would be abandoned. Besides the global maximum

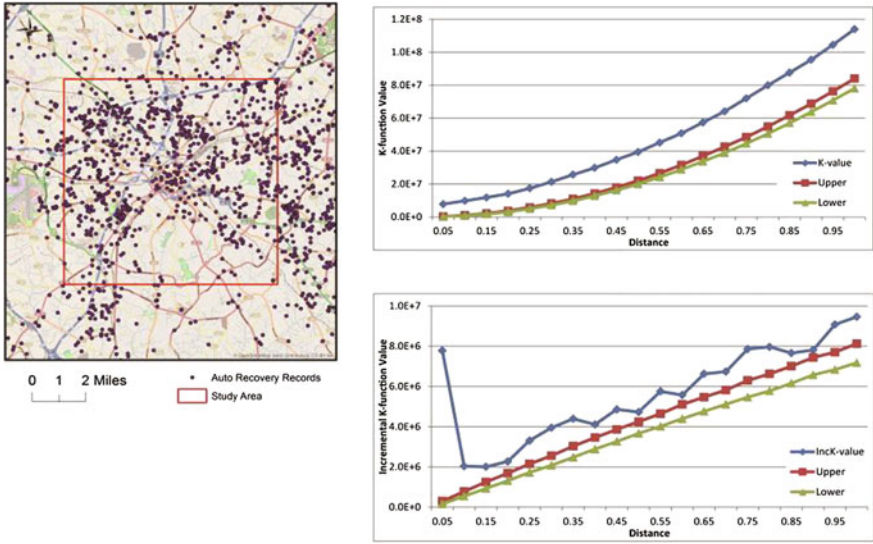


Fig. 7 Global results of Charlotte auto-recovery data

at the 0.05 scale, other local maxima in the Incremental K-function are found at 0.35 and 0.8 mile, as well as several less prominent ones. These could be explained as thieves’ favorite car dumping areas, for example unsupervised neighborhoods or the vicinity of the airport.

Beside the analysis of the global spatial pattern in the car recovery locations, we also apply the local version of two K-functions to investigate clusters locally. We implement the local functions following the same process as the global ones, but change the significance level to 0.1 %. While the local analysis can be conducted at any and all discrete location in the study region, for the sake of the illustration we pick out one specific record located to the Northeast of the city center. Figure 8 includes a map of this localized area and the plots of two local K-functions. This time the local Incremental K-function returns two significant peaks with one at the 0.05 mile scale and the other at 0.75 mile. On the map at left-hand side, the scales of 0.05, 0.1, 0.7, and 0.75 mile are highlighted using red circles. The smallest circle encompasses 17 events, while the second one includes no additional one. Comparing the results of two local K-functions, the ‘Local IncK-value’ has captured this situation rather crisply as it indicates clustering only at the first scale. Conversely, the ‘Local K-value’ shows clustering at both of the first two scales and even beyond (although at a decreasingly level of significance), which is a clear case of false positive error and gives a misleading message that the local cluster keeps enlarging. The ring-like area between 0.7 and 0.75 mile scale includes 13 neighboring events and it results in another significant peak in the ‘local IncK-value’. Examination of the map of events around the selected focal point suggests that no clustering tendency is detectable between these two scales; this is corroborated by

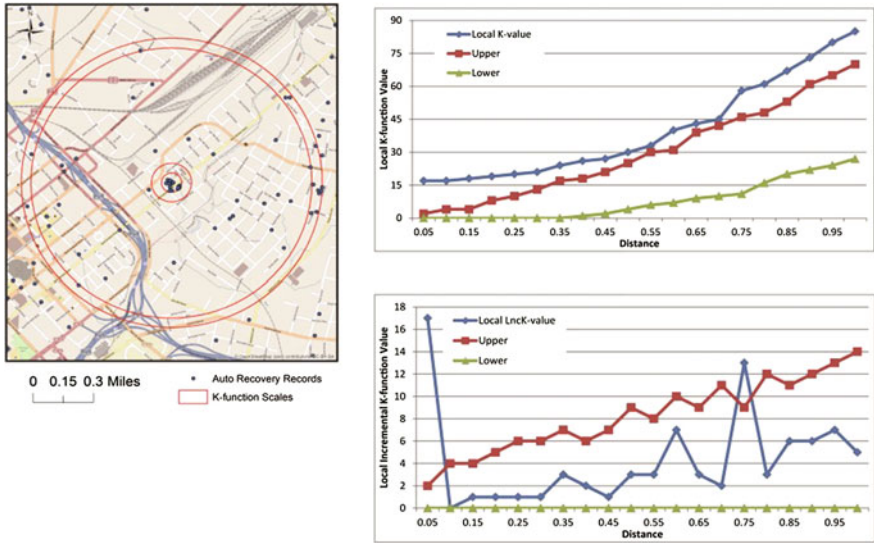


Fig. 8 Local results of Charlotte auto-recovery data

the local Incremental K-function, while the local K-function overdetects a clustering tendency. By looking at the global and local results side by side, it is clear that local point patterns have contribution to the global pattern. Both the global and local Incremental K-functions are capable to reflect the heterogeneous point process and their results can certainly give more meaningful guidance than the original K-functions.

