QÜESTIIÓ, vol. 26, 1-2, p. 29-59, 2002

STATISTICAL PROCEDURES FOR SPATIAL POINT PATTERN RECOGNITION

J. MATEU*

Spatial structures in the form of point patterns arise in many different contexts, and in most of them the key goal concerns the detection and recognition of the underlying spatial pattern. Particularly interesting is the case of pattern analysis with replicated data in two or more experimental groups. This paper compares design-based and model-based approaches to the analysis of this kind of spatial data. Basic questions about pattern detection concern estimating the properties of the underlying spatial point process within each experimental group, and comparing the properties between groups. The paper discusses how either approach can be implemented in the specific context of a single-factor replicated experiment and uses simulations to show how the model-based approach can be more efficient when the underlying model assumptions hold, but potentially misleading otherwise.

Keywords: Expected significance levels, Replicated patterns, Spatial point structures

AMS Classification (MSC 2000): 62M05, 60G55

* Department of Mathematics, Universitat Jaume I, E-12071, Castellón, Spain.
Correspondence to: Jorge Mateu, Department of Mathematics, Universitat Jaume I, E-12071, Castellón, Spain. Email:mateu@mat.uji.es.

– Received July 2000.
– Accepted January 2001.

1. INTRODUCTION

For the single replicate point patterns which dominate the spatial point pattern literature (Cressie (1993), Diggle (1983) and Ripley (1981)), there has been a strong emphasis on fitting models to detect and recognize the spatial structures.

Given the wealth of opportunities for collecting replicated spatial point pattern data, for example in clinical neuroanatomy, or in materials science, where replicate images can easily be obtained using standard microscopical equipment, it is surprising that few appropriate methods of analysis have been suggested. Design-based methods have been suggested by Diggle *et al.* (1991), Wilson (1998), Wilson *et al.* (1998) and Baddeley *et al.* (1985, 1993). The first of these presents a method, motivated by analysis of variance, for the assessment of replicated spatial point patterns, allowing for a single replicate from each individual of interest. The authors illustrate their method with an example of two-dimensional replicated data from clinical neuroanatomy. The second extends the methodology developed by Diggle *et al.* (1991), to allow for multiple replicates per individual, and to incorporate more complex experimental set-ups. Again, the example data is from clinical neuro-anatomy, but the data is three-dimensional. Baddeley *et al.* (1993) present an alternative method, which is again motivated by analysis of variance, but uses a ratio regression approach, and again allows for replication within individuals as well as between individuals. In this paper, we review in detail the methods developed by Wilson (1998).

In many applications such as biological or neuroanatomical applications, the points of interest are the centres of cells, and it is a key point the modelling of cell centre positions. Figure 1 represents one such example, where each one of the 12 plots shows the spatial positions (in form of a point in the square window) of pyramidal neurons of the Cingulate Cortex of a particular selected individual. This data set was analyzed by Diggle *et al.* (1991) and further information on the data can be found there. At one extreme, the cell centres might be thought of as a hard-core process, with the hard-core parameter being equal to the cell diameter. A pairwise interaction process (Ripley, 1977; Diggle *et al.*, 1994) might be a more appropriate model for cell centre data; this class of point process is described at length in Section 2. Broadly speaking, cells are separated by a particular distance δ with a given probability, where small distances may be assigned small probability (a hard-core process of hard-core radius ρ is a special case of this; the separation probability is 0 for distances less than ρ and 1 for distances greater than ρ). Through this framework, the rigid regularity of the hard-core process can be relaxed, presenting a more realistic description of cell centre behaviour. There are various ways in which we could choose to fit models to the replicated spatial point pattern data. Appropriate single replicate methods are reviewed by Diggle *et al.* (1994). Our chosen approach is an extension of the maximum pseudo-likelihood method.

We have thus highlighted two contrasting approaches to the assessment of replicated spatial point patterns, the design-based ANOVA approach and the model-based maximum pseudo-likelihood approach. Both methodologies are competent and useful at detecting pattern structure though their practical performance is little known.

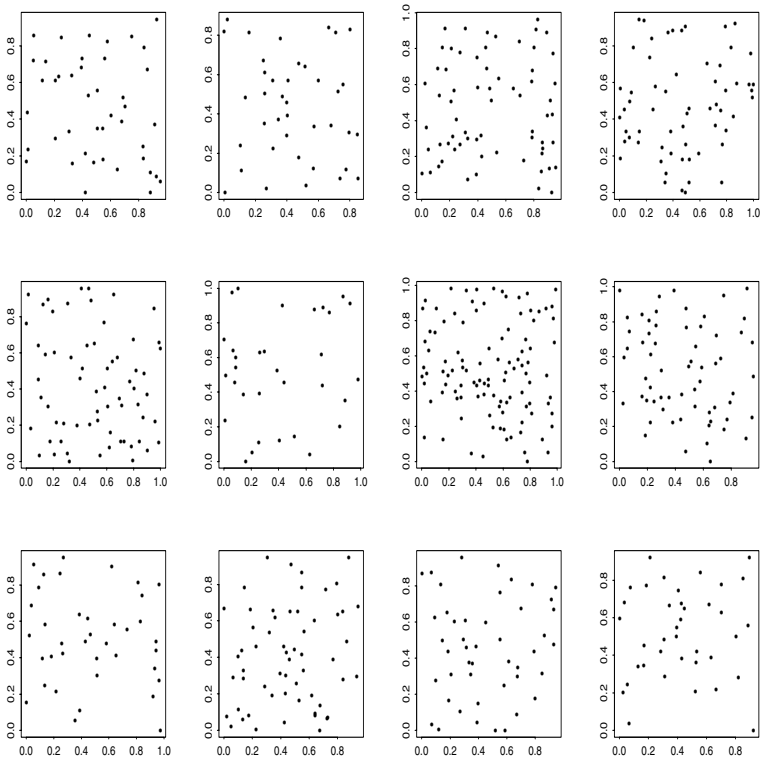

Figure 1. Digitized pyramidal neuron positions.

In this paper, we compare and contrast the two approaches under a variety of experimental conditions. As part of the study, we investigate the effect of fitting various pairwise interaction models to different kinds of replicated spatial point pattern data. There are two questions which we seek in particular to answer:

- When the model specification is correct, is the model-based approach more powerful than the design-based approach for a given set of data?
- When the model specification is incorrect, does the model-based procedure break down, to give misleading results?

In Section 2, we introduce formally the pairwise interaction process, and highlight the classes of this process upon which we base this study. In Section 3, we review the design-based ANOVA approaches of Diggle *et al.* (1991) and Wilson (1998). In Section 4, we review the maximum pseudo-likelihood method for replicated spatial data. Section 5 defines the concept of Expected Significance Level used throught the simulation study. Finally, in Section 6, we give the results of a simulation study which investigates the above two questions.

2. PAIRWISE INTERACTION POINT PROCESSES

2.1. Theory setup

Assume we have observed a pattern of points $X = \{x_i \in A : i = 1,\ldots,n\}$ in a planar region A. The x_i are called *events* to distinguish them from generic points in $x \in A$. Each point process on A can be defined by a sequence of intensity functions $\lambda_n(x_1,x_2,\ldots,x_n)$ which are measurable functions that do not depend on the order of the events. Then, the point process will contain n events with probability p_n,

$$p_n = \begin{cases} \exp(-\nu(A)) & \text{if } n = 0 \\ \frac{\exp(-\nu(A))}{n!}\int_{A^n}\lambda_n(x_1,x_2,\ldots,x_n)dx_1\cdots dx_n & \text{if } n \geq 1 \end{cases} \tag{2.1}$$

where $\nu(A)$ represents the Lebesgue measure of A and $\sum_{n=0}^{\infty} p_n = 1$ (Kelly & Ripley, 1976).

The joint density function for X,

$$f[(x_1,x_2,\ldots,x_n),n] = \frac{\exp(-\nu(A))}{n!}\lambda_n(x_1,x_2,\ldots,x_n) \tag{2.2}$$

can be factored uniquely for $n = 1,2,\ldots$ as (Daley & Vere-Jones, 1988)

$$\begin{aligned} f[(x_1,x_2,\ldots,x_n),n] &= \frac{\exp(-\nu(A))}{\alpha n!}\exp\left\{\sum_{i=1}^{n} g_1(x_i) + \sum_{i=1}^{n}\sum_{j>i} g_{12}(x_i,x_j) + \right. \\ &\quad \left. \cdots + g_{12\cdots n}(x_1,x_2,\ldots,x_n)\right\} \end{aligned} \tag{2.3}$$

where α is the normalising constant which usually does not have a closed-form expression.

