

Inhomogeneous Markov point processes by transformation

EVA B. VEDEL JENSEN* and LINDA STOUGAARD NIELSEN**

*Laboratory for Computational Stochastics, Department of Mathematical Sciences, University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark. E-mail: *eva@imf.au.dk;*

***lins@imf.au.dk*

We construct parametrized models for point processes, allowing for both inhomogeneity and interaction. The inhomogeneity is obtained by applying parametrized transformations to homogeneous Markov point processes. An interesting model class, which can be constructed by this transformation approach, is that of exponential inhomogeneous Markov point processes. Statistical inference for such processes is discussed in some detail.

Keywords: coarea formula; Hammersley–Clifford theorem; Hausdorff measure; inhomogeneity; interaction; manifolds; Markov chain Monte Carlo; Markov point processes; maximum likelihood estimation; relation; Strauss process; testing

1. Introduction

Models for spatial point processes, describing inhomogeneity as well as interaction between the points, have recently attracted considerable attention; see Baddeley and Turner (2000), Baddeley *et al.* (2000), Brix and Møller (1998) and Stoyan and Stoyan (1998). This appears to be a very natural step towards more realistic modelling, where both first- and second-order properties of the point pattern (such as the mean and variance in a univariate setting) are taken into account.

At least two of these model classes can be derived from homogeneous Markov point processes. (In the present paper, we will add another of this type.) According to the Hammersley–Clifford theorem (Ripley and Kelly 1977), such a process has a density with respect to a Poisson point process of the form

$$f(x) = \prod_{y \subseteq x} \varphi(y),$$

where $x = \{x_1, \dots, x_n\}$, $x_i \in \mathcal{S}$ and \mathcal{S} is a bounded Borel subset of \mathbb{R}^m , say. The integer n may be 0 to allow for $x = \emptyset$. The function φ is an interaction function, i.e. $\varphi(y) = 1$, unless all pairs of points in y are neighbours. The Markov point process is called homogeneous if φ is constant on all sets consisting of one point; see, for example, Ogata and Tanemura (1986) and Stoyan and Stoyan (1998).

The first type of models relates directly to the Hammersley–Clifford decomposition of the density and has been suggested by, among others, Ogata and Tanemura (1986). The idea is here to let the main effects in the decomposition (the interaction of one-point sets) be

non-constant. Under regularity conditions, an approximation to the likelihood function can be derived using methods from statistical physics. Stoyan and Stoyan (1998) have recently discussed this model in a forestry setting. See also Ripley (1990) and Møller *et al.* (1998).

Another type of models is constructed by using an independent thinning of a homogeneous Markov point process; see Baddeley *et al.* (1998). This is a well-known procedure for generating an inhomogeneous Poisson point process from a homogeneous one; see, for example, Stoyan *et al.* (1995). In Baddeley *et al.* (1998), semi-parametric inference of the thinned Markov point process is discussed.

In the present paper, yet another construction is suggested. The basic idea here is to introduce the inhomogeneity by applying a transformation to a homogeneous Markov point process. Inhomogeneous Poisson point processes as well as homogeneous Markov point processes can be included in such a model. The transformed point process is still a Markov point process with respect to the induced relation. This approach yields not only inhomogeneity in the intensity of the point process, but also inhomogeneity in the strengths of interactions among events. In particular, interactions are weaker among events in regions of high intensity than in regions of low intensity. This is appealing from an ecological perspective since competition among plants is weaker in regions where resources are abundant than in regions where resources are limited.

One of the useful properties of our model class is that the inhomogeneity and interaction can be separated. The statistical inference is based on the estimation of the transformation which can remove the inhomogeneity. After application of this transformation we are left with a homogeneous point pattern which can be analysed by known tools; see, for example, Geyer (1999) and Møller (1999).

The idea of modelling inhomogeneity by transformations has been applied in other areas of spatial statistics; for instance, for modelling the covariance structure of a non-stationary spatial process (Sampson and Guttorp 1992; Smith 1997). Related work can also be found in Monestiez *et al.* (1993), Meiring (1995), Perrin (1997) and references therein.

In Section 2, the basic concepts relating to Markov point processes are outlined. In Section 3, transformations of point processes are introduced and studied for Markov models. Parametrized transformations are considered in Section 4, resulting in models for point processes allowing for both inhomogeneity and interaction. An important particular case is that of the exponential inhomogeneous Markov point processes for which explicit expressions for the parametrized transformation can be found in the unit cube in \mathbb{R}^m and on the unit sphere in \mathbb{R}^3 . In Section 5, maximum likelihood estimation for the models described in Section 4 is discussed and the actual estimation procedure is applied to a simulated inhomogeneous point pattern on the unit sphere. In this section, tests for simple hypotheses are also derived. Section 6 discusses open questions and future work.

2. Markov point processes

In this section, we summarize some of the basic terminology for Markov point processes. A more detailed account of the notation and set-up can be found in Baddeley and Møller (1989) and Møller (1999).

Let $(\mathcal{S}, \mathcal{B}, \mu)$ be a measure space where $0 < \mu(\mathcal{S}) < \infty$ and \mathcal{B} is separable and contains all singletons. Let Ω be the set of finite subsets of \mathcal{S} , equipped with the σ -field \mathcal{F} , as defined in Møller (1999). Then, a finite point process X is a measurable mapping defined on some probability space and taking values in (Ω, \mathcal{F}) . In what follows, it will be assumed that X has a density f with respect to the Poisson point process on \mathcal{S} with intensity measure μ .

In order to define a Markov point process, we need a reflexive and symmetric relation \sim on \mathcal{S} . Two points $\xi, \eta \in \mathcal{S}$ are called neighbours if $\xi \sim \eta$. For $\eta \in \mathcal{S}$, the neighbourhood of η is the set

$$\partial\eta = \{\xi \in \mathcal{S} : \xi \sim \eta\}.$$

A finite subset x of \mathcal{S} is called a *clique* if all points in x are neighbours. A singleton is a clique by the requirement that \sim is reflexive. By convention, the empty set is a clique. The set of cliques is denoted \mathcal{C} .

A finite point process X is said to be a Markov point process if (Ripley and Kelly 1977):

(M1) $f(x) > 0 \Rightarrow f(y) > 0$ for all $y \subseteq x, x \in \Omega$;

(M2) if $f(x) > 0$, then

$$\lambda(\eta; x) = f(x \cup \{\eta\})/f(x), \quad \eta \in \mathcal{S}, x \in \Omega, \eta \notin x,$$

depends only on η and $\partial\eta \cap x$.

Note that $\lambda(\eta; x)$ can be regarded as the conditional ‘intensity’ of adding an extra point η to the point configuration x .

One way of introducing inhomogeneity into the model is to use a non-homogeneous intensity measure μ of the reference Poisson point process. This is equivalent to using a non-constant interaction function on singletons, but a homogeneous (Lebesgue, Hausdorff) intensity measure μ .

The Hammersley–Clifford theorem gives a factorization of a Markov density in terms of interactions which are only allowed between points in cliques.

Theorem 2.1 (Hammersley–Clifford). *A density f defines a Markov point process with respect to \sim if and only if there exists a function $\varphi : \Omega \rightarrow [0, \infty)$, such that $\varphi(y) \neq 1$ implies that $y \in \mathcal{C}$, and such that*

$$f(x) = \prod_{y \subseteq x} \varphi(y)$$

for all $x \in \Omega$. The function φ is called the clique interaction function.

In the present paper, a Markov point process is called inhomogeneous if φ is non-constant on sets consisting of one point. Other definitions of inhomogeneity are of course possible (Stoyan *et al.* 1995), but the definition given here suffices for the purposes of our studies.