5 Conclusions

The K-function is regarded as one of the most effective methods to detect non-random tendencies in a geospatial distribution of points and particularly high concentrations of point events. Although it is designed as a ‘Multi-Distance Spatial Cluster Analysis’, we argue that it cannot reflect the cross-scale changes of point pattern very well. Instead, we bring up again one of its variant, namely the Incremental K-function, as a solution to this problem. In this paper, we presented the results of a series of comparison studies on both simulated datasets and a real-world dataset. The results from simulated datasets indicate that the Incremental K-function can pick out the scales at which it shows most significantly clustering patterns without the false positive errors that can be so pervasive with the K-function. These peak scales coincide with the true scales designed in our data simulation processes. In addition, the Incremental K-function is also capable to deal with more complex situations such as two-stage clusters and random located

clusters of various scales. Meaningful information about how the point pattern varying across scales can be extracted. In contrast, the standard K-function can only offer us a coarse clustering pattern from scale to scale. Given the controlled conditions of the experiments done on simulated point patterns with known properties, it has been demonstrated that the K-function is afflicted by false positive error flaws and incapable of capturing the true scale of point processes.

Moreover, the results from the Charlotte vehicle recovery dataset provide real-world evidence that the Incremental K-function can accurately reflect the underlying heterogeneous point process, and that it does so more reliably than the K-function. Practical usefulness can also be obtained. For instance, the very compact cluster size (80 m) directs that there exist some individual facilities or locations that concentrate a significant number of stolen vehicles. The local Incremental K-function can serve to pinpoint these locations on the city map for further investigation. To conclude, we hope this work can bring the Incremental K-function to scholars' attention as an effective hot-spot detection method especially dealing with complex spatial point processes.

References

1. Fotheringham AS, Zhan FB (1996) A comparison of three exploratory methods for cluster detection in spatial point patterns. *Geogr Anal* 28(3):200–218
2. Yamada I, Thill J-C (2007) Local indicators of network-constrained clusters in spatial point patterns. *Geogr Anal* 39(3):268–292
3. Getis A, Boots BN (1978) Models of spatial processes: an approach to the study of point, line and area patterns, vol 198. Cambridge University Press, Cambridge
4. Laney D (2001) 3D data management: controlling data volume, velocity and variety. META Group Research Note, 6
5. Miller HJ (2010) The data avalanche is here. Shouldn't we be digging? *J Reg Sci* 50(1):181–201
6. Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. *Econ Geogr* 46(2):234–240
7. Waller LA (2009) Detection of clustering in spatial data. *The Sage Handbook of Spatial Analysis*. London, pp 299–320
8. Ripley BD (1976) The second-order analysis of stationary point processes. *J Appl Probab* 13:255–266
9. Getis A, Franklin J (1987) Second-order neighborhood analysis of mapped point patterns. *Ecology* 68:473–477
10. Fotheringham AS (1997) Trends in quantitative methods I: stressing the local. *Prog Hum Geogr* 21(1):88–96
11. Openshaw S, Charlton M, Wymer C, Craft A (1987) A mark 1 geographical analysis machine for the automated analysis of point data sets. *Int J Geogr Inf Syst* 1(4):335–358
12. Besag J, Newell J (1991) The detection of clusters in rare diseases. *J Roy Stat Soc A (Stat Soc)* 154:143–155
13. Anselin L (1995) Local indicators of spatial association—LISA. *Geogr Anal* 27(2):93–115
14. Getis A, Ord JK (1992) The analysis of spatial association by use of distance statistics. *Geogr Anal* 24(3):189–206

15. Ord JK, Getis A (1995) Local spatial autocorrelation statistics: distributional issues and an application. *Geogr Anal* 27(4):286–306
16. Aldstadt J, Getis A (2006) Using AMOEBA to create a spatial weights matrix and identify spatial clusters. *Geogr Anal* 38(4):327–343
17. Widener MJ, Crago NC, Aldstadt J (2012) Developing a parallel computational implementation of AMOEBA. *Int J Geogr Inf Sci* 26(9):1707–1723
18. Tang W, Feng W, Jia M (2014) Massively parallel spatial point pattern analysis: Ripley's K function accelerated using graphics processing units. *Int J Geogr Inf Sci* 28(5):1107–1127 (Accepted)
19. Okabe A, Boots B, Satoh T (2010) A class of local and global k functions and their exact statistical methods. In: *Perspectives on spatial data analysis*. Springer, Berlin, pp 101–112
20. Boots B, Okabe A (2007) Local statistical spatial analysis: inventory and prospect. *Int J Geogr Inf Sci* 21(4):355–375