The likelihood (Janossy density) is given by

$$l(x_1,x_2,\ldots,x_n) = n!\, f[(x_1,x_2,\ldots,x_n),n] \tag{2.4}$$

The point process defined by (2.3) and (2.4) is called *Gibbs process*. Adding a nearest-neighbour condition to a Gibbs process we have a Markov process. The nearest-neighbour

condition is usually defined in terms of the Euclidean distance. Two events x_i and x_j are said to be neighbours if $\|x_i - x_j\| \leq r$, for some $r > 0$. A *clique* is defined to be a single event or a set of events, all of which are neighbours of each other. By the Hammersley-Clifford Theorem, $g_{1,2,\ldots,k}(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = 0$ unless the events $x_{i_1}, x_{i_2}, \ldots, x_{i_k}$ form a clique. Then, the point process defined by (2.3) is Markov of range r. Markov point processes were introduced by Ripley & Kelly (1977), whereas the concept of Gibbs distributions has been used in statistical physics for a longer time (Preston, 1977). Since the introduction of Markov point processes in spatial statistics attention has focussed on the special case of *pairwise interaction models* in which each configuration of events interacts only via pairs of points from this configuration. These provide a large variety of complex patterns starting from simple potential functions which are interpretable as attractive and/or repulsive forces acting among events. Pairwise interaction models are simple exponential families and are widely used as they are very amenable to simulation and iterative statistical methods.

For a *pairwise interaction point process*, X, of n events in a bounded region A, the joint density (2.3) is of the form

$$f(X) = C^{-1}\frac{\beta^n}{n!}\exp\left\{-\sum_{i=1}^{n}\sum_{j>i}\Phi\left(\|x_i - x_j\|;\theta\right)\right\} \tag{2.5}$$

where C is a normalising constant depending on β and Φ, $||.||$ denotes the Euclidean distance, Φ is a *potential function* depending on a set of parameters θ and β is a parameter which determines the intensity of the process. Note that (2.5) is (2.3) with $g_1(x_i) = \log\beta$, $g_{12}(x_i, x_j) = -\Phi(.)$, $g_{12\cdots k}(.) = 0$ for $k \geq 3$ and $C = \alpha/\exp(-\nu(A))$. The function

$$U_n(X) = \sum_{i=1}^{n}\sum_{j>i}\Phi\left(\|x_i - x_j\|;\theta\right) \tag{2.6}$$

is usually called *total potential energy*. Often, an *interaction function* is used instead of the potential function,

$$e(t) = \exp(-\Phi(t)) \tag{2.7}$$

Using (2.7), then (2.5) is usually represented as (Baddeley & Møller, 1989)

$$f(X) = \alpha\prod_i b(x_i)\prod_{i<j} e(x_i, x_j) \tag{2.8}$$

where α stands for the normalising constant and b, e (the interaction function) are non-negative measurable functions.

Note that restrictions on the form of the potential function are needed to ensure that the normalising constant is finite. A sufficient condition is that $\Phi(t) \geq 0$ for all t and decreases as t increases.

The resulting processes are called inhibitory and generate patterns with varying degrees of spatial regularity according to the specific specification of the potential/interaction function. They do not seem to be able to produce clustered patterns in sufficient variety. The original clustering model of Strauss (Strauss, 1975) turned out (Kelly & Ripley, 1976) to be non-integrable for those parameter values such that $\Phi(t) < 0$ for $t > h$, where $h > 0$ is called the *hard-core distance*. Gates & Westcott (1986) showed that partly-attractive potentials may violate a stability condition, implying that they produce extremely clustered patterns with high probability. Also, simulations by Møller (1993) suggest that the behaviour of the Strauss model with fixed n undergoes an abrupt transition from «Poisson-like» patterns to tightly clustered patterns rather than exhibiting intermediate, moderately clustered patterns.

It is known that the distribution given by (2.5) coincides with the stationary distribution of a spatial birth-death process on A which provides one method for simulating realisations of the point process so defined.

2.2. Models

Attention in this paper is focussed on three different parametric families of models, each having a specific interaction function depending on a scalar parameter.

Our first model has interaction function

$$e(t) = \begin{cases} 1-(1-t^2/\theta^2)^2 & \text{if } t \le \theta \\ 1 & \text{if } t > \theta. \end{cases} \tag{2.9}$$

The interaction function is continuously differentiable with respect to t and θ defines the range of interaction. This model was used by Diggle (1986) in a comparison between approximate maximum likelihood and maximum pseudo-likelihood estimators for θ. It was also used by Diggle *et al.* (1994) to compare several parametric estimation techniques. Small values of θ are indicative of weak interaction whereas larger values indicate strong interaction. This model will be called *Diggle* model throughout the paper.

The second model has interaction function

$$e(t) = 1 - \exp\left(-\frac{t^2}{\theta^2}\right). \tag{2.10}$$

It is called *Very-Soft-Core* (VSC) model and was used in Ogata & Tanemura (1984) and Diggle *et al.* (1994). The range of interaction is infinite, i.e., $e(t) < 1$ for all finite t. Again, small values of θ are indicative of weak interaction whereas larger values indicate strong interaction.

Finally, we have used a third model which is qualitatively different from the other first two. It has interaction function

$$e(t) = \begin{cases} \frac{t}{\theta} & \text{if } t \leq \theta \\ 1 & \text{if } t > \theta \end{cases} \tag{2.11}$$

and due to its lineal form will be called *linear* model. As with the Diggle model, the interaction function is continuously differentiable with respect to t and θ defines the range of interaction. Small values of θ are indicative of weak interaction whereas larger values indicate strong interaction.

2.3. Simulation method

The spatial birth-and-death process provides the framework under which Ripley (1977, 1981) proposes to simulate a Markov point process on the bounded Borel set $A \subset \Re^d$ with n fixed. The method is related to Markov processes used in statistical mechanics and surveyed originally by Hastings (1970). Other techniques such as Metropolis-Hastings algorithms (Geyer & Møller, 1994; Møller, 1992) are clearly possible.

Consider a set of particles interacting according to a certain potential function, on a square A with periodic boundary, i.e., A is identified with a torus. The algorithm is as follows.

First, select n events from a uniform distribution on A and call this initial point pattern $X^n(0)$. At step $(t+1)$, delete systematically in turn one of the n events of $X^n(t) = \{x_1, x_2, \ldots, x_n\}$, say event x_i, and let $X^n(t) \setminus \{x_i\}$ denote the point pattern formed by removing x_i from $X^n(t)$. Let

$$p(u; X^n(t) \setminus \{x_i\}) = \frac{f(X^n(t) \setminus \{x_i\}, u)}{f(X^n(t) \setminus \{x_i\})} \tag{2.12}$$

denote the conditional intensity at $u \in A$ given $X^n(t) \setminus \{x_i\}$. Define

$$M = \sup_{u \in A} p(u; X^n(t) \setminus \{x_i\}).$$

Finally, select an event u from a uniform distribution on A and set $X^n(t+1) = \{X^n(t) \setminus \{x_i\}, u\}$ with probability $p(u; X^n(t) \setminus \{x_i\})/M$; otherwise, selection is repeated until a qualifying u is found. This method ensures that samples taken every n steps have no points in common. Ultimately, convergence to a Markov point process with likelihood $f(.)$ will occur. This algorithm is analogous to the Gibbs sampler on the spatial lattice (Geman & Geman, 1984).

In the simulation study we simulate on a larger polygon $(-0.5, 1.5) \times (-0.5, 1.5)$ and use a toroidal shift to make our point process compatible with the MPL method discussed in Section 4. We then extract the central unit square from the polygon, and use this part of the data for further analysis.

3. A DESIGN-BASED APPROACH

A descriptor which makes use of the K-function (the reduced second moment measure (Ripley, 1977, 1988) is suggested. For a stationary, isotropic, orderly point process, the K-function may be defined as

$$K(t) = 2\pi\lambda^{-2}\int_0^t \lambda_2(r)dr \tag{3.1}$$

where λ is the point process intensity (the expected number of events per unit area in two dimensions, volume in three dimensions and so on), and $\lambda_2(r)$ is the corresponding second order intensity function. Heuristically, the K-function describes the expected number of cells within a distance t of an arbitrarily chosen point of the process, and, as such, provides a good indicator of regularity or clustering in a given point pattern.

Diggle *et al.* (1991) develop a method for investigating whether schizophrenic patients suffer from some structural defect of the cerebral cortex. Specifically, the authors set out to establish whether the brain of the schizophrenic demonstrates unusual arrangements of neurons. A single, two dimensional neuronal cell pattern is obtained from each of several schizophrenic, schizoaffective and control individuals.