One of the most well-known homogeneous Markov point processes is the Strauss process

(Strauss 1975). If we let $n(x)$ be the number of elements in x , this process is characterized by the clique interaction function

$$\varphi(x) = \begin{cases} \alpha & \text{if } n(x) = 0, \\ \beta & \text{if } n(x) = 1, \\ \gamma & \text{if } n(x) = 2, x \in \mathcal{C}, \\ 1 & \text{otherwise,} \end{cases}$$

such that

$$f(x) = \alpha\beta^{n(x)}\gamma^{s(x)}, \quad x \in \Omega,$$

where $s(x)$ is the number of neighbour pairs in x ,

$$s(x) = \sum_{z \subseteq x} 1[n(z) = 2, z \in \mathcal{C}].$$

Note that $\alpha = \alpha(\beta, \gamma)$ is a function of $\beta, \gamma > 0$. Usually, it is also assumed that $\gamma \leq 1$.

3. Transformations of point processes

In this section, attention is restricted to the case where \mathcal{S} is a k -dimensional differentiable manifold $\mathcal{B} \subseteq \mathbb{R}^m$ and μ is the k -dimensional Hausdorff measure λ_m^k in \mathbb{R}^m ; see, for example, Jensen (1998, Chapter 2) for a formal definition of Hausdorff measures. Intuitively, λ_m^k measures k -dimensional volume in \mathbb{R}^m . We will study smooth transformations of a point process X on \mathcal{B} . In Figure 1, an example of such a transformation is shown.

If it is important to emphasize the containing space \mathcal{B} , the set of finite subsets of \mathcal{B} is from now on denoted $\Omega_{\mathcal{B}}$ and the associated σ -field $\mathcal{F}_{\mathcal{B}}$. Likewise for other manifolds appearing below.

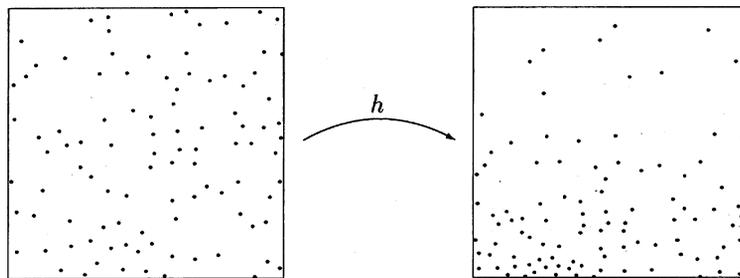


Figure 1. Adding inhomogeneity through transformation. The original point process is a conditional Strauss process on $[0, 1]^2$ with 100 points and $\gamma = 0.01$. Two points are neighbours if their mutual distance is less than $R = 0.05$.

The coarea formula gives a useful transformation result for a mapping between two manifolds; see Jensen (1998, Theorem 2.1).

Lemma 3.1 (Coarea formula). *Let $\mathcal{X} \subseteq \mathbb{R}^m$ and $\mathcal{Y} \subseteq \mathbb{R}^d$ be differentiable manifolds of dimension k . Let $h: \mathcal{X} \rightarrow \mathcal{Y}$ be a one-to-one differentiable mapping of \mathcal{X} onto \mathcal{Y} . Then, there exists a function $Jh: \mathcal{X} \rightarrow [0, \infty)$, called the Jacobian, such that for any non-negative measurable function g on \mathcal{X} ,*

$$\int_{\mathcal{X}} g(x)Jh(x) \, dx^k = \int_{\mathcal{Y}} g(h^{-1}(y)) \, dy^k,$$

where dx^k and dy^k are abbreviations for $\lambda_m^k(dx)$ and $\lambda_d^k(dy)$, respectively.

Below, the coarea formula is used to find the density of a transformed point process $h(X) = \{h(\xi): \xi \in X\}$.

Proposition 3.2. *Let \mathcal{X} , \mathcal{Y} and h be as in Lemma 3.1. Furthermore, suppose that X is a point process on \mathcal{X} with density f_X with respect to the Poisson point process on \mathcal{X} with intensity measure λ_m^k . Then, $h(X)$ is the point process on \mathcal{Y} with density, with respect to the Poisson point process on \mathcal{Y} with intensity measure λ_d^k , of the form*

$$f_{h(X)}(y) = f_X(h^{-1}(y))e^{\lambda_d^k(\mathcal{Y}) - \lambda_m^k(\mathcal{X})} \prod_{\eta \in y} Jh^{-1}(\eta), \quad y \in \Omega_{\mathcal{Y}}.$$

Proof. Let $F \in \mathcal{F}_{\mathcal{Y}}$. Using the well-known expansion of the distribution of the Poisson point process (see, for example, Møller 1999, Section 2), we obtain

$$\begin{aligned} P(h(X) \in F) &= \sum_{n=0}^{\infty} e^{-\lambda_m^k(\mathcal{X})} \frac{1}{n!} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} 1[\{h(x_1), \dots, h(x_n)\} \in F] \\ &\quad \times f_X(\{x_1, \dots, x_n\}) \, dx_1^k \cdots dx_n^k \\ &= \sum_{n=0}^{\infty} e^{-\lambda_m^k(\mathcal{X})} \frac{1}{n!} \int_{\mathcal{Y}} \cdots \int_{\mathcal{Y}} 1[\{y_1, \dots, y_n\} \in F] \\ &\quad \times f_X(\{h^{-1}(y_1), \dots, h^{-1}(y_n)\}) \prod_{i=1}^n Jh^{-1}(y_i) \, dy_1^k \cdots dy_n^k. \end{aligned}$$

In the latter expression, the coarea formula has been used on h^{-1} . For the term indexed by n , the formula has been used n times. The result now follows immediately. \square

Next, attention will be restricted to transformations of a point process X , which is Markov with respect to a relation \sim on \mathcal{X} . In the corollary below it is shown that the transformed process is again Markov with respect to the induced relation. As will be apparent later in this paper, it is very important from a technical point of view to use the

induced relation. It is also in many cases very natural, for instance in ecological applications, because the criterion for being neighbours in the transformed point pattern is stricter in regions where the transformation has attracted the points.

Corollary 3.3. *Let \mathcal{X} , \mathcal{Y} and h be as in Lemma 3.1. Furthermore, suppose that X is a Markov point process with respect to \sim such that*

$$f_X(x) = \prod_{y \subseteq x} \varphi(y), \quad x \in \Omega_{\mathcal{X}},$$

where φ is a clique interaction function. Then, $Y = h(X)$ is a Markov point process on \mathcal{Y} with respect to the induced relation \approx , defined for $\eta_1, \eta_2 \in \mathcal{Y}$ by

$$\eta_1 \approx \eta_2 \Leftrightarrow h^{-1}(\eta_1) \sim h^{-1}(\eta_2).$$

Furthermore, the density of Y is of the form

$$f_Y(y) = \prod_{z \subseteq y} \psi(z), \quad y \in \Omega_{\mathcal{Y}},$$

where ψ is the clique interaction function

$$\psi(z) = \begin{cases} \varphi(\emptyset) e^{\lambda_d^k(\mathcal{Y}) - \lambda_m^k(\mathcal{X})} & \text{if } n(z) = 0 \\ \varphi(h^{-1}(\eta)) J h^{-1}(\eta) & \text{if } n(z) = 1, z = \{\eta\}, \\ \varphi(h^{-1}(z)) & \text{otherwise.} \end{cases}$$

For the transformed point pattern it is worthwhile to notice that, except for the fact that all interactions are evaluated on the inversely transformed point pattern, only the main effects of the interaction function change. Furthermore, these values change with the same factor, as if a single point is transformed from one manifold to another.

