Spatial Point Pattern Analysis

Some Useful Tools for Analysing Locational Data

MANFRED M. FISCHER*, MICHAEL LEITNER** and PETRA STAUFER-STEINNOCHER***

with 6 Figures

Content ***Page***

1 General Orientation

The proliferation and dissemination of digital spatial databases, coupled with the ever wider use of geographic information systems, is stimulating interest in spatial data analysis from outside the spatial sciences. Spatial data analysis focuses on detecting patterns, and exploring and modelling relationships between such patterns in order to understand processes responsible for observed patterns. In this way spatial data analysis emphasizes the role of space as a potentially important explicator of socioeconomic systems, and attempts to enhance understanding of the working and representation of space, spatial patterns and processes (FISCHER 2001).

Empirical studies in the spatial sciences routinely employ data for which locational attributes are an important source of information. Such data

* Professor and Head, Department of Economic Geography & Geoinformatics,
Director, Institute for Economic Geography, Regional Development and Evironmental Economics,
Vienna University of Economics and Business Administration, A-1090 Vienna, Rossauer Lände 23/1

** Research Scholar, Department of Economic Geography & Geoinformatics, Vienna University of Economics and Business Administration, A-1090 Vienna, Rossauer Lände 23/1

*** Assistant Professor, Department of Economic Geography & Geoinformatics, Vienna University of Economics and Business Administration, A-1090 Vienna, Rossauer Lände 23/1

characteristically consist of one or few cross-sections of observations for either micro-units such as individuals, households or firms at specific points in space, or aggregate spatial entities such as census tracts, regions, or even countries. Observations such as these, for which the absolute location and/or relative positioning (spatial arrangement) is explicitly taken into account, are termed *spatial data* (FISCHER 1999).

In the socio-economic realm, points, lines and areal units are the fundamental entities for representing spatial phenomena. This form of spatial referencing is also a salient feature of geographic information systems. Four classes of spatial data may be distinguished (FISCHER 2001):

(i) *spatial point patterns* [also termed locational data], that is, point locations at which events of interest have occurred,
(ii) *area data* [also termed lattice data] that are defined as discrete variations of attributes over space,
(iii) *field data* [also termed geostatistical or spatially continuous data], that is, observations associated with a continuous variation over space, and
(iv) *spatial interaction data* [sometimes also termed link or flow data] that consist of measurements, each of which is associated with a pair of locations representing points or areas.

The analysis of spatial interaction data has a long tradition in transportation and migration analysis. Field data play an important role in the environmental sciences, but are less important in the social sciences. Point patterns and area data are the most appropriate perspective for spatial data analysis applications in the social sciences. It is noteworthy that point patterns can be converted to area data, and area data can be represented by point reference.

Point patterns arise when the important variable to be analysed is the location of events. A major reason to analyse point patterns is the belief that data represent one source of evidence that may be helpful in learning more about the phenomenon represented and the processes responsible for generating it. Even when our knowledge of a phenomenon is very rudimentary, information gained from point pattern analysis may enable us to give some initial insights into the phenomenon. For example, the knowledge that incidences of cases of a little understood disease are widely dispersed over a region might lead us to consider that it was not spread by contagion.

The purpose of this contribution is to introduce a sample of the procedures that have been designed to analyse point patterns. In order to keep things as simple as possible we do neither consider multivariate nor marked point processes. The objective is to enable readers to evaluate the appropriateness of existing applications of univariate point pattern analysis and to provide sufficient background for the readers to pursue examples of their own. In terms of the techniques presented, our presentation is selective and emphasizes typical problems, that the reader is most likely to encounter in practice. Those readers interested in more exhaustive and sophisticated treatment of point pattern analysis are referred to RIPLEY (1981), DIGGLE (1983), BOOTS and GETIS (1988), CRESSIE (1993), BAILEY and GATRELL (1995).

In the reminder of this contribution we first describe some fundamental characteristics of point patterns and the processes that generate them (see Section 2). Then we begin our discussion of particular techniques of point pattern analysis. Section 3 focuses on techniques for exploring spatial point patterns, while Section 4 deals with modelling spatial point patterns. This distinction is useful but not clear-cut. In practice, there is usually a close interplay between the two, with patterns and relationships explored first and then possibly followed by some modelling. The final section contains advice in choosing one technique over another.

2 General Issues

2.1 Basic Properties of Point Patterns

A point pattern map contains two types of components: the points representing the events being studied and the geographical area in which they are located. We shall refer to these components as the *point pattern* and the *study region*, respectively. A spatial point pattern is a simple example of spatial data because the data comprise only the coordinates of events, at least at the most basic level which we consider in this contribution. Thus, a point pattern data set consists of a series of point locations $\{s_1,..., s_N\}$ in some study region R. We use the term 'event' in a very general sense since the events in question could relate to a wide variety of spatial phenomena that can be regarded to occur at point locations. One advantage of referring to observations on a spatial point pattern as events is that often we will need to distinguish between observed occurrences and other arbitrary locations in the study region. For these other locations we will reserve the term points.

Before discussing techniques for analysing point pattern maps it is useful to become familiar with some basic characteristics of both point patterns and study regions. They are important because they can influence the choice of an appropriate technique and particular decisions necessary to make individual techniques operational. One of the most obvious characteristics of a point pattern is its *size*. This is simply the number of points, N, in the pattern. The study region may be represented by features of various dimensions. In this contribution we limit our attention to [study] regions of *two dimensions*. Two-dimensional regions may be bounded in various ways. We shall refer to the figure enclosed by the boundary of the region as the *shape* of the study region. The most important aspect of the shape is whether it is *regular* or *irregular*. Some of the techniques we discuss can be applied only to regions which are square or rectangular. In this case we assume that a suitable subarea of the original study region has been chosen. In addition, we may have to work with a subarea to avoid *edge effects* by leaving a suitable guard area between the perimeter of the original study region and the subregion within which the analysis is performed.

2.2 Spatial Point Process

Informally speaking, a spatial point process is a stochastic model governing the location of events $\{s_i\}$ in $\mathbf{R}$, $\mathbf{R} \subseteq \mathbb{R}^2$ where $\mathbb{R}^2$ is the two-dimensional real space. The process is represented by a set of random variables $\{Y(\mathbf{A}), \mathbf{A} \subseteq \mathbf{R}\}$ where $Y(\mathbf{A})$ is the number of events ocurring in area $\mathbf{A}$. Hence, our spatial point pattern is a single realization of a spatial point process.

The behaviour of spatial phenomena is often the result of a mixture of both *first order effects* (global or large scale trend) and *second order effects* (local or small scale effects). First order effects are related to first order properties of the spatial point process. They may be described in terms of the intensity $\lambda(\boldsymbol{s})$ of the process which is the mean number of events per unit area at point $\boldsymbol{s}$. Formally, this is defined on the basis of the limiting behaviour of the quantity 'per unit area' (see, for example, DIGGLE 1983):

$$\lambda(\boldsymbol{s}) = \lim_{ds \to 0} \left\{ \frac{E(Y(\boldsymbol{ds}))}{ds} \right\} \tag{1}$$

where $\boldsymbol{ds}$ is a small region around point s, and ds is the area of this region. For a stationary point process[1] $\lambda(\boldsymbol{s})$ is a constant over $\mathbf{R}$, say λ, and then $E(Y(\mathbf{A})) = \lambda A$, where A is the area of $\mathbf{A}$.

The *second order properties*, or spatial dependence, of a spatial point process involve the relationship between events in pairs of areas in $\mathbf{R}$. Formally, this can be described by the *second order intensity*, $\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j)$, of the spatial process that again involves events per unit area and may be defined formally as the mathematical limit

$$\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j) = \lim_{ds_i \, ds_j \to 0} \left\{ \frac{E(Y(\boldsymbol{ds}_i), Y(\boldsymbol{ds}_j))}{ds_i \, ds_j} \right\} \tag{2}$$

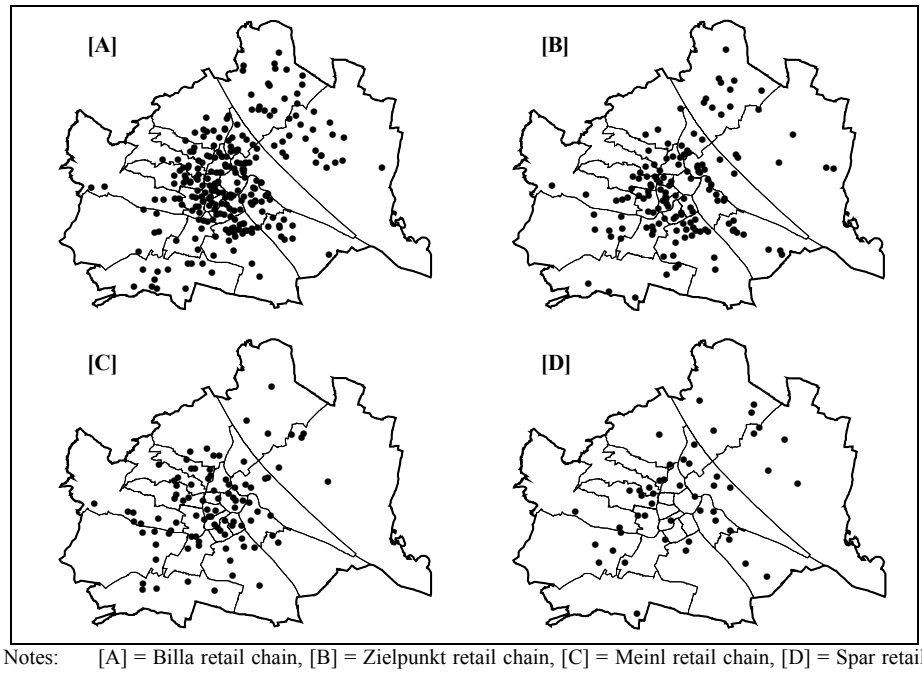
with notation as above. For a stationary process $\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j) = \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) = \gamma(\boldsymbol{h})$. That is, the second order intensity depends only on the vector difference, $\boldsymbol{h}$ (direction and distance) between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, but not on their absolute locations. The process is *isotropic* if, in addition to stationarity, the covariance depends only on the distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ and not on the direction in which they are separated. Then $\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j) = \gamma(\boldsymbol{h})$ (see BAILEY and GATRELL 1995).

[1] A spatial process is termed stationary and homogenous, if its statistical properties are independent of absolute locations in $\mathbf{R}$.

2.3 Case Study

To illustrate the use of spatial point pattern analysis we use a data set consisting of the locations of the outlets of the four largest retail chains in the city of Vienna. Fig. 1 provides an initial impression of the spatial distribution of these outlets in form of dot maps. Do the spatial locations in Fig. 1 appear completely spatially random? Or, are they clustered? Such intuitive notions need quantification because different observers may disagree as to the amount of clustering or randomness in a spatial point pattern.

Fig. 1: Spatial distribution of the outlets of the four largest retail chains in the city of Vienna



Notes: [A] = Billa retail chain, [B] = Zielpunkt retail chain, [C] = Meinl retail chain, [D] = Spar retail chain

Source: ArcAustria-Data Program, WIGeoGIS Vienna, 2000

The standard against which spatial point patterns are generally compared is a realization from a completely spatially random (csr) point process. Clustered patterns are those in which the points are significantly more grouped in the study area than they are in csr, whereas regular patterns are those in which the points are more spread out over the environment than they would be in csr. DIGGLE (1983) suggests three reasons for using csr as the benchmark. *First*, if the null hypothesis of csr is not rejected, further formal statistical analysis is not warranted. *Second*, a null hypothesis of csr provides a dividing hypothesis between clustered and regular patterns. Finally,

even when we anticipate that a null hypothesis of csr will be rejected, the results of the test can be utilized to formulate new null hypotheses.

3 Exploratory Techniques for Spatial Point Patterns

Point pattern analysis in the spatial sciences grew out of a hypothesis-testing tradition, not out of the pattern recognition tradition (see GETIS 1999). Exploratory techniques aim to search for data characteristics such as trends, patterns and outliers. This is important when the data are of poor quality or genuine a priori hypotheses are lacking. We consider methods where we derive summary statistics or plots from the observed distribution of events and attempt to use these informally to analyse hypotheses of interest or to suggest possible models. Some of the methods such as quadrat methods and kernel estimations are more concerned with analysing first order effects in the spatial point process, others such as nearest neighbour distances and the K-function look at the possibility of spatial dependence or second order effects.

3.1 Quadrat Methods

One simple way of summarizing the pattern in the locations of events in a study region R is to partition R into subregions of equal area or quadrats and to utilize the counts of the number of events in each of the quadrats to summarize the spatial pattern (see, RIPLEY 1981, BOOTS and GETIS 1988 for an overview and GETIS 1964 for an early example). Essentially, we are creating a two-dimensional histogram of the observed event occurrencies. We impose a regular grid over R, count the number of events falling into each of the grid squares and convert this to an intensity measure by dividing the area of each of the squares. The result will give some indication of whether and how the intensity of the spatial point process $\lambda(\boldsymbol{s})$ is changing over R (CRESSIE 1993). It is noteworthy that we have effectively transformed the original point pattern into a set of area data. Thus, we could use many of the powerful analysis tools designed for area data (see, for example, FISCHER 2001).

There are several problems that have to be resolved when using quadrat methods. The choice of quadrat size is a difficult issue. Large quadrats may give a global idea of subregions with high or low intensity but throw away much of the spatial detail in the observed pattern. As the quadrats are made smaller to retain more spatial information we get a high variability in quadrat counts. This may finally degenerate into a mosaic with many empty quadrats making meaningful interpretation impossible (see BOOTS and GETIS 1988, BAILEY and GATRELL 1995).

This problem may be circumvented by the use of counts per unit area in a moving window. One defines a suitable size of the window that is then moved over a grid of locations in R. The intensity at each grid point is estimated from the event count per unit area within the window centered around that point. This generates a more spatially smooth estimate of the way in which $\lambda(\boldsymbol{s})$ is varying. Of course the problem remains that the relative location of events within the particular window is not taken into account. Therefore, it is not easy to decide which size of the window to utilize (see BAILEY and GATRELL 1995).

3.2 Kernel Estimators

Kernel estimators of density functions can be extended to obtain nonparametric estimators for $\lambda(.)$ (see DIGGLE 1985, CRESSIE 1993). Let $\{s_1,..., s_N\}$ be the spatial locations of N events in a bounded study region $R \subset \mathbb{R}^2$, then the intensity, $\lambda(\boldsymbol{s})$, at $\boldsymbol{s} \in R$ may be estimated by

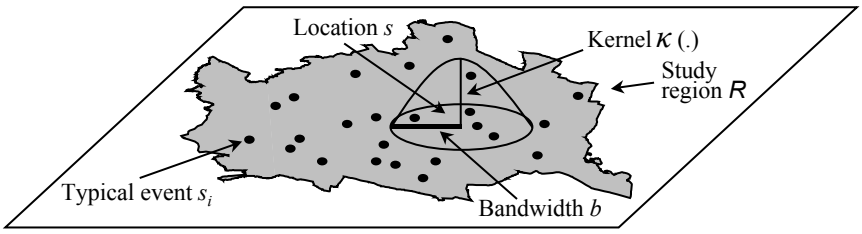
$$\hat{\lambda}_b(\boldsymbol{s}) = \frac{1}{p_b(\boldsymbol{s})} \sum_{i=1}^{N} \frac{1}{b^2} \kappa\left\{\frac{\boldsymbol{s}-\boldsymbol{s}_i}{b}\right\} \tag{3}$$

where $\kappa(.)$ is a suitable chosen probability density (kernel) function symmetric about the origin, $b > 0$ is the bandwidth and determines the amount of smoothing. The function

$$p_b(\boldsymbol{s}) = \int_R \frac{1}{b^2} \kappa\left\{\frac{\boldsymbol{s}-\boldsymbol{u}}{b}\right\} d\boldsymbol{u} \tag{4}$$

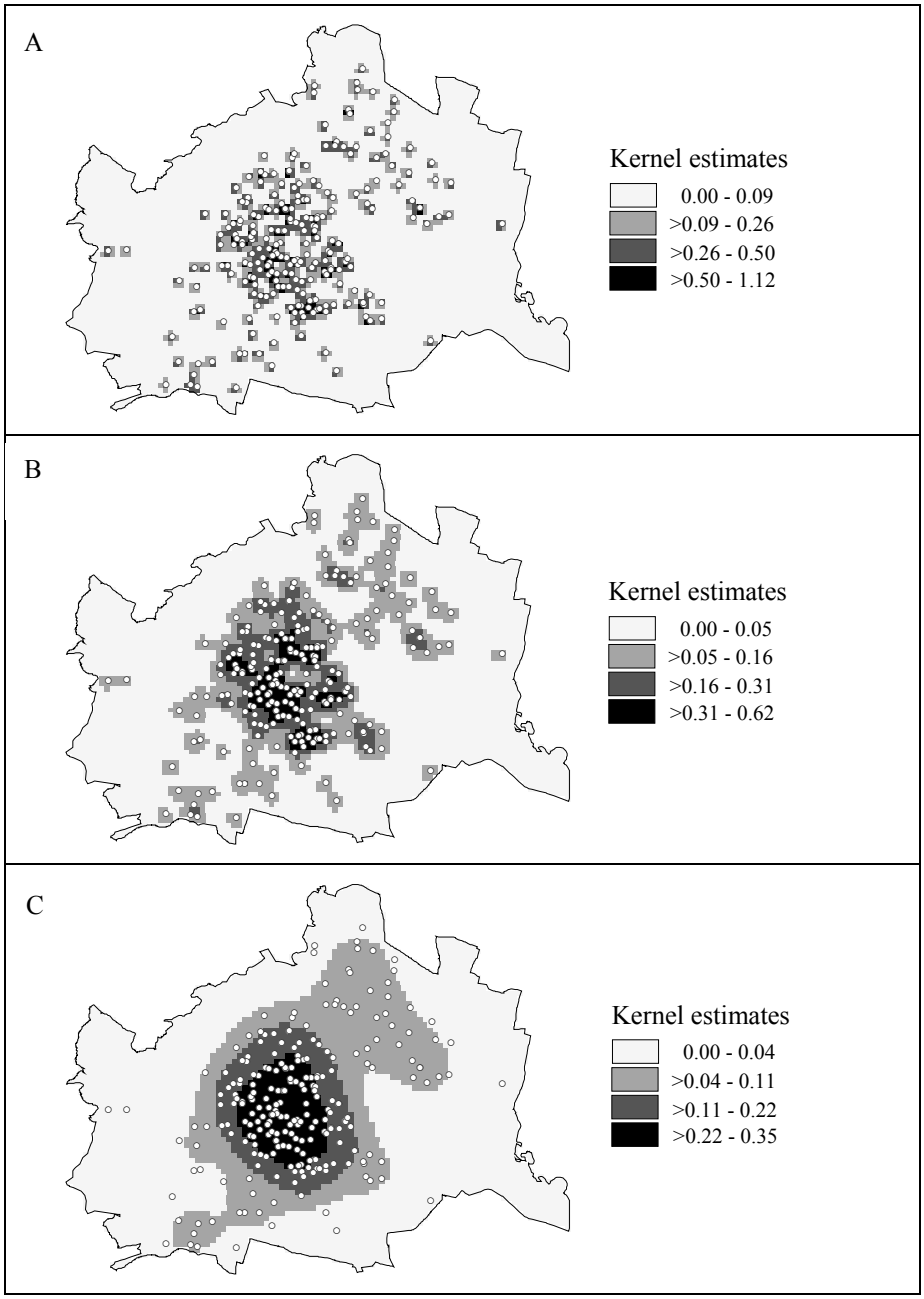
is an edge correction (DIGGLE 1985) - the volume under the scaled kernel centred on $\boldsymbol{s}$ which lies inside R. For any chosen kernel and bandwidth, values of $\lambda_b(\boldsymbol{s})$ can be examined at locations on a suitably chosen fine grid over R to provide a useful visual indication of the variation in the intensity, $\lambda(\boldsymbol{s})$, over the study region (see Fig. 2).

Fig. 2: Kernel estimation of a point pattern



The choice of an appropriate bandwidth or smoothing constant $b > 0$ is of primary concern when estimating $\lambda(.)$ (SILVERMAN 1978). The effect of increasing the bandwidth, b, is to stretch the region around $\boldsymbol{s}$ within which observed events influence the intensity estimate at $\boldsymbol{s}$. For very large b, $\lambda_b(\boldsymbol{s})$ will appear flat and local features will be obscured. If b is small, then $\lambda_b(\boldsymbol{s})$ tends to a collection of spikes centered on the $\boldsymbol{s}_i$. The choice of the kernel function $\kappa(.)$ is of secondary importance.

Fig. 3: Kernel estimates of intensity of the outlets of the Billa retail chain: A: b = 500 meters, B: b = 1,000 meters, and C: b = 3,000 meters



Source: ArcAustria-Data Program, WIGeoGIS Vienna, 2000

Any reasonable choice gives close to optimal results. A typical choice for $\kappa(.)$ might be the quartic kernel

$$\kappa(\boldsymbol{u}) = \begin{cases} \frac{3}{\pi}(1 - \boldsymbol{u}^T\boldsymbol{u})^2 & \text{for } \boldsymbol{u}^T\boldsymbol{u} \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where *T* denotes the transpose operator. The value of the kernel estimation is that one can experiment with different values of *b*, exploring the surface $\lambda_b(\boldsymbol{s})$ using different degrees of smoothing in order to look at the variation in $\lambda(\boldsymbol{s})$ at different scales. Local adjustment of bandwidth may be achieved by means of a technique known as adaptive kernel estimation (see, for example, BAILEY and GATRELL 1995 for more details).

Fig. 3 illustrates kernel estimations by estimating $\lambda_b(\boldsymbol{s})$ with the quartic kernel as defined by (5), but without edge correction. Three examples are presented, where values of *b* = 500 meters, *b* = 1,000 meters, and *b* = 3,000 meters reveal that the higher *b*-value of 3,000 meters smooths the distribution too much, while the lower *b*-value of 500 meters gives a rather spiky impression of the spatial distribution. A value of *b* = 1,000 meters gives the most adequate representation of the spatial distribution.

3.3 Distance Methods

Distance methods make use of precise information on the location of events and have the advantage of not depending on arbitrary choices of quadrat size or shape. They are methods designed more to investigate second order properties using distances between event and/or point locations in the study region R. In general one can define several different kinds of inter-event distances that can throw some light on second order properties of the spatial point process. The methods that we discuss will be based on nearest neighbour distances. In particular, we are interested in distances measured between events and nearest neighbouring events (*W*) which we take as being the distance between a randomly chosen event and the nearest neighbouring event, and also the nearest neighbour point-event distances (*X*), the distance between a randomly selected point in R and the nearest neighbouring observed event. Both measures may be utilized when all events in R have been enumerated. The reader interested in higher-order neighbour distance analysis should consult BOOTS and GETIS (1988).

One way of investigating the degree of spatial dependence in a point pattern is to analyse the observed distribution of one or both of these nearest neighbour distances. It is important to note that nearest neighbour distances only provide information at the smallest scale of pattern. But this may be a very sensible approach if there is also the possibility of large scale variation in the intensity of the given point pattern over R. In this case it is only sensible to analyse second order effects at a scale small enough to avoid variation in intensity becoming confounded with interaction (BAILEY and GATRELL 1995).

A simple approach to summarize patterns using event-to-event nearest neighbour distances (W) or those for point-event distances (X) is to estimate the empirical cumulative probability distribution, $\hat{G}(w)$ of W or $\hat{F}(x)$ of X from the data as (see, for example, DIGGLE 1983)

$$\hat{G}(w) = \frac{1}{N}\left\{\text{number of } (w_i \leq w)\right\} \tag{6}$$

and

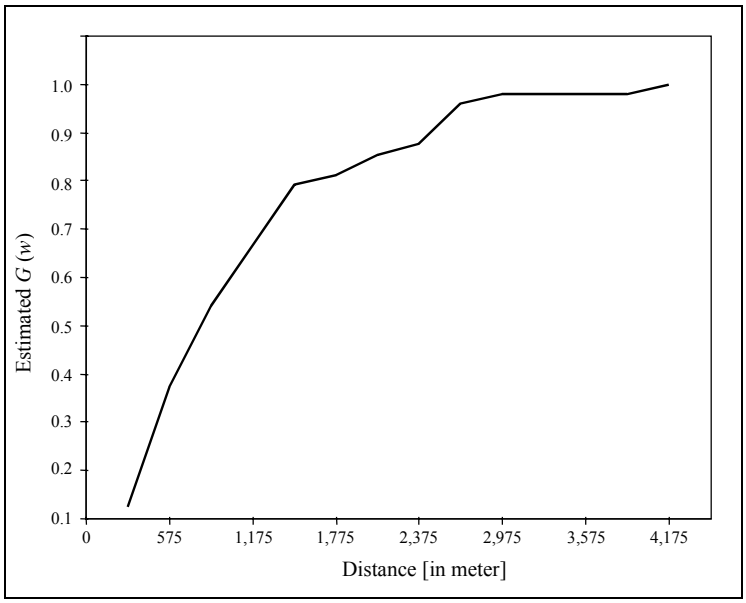
$$\hat{F}(x) = \frac{1}{M}\left\{\text{number of } (x_i \leq x)\right\} \tag{7}$$

where N is the number of events in R and M the number of random points sampled. The resulting empirical distribution function, either $G(w)$ or $F(x)$, can then be plotted against appropriate values of w or x and analysed in an exploratory manner to throw light on possible evidence of inter-event or point-event interaction. If the distribution function climbs very steeply in the early part of its range before flattening out, then the indication would be an observed high probability of short as opposed to long nearest neighbour distances. This would suggest clustering due to inter-event attraction. Alternatively, if it climbs very steeply in the latter part of its range, then this would be an indication for inter-event repulsion or regularity. Fig. 4 illustrates these ideas by estimating $G(w)$ for the outlets of the Spar retail chain. The probability distribution climbs very steeply until a distance of about 2,500 meters and subsequently flattens out, which clearly indicates a higher probability of short as opposed to long nearest neighbour distances and indicates clustering rather than regularity.

The reduction of complex point patterns to an one-dimensional nearest neighbour summary statistic results in a considerable loss of information. Information on individual nearest neighbour distances is lost. Because distances are measured only to the closest events. Only the smallest scales of patterns are taken into account. Unlike quadrat methods, these statistics do not depend on some arbitrary choice of quadrat size. But choosing to measure distances to nearest neighbours as opposed to second, third etc. nearest neighbours is also arbitrary.

Note that the estimation of $G(x)$ or $F(x)$ discussed above makes no correction for edge effects. Thus, the nearest neighbour distance for an event near the boundary of R will be biased, tending to be greater than that for one well inside R. One way to circumvent this problem is to construct a guard area inside the perimeter of R. Nearest neighbour distances are not used for events within the guard area, but events in the guard area are allowed as neighbours of any events (or points) in the rest of R (for more details see BOOTS and GETIS 1988, BAILEY and GATRELL 1995).

Fig. 4: Cumulative probability distribution, $G(w)$ of W, for the outlets of the Spar retail chain



Source: ArcAustria-Data Program, WIGeoGIS Vienna, 2000

3.4 The K-Function

The origin of the K-function can be found in BARTLETT (1964), but its importance as an effective summary of spatial dependence over a wide range of scales was first realized and developed by RIPLEY (1976, 1977). Its definition for mapped data is

$$K(d) = \lambda^{-1} E \text{ (number of events within distance } d \text{ of an arbitrary event; } d \geq 0) \qquad (8)$$

where E is the normal expectation operator, λ the intensity or mean number of events per unit area, assumed constant throughout R. It is sometimes called the reduced second moment measure because it is closely related to the second-ordered intensity of a stationary isotropic point process (see RIPLEY 1976). Its estimation is based on an empirical average replacing the expectation operator. Because the estimation counts numbers of events within a range of distances, it is not very efficient to work with sampled events.

From the complete map of events, let $\{s_1,..., s_N\}$ denote the N locations of all events in R, and assume $\{d_1,..., d_N\}$ are the associated distances from events to the nearest boundary of R. Assuming stationarity and isotropy RIPLEY (1976) has proposed an edge-corrected estimator that uses information on events for which $d_j \leq d$:

$$\hat{K}(d) = \hat{\lambda}^{-1} \sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{j \neq 1}^{N} \boldsymbol{w}(\boldsymbol{s}_i, \boldsymbol{s}_j)^{-1} \boldsymbol{I}(\| \boldsymbol{s}_i - \boldsymbol{s}_j \| \leq d) / N \quad \text{for } d > 0 \tag{9}$$

where $\boldsymbol{w}(\boldsymbol{s}_i, \boldsymbol{s}_j)$ is the proportion of the circumference of a circle centered at $\boldsymbol{s}_i$, passing through $\boldsymbol{s}_j$, and that is inside the study region $\boldsymbol{R}$. $\boldsymbol{I}(\|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq d)$ is an indicator function which is 1 if $\|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq d$ and 0 otherwise. $K(d)$ is unbiased when an estimate of the intensity λ is known. If R is the area of $\boldsymbol{R}$ then the expected number of events in $\boldsymbol{R}$ is λR. Thus, use $\hat{\lambda} = N/R$. It seems that the use of $\hat{\lambda}$ does not upset the unbiasedness too much (see RIPLEY 1981, 159).

The edge corrected estimate of the K-function for an observed point pattern will be reasonable as long as d is not taken too large relative to the size of $\boldsymbol{R}$. This restriction on d is necessary because the weights $\boldsymbol{w}(\boldsymbol{s}_i, \boldsymbol{s}_j)$ can become unbounded as d increases. In practice, this is not a serious problem because of the interest in relatively small values of d. It would be not realistic to explore second order effects that operate on the same scale as the dimensions of $\boldsymbol{R}$.

Fig. 5: Estimation of the K-function

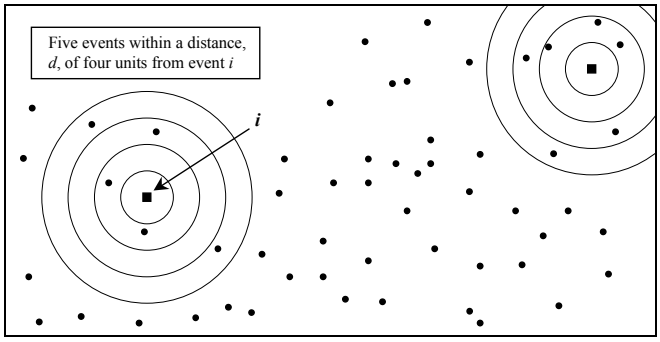


Fig. 5 provides an intuitive graphical idea of what is embodied in the notion of a K-function. Imagine that an event is visited and that around this event a set of concentric circles at a fine spacing is constructed. The cumulative number of events within each of these distance 'bands' is counted. Every other event is similarly visited and the cumulative number of events within distance bands up to a radius d around all the events becomes the estimate of $K(d)$ when scaled by R/N^2. Note that this simple explanation ignores the edge correction factor $\boldsymbol{w}(\boldsymbol{s}_i, \boldsymbol{s}_j)^{-1}$.

Explicit formulae for $\boldsymbol{w}(\boldsymbol{s}_i, \boldsymbol{s}_j)$ can be written down for simple shapes such as rectangular or circular $\boldsymbol{R}$. In other cases the derivation will require quite complex algorithms and can be computationally intensive. Once obtained, $K(d)$ can be plotted and analysed for suggestions of spatial dependence in the point process at different values of d. Under the assumption of csr we would expect $K(d) = \pi d^2$ $(d \geq 0)$ for a homogeneous process with no spatial dependence. Under regularity $K(d)$ tends to be less than πd^2, whereas under clustering $K(d)$ tends to be greater than πd^2. This

suggests an obvious approach to exploratory use of the K-function: Compare $K(d)$ estimated from the observed data with πd^2. One way of doing this is through a plot $L(d)$ against d where

$$\hat{L}(d) = (\hat{K}(d)\pi^{-1})^{\frac{1}{2}} - d. \tag{10}$$

In this plot peaks in positive values tend to indicate spatial attraction of events or clustering, while troughs of negative values indicate spatial repulsion or regularity, at corresponding scales of distance d in each case (see Fig. 6, Section 4.3).

The K-function has several attractive features as a method of data summary and exploration. *First*, it presents information at various scales of pattern, involves the use of the precise location of events and incorporates all event-event distances, not just nearest neighbour distances. *Second*, the theoretical form of $K(d)$ is known for various possible spatial point process models. Thus, $K(d)$ can be used not only to explore spatial dependence, but also to suggest specific models to represent it and to estimate the parameters of such models.

4 Modelling Spatial Point Patterns

Exploratory spatial data analysis is a preliminary step in spatial analysis to more formal modelling approaches. Model-driven analysis of spatial data relies on testing hypotheses about patterns, utilising a range of techniques and methodologies for hypothesis testing, the determination of confidence intervals, estimation of spatial models, simulation, prediction, and assessment of model fit. Testing for complete spatial randomness (csr) is generally the first step in the analysis of a spatial point pattern.

4.1 Complete Spatial Randomness

Complete spatial randomness (csr) is the white noise of spatial point processes. It characterizes the absence of structure in the data. As such, it is usually the null hypothesis in a statistical test to determine whether there is a spatial structure in a given point pattern. Because all stochastic processes have components of randomness in them, the term *complete* spatial randomness is used and is synonymous with a *homogeneous* Poisson process. There are several equivalent definitions of csr, via counting measures, the zero-probability functional, the Laplace generating functional and the probability generating functional (see CRESSIE 1993 for details). Informally speaking, this implies that the number of events in any bounded region R has a Poisson distribution with mean λR where R is the area of R. Given that there are N events in R, those events are independent and form a random sample of uniform distribution in R. This amounts to saying that any event has an equal probability of occurring at any position in R, and that the position of any event is independent of any other [i.e., events do not interact with one another].

If the null hypothesis is rejected, the next obvious step is to fit some alternative (parametric) model to the data. After a model has been fitted, diagnostic tests should be performed to assess its goodness-of-fit. Finally, inference for the estimated parameters is often needed in response to a specific research question. The necessary distribution theory for the estimators can be difficult to obtain. In this case approximations may be necessary.

Departure from csr is usually towards regularity or clustering of events. Clustering may be modelled through an inhomogeneous Poisson point process, a Cox process, or a Poisson cluster process. Inhibition processes can be used to model regular point processes, while Markov point processes can incorporate both elements through small-scale regularity and large-scale clustering. Methods used to fit models to data may change according to the class of models under consideration (see DIGGLE 1983, CRESSIE 1993, for details on this issue).

In the subsections that follow we focus our attention to some simple tests for csr based on nearest neighbour distances and the K-function.

4.2 Nearest Neighbour Tests for CSR

Recall the two basic distances W (event-to-event) and X (sample point-to-event) from Section 3.3. Their distribution theory under csr is well known (see, for example, UPTON and FINGLETON 1985). For a homogeneous Poisson process the probability that there are no events within distance x of an arbitrary point is $\exp(-\lambda \pi x^2)$. Therefore, the distribution function $F(x)$ of the point-to-nearest-event distances X for csr is given by

$$F(x) = Pr(X \leq x) = 1 - exp(-\lambda \pi x^2) \quad \text{for } x \geq 0. \tag{11}$$

This implies that πX^2 follows an exponential distribution with parameter λ, or equivalently, that $2 \pi \lambda X^2$ is distributed as χ_2^2. Thus, we may deduce that

$$E(X) = (2 \lambda^{\frac{1}{2}})^{-1} \tag{12}$$

$$Var(X) = (4 - \pi)(4 \lambda \pi)^{-1}. \tag{13}$$

It also follows that if $X_1, \ldots, X_n$ are independent nearest neighbour point-event distances then $2 \pi \lambda \sum X_i^2$ is distributed as χ_2^2. The same arguments apply for the distribution function $G(w)$ of the event-to-event distance W

$$G(w) = Pr(W \leq w) = 1 - exp(-\lambda \pi w^2) \quad \text{for } w \geq 0 \tag{14}$$

and $E(W)$, $Var(W)$ are given as above for X.

The reader should be cautioned against applying nearest neighbour tests without appropriate corrections for edge effects. Nearest neighbour distances for events near the boundary of R will be biased, tending to be greater than those for events well inside the region. Some tests have explicit edge corrections, built in, but in general these only apply to a restricted set of regularly shaped study regions. In order to avoid this problem, one has to construct a guard area inside the perimeter of R, or employing a torroidal edge correction when the study region is rectangular (BAILEY and GATRELL 1995).

Many statistics have been proposed for testing csr, usually based on a random sample of M points or a random sample of N events. A summary of test statistics and their asymptotic distribution under csr can be found in UPTON and FINGLETON (1985), and CRESSIE (1993). Distribution theory for these tests is based on independence of M [N] nearest neighbour measurements randomly sampled from the study region R, an assumption that is unlikely to hold when R is intensively sampled. Monte Carlo procedures should be used to circumvent this problem. These tests only indicate departure from csr, but do not provide information as to the form of any alternative model when csr is rejected. Thus one would not recommend this type of approach when a completely mapped pattern is available, except perhaps as a preliminary procedure in analysis.

For completely enumerated point patterns there is the possibility to look at the complete estimated distribution function of W or X rather than just a simple summary statistic. In Section 3.3 we discussed estimation of the empirical distribution functions $G(w)$ of W or $F(x)$ of X from an observed point pattern. We follow BAILEY and GATRELL (1995) to present an approach for comparing the whole of this distribution with its theoretical form under csr, for the case of W. An analogous approach would be applicable for X.

The simulation estimate for $G(w)$ under csr is computed as $\overline{G}(w) = \sum \hat{G}_i(w)/M$ where $G(w)$ for $i=1,\ldots, M$ are empirical distribution functions. Each is estimated – without edge correction – from one of the M independent simulations of N events under csr in R. For assessing the significance of departures between the simulated csr distribution $\overline{G}(w)$ and that which is actually observed, $G(w)$, we define upper and lower simulation envelopes (see DIGGLE 1983):

$$U(w) = \max_{i=1,\ldots,M} \left\{\hat{G}_i(w)\right\} \tag{15}$$

$$L(w) = \min_{i=1,\ldots,M} \left\{\hat{G}_i(w)\right\}. \tag{16}$$

We plot $G(w)$ against $\overline{G}(w)$ and add the envelopes $U(w)$ and $L(w)$ to the plot. If the data are compatible with csr then the plot should be roughly linear and at 45 degrees. In the case of clustering the plot will be above the line and in the case of regularity below. $U(w)$ and $L(w)$ assist to assess the significance of departures from the 45 degree line in the plot since

$$Pr(\hat{G}(w) > U(w)) = Pr(\hat{G}(w) < L(w)) = (M+1)^{-1}. \quad (17)$$

4.3 K-Function Tests for CSR

The K-function has obvious advantages. RIPLEY'S (1977) adaption of BARTLETT'S (1964) original suggestion guarantees that it does not depend on the shape of the study region. In addition it presents spatial information at all scales of pattern, and precise spatial locations of events are used in its estimation. We present now a way to compare the empirical K-function, $K(d)$, estimated from the observed data with its theoretical value. This may be done via a plot of $L(d)$ against d where

$$\hat{L}(d) = \pi^{-\frac{1}{2}}(K(d))^{\frac{1}{2}} - d. \quad (18)$$

In this plot positive peaks tend to point to clustering and negative troughs to regularity at different scales of d. Formal assessment of the significance of observed peaks and troughs demands knowledge of $L(d)$ and thus $K(d)$, under csr. This is unknown and complex because of the edge correction built into $K(d)$. But it is possible to use an approach analogous to that used in the case of the nearest neighbour distributions to obtain a simulation estimation of this sampling distribution. The approach is to construct upper and lower simulation envelopes

$$U(d) = \max_{i=1,\ldots,M} \left\{\hat{L}_i(d)\right\} \quad (19)$$

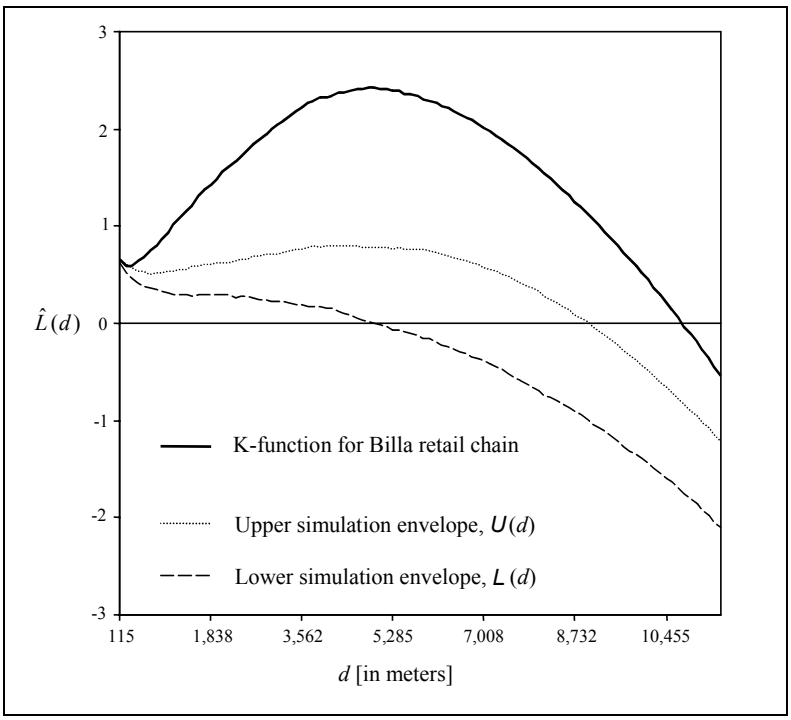
$$L(d) = \min_{i=1,\ldots,M} \left\{\hat{L}_i(d)\right\} \quad (20)$$

from M independent simulations of N events in the study region under csr and their associated $L_i(d)$. These envelopes are included in the plot. The significance of peaks and troughs is assessed on the basis that

$$Pr(\hat{L}(d) > U(d)) = Pr(\hat{L}(d) < L(d)) = (M+1)^{-1}. \quad (21)$$

In the case of significant clustering the plot of $L(d)$ will be above the envelope and in the case of significant regularity below. A point pattern that does not significantly departure from csr will have a $L(d)$ falling in between the plots of $U(w)$ and $L(w)$. The outlets of the Billa retail chain are spatially clustered at all distances with a single peak at d of approximately 5,000 meters (see Fig. 6).

Fig. 6: K-function with upper and lower simulation envelopes for the outlets of the Billa retail chain



Source: ArcAustria-Data Program, WIGeoGIS Vienna, 2000

5 Some Final Remarks

Early studies of spatial point patterns were primarily concerned with comparing quadrat counts to a Poisson distribution. Departures indicate that the problem is not completely spatially random. The degree of departure was generally measured by an index based on the quadrat counts or based on distance measures between events, points and events. These indices do not show characteristics of the pattern at multiple scales very well.

Perhaps the most important developments in recent years are the applications of K-function analysis to the study of point patterns. The K-function has obvious advantages. RIPLEY'S (1977) adaption of BARTLETT'S (1964) original suggestion ensures that it does not depend on the shape of the study region. Moreover, it presents spatial information at all scales of pattern, and precise spatial locations of events are used in its estimation.

The function by RIPLEY (1977) was modified by BESAG (1977) to take into account the need to stabilize variance, and GETIS (1984) generalized the formula to include the weighting of points, such that the sum of pairs of points becomes the sum of the multiples of the weights associated with each marker of the weights of a pair of

points. DIGGLE (1983) has done much to exploit this formulation to show many new features of patterns. For example, one can easily show the difference between an existing pattern and a random pattern. Recently, GATRELL et al. (1996) showed that the K-function can be used as indicator of time-space clustering. This is especially useful for identifying disease clustering over time. Some themes of current interest in pattern analysis are

- the development and testing of time-space pattern models (GATRELL et al. 1996),
- search for pockets of extreme values in large data sets (ORD and GETIS 1995, HASLETT, WILLS and UNWIN 1990),
- The development of pattern models based on differences, absolute differences and similiarities between nearby observations (GETIS and ORD 1996).

6 References

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

BARTLETT, M.S. (1964): The Spectral Analysis of Two-dimensional Point Processes. In: Biometrika, Vol. 51, pp. 299-311

BESAG, J.E. (1977): Discussion Following Ripley. In: Journal of the Royal Statistical Society B, Vol. 39, pp. 193-95

BOOTS, B.N. and GETIS, A. (1988): Point Pattern Analysis. Sage Scientific Geography Series, Vol. 8, Sage Publications, London

CRESSIE, N.A.C. (1993): Statistics for Spatial Data. John Wiley, Chichester

DIGGLE, P.J. (1983): Statistical Analysis of Spatial Point Patterns. Academic Press, London

DIGGLE, P.J. (1985): A Kernel Method for Smoothing Point Process Data. In: Applied Statistics, Vol. 34, pp. 138-47

FISCHER, M.M. (1999): Spatial Analysis: Retrospect and Prospect. In: LONGLEY, P.A., GOODCHILD, M.F., MAGUIRE, P., RHIND, D.W. (Eds.) Geographical Information Systems: Principles, Technical Issues, Management Issues and Applications. Wiley, New York, pp. 283-92

FISCHER, M.M. (2001): Spatial Analysis. In: SMELSER, N.J. and BALTES, P.B. (Eds.) International Encyclopedia of the Social & Behavioural Sciences. Pergamon, Amsterdam [forthcoming]

GATRELL, A.C., BAILEY, T.C., DIGGLE, P.J. and ROWLINGSON, B.S. (1996): Spatial Point Pattern Analysis and its Application in Geographical Epidemiology. In: Transactions of the Institute of British Geographers New Series, Vol. 21, pp. 256-74

GETIS, A. (1964): Temporal Land-Use Pattern Analysis with the Use of Nearest Neighbour and Quadrat Methods. In: Annals of the Association of American Geographers, Vol. 54, pp. 391-99

GETIS, A. (1984): Interaction Modelling Using Second-Order Analysis. In: Environment and Planning A, Vol. 16, pp. 173-83

GETIS, A. (1999): Spatial Statistics. In: LONGLEY, P.A., GOODCHILD, M.F., MAGUIRE, P., RHIND, D.W. (Eds.) Geographical Information Systems: Principles,

Technical Issues, Management Issues and Applications. Wiley, New York, pp. 239-51

GETIS, A. and ORD, J.K. (1996): Local Spatial Statistics: an Overview. In: LONGLEY, P. AND BATTY, M. (Eds.) Spatial Analysis: Modelling in a GIS Environment. Cambridge (UK), GeoInformation International, pp. 269-85

HASLETT, J., WILLS, G. and UNWIN, A. (1990): SPIDER – an Interactive Statistical Tool for the Analysis of Spatially Distributed Data. In: International Journal of Geographical Information Systems, Vol. 4, pp. 285-96

ORD, J.K. and GETIS, A. (1995): Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. In: Geographical Analysis, Vol. 27, pp. 286-306

RIPLEY, B.D. (1976): The Second-Order Analysis of Stationary Point Processes. In: Journal of Applied Probability, Vol. 13, pp. 255-66

RIPLEY, B.D. (1977): Modelling Spatial Patterns. In: Journal of the Royal Statistical Society B, Vol. 39, pp. 172-212

RIPLEY, B.D. (1981): Spatial Statistics. Wiley, New York

SILVERMAN, B.W. (1978): Choosing the Window Width when Estimating a Density. In: Biometrica, Vol. 65, pp. 1-11

UPTON, G. and FINGLETON, B. (1985): Spatial Data Analysis by Example: Volume 1: Point Pattern and Quantitative Data. Wiley, New York