The statistic defined by Diggle *et al.* (1991) assumes the simplest case scenario, namely that there is a one-way experimental set-up, with g groups of replicated spatial data, and within each of those g groups, each individual provides a single point pattern realisation. Their proposed statistic is given by the following expression

$$D_g = \sum_{i=1}^{g}\int_0^{t_0}\left[\bar{K}_i^{1/2}(t) - \bar{K}^{1/2}(t)\right]^2 dt, \tag{3.2}$$

where

$$\bar{K}_i(t) = \frac{\sum_{j=1}^{m_i} n_{ij}\hat{K}_{ij}(t)}{\sum_{j=1}^{m_i} n_{ij}} \tag{3.3}$$

and

$$\bar{K}(t) = \frac{1}{n}\sum_{i=1}^{g} n_i\bar{K}_i(t). \tag{3.4}$$

In addition, n_{ij} is the number of points in the sampling window from individual j in group i, $n_i = \sum_{j=1}^{m_i} n_{ij}$ and $n = \sum_{i=1}^{g} n_i$. We see that $\bar{K}_i(t)$ is the weighted mean K-function in group i, and $\bar{K}(t)$ is the weighted mean K-function across all groups.

The statistic D_g is similar to a between-treatment sum of squares in an ordinary one way analysis of variance, in that it compares group mean K-functions with the overall

mean K-function, in an attempt to quantify between group differences. This test statistic in combination with a Monte-Carlo procedure is used to assess the significance of between-group differences in the spatial patterns.

3.1. A new test statistic

In a further step, Wilson (1998) suggested a new test statistic for comparing groups of spatial data, which is a modification of the Diggle *et al.* (1991) test statistic D_g.

To motivate the new test statistic, first recall a result from the analysis of variance literature. Suppose that we have a simple, one-way experiment, with observations $y_{i1}, \ldots, y_{im_i}$ in g groups (so that group i contains m_i observations). Then under the assumption that

$$\mathrm{Var}(y_{ij}) = \frac{\sigma^2}{n_{ij}},$$

it can be shown that the best estimate of the between-treatment sum of squares is given by expression (3.5)

$$\mathrm{BTSS} = \left[\sum_{i=1}^{g} n_i \left(\bar{y_i} - \bar{y}\right)^2 \right], \tag{3.5}$$

where

$$\bar{y_i} = \frac{\sum_{j=1}^{m_i} n_{ij} y_{ij}}{n_i} \tag{3.6}$$

and

$$\bar{y} = \frac{\sum_{i=1}^{g} n_i \bar{y_i}}{n}, \tag{3.7}$$

with $n = \sum_{i=1}^{g} n_i$. Returning to spatial point patterns, consider now an estimated K-function from individual j in one of the groups i, $\hat{K}_{ij}(t)$, say. Fix t at a particular distance t_d, say. Then $\hat{K}_{ij}(t_d)$ represents a single realised value. Furthermore, assume that $\mathrm{Var}(\hat{K}_{ij}(t_d))$ is proportional to $1/n_{ij}$, to reflect the fact that a K-function estimated from a large number of points has a small variance, and vice-versa. The variance structure here is identical to that defined above, and so we use this analysis of variance result as the basis for a new statistic.

We define

$$\bar{K_i}(t_d) = \frac{\sum_{j=1}^{m_i} n_{ij} \hat{K}_{ij}(t_d)}{n_i}$$

and extend this to cover the range of values for t_d

$$\bar{K}_i(t) = \frac{\sum_{j=1}^{m_i} n_{ij}\hat{K}_{ij}(t)}{n_i}. \tag{3.8}$$

In a similar way, we define

$$\bar{K}(t) = \frac{\sum_{i=1}^{g} n_i\bar{K}_i(t)}{n}, \tag{3.9}$$

where n_i and n are defined as before.

This new test statistic is defined as follows

$$D_{g2} = \sum_{i=1}^{g}\int_0^{t_0} w(t)n_i\left[\bar{K}_i(t) - \bar{K}(t)\right]^2 dt. \tag{3.10}$$

Note that, in addition to the new factor n_i, the statistic does not have the square root transformation, suggested by Diggle *et al.* (1991) as a variance-stabilising mechanism. This has been replaced with a weighting function $w(t)$, which down-weights the variance of the K-function estimates at large t. We select an appropriate weighting function by considering a spatially random patterns; the estimated K-function for each pattern behaves like a Poisson random variable. This follows because the K-function looks at count data within discs (in two dimensions). As the K-function itself is Poisson, then by the definition of a Poisson distribution, the variance of $\hat{K}$ is proportional to K.

Furthermore, for a spatially random process, $K(t) = \pi t^2$, i.e. $K(t) \propto t^2$. We therefore select the reciprocal of t^2 as an appropriate weighting function (for one-dimensional data we would use $w(t) = 1/t$, and for three-dimensional data we would use $w(t) = 1/t^3$, by the same argument).

3.2. A Monte-Carlo procedure

We adopt the Monte-Carlo procedure suggested by Diggle *et al.* (1991) as a means of assigning significance to observed values of the new test statistic. The procedure is reproduced from Diggle *et al.* (1991).

Suppose we have a set of point process data, divided into three groups. Suppose also that there are r_i individuals in group i. We maintain the definitions of n_{ij} and n_i given earlier.

- We begin by defining the *residual K-function*. This may be written as

$$\hat{R}_{ij}(t) = n_{ij}^{\frac{1}{2}}\left[\hat{K}_{ij}(t) - \bar{K}_i(t)\right].$$

- Using this definition, we can further obtain new functions $\hat{K}^*(t)$, which we define as

$$\hat{K}^*_{ij}(t) = \bar{K}(t) + n_{ij}^{-\frac{1}{2}} \hat{R}^*_{ij}(t),$$

  where the $\hat{R}^*_{ij}(t)$ are obtained by permuting at random, across groups, the residual K-functions, so that $\hat{R}^*_{ij}(t)$ is one of the residual K-functions, drawn at random and without replacement.
- Recompute $\bar{K}^*_i(t)$ and $\bar{K}^*(t)$ from this permutation sample, and compute the statistic D^*_{g2}.
- Repeat this procedure n_s times, and each time obtain a new value for D^*_{g2}.
- Rank each statistic from the original data amongst the n_s statistics obtained from the permutation test, and from the rank compute a p-value for each statistic. A large statistic corresponds to a small p-value and *vice-versa*.

We thus have a test statistic and a method for assigning significance to observed values from data sets. Several simulation studies in the thesis of Wilson (1998) demonstrate the improved performance of D_{g2} in comparison with D_g. Particularly, D_{g2} detects differences between groups of replicated spatial data which are not obvious from visual inspection alone.

4. A MODEL-BASED APPROACH

4.1. Parameter estimation

In this paper we focus on the edge-corrected *pseudo-likelihood* estimation method as it can be used routinely in applications and does not place artificial restrictions on the parametric form of the potential function. Moreover, this method can be easily implemented for the case where replicated spatial point pattern data is available. Other general methods such as approximations to maximum likelihood require numerical or Monte Carlo approximations to the normalising constant (Ogata & Tanemura, 1981, 1984, 1989; Penttinen, 1984) or recursive approximation methods (Moyeed & Baddeley, 1991). The Takacs-Fiksel method (Takacs, 1986; Fiksel, 1984, 1988) has been proposed in literature but will not be explored further in this paper. For a more detailed review see (Diggle *et al.*, 1994; Ripley, 1988; Geyer & Thompson, 1992).

The pseudo-likelihood for Markov processes is defined (Besag, 1977; Jensen & Møller, 1991) by

$$PL(\theta, \beta; X) = \exp\left\{ -\int_A \lambda(u; X) du \right\} \prod_{i=1}^{n} \lambda(x_i; X^i) \tag{4.1}$$

where $X^i = X - \{x_i\}$ and $\lambda(u;X)$ represents the *conditional intensity* (Papangelou conditional intensity, see Daley & Vere-Jones, 1988) of an event u, given the pattern X and is defined by $\lambda(u;X) = \frac{f(X\cup\{u\})}{f(X)}$. Usually, (4.1) is re-cast in terms of its logarithm. Maximisation of (4.1) with respect to parameters θ, β yields the maximum pseudo-likelihood estimators. The estimating equation for parameter β is (Besag, 1977)

$$n = \beta \int_A \lambda_0(u;X)\,du \tag{4.2}$$

where $\lambda_0(u;X)$ is derived as (Diggle *et al.*, 1994)

$$\lambda(u;X) = \beta \exp\left\{ -\sum_{j=1}^{n} \Phi\left(\left\|u - x_j\right\|;\theta\right)\right\} = \beta\lambda_0(u;X). \tag{4.3}$$

Then, the estimating function for parameter θ becomes

$$PL(\theta) = \sum_{i=1}^{n} \log\left\{\lambda_0(x_i;X^i)\right\} - n\log\left\{\int_A \lambda_0(u;X)\,du\right\}. \tag{4.4}$$

For inhibitory pairwise interaction models it is known that pseudo-likelihood estimation is a special case of the Takacs-Fiksel method when the interaction radius is fixed (Ripley, 1988; Diggle *et al.*, 1994). The pseudo-likelihood equation (4.1) can be interpreted as the limit case of pseudo-likelihood for lattice processes (Besag, 1977; Besag *et al.*, 1982). For Markov processes of finite range, maximum pseudo-likelihood estimators are consistent (Jensen & Møller, 1991). Asymptotic normality is considered in Jensen (1993).