Proof of Corollary 3.3. Most of the results of the corollary follow from Proposition 3.2. It only remains to verify that ψ is a clique interaction function with respect to \approx . Thus, let us suppose that $\psi(z) \neq 1$. We want to show that z is a clique with respect to \approx . Since every set with at most one point is a clique by convention, it suffices to consider the case where $n(z) \geq 2$. Then, $\psi(z) = \varphi(h^{-1}(z))$ and $h^{-1}(z)$ is a clique with respect to \sim . Accordingly, z is a clique with respect to \approx . \square

4. Exponential inhomogeneous Markov point processes

The transformation result from the previous section can be used to develop a new approach to inhomogeneity. Transforming homogeneous Markov point processes by a suitable bijective mapping, Markov point processes allowing for both interaction and inhomogeneity can be

constructed. In what follows, we will restrict attention to the case where the transformation h maps \mathcal{X} onto itself, that is, $\mathcal{Y} = \mathcal{X}$.

Let X be a Markov point process on \mathcal{X} with respect to a relation \sim and with clique interaction function φ . Furthermore, let $g_\theta : \mathcal{X} \rightarrow [0; \infty)$ be a parametrized model of the inhomogeneity where $\theta \in \Theta \subseteq \mathbb{R}^l$. Suppose that we can find, for each $\theta \in \Theta$, a differentiable one-to-one transformation h_θ of \mathcal{X} onto \mathcal{X} such that

$$Jh_\theta^{-1}(\eta) = g_\theta(\eta), \quad \eta \in \mathcal{X}. \tag{1}$$

Corollary 3.3 now gives that $Y = h_\theta(X)$ is a Markov point process on \mathcal{X} with respect to the induced relation \approx and with density

$$f_Y(y; \theta) = \prod_{\eta \in y} g_\theta(\eta) \prod_{z \subseteq h_\theta^{-1}(y)} \varphi(z), \quad y \in \Omega. \tag{2}$$

Notice that the inhomogeneity has been introduced by the transformation while the interaction has been inherited from the original homogeneous Markov point process. This has important consequences for statistical inference, as shown in the next section.

Using the coarea formula, it is easily seen that (1) implies that

$$\int_{\mathcal{X}} g_\theta(\eta) d\eta^k = \lambda_m^k(\mathcal{X}), \quad \text{for all } \theta \in \Theta. \tag{3}$$

Therefore, $\{g_\theta : \theta \in \Theta\}$ can be regarded as a parametrized class of densities on \mathcal{X} with respect to the uniform distribution on \mathcal{X} (density $d\eta^k / \lambda_m^k(\mathcal{X})$). A natural and useful choice is an exponential family model of the inhomogeneity

$$g_\theta(\eta) = \beta(\theta) e^{\theta \cdot \tau(\eta)}, \tag{4}$$

where \cdot indicates inner product in \mathbb{R}^l and $\tau : \mathcal{X} \rightarrow \mathbb{R}^l$. Note that (3) then implies that

$$\beta(\theta) = \lambda_m^k(\mathcal{X}) / \int_{\mathcal{X}} e^{\theta \cdot \tau(\eta)} d\eta^k.$$

A Markov point process with density (2) and g_θ given in (4) is called an *exponential inhomogeneous Markov point process*. Such a process has density

$$f_Y(y; \theta) = \beta(\theta)^{n(y)} e^{\theta \cdot t(y)} \prod_{z \subseteq h_\theta^{-1}(y)} \varphi(z), \quad y \in \Omega, \tag{5}$$

where $t(y) = \sum_{\eta \in y} \tau(\eta)$.

The problem left is to find a bijective mapping h_θ that has inverse Jacobian g_θ . This problem is equivalent to that of solving a differential equation, which is not always an easy task. However, considering simple but still flexible types of inhomogeneity, it is possible to solve the equation. In the following, two such examples will be studied. Both examples are exponential inhomogeneous Markov point processes.

Example 4.1 *The unit cube in \mathbb{R}^m .* Let $\mathcal{X} = [0, 1]^m$. Suppose that we are interested in adding independent exponential inhomogeneity on each axis. That is, for $\theta \in \mathbb{R}^m$, we

consider

$$g_\theta(\eta) = \beta(\theta)e^{\sum_{i=1}^m \theta_i \tau_i(\eta_i)}, \quad \eta \in [0, 1]^m,$$

where $\tau_i: \mathbb{R} \rightarrow \mathbb{R}$ and η_i is the i th coordinate of η . For this choice of g_θ , there is a unique solution h_θ to (1) among differentiable transformations of the form

$$h_\theta(\eta) = (h_{\theta 1}(\eta_1), \dots, h_{\theta m}(\eta_m)), \quad \eta \in [0, 1]^m,$$

where $h_{\theta i}$ is an increasing function of $[0, 1]$ onto itself. The uniqueness can be seen, using the fact that then h_θ^{-1} is of a similar form and

$$Jh_\theta^{-1}(\eta) = \prod_{i=1}^m \frac{\partial h_{\theta i}^{-1}}{\partial \eta_i}(\eta_i), \quad \eta \in [0, 1]^m.$$

For any $\theta \in \mathbb{R}^m$, the unique solution to (1) is given by

$$h_\theta^{-1}(\eta) = \left(\beta_1(\theta_1) \int_0^{\eta_1} e^{\theta_1 \tau_1(u)} du, \dots, \beta_m(\theta_m) \int_0^{\eta_m} e^{\theta_m \tau_m(u)} du \right), \quad \eta \in [0, 1]^m, \quad (6)$$

where

$$\beta_i(\theta_i) = \left(\int_0^1 e^{\theta_i \tau_i(u)} du \right)^{-1}, \quad i = 1, \dots, m,$$

and $\beta(\theta) = \prod_{i=1}^m \beta_i(\theta_i)$. Note that for $\theta = (0, \dots, 0)$, h_θ is the identity mapping.

In particular, if $\tau_i(u) = u$ for all i , the integrals in (6) can be calculated explicitly and

$$h_\theta(\eta) = \left(\frac{1}{\theta_1} \log(1 + (e^{\theta_1} - 1)\eta_1), \dots, \frac{1}{\theta_m} \log(1 + (e^{\theta_m} - 1)\eta_m) \right), \quad \eta \in [0, 1]^m. \quad (7)$$

In Figure 2, some simulated realizations of the exponential inhomogeneous Strauss process are shown.

Example 4.2 *The unit sphere in \mathbb{R}^3 .* Let $\mathcal{X} = S^2$, the unit sphere in \mathbb{R}^3 , and let us consider exponential inhomogeneity which depends on $m \cdot \eta$, where $m \in S^2$ is fixed. So the aim is for $\theta \in \mathbb{R}$ to find a differentiable one-to-one mapping h_θ of S^2 onto S^2 with inverse Jacobian

$$Jh_\theta^{-1}(\eta) = \beta(\theta)e^{\theta \tau(m \cdot \eta)}, \quad \eta \in S^2, \quad (8)$$

where $\tau: \mathbb{R} \rightarrow \mathbb{R}$. Note that for $\tau(u) = u$, (8) is the density of the Fisher distribution in directional statistics, whereas for $\tau(u) = u^2$, (8) is the density of the Dimroth–Watson distribution (Mardia 1972).

