Characterising Linear Point Patterns

J.A. Downs, M.W. Horner

Department of Geography, The Florida State University, 323 Bellamy Bldg., Tallahassee, FL 32306 USA
Telephone: 1+ 740 603 3729
Fax: 1+ 850 644 5913
Email: jad05m@fsu.edu, mhorner@fsu.edu

1. Introduction

Point pattern analysis often involves delineating the areal region occupied by a set of sample points (Galton and Duckham, 2006). These areal boundaries are generally derived using either polygon, or hull-based, methods or statistical techniques. Kernel density estimation (KDE), which produces continuous estimates of the spatial intensity of a point pattern (Silverman, 1986), is widely used in many applications, including crime hot spot analysis (Bailey and Gatrell, 1995) and wildlife home range estimation (Worton, 1987).

While KDE is widely used to characterise point patterns, the technique has received recent criticism in the literature, particularly when applied to wildlife tracking data. KDE is sensitive to selection of the bandwidth (Seaman and Powell, 1996; Gitzen and Millspaugh, 2003), which is generally chosen using a least-squares cross validation procedure (LSCV; Silverman, 1986). Several authors have argued that KDE using LSCV bandwidth selection fails to accurately delineate point pattern boundaries, particularly when sample sizes are small (Seaman et al., 1999; Girard et al., 2002), when sample points are spatially or serially autocorrelated (de Solla et al., 1999), or when the distribution of points is linear or contains a large amount empty space (Blundell et al., 2001; Hemson et al., 2005). While sample size and autocorrelation can be addressed by choice of sampling design (Borger et al., 2006; Katajisto and Moilanen, 2006), impact of the shape of the distribution has not been adequately considered in the literature.

Examination of a fundamental assumption of KDE can explain why the overall shape of a point distribution can impact the accuracy of the estimates. Namely, KDE generates continuous measures of spatial intensity using a weighted distance function, typically a Gaussian kernel, where distances are measured in Euclidean space. Many processes, however, operate in network space (Miller, 2005), resulting in linear or irregularly shaped spatial distributions. If a set of points is generated by a network-related phenomenon, then using KDE based on Euclidean distances cannot be expected to yield accurate results, regardless of the method used to select the bandwidth. Rather, in these cases, we suggest an approach where KDE is performed using network distances. Borruso (2005) used a similar approach to identify clusters of node intersections in a street network, although he did not apply a distance weighting function as is used in KDE. Additionally, the use of network distances has been recently applied to *K*-functions (Yamada and Thill, 2004). This paper describes a preliminary implementation of network-based kernel density estimation (NKDE).

2. Methodology

The kernel density estimate (f) at the point x is formulated as:

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \qquad (1)$$

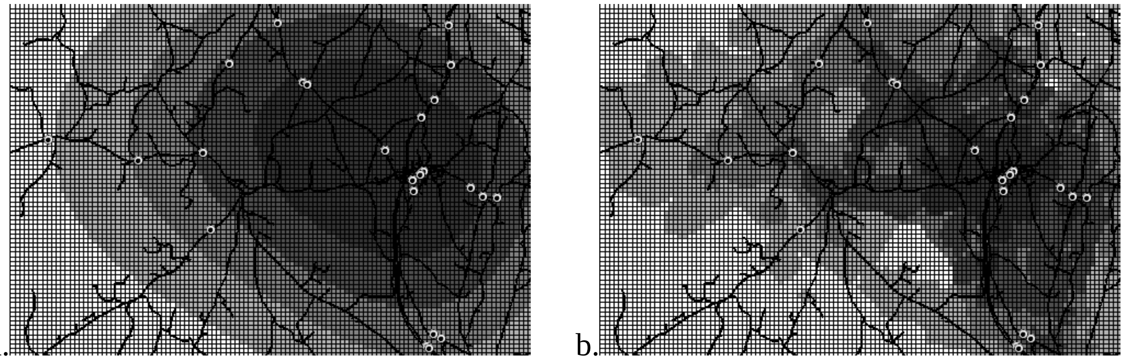
where sample size n contains points X_1, X_2,…, X_n, h is the bandwidth, and $K(y)$ is the kernel (Silverman, 1986). NKDE is implemented in a GIS environment similar to KDE, except distances between the grid points where the spatial intensity is to be evaluated and the individual event points are calculated along a specified network. Network distances are calculated by first connecting each evaluation grid point to the nearest node in the network and then computing the distance to each sample point. A distance weighting kernel is then applied to the network distances and substituted into the density function f. Any distance weighting function can be used, although we selected the Gaussian kernel. For comparison, we computed both KDE and NKDE estimates for two scenarios, one where a known network could be specified and another where the network was unknown. We calculated the density estimates using TransCAD GIS and mapped the results using ArcGIS 9.1.

First, we characterised the pattern of traffic accidents in a portion of Orleans County, Vermont, USA, which contained 27 unique traffic accident sites during 1998-2001. We obtained digital traffic accident and road data from the Vermont Center for Geographic Information. We applied NKDE using road network distances and specified a bandwidth of 4 km for both the NKDE and KDE analyses.

Next, we applied KDE and NKDE to characterise the area occupied by an individual animal, termed the home range. We used a hypothetical dataset of 50 sample points, similar to what would be obtained during a radio-tracking study (Amstrup et al., 2004). Since the exact paths travelled by wildlife are not known from typical tracking studies, we estimated a travel network by creating a two-dimensional triangular irregular network (TIN) from the set of sample points. The TIN approximates paths used by wildlife in their home range, as paths between neighbouring points are represented by straight lines, while paths between more distant points are represented as pathways that connect intermediate points. For illustration, we computed both the KDE and NKDE home ranges using a single arbitrary bandwidth.

3. Results and Analysis

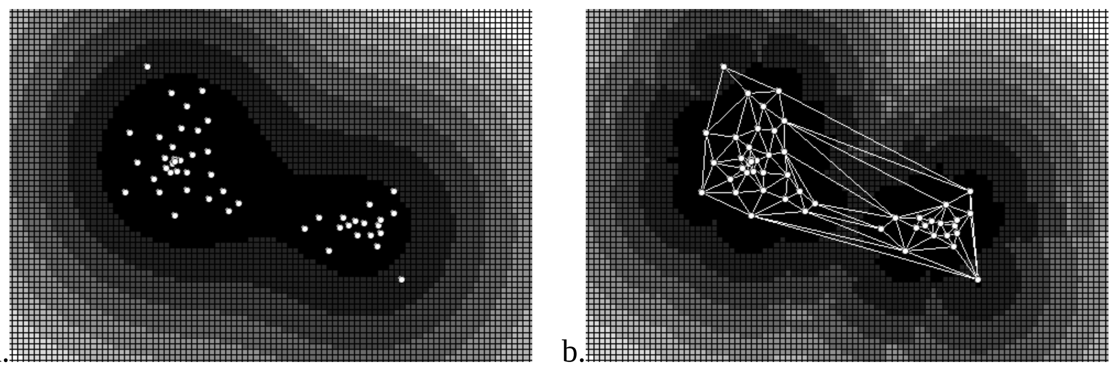
The characterised patterns of traffic accident densities differed markedly between KDE and NKDE applications (Figure 1). Use of Euclidean distances produced a single high-intensity core area, with intensity decreasing radially outward. However, the network analysis produced highest densities along road segments with several accident locations, with intermediate levels corresponding to connecting roads or those with fewer accidents. These results illustrate that NKDE provided a much more realistic characterisation of traffic accident density than Euclidean-based KDE, given that the accidents were associated with the road network.



a. b.

Figure 1. Euclidean (a) and network-based (b) kernel density estimates for 27 traffic accidents in Vermont, USA.

Results of the wildlife home range analysis were similar to the traffic example, although the differences between the KDE and NKDE results were not as dramatic (Figure 2). The NKDE analysis produced two distinct core areas separated by a lower intensity area. Using the same quantile classification, the Euclidean-based analysis did not distinguish between the two clusters of points and the empty space between them. Additionally, the NKDE estimate appeared to generate a better fit to the shape of the point distribution, containing less empty space within the high intensity area.



a. b.

Figure 2. Euclidean (a) and TIN-based (b) kernel density estimates of the home range of an animal from 50 sample locations.

The presented analyses were sensitive to both the specified networks and bandwidths. While the road network was an obvious choice for the accident data, the TIN served as an approximation of the animal's network of travel paths. While the TIN generated will vary with different sets of sample points, we suspect that NKDE estimates will improve as the number of points increases, as the TIN should better represent actual paths travelled. While we did not conduct the analyses using LSCV-selected bandwidths—as we have yet to develop such a procedure—the chosen bandwidths still served to illustrate NKDE. The preliminary results suggest that NKDE provides an improved method for characterising point patterns generated in network space.

4. References

AMSTRUP S., MCDONALD T., and DURNER G., 2004, Using satellite radiotelemetry data to delineate and manage wildlife populations, Wildlife Society Bulletin, 32, pp. 661-679.

BAILEY, T., and GATRELL, A., 1995. Interactive Spatial Data Analysis (Harlow: Longman).

BORGER L., FRANCONI N., DE MICHELE G., GANTZ A., MESCHI F., MANICA A., LOVARI S., and COULSON T., 2006, Effects of sampling regime on the mean and variance of home range size estimates, Journal of Animal Ecology, 75, pp. 1493-1405.

BLUNDELL G., MAIER J., and DEBEVEC E., 2001, Linear home ranges: Effects of smoothing, sample size, and autocorrelation on kernel estimates, Ecological Monographs, 71, pp. 469-489.

BORRUSO G., 2005, Network density estimation: Analysis of point patterns over a network, Lecture Notes in Computer Science, 3482, pp. 126-132.

DE SOLLA S., BONDURIANSKY R., and BROOKS R., 1999, Eliminating autocorrelation reduces biological relevance of home range estimates, Journal of Animal Ecology, 68, pp. 221-234.

GALTON A., and DUCKHAM, M., 2006, What is the region occupied by a set of points? Lecture Notes in Computer Science, 4197, pp. 81-98.

GIRARD I., OUELLET J., COURTOIS R., DUSSAULT C., and BRETON L., 2002, Effects of sampling effort based on GPS telemetry on home-range size estimations, Journal Of Wildlife Management, 66, pp. 1290-1300.

GITZEN R., MILLSPAUGH J., 2003, Comparison of least-squares cross-validation bandwidth options for kernel home-range estimation, Wildlife Society Bulletin, 31, pp. 823-831.

HEMSON G., JOHNSON P., SOUTH A., KENWARD R., RIPLEY R., and MACDONALD D., 2005, Are kernels the mustard? Data from global positioning system (GPS) collars suggests problems for kernel home-range analyses with least-squares cross-validation, Journal Of Animal Ecology, 74, pp. 455-463.

KATAJISTO J., and MOILANEN A., 2006, Kernel-based home range method for data with irregular sampling intervals, Ecological Modelling, 194, pp. 405-413.

MILLER H., 2005, A measurement theory for time geography, Geographical Analysis, 37, pp. 17-45.

MILLSPAUGH J., GITZEN R., KERNOHAN B., LARSON J., and CLAY C., 2004, Comparability of three analytical techniques to assess joint space use, Wildlife Society Bulletin, 32, pp. 148-157.

SILVERMAN B., 1986. Density estimation for statistics and data analysis (London: Chapman and Hall).

WORTON B., 1987, A Review Of Models Of Home Range For Animal Movement, Ecological Modelling, 38, pp. 277-298.

YAMADA I., and THILL, J-C., 2004, Comparison of planar and network K-functions in traffic accident analysis, Journal of Transport Geography, 12, pp. 149-158.

Biography

Joni Downs is currently a second-year Ph.D. student in the FSU Department of