The extension to the replicated case is straightforward. Let $PL_i(\theta,\beta;X_i)$ denote the pseudo-likelihood equation (4.1) for the i-th pattern, $i = 1,\ldots,k$, k denoting the number of replicates. Then, the pooled pseudo-likelihood function for the replicated patterns is given by

$$PL^p(\theta,\beta;X^p) = \prod_{i=1}^{k} PL_i(\theta,\beta;X_i), \tag{4.5}$$

where X^p denotes the whole set of point patterns and X_i stands for the i-th pattern. If we use $log\ PL^p$, then we sum up the k terms instead of multiplying them.

Usually, in applications, the region A is a sampled sub-region of a much larger region within which the phenomenon of interest operates, and some form of edge-correction is vital. We use here an adaptation of Ripley's edge-correction (Ripley, 1977, 1988). See also Diggle *et al.* (1994). The idea is to replace those summations of the form

$$\sum_{j>i} \Phi\left(\left\|x_i - x_j\right\|;\theta\right)$$

appearing, for example, in the density function (2.5), by

$$\frac{1}{2}\sum_{j\neq i} w_{ij}^{-1}\Phi\left(\left\|x_i - x_j\right\|;\theta\right)$$

where w_{ij} is the proportion of the circumference of the circle with centre x_i and radius $\left\|x_i - x_j\right\|$ which is contained within A. This edge-correction compensates for the omission of contributions to this total potential from unobserved events outside A.

4.2. A formal test: Testing for group differences

Let θ denote the interaction parameter in a pairwise interaction model. Then to test for group differences for replicated data via the model-based or parametric method, we use the following procedure:

- Fit a different θ to each one of the g groups and get the maximised value of the log-pseudo-likelihood, PL^1, where $PL^1 = \sum_{j=1}^{g} PL_j^p$ and PL_j^p equals the maximised value of the log-pseudo-likelihood of the j-th group.
- Pool all the data, ignoring groupings, and fit a common θ' to all the data. Obtain the maximised value of the log-pseudo-likelihood, PL^0.
- Define a test statistic $T = PL^1 - PL^0$.

To assess between-group spatial differences, we use the following Monte Carlo procedure:

- Condition on the number of groups g, number of replicates per group r_i and the number of events per replicate n_{ij}. Treat this latter as the expected number of events per replicate. Simulate a new data set using these parameters and the estimate of the pooled θ' from above as the interaction parameter for all replicates.
- Compute the test statistic using the method outlined above, and get a value for this new data set.
- Repeat this procedure 99 times.
- Rank the test statistic from original data amongst these simulated test statistics, to get a p-value. A high rank of the test statistic from original data is evidence against the hypothesis of a common value of θ in all the groups.

5. ASSIGNING SIGNIFICANCE TO OBSERVED TEST STATISTICS: EXPECTED SIGNIFICANCE LEVEL (ESL)

Rather than carrying out a conventional power study to compare the performance of the two approaches, we assess the performance of each approach using the concept of Expected Significance Level (ESL), which was first introduced by Dempster and Schatzoff (1965). The theory of power is based upon decision rules, that is, in some sense arbitrary choices of cut-offs; for a test of size α, where α represents the probability of a type I error, we select a value of power to be $(1-\beta)$, where β is the probability of a type II error (so that the power represents the probability of detecting an effect given that it exists).

The idea of ESL is more flexible than this; suppose that we have a test statistic T, and in a set of n simulations under the alternative hypothesis H_A, we observe p-values $\alpha_1, \alpha_2, \ldots, \alpha_n$. In a classical power framework, we would estimate the power of the statistic for a given level α' to be

$$(1-\beta) = 1 - \frac{(\text{number of } \alpha_i : \alpha_i < \alpha')}{n}.$$

This requires the (somewhat arbitrary) specification of α'. Instead, we could take the ESL approach, and estimate the ESL by

$$\bar{\alpha} = \frac{\sum_{i=1}^{n} \alpha_i}{n}.$$

Immediately we are preserving information about the observed significance levels which is lost in conventional power summaries; in addition, there is scope to look at the range and distribution of α_i, and we make use of this in this paper. Formally, the ESL is equivalent to a uniform weighting of power over $\{\alpha : 0 \leq \alpha \leq 1\}$.

6. COMPARISON BETWEEN THE STATISTICAL PROCEDURES

We develop a simulation study to quantify and compare the parametric model-based and the non-parametric design-based approaches under different experimental circumstances. Consider several scenarios:

- Suppose we assume a certain pairwise interaction model for the data which is correct. Then we would expect the model-based approach to be better at detecting group differences than the design-based approach, as a consequence of the fact that we allow ourselves to make more assumptions about the model which generates the data. We must verify that the model-based approach gives smaller estimated Expected Significance Levels than the design-based approach under comparable experimental circumstances.

- Suppose instead that we assume a certain pairwise interaction model for the data which is incorrect. Then we might expect the model-based procedure to give misleading results; specifically,
  - significance levels obtained from the model-based procedure may or may not be wrong; to test whether this is the case, we must check that the distribution of the observed p-values under the null hypothesis of no difference between groups is Uniform on the range $(0,1)$.
  - the parametric approach may or may not be better than the non-parametric approach in this situation. Again, we assess this by looking at Expected Significance Levels.

We must address these two questions separately, as the former requires simulations of situations where the null hypothesis is false, namely where there are differences between the groups, and the latter requires simulation of data from the null, where no between-group differences exist.

6.1. Description of simulation procedures

The simulation study proceeds in the following way:

- Simulate some replicated spatial data from model **A**, where there are r_1 replicates in group 1, r_2 replicates in group 2, and n_1 is the expected number of points per replicate in group 1, n_2 is the expected number of points per replicate in group 2. When we address the first question of the model being correct, we set $\theta_1 \neq \theta_2$, to allow us to see whether the parametric model-based procedure is superior when a difference exists between groups. When we address the issue of incorrect model specification, we set $\theta_1 = \theta_2$, to allow us to investigate the behaviour of the parametric procedure under the null hypothesis. (The non-parametric design-based procedure must perform correctly under the null, as no modelling assumptions are required by definition, and so we do not strictly need to test this).
- Perform the spatial ANOVA on the data. Obtain a p-value for the data.
- Propose a model **B** to fit the data, where **B** may or may not be the same as **A**, depending on the set-up of interest. When we address the question of the model being correct, the model specification matches that which generates the data; when we address the incorrect model specification question, the opposite is true.
- Fit the model to the data. Obtain a test statistic via the MPL approach outlined above.
- Now simulate a new data set from model **B**, and fit model **B** to the data realization. Obtain a test statistic. Repeat this 99 times. This constitutes the Monte Carlo procedure.

- Obtain a second p-value for the data.
- Repeat all of these steps many times to obtain two sets of p-values, one from the design-based approach, and one from the model-based approach.

6.2. Fundamentals for the simulation study

We fix the number of groups as 2, the number of replicates per group as 10, and the expected number of events per replicate, $E(n)$, as 30. Each replicate is fixed on a unit square. We then simulate the observed number of events to be

$$n_{ij} \sim Poisson(30). \tag{6.1}$$

We select these numbers as a compromise between an ideal simulation set-up and feasibility. If it were possible, we would simulate large numbers of replicates per group, and would look at wide varieties of combinations of r, n and g. However, computing time imposes certain constraints, and so we focus upon the above specific case. A simulation study carried out in Wilson (1998) demonstrated that ten replicates per group in the two group case is adecuate for the non-parametric procedure to give reliable results.

6.3. Case 1: Model is correct

We specify here the combinations of models **A** and **B** which are used in this case, in combination with the values of r, n and g given in the previous Section:

Table 1. (a) Data model and (b) fitted model combinations when model is correct.

(a)	(b)
Diggle	Diggle
VSC	VSC
Linear	Linear

The parameter combinations for the three models we look at are given in Table 2. Note at this stage that a parameter θ equal to 0.03, say, for the Diggle interaction process does not have the same effect as a parameter θ equal to 0.03 for the very soft core interaction process. We consider the same range of parameters principally for consistency across models.

We seek here to answer the following questions:

- When the discrepancy between θ_1 and θ_2 is large, in some sense, do both the model-based and design-based procedure perform well?
- When the discrepancy between θ_1 and θ_2 is more subtle, is the model-based procedure better at detecting the group difference?

Table 2. Data model and fitted model parameter combinations. * indicates that combination is considered, - indicates that combination is not considered.

		θ_2				
		0.03	0.05	0.07	0.08	0.09
θ_1	0.03	-	-	-	-	-
	0.05	*	-	-	-	-
	0.07	*	*	-	-	-
	0.08	*	*	*	-	-
	0.09	*	*	*	*	-

6.3.1. Results: Data = Diggle, Model = Diggle

We simulate the set-up described in Sections 6.1 and 6.2, for each pair (θ_1, θ_2) in turn. For each set-up, we estimate (a) the expected significance level for the parametric model-based procedure, (b) the expected significance level for the non-parametric design-based procedure, (c) the difference in expected significance levels for the two procedures, (d) the standard error of this difference and (e) upper and lower 95% confidence limits for the difference. The results for the Diggle data, Diggle model set-up are summarised in Table 7.

There are several cases where the parametric procedure outperforms the non-parametric (starred in the table). Moreover, in each of these cases, the absolute ESL for the parametric procedure is much smaller than that for the non-parametric. Examining these cases more closely, we see that four out of the five estimated ESLs for the parametric procedure achieve significance at the 10% level, whereas only one in the non-parametric case crosses the 10% boundary.

The initial suggestion from this, therefore, is two-fold:

- the model-based procedure is able to detect differences which are too subtle for the design-based procedure to pick up;
- when the Diggle interaction model is correctly specified, the model-based procedure performs appreciably better than the equivalent design-based procedure.

Table 7. Data model (Diggle) and fitted model (Diggle) parameter combinations. L and U indicate lower and upper 95% confidence limits respectively, on the difference in p-values from the model-based and design-based procedures. $*$ indicates a significant result in favour of the model-based procedure. Each mean is obtained from 12 simulations.

θ_1	θ_2	$E(p)$	$E(np)$	$E(p-np)$	$SE(\overline{p}-\overline{np})$	L	U
0.03	0.05	0.1292	0.3633	−0.2342	0.0911	−0.4126	−0.0557*
0.03	0.07	0.0967	0.2425	−0.1458	0.0498	−0.2434	−0.0483*
0.03	0.08	0.0925	0.2150	−0.1225	0.0591	−0.2384	−0.0066*
0.03	0.09	0.0100	0.0892	−0.0792	0.0278	−0.1336	−0.0247*
0.05	0.07	0.3892	0.2641	0.1250	0.1147	−0.0998	0.3498
0.05	0.08	0.2350	0.3112	−0.0767	0.0865	−0.2462	0.0929
0.05	0.09	0.0667	0.1550	−0.0883	0.0398	−0.1664	−0.0102*
0.07	0.08	0.3675	0.5383	−0.1708	0.1173	−0.4008	0.0591
0.07	0.09	0.2500	0.2833	−0.0333	0.0569	−0.1448	0.0782
0.08	0.09	0.4533	0.5067	−0.0533	0.0863	−0.2224	0.1158

6.3.2. Results: Data = very soft core, Model = very soft core

We look to other pairwise interaction process models, to confirm that the results observed in Section 6.3.1 are not an artefact of the Diggle process. First, we consider the very soft core model described in Section 2.2. Equivalent simulations are performed, and the results displayed in Table 8.

Table 8. Data model (VSC) and fitted model (VSC) parameter combinations. L and U indicate lower and upper 95% confidence limits respectively, on the difference in p-values from the model-based and design-based procedures. $*$ indicates a significant result in favour of the model-based procedure. Each mean is obtained from 12 simulations.

θ_1	θ_2	$E(p)$	$E(np)$	$E(p-np)$	$S.E.(\overline{p}-\overline{np})$	L	U
0.03	0.05	0.1208	0.1942	−0.0733	0.0536	−0.1783	0.0317
0.03	0.07	0.0117	0.0200	−0.0083	0.0055	−0.0191	0.0024
0.03	0.08	0.0100	0.0117	−0.0017	0.0017	−0.0049	0.0016
0.03	0.09	0.0100	0.0100	0.0000	0.0000	0.0000	0.0000
0.05	0.07	0.1700	0.2342	−0.0642	0.0675	−0.1966	0.0682
0.05	0.08	0.1442	0.2633	−0.1192	0.0492	−0.2155	−0.0228*
0.05	0.09	0.0217	0.0533	−0.0316	0.0168	−0.0646	0.0012
0.07	0.08	0.3325	0.4575	−0.1250	0.0795	−0.2808	0.0308
0.07	0.09	0.2475	0.3650	−0.1175	0.0706	−0.2558	0.0208
0.08	0.09	0.3150	0.3292	−0.0147	0.0722	−0.1561	0.1268

The absolute ESLs estimated for the design-based procedure are in general smaller than they were for the Diggle model in the previous section. We suggest that this is because the very soft core process has infinite range, whereas the Diggle interaction function is 1 beyond θ for any given θ.

6.3.3. Results: Data = Linear, Model = Linear

Again, we repeat the simulations, this time setting the data generating model and the fitted model to have the linear interaction function defined in Section 2.2. Broadly similar results are observed (see Table 9), namely that the parametric model-based procedure gives lower ESLs than the non-parametric design-based, and the parametric procedure detects differences at a reasonable level of significance when the non-parametric fails to do so.

Table 9. Data model (Linear) and fitted model (Linear) parameter combinations. L and U indicate lower and upper 95% confidence limits respectively, on the difference in p-values from the model-based and design-based procedures. $*$ indicates a significant result in favour of the model-based procedure. Each mean is obtained from 12 simulations.

θ_1	θ_2	$E(p)$	$E(np)$	$E(p-np)$	$S.E.(\overline{p}-\overline{np})$	L	U
0.03	0.05	0.3625	0.4208	-0.0583	0.1477	-0.3479	0.2312
0.03	0.07	0.1192	0.3225	-0.2033	0.0668	-0.3344	-0.0723^*
0.03	0.08	0.0350	0.1067	-0.0717	0.0354	-0.1411	-0.0022^*
0.03	0.09	0.0450	0.0700	-0.0250	0.0126	-0.0498	-0.0002^*
0.05	0.07	0.2108	0.3892	-0.1783	0.0640	-0.3038	-0.0528^*
0.05	0.08	0.0992	0.2083	-0.1092	0.0332	-0.1742	-0.0441^*
0.05	0.09	0.1783	0.1983	-0.0200	0.0599	-0.1375	0.0975
0.07	0.08	0.4950	0.5342	-0.0392	0.0713	-0.1789	0.1006
0.07	0.09	0.5600	0.4800	0.0800	0.0873	-0.0911	0.2512
0.08	0.09	0.4120	0.4120	0.0000	0.0781	-0.1531	0.1531

6.3.4. Conclusions

We have thus shown that, under correct model specification, the parametric model-based procedure outperforms the non-parametric ANOVA, for a variety of classes of pairwise interaction process. There are situations where both procedures do equivalently well; however, when this is the case, the difference between the θ parameters in both groups tends to be non-trivial. Similarly, as we might expect, there are also situations where neither procedure detects a difference which exists; this is generally in cases where the difference between θs is very small.

6.4. Case 2: Model is incorrect

The alternative question of interest concerns the relative performance of the two procedures when the underlying model is mis-specified; specifically, does the model-based procedure give misleading results under this case scenario?

We specify again the combinations of models **A** and **B** which are used in this case in Table 3.

Table 3. (a) Data model and (b) fitted model combinations when model is incorrect.

(a)	(b)
Diggle	VSC
Linear	VSC
VSC	Linear

The parameter combinations we look at are given in Tables 4, 5 and 6.

6.4.1. Data Diggle, fitted model very soft core

We begin by looking at the results from the case where the data is simulated from a Diggle model, and the very soft core model is fitted. The examined parameter values are given in Table 4.

Table 4. Data model (Diggle) and fitted model (VSC) parameter combinations.

θ_1	0.03	0.05	0.07	0.09	0.11
θ_2	0.03	0.05	0.07	0.09	0.11

The plots in Figures 2.1 to 2.5 demonstrate no departure from Uniformity over the required $(0,1)$ range, for either the parametric or non-parametric procedure. This indicates that the correct result holds under the null hypothesis, when a very soft core model is fitted to data with the Diggle potential function, with parameter 0.03, 0.05 or 0.07. We suggest that the very soft core approximation is good as a consequence of the fact that the interaction functions are somehow *similar*, in that they are both smooth, and one can follow the shape of the other quite closely (*viz.* Figures 3.1 to 3.5).

We note, however, that for the case of $\theta = 0.09$, the parametric model-based procedure yields a marginally significant departure from Uniformity over the required range. However, the expected uniformity is once again observed for the case where $\theta = 0.11$, and so the departure in the $\theta = 0.09$ may be spurious.

Figure (2.1) $\theta_1 = \theta_2 = 0.03$. For parametric, K-S = 0.135 ($p = 0.4223$); for non-parametric, K-S = 0.165 ($p = 0.2024$).

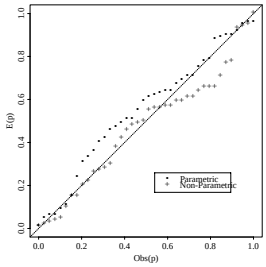

Figure (2.2) $\theta_1 = \theta_2 = 0.05$. For parametric, K-S = 0.100 ($p = 0.6623$); for non-parametric, K-S = 0.140 ($p = 0.5614$).

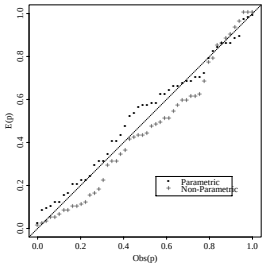

Figure (2.3) $\theta_1 = \theta_2 = 0.07$. For parametric, K-S = 0.090 ($p = 0.7793$); for non-parametric, K-S = 0.120 ($p = 0.4338$).

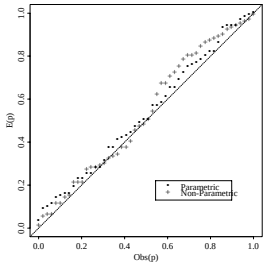

Figure (2.4) $\theta_1 = \theta_2 = 0.09$. For parametric, K-S = 0.180 ($p = 0.0688$); for non-parametric, K-S = 0.140 ($p = 0.2561$).

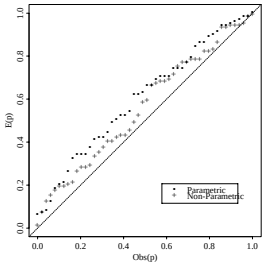

Figure (2.5) $\theta_1 = \theta_2 = 0.11$. For parametric, K-S = 0.100 ($p = 0.6623$); for non-parametric, K-S = 0.180 ($p = 0.0688$).

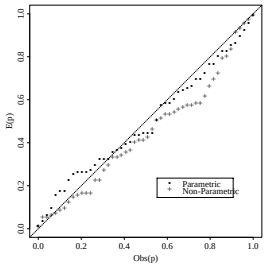

Figure 2. Observed versus Expected p-values under H_o, in the case where the data are simulated from the Diggle model, and a very soft core model is fitted, with the parameters indicated. In each case, «K-S» indicates the Kolmogorov-Smirnov test statistic (associated p-value in parentheses). (a) is the value in the model-based case, (b) in the design-based.

Figure (3.1) $\theta = 0.03$.

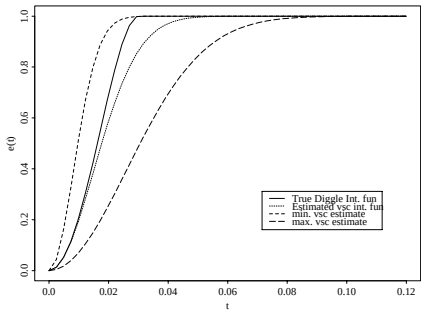

Figure (3.2) $\theta = 0.05$.

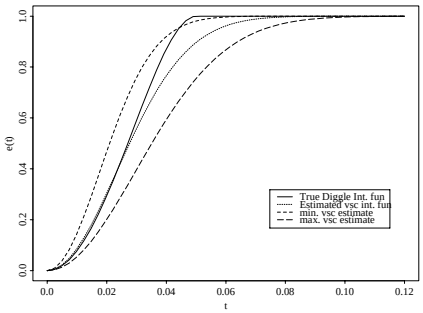

Figure (3.3) $\theta = 0.07$.

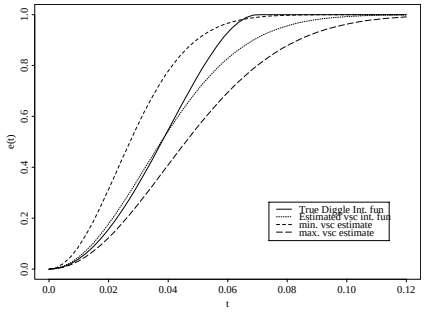

Figure (3.4) $\theta = 0.09$.

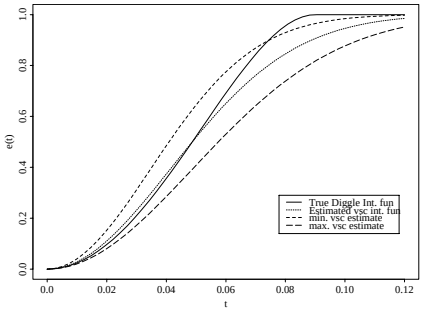

Figure (3.5) $\theta = 0.11$.

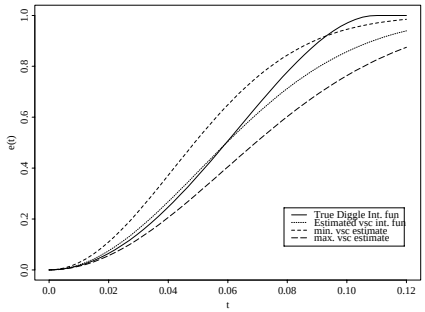

Figure 3. Interaction functions for true underlying process (Diggle potential with indicated parameter) and mean and range of fitted very soft core parameters, in each case from repeated simulations.

We conclude from this part of the study that the very soft core process provides a good model for data which has an underlying Diggle potential function.

6.4.2. Data Linear, fitted model very soft core

To provide a contrast, we now look at how well the very soft core process can model data for which the underlying interaction function is linear. We hypothesise that the model-based procedure is more likely to break down here than in the Diggle/very soft core case, as a consequence of the fact that the interaction functions are more dissimilar.

The parameter values we examine are given in Table 5.

Table 5. Data model (Linear) and fitted model (VSC) parameter combinations.

θ_1	0.03	0.05	0.06	0.08	0.10	0.11	0.12	0.13	0.14
θ_2	0.03	0.05	0.06	0.08	0.10	0.11	0.12	0.13	0.14

Again, we fix the n_{ij} and r_i. From repeated simulations of the experimental set-ups described above, we obtain the sets of p-values and associated Kolmogorov-Smirnov statistics illustrated in Figures 4.1 to 4.9.

The situation here is different in that, as suspected, the model-based procedure breaks down for certain parameter values. For the smaller parameters, clear departures from uniformity are demonstrated. As the parameters become larger, the p-values tend towards uniformity, however; we suggest that this is because as the parameter gets larger, the linear potential function becomes smoother in some sense, and so the very soft core model is better equipped to approximate the linear in this case (*viz.* Figures 5.1 to 5.9).

6.4.3. Data Very Soft Core, fitted model Linear

We look at the opposite situation, namely where the data is from a very soft core model and we attempt to fit a linear interaction function to it. The parameter values examined are displayed in Table 6.

Table 6. Data model (VSC) and fitted model (Linear) parameter combinations.

θ_1	0.10	0.11	0.12	0.13	0.14
θ_2	0.10	0.11	0.12	0.13	0.14

We repeat the procedure, and obtain the p-values and associated Kolmogorov-Smirnov statistics shown in Figures 6.1 to 6.5.

Figure (4.1) $\theta_1 = \theta_2 =$ 0.03. (a) K-S = 0.303 ($p = 0.0024$); (b) K-S = 0.094 ($p = 0.8860$).

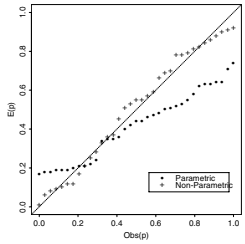

Figure (4.2) $\theta_1 = \theta_2 =$ 0.05. (a) K-S = 0.370 ($p = 0.0238$); (b) K-S = 0.223 ($p = 0.3854$).

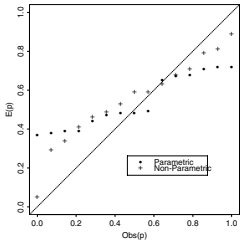

Figure (4.3) $\theta_1 = \theta_2 =$ 0.06. (a) K-S = 0.480 ($p = 0.0122$); (b) K-S = 0.290 ($p = 0.3067$).

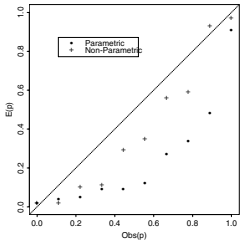

Figure (4.4) $\theta_1 = \theta_2 =$ 0.08. (a) K-S = 0.273 ($p = 0.0248$); (b) K-S = 0.204 ($p = 0.1678$).

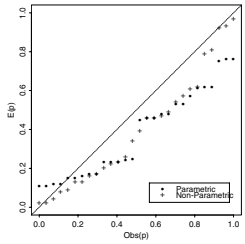

Figure (4.5) $\theta_1 = \theta_2 =$ 0.10. (a) K-S = 0.270 ($p = 0.0129$); (b) K-S = 0.1964 ($p = 0.1369$).

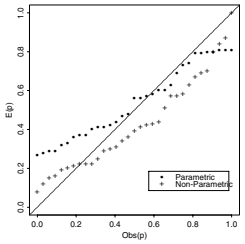

Figure (4.6) $\theta_1 = \theta_2 =$ 0.11. (a) K-S = 0.260 ($p = 0.0281$); (b) K-S = 0.090 ($p = 0.9502$).

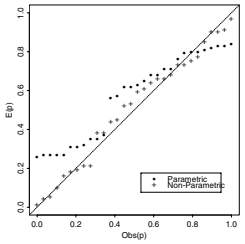

Figure (4.7) $\theta_1 = \theta_2 =$ 0.12. (a) K-S = 0.120 ($p = 0.4895$); (b) K-S = 0.0948 ($p = 0.7676$).

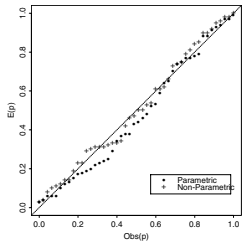

Figure (4.8) $\theta_1 = \theta_2 =$ 0.13. (a) K-S = 0.125 ($p = 0.3394$); (b) K-S = 0.147 ($p = 0.1721$).

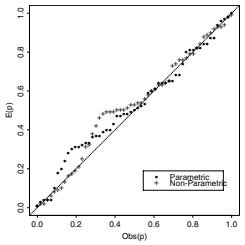

Figure (4.9) $\theta_1 = \theta_2 =$ 0.14. (a) K-S = 0.170 ($p = 0.1262$); (b) K-S = 0.092 ($p = 0.7955$).

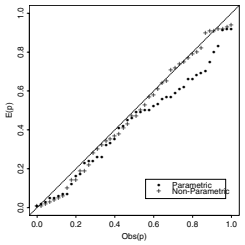

Figure 4. Observed versus Expected *p*-values under H_o, in the case where the data are simulated from the Linear model, and a very soft core model is fitted, with the parameters indicated. In each case, «K-S» indicates the Kolmogorov-Smirnov test statistic (associated *p*-value in parentheses). (a) is the value in the model-based case, (b) in the design-based.

Figure(5.1) $\theta = 0.03$.

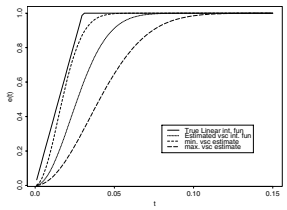

Figure (5.2) $\theta = 0.05$.

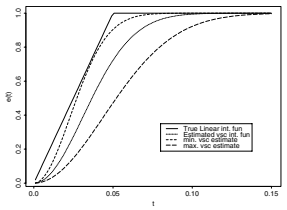

Figure (5.3) $\theta = 0.06$.

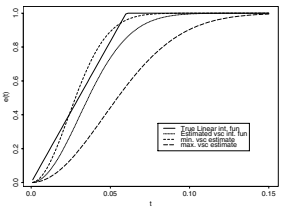

Figure (5.4) $\theta = 0.08$.

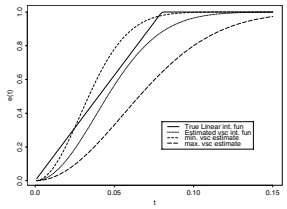

Figure (5.5) $\theta = 0.10$.

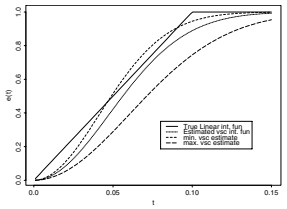

Figure (5.6) $\theta = 0.11$.

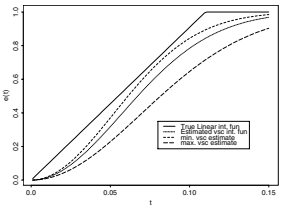

Figure (5.7) $\theta = 0.12$.

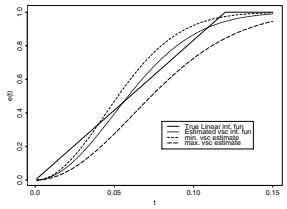

Figure (5.8) $\theta = 0.13$.

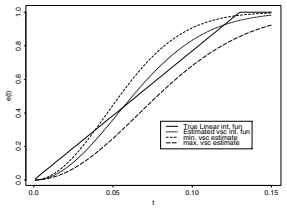

Figure (5.9) $\theta = 0.14$.

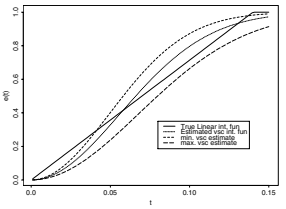

Figure 5. Interaction functions for true underlying process (Linear potential with indicated parameter) and mean and range of fitted very soft core parameters, in each case from repeated simulations.

Figure (6.1) $\theta_1 = \theta_2 = 0.10$. (a) K-S = 0.333 ($p = 0.1087$); (b) K-S = 0.343 ($p = 0.0912$).

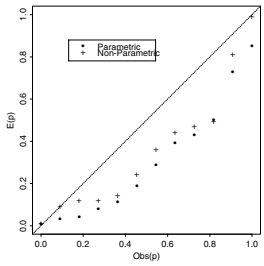

Figure (6.2) $\theta_1 = \theta_2 = 0.11$. (a) K-S = 0.1900 ($p = 0.1741$); (b) K-S = 0.141 ($p = 0.5013$).

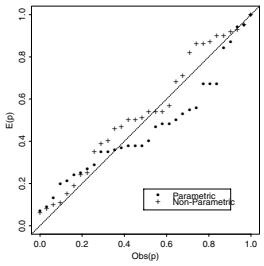

Figure (6.3) $\theta_1 = \theta_2 = 0.12$. (a) K-S = 0.340 ($p = 0.0145$); (b) K-S = 0.150 ($p = 0.7045$).

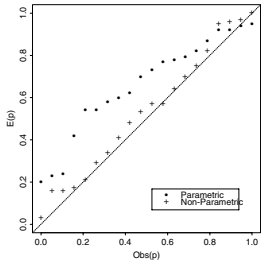

Figure (6.4) $\theta_1 = \theta_2 = 0.13$. (a) K-S = 0.237 ($p = 0.0294$); (b) K-S = 0.090 ($p = 0.9073$).

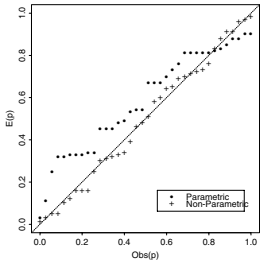

Figure (6.5) $\theta_1 = \theta_2 = 0.14$. (a) K-S = 0.302 ($p = 0.0366$); (b) K-S = 0.1995 ($p = 0.3284$).

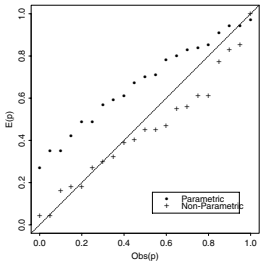

Figure 6. Observed versus Expected *p*-values under H_o, in the case where the data are simulated from the very soft core model, and a linear model is fitted, with the parameters indicated. In each case, «K-S» indicates the Kolmogorov-Smirnov test statistic (associated *p*-value in parentheses). (a) is the value in the model-based case, (b) in the design-based.

Figure (7.1) $\theta = 0.10$.

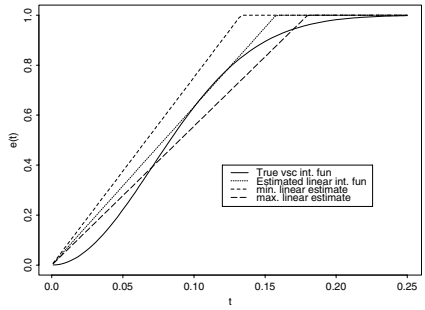

Figure (7.2) $\theta = 0.11$.

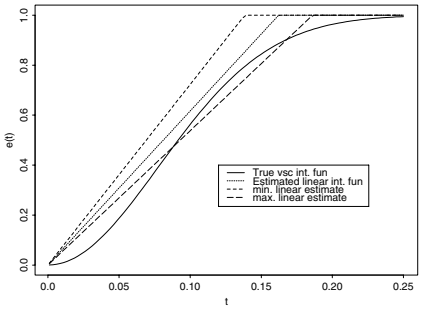

Figure (7.3) $\theta = 0.12$.

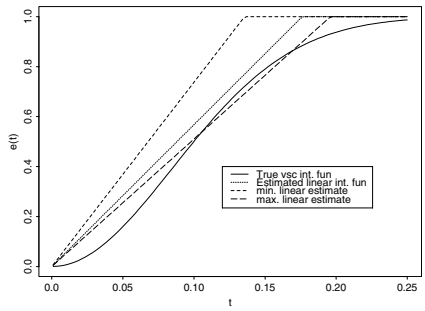

Figure (7.4) $\theta = 0.13$.

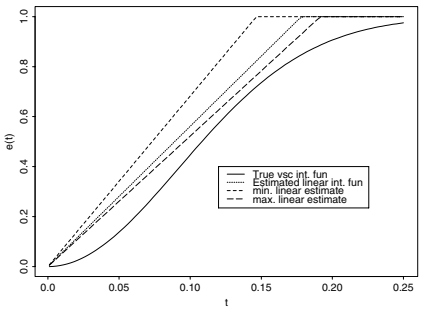

Figure (7.5) $\theta = 0.14$.

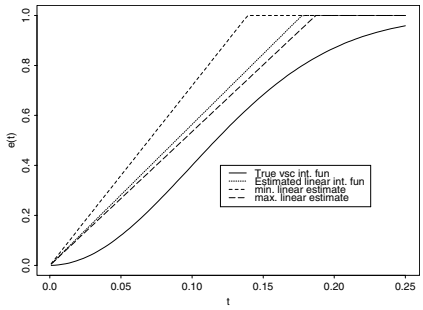

Figure 7. Interaction functions for true underlying process (Very soft core potential with indicated parameter) and mean and range of fitted linear parameters, in each case from repeated simulations.

The results can be explained once again by the relationship between the interaction functions in the two types of process. Now the parametric model-based procedure breaks down for the larger parameters, for which clear departures from uniformity are again demonstrated. Note again that as the parameter gets larger, the linear potential function is no longer equipped to fit the very soft core model (*viz.* Figures 7.1 to 7.5).

6.4.4. Conclusions

When both the true and the assumed model were either VSC or Diggle model, the parametric test led to empirical distributions of *p*-values which were well fitted by the uniform distribution on $(0,1)$; note that both models have differentiable interaction functions. Significant departures from uniformity arose when the non-differentiable linear model was used to simulate the data but VSC model was assumed, and *vice versa*. Note that the number of simulations contributing to Case 2 situation varied between 20 and 50 because of limitations on the available computing time; in general, simulations of pairwise interaction point processes become time-consuming as the strength of the interaction between points increases.

7. DISCUSSION

Pairwise interaction point processes are widely used as descriptive models for spatial pattern recognition. Approximate likelihood-based methods of inference are available for these models, but rely on computer-intensive Monte Carlo methods of inference. Whether this matters in practice depends on practical constraints which vary between applications, but it is a potentially limiting factor for data consisting of large numbers of replicate patterns. There are many interesting areas, such as medicine, biology or neuroanatomy, in which replicated point pattern data frequently arise. In this context, individual patterns often consist of spatial point coordinates, where it is straightforward to obtain large numbers of such patterns by repeated sectioning, either within or between experimental subjects.

In this study, we have confirmed our suspicion that the parametric model-based approach is more efficient at recognizing the spatial structure than an equivalent non-parametric design-based approach under correct model specification. In addition, we have ascertained that the performance of the parametric procedure under model misspecification depends upon the type of model being considered. Specifically, we have determined that the pairwise interaction process with Diggle potential and that with very soft core potential can generally model each other quite well. In contrast, neither of these two models can be applied successfully to data with linear potential function, and the reverse applies. In summary, when the model driving the data is unable to be approximated, in some sense, by the model being fitted, the procedure appears to break down.

ACKNOWLEDGEMENTS

This paper has been written after several visits of the author to Lancaster University, and it is a consequence of the fruitful discussions with Prof. Peter Diggle and Dr. Helen Wilson. The referees are also acknowledged for their suggestions on an earlier draft.

REFERENCES

Baddeley, A. J., Reid, S., Howard, V. & Boyde, A. (1985). «Unbiased estimation of particle density in the tandem scanning reflected light microscope». *Journal of Microscopy*, 138, 203-212.

Baddeley, A. J. & Møller, J. (1989). «Nearest-neighbour Markov point processes and random sets». *International Statistical Review*, 57, 89-121.

Baddeley, A. J., Moyeed, R., Howard, C. & Boyde, A. (1993). «Analysis of a three-dimensional point pattern with replication». *Journal of Applied Statistics*, 42, 641-646.

Besag, J. (1977). «Some methods of statistical analysis for spatial data». *Bulletin of the International Statistical Institute*, 47, 77-92.

Besag, J., Milne, R. & Zachary, S. (1982). «Point process limits of lattice processes». *Journal of Applied Probability*, 19, 210-216.

Cressie, N. A. (1993). *Statistics for spatial data*, Wiley, Revised edition.

Daley, D. J. & Vere-Jones, D. (1988). *Introduction to the theory of point processes*. New York: Springer.

Dempster, A. P. & Schatzoff, M. (1965). «Expected Significance Level as a sensitivity index for test statistics». *Journal of the American Statistical Association*, 60, 420-436.

Diggle, P. J. (1983). *Statistical analysis of spatial point patterns*. Academic Press, London.

Diggle, P. J. (1986). «Parametric and non-parametric estimation for pairwise interaction point processes». *Proceedings of the 1st World Congress of the Bernoulli Society*.

Diggle, P. J., Lange, N. & Beněs, F. (1991). «Analysis of variance for replicated spatial point patterns in clinical neuroanatomy». *Journal of the American Statistical Association*, 86, 415, 618-625.

Diggle, P. J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. & Tanemura, M. (1994). «On parameter estimation for pairwise interaction point processes». *International Statistical Review*, 62, 99-117.

Fiksel, T. (1984). «Estimation of parametrized pair potentials of marked and non-marked Gibbsian point processes». *Elektron. Inform. Kybernet.*, 20, 270-278.

Fiksel, T. (1988). obrir Estimation of interaction potentials of Gibbsian point processes». *Math. Operationsf. Statist. Ser. Statist.*, 19, 77-86.

Gates, D. J. & Westcott, M. (1986). «Clustering estimates for spatial point distributions with unstable potentials». *Ann. Inst. Statist. Math.*, 38, 123-135.

Geman, S. & Geman, D. (1984). «Stochastic relaxation, Gibbs distributions and the bayesian restoration of images». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

Geyer, C. J. & Thompson, E. A. (1992). «Constrained Monte Carlo maximum likelihood for dependent data (with discussion)». *Journal of the Royal Statistical Society*, B 54, 657-699.

Geyer, C. J. & Møller, J. (1994). «Simulation and likelihood inference for spatial point processes». *Scandinavian Journal of Statistics*, 21, 359-373.

Hastings, W. K. (1970). «Monte Carlo sampling methods using Markov chains and their applications». *Biometrika*, 57, 97-109.

Jensen, J. L. (1993). «Asymptotic normality of estimates in spatial point processes». *Scandinavian Journal of Statistics*, 20, 97-109.

Jensen, J. L. & Møller, J. (1991). «Pseudolikelihood for exponential family models of spatial point processes». *Annals of Applied Probability*, 1, 445-461.

Kelly, F. P. & Ripley, B. D. (1976). «On Strauss' model for clustering». *Biometrika*, 63, 357-60.

Møller, J. (1992). «Discussion contribution». *Journal of the Royal Statistical Society*, B 54, 692-693.

Møller, J. (1993). «Spatial point processes and Markov chain Monte Carlo methods». Lecture at *Conference on Stochastic Processes and their Applications*, Amsterdam.

Moyeed, R. A. & Baddeley, A. J. (1991). «Stochastic approximation for the MLE of a spatial point process». *Scandinavian Journal of Statistics*, 18, 39-50.

Ogata, Y. & Tanemura, M. (1981). «Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure». *Annals of the Institute of Statistical Mathematics*, 33 B, 315-338.

Ogata, Y. & Tanemura, M. (1984). «Likelihood analysis of spatial point patterns». *Journal of the Royal Statistical Society*, B 46, 496-518.

Ogata, Y. & Tanemura, M. (1989). «Likelihood estimation of soft-core interaction potentials for Gibbsian point patterns». *Annals of the Institute of Statistical Mathematics*, 41 B, 583-600.

Penttinen, A. (1984). «Modelling interaction in spatial point patterns: parameter estimation by the maximum likelihood method». Number **7** in *Jyvaskyla Studies in Computer Science, Economics and Statistics*.

Preston, C. J. (1977). «Spatial birth-and-death processes». *Bull. Inst. Intern. Statist.*, 46, 371-391.

Ripley, B. D. (1977). «Modelling spatial patterns (with Discussion)». *Journal of Royal Statistical Society*, B 39, 172-212.

Ripley, B. D. (1981). *Spatial Statistics*. Wiley, New York.

Ripley, B. D. (1988). *Statistical inference for spatial processes*. Cambridge, Cambridge University Press.

Ripley, B. D. & Kelly, F. P. (1977). «Markov point processes». *J. London Math. Soc.*, 15, 188-192.

Strauss, D. J. (1975). «A model for clustering». *Biometrika*, 63, 467-475.

Takacs, R. (1986). «Estimator for the pair-potential of a Gibbsian point process». *Math. Operationsf. Statist. Ser. Statist.*, 17, 429-433.

Wilson, H. E. (1998). «Statistical analysis of replicated spatial point patterns». *Unpublished PhD. thesis*, University of Lancaster, U.K.

Wilson, H. E., Diggle, P. J. and Howard, C. V. (1998). «Methods for the analysis of replicated spatial point patterns in clinical neuroanatomy». *Advances in Applied Probability*, 30(2). Abstract in proceedings of 9th International Workshop in Stereology, Stochastic Geometry and Image Analysis, Comillas, October 1997.