Let us choose a coordinate system in \mathbb{R}^3 such that $m = (0, 0, 1)$. In the appendix, it is shown that there is a unique solution to (8) among differentiable one-to-one mappings h_θ of S^2 onto S^2 of the form

$$h_\theta(\eta_1, \eta_2, \eta_3) = \left(\frac{\sqrt{1 - r_\theta(\eta_3)^2}}{\sqrt{1 - \eta_3^2}} \eta_1, \frac{\sqrt{1 - r_\theta(\eta_3)^2}}{\sqrt{1 - \eta_3^2}} \eta_2, r_\theta(\eta_3) \right), \quad (9)$$

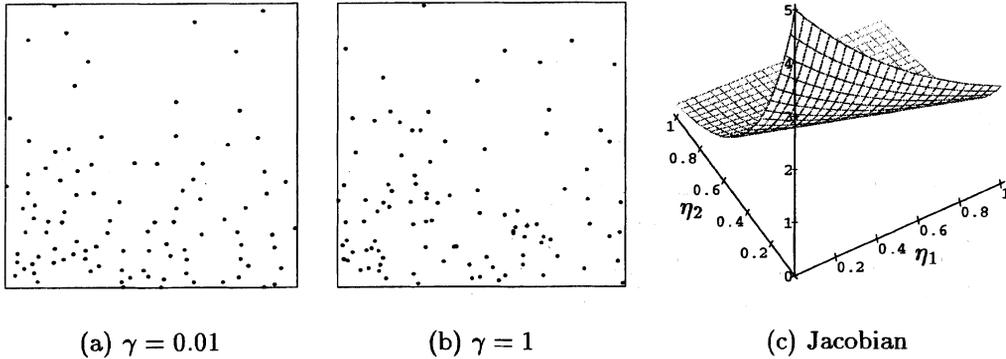


Figure 2. Conditional simulations of the exponential inhomogeneous Strauss process on $[0, 1]^2$ with 100 points, $R = 0.05$ and γ as indicated. The inhomogeneity is introduced by the transformation (7) with $m = 2$ and $(\theta_1, \theta_2) = (-1, -3)$. The Jacobian of the inverse transformation is shown on the right. Another realization of the process with parameters as in (a) is shown on the right in Figure 1.

where r_θ is an increasing differentiable bijection on $[-1, 1]$. Note that such a transformation only changes the angle between $\eta = (\eta_1, \eta_2, \eta_3)$ and $m = (0, 0, 1)$.

The solution is most easily expressed in terms of the inverse. We find (see the Appendix), that the unique solution to (8) is given by

$$h_\theta^{-1}(\eta_1, \eta_2, \eta_3) = \left(\frac{\sqrt{1 - g_\theta(\eta_3)^2}}{\sqrt{1 - \eta_3^2}} \eta_1, \frac{\sqrt{1 - g_\theta(\eta_3)^2}}{\sqrt{1 - \eta_3^2}} \eta_2, g_\theta(\eta_3) \right),$$

where

$$g_\theta(u) = 1 - \int_u^1 \beta(\theta) e^{\theta \tau(v)} dv, \quad -1 \leq u \leq 1,$$

and

$$\beta(\theta) = \frac{2}{\int_{-1}^1 e^{\theta \tau(v)} dv}.$$

Note that, for $\theta = 0$, h_θ is the identity mapping.

For τ equal to the identity mapping,

$$g_\theta(u) = 1 + \frac{\beta(\theta)}{\theta} (e^{\theta u} - e^\theta), \quad -1 \leq u \leq 1,$$

and

$$\beta(\theta) = \frac{2\theta}{e^\theta - e^{-\theta}}.$$

Hence,

$$r_\theta(u) = \frac{1}{\theta} \log \left(\frac{u-1}{2} (e^\theta - e^{-\theta}) + e^\theta \right), \quad -1 \leq u \leq 1. \quad (10)$$

In Figure 3, conditional simulations of the exponential inhomogeneous Strauss process are shown with $\tau(u) = u$ and $m = (0, 0, 1)$. The relation for the untransformed process is given by

$$\eta \sim \xi \Leftrightarrow d_3(\eta, \xi) < R, \quad (11)$$

where d_3 is the spatial distance. Note that the spatial distance is proportional to the geodesic distance.

In Figure 4, a realization of the continuum random cluster process with tendency to cluster (Møller 1999) is transformed with three different transformation parameters. The original untransformed process is shown in Figure 4(a). The relation is the one from (11). \square

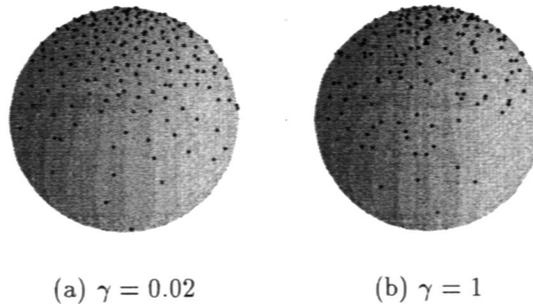


Figure 3. Conditional simulations of the exponential inhomogeneous Strauss process on S^2 with 200 points, $R = 0.02$ and γ as indicated. The inhomogeneity is introduced by the transformation (9) with $\theta = 3$ and r_θ given in (10).

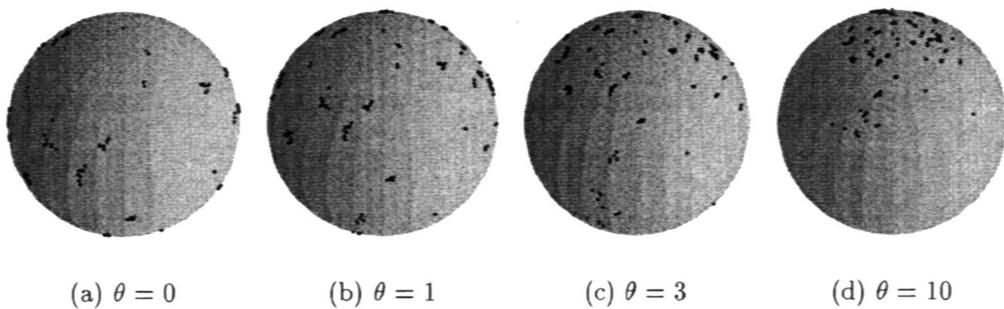


Figure 4. Conditional simulations of the exponential inhomogeneous continuum random cluster process on S^2 with 200 points, $R = 0.05$ and $\gamma = 1000$. As in Figure 3, the inhomogeneity is introduced by the transformation (9), here with θ as indicated.

5. Statistical inference for exponential inhomogeneous Markov point processes

An exponential inhomogeneous Markov point process has a density of the form (5). Note that the Jacobian part

$$\beta(\theta)^{n(y)} e^{\theta \cdot t(y)}$$

of (5) is the density of $h_\theta(X)$ when X is a homogeneous Poisson point process on \mathcal{X} with intensity measure λ_m^k .

We will mainly discuss statistical inference conditional on $n(Y) = n$, the observed number of points. Using the fact that h_θ is one-to-one so that $n(Y) = n(h_\theta(X)) = n(X)$, it is easy to see that the conditional density of Y given $n(Y) = n$ is of the form

$$f_n(y; \theta) = \beta(\theta)^n e^{\theta \cdot t(y)} \prod_{z \subseteq h_\theta^{-1}(y)} \psi(z), \quad n(y) = n,$$

where

$$\psi(z) = \begin{cases} \varphi(z) & \text{if } n(z) > 0, \\ \frac{\varphi(\emptyset)}{\mathbb{P}(n(X) = n)} & \text{if } n(z) = 0. \end{cases}$$

Let us suppose that the interaction function ψ can be parametrized by some parameter $\gamma \in \Gamma \subseteq \mathbb{R}^p$. Let $L(\theta, \gamma; y)$ be the conditional likelihood function based on the inhomogeneous data y . Furthermore, let

$$L_0(\theta; y) = \beta(\theta)^n e^{\theta \cdot t(y)}$$

be the conditional likelihood function of θ based on y , when disregarding the interaction, and let

$$L_1(\gamma; x) = \prod_{z \subseteq x} \psi(z; \gamma)$$

be the likelihood of γ , when observing x in the homogeneous model. Then,

$$L(\theta, \gamma; y) = L_0(\theta; y) L_1(\gamma; h_\theta^{-1}(y)).$$

This decomposition of the likelihood function has important consequences for the statistical inference. In particular, for each fixed θ , the maximum of L with respect to γ can be found, using an algorithm developed for the homogeneous case with data $h_\theta^{-1}(y)$. Further results concerning estimation can be obtained if we choose a specific model for the interaction.

Let us assume that

$$L_1(\gamma; x) = \alpha_n(\gamma) \gamma^{u(x)}, \quad n(x) = n, \gamma > 0, \tag{12}$$

such that the interaction model is a regular exponential family of order 1 (Barndorff-Nielsen 1978). There are three interesting special cases of this model:

- Strauss model (Strauss 1975),

$$u(x) = s(x) = \sum_{z \subseteq x} 1(n(z) = 2, z \in \mathcal{C}).$$

- Continuum random cluster model (Møller 1999),

$$u(x) = -c(x),$$

where $c(x)$ is the number of path-connected components in x .

- Area-interaction model (Baddeley and van Lieshout 1995),

$$u(x) = -\lambda_m(\cup_{\eta \in x} B(\eta, R)),$$

where $B(\eta, R) \subseteq \mathbb{R}^m$ is a ball with centre η and radius R , and $\lambda_m = \lambda_m^m$ is the Lebesgue measure in \mathbb{R}^m .

The reason for using a minus in the two latter models is that then the interaction parameter γ has the same qualitative interpretation in all three models: $\gamma < 1$ corresponds to inhibition, $\gamma = 1$ to independence and $\gamma > 1$ to clustering.

The continuum random cluster model is not a Markov model in the sense of Ripley and Kelly (1977), but a nearest-neighbour Markov model (Baddeley and Møller 1989), but this distinction is not important here when discussing likelihood inference.

If a maximum likelihood estimate $(\hat{\theta}, \hat{\gamma})$ exists of (θ, γ) under the interaction model (12), then $u(h_{\hat{\theta}}^{-1}(y)) \in \text{int } C$, where C is the convex support of $u(X)$; see Barndorff-Nielsen (1978, p. 151). It therefore suffices to restrict attention to transformations in

$$\Theta^* = \{\theta \in \Theta : u(h_{\theta}^{-1}(y)) \in \text{int } C\}.$$

For $\theta \in \Theta^*$, there is a unique $\gamma = \gamma(\theta)$ for which $L(\theta, \cdot; y)$ attains its maximum, viz. the unique solution of

$$E_{\gamma(\theta)} u(X) = u(h_{\theta}^{-1}(y));$$

see Barndorff-Nielsen (1978, p. 152). This solution can be found using Markov chain Monte Carlo (MCMC) simulations; see, for example, Geyer (1999) and Møller (1999). An example is given later in this section.

The next step is to evaluate the partially maximized likelihood function for $\theta \in \Theta^*$,

$$\bar{L}(\theta; y) = \max_{\gamma > 0} L(\theta, \gamma; y) = L_0(\theta; y) \alpha_n(\gamma(\theta)) \gamma(\theta)^{u(h_{\theta}^{-1}(y))}. \quad (13)$$

This step also requires MCMC simulation since the normalizing constant $\alpha_n(\gamma(\theta))$ is not known explicitly. In order to obtain a stable calculation, it is very important to evaluate a ratio of likelihoods instead of a likelihood directly; see, for example, Geyer (1999). Typically, we want to determine, up to a constant, $\bar{L}(\theta; y)$ on a grid of θ values. For this purpose, it suffices, since $L_0(\theta; y)$ can be calculated directly, to calculate for pairs of neighbour grid points $\theta, \tilde{\theta} \in \Theta^*$,

$$\log \left(\frac{\bar{L}(\theta; y)/L_0(\theta; y)}{\bar{L}(\tilde{\theta}; y)/L_0(\tilde{\theta}; y)} \right) = (u(h_{\theta}^{-1}(y)) - u(h_{\tilde{\theta}}^{-1}(y))) \log \gamma(\tilde{\theta}) + \log E_{\gamma(\theta)} \left(\frac{\gamma(\tilde{\theta})}{\gamma(\theta)} \right)^{u(X) - u(h_{\tilde{\theta}}^{-1}(y))} . \tag{14}$$

The mean value on the right-hand side of this formula can be evaluated using MCMC simulations.

Being able to calculate the partially maximized likelihood function \bar{L} , a maximum of \bar{L} can be searched for. Note that if $\bar{L}(\cdot; y)$ is maximal at $\tilde{\theta}$, then $L(\cdot, \cdot; y)$ is maximal at $(\tilde{\theta}, \gamma(\tilde{\theta}))$.

For the Strauss model and the continuum random cluster model, $u(X)$ is discrete and the function $\theta \rightarrow \gamma(\theta)$ is actually a step function. Let the support of $u(X)$ be

$$S = \{u_-, u_- + 1, \dots, u_+\}$$

such that $\text{int } C = \text{int}[u_-, u_+] = (u_-, u_+)$, and let, for $i \in S$,

$$\Theta_i = \{\theta \in \Theta : u(h_{\theta}^{-1}(y)) = i\}.$$

Then,

$$\Theta^* = \bigcup_{i=u_-+1}^{u_+-1} \Theta_i,$$

and for $\theta \in \Theta_i$, $\gamma(\theta) = \gamma_i$, say. The partially maximized likelihood function becomes

$$\bar{L}(\theta; y) = L_0(\theta; y) \alpha_n(\gamma_i) \gamma_i^i, \quad \theta \in \Theta_i,$$

$i = u_- + 1, \dots, u_+ - 1$. Accordingly, in the subregion Θ_i , $\bar{L}(\theta; y)$ will be a rescaling of $L_0(\theta; y)$ with factor $\alpha_n(\gamma_i) \gamma_i^i$. Therefore, $\bar{L}(\cdot; y)$ is not continuous at $\theta \in \partial\Theta_i \cap \partial\Theta_{i'}$, $i \neq i'$, and traditional iterative procedures such as Newton–Raphson do not seem to be appropriate for seeking a maximum of \bar{L} . Instead, tabulating \bar{L} in a reduced parameter set $\Theta_{\text{red}} \subseteq \Theta^*$, is a better idea, when $u(X)$ is discrete. In order to be able to disregard parameter values outside Θ_{red} , this reduced set should have the property that, for any $\theta \in \Theta^* \setminus \Theta_{\text{red}}$, there exists $\theta' \in \Theta_{\text{red}}$ such that

$$\bar{L}(\theta; y) \leq \bar{L}(\theta'; y).$$

The first procedure for reducing the parameter set is based on the proportionality of \bar{L} and L_0 in the subregions Θ_i . Note that Θ_i does not need to be connected. Since $L_0(\cdot; y)$ is the likelihood function for an exponential family model, it is log-concave and thereby unimodal. Let $\hat{\theta}_0$ be the maximum likelihood estimate of θ based on $L_0(\cdot; y)$. (If L_0 is the likelihood for a regular exponential family, then $\hat{\theta}_0$ exists and is unique if and only if $t(y)$ lies in the interior of its convex support.) If θ is one-dimensional, all points in Θ_i can be excluded except for points in the two sets

$$\Theta_{i-} = \{\max\{\theta \in \Theta_i : \theta \leq \hat{\theta}_0\}\}$$

and

$$\Theta_{i+} = \{\min\{\theta \in \Theta_i : \theta \geq \hat{\theta}_0\}\}.$$

Each of these two sets consists of at most one point. See also the illustration in Figure 5, where $\Theta_{i-} = \{\theta_{i-}\}$ and $\Theta_{i+} = \{\theta_{i+}\}$. So any set including

$$\bigcup_{i=u_-+1}^{u_+-1} \Theta_{i-} \cup \Theta_{i+}$$

can be used as Θ_{red} . Note that Θ_{i-} and Θ_{i+} can be found by tabulating the function

$$\theta \rightarrow u(h_\theta^{-1}(y))$$

and using the determined value of $\hat{\theta}_0$.

Similar procedures must be possible if the dimension of θ is larger than 1.

The second procedure for reducing the parameter set requires that the only interaction considered is inhibition, i.e. $\gamma \leq 1$. Such a restriction may be quite natural, however, since we expect that it is going to be difficult to distinguish between inhomogeneity and clustering. For $\theta \in \Theta^*$, let $\tilde{\gamma}(\theta)$ be the unique $\gamma \in (0, 1]$ maximizing $L(\theta, \cdot; y)$. Note that

$$\tilde{\gamma}(\theta) = \begin{cases} \gamma(\theta) & \gamma(\theta) \leq 1, \\ 1 & \gamma(\theta) > 1. \end{cases}$$

Furthermore, let $\tilde{L}(\theta; y) = \max_{\gamma \in (0,1]} L(\theta, \gamma; y)$. Then we have the following proposition.

Proposition 5.1. *Let $\hat{\theta}_0$ be the maximum likelihood estimate of θ based on L_0 . Suppose that $\hat{\theta}_0 \in \Theta^*$. Then, for any $\theta \in \Theta^*$,*

$$u(h_\theta^{-1}(y)) \geq u(h_{\hat{\theta}_0}^{-1}(y)) \Rightarrow \tilde{L}(\theta; y) \leq \tilde{L}(\hat{\theta}_0; y).$$

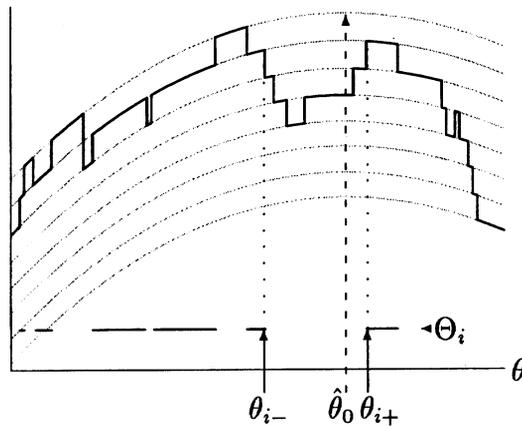


Figure 5. The partial log-likelihood function $\log \bar{L}$. The parallel curves are translations of $\log L_0(\theta; y)$. The intervals marked horizontally constitute Θ_i for the value of i corresponding to the second curve from the top.

Proof. Since $\hat{\theta}_0$ is the maximum likelihood estimate of θ based on $L_0(\cdot; y)$,

$$\beta(\theta)^n e^{\theta \cdot t(y)} \leq \beta(\hat{\theta}_0)^n e^{\hat{\theta}_0 \cdot t(y)}.$$

Therefore, since $\gamma^{u(h_{\hat{\theta}_0}^{-1}(y))} \leq \gamma^{u(h_{\theta}^{-1}(y))}$ for $\gamma \leq 1$,

$$L(\theta, \gamma; y) \leq L(\hat{\theta}_0, \gamma; y), \quad \gamma \leq 1,$$

and accordingly, the corresponding relation holds for the partially maximized likelihood function \tilde{L} . \square

According to Proposition 5.1, when seeking a maximum of \tilde{L} , it is enough to search in

$$\{\theta \in \Theta^* : u(h_{\theta}^{-1}(y)) \leq u(h_{\hat{\theta}_0}^{-1}(y))\}.$$

As will be seen in the example below, this may result in a drastic reduction in the number of θ values at which \tilde{L} has to be evaluated.

Example 5.2 *An application of the estimation procedure.* In this example we will show how the estimation procedure can be carried out, using a simulated point pattern y which is a realization of an exponential inhomogeneous Strauss process on the unit sphere; see the right-hand part of Figure 6. The density is given by

$$f_n(y) = \left(\frac{2\theta}{e^\theta - e^{-\theta}} \right)^n e^{\theta \sum_{i=1}^n y_{i3}} \alpha_n(\gamma) \gamma^{s(h_{\theta}^{-1}(y))}, \tag{15}$$

where $y = \{y_1, \dots, y_n\}$ and $y_i = (y_{i1}, y_{i2}, y_{i3})$, $i = 1, \dots, n$. The transformation h_{θ} is given in (9) and (10) and the relation \sim is given in (11).

The aim is to estimate (θ, γ) on the basis of the inhomogeneous data set y . We will assume that $\gamma \leq 1$. As suggested earlier in this section, the estimation is based on a tabulation of the partially maximized likelihood function $\tilde{L}(\theta; y)$ in a reduced region Θ_{red} which is known to contain $\hat{\theta}$.

In order to construct Θ_{red} , we first determine $\hat{\theta}_0$. For the example, $\hat{\theta}_0 = 5.24$ whereas the true value is $\theta = 5$. Secondly, the mapping $\theta \rightarrow s(h_{\theta}^{-1}(y))$ is tabulated (Figure 7), and using this information we find

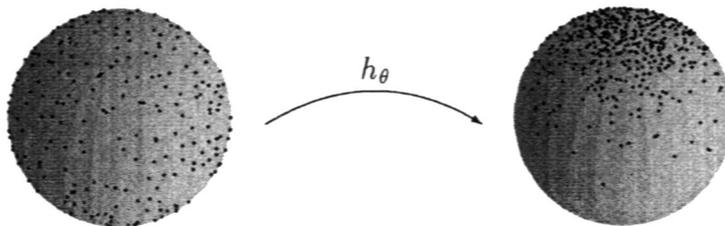


Figure 6. Transformation into the particular realization of the exponential inhomogeneous Strauss process studied in Example 5.2. Here $n = 400$, $\gamma = 0.5$, $R = 0.1$ and $\theta = 5$.

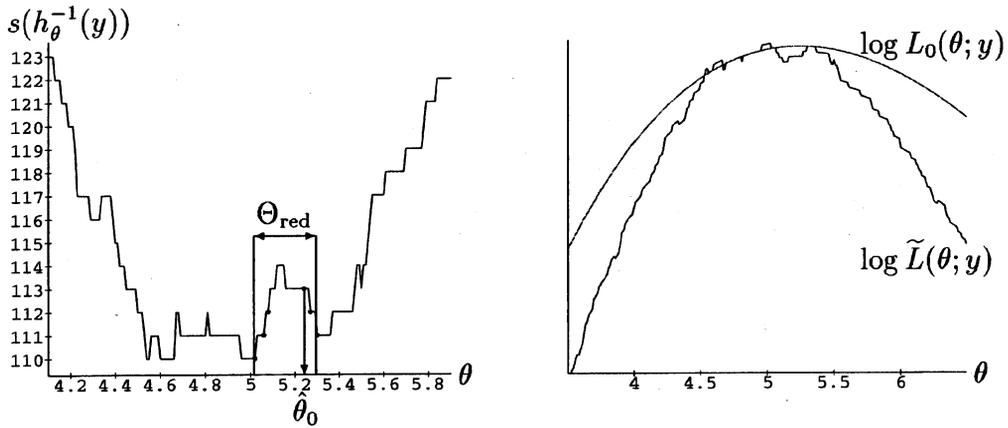


Figure 7. The number of neighbours in the inversely transformed process and the likelihood functions with and without the information about the interaction.

$$\{\theta : s(h_\theta^{-1}(y)) \leq s(h_{\hat{\theta}_0}^{-1}(y))\} = \{\theta : s(h_\theta^{-1}(y)) \in \{110, 111, 112, 113\}\} \subseteq [4.45, 5.50].$$

Then, according to Proposition 5.1, $\hat{\theta} \in [4.45, 5.50]$. Furthermore, we know that θ values with the same value of $s(h_\theta^{-1}(y))$ lie in the same Θ_i region and here we can restrict attention to θ values closest to $\hat{\theta}_0$. In the example, this means that $\hat{\theta} \in \{5.02, 5.06, 5.08, 5.24, 5.27, 5.30\}$; see Figure 7. We can take any set containing these 6 values as Θ_{red} . Since $\tilde{L}(\theta; y)$ is evaluated by calculating ratios at pairs of close θ values – see (14) with \bar{L} replaced by \tilde{L} and $\gamma(\theta)$ by $\tilde{\gamma}(\theta)$ – we take

$$\Theta_{red} = \{5.02, 5.03, \dots, 5.30\}.$$

The next step is to determine $\theta \rightarrow \gamma(\theta)$. Using MCMC simulations, the mapping

$$\gamma \rightarrow E_\gamma s(X) \tag{16}$$

is tabulated on a coarse grid of γ values in $(0, 1]$. The mean value in (16) is calculated in the homogeneous model. The function (16) is tabulated once more on a finer reduced grid, $\Gamma_{red} \subseteq (0, 1]$, of γ values. This reduced set is chosen such that

$$\{s(h_\theta^{-1}(y)) : \theta \in \Theta_{red}\} \subseteq \{E_\gamma(s(X)) : \gamma \in \Gamma_{red}\},$$

as we are only interested in these particular values of neighbour pairs.

Regression analysis gives the relation

$$\log(E_\gamma(s(X))) = \alpha + \beta \log(\gamma), \quad \gamma \in \Gamma_{red},$$

and, replacing $E_\gamma(s(X))$ with $s(h_\theta^{-1}(y))$, we obtain the approximation

$$\tilde{\gamma}(\theta) = \exp\left(\frac{\log(s(h_\theta^{-1}(y))) - \alpha}{\beta}\right), \quad \theta \in \Theta_{red}.$$

In the example studied, $\alpha = 5.254$ and $\beta = 0.751$. Note that $\gamma(\theta)$ only depends on $s(h_\theta^{-1}(y))$ which is discrete and constant in Θ_i regions. Therefore, these two functions jump at the same time.

Now we are ready to find the partially maximized likelihood function. If we let

$$\Theta_{\text{red}} = \{\theta_0, \theta_1, \dots, \theta_d\},$$

$\tilde{L}(\theta; y)$, $\theta \in \Theta_{\text{red}}$, can be determined up to a constant by using the recursive formula

$$\log \left(\frac{\tilde{L}(\theta_l; y)/L_0(\theta_l; y)}{\tilde{L}(\theta_0; y)/L_0(\theta_0; y)} \right) = \sum_{j=1}^l \log \left(\frac{\tilde{L}(\theta_j; y)/L_0(\theta_j; y)}{\tilde{L}(\theta_{j-1}; y)/L_0(\theta_{j-1}; y)} \right),$$

for $l = 1, \dots, d$. The terms in the sum can be found by MCMC simulation using (14). Denoting the above expression by Δ_l , we have that

$$\log \tilde{L}(\theta_l; y) = \log L_0(\theta_l; y) + \Delta_l + \text{constant}.$$

Note that Δ_l is constant on Θ_i .

In Figure 7, $\log \tilde{L}(\theta; y)$ is shown for a larger range of θ values than Θ_{red} , in order to form a general impression of the function. Note that this function is in fact the same as the one shown in Figure 5. We find $(\hat{\theta}, \hat{\gamma}) = (\hat{\theta}, \gamma(\hat{\theta})) = (5.02, 0.48)$.

Let us finally discuss two types of tests. Let us first consider the hypothesis of no interaction, $H_0: \gamma = \gamma_0$, where $\gamma_0 = 1$. Since $L_1(\gamma_0; \cdot) \equiv 1$, we find

$$Q = \frac{L(\hat{\theta}_0, \gamma_0; y)}{L(\hat{\theta}, \hat{\gamma}; y)} = \frac{L_0(\hat{\theta}_0; y) L_1(\gamma_0; h_{\hat{\theta}_0}^{-1}(y))}{L_0(\hat{\theta}; y) L_1(\hat{\gamma}; h_{\hat{\theta}}^{-1}(y))} = Q_0 \cdot Q_1,$$

where Q_0 represents a comparison between $\hat{\theta}_0$ and $\hat{\theta}$, while Q_1 is a test for $\gamma = \gamma_0$ on the basis of $x = h_{\hat{\theta}}^{-1}(y)$. Note that

$$Q_1 = \left[E_{\hat{\gamma}} \left(\frac{\gamma_0}{\hat{\gamma}} \right)^{u(X) - u(x)} \right]^{-1}$$

and can be calculated, using MCMC simulation.

It is also of interest to test for homogeneity. Let us suppose that this corresponds to $H_0: \theta = \theta_0$, where $\theta_0 = 0$ and h_{θ_0} is the identity. Furthermore, let $\hat{\gamma}_0$ be the estimate of γ under homogeneity. Then,

$$\begin{aligned} Q &= \frac{L(\theta_0, \hat{\gamma}_0; y)}{L(\hat{\theta}, \hat{\gamma}; y)} = \frac{L_0(\theta_0; y) L_1(\hat{\gamma}_0; y)}{L_0(\hat{\theta}; y) L_1(\hat{\gamma}; h_{\hat{\theta}}^{-1}(y))} \\ &= \frac{L_0(\theta_0; y) L_1(\hat{\gamma}_0; h_{\hat{\theta}}^{-1}(y))}{L_0(\hat{\theta}; y) L_1(\hat{\gamma}; h_{\hat{\theta}}^{-1}(y))} \frac{L_1(\hat{\gamma}_0; y)}{L_1(\hat{\gamma}_0; h_{\hat{\theta}}^{-1}(y))} \\ &= \tilde{Q}_0 \cdot \tilde{Q}_1 \cdot \hat{\gamma}_0^{u(y) - u(h_{\hat{\theta}}^{-1}(y))}, \end{aligned}$$

say. The intermediate ratio \tilde{Q}_1 can be calculated using MCMC.

Large values of the test statistics Q are critical. Since their distributions are not known under the respective null hypotheses, simulations are needed in order to evaluate the observed values.

6. Remarks, open questions and related work

The basic idea of the present paper is that of introducing inhomogeneity by transformation. Observing an inhomogeneous point pattern, the problem is then to construct the inverse transformation which can compensate for the inhomogeneity. Similar approaches can be found in a number of related areas. In addition to the examples presented in the introduction, one could mention that Baddeley and van Lieshout (1995, p. 605) discuss the possibility of letting the balls, appearing in the area-interaction model, depend on a parameter θ .

Note that our model may be extended to consider densities of the form

$$f_Y(\{y_1, \dots, y_n\}) \propto \prod_i \psi(h_{\theta_1}^{-1}(y_i)) \prod_{i < j} \varphi(h_{\theta_2}^{-1}(y_i, y_j)) \dots$$

(Jesper Møller, personal communication). The approach by Ogata and Tanemura (1986) is then obtained by letting h_{θ_2} be the identity, while our transformation approach corresponds to $h_{\theta_1} = h_{\theta_2}$ and

$$\psi(z) = \varphi(z)/Jh(z).$$

Note also that the model suggested by Ogata and Tanemura (1986) can be used to describe a system of hard discs of fixed diameters with location-dependent intensity. Such a system cannot be described by the transformation set-up discussed in the present paper.

The differential equation (1) makes some restrictions on the kind of inhomogeneity which can be described by the transformation approach. It will be of interest to characterize the class of point processes that can be described by this approach.

From a practical point of view, it is going to be important to use concomitant environmental variables to explain the inhomogeneity. In fact, the exponential inhomogeneity as described by (4) is identical to that considered in Rathbun (1996) if

$$\tau(\eta) = (\tau_1(\eta), \dots, \tau_l(\eta))$$

is a vector of explanatory variables evaluated at η . Rathbun (1996) considers, however, only the Poisson case. Our future plans include the analysis of concrete data sets of this type where the interaction is also taken into account.

The estimation procedure developed in Section 5 worked surprisingly well on the simulated example in Example 5.2. The main reason was that the mapping $\theta \rightarrow s(h_{\theta}^{-1}(y))$ was first essentially decreasing and then essentially increasing with a minimum near $\hat{\theta}_0$. Whether this is true more generally needs to be investigated.

Maximum likelihood estimation is somewhat involved and it is therefore of interest to investigate alternative procedures such as pseudolikelihood estimation. See Baddeley and Turner (2000) for its implementation in cases where both inhomogeneity and interaction are present. An even simpler procedure would be to use $\hat{\theta}_0$ as the estimate of the

inhomogeneity parameter. In the example, this appeared to work well. Such a procedure would be justified if the distribution of $t(Y)$ does not depend very much on θ . If so, the extensive work on estimating intensity functions in inhomogeneous Poisson models could then also be applied. These estimation problems are currently under investigation (Nielsen *et al.* 2000).

On the theoretical side, conditions still need to be found which ensure the existence and uniqueness of the maximum likelihood estimates, an asymptotic distribution theory needs to be developed for the maximum likelihood estimators as well as for the likelihood ratio tests.

Appendix

In this appendix, we show that, for $\theta \in \mathbb{R}$,

$$Jh_\theta^{-1}(\eta) = \beta(\theta)e^{\theta\tau(m\eta)}, \quad \eta \in S^2, \tag{A.1}$$

has a unique solution among mappings of the form

$$h_\theta^{-1}(\eta_1, \eta_2, \eta_3) = \left(\frac{\sqrt{1 - g_\theta(\eta_3)^2}}{\sqrt{1 - \eta_3^2}} \eta_1, \frac{\sqrt{1 - g_\theta(\eta_3)^2}}{\sqrt{1 - \eta_3^2}} \eta_2, g_\theta(\eta_3) \right), \tag{A.2}$$

where g_θ is an increasing differentiable function from $[-1, 1]$ onto $[-1, 1]$. This result is used in Example 4.2.

Let

$$p: [0, \pi) \times [0, 2\pi) \mapsto S^2$$

$$(\omega_1, \omega_2) \rightarrow (\sin(\omega_1)\cos(\omega_2), \sin(\omega_1)\sin(\omega_2), \cos(\omega_1))$$

be the polar coordinate mapping. Functions of the form (A.2) can then equivalently be described as

$$h_\theta^{-1} = p \circ k_\theta \circ p^{-1}, \tag{A.3}$$

where $k_\theta(\omega_1, \omega_2) = (l_\theta(\omega_1), \omega_2)$ and l_θ is an increasing differentiable bijection on the interval $[0, \pi)$. Here \circ denotes the composition of mappings.

Below, we show the result

$$Jh_\theta^{-1}(p(\omega_1, \omega_2)) = Jk_\theta(\omega_1, \omega_2) \frac{\sin l_\theta(\omega_1)}{\sin \omega_1} = l'_\theta(\omega_1) \frac{\sin l_\theta(\omega_1)}{\sin \omega_1}. \tag{A.4}$$

In order to prove the first equality, we use the coarea formula and obtain, for an arbitrary function f on S^2 , that

$$\begin{aligned}
\int_{S^2} f(h_\theta(\eta)) \, d\eta^2 &= \int_{S^2} f(p \circ k_\theta^{-1} \circ p^{-1}(\eta)) \, d\eta^2 \\
&= \int_0^\pi \int_0^{2\pi} f(p \circ k_\theta^{-1}(\omega_1, \omega_2)) \sin \omega_1 \, d\omega_2 \, d\omega_1 \\
&= \int_0^\pi \int_0^{2\pi} f(p(\omega_1, \omega_2)) Jk_\theta(\omega_1, \omega_2) \sin(l_\theta(\omega_1)) \, d\omega_2 \, d\omega_1 \\
&= \int_{S^2} f(\eta) Jk_\theta(p^{-1}(\eta)) \frac{\sin((k_\theta \circ p^{-1}(\eta))_1)}{\sin(p^{-1}(\eta)_1)} \, d\eta^2.
\end{aligned}$$

From these results, the first equality of (A.4) follows. The next equality follows from the fact that k_θ is a bijection on a subset of \mathbb{R}^2 of full dimension and

$$Dk_\theta(\omega_1, \omega_2) = \begin{bmatrix} l'_\theta(\omega_1) & 0 \\ 0 & 1 \end{bmatrix}.$$

Combining (A.1) and (A.4), we find

$$\beta(\theta)e^{\theta\tau(\cos \omega_1)} = l'_\theta(\omega_1) \frac{\sin l_\theta(\omega_1)}{\sin \omega_1},$$

or

$$(\cos l_\theta(\omega_1))' = -\beta(\theta)\sin(\omega_1)e^{\theta\tau(\cos \omega_1)}.$$

This equation has the unique solution, among increasing bijections on $[0, \pi)$,

$$\cos l_\theta(\omega_1) = 1 - \int_{\cos(\omega_1)}^1 \beta(\theta)e^{\theta\tau(u)} \, du = g_\theta(\cos \omega_1),$$

say, where

$$\beta(\theta) = \frac{2}{\int_{-1}^1 e^{\theta\tau(u)} \, du}.$$

Therefore, according to (A.3), h_θ^{-1} is of the form stated in (A.2).

Acknowledgement

This work was supported in part by MaPhySto – Centre for Mathematical Physics and Stochastics, funded by a grant from the Danish National Research Foundation.

References

- Baddeley, A.J. and Møller, J. (1989) Nearest-neighbour Markov point processes and random sets. *Internat. Statist. Rev.*, **57**, 89–121.

- Baddeley, A.J. and Turner, R. (2000) Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics*. To appear.
- Baddeley, A.J. and van Lieshout, M.N.M. (1995) Area-interaction point processes. *Ann. Inst. Statist. Math.*, **47**, 601–619.
- Baddeley, A.J., Møller, J. and Waagepetersen, R. (2000) Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*. To appear.
- Barndorff-Nielsen, O.E. (1978) *Information and Exponential Families in Statistical Theory*. Chichester: Wiley.
- Brix, A. and Møller, J. (1998) Space-time multitype log Gaussian Cox processes with a view to modelling weed data. Research report no. R-98-2012, Department of Mathematics, Aalborg University, Denmark.
- Geyer, C.J. (1999) Likelihood inference for spatial processes. In O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout (eds), *Stochastic Geometry, Likelihood and Computation*, pp. 79–140. Boca Raton, FL: Chapman and Hall/CRC.
- Jensen, E.B.V. (1998) *Local Stereology*, Singapore: World Scientific.
- Mardia, K.V. (1972) *Statistics of Directional Data*. London: Academic Press.
- Meiring, W. (1995) Estimation of heterogeneous space-time covariance. PhD thesis, University of Washington.
- Møller, J. (1999) Markov chain Monte Carlo and spatial point processes. In O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout (eds), *Stochastic Geometry, Likelihood and Computation*, pp. 141–172. Boca Raton, FL: Chapman and Hall/CRC.
- Møller, J., Syversveen, A.R. and Waagepetersen, R. (1998) Log Gaussian Cox processes. *Scand. J. Statist.*, **25**, 451–482.
- Monestiez, P., Sampson, P.D. and Guttorp, P. (1993) Modelling of heterogeneous spatial correlation structure by spatial deformation. *Cahiers de Géostatistique*, **3**, 1–12.
- Nielsen, L.S., Baddeley, A.J. and Jensen, E.B.V. (2000) Estimation in models for inhomogeneous point processes. Research report 12, Laboratory for Computational Stochastics, Department of Mathematical Sciences, University of Aarhus. In preