

**PRACTICAL MAXIMUM PSEUDOLIKELIHOOD
FOR SPATIAL POINT PATTERNS**
(with Discussion)

ADRIAN BADDELEY^{1*} AND ROLF TURNER²

University of Western Australia

Summary

This paper describes a technique for computing approximate maximum pseudolikelihood estimates of the parameters of a spatial point process. The method is an extension of Berman & Turner's (1992) device for maximizing the likelihoods of inhomogeneous spatial Poisson processes. For a very wide class of spatial point process models the likelihood is intractable, while the pseudolikelihood is known explicitly, except for the computation of an integral over the sampling region. Approximation of this integral by a finite sum in a special way yields an approximate pseudolikelihood which is formally equivalent to the (weighted) likelihood of a loglinear model with Poisson responses. This can be maximized using standard statistical software for generalized linear or additive models, provided the conditional intensity of the process takes an 'exponential family' form. Using this approach a wide variety of spatial point process models of Gibbs type can be fitted rapidly, incorporating spatial trends, interaction between points, dependence on spatial covariates, and mark information.

Key words: area-interaction process; Berman–Turner device; Dirichlet tessellation; edge effects; generalized additive models; generalized linear models; Gibbs point processes; GLIM; hard core process; inhomogeneous point process; marked point processes; Markov spatial point processes; Ord's process; pairwise interaction; profile pseudolikelihood; spatial clustering; soft core process; spatial trend; S-PLUS; Strauss process; Widom–Rowlinson model.

1. Introduction

This paper describes a computational device for rapidly fitting statistical models to spatial point patterns. Applications are shown in Section 10. Datasets may consist of points in two or three dimensions or in space–time; the points may be classified into different types or carry auxiliary observations ('marks'). Additionally there may be spatial covariates, such as topography or another spatial pattern observed in the same region. Realistic models for such data should incorporate both spatial inhomogeneity ('trend') and dependence between points ('interaction' such as clustering or regularity). Ogata & Tanemura (1981, 1984, 1985, 1986)

Received August 1998; revised May 1999; accepted June 1999.

* Author to whom correspondence should be addressed.

¹ Dept of Mathematics and Statistics, University of Western Australia, Nedlands, WA 6907.

e-mail: adrian@maths.uwa.edu.au

² Dept of Mathematics and Statistics, University of New Brunswick, Fredericton, NB, Canada E3B 5A3.

Acknowledgments. The authors warmly thank an anonymous referee and Y.C. Chin, M. Tanemura, H. Wilson, A. Särkkä, A. Penttinen and M. Van Lieshout for helpful comments, P.J. Diggle, V. Isham, Y. Ogata and A. Särkkä for supplying data files, and R.D. Harkness, M. Numata and M. Tanemura for their kind permission to use the data. Data for Figure 12 were supplied by Professor R.D. Harkness, Professor V. Isham and Dr A. Särkkä. The authors' research was supported by the Australian Research Council and the Natural Sciences and Engineering Research Council of Canada.

and Penttinen (1984) developed methods for maximum likelihood estimation for such models, and applied them to real data. Recent advances have been made by Geyer & Møller (1994), Geyer (1998) and others. However, maximum likelihood is computationally intensive, and employs simulation algorithms that are specific to the chosen model. It is even more costly for inhomogeneous spatial patterns because of increased parameter dimensionality and complexity of simulation. This militates against the modern statistical practice of fitting several alternative models to the same dataset and introducing smooth functions as model terms. Few writers apart from Ogata & Tanemura (1986) have fitted inhomogeneous point process models, other than the inhomogeneous Poisson process, to real spatial data.

Berman & Turner (1992) introduced a technique for maximizing the likelihoods of (a) general point processes in time, and (b) inhomogeneous Poisson processes in d -dimensional space. The intensity or conditional intensity of the process was assumed to be loglinear in the parameters. They approximated the log-likelihood by a finite sum that had the same analytical form as the (weighted) log-likelihood of a generalized linear model with Poisson responses. The approximate likelihood could then be maximized using existing software for generalized linear models. Related ideas have been explored by Lindsey (1992, 1995, 1996) and Lindsey & Mersch (1992).

In this paper we extend the Berman–Turner device to a much larger class of spatial point process models, namely Gibbs point processes with exponential family likelihoods. We obtain an approximation to the pseudolikelihood (Besag, 1975; Besag, 1977; Jensen & Møller, 1991) rather than to the likelihood. The maximum pseudolikelihood estimator is a practical alternative to the maximum likelihood estimator (MLE), satisfies unbiased estimating equations, and is consistent and asymptotically normal under suitable conditions. The MLE is not necessarily optimal here because the usual asymptotic theory is not applicable. Under reasonable assumptions (Diggle *et al.*, 1994) the maximum pseudolikelihood normal equations are a special case of the Takacs–Fiksel estimating equations, an application of the method of moments (Fiksel, 1984; Takacs, 1986; Fiksel, 1988). Using the extended Berman–Turner device and standard statistical software, we can rapidly fit quite complex spatial stochastic models involving spatial trends and spatial covariates as well as interactions between points.

The plan of the paper is as follows. Sections 2 and 3 give definitions and background. Section 4 presents our extension of the Berman–Turner computational device. Section 5 treats a simple example. Section 6 develops applications of the method to specific models of spatial interaction; Section 7 develops applications for spatial inhomogeneity, and Section 8 develops applications for marked point patterns. Section 9 treats some issues in estimation and inference. The method is applied to real datasets in Section 10. Section 11 reports a simulation study of the accuracy of the technique.

2. Background and definitions

2.1. Likelihoods

The data consist of a spatial point pattern \mathbf{x} observed in a bounded region W of space. Thus

$$\mathbf{x} = \{x_1, \dots, x_n\},$$

where the number of points $n \geq 0$ is not fixed, and each x_i is a point in W . The region W is a known, bounded subset of d -dimensional space \mathbb{R}^d , where $d \geq 1$. Sections 7.1–7.2 and 8

discuss extensions of this basic setup to incorporate spatial covariates and marked points, respectively.

The data \mathbf{x} are assumed to be a realization of a random point process X in W . Typically the null model (or the null hypothesis) is the homogeneous Poisson point process (Cox & Isham, 1980; Kingman, 1993). Other models are specified by their likelihood with respect to the Poisson process. Thus we assume X has a probability density $f(\mathbf{x}; \theta)$ with respect to the distribution of the Poisson process with intensity 1 on W . Additionally we assume $f(\mathbf{x}; \theta) > 0$ implies $f(\mathbf{y}; \theta) > 0$ for all subsets $\mathbf{y} \subset \mathbf{x}$. This is the class of Gibbs processes on W (see Preston, 1976; Ripley, 1989; Stoyan, Kendall & Mecke, 1995). The distribution is governed by a vector parameter θ ranging over a set $\Theta \subseteq \mathbb{R}^p$, see Cox & Isham (1980), Geyer (1998).

2.2. Basic models

Specific models are detailed in Sections 6–8, but it is instructive to list three important examples. First, the homogeneous Poisson process with intensity $\lambda > 0$ has density

$$f(\mathbf{x}; \lambda) = e^{-(\lambda-1)|W|} \lambda^{n(\mathbf{x})},$$

where $n(\mathbf{x})$ denotes the number of points in \mathbf{x} and $|W|$ is the volume of W . This yields the maximum likelihood estimate $\hat{\lambda} = n(\mathbf{x})/|W|$.

Second, consider the inhomogeneous Poisson process on W with rate or intensity function $\lambda : W \rightarrow \mathbb{R}$; see Cox & Isham (1980), Kingman (1993). In statistical models, the intensity $\lambda_\theta(u)$ will depend on θ to reflect ‘spatial trend’ (a change in intensity across the region of observation) or dependence on a covariate. The density is

$$f(\mathbf{x}; \theta) = \prod_{i=1}^{n(\mathbf{x})} \lambda_\theta(x_i) \exp\left(-\int_W [\lambda_\theta(u) - 1] du\right). \tag{1}$$

Maximization of (1) generally requires iterative optimization methods.

Third, the pairwise interaction process on W with trend or activity function $b_\theta : W \rightarrow \mathbb{R}_+$ and interaction function $h_\theta : W \times W \rightarrow \mathbb{R}_+$ has density

$$f(\mathbf{x}; \theta) = \alpha(\theta) \prod_{i=1}^{n(\mathbf{x})} b_\theta(x_i) \prod_{i < j} h_\theta(x_i, x_j), \tag{2}$$

where $\alpha(\theta) > 0$ is the normalizing constant. Conditions must be imposed on b_θ and h_θ to ensure the density is well defined and integrable: in particular $h_\theta(u, v) = h_\theta(v, u)$. Examples are given in Section 6. See the excellent surveys by Ripley (1988, 1989). Pairwise interaction models are suitable for the data in Figures 6 and 12, as shown by Ogata & Tanemura (1981, 1984, 1985, 1986), Särkkä (1993), and Takacs & Fiksel (1986). The terms $b_\theta(x_i)$ in (2) influence the intensity of points and introduce a spatial trend if $b_\theta(\cdot)$ is not constant. The terms $h_\theta(x_i, x_j)$ introduce dependence (‘interaction’) between different points of the process X . If $h_\theta \equiv 1$ the model reduces to an inhomogeneous Poisson process with intensity function $b_\theta(u)$.

The normalizing constant $\alpha(\theta)$ in (2) is generally an intractable function of θ . Methods for approximating $\alpha(\cdot)$ and maximizing likelihood include functional expansions of $\alpha(\cdot)$,

Monte Carlo integration, and analogues of E-M and stochastic approximation (Ogata & Tanemura, 1981, 1984, 1985, 1986; Penttinen, 1984; Moeed & Baddeley, 1991; Geyer, 1998).

Most models considered in this paper are pairwise interaction processes, but we also discuss the Widom–Rowlinson (‘area-interaction’) model (Section 6.2) and Ord’s model (Section 6.3).

3. Pseudolikelihood

It is generally difficult to evaluate and maximize the likelihoods of point processes other than the inhomogeneous Poisson process (1). Even simple exponential family models such as the pairwise interaction processes (2) include a normalizing constant which is an intractable function of θ . An alternative to the likelihood function is the pseudolikelihood (Besag, 1975, 1977; Besag, Milne & Zachary, 1982; Jensen & Møller, 1991) which we describe here; see Fiksel (1984, 1988); Takacs (1986); Ripley (1988, 1989); Särkkä (1993); Diggle *et al.* (1994) for other applications.

Originally Besag (1975, 1977) defined the pseudolikelihood of a finite set of random variables X_1, \dots, X_n as the product of the conditional likelihoods of each X_i given the other variables $\{X_j, j \neq i\}$. This was extended (Besag, 1977; Besag *et al.*, 1982) to point processes, for which it can be viewed as an infinite product of infinitesimal conditional probabilities.

3.1. Conditional intensity

To construct the pseudolikelihood we need the (Papangelou) conditional intensity $\lambda(u; \mathbf{x})$ of X at a location $u \in W$. This may be loosely interpreted as giving the conditional probability that X has a point at u given that the rest of the process coincides with \mathbf{x} . See Kallenberg (1984) for an informal introduction, or Glötzl (1980a,b), Kallenberg (1983), Kozlov (1976) for details.

For any Gibbs process on W (see Section 2) with density f , the conditional intensity at a point $u \in W$ is

$$\lambda(u; \mathbf{x}) = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x})} \quad (u \notin \mathbf{x}), \quad (3)$$

$$\lambda(x_i; \mathbf{x}) = \frac{f(\mathbf{x})}{f(\mathbf{x} \setminus \{x_i\})} \quad (x_i \in \mathbf{x}). \quad (4)$$

For example, the inhomogeneous Poisson process with intensity function $\lambda(\cdot)$ has conditional intensity $\lambda(u; \mathbf{x}) = \lambda(u)$ at all points u . The fact that this does not depend on \mathbf{x} is a consequence of the independence properties of the Poisson process. For a general Gibbs point process, $\lambda(u; \mathbf{x})$ depends on \mathbf{x} . The general pairwise interaction process (2) has conditional intensity

$$\lambda_\theta(u; \mathbf{x}) = b_\theta(u) \prod_{\substack{i=1 \\ x_i \neq u}}^{n(\mathbf{x})} h_\theta(u, x_i). \quad (5)$$

Note that $\lambda_\theta(\cdot; \mathbf{x})$ is discontinuous at the data points x_i , and that the intractable normalizing constant in (2) has been eliminated in the conditional intensity.

3.2. Definition of pseudolikelihood

Besag (1977) defined the pseudolikelihood of a point process with conditional intensity $\lambda_\theta(u; \mathbf{x})$ over a subset $A \subseteq W$ to be

$$PL_A(\theta; \mathbf{x}) = \left(\prod_{x_i \in A} \lambda_\theta(x_i; \mathbf{x}) \right) \exp \left(- \int_A \lambda_\theta(u; \mathbf{x}) du \right) \tag{6}$$

and gave examples of the utility of maximum pseudolikelihood estimates. Further theory was developed in Besag *et al.* (1982), Jensen & Møller (1991) and Jensen & Künsch (1994).

If the process is Poisson, the pseudolikelihood coincides with the likelihood (1) up to the factor $\exp(|W|)$. For a pairwise interaction process (2), the pseudolikelihood is

$$PL(\theta; \mathbf{x}) = \left(\prod_{i=1}^{n(\mathbf{x})} b_\theta(x_i) \prod_{i \neq j} h_\theta(x_i, x_j) \right) \exp \left(- \int_W b_\theta(u) \prod_{i=1}^{n(\mathbf{x})} h_\theta(u, x_i) du \right); \tag{7}$$

the intractable normalizing constant $\alpha(\theta)$ appearing in the likelihood (2) has been replaced by an exponential integral in (7) as if the process were Poisson. We give other examples in Sections 6 and 7 below.

For processes with ‘weak interaction’ in the sense that $\lambda_\theta(u; \mathbf{x})$ can be approximated well by a function of u only, the process is approximately Poisson and the pseudolikelihood is an approximation to the likelihood. Hence the maximum pseudolikelihood estimator (MPLE) should be efficient if interaction is weak. Folklore holds that it is inefficient for strong interactions.

3.3. Loglinear case

In this paper we focus on Gibbs point process models for which the conditional intensity is loglinear:

$$\lambda_\theta(u; \mathbf{x}) = \exp(\theta^\top S(u; \mathbf{x})), \tag{8}$$

where $S(u; \mathbf{x})$ is a vector of spatial covariates defined at each point u in W . This includes exponential family likelihoods with canonical parameter θ .

Assume $\|S(u; \mathbf{x})\| \exp(\theta^\top S(u; \mathbf{x}))$ is uniformly bounded in $u \in W$ and $\theta \in \Theta$, for each fixed \mathbf{x} . Then the maximum pseudolikelihood normal equations

$$\frac{\partial}{\partial \theta} \log PL_A(\theta; \mathbf{x}) = 0$$

become

$$\sum_{x_i \in A} S(x_i; \mathbf{x}) = \int_A S(u; \mathbf{x}) \exp(\theta^\top S(u; \mathbf{x})) du. \tag{9}$$

Numerical solution of (9) usually requires iterative algorithms.

Equation (9) is an unbiased estimating equation, i.e. the expectations of the left and right sides of (9) under θ are equal. The proof is an application of a non-stationary form of the Nguyen–Zessin formula (Nguyen & Zessin, 1976), namely

$$E \left(\sum_{x_i \in X \cap A} h(x_i, X) \right) = \int_A E(\lambda(u; X)h(u, X)) du, \tag{10}$$

holding for all non-negative bounded measurable functions $h(u, \mathbf{x})$. This extends a result of Diggle *et al.* (1994) that, under reasonable conditions, the normal equations in the stationary case are a special case of the Takacs–Fiksel estimating equations, themselves an application of the method of moments (Fiksel, 1984, 1988; Takacs, 1986).

The MPLE is known to be consistent and asymptotically normal (Jensen & Møller, 1991; Jensen & Künsch, 1994), at least for stationary pairwise interaction processes whose interaction functions satisfy suitable regularity conditions.

If (8) holds, the pseudolikelihood is log-convex. If, moreover, the parameter space is a convex set $\Theta \subset \mathbb{R}^p$, it follows that the maximum exists and occurs either at an interior point of Θ where the normal equations are satisfied, or on the convex boundary $\partial\Theta$ of Θ .

4. Berman–Turner device for maximum pseudolikelihood

This section describes the computational device that we propose for computing approximate maximum pseudolikelihood estimates. The method is an adaptation of an earlier technique of Berman & Turner (1992) for approximate maximum likelihood estimation for the inhomogeneous Poisson point process. Related ideas have been explored by Lindsey (1992, 1995, 1996) and Lindsey & Mersch (1992).

4.1. Derivation

Let X be a Gibbs point process with conditional intensity $\lambda_\theta(u; \mathbf{x})$ and consider the pseudolikelihood (6) for X , taking $A = W$ for simplicity. Approximate the integral in (6) by a finite sum using any quadrature rule,

$$\int_W \lambda_\theta(u; \mathbf{x}) du \approx \sum_{j=1}^m \lambda_\theta(u_j; \mathbf{x}) w_j, \quad (11)$$

where u_j , $j = 1, \dots, m$, are points in W and $w_j > 0$ are quadrature weights summing to $|W|$. This yields an approximation to the log-pseudolikelihood,

$$\log \text{PL}(\theta; \mathbf{x}) \approx \sum_{i=1}^{n(\mathbf{x})} \log \lambda_\theta(x_i; \mathbf{x}) - \sum_{j=1}^m \lambda_\theta(u_j; \mathbf{x}) w_j. \quad (12)$$

Extending an observation of Berman and Turner, we note that if the list of points $\{u_j, j = 1, \dots, m\}$ includes all the data points $\{x_i, i = 1, \dots, n\}$, we can rewrite (12) as

$$\log \text{PL}(\theta; \mathbf{x}) \approx \sum_{j=1}^m (y_j \log \lambda_j - \lambda_j) w_j, \quad (13)$$

where $\lambda_j = \lambda_\theta(u_j)$ and $y_j = z_j/w_j$, and

$$z_j = \begin{cases} 1 & \text{if } u_j \text{ is a data point, } u_j \in \mathbf{x}, \\ 0 & \text{if } u_j \text{ is a dummy point, } u_j \notin \mathbf{x}. \end{cases} \quad (14)$$

The right side of (13), for fixed \mathbf{x} , is formally equivalent to the log-likelihood of independent Poisson variables Y_k with means λ_k taken with weights w_k .

The expression (13) can therefore be maximized using standard software for fitting generalized linear models (McCullagh & Nelder, 1989) provided that (a) the software handles weighted likelihoods (with weights not necessarily summing to 1); (b) the software accepts non-integer values of the responses y_j in Poisson loglinear regression and correctly maximizes the log-likelihood expression; (c) the conditional intensity function $\lambda_\theta(\cdot; \mathbf{x})$, for fixed \mathbf{x} , is related to any explanatory variables by

$$g(\lambda_\theta(u; \mathbf{x})) = \theta^\top S(u, \mathbf{x}), \tag{15}$$

where g is a link function implemented in the software, and $S(u, \mathbf{x})$ is a vector of spatial covariates (possibly depending on \mathbf{x}) defined at each point u in W .

Software packages satisfying these criteria include GLIM (Aitkin *et al.*, 1989) and S-PLUS (Becker, Chambers & Wilks, 1988; Chambers & Hastie, 1992; Venables & Ripley, 1994). The only choice of g in (15) we consider is the log link, giving rise to the ‘loglinear model’ (8).

The key reason for adopting this approach is that the use of standard statistical packages rather than *ad hoc* software confers great advantages in applications. Modern statistical packages have a convenient notation for statistical models (Aitkin *et al.*, 1989; Chambers & Hastie, 1992; Venables & Ripley, 1994) which makes it very easy to specify and fit a wide variety of models of the type (8). Algorithms in the package may allow one to fit very flexible model terms such as the smooth functions in a generalized additive model (Hastie & Tibshirani, 1990). Interactive software allows great freedom to re-analyse the data. The fitting algorithms are typically more reliable and stable than in home-grown software.

4.2. Procedure

In summary, the procedure is as follows:

1. generate a set of dummy points, and combine it with the data points x_i to form the set of quadrature points u_j ;
2. compute the quadrature weights w_j ;
3. form the indicators z_j as in (14) and calculate $y_j = z_j/w_j$;
4. compute the (possibly vector) values $v_j = S(u_j, \mathbf{x})$ of the sufficient statistic at each quadrature point;
5. invoke the model-fitting software, specifying that the model is a loglinear Poisson regression

$$\log \lambda_j = \theta^\top v_j$$

to be fitted to the responses y_j and covariate values v_j , with weights w_j .

The coefficient estimates returned by the software give the (approximate) MPLE $\hat{\theta}$ of θ . The estimates of standard errors are not applicable, since they assume independent and identically distributed (iid) Poisson observations. The software also typically returns the deviance D of the fitted model; this is related to the log-pseudolikelihood of the fitted model by

$$-\log \text{PL}(\hat{\theta}; \mathbf{x}) = \frac{1}{2}D + \sum_{i=1}^{n(\mathbf{x})} \log w_i + n(\mathbf{x}). \tag{16}$$

Note that the sum is over data points only. Conveniently, the null model $\lambda_j \equiv \lambda$ in the loglinear Poisson regression corresponds to the uniform Poisson point process with intensity λ . The MPLE is $\hat{\lambda} = n(\mathbf{x}) / \sum_j w_j = n(\mathbf{x}) / |W|$ with corresponding log-pseudolikelihood $\log \text{PL}(\hat{\lambda}; \mathbf{x}) = n(\mathbf{x})(\log n(\mathbf{x}) - \log |W| - 1)$.

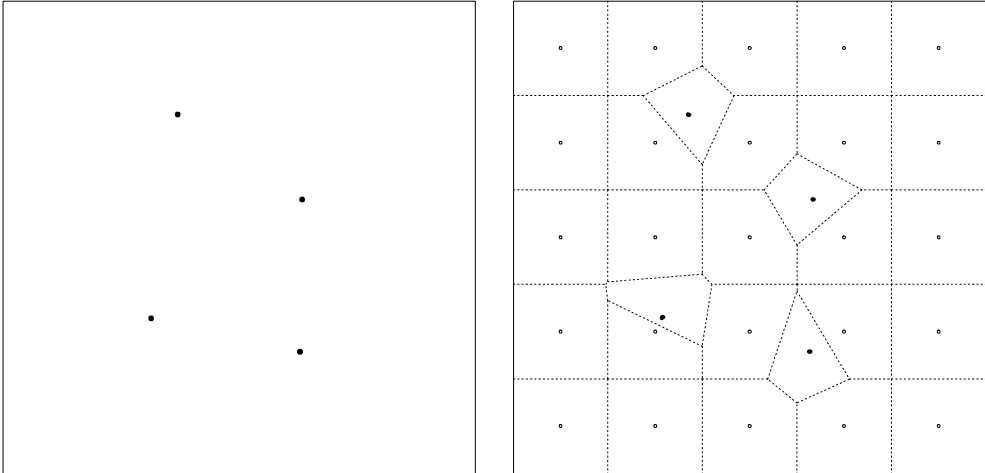


Figure 1. Quadrature using the Dirichlet tessellation (Berman & Turner, 1992). *Left:* Illustrative example of a point pattern dataset in the unit square W . *Right:* The Dirichlet tessellation of W based on the data points and a 5×5 grid of dummy points. Data points are marked by filled dots. The quadrature weight w_j is the area of the Dirichlet tile.

This formulation assumes $\lambda(u; \mathbf{x})$ is positive everywhere. Zero values are also permissible, provided the set of zeroes does not depend on θ . Thus we formally allow negative infinite values for $S(u; \mathbf{x})$. In the approximation (13) all points u_j with $\lambda(u_j; \mathbf{x}) = 0$ are dummy points. Their contribution is zero and so they should be omitted in all contexts.

4.3. Quadrature schemes and their accuracy

Berman & Turner (1992) used the Dirichlet tessellation or Voronoi diagram (Okabe, Boots & Sugihara, 1992) to generate quadrature weights for the analogue of (11). The data points are augmented by a list of dummy points, then the Dirichlet tessellation of the combined set of points is computed as sketched in Figure 1. The quadrature weight w_j associated with a (data or dummy) point u_j is the area of the corresponding Dirichlet tile.

A computationally cheaper scheme is to partition W into tiles T_k of equal area, and in each tile place exactly one dummy point, either systematically or randomly. Ascribe to each dummy or data point u_j a weight $w_j = a/n_j$ where a is the area of each tile, and n_j is the number of (dummy or data) points in the same tile as u_j . We call these the counting weights.

Note that for non-Poisson processes the conditional intensity $\lambda_\theta(u; \mathbf{x})$ is typically a discontinuous function of u at the data points x_i , while generically the limit as $u \rightarrow x_i$ exists. Thus the approximation (11) involves a ‘discontinuity error’ of size

$$\sum_{i=1}^{n(\mathbf{x})} (\lambda_\theta(x_i; \mathbf{x}) - \lim_{u \rightarrow x_i} \lambda_\theta(u; \mathbf{x}))w_i \tag{17}$$

(a sum of contributions from data points only) in addition to the ‘quadrature error’ associated with the finite approximation to the integral. The discontinuity error is controlled by reducing $\sum_i w_i$, the total quadrature weight of the contributions from the data points, usually by increasing the number $m - n$ of dummy points. See further comments at the end of Section 5.

5. Example: Strauss process

Next we illustrate the method as it applies to the simple Strauss process model (Strauss, 1975; Kelly & Ripley, 1976). This is a pairwise interaction process (2) in which $b_\theta(u) \equiv \beta$ is constant and $h_\theta(u, v) = \gamma$ if $\|u - v\| \leq r$, and $h_\theta(u, v) = 1$ otherwise. Here $\beta > 0$ and $0 \leq \gamma \leq 1$ are parameters and $r > 0$ is a fixed ‘interaction distance’. Each pair of points closer than r units apart contributes a penalty of γ to the likelihood,

$$\text{lik}(\beta, \gamma; \mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \gamma^{s(\mathbf{x})} \tag{18}$$

(taking $0^0 = 1$), where $\alpha = \alpha(\beta, \gamma)$ is the normalizing constant, and

$$s(\mathbf{x}) = \#\{(i, j) : i < j, \|x_i - x_j\| \leq r\}$$

is the number of unordered pairs of points which lie closer than r units apart. The Strauss process is well-defined for all $\gamma \in [0, 1]$. If $\gamma = 1$, it reduces to the homogeneous Poisson process with intensity β . For $\gamma = 0$ it is a ‘hard core’ process in which no two points ever lie closer than r units apart. For $0 < \gamma < 1$ there is inhibition between close pairs of points. The conditional intensity is

$$\lambda_{\beta, \gamma}(u; \mathbf{x}) = \beta \gamma^{t(u, \mathbf{x})},$$

where

$$t(u, \mathbf{x}) = \#\{x_i \in \mathbf{x} : 0 < \|x_i - u\| \leq r\} \tag{19}$$

is the number of points $x_i \in \mathbf{x}$ which are close to u , other than u itself.

The pseudolikelihood is

$$\text{PL}(\beta, \gamma; \mathbf{x}) = \beta^{n(\mathbf{x})} \gamma^{2s(\mathbf{x})} \exp\left(-\beta \int_W \gamma^{t(u, \mathbf{x})} du\right), \tag{20}$$

which is in the required loglinear form (8) with $\theta = (\ln \beta, \ln \gamma)^\top$ and $S(u; \mathbf{x}) = (1, t(u, \mathbf{x}))^\top$. The MPLS normal equations (9) are

$$n(\mathbf{x}) = \beta \int_W \gamma^{t(u, \mathbf{x})} du \tag{21}$$

$$2s(\mathbf{x}) \int_W \gamma^{t(u, \mathbf{x})} du = n(\mathbf{x}) \int_W t(u, \mathbf{x}) \gamma^{t(u, \mathbf{x})} du. \tag{22}$$

The maximum of the pseudolikelihood may occur either at a solution of these equations or at $\gamma = 0, 1$. If r is less than the minimum interpoint distance, then $s(\mathbf{x}) = 0$ and the pseudolikelihood is maximized when $\gamma = 0$. Otherwise $\hat{\gamma} > 0$. If the solution of (21)–(22) occurs where $\gamma > 1$, then, since the log-pseudolikelihood is concave, $\hat{\gamma} = 1$. The maximized log-pseudolikelihood is

$$\log \text{PL}(\hat{\beta}, \hat{\gamma}; \mathbf{x}) = n(\mathbf{x}) \log \hat{\beta} + 2s(\mathbf{x}) \log \hat{\gamma} - \hat{\beta} p(\hat{\gamma}). \tag{23}$$

Consistency and asymptotic normality of the MPLS follow from Jensen & Künsch (1994) and Jensen & Møller (1991).

To compute the approximate MPLE using the Berman–Turner device we would follow the procedure in Section 4.2, fitting the loglinear model

$$\log \lambda_j = \theta_1 + \theta_2 v_j,$$

where $v_j = t(u_j, \mathbf{x})$, with $t(u, \mathbf{x})$ as defined in (19). A suitable S-PLUS invocation would be

```
glm(y ~ v, family=poisson, link=log, weights=w)
```

where y, v, w are S-PLUS vectors of equal length containing the responses y_j , the ‘explanatory variable’ values v_j , and the weights w_j , respectively, for each quadrature point u_j . If `glm()` yields a solution $\theta_2 > 0$, i.e. $\gamma > 1$, then the MPLE is $\hat{\gamma} = 1$, $\hat{\beta} = n(\mathbf{x})/|W|$.

Note that the integrals in (20)–(22) are polynomials in γ :

$$p(\gamma) = \int_W \gamma^{t(u,\mathbf{x})} du = a_0 + a_1\gamma + \dots + a_K\gamma^K, \tag{24}$$

say, where $a_k = |A_k|$ is the area of the region $A_k = \{u \in W : t(u, \mathbf{x}) = k\}$. Thus (21)–(22) can be rewritten

$$\beta p(\gamma) = n(\mathbf{x}) \tag{25}$$

$$\frac{\gamma p'(\gamma)}{p(\gamma)} = \frac{2s(\mathbf{x})}{n(\mathbf{x})}. \tag{26}$$

In this simple case, the MPLE can be computed by solving (25)–(26) directly, although this still requires evaluation of the coefficients a_j , which calls for numerical integration or computational geometry. We use this ‘polynomial’ approach to check the accuracy of our method in Section 10.

The quadrature approximation (11) consists of replacing $p(\gamma)$ by

$$q(\gamma) = \sum_{j=1}^m \gamma^{t(u_j,\mathbf{x})} w_j = \sum_{k=1}^K b_k \gamma^k,$$

where $b_k = \sum_{u_j \in A_k} w_j$ are approximations to the areas of the sets A_k . The approximation includes discontinuity error (17) arising because the weight w_i for a data point x_i , with $t(x_i; \mathbf{x}) = k$ but $\lim_{u \rightarrow x_i} t(u; \mathbf{x}) = k + 1$, is ascribed to b_k rather than to b_{k+1} .

The total error in approximating $p(\gamma)$ by $q(\gamma)$ is bounded by $E_1 = \sum_{k=1}^K |a_k - b_k|$, the sum of the errors in approximating the area a_k by b_k . The error in approximating $\gamma p'(\gamma)$ by $\gamma q'(\gamma)$ is bounded by $E_2 = \sum_{k=1}^K k |a_k - b_k|$. To control both E_1 and E_2 , dummy points must be sufficiently dense throughout W and sufficiently dense where $t(u; \mathbf{x})$ is high; that is, near the data points.

6. Spatial interaction terms

Sections 6–8 present further examples of point processes, and examine the computational requirements for applying our method. The present section concerns point processes with various kinds of interpoint interaction (pairwise interaction and other). Inhomogeneous models are discussed in Section 7 and marked point processes in Section 8.

6.1. Pairwise interaction models

6.1.1. General loglinear form

Consider first the general pairwise interaction process (2) and assume

$$b_\theta(u) = \exp(\theta^T B(u)), \tag{27}$$

$$h_\theta(u, v) = \exp(\theta^T H(u, v)), \tag{28}$$

where $B(u)$ and $H(u, v)$ are vectors defined for every $u, v \in W$. Note $H(u, v)$ should be a symmetric function of u and v . The conditional intensity (5) becomes

$$\lambda_\theta(u; \mathbf{x}) = \exp\left(\theta^T B(u) + \theta^T \sum_{i=1}^{n(\mathbf{x})} H(u, x_i)\right). \tag{29}$$

This is of the loglinear form (8) required for our approximation, with

$$S(u, \mathbf{x}) = B(u) + \sum_{\substack{i=1 \\ x_i \neq u}}^{n(\mathbf{x})} H(u, x_i), \tag{30}$$

and the procedure of Section 4.2 may be applied. The log-pseudolikelihood is concave in θ so the MPLE values form a non-empty convex set. Consistency of the MPLE is not guaranteed in this generality.

In the rest of this section we assume $B(u)$ is constant; Section 7 discusses models for spatial inhomogeneity. Here it is important to note that the general form (27) assumed for b_θ embraces not only parametric models but also generalized additive models (GAM; Hastie & Tibshirani, 1990) in which $B(u)$ would be a vector of spline basis functions. However, this apparently does not extend to GAM type models for h_θ , since the sufficient statistic (30) is a sum of a variable number of terms which is beyond the scope of current GAM fitting algorithms. Hence we are currently forced to consider only parametric models for interpoint interaction, such as the Strauss process.

6.1.2. Soft core process

The ‘soft core’ model discussed by Ogata & Tanemura (1981) is a pairwise interaction process (2) with $b(u) \equiv \beta$ and

$$h(u, v) = \exp\left(-\left(\frac{\sigma}{\|u - v\|}\right)^{2/\kappa}\right) \quad (u \neq v),$$

where $\beta > 0$ and $0 \leq \sigma < \infty$ are parameters and $0 < \kappa < 1$ is an irregular parameter which we assume known for the moment. The limit as $\kappa \rightarrow 0$ is the hard core process (Strauss with $\gamma = 0, r = \sigma$); the density is not integrable for $\kappa \geq 1$. Thus $\log \lambda_\theta(u; \mathbf{x}) = \exp(\theta^T S(u, \mathbf{x}))$ where $S(u, \mathbf{x}) = (1, V(u, \mathbf{x}))^T$ and $\theta = (\log \beta, \sigma^{2/\kappa})^T$ with

$$V(u, \mathbf{x}) = - \sum_{\substack{i=1 \\ x_i \neq u}}^n \|u - x_i\|^{-2/\kappa}. \tag{31}$$

The log-pseudolikelihood is concave in θ and the MPLE is well defined, consistent and asymptotically normal (Jensen & Künsch, 1994).

The conditional intensity is loglinear in θ . To estimate β and σ (given a value of κ) one would execute the S-PLUS command

```
glm(y ~ v, family=poisson, weights=w)
```

where y, v, w are S-PLUS vectors containing the responses $y_j = z_j/w_j$, explanatory variable $v_j = V(u_j, \mathbf{x})$, and weights w_j respectively. Then $\hat{\beta} = \exp \hat{\theta}_1$ and $\hat{\sigma} = \hat{\theta}_2^{\kappa/2}$ where $\hat{\theta}_1$ and $\hat{\theta}_2$ are the estimates of the linear coefficients returned by the `glm()` function of S-PLUS.

6.1.3. Step function interaction

In the absence of non-parametric estimators of h_θ , there is much interest (Fiksel, 1984, 1988; Penttinen, 1984; Takacs, 1986) in fitting a stationary pairwise interaction process with a piecewise constant interaction function h_θ . Thus $b_\theta(u) \equiv \beta$ and $h_\theta(u, v)$ is a step function of $\|u-v\|$, say $\log h_\theta(u, v) = \theta_\ell$ if $r_{\ell-1} < \|u-v\| \leq r_\ell$, and $\log h_\theta(u, v) = 0$ if $\|u-v\| > r_k$, where $0 = r_0 < r_1 < r_2 < \dots < r_k$ are parameters. This is a special case of (27)–(28) with $\theta = (\log \beta, \theta_1, \theta_2, \dots, \theta_k)^\top$, say, and

$$B(u) = (1, 0, 0, \dots, 0), \quad H(u, v) = (0, I_1(\|u-v\|), \dots, I_k(\|u-v\|)),$$

where $I_\ell(d) = \mathbf{1}\{r_{\ell-1} < d \leq r_\ell\}$ for $\ell = 1, 2, \dots, k$. Thus

$$S(u, \mathbf{x}) = (1, t_1(u, \mathbf{x}), \dots, t_k(u, \mathbf{x}))^\top,$$

where $t_\ell(u, \mathbf{x}) = \#\{x_i \in \mathbf{x} : r_{\ell-1} < \|x_i - u\| \leq r_\ell\}$ ($\ell = 1, 2, \dots, k$),

is the number of points $x_i \in \mathbf{x}$ whose distance from u lies in the interval $(r_{\ell-1}, r_\ell]$. The MPLE is consistent by Jensen & Møller (1991 Theorem 3.1) if either $\theta_\ell \leq 0$ for all $\ell = 1, \dots, k$, or the θ_ℓ are uniformly bounded from above and $\theta_1 = -\infty$ (the process has a hard core).

In our approach it is easy to fit this model, analogously to the Strauss process. The associated loglinear model is

$$\log \lambda_j = \log \beta + \theta_1 v_{1j} + \dots + \theta_k v_{kj}, \quad \text{where } v_{\ell j} = t_\ell(u_j, \mathbf{x}).$$

6.2. Area-interaction process

The Widom–Rowlinson ‘penetrable sphere model’ of liquid–vapour equilibrium (Widom & Rowlinson, 1970; Hammersley, Lewis & Rowlinson, 1975; Rowlinson, 1980) also known as the ‘area-interaction process’ (Baddeley & van Lieshout, 1995), has probability density (in the simplest case)

$$p(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \gamma^{-A(\mathbf{x})}, \tag{32}$$

where $A(\mathbf{x})$ is the area of the union of discs of radius r centred at x_i . Here $\beta, \gamma, r > 0$ are parameters and $\alpha = \alpha(\beta, \gamma, r)$ is the normalizing constant. Generalizations are given in Baddeley & van Lieshout (1995). The process is well defined, i.e. (32) is integrable, for all values of $\gamma > 0$ and for all compact $W \subset \mathbb{R}^2$. It reduces to a Poisson process when

$\gamma = 1$, exhibits ordered patterns for $0 < \gamma < 1$ and produces clustering when $\gamma > 1$. Other properties and maximum likelihood estimation are being investigated.

The conditional intensity is of the desired form $\log \lambda_\theta(u; \mathbf{x}) = \theta^\top S(u; \mathbf{x})$ putting $\theta = (\log \beta, \log \gamma)^\top$ and $S(u; \mathbf{x}) = (1, A(\mathbf{x} \cup \{u\}) - A(\mathbf{x}))^\top$. Results in Jensen & Møller (1991) imply that if r is known, the MPLE of (β, γ) is consistent. We do not know whether a central limit theorem is available; the results of Jensen & Künsch (1994) do not apply.

Another advantage of the Berman–Turner approach here is the reduction in the computational cost because the values of $D(u; \mathbf{x}) = A(\mathbf{x} \cup \{u\}) - A(\mathbf{x})$ are only required for a relatively small number of points, i.e. the quadrature points u_j .

6.3. Ord’s process

Ord (Ripley, 1977 Discussion) suggested a model for regular patterns of points representing entities which compete for resources, such as trees or towns. The Dirichlet tile associated with a point can be interpreted as the territory from which it draws resources. Ord suggested densities of the form

$$f(\mathbf{x}; \theta) \propto \prod_{i=1}^n g_\theta(A(x_i, \mathbf{x})), \tag{33}$$

where $A(x_i, \mathbf{x})$ denotes the area of the Dirichlet tile associated with x_i in the pattern \mathbf{x} , and $g_\theta : \mathbb{R} \rightarrow [0, \infty)$ is a function combining the roles of the spatial interaction and intensity terms in other models. The special case $g_\theta(v) \equiv \lambda$ is the uniform Poisson process with intensity λ . Typically $g_\theta(\cdot)$ would be an increasing function, so that small tiles are penalized.

Ripley (1981 p. 175) concludes his analysis of the Swedish pines data (Section 10.1) with a comment that fitting Ord’s process would be an interesting alternative analysis. To our knowledge, this has not been attempted and Ord’s model has not been investigated or mentioned further, except in Baddeley & Møller (1989).

The process (33) exists (i.e. f is integrable) under reasonable conditions: for example, whenever $g_\theta(\cdot)$ is uniformly bounded. The conditional intensity is

$$\lambda_\theta(u, \mathbf{x}) = g_\theta(A(u, \mathbf{x} \cup \{u\})) \prod_{x_i \sim u} \frac{g_\theta(A(x_i, \mathbf{x} \cup \{u\}))}{g_\theta(A(x_i, \mathbf{x}))},$$

where the product is over all points x_i that are Dirichlet neighbours of u in the pattern $\mathbf{x} \cup \{u\}$, and $A(u, \mathbf{x} \cup \{u\})$ is the area of the Dirichlet tile with centre u in this pattern. Explicit analytic expressions for $A(u, \mathbf{x} \cup \{u\})$, the pseudolikelihood, or the MPLE are not available. Geometric computation of $A(u, \mathbf{x})$ is time-consuming, so a discrete approximation to the pseudolikelihood becomes a necessity.

The Berman–Turner device (Section 4) can be applied if the kernel is modelled in log-linear form $g_\theta(v) = \exp(\theta^\top G(v))$. Then $\log \lambda(u, \mathbf{x}) = \theta^\top V(u, \mathbf{x})$, where

$$V(u, \mathbf{x}) = G(A(u, \mathbf{x} \cup \{u\})) + \sum_{x_i \sim u} [G(A(x_i, \mathbf{x} \cup \{u\})) - G(A(x_i, \mathbf{x}))]$$

is the regression variable. Evaluating $v_j = V(u_j, \mathbf{x})$ for all j requires computation of $m + 1$ different Dirichlet tessellations.

7. Inhomogeneous models

Few writers to date, apart from Ogata & Tanemura (1986), have fitted explicit models to point pattern data that incorporate both spatial inhomogeneity and interpoint interactions.

In the context of our method, it is easy to introduce a spatial trend or dependence on spatial covariates. This is simply a matter of adding more terms to the linear predictor $S(u; \mathbf{x})$ in the associated Poisson loglinear regression model.

7.1. Spatial trend

A straightforward model of spatial trend in a pairwise interaction process (2) is the log-linear form $b_\theta(u) = \exp(\theta^\top B(u))$ and $h_\theta(u, v) = h_\theta(u - v) = \exp(\theta^\top H(u - v))$ as in (27)–(28), but with the assumption that $H(u, v) = H(u - v)$ depends only on $u - v$.

The spatial trend is expressed by the dependence of $b_\theta(u)$ on location u , while the inter-point interaction does not exhibit trend. Typically H would be one of the pairwise interaction functions considered in Sections 5–6, while $B(u) = (B_1(u), \dots, B_k(u))^\top$ would be a vector of convenient scalar functions of location, such as polynomials or orthonormal functions of the coordinates. It is also possible to use the GAM approach (Hastie & Tibshirani, 1990) to model each $B_\ell(u)$ by a smooth function of one coordinate.

By (29) we may fit these models using the method of Section 4.2 by adding the term $\theta^\top B(u)$ to the linear predictor in one of the models discussed in previous subsections.

Ogata & Tanemura (1981) developed maximum likelihood estimation techniques for models of this form, in particular combining a spatial trend with the soft core interaction of Section 6.1.2. The trend term $\theta^\top B(u)$ was a polynomial in the Cartesian coordinates. Details are given in Section 10.3.

More generally, the spatial interaction can also depend on location. Loglinearity is usually lost, however, and we cannot apply the method of Section 4 directly.

An effective alternative way to fit models with spatially-varying interaction range is proposed by Nielsen & Jensen (1998).

7.2. Spatial covariates

The data may include spatial covariates such as topographic elevation, soil pH, or another observed spatial pattern. Covariates may serve to eliminate spurious trend, explain variation in intensity, or make inferences conditional upon another spatial pattern. For our purposes the spatial covariate must be incorporated as a function $Z(u)$, $u \in W$, observed at each of the quadrature points u_j . We add terms in $Z(u_j)$ to the linear predictor. The covariate value $Z(u)$ may be simply an observation such as pH or elevation, but often the covariate data can be transformed to yield $Z(u)$. For example in spatial epidemiology $Z(u)$ could be a kernel smoothed estimate of the density of the population at risk (Cuzick & Edwards, 1990).

Another observed spatial pattern can be included as a spatial covariate by computing a suitable function $Z(u)$ associated with the pattern. Berman (1986) proposed modelling the dependence of a point process X on a line segment process Y by conditioning on Y and testing whether X is inhomogeneous Poisson with an intensity $\lambda(u)$ depending on the minimum distance $Z(u)$ from location u to the nearest line segment.

8. Marked point patterns

8.1. General

The observed points may also carry ‘marks’, i.e. observations m_i associated with each point x_i of the pattern. The full dataset is a list

$$\mathbf{v} = \{(x_1, m_1), \dots, (x_n, m_n)\} \quad (x_i \in W \text{ and } m_i \in \mathcal{M}),$$

where \mathcal{M} is the space of possible marks. The marks may be observations of any kind; commonly \mathcal{M} is either a discrete set of ‘labels’ $\mathcal{M} = \{1, 2, \dots, c\}$ or the positive real line $\mathcal{M} = [0, \infty)$. In the discrete case, the data points are effectively classified into m different types or colours, and the mark attached to each point indicates its type. In the continuous case m_i is usually a physical measurement such as the height or diameter of a tree whose location is x_i . See Diggle (1983 Chapters 6, 7), Cressie (1991 Sections 8.6–8.7), Baddeley & Møller (1989), Frey & Schmidt (1998), Stoyan *et al.* (1995). Jensen & Møller (1991) formally treat the pseudolikelihood of a marked point process and prove consistency of the MPLE; Goulard, Särkkä & Grabarnik (1996) investigate further statistical properties. See also Särkkä & Högmander (1998).

The reference process for likelihoods is the Poisson marked point process constructed by attaching iid random marks to the points of a Poisson point process on W with unit intensity (Kingman, 1993). The distribution of the marks in this reference process is an arbitrary probability distribution Q on \mathcal{M} .

The inhomogeneous Poisson marked point process with intensity function $b_\theta : W \times \mathcal{M} \rightarrow \mathbb{R}_+$, is the analogue of (1), with density

$$f(\mathbf{v}; \theta) = \alpha \prod_{i=1}^{n(\mathbf{x})} b_\theta(x_i, m_i),$$

where $\alpha = \alpha(\theta) = \exp(-\int_{W \times \mathcal{M}} (b_\theta(u, m) - 1) dQ(m) du)$ is the normalizing constant.

The pairwise interaction marked point process is the analogue of (2), with density

$$f(\mathbf{v}; \theta) = \alpha \left(\prod_{i=1}^{n(\mathbf{x})} b_\theta(x_i, m_i) \right) \left(\prod_{i < j} h_\theta((x_i, m_i), (x_j, m_j)) \right), \tag{34}$$

where $\alpha = \alpha(\theta)$ is the normalizing constant, $b_\theta : W \times \mathcal{M} \rightarrow \mathbb{R}_+$ is the activity/trend function and $h_\theta : (W \times \mathcal{M})^2 \rightarrow \mathbb{R}_+$ the interaction function. The function h_θ is symmetric in the sense that $h_\theta((u, m), (u', m')) = h_\theta((u', m'), (u, m))$ for $u, u' \in W$ and $m, m' \in \mathcal{M}$. Conditions must be imposed on b_θ, h_θ to ensure the density is integrable.

8.2. Pseudolikelihood

The conditional intensity of a Gibbs marked point process, analogous to (3)–(4), is a function $\lambda((u, m); \mathbf{v})$ of the marked pattern \mathbf{v} and of a marked point (u, m) with $u \in W$ and $m \in \mathcal{M}$. For example the pairwise interaction marked point process (34) has

$$\lambda_\theta((u, m), \mathbf{v}) = b_\theta(u, m) \prod_{\substack{i=1 \\ (u, m) \neq (x_i, m_i)}}^{n(\mathbf{x})} h_\theta((u, m), (x_i, m_i)).$$

The pseudolikelihood of a Gibbs marked point process is (Jensen & Møller, 1991; Goulard *et al.*, 1996)

$$PL(\theta; \mathbf{v}) = \left(\prod_{i=1}^{n(\mathbf{x})} \lambda_\theta((x_i, m_i); \mathbf{v}) \right) \exp \left(- \int_W \int_{\mathcal{M}} \lambda_\theta((u, m); \mathbf{v}) dQ(m) du \right) \tag{35}$$

In the case of a multitype point process with c different types, we have $\mathcal{M} = \{1, 2, \dots, c\}$ and the pseudolikelihood is usually defined by

$$PL(\theta; \mathbf{v}) = \left(\prod_{i=1}^{n(\mathbf{x})} \lambda_{\theta}((x_i, m_i); \mathbf{v}) \right) \exp \left(- \sum_{m=1}^c \int_W \lambda_{\theta}((u, m); \mathbf{v}) du \right). \tag{36}$$

8.3. Berman–Turner device

To apply our approximation method to (35) we create a set of marked points (u_j, k_j) , $j = 1, \dots, M$, which include the data (x_i, m_i) , $i = 1, \dots, n$, and form a good quadrature rule for $W \times \mathcal{M}$. It is usually convenient to take the Cartesian product of a set of quadrature points in W and a set of elements of \mathcal{M} . We assume this and write the marked points as (u_j, k_{ℓ}) for $j = 1, \dots, J$ and $\ell = 1, \dots, L$ where $u_j \in W, k_{\ell} \in \mathcal{M}$. Then we define the indicator $z_{j\ell}$ to equal 1 if (u_j, k_{ℓ}) is a data point and 0 if it is a dummy point. Let $w_{j\ell}$ be the corresponding weights for a linear quadrature rule in $W \times \mathcal{M}$. Then the pseudolikelihood is approximated by

$$\log PL(\theta; \mathbf{v}) \approx \sum_{\ell=1}^L \sum_{j=1}^J (y_{j\ell} \log \lambda_{j\ell} - \lambda_{j\ell}) w_{j\ell},$$

where $\lambda_{j\ell} = \lambda_{\theta}((u_j, k_{\ell}); \mathbf{v})$ and $y_{j\ell} = z_{j\ell}/w_{j\ell}$. For discrete marks as in (36), the weights may simply be those for a quadrature rule in W corresponding to the points u_j .

8.4. Example: 2-type Strauss process

This is the special case of the pairwise interaction marked point process (34) in which $\mathcal{M} = \{1, 2\}$, i.e. points belong to one of two types, and $b_{\theta}(u, m) = \beta_m$,

$$h_{\theta}((u, m), (u', m')) = \begin{cases} \gamma_{m,m'} & \text{if } 0 < \|u - u'\| < r_{m,m'}, \\ 1 & \text{otherwise,} \end{cases}$$

where $\beta_1, \beta_2 > 0$ are intensity parameters, $\gamma_{11}, \gamma_{22}, \gamma_{12}, \gamma_{21} \in [0, 1]$ are interaction parameters, and $r_{11}, r_{22}, r_{12}, r_{21} > 0$ are interaction distances, with $\gamma_{12} = \gamma_{21}$ and $r_{12} = r_{21}$. The density may be expressed analogously to (18) as

$$f(\mathbf{v}; \theta) = \alpha \beta_1^{n_1(\mathbf{v})} \beta_2^{n_2(\mathbf{v})} \gamma_{11}^{s_{11}(\mathbf{v})} \gamma_{12}^{s_{12}(\mathbf{v})} \gamma_{22}^{s_{22}(\mathbf{v})},$$

where $n_1(\mathbf{v}), n_2(\mathbf{v})$ are the numbers of points of type 1 and 2 respectively, and $s_{m,m'}(\mathbf{v})$ is the number of pairs of distinct marked points of types m and m' respectively within a distance $r_{m,m'}$ of each other. The conditional intensity is

$$\lambda_{\theta}((u, m), \mathbf{v}) = \beta_m \gamma_{m1}^{t_1((u,m),\mathbf{v})} \gamma_{m2}^{t_2((u,m),\mathbf{v})} \quad (u \in W, m \in \mathcal{M},)$$

where $t_{m'}((u, m), \mathbf{v}) = \#\{i : 0 < \|u - x_i\| \leq r_{m,m'}, m_i = m'\}$

is the number of type m' points within the required distance $r_{m,m'}$ of a point u with type m .

This model may be cast in the loglinear form (8) with parameter vector $\theta = (\log \beta_1, \log \beta_2, \log \gamma_{11}, \log \gamma_{12}, \log \gamma_{22})$ and five ‘explanatory variables’, namely $I_1(m)$, $I_2(m)$, $I_1(m)t_1((u, m), \mathbf{v})$, $I_1(m)t_2((u, m), \mathbf{v}) + I_2(m)t_1((u, m), \mathbf{v})$ and $I_2(m)t_2((u, m), \mathbf{v})$, respectively, where $I_k(m) = \mathbf{1}\{m = k\}$. Equivalently it may be described as a nested model with one factor and two covariates, one of which is nested within the factor.

The pseudolikelihood estimate of θ is consistent (Jensen & Møller, 1991). The central limit theorem of Jensen & Künsch (1994) does not strictly apply here because there are more than two parameters, but Jensen and Künsch conjecture (1994 p.477) that a generalization does hold.

9. Estimation and inference issues

9.1. Edge effects

For inferential purposes it matters whether we assume the data \mathbf{x} are a realization of a finite point process X defined only inside W (‘bounded case’) or a partially observed realization $\mathbf{y} \cap W$ of a point process Y extending throughout \mathbb{R}^d only through the ‘window’ W (‘unbounded case’). In the unbounded case we have an ‘edge effect’ problem: the conditional intensity $\lambda_\theta(u; \mathbf{y})$ of Y may not be observable from the data $\mathbf{x} = \mathbf{y} \cap W$, since the required information may involve points outside the observation window W . Ripley (1988), Stoyan *et al.* (1995) and Baddeley (1998) survey remedies for edge effects. Following are some possible strategies.

9.1.1. Periodic boundary conditions

If the window W is rectangular one may apply ‘periodic boundary conditions’ (Ripley, 1977) by identifying opposite sides of W so that points near the right edge (say) have neighbours near the left edge. This is also called the ‘torus correction’. It typically reduces bias but inflates variance, and is only applicable to certain shapes of W . It seems inadvisable for inhomogeneous patterns.

9.1.2. Border method

The border method applies (Ripley, 1988) to any process with finite interaction range r , in the sense that $\lambda(u; \mathbf{x})$ depends only on data points x_i lying within a distance r of u . An example is the Strauss process with fixed r . Form the pseudolikelihood over the subregion $W_{\ominus r} = \{u \in W : b(u, r) \subset W\}$ of all points of W lying at least r units from the boundary. For $u \in W_{\ominus r}$ the conditional intensity is observable, $\lambda_\theta(u; \mathbf{y}) = \lambda_\theta(u; \mathbf{y} \cap W)$, so the pseudolikelihood over $W_{\ominus r}$ can be calculated from the data. This applies both to stationary and non-stationary processes.

The main drawback is that the method discards appreciable amounts of data. Also, if r is unknown, one must be wary of comparing pseudolikelihoods based on different subsets $W_{\ominus r}$. One strategy is to compute all pseudolikelihoods over the same domain $W_{\ominus R}$ where R is the maximum r value contemplated.

9.1.3. Ripley’s hybrid method

Edge correction weights (Ripley, 1988; Stoyan *et al.*, 1995; Baddeley, 1998) are widely used for estimation in stationary models. Ripley (1988 p.67) and Venables & Ripley (1994 p.396) extended this to maximum pseudolikelihood estimation for the Strauss process. Ripley

proposed that the right side of (26), which cannot be observed due to edge effects, be estimated by $n(\mathbf{x})\hat{K}(r)/|W|$, a quantity which has approximately the same expectation. Here $\hat{K}(r)$ is the estimate of $K(r)$, the second moment function K of the process (Ripley, 1988),

$$\hat{K}(r) = \frac{|W|}{n(\mathbf{x})^2} \sum_{i \neq i'} \mathbf{1}\{\|x_i - x_{i'}\| \leq r\} e(x_i, x_{i'}, W),$$

where $e(u, v, W)$ is an edge effect correction factor ensuring unbiased estimation of $\lambda^2 K(r)$ if the point process is stationary and isotropic. The left side of (26) is also subject to edge effects, and is modified by using the eroded domain $W_{\ominus r}$ instead of W in (24). This is a ‘hybrid’ of the border method and the edge correction weights strategies.

9.1.4. *Edge corrected pseudolikelihood*

An alternative, that may be new, is to introduce edge correction weights into the pseudolikelihood itself. Consider a stationary pairwise interaction process with $b_\theta(u) \equiv \beta$ and $h_\theta(u, v) = \exp(\theta^\top H(u - v))$. Modify the model, replacing the pairwise interaction function by

$$h_\theta^E(u, v) = \exp(e(u, v, W)\theta^\top H(u - v)),$$

where $e(u, v, W)$ is an edge correction weight as in Ripley (1988), Stoyan *et al.* (1995) and Baddeley (1998) which must be symmetric in u and v . This modified model has conditional intensity $\lambda_\theta^E(u; \mathbf{x}) = \beta \exp(\theta^\top S^E(u; \mathbf{x}))$, where

$$S^E(u; \mathbf{x}) = \sum_{i=1}^{n(\mathbf{x})} e(u, x_i, W)H(u - x_i)$$

is a ‘plug-in’ estimator of the unobservable potential $S(u; \mathbf{y})$ for the original model. Forming the pseudolikelihood for the modified model and deriving the normal equations, we obtain (9) with $S(u; \mathbf{x})$ replaced by $S^E(u; \mathbf{x})$ throughout. By (10), this is an unbiased estimating equation for the *modified* model. It is approximately unbiased for the original model, when $\|\theta\|$ is small. This model can be fitted by our method.

9.1.5. *Data augmentation*

The unobserved points of \mathbf{y} outside the window which affect the value of $\lambda_\theta(u; \mathbf{y})$ for $u \in W$ can alternatively be regarded as missing data. One approach is data augmentation Tanner (1996 Chapter 5) which has been applied to maximum likelihood inference for point processes by Geyer (1998). This can also be applied in our context.

9.2. **Irregular parameters and profile pseudolikelihood**

The point process models considered above contain ‘irregular’ parameters which do not enter in the loglinear form (8) required for our method. A possible approach to estimation is by analogy with profile likelihood. Write $\theta = (\varphi, \psi)$ where ψ are the irregular parameters, so that we assume

$$\lambda_\theta(u; \mathbf{x}) = \exp(\varphi^\top S(u, \mathbf{x}, \psi))$$

instead of (8). For each fixed value of ψ the model is loglinear in φ , so we can apply our approximation method to maximize the pseudolikelihood over φ , yielding an MPLE $\hat{\varphi}(\psi)$

for fixed ψ . Computing the maximized pseudolikelihood as a function of ψ yields the profile pseudolikelihood

$$PL((\hat{\varphi}(\psi), \psi); \mathbf{x}) = \max_{\varphi} PL(\varphi, \psi).$$

The global MPLE of θ is then obtained by maximizing this profile pseudolikelihood over ψ . We examine this approach for the Strauss and soft core processes in Section 10.

9.3. Parametric inference and model choice

We use the parametric bootstrap for inference and model choice. We assume in practice that the MPLE $\hat{\theta}$ is approximately unbiased and approximately normal, although this has only been established in certain cases (Jensen & Møller, 1991; Jensen & Künsch, 1994).

To obtain confidence intervals for θ , we simulate from the *fitted* model $\theta = \hat{\theta}$, obtaining M simulated values $\hat{\theta}_{(1)}^*, \dots, \hat{\theta}_{(M)}^*$ from the distribution of the MPLE under the fitted model. We estimate the mean vector and covariance matrix of this distribution from the simulated values, then construct confidence intervals using location models based on the multivariate normal or the bootstrap distribution. Similarly for model choice we use the bootstrap distribution of the deviance between two (nested) models.

10. Examples of applications

Our analyses were performed in S-PLUS (Becker *et al.*, 1988; Chambers & Hastie, 1992; Venables & Ripley, 1994) using the generalized linear model fitting function `glm()` and occasionally the generalized additive model function `gam()`. Some analyses were cross-checked using GLIM (Aitkin *et al.*, 1989).

10.1. Swedish pines data

Figure 2 depicts the Swedish pines data of Strand (1972) which give the locations of 71 pine saplings in a 10 m \times 10 m square. Ripley’s pioneering analysis (Ripley, 1981 Section 8.6, pp. 172–175) plotted $L(t) = \sqrt{K(t)/\pi}$ and rejected the hypothesis of a homogeneous Poisson process at the 1% level by a Monte Carlo test based on $D = \sup_t |L(t) - t|$. Ripley then fitted a Strauss process manually, obtaining $r = 0.7$ m and $\gamma = 0.20$. In the latest analysis, by Venables & Ripley (1994 p. 396), γ was estimated to be 0.15 using maximum pseudolikelihood (using a procedure essentially the same as our ‘polynomial approach’ (25)–(26)) with Ripley’s hybrid edge correction (Section 9.1.3).

We fitted a Strauss process to these data by maximum pseudolikelihood, using both the Berman–Turner device and the polynomial approach via (25)–(26). We estimated β and γ , but initially held r fixed at 0.7. For the Berman–Turner method, we tried varying densities of dummy points, with various edge corrections, and computed the quadrature weights using the Dirichlet and counting methods (Section 4.3). Estimates obtained for γ ranged from 0.29 to 0.20, and for β from 1.49 to 2.12. A finer quadrature scheme always led to a smaller value of γ and a larger value of β . Both the Berman–Turner and polynomial methods gave $\gamma = 0.21$ using a 50 \times 50 grid of dummy points. This is close to the value obtained by Ripley (1981). The corresponding value of β was 1.98 by the Berman–Turner method and 2.01 by the polynomial method.

Various edge corrections (Section 9.1) were tried, all using a 50 \times 50 grid of dummy points. Using the border method, eroding the window by a distance $r = 0.7$ m, we obtained

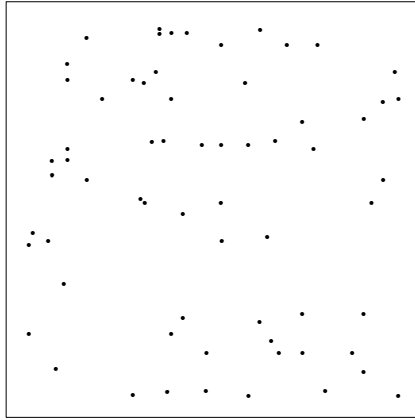


Figure 2. The Swedish pines data: locations of 71 pine saplings in a 10 m \times 10 m square. Extracted by Ripley (1981) from Strand (1972). Data obtained from the MASS library accompanying Venables & Ripley (1994).

$\hat{\gamma} = 0.13$ and β values of 3.24 (Berman–Turner) and 3.29 (polynomial). Periodic edge correction yielded $\hat{\beta} = 2.09$, $\hat{\gamma} = 0.24$ (Berman–Turner) and $\hat{\beta} = 2.24$, $\hat{\gamma} = 0.22$ (polynomial). Our proposed edge corrected pseudolikelihood method was also applied, using the translation correction (Ripley, 1988; Baddeley, 1998) as the edge correction factor $e(u, v, W)$. The parameter estimates were $\hat{\beta} = 1.97$, $\hat{\gamma} = 0.25$. The latter two edge corrections inflated the estimate of γ while the border correction deflated it.

A graph of the profile log-pseudolikelihood of the interaction distance r (Figure 3) yields $\hat{r} = 0.7$, which agrees with Venables & Ripley (1994 p.396). The jaggedness of the plot is due to the discontinuity of the interpoint interaction: $\mathbf{1}\{\|u - x_i\| \leq r\}$ and hence $s(\mathbf{x})$ are discontinuous functions of r , while the left sides of (25)–(26) are differentiable with respect to r . There seems little prospect of a convenient limit theory for \hat{r} .

Next we estimated the covariance matrix of the parameter estimates using the parametric bootstrap (Section 9.3). To reduce the amount of computation we did not apply edge correction and looked at only one set of quadrature weights (based on a 50×50 regular grid). However, r was estimated by profile pseudolikelihood. This version of the estimation algorithm was first applied to the data yielding $(\hat{\beta}, \hat{\gamma}, \hat{r}) = (1.9781, 0.2131, 0.7000)$. A Metropolis–Hastings birth–death–shift algorithm (Geyer & Møller, 1994) generated 500 simulated realizations from the Strauss process with the same parameter values. The bootstrap covariance matrix, based on 500 parametric bootstrap replicates, was

$$\hat{C} = \begin{bmatrix} 0.1938 & -0.0155 & 0.0036 \\ -0.0155 & 0.0063 & 0.0008 \\ 0.0036 & 0.0008 & 0.0014 \end{bmatrix},$$

yielding corresponding (normal-based) 95% confidence intervals of [1.1153, 2.8410], [0.0575, 0.3686], and [0.6267, 0.7733] for β , γ , and r , respectively.

Normality of the estimates is suspect. Chi-squared tests for normality on the sequences of bootstrap replicates of estimates gave P -values of 0.02, 0.22, and 0.00, respectively, for the normality of $\hat{\beta}$, $\hat{\gamma}$, and \hat{r} ; so $\hat{\gamma}$ is the only estimate which may legitimately be assumed normal. However, rough 95% confidence intervals based on the empirical quantiles

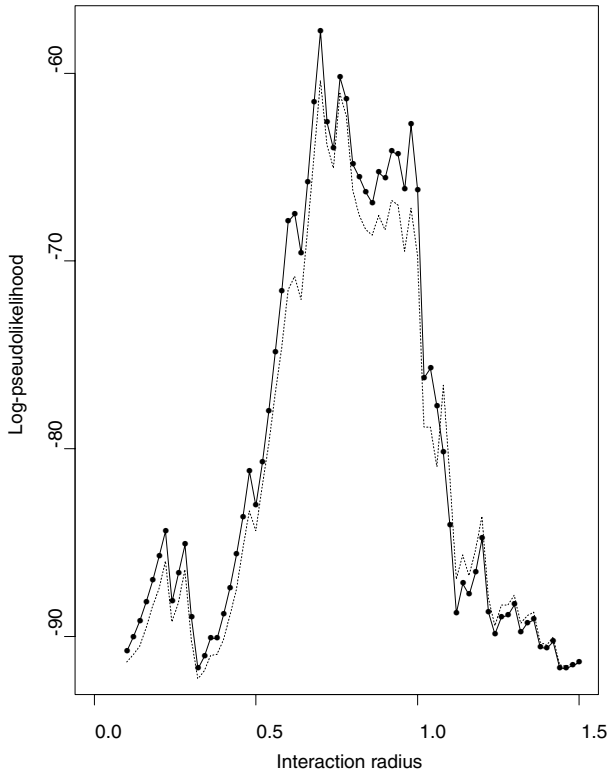


Figure 3. Profile log-pseudolikelihood of the Strauss process irregular parameter r for the Swedish pines data. — : polynomial method and (23); ····: Berman–Turner device and (16). For comparison, the homogeneous Poisson process achieves a maximum log-pseudolikelihood of -92.4 (see Section 4.2).

of the bootstrap replicates were calculated as $[1.29, 2.73]$, $[0.09, 0.39]$, and $[0.62, 0.81]$, respectively. These are in broad agreement with the normal-based intervals.

The confidence interval for γ easily captures Ripley’s (edge-corrected) value of 0.15. However, the β interval does not embrace the corresponding β -value of 3.11. Thus β appears to be more sensitive to the estimation method than does γ .

10.2. Swedish pines data — Ord’s model

Following remarks of Ord in Ripley (1977 Discussion) and Ripley (1981 p.175) we attempted to fit Ord’s model (Section 6.3) to the Swedish pines data. For simplicity we took a ‘Strauss-type’ kernel

$$g_{\theta}(v) = \begin{cases} \beta\gamma & \text{if } v \leq v_0, \\ \beta & \text{if } v > v_0, \end{cases} \tag{37}$$

where $\beta, \gamma > 0$ are parameters to be estimated and the threshold $v_0 > 0$ is an irregular parameter. Figure 4 shows the profile log-pseudolikelihood of v_0 for the Swedish pines data. The calculations were done using the Berman–Turner technique with Dirichlet weights, and the border edge correction.

The jaggedness of the graph may again be explained by discontinuity of the kernel. There is a sharp peak at $v_0 = 1.10 \text{ m}^2$. Adopting this value as the threshold, the parameter estimates

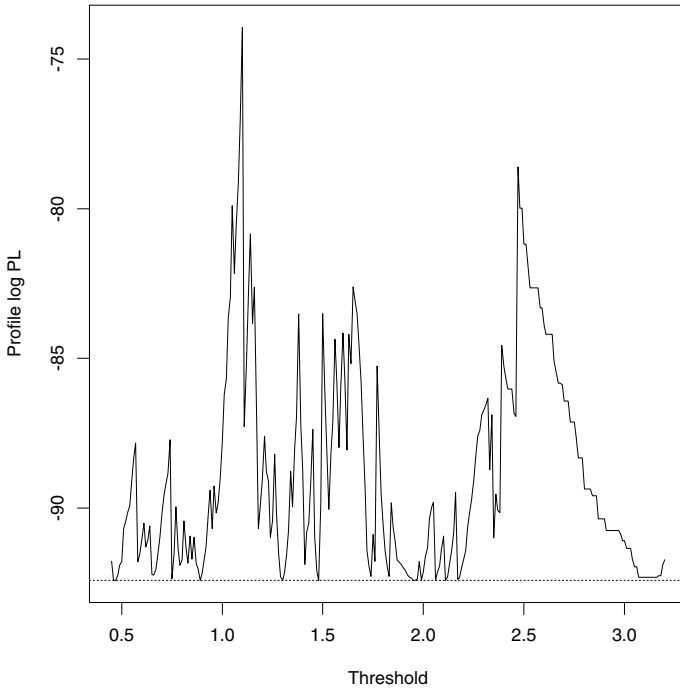


Figure 4. Profile log-pseudolikelihood (PL) for the threshold parameter v_0 of the Strauss-type kernel (37) in the Ord model for the Swedish pines data. The peak is at $v_0 = 1.10 \text{ m}^2$. Horizontal line shows log-pseudolikelihood for the homogeneous Poisson model.

for the Swedish pines data are $\beta = 1.70$ and $\gamma = 0.43$. Figure 5 shows the result of fitting the same model (with v_0 fixed at 1.10) to 100 simulations of a binomial process, i.e. 71 independent uniformly distributed points in the same region. It indicates very strong dependence between $\hat{\beta}$ and $\hat{\gamma}$ for the binomial process. The graph confirms that the Swedish pines data appear to be strongly ordered.

10.3. Japanese black pines data

Figure 6 depicts the Japanese black pines data of Numata (1964) giving the locations of 204 seedlings in a $10 \text{ m} \times 10 \text{ m}$ square. Ogata & Tanemura (1986) used approximate maximum likelihood estimation to fit a soft core model with log-polynomial trend (i.e. where $B(u)$ in (27) is a polynomial in the Cartesian coordinates), and found that a cubic polynomial gave the optimal fit.

In our analysis we first fitted a soft core model with log-cubic trend. The homogeneous soft core model has already been discussed in Section 6.1.2. Adding the polynomial trend to the model is trivial using our approach; it is simply a matter of adding polynomial terms in the Cartesian coordinates to the linear predictor in the associated loglinear model. The estimation of the irregular parameter κ is problematic, so we initially set $\kappa = 0.5$ arbitrarily. Both Dirichlet and counting weights were used; the results were visually indistinguishable. The fitted trend surface shown (Figure 7) has contours similar to those obtained by Ogata & Tanemura (1986). Edge corrections had little effect on the fit, suggesting that edge effects are negligible.

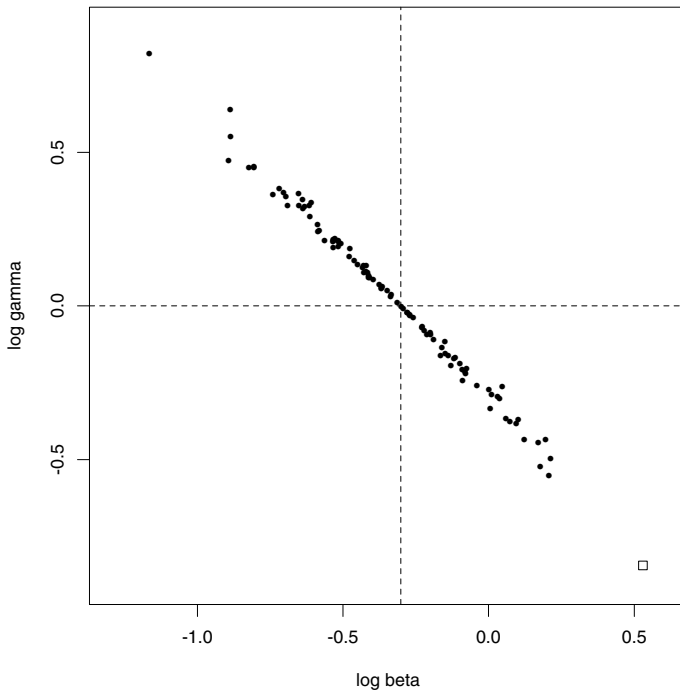


Figure 5. Estimated parameters $\log \beta$ and $\log \gamma$ for the Strauss-type Ord kernel (37) for the Swedish pines data (□ — lower right hand corner) and for 100 simulations of a binomial process with the same number of points (●). Dashed lines indicate $\beta = \hat{\lambda}$ (the estimated intensity) and $\gamma = 1$ (corresponding to a Poisson process).

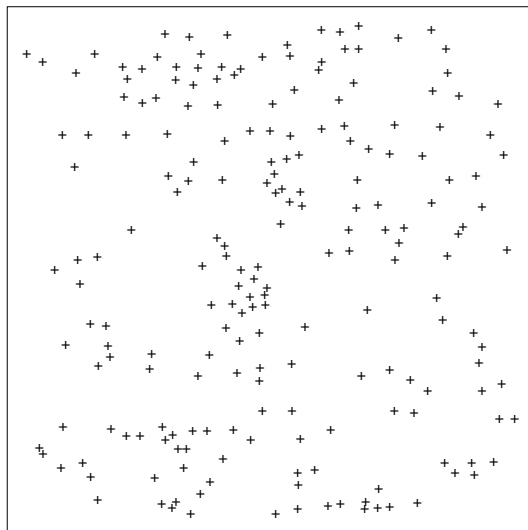


Figure 6. Map of a natural stand of seedlings and saplings of Japanese black pine (*Pinus thunbergii*), 204 seedlings and saplings in a sampling rectangle 10 m \times 10 m. Source: Numata (1964). Data kindly supplied by Professor Y. Ogata and Professor M. Tanemura.

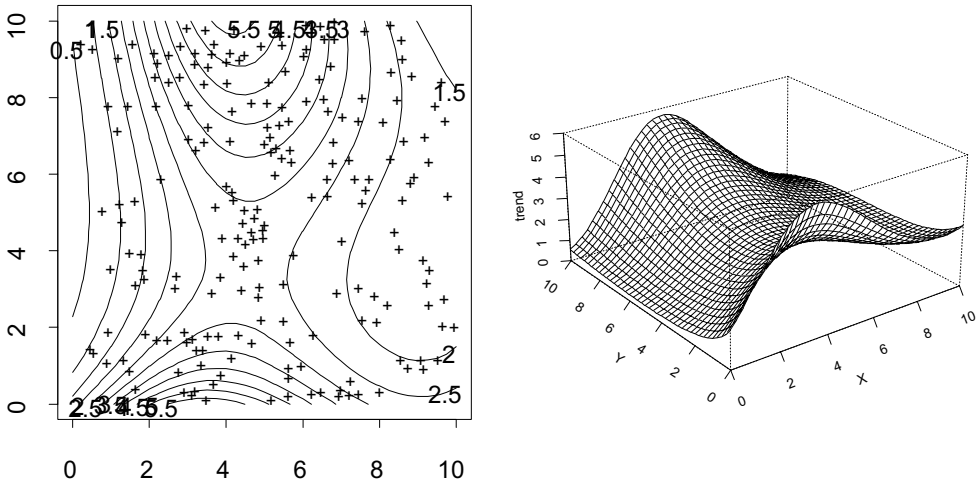


Figure 7. Fitted log-cubic trend surface $\exp(\hat{\theta}^\top B(u))$ for the Japanese black pines data with soft core interaction model: perspective plot (right); contour plot (left)

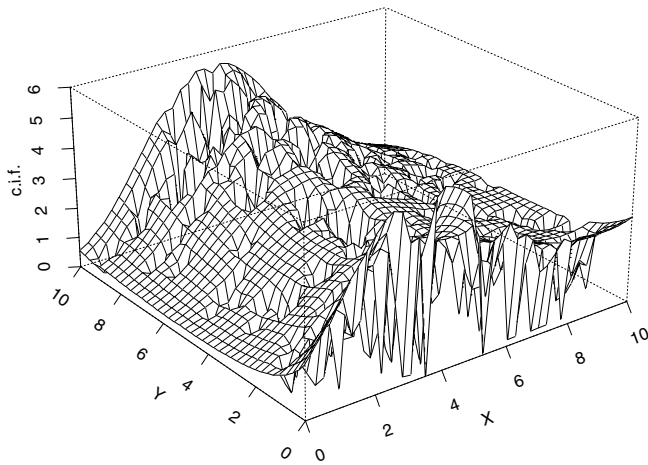


Figure 8. Fitted conditional intensity function $\lambda_{\hat{\theta}}(\cdot; \mathbf{x})$ for the Japanese black pines data with soft core interaction model

We also find it helpful to graph the fitted conditional intensity function $\lambda_{\hat{\theta}}(\cdot; \mathbf{x})$ (Figure 8). This is not a substitute for plotting the trend surface, because the conditional intensity depends on the realized pattern \mathbf{x} . Its usefulness lies in visualizing the effect of the fitted interaction model on the underlying trend, the relative magnitudes and ranges of the trend and interaction terms, and the trade-off between these two (when comparing different models). The graph also helps in checking discretization effects.

Other interaction terms and trend terms can be fitted at little extra cost using the Berman-Turner device, in contrast to the extra effort required for maximum likelihood or simulation-based approaches. It is of interest to compare the foregoing fit with that obtained using a Strauss model for the interaction (along with a log-cubic polynomial spatial trend). In obtaining the Strauss fit we estimated the interaction radius r by maximizing the profile pseudolikelihood, as well as estimating the parameters β and γ .

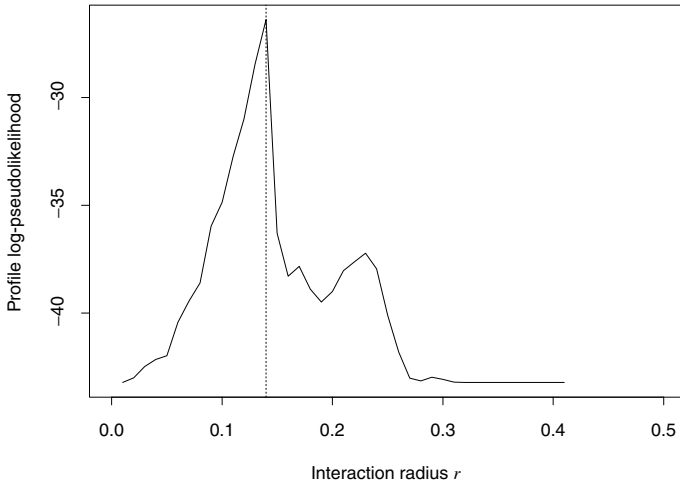


Figure 9. Profile log-pseudolikelihood of the Strauss interaction range r for the Japanese black pines data; Strauss process, with log-cubic polynomial trend, fitted using the Berman–Turner device. Maximum occurs at $r = 0.14$.

The Strauss parameter γ must satisfy $0 \leq \gamma \leq 1$. In this analysis we had to impose the constraint explicitly, i.e. for some values of r an estimate $\hat{\gamma} > 1$ was obtained, whereupon we set $\hat{\gamma} = 1$, and adjusted β and the pseudolikelihood accordingly, as noted in Section 3.3. A graph of the resulting profile log-pseudolikelihood (Figure 9) yields $\hat{r} = 0.14$. This value is just less than the minimum interpoint distance for the Japanese black pines dataset. That is, when a spatial trend is allowed for, the optimal Strauss model for the interaction is the hard core model. It is probably a result of this fact that the profile here is so much smoother than that depicted in Figure 3.

Using $r = 0.14$ we fitted the Strauss model with log-cubic polynomial trend. The fitted trend was visually identical to that obtained for the inhomogeneous soft core model, and is not shown. The fitted conditional intensity function (Figure 10) is essentially the trend surface ‘with holes of radius 0.14 punched in it’ at each data point.

Although the trend surfaces are visually identical, one might ask for a more objective assessment of the difference between the two trends. The differences between the corresponding polynomial coefficients appear to be relatively small. The maximum percentage difference,

$$\frac{|\text{est}_1 - \text{est}_2|}{\frac{1}{2}(|\text{est}_1| + |\text{est}_2|)} \times 100\%,$$

is about 7%. Yet it is not clear how to assess the magnitude of these differences. A rough idea might be given by dividing the differences by an estimate of the standard deviations of, say, the Strauss fits, obtained by bootstrapping. When this is done, the maximum absolute value of the resulting ratios is 0.0956 (corresponding to the x^3 coefficient). Intuitively this confirms the visual impression that there is no evidence of a difference between the two fitted trends.

A trend was also fitted to the Japanese black pines data in the form of a general non-parametric smooth function, as foreshadowed in Section 7.1. The smooth function was provided by the S-PLUS function `lo()`, and fitted using the function `gam()` in place of `glm()`. Both Strauss and soft core models were used for the interaction. When the Strauss model was

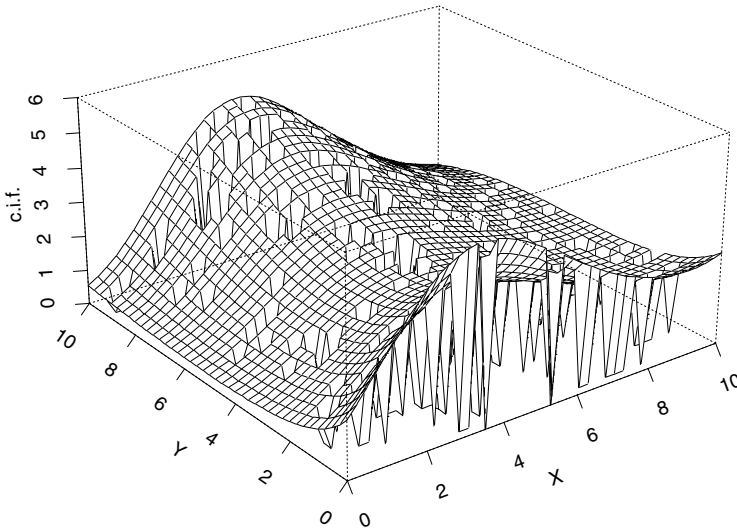


Figure 10. Conditional intensity function for the Japanese black pines data with the interaction modelled as a Strauss process

used, profile pseudolikelihood indicated a value of 0.14 for the r parameter — i.e. a hard core model — the same as for the cubic polynomial trend. For the Strauss (hard core) model the intensity function and the trend surfaces were visually indistinguishable from those obtained using the cubic polynomial trend. For the soft core model the interaction seemed much more subdued when the trend was modelled using $1_0(\cdot)$; the plot of the conditional intensity showed not much more than dimples in the trend surface.

This finding reinforces the principle that the more freedom we allow for the trend, the closer the trend fits the actual data, so the less interaction is needed to explain the data. This effect is not noticeable with the hard core interaction which cannot adapt itself to the smooth trend surface. The $1_0(\cdot)$ trend itself, with soft core interaction, was visually very similar to the cubic polynomial trend, but slightly lower.

Next we attempted to estimate the soft core irregular parameter κ . The approximate profile log-pseudolikelihood can be calculated from the output of the GLM fitting algorithm, via (16). However, graphs of this quantity and of the parameter estimates suggested that small values of κ lead to numerical instability. This persisted when different starting values and different statistical packages (S-PLUS, GLIM) were used. Note that the interaction potential (31) is unbounded, with infinities at the data points, and the approximate pseudolikelihood is not uniformly continuous in κ , even for fixed data and dummy points. Hence the quadrature schemes advocated in Section 4.3 appear to be inadequate for the profile pseudolikelihood.

An alternative numerical integration procedure was then implemented using the mid-point rule and a fine array of integration points. Figure 11 shows the resulting approximate profile log-pseudolikelihood. It suggests that the maximum occurs very close to $\kappa = 0$. In fact the pseudolikelihood can have an infinite maximum at a point where $\kappa = 0$ in certain circumstances. Let $d = \min_{i \neq i'} \|x_i - x_{i'}\|$ be the minimum interpoint distance and $b = \sup_{u \in W} \min_{i=1, \dots, n(x)} \|u - x_i\|$ be the maximum distance from a location in W to the nearest data point. If $b \leq d$ then the soft core pseudolikelihood has an infinite maximum at $\sigma = d$ and $\kappa = 0$.

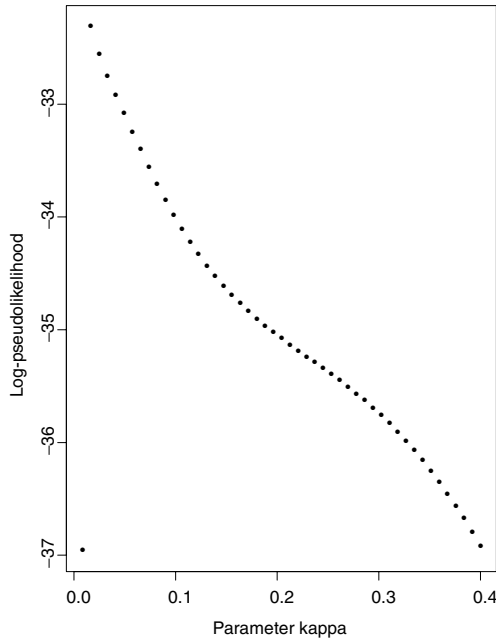


Figure 11. Profile log-pseudolikelihood of the soft core process irregular parameter κ for the Japanese black pines data, calculated by the ‘exact’ method

TABLE 1
Experiments exploring the sensitivity of estimates of the soft core model parameters β, σ to the irregular parameter κ

	$\hat{\beta}$	$\hat{\sigma}$	True κ	κ used
mean	2.29	0.10	0.75	0.75
sd	0.24	0.02		
mean	2.01	0.10	0.75	0.25
sd	0.16	0.02		
mean	1.87	0.11	0.9	0.5
sd	0.18	0.02		
mean	2.46	0.10	0.1	0.5
sd	0.20	0.01		

This finding reinforces the claim (Section 3.2) that pseudolikelihood is inefficient for strong interactions. On the other hand, the estimates of β and σ are not overly sensitive to κ . This sensitivity was investigated in a small simulation study. Realizations of a stationary soft core process were generated in a 10×10 window using parameter values $\beta = 2.3, \sigma = 0.1$ close to those fitted to the Japanese pines data, but with various values of κ . In four separate experiments, 100 realizations of the process with a particular ‘true’ κ value were generated, and from each realization β and σ were estimated using another value of κ . The results are summarized in Table 1. In the second row, β is substantially underestimated when the wrong value of κ is used, but the estimates appear to be still tolerable. In the third and fourth rows, β is very substantially underestimated when the true value of κ is large, and slightly overestimated when the true value of κ is small, but is still accurate to one significant figure. The estimates of σ (arguably the more important parameter) are good in all instances.

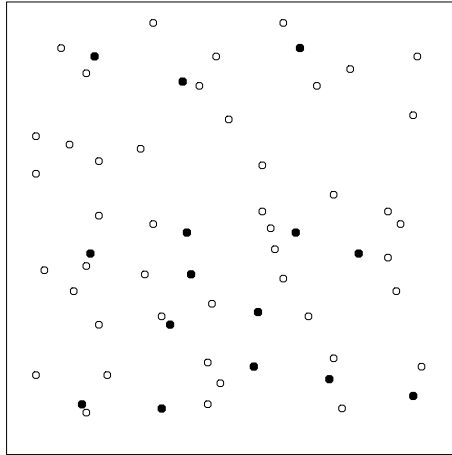


Figure 12. Map of the locations of nests of two species of ants, *Messor wasmanni* (○) and *Cataglyphis bicolor* (●) in a 250 metre square area. Rotated subset of original data of Harkness & Isham (1983) demarcated by Särkkä (1993).

10.4. Harkness–Isham ant data

Figure 12 shows the locations of nests of two species of ants, *Messor wasmanni* and *Cataglyphis bicolor* in a 250 metre square area. These are a subset of the original data collected by Professor R.D. Harkness and analysed in Harkness & Isham (1983), Isham (1984) and Takacs & Fiksel (1986). We follow Särkkä (1993) in restricting attention to this subset.

For comparison we fitted the Strauss–hard core model used by Takacs & Fiksel (1986), Särkkä (1993) and others. This is a pairwise interaction two-type point process (34) with $b_\theta(x, m) = \beta_m$ and an interaction that is a combination of the Strauss type with a hard core:

$$h_\theta((x, i), (y, j)) = \begin{cases} 0 & \text{if } \|x - y\| \leq r_{ij}, \\ \gamma_{ij} & \text{if } r_{ij} < \|x - y\| \leq R_{ij}, \\ 1 & \text{otherwise.} \end{cases}$$

Here $\beta_1, \beta_2, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}$ are parameters with $\gamma_{12} = \gamma_{21}$, and $r_{ij} < R_{ij}$ are non-zero interaction distances with $r_{12} = r_{21}$, $R_{12} = R_{21}$. We define

$$t_\ell((u, m), \mathbf{v}) = \#\{i : r_{m\ell} < \|u - x_i\| \leq R_{m\ell} \text{ and } m_i = \ell\}.$$

The conditional intensity is

$$\lambda((u, m), \mathbf{v}) = \beta_m \gamma_{m1}^{t_1((u,m),\mathbf{v})} \gamma_{m2}^{t_2((u,m),\mathbf{v})},$$

provided that $\|u - x_i\| > r_{m,m'}$ for all $(x_i, m') \in \mathbf{v}$; and $\lambda((u, m), \mathbf{v}) = 0$ otherwise. That is, the conditional intensity is zero whenever there is a marked point of the pattern lying within the relevant hard core distance of the marked point in question. As discussed in Section 4, the appropriate action is to delete from our set of dummy points all those points (x, m) for which there is a data point of type 1 within distance r_{m1} of x or a data point of type 2 within distance r_{m2} of x .

TABLE 2

Parameter estimates for the Strauss or hard core model for the ant data

Method	β_1	β_2	γ_{11}	γ_{12}	γ_{22}
Fixed, 50 × 50	0.026	0.001	0.431	0.977	0.360
Fixed, 20 × 20	0.018	0.001	0.464	0.996	0.383
Random, 20 × 20	0.026	0.001	0.443	0.910	0.352
Särkkä			0.43	0.88	0.41

There are five ‘regular’ parameters $\beta_m, \gamma_{m,m'}$ (which appear in the conditional intensity in loglinear form (8)), and six ‘irregular’ parameters $r_{m,m'}, R_{m,m'}$. The irregular parameters could be estimated by maximum profile pseudolikelihood, but the dimensionality makes this impractical. Instead we adopted Särkkä’s (1993) values of the interaction radii, $r_{11} = 9.1; r_{12} = 5.0; r_{22} = 2.5; R_{11} = R_{12} = R_{22} = 45.0$ except that we changed the value of r_{12} from 6.1 to 5.0 to accommodate an interpoint distance of 5.6 in the data.

The S-PLUS invocation to fit the model is of the form

```
glm(y ~ marks + t11 + t12 + t22, family=poisson,weights=w)
```

where y and w are S-PLUS vectors containing the responses $y_j = z_j/w_j$, and the quadrature weights w_j respectively, $t11, t12$ and $t22$ are the ‘explanatory’ variables containing the values of $t_{ij}(x, \mathbf{v})$, and $marks$ is a factor with levels 1 and 2 containing the marks for the data or dummy points.

To allow direct comparison with the analysis by Särkkä (1993), we used periodic edge correction (Section 9.1.1). Results are shown in Table 2 together with Särkkä’s estimates of the γ parameters (estimates of β_i were not stated). The estimation procedures used all relied upon the Berman–Turner approximation and the counting quadrature weights, with (a) fixed 50 × 50 grid of dummy points, (b) fixed 20 × 20 grid of dummy points, and (c) random dummy points each of which was uniformly distributed on each of a 20 × 20 grid of cells covering the window. This last was done so as to emulate Särkkä’s procedure as closely as possible. The values obtained are reasonably commensurate with those of Särkkä (1993).

11. Simulation study

To assess the performance of the maximum pseudolikelihood procedure, we generated 500 simulated realizations from the Strauss process in a 10 × 10 square, with $(\beta, \gamma, r) = (3.0, 0.15, 0.7)$ suggested by the fit in Section 10.1. To imitate the unbounded process, data were actually generated in an expanded square of area 200 and clipped to the given square. We produced each realization by 600 000 iterations of a Metropolis–Hastings birth–death–shift procedure (Geyer & Møller, 1994). We chose this number by inspecting the autocorrelation of the sufficient statistics $n(\mathbf{x}), s(\mathbf{x})$ which fell to 0 at a lag of about 15 000.

Estimates of β and γ were computed from each simulated pattern. To reduce the computational effort, r was fixed at the true value of 0.7. We applied the Berman–Turner device (border correction, counting weights), the polynomial method (border correction) and the polynomial method with Ripley’s hybrid correction. Table 3 shows the sample means of the various estimates, and their standard errors. The results using Ripley’s hybrid correction were obtained using our software, but with \hat{K} calculated by Venables and Ripley’s S-PLUS function `Kfn()`.

TABLE 3

Means of parameter estimates from 500 simulated realizations of the Strauss process with true values $\beta = 3.0$ and $\gamma = 0.15$. Numbers in parentheses are standard errors.

Method	Grid size	Not edge corrected		Edge corrected	
		$\hat{\beta}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\gamma}$
Berman–Turner	25	2.270 (0.019)	0.195 (0.003)	2.643 (0.028)	0.179 (0.003)
Polynomial	25	2.447 (0.023)	0.181 (0.003)	3.139 (0.039)	0.152 (0.003)
Ripley	25			3.137 (0.039)	0.152 (0.003)
Berman–Turner	50	2.607 (0.025)	0.170 (0.003)	2.995 (0.036)	0.160 (0.003)
Polynomial	50	2.601 (0.025)	0.170 (0.003)	3.134 (0.040)	0.153 (0.003)
Ripley	50			3.132 (0.039)	0.152 (0.003)
Berman–Turner	100	2.710 (0.027)	0.163 (0.003)	3.098 (0.039)	0.155 (0.003)
Polynomial	100	2.672 (0.027)	0.165 (0.003)	3.136 (0.040)	0.153 (0.003)
Ripley	100			3.134 (0.040)	0.153 (0.003)

On the basis of the mean values, Ripley's hybrid correction and the border corrected polynomial method do well even for a grid size of 25. For grid sizes of 50 and 100 all methods are comparable as long as edge correction is used. The uncorrected versions appear to underestimate β by 10–25%, and to overestimate γ by similar amounts. The edge corrected versions overestimate both β and γ by about 3–4% in general.

When the performance of the estimates is compared on the basis of root mean squared error (RMSE), a somewhat different story is told. The RMSE values as fractions of the true parameter values are displayed in Table 4. In this table the RMSE is also broken down into absolute bias and standard deviation, also expressed as fractions of the true parameters.

Edge correction appears to reduce the RMSE only when a coarse grid is used. The RMSE in the estimates of β has its smallest overall value when the estimate is *not* edge corrected. It is around 23% of the true value (for both the 50×50 and the 100×100 grids) when no edge correction is used, and goes up by about 6 or 7 percentage points with edge correction. We observe that the bias is substantially diminished by edge correction, but that the standard deviation increases. This effect is much less pronounced for γ . The RMSE (for the estimates from the finer grids) goes up under edge correction by only about 0.5 to 1.5 percentage points, and in one instance (for the polynomial method, 50×50 grid) diminishes slightly. Thus in this one instance, edge correction seems to improve efficiency.

12. Discussion and conclusions

It appears that this extension of the Berman–Turner device works well (and has the advantage of being easy to implement and extremely versatile). Edge corrections require further investigation.

The computation times depend, of course, upon the quadrature scheme used and upon the number of quadrature points — data points and dummy points. The following are typical

TABLE 4

Root mean squared error (RMSE), bias, and standard deviation of parameter estimates in the simulation study, expressed relative to true values

Method	Grid size	Not edge corrected		Edge corrected	
		$\hat{\beta}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\gamma}$
Relative RMSE					
Berman-Turner	25	0.281	0.502	0.240	0.474
Polynomial		0.253	0.456	0.297	0.402
Ripley				0.294	0.379
Berman-Turner	50	0.228	0.403	0.270	0.419
Polynomial		0.231	0.408	0.296	0.405
Ripley				0.294	0.381
Berman-Turner	100	0.224	0.385	0.293	0.409
Polynomial		0.226	0.393	0.300	0.407
Ripley				0.298	0.383
Relative bias					
Berman-Turner	25	-0.243	+0.298	-0.119	+0.191
Polynomial		-0.184	+0.203	+0.046	+0.015
Ripley				+0.046	+0.012
Berman-Turner	50	-0.131	+0.132	-0.002	+0.063
Polynomial		-0.133	+0.131	+0.045	+0.019
Ripley				+0.044	+0.016
Berman-Turner	100	-0.097	+0.089	+0.033	+0.031
Polynomial		-0.109	+0.102	+0.045	+0.019
Ripley				+0.045	+0.017
Relative standard deviation					
Berman-Turner	25	0.141	0.404	0.209	0.434
Polynomial		0.174	0.409	0.293	0.402
Ripley				0.290	0.379
Berman-Turner	50	0.187	0.381	0.270	0.414
Polynomial		0.189	0.387	0.293	0.404
Ripley				0.291	0.381
Berman-Turner	100	0.202	0.375	0.291	0.408
Polynomial		0.198	0.379	0.296	0.406
Ripley				0.294	0.383

times for a modest (133 MHz, 96 Mb) workstation. Fitting a Strauss model (with 'known' r) to the Swedish pines data took about 2.9 cpu seconds, using Dirichlet weights and 625 dummy points (for a total of 696 points). With 3600 dummy points for a total of 3671 points, the fit took about 32 seconds. Using counting weights and a 50×50 grid (for a total of 2571 points), the fit took 1.6 seconds. With a 100×100 grid (10 071 points) it took about 5.2 seconds.

Addition of a spatial trend naturally increases the computational burden. It took 6.1 seconds to fit a soft core model to the Numata pines data, with counting weights and a 50×50 grid. With a log-cubic polynomial trend added to the model, the fit took 12.3 seconds. The smooth trends fitted using `lo()` and `gam()` were much more time consuming. A soft core model as above, with a smooth trend, took about 211 seconds to fit. Likewise to fit an Ord model is very time consuming. To fit such a model to the Swedish pines data with a 20×20 grid of dummy points (471 points in all) took about 350 seconds of cpu time, or about 0.75 seconds per point. Note, however, that this computation was done in a very sub-optimal way — the calculation of the conditional intensity at each data and dummy point involved recomputing the Dirichlet tessellation from scratch.

This paper has not dealt with replicated patterns, or more generally, designed experiments where the response is a point pattern. Such data can be analysed using the same approach.

It would be of great interest to fit smooth curves (in the style of generalized additive models) to the pair potential function $H(\|u - v\|)$ in a pairwise interaction process. It appears to us that existing software cannot be applied directly.

We have implemented the methods of this paper in an SPLUS software package which is available at <http://www.maths.uwa.edu.au/~adrian/spatstat.html>.

References

- AITKIN, M., ANDERSON, D.A., FRANCIS, B.J. & HINDE, J.P. (1989). *Statistical Modelling in GLIM*. Oxford University Press.
- BADDELEY, A.J. (1998). Spatial sampling and censoring. In *Stochastic Geometry: Likelihood and Computation*, eds O.E. Barndorff-Nielsen, W.S. Kendall & M.N.M. van Lieshout, Chapter 2. London: Chapman & Hall.
- BADDELEY, A.J. & MØLLER, J. (1989). Nearest-neighbour Markov point processes and random sets. *Internat. Statist. Rev.* **57**, 89–121.
- BADDELEY, A.J. & VAN LIESHOUT, M.N.M. (1995). Area-interaction point processes. *Ann. Inst. Statist. Math.* **47**, 601–619.
- BECKER, R.A., CHAMBERS, J.M. & WILKS, A.R. (1988). *The NEW S Language*. Chapman & Hall.
- BERMAN, M. (1986). Testing for spatial association between a point process and another stochastic process. *Appl. Statist.* **35**, 54–62.
- BERMAN, M. & TURNER, T.R. (1992). Approximating point process likelihoods with GLIM. *Appl. Statist.* **41**, 31–38.
- BESAG, J. (1975). Statistical analysis of non-lattice data. *The Statistician* **24**, 179–195.
- BESAG, J. (1977). Some methods of statistical analysis for spatial data. *Bull. Internat. Statist. Inst.* **47**, 77–92.
- BESAG, J., MILNE, R. & ZACHARY, S. (1982). Point process limits of lattice processes. *J. Appl. Probab.* **19**, 210–216.
- CHAMBERS, R.M. & HASTIE, T. (Eds). (1992). *Statistical Models in S*. Wadsworth & Brooks/Cole.
- COX, D.R. & ISHAM, V. (1980). *Point Processes*. London: Chapman & Hall.
- CRESSIE, N.A.C. (1991). *Statistics for Spatial Data*. New York: John Wiley & Sons.
- CUZICK, J. & EDWARDS, R. (1990). Spatial clustering for inhomogeneous populations. *J. Roy. Statist. Soc. Ser. B* **52**, 73–104.
- DIGGLE, P.J. (1983). *Statistical Analysis of Spatial Point Patterns*. London: Academic Press.
- DIGGLE, P.J., FIKSEL, T., GRABARNIK, P., OGATA, Y., STOYAN, D. & TANEMURA, M. (1994). On parameter estimation for pairwise interaction. *Internat. Statist. Rev.* **62**, 99–117.
- FIKSEL, T. (1984). Estimation of parameterized pair potentials of marked and non-marked Gibbsian point processes. *Elektron. Informationsverarb. Kybernet.* **20**, 270–278.
- FIKSEL, T. (1988). Estimation of interaction potentials of Gibbsian point processes. *Statistics* **19**, 77–86.
- FREY, A. & SCHMIDT, V. (1998). Marked point processes in the plane I — a survey with applications to spatial modeling of communication networks. *Adv. in Performance Analysis* **1**, 65–110.
- GEYER, C.J. (1999). Likelihood inference for spatial point processes. In *Stochastic Geometry: Likelihood and Computation*, eds O.E. Barndorff-Nielsen, W.S. Kendall & M.N.M. van Lieshout, number 80 in Monographs on Statistics and Applied Probability, Chapter 3. CRC, Boca Raton: Chapman & Hall.
- GEYER, C.J. & MØLLER, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Statist.* **21**, 359–373.
- GLÖTZL, E. (1980a). Bemerkungen zu einer Arbeit von O.K. Kozlov. *Math. Nachr.* **94**, 277–289.
- GLÖTZL, E. (1980b). Lokale Energien und Potentiale für Punktprozessen. *Math. Nachr.* **96**, 195–206.
- GOULARD, M., SÄRKKÄ, A. & GRABARNIK, P. (1996). Parameter estimation for marked Gibbs point processes through the maximum pseudolikelihood method. *Scand. J. Statist.* **23**, 365–379.
- HAMMERSLEY, J.M., LEWIS, J.W.E. & ROWLINSON, J.S. (1975). Relationships between the multinomial and Poisson models of stochastic processes, and between the canonical and grand canonical ensembles in statistical mechanics, with illustrations and Monte Carlo methods for the penetrable sphere model of liquid-vapour equilibrium. *Sankhyā Ser. A* **37**, 457–491.
- HARKNESS, R.D. & ISHAM, V. (1983). A bivariate spatial point pattern of ants' nests. *Appl. Statist.* **32**, 293–303.
- HASTIE, T. & TIBSHIRANI, R. (1990). *Generalized Additive Models*. London: Chapman & Hall.
- ISHAM, V.S. (1984). Multitype Markov point processes: some approximations. *Proc. Roy. Soc. London Ser. A* **39**, 39–53.
- JENSEN, J.L. & KÜNSCH, H.R. (1994). On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Ann. Inst. Statist. Math.* **46**, 475–486.
- JENSEN, J.L. & MØLLER, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.* **1**, 445–461.

- KALLENBERG, O. (1983). *Random Measures*. 3rd edn. Berlin/New York: Akademie Verlag/Academic Press.
- KALLENBERG, O. (1984). An informal guide to the theory of conditioning in point processes. *Internat. Statist. Rev.* **52**, 151–164.
- KELLY, F.P. & RIPLEY, B.D. (1976). On Strauss's model for clustering. *Biometrika* **63**, 357–360.
- KINGMAN, J.F.C. (1993). *Poisson Processes*. Oxford University Press.
- KOZLOV, O.K. (1976). Gibbsian description of point random fields. *Theory Probab. Appl.* **21**, 339–355.
- LINDSEY, J.K. (1992). *The Analysis of Stochastic Processes using GLIM*. Berlin: Springer.
- LINDSEY, J.K. (1995). Fitting parametric counting processes by using linear models. *Appl. Statist.* **44**, 201–212.
- LINDSEY, J.K. (1996). Fitting bivariate intensity functions, with an application to modelling delays in reporting acquired immune deficiency syndrome. *J. Roy. Statist. Soc. Ser. A*, **159**, 125–131.
- LINDSEY, J.K. & MERSCH, G. (1992). Fitting and comparing probability distributions with log linear models. *Comput. Statist. Data Anal.* **13**, 373–384.
- MCCULLAGH, P. & NELDER, J.A. (1989). *Generalized Linear Models*. 2nd edn. Chapman & Hall.
- MOYEED, R.A. & BADDELEY, A.J. (1991). Stochastic approximation for the MLE of a spatial point process. *Scand. J. Statist.* **18**, 39–50.
- NGUYEN, X.X. & ZESSIN, H. (1976). Punktprozesse mit Wechselwirkung. *Z. Wahrsch. Verw. Gebiete* **37**, 91–126.
- NIELSEN, L.S. & JENSEN, E.B.V. (1998). *Inhomogeneous Markov Point Processes by Transformation*. Research Report 2, Laboratory for Computational Stochastics. Department of Mathematical Sciences, University of Aarhus.
- NUMATA, M. (1964). Forest vegetation, particularly pine stems in the vicinity of Choshi — coastal flora and vegetation in Choshi, Chiba prefecture, IV. (In Japanese) *Bulletin of the Choshi Marine Laboratory* **6**, 27–37. Chiba University.
- OGATA, Y. & TANEMURA, M. (1981). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Ann. Inst. Statist. Math.* **33**, 315–338.
- OGATA, Y. & TANEMURA, M. (1984). Likelihood analysis of spatial point patterns. *J. Roy. Statist. Soc. Ser. B* **46**, 496–518.
- OGATA, Y. & TANEMURA, M. (1985). Estimation of interaction potentials of marked spatial point processes through the maximum likelihood method. *Biometrics* **41**, 421–433.
- OGATA, Y. & TANEMURA, M. (1986). Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In *Pacific Statistical Congress*, eds I.S. Francis, B.J.F. Manly & F.C. Lam, pp.150–154. Elsevier.
- OKABE, A., BOOTS, B. & SUGIHARA, K. (1992). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Wiley series in probability and mathematical statistics. Chichester, England; New York: John Wiley & Sons.
- PENTTINEN, A. (1984). *Modelling Interaction in Spatial Point Patterns: Parameter Estimation by the Maximum Likelihood Method*. Number 7 in Jyväskylä Studies in Computer Science, Economics and Statistics. University of Jyväskylä.
- PRESTON, C.J. (1976). *Random Fields*. Berlin; Heidelberg; New York: Springer Verlag.
- RIPLEY, B.D. (1977). Modelling spatial patterns (with discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 172–212.
- RIPLEY, B.D. (1981). *Spatial Statistics*. New York: John Wiley & Sons.
- RIPLEY, B.D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press.
- RIPLEY, B.D. (1989). Gibbsian interaction models. In *Spatial Statistics: Past, Present and Future*, ed. D.A. Griffiths, pp. 1–19. New York: Image.
- ROWLINSON, J.S. (1980). Penetrable sphere models of liquid–vapor equilibrium. *Adv. Chem. Phys.* **41**, 1–57.
- SÄRKKÄ, A. (1993). Modelling interaction in spatial point patterns: parameter estimation by the maximum likelihood method. Number 22 in Jyväskylä Studies in Computer Science, Economics and Statistics. University of Jyväskylä.
- SÄRKKÄ, A. & HÖGMANDER, H. (1998). *Multitype Spatial Point Patterns with Hierarchical Interactions*. Research Report 1998: 1. Sweden: Department of Statistics, Göteborg University.
- STOYAN, D., KENDALL, W.S., & MECKE, J. (1995). *Stochastic Geometry and its Applications*. 2nd edn. Chichester: John Wiley & Sons.
- STRAND, L. (1972). A model for stand growth. IUFRO *Third Conference Advisory Group of Forest Statisticians*. INRA, Institut National de la Recherche Agronomique, Paris. pp. 207–216.
- 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. *Statistics* **17**, 429–433.
- TAKACS, R. & FIKSEL, T. (1986). Interaction pair-potentials for a system of ants' nests. *Biometrical J.* **28**, 1007–1013.
- TANNER, M.A. (1993). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. 2nd edn. New York: Springer-Verlag.
- VENABLES, W.N. & RIPLEY, B.D. (1994). *Modern Applied Statistics with S-Plus*. Springer.
- WIDOM, B. & ROWLINSON, J.S. (1970). New model for the study of liquid–vapor phase transitions. *J. Chem. Phys.* **52**, 1670–1684.

Discussion of the paper by Adrian Baddeley and Rolf Turner

Peter Diggle (*Medical Statistics Unit, Lancaster University*): I very much enjoyed this paper, which offers a routine, computationally easy method of point estimation for a very wide class of point process models. With the progressive development of Markov chain Monte Carlo methods, one might ask whether ‘quick and easy’ methods of estimation are still needed. My answer is yes.

In my experience, MCMC methods need careful, and sometimes laborious, re-tuning to each new application if misleading results are to be avoided. Exact MCMC methods, which seem well suited to point process models (see e.g. Kendall & Moller, 1999), could eventually change this assessment, but in their current state of development do not offer the flexibility required for routine use.

Simpler methods of the kind described by Baddeley and Turner can help to suggest good proposal distributions for Metropolis-type MCMC algorithms, but are also valuable in their own right in situations where inferential validity is more important than theoretical efficiency. One such situation, to which the authors briefly refer in their discussion, is in the analysis of replicated point patterns from designed experiments, where the design itself can provide the basis for a valid inference based on an appropriate summary of each replicate. Diggle, Lange & Benes (1991) develop this approach using Ripley’s (1977) K -function as a non-parametric summary, and Diggle, Mateu & Clough (2000) describe the analogous approach using pseudo-likelihood estimates as summarizing statistics for a parametric, pairwise interaction process.

Baddeley and Turner place some importance on the fact that their neat computational trick of turning the point process pseudo-likelihood into a classical Poisson likelihood allows the pseudo-likelihood method to be implemented within standard generalized linear modelling software. Their arguments are reasonable, but the advantages seem to be bought at the cost of at least two potential disadvantages. First, by using a mixture of data-locations and non-data-locations as quadrature points, the authors introduce a discontinuity error as discussed in Section 3.3. Second, if we were to approach the evaluation of (11) as a standard quadrature problem, we might prefer to proceed by computing successive approximations based on an increasing sequence of values of m without recomputation of any of the $\lambda(u_j; \mathbf{x})$. Agreement of successive approximations would then provide a direct check on the accuracy of the quadrature. Diggle & Gratton (1984) proposed a similar strategy in the context of Monte Carlo likelihood maximization. This has the advantage that a small value of m can be used initially, with larger values (and hence improved approximations) being introduced as the optimization algorithm homes in on the maximum likelihood estimate.

In Section 6.1.3, the authors consider step-function interactions, apparently as a surrogate for non-parametric estimation of $h(\cdot)$. Diggle, Gates & Stibbard (1987) propose a smooth, non-parametric estimator for $h(\cdot)$, although their method involves some subjective judgement, and I would agree that fitting a piecewise constant interaction function may well be more robust for routine use.

The edge-corrected version of pseudo-likelihood that the authors suggest in Section 9.1.4 is not new. It was proposed in Diggle *et al.* (1994).

The examples in the paper are of unreplicated patterns with relatively small numbers of points in each. It would be interesting to see the method applied to larger, more complex datasets. One rich source of such data is environmental epidemiology, where each point represents an individual case of a particular disease, or an individual control sampled from the

population at risk; the complexity arises because of the considerable spatial heterogeneity in most human settlement patterns. The capacity of the authors' methods to handle the combination of spatial heterogeneity and interactions between points opens up interesting possibilities. Elliott, Martuzzi & Shaddick (1995) give a thoughtful discussion of the role of spatial statistical methods in environmental epidemiology, whilst Elliott *et al.* (2000) review more recent methodological developments including, but not restricted to, point process methods.

My final comments are only very indirectly related to the paper, and concern statistical methods for marked point processes. This seems to me to be an area where formal mathematical theory has been developed with rather little discussion of what underlying scientific problems the analysis of such data is intended to address. I would argue that, from a scientific perspective, the subsuming of what I would call multivariate point processes within the wider class of marked point processes is unhelpful.

In a general sense, a marked point process is a model for two random processes, the locations P and the marks M . By analogy with simple bivariate distributions, the joint probabilistic structure of P and M can be expressed in either of two equivalent ways, namely

$$[P, M] = [P][M | P] = [M][P | M].$$

Case-control data in environmental epidemiology can be modelled formally as a marked point process with a binary mark process M identifying whether or not an individual member of the population at risk acquires the disease in question. The unmarked process P of individual locations irrespective of their disease status is an object which makes physical sense as a description of the population at risk, and could be modelled in its own right. Thus, the factorization $[P, M] = [P][M | P]$ provides a reasonable foundation for modelling. However, the questions to which epidemiologists require answers concern the conditional distribution $[M | P]$, and there is no obvious scientific value in devising elaborate models for $[P]$.

On the other hand, in the area of spatial statistics known as geostatistics the mark process M is derived from an underlying random field $M(x): x \in A$ on a continuous spatial region A , with values of $M(x)$ being observed only at a set of sampling locations which define a point process P . Because $M(x)$ exists in its own right, the more natural factorization on which to base a model is $[P, M] = [M][P | M]$. In many applications, the choice of sampling locations is influenced by prior knowledge of the spatial variation in $M(x)$ and the conditioning of P on M is non-trivial; as a specific example, air pollution monitoring networks often deliberately site monitoring stations close to known or suspected sources of pollution. However, geostatistical analyses typically ignore any stochastic element in P and analyse the data on M under the tacit assumption that P and M are independent. When this is not the case, inferential validity is potentially compromised. By construction, the observed data on $M(x)$ are a realization from the conditional $[M | P]$, not from $[M]$.

Finally, the data in Baddeley and Turner's third example are generated from a pair of potentially interdependent point processes, one for each of two species of ant. In this setting, neither of the two factorizations of $[P, M]$ seems particularly helpful. It would make no scientific (as opposed to mathematical) sense to model either a point process of ants of indeterminate species, or a binary spatial process of the species of a hypothetical ant at an arbitrary location. The model which the authors fit to these data, in which the interaction between a pair of ants depends both on their relative locations and on their respective species, seems sensible for this kind of bivariate point process.

References

- DIGGLE, P.J. & GRATTON, R.J. (1984). Monte Carlo methods of inference for implicit statistical models. *J. Roy. Statist. Soc.* **46**, 193–227.

- DIGGLE, P.J., GATES, D.J. & STIBBARD, A. (1987). A non-parametric estimator for pairwise-interaction point processes. *Biometrika* **74**, 763–770.
- DIGGLE, P.J., LANGE, N. & BENES, F.M. (1991). Analysis of variance for replicated spatial point patterns in clinical neuroanatomy. *J. Amer. Statist. Assoc.* **86**, 618–625.
- DIGGLE, P.J., MATEU, J. & CLOUGH, H.E. (2000). A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns. *Adv. in Appl. Probab.* (to appear).
- DIGGLE, P.J., FIKSEL, T., GRABARNIK, P., OGATA, Y., STOYAN, D. & TANEMURA, M. (1994). On parameter estimation for pairwise interaction. *Internat. Statist. Rev.* **62**, 99–117.
- ELLIOTT, P., MARTUZZI, M. & SHADDICK, G. (1995). Spatial statistical methods in environmental epidemiology: a critique. *Statistical Methods in Medical Research* **4**, 139–161.
- ELLIOTT, P., WAKEFIELD, J.C., BEST, N.G. & BRIGGS, D.J. (2000). *Spatial Epidemiology: Methods and Applications*. Oxford: Oxford University Press.
- KENDALL, W.S. & MOLLER, J. (1999). *Perfect Metropolis–Hastings Simulation of Locally Stable Spatial Point Processes*. Research Report 99-2001. Aalborg: Department of Mathematical Sciences, Aalborg University.
- RIPLEY, B.D. (1977). Modelling spatial patterns (with discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 172–212.

Mark Berman (*CSIRO Mathematical & Information Sciences*): It is a pleasure to be invited to discuss this very comprehensive and useful paper.

The analysis of spatial point patterns has been notoriously difficult, mainly because most of the models require spatial interactions between the points themselves. Many of the more interesting models involving spatial interaction belong to the class of Gibbs processes, which is the class considered by the authors. Because there is no natural ordering in two (or more) dimensions, it is usually impossible to calculate analytically the normalization constants for even the simplest (non-Poisson) Gibbs models (typically because a complicated multiple integral is involved (Cox & Isham, 1980 Equation 6.17)), and hence maximum likelihood estimation (MLE) becomes impracticable in most cases. For this reason, many authors try to maximize Besag's pseudolikelihood, which the authors point out 'can be viewed as an infinite product of infinitesimal conditional probabilities'. The computational advantage of pseudolikelihood over maximum likelihood is that only a two-dimensional integral needs to be calculated, either analytically or numerically. Although maximum pseudolikelihood estimation (MPLE) is usually sub-optimal, in many cases it is at least consistent and asymptotically normal. Unfortunately, until the advent of this paper, each model required its own particular method for solving the MPLE equations. In turn this meant that only the cognoscenti were able to estimate the parameters of most Gibbs point process models.

Adrian Baddeley and Rolf Turner have built on earlier work by Rolf and myself, in which we approximated MLE for general point processes on the line (which are more tractable because a natural ordering exists) and for spatial Poisson processes (which are more tractable because of independence). They have shown how MPLE can be well approximated relatively quickly and easily using simple quadrature schemes and generalized linear or additive models. Their approximation to the pseudolikelihood is formally equivalent to the weighted likelihood of independent Poisson variables, and so software packages such as GLIM and S-PLUS can be used to obtain the MPL estimators. (It is, however, worth repeating here the authors' point that the standard errors produced by these packages are not correct, because the full likelihood is not being maximized. The bootstrap is needed for inference and model selection purposes.)

The power of the authors' relatively simple observation is firstly in the breadth of its application, which is amply demonstrated in the paper, and secondly in the relative speed with which the MPL estimators are calculated for most of the models. This means that the authors have made a significant step towards making the analysis of Gibbs point process models a routine procedure. Such analyses would become truly routine (at least for S-PLUS users) if there were a generic S routine to carry out the bootstrap-based inference for models with the intensity (8). The authors may care to comment on this possibility.

One might ask: how is it that the simple numerical approximation (11) to an integral leads to the Poisson log-likelihood (13) in so many cases? In fact, there is a basic probabilistic explanation for all this, and it is closely related to the way in which likelihoods of many point process models are derived. Divide the bounded observation region W into tiles of equal area Δ . Let X_j denote the number of observations in tile j , and let \mathbf{x}_j represent all the X_i in W except for X_j . Assuming that the point process is *orderly* (Diggle, 1983 p.47), then provided Δ is sufficiently small,

$$\begin{aligned} \Pr(X_j = 0 \mid \mathbf{x}_j) &= 1 - \lambda(j; \mathbf{x}_j) + o(\Delta), \\ \Pr(X_j = 1 \mid \mathbf{x}_j) &= \lambda(j; \mathbf{x}_j) + o(\Delta), \end{aligned}$$

where $\lambda(j; \mathbf{x}_j)$ is the conditional intensity for the j th tile. The (discretized) pseudolikelihood is just $\prod_{j \in W} \Pr(X_j \mid \mathbf{x}_j)$. If we now let Δ approach zero, then under mild regularity conditions ensuring that, in the limit, the $\lambda(j; \mathbf{x}_j)$ do not fluctuate too wildly, the product of conditional Bernoulli probabilities above will converge to (6), which is a variant of the usual Poisson-type approximation to the Binomial. The approximation (11) rediscrretizes the continuous Poisson approximation to a discrete Poisson approximation instead of a Bernoulli one. It therefore seems plausible that we could also approximate (6) by a product of conditional Bernoulli probabilities. Presumably, this approximation could also be analysed using GLIM with Binomial rather than Poisson ‘errors’, although it is not obvious that there are any advantages in doing so.

In our earlier paper (Berman & Turner, 1992) Rolf Turner and I used a quadrature scheme based on the Dirichlet tessellation. This was done so as to ensure that there was at most one point in every tile, in sympathy with the notion of orderliness. The authors describe a computationally cheaper scheme, in which tiles are of equal size and the weight in tile j is a/n_j , where a is the area of each tile, and n_j is the number of dummy or data points in the same tile. Without loss of generality, we can assume that $a = 1$; this just involves a rescaling of our axes. If we make a further approximation that the conditional intensity is constant within a tile, then it is easily shown that the fundamental approximation (13) can be rewritten as

$$\log \text{PL}(\theta; \mathbf{x}) \approx \sum_k (m_k \log \lambda_k - \lambda_k),$$

where the summation is now over tiles (rather than points) and m_k is the number of data points in tile k . The above approximation is formally equivalent to the log likelihood of independent and *unweighted* Poisson variables $Y_k \sim \text{Poisson}(\lambda_k)$. Provided that the tiles are sufficiently small, all the approximations should give similar results. The slight advantage of this one is that all the weights are equal.

Finally, I would like to make a brief observation about the speed of parameter estimation for the Ord model, which is computationally expensive, because many Dirichlet tessellations need to be computed whether for the ‘exact’ pseudolikelihood (6) or its approximation (13). Specifically, the tessellation involving the points $\mathbf{x} \cup u$, where u is a typical dummy point, needs to be computed *for each dummy point*, of which there are typically several hundred or several thousand. In their example (Section 10.2), the authors have recomputed new Dirichlet tessellations for each dummy point. This has made the complete estimation quite slow. Clearly, as one moves systematically and locally from one dummy point to the next (e.g. if they are on a regular grid), there are significant speed gains to be made, using the fact that the tessellation is only likely to change locally, by removing one dummy point and inserting

the next. Efficient algorithms based on successively inserting points into the tessellation are widely available (Okabe, Boots & Sugihara, 1992 Section 4.3), although I am unaware of algorithms that involve both insertion and deletion. Presumably, such an algorithm would involve only a minor modification of existing ones based on insertions alone.

References

- BERMAN, M. & TURNER, T.R. (1992). Approximating point process likelihoods with GLIM. *Appl. Statist.* **41**, 31–38.
- COX, D.R. & ISHAM, V. (1980). *Point Processes*. London: Chapman and Hall.
- DIGGLE, P.J. (1983). *Statistical Analysis of Spatial Point Patterns*. London: Academic Press.
- OKABE, A., BOOTS, B. & SUGIHARA, K. (1992). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Chichester, England: John Wiley & Sons.

Authors' rejoinder: We are extremely grateful to Mark Berman and Peter Diggle for their insightful commentaries on the paper, and delighted at their encouraging remarks on its usefulness for statistical practice.

Before responding we would like to acknowledge the work of Merlise Clyde and David Strauss (Clyde & Strauss, 1991) which has much in common with our own approach. We apologise for this omission in the paper. We also acknowledge Diggle's *et al.* (1994) priority for the 'edge-corrected pseudolikelihood' (Section 9.1.4).

Mark Berman gives an attractive argument for approximating the pseudolikelihood of a point process by the likelihood of a (Bernoulli) logistic model, rather than by a (Poisson) loglinear model as we have done. Indeed this is essentially the approach of Clyde & Strauss (1991): the window of observation is discretized into a fine array of cells, with binary responses for the presence or absence of a data point in each cell. The continuous pseudolikelihood is well approximated by the (rescaled) discrete pseudolikelihood (Besag *et al.*, 1982), which is formally identical to the likelihood of a logistic regression.

The main difference between our approach and the logistic regression strategy is that the latter requires a *fine* discretization of the window of observation, while our approach uses a coarse quadrature approximation. Indeed our own approach effectively degenerates to the logistic regression approach when a fine discretization is used.

In our experience, fine discretizations become computationally prohibitive as soon as the model contains several explanatory variables, and particularly when it is a generalized additive model. Our S-PLUS routines were unable to fit the stationary Strauss process using regular grids of more than about 5000 dummy points. This may well be a limitation of our current hardware or software, rather than an essential problem. In any case, computational load is an important practical consideration.

Fine discretization would remove the 'discontinuity errors' mentioned in our paper (Sections 4.3 and 5). On the other hand, it may render the parameter estimation and deviance differences susceptible to the Hauck–Donner effect (Hauck & Donner, 1977, 1980; Væth, 1985) which arises in logistic regression when the fitted probabilities are small.

Mark Berman points out that the computationally cheap 'counting weights' scheme (Section 4.3) yields approximately an unweighted Poisson likelihood. This is a nice point, and we agree it may well confer a slight computational advantage, when the conditional intensity is approximately flat.

Peter Diggle aptly describes our approach as a computationally effective model fitting procedure bought 'at a price'. Part of the price is the discontinuity error (17). We of course

believe that the price is worth paying. For techniques to be genuinely useful to the statistical community they must be amenable to implementation within readily available software. Still, it is of considerable interest to assess the magnitude of the discontinuity error (17) in practical cases. Consider the Swedish pines data (Section 10.1). For the Strauss process with parameters β and γ , a simple upper bound on (17) is $n\beta(1 - \gamma) \max_j w_j$. The following table shows the worst case bound (largest β , smallest γ) amongst the different parameter estimates for the Swedish pines data, using the ‘counting weights’. Also shown is the actual absolute discontinuity error.

Grid size	log PL (disc.)	log PL (no disc.)	Difference	Bound
50	-67.5021	-64.6743	2.8278	8.15
100	-62.1381	-61.4266	0.7115	2.04
200	-62.1594	-61.9815	0.1779	0.51

The reduction of grid size from 100×100 to 50×50 causes a difference of 3.2477 in the log pseudolikelihood computed without discontinuity error. This has the same order of magnitude as the discontinuity error of 2.8278 for the 50×50 grid size. Unfortunately the two errors have the same sign. It is more difficult to ascertain the impact of the discontinuity error upon the actual parameter estimates.

We are not sure how to respond to Peter Diggle’s comments on the ‘second disadvantage’ of our approach. We grant that successive approximations of the exact value of pseudolikelihood for a given value of θ could be computed by taking successively finer grids of quadrature points. However, re-computation of *all* the $\lambda_\theta(u_k; \mathbf{x})$ is required each time we update θ .

Peter Diggle’s illuminating and wise comments on the analysis of marked point processes are well taken. He identifies a trichotomy of possible goals for inference. We agree that the scientific context of the Harkness–Isham ant data is such that the analysis we have reported is the appropriate one. Readers seeking to apply the same methods to other marked point pattern datasets should take note of Peter Diggle’s comments. Other relevant discussion is in Goulard *et al.* (1996) and Särkkä & Hogmander (1998). This is clearly an area for further investigation.

Mark Berman asks us to comment on the possibility of creating a generic S-PLUS function for bootstrap inference for models with the intensity (8). This would presumably be a parametric bootstrap approach, involving the generation of simulated realizations from the fitted parametric model. The sticking point is the generation of such simulations for a virtually arbitrary point process model. Markov chain Monte Carlo (MCMC) methods make such simulations possible for specific models, but, as we have noted in the paper, the simulation algorithms are very specific to the model and require careful tuning and monitoring in execution. At the time of writing, most generic MCMC algorithms still require large ‘burn in’ times of the order of millions of iterations. Bootstrapping requires of the order of 500 such runs, which makes the whole procedure computationally expensive. Moreover it is completely unfeasible to conduct such simulations in interpreted languages such as the S-PLUS command language — they must be coded in a low-level language such as C or FORTRAN and loaded into the package. A user who wished to perform inference for a new model would need to write low-level code for the conditional intensity of the new model.

We are not convinced that the bootstrap is essential for correct inference and model selection. We believe that further work should be done on model selection using the pseudolikelihood ratio. Another interesting approach to model fitting, recently proposed by Huang and Ogata, is to compute the maximum pseudolikelihood estimate of the parameters, and then

to perform just a *single* Newton–Raphson step towards maximization of the likelihood itself. Typically the Newton–Raphson step requires MCMC estimation of the normalizing constant for the likelihood; in this context we only need to simulate from a single model, namely the maximum pseudolikelihood fit. Similar remarks apply to inference and model selection.

The algorithm we used to fit Ord’s model is capable of vast acceleration, along the lines suggested by Mark Berman. We did not have access at the time to a Dirichlet tessellation algorithm which allowed the *deletion* of points.

References

- BESAG, J., MILNE, R. & ZACHARY, S. (1982). Point process limits of lattice processes. *J. Appl. Probab.* **19**, 210–216.
- CLYDE, M. & STRAUSS, D. (1991). Logistic regression for spatial pair-potential models. In *Spatial Statistics and Imaging*, ed. A. Possolo, Vol. 20 of Lecture Notes — Monograph series, Chapter II, pp. 14–30. Institute of Mathematical Statistics.
- DIGGLE, P.J., FIKSEL, T., GRABARNIK, P., OGATA, Y., STOYAN, D. & TANEMURA, M. (1994). On parameter estimation for pairwise interaction. *Internat. Statist. Rev.* **62**, 99–117.
- GOULARD, M., SÄRKKÄ, A. & GRABARNIK, P. (1996). Parameter estimation for marked Gibbs point processes through the maximum pseudolikelihood method. *Scand. J. Statist.* **23**, 365–379.
- HAUCK, W.W., Jr & DONNER, A. (1977). Wald’s test as applied to hypotheses in logit analysis. *J. Amer. Statist. Assoc.* **72**, 851–853.
- HAUCK, W.W., Jr & DONNER, A. (1980). Corrigenda. *J. Amer. Statist. Assoc.* **75**, 482.
- SÄRKKÄ, A. & HÖGMANDER, H. (1998). Multitype spatial point patterns with hierarchical interactions. Research Report 1998:1. Göteborg University, Sweden, Department of Statistics.
- VÆTH, M. (1985). On the use of Wald’s test in exponential families. *Internat. Statist. Rev.* **53**, 199–214.