

Extending point pattern analysis to the temporal comparison of point patterns

Li Li

Department of Political Science and Geography Old Dominion University Norfolk, Virginia

Abstract. Point pattern analysis has been widely applied to studies in many disciplines. Different approaches of point pattern analysis have been developed to study different aspects of point patterns. However, there is a gap in the current point pattern analysis. Few approaches have been developed to compare two point patterns representing the same objects. This study proposes a simulation based approach to compare the clustering levels of two point patterns. The results indicate that this approach can provide sufficient details on differences of the clustering levels of two point patterns at multiple scales.

Keywords: Point Pattern; Monte Carlo; K-function; Spatial Statistics

1. Introduction

Spatial analysis has been increasingly applied to the data in many scientific disciplines [1]–[3]. Point pattern analysis is one of the important topics in spatial analysis, since points are the simplest representation of spatial objects in a geographic space, and many objects have to be studied as points for the lack of data on their dimensions.

A collection of points representing the same objects is often referred as a point pattern. Point pattern analysis attempts to identify patterns that are indicative of underlying processes for point patterns. Most of the point pattern analysis approaches classify a point pattern as regular, random or clustered [4]. Among these three types, the clustered pattern helps most for a study to raise hypotheses for underlying processes. The reason is that this type indicates points are “closer” than they would be in a reference pattern that often represents the absence of underlying processes [5]. Spatial randomness and regular pattern attract less attention, since it was assumed to indicate no underlying process.

The temporal aspect of a study is often important for identifying underlying processes. The differences in clustering level of point patterns may reflect the changes in the underlying processes. Tools have been developed to analyze the temporal and spatial clusters [6]. However, to use these tools, the point data have to be spatially and temporally extensive. In some studies, the data are not temporally extensive due to logistic reasons (e.g., limited funding supports) [7].

Approaches that can be used compare the clustering levels of two or more point patterns are needed. There are many challenges associated with the development of such approaches. First, a variety approaches are available to quantify the clustering levels of point patterns [8]–[10]. Finding the right approach and extending it for the comparison of two or more point patterns is challenging. Second, from a probabilistic point of view, a clustering point pattern is just one realization of a point process [10]. The deviation of the clustering level of one point pattern from another could be a result of the stochastic deviation of one point process. Confirming the statistical significance of the detected difference in the clustering levels of two point patterns is also a challenging task. Finally, two point patterns, as well as their reference point patterns often consist of different number of points. This may complicate the point pattern comparison.

In this study, a simulation approach is proposed to compare the clustering levels of two point patterns. The next three sections first provide a brief description of point pattern analysis. Then an approach to compare point patterns is proposed. A case study is provided to illustrate the proposed approach. In the end, the results are presented and discussed.

2. Point pattern analysis

2.1. First-order and second-order properties

A point pattern is often characterized in terms of its first-order and second-order properties. The first order properties concern how the expected value (mean or average) of the process varies across space [4]. The second-order properties describe the correlation between values of the points at different regions [11]. The advantage of point pattern analysis approaches that focus on the second-order properties is that they can identify the inter-point interactions, which has been a primary interest for many ecological studies.

The K-function [10] is also based on the second-order properties and it measures the clustering pattern of points at multiple scales. Over the last several years, Ripley's K-function has been widely used and generally praised for its ability to count for the impact of scale of the analysis. Scale is an important concept because virtually all ecological processes are scale dependent and their characteristics may change across scales [12]. This method has been used in studying the spatial distribution patterns of various objects [13]-[14]. The following subsection provides detailed explanation of this method.

2.2. K-function

As shown in Function (1), K-function ($K(t)$) characterize the clustering of a point pattern at different distance scales (t). The essential idea of this function is to compare the observed number of points within distance t of an arbitrary point to an expected number. To calculate K function at distance t , hypothetical circles (with radius t , often called a lag distance) were placed around each point location, and the average numbers of points that lie within those circles were calculated. When a point is close to the edge of the region, a hypothetical circle may occupy area what is outside the study region. The following function is an edge effect corrected K-function.

$$K(t) = \frac{A}{N^2} \sum_i \sum_j \frac{I_{d(ij)} w_{ij}}{w_{ij}}, \quad \text{for } (i \neq j), \quad (1)$$

where A is the area of the study region, N is the number of points in the study region, $I_{d(ij)} = 1$ if the distance between point i and point j is shorter than t , otherwise $I_{d(ij)} = 0$, w_{ij} is used to take into account of the edge effect. A widely accepted algorithm of w_{ij} is to calculate the portion of the hypothetical circle that fall within the study region. This edge-corrected version of K-function and provides an approximately unbiased estimator of K-function. This procedure is repeated with increasing t , and $K(t)$ is calculated and recorded at every selected distance t .

A process with no spatial dependence between points (regular pattern) has $K(t)$ equal to πt^2 ; a clustering pattern has $K(t) > \pi t^2$; and a pattern with points expelling each other has $K(t) < \pi t^2$. Depending on the relationship between $K(t)$ and πt^2 , the $K(t)$ function can be transformed into an L-function (Equation (2)), which normalizes the K-function [15].

$$L(t) = \sqrt{\frac{K(t)}{\pi}} - t \quad (2)$$

The L-function value range is $[-\infty, +\infty]$. A regular point pattern has L-function equal to zero. If a L-function value is larger than zero at one scale, it indicates the existence of clustering pattern at this scale.

Statistical significance of the result of K-function analysis is usually evaluated via the Monte Carlo simulation of a null model [10]. The common null model is based on the homogeneous Poisson process. The purpose of Monte Carlo Methods in K-function analysis is to combine the data from an arbitrary number of replications to obtain information over a wide range of L-function values. The 5% and 95% envelope which enclose the 90% of the L-function values from the simulated point patterns are often obtained. If values of the observed point pattern are contained by this envelope, this indicates that the observed point pattern is a random point pattern with 95% confidence.

3. A simulation based approach for point pattern comparison

The K-function results simply label a pattern as either "clustering" or "non-clustering", and offer little insights in the differences in two clustered point patterns. The simplicity in the results is because of the

simple Monte Carlo simulation process. In this study, an extended point pattern analysis is proposed and illustrated in this section.

3.1. Data

The study area is a 1000 m by 1000 rectangular area. To demonstrate the proposed approach, three point patterns are created with known Matern cluster point processes to serve as examples: point pattern A (Figure 1A), point pattern B (Figure 1B), and point pattern C (Figure 1C). The benefit of using the point pattern with a known point process is that it enables an informed demonstration of the proposed approach. The Matern cluster point process is a widely recognized hierarchical clustering point process [16]. It consists of a mother process and a child process. To create a Matern cluster point pattern, a homogeneous Poisson point pattern is firstly created. Then, certain number of points ($n=N$) is randomly selected from the created point pattern. These points are defined by a process with the intensity λ_m , and they are referred as mother points. Around each mother point, the points within certain distance from the mother are randomly selected from the points created in the first step. These points are called child points, which have the intensity λ_c . The intensity of a Matern clustering process depends on both λ_m and λ_c . Its intensity function (λ_{MC}) is the following:

$$\lambda_{MC} = R^2 \pi \lambda_m \lambda_c \quad (3)$$

As shown in Fig. 1, the point pattern A has 600 points, while the point patterns B and C each have 400 points. The concept of scale is incorporated into this study. The point pattern A and B have the same scale. The maximum distance from the child points to their mother points is 50 m. This distance is 45 m for point pattern C. The intensities of point pattern A and C should be similar at the maximum distance. The λ_c of the point pattern B should be larger than that of the point pattern A and C for these two point patterns.

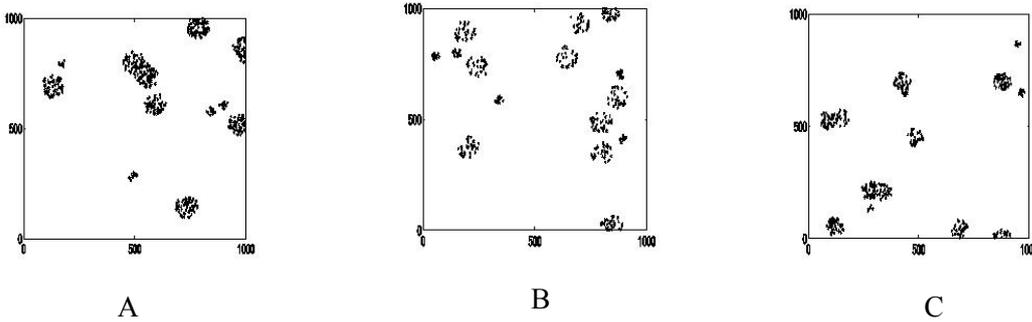


Fig. 1: Point patterns A (600 points) (A), B (400 points) (B), and C (400 points) (C).

3.2. Analysis

For each point pattern, the L-functions (Equation (1) and Equation (2)) are calculated for 15 arbitrarily selected scales. These scales range from 20 m to 150 m with a 10 m increment. For each scale, the difference in the L-functions of three pairs are obtained: point pattern A vs. point pattern B, point pattern B vs. point pattern C, and point pattern A vs. point pattern C. When the Monte Carlo approach was implemented in the K-function analysis, one replication usually consists of two steps: in step one, one point pattern is created by randomly positioning points in the study area; in step two, K-function is calculated for this point pattern.

To investigate the statistical significance in the differences of L-function values of different point patterns, a three-step Monte Carlo simulation is proposed. In the first step, three random point patterns are created. One of these three point patterns has 600 points and two of them have 400 points. These three point patterns are all generated by the homogeneous Poisson point process. They serve as the null models for the previous created three clustering point patterns. In the step two, K-function, via L-function, is calculated for every point pattern. In the step three, differences in the L-functions are calculated between any two of three point patterns representing null models. It is typically to have 99 replications of K-function calculations to get a reliable estimation. One important principle of the Monte Carlo simulation is that the number of replication should increase when there are parameters involved in each replication. In this study, 400 replications are carried out in the Monte Carlo simulation [10]. All the calculations, including Monte Carlo simulation are implemented in a Matlab environment.

4. Results and Conclusion

4.1. L-function values Data

As explained in the previous section, for each of the three point patterns (A., B and C), L-function values are calculated for 15 scales. The values of the point patterns A, B and C are displayed against their 95% confidence envelopes in Figs. 2P1A, 2P1B and 2P1C. It is clear that all three point patterns are clustered at all the selected scales, since the calculated L-function values are well above their envelope. For point patterns A and C, the L-function values peak around distance 60 m, and decrease rapidly with increasing distances, indicating that this is the point patterns achieve their highest level of clustering at 75 m. For point pattern B, the highest level of clustering achieved around 80 m, and decrease slowly with increasing distances. The direct observation indicates that the point patterns A and C have similar level of clustering, and they are more clustered than the point pattern B. This is consistent with the true nature of these two patterns, since point patterns A and C were expected to have similar intensities.

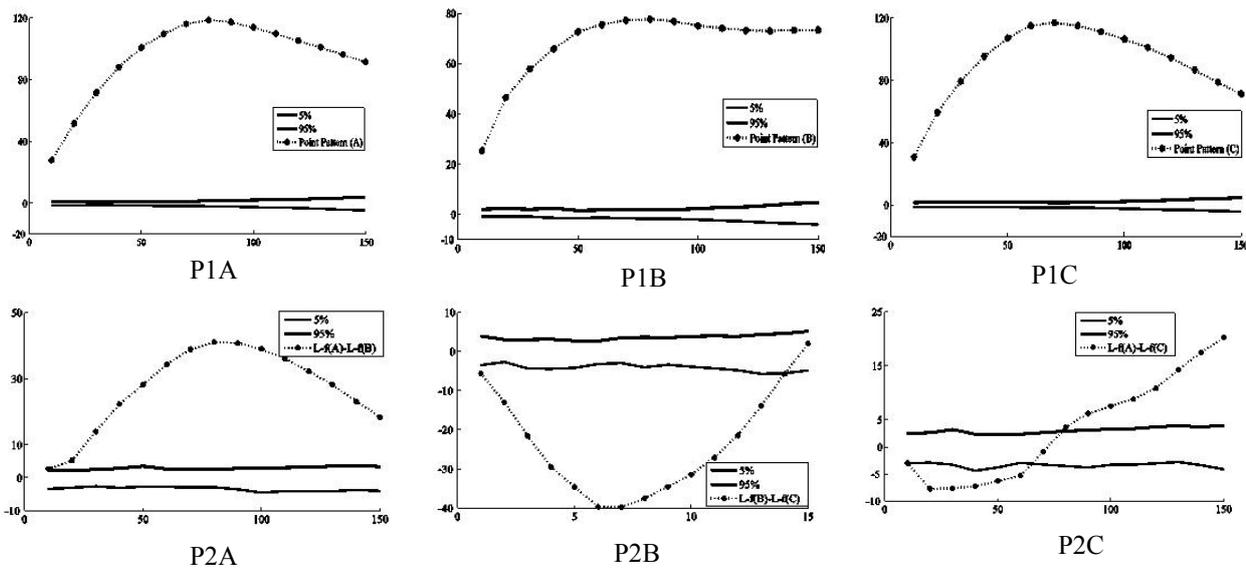


Fig. 2: Part 1: L-functions and their 95% confidence envelope for Point patterns A (600 points) (P1A), B (400 points) (P1B), and C (400 points) (P1C). Part 2: Differences in the L-functions of three pairs of point patterns: A-B (P2A), B-C (P2B) and A-C (P2C).

4.2. Difference in L-function values

To further investigate the differences in the clustering levels of two point patterns, the differences of L-function values of three pairs were compared against the 95% confidence envelopes (Fig. 2). As shown in the part 2 of Fig. 2, the differences in the L-function values of any point patterns are generally larger than the 95% confidence envelopes of such between two random Poisson point patterns. The ranges of the difference values are similar for two of the pairs (point patterns A and B, and point patterns A and C), and they are larger than such range for the other pair (point patterns B and C). As shown in Figs. 2P2A and 2P2C, even these two pairs include point patterns with different numbers of points, their confidence envelopes of the difference in the L-functions do not differ that much from each other.

As expected, the results show a significant difference in the clustering levels between point patterns A and B and between point patterns B and C. This indicates that the proposed approach can reveal the detailed differences in two point patterns. As expected, the results also show that point patterns A and C have similar clustering levels from approximately 60 m to 80m. This further indicates the proposed approach can identify the scales that two point patterns have similar level of clustering.

4.3. Conclusions

Based on Figure 2, several conclusions have important implications. First, K-function is a valuable tool for the comparison of the clustering levels of two point patterns. Second, the direct visual comparison of the clustering measures of two point patterns may not provide a faulty answer. But it could not reveal detailed

differences in the clustering levels. It also cannot confirm the statistical significance of such a comparison. Finally, the number of points in null models does not seem to have significant impact of the Monte Carlo simulation envelopes. This implied that although the reference point patterns in different point pattern analysis are different, the Monte Carlo simulation envelopes could be similar. The variation in the envelopes may be caused by insufficient number of replications.

One of the limitations in the Monte Carlo analysis is that it is computationally extensive [17]. Most of the existing softwares that calculate L-function provide limited flexibility in their simulation procedures. For instance, it often limits number of replications or number of points that are to be examined. The limitation of this proposed approach is that the software supports this approach is lacking and programming is required to implement this approach. This approach also requires longer running time than the typical simulation approach used in the current point pattern analysis softwares.

5. References

- [1] F. Fleischer, M. Beil, M. Kazda, and V. Schmidt, "Case studies on spatial point processes models," *Lecture Notes in Statistics*, vol. 185, pp. 235–60, 2006.
- [2] R. Foxall and A. Baddeley, "Nonparametric measures of association between a spatial point process and a random set, with geological applications," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 51, no. 2, pp. 165–182, 2002.
- [3] T. Mattfeldt, S. Eckel, F. Fleischer, and V. Schmidt, "Statistical analysis of reduced pair correlation functions of capillaries in the prostate gland," *Journal of Microscopy*, vol. 223, no. Pt 2, pp. 107–19, 2006.
- [4] B. Ripley, *Spatial Statistics*, B. Ripley, Ed. New York: Wiley, 1981.
- [5] L. Anselin, "Review of cluster analysis software," *Urbana*, vol. 51, p. 61801, 2003.
- [6] M. Kulldorff, "A spatial scan statistic," *Communications in Statistics- Theory and Methods*, vol. 26, no. 6, pp. 1481–1496, 1997.
- [7] S. Martens, D. Breshears, C. Meyer, and F. Barnes, Scales of Above- Ground and Below-Ground competition in a Semi-Arid woodland detected from spatial pattern," *Journal of Vegetation Science*, vol. 8, no. 5 pp. 655–664, Nov. 1997.
- [8] R. Thomas, *An Introduction to Quadrat Analysis*. Geo Abstracts Ltd., *University of East Anglia*, 1977.
- [9] D. Moore and T. Carpenter, "Spatial analytical methods and geographic information systems: Use in health research and epidemiology," *Epidemiologic Reviews*, vol. 21, pp. 143–161, 2002.
- [10] B. Ripley, "The second-order analysis of stationary point processes," *Journal of Applied Probability*, vol. 13, pp. 255–266, 1976.
- [11] A. Gatrell, T. Bailey, P. Diggle, and B. Rowlingson, "Spatial point pattern analysis and its application in geographical epidemiology," *Transactions of the Institute of British Geographers*, vol. 21, pp. 256–274, 1996.
- [12] R. Levine, A. Peterson, and M. Benedict, "Geographic and ecologic distributions of the anopheles gambiae complex predicted using a genetic algorithm," *American Journal of Tropical Medicine and Hygiene*, vol. 70, pp. 105–109, 2004.
- [13] [13] W. Khaemba, "Use of gis for a spatial and temporal analysis of Kenyan wildlife with generalized linear modelling," *International Journal of Geographical Information Science*, vol. 14, pp. 833–853, 2002.
- [14] O. Munyekenye, A. Githeko, G. Zhou, E. Mushinzimana, N. Minakawa, and G. Yan, "Plasmodium falciparum spatial analysis, western Kenya highlands," *Emerging infectious disease*, vol. 11, no. 10, pp. 1570–1577, 2005.
- [15] J. Besag and P. J. Diggle, "Simple monte carlo tests for spatial pattern," *Applied Statistics*, pp. 327–333, 1977.
- [16] D. Stoyan, U. Bertram, and H. Wendrock, "Estimation variances for estimators of product densities and pair correlation functions of planar point processes," *Annals of the Institute of Statistical Mathematics*, vol. 45, no. 2, pp. 211–221, 1993.
- [17] R. Swendsen and J. Wang, "Nonuniversal critical dynamics in Monte Carlo simulations," *Physical Review Letters*, vol. 58, no. 2, pp. 86–88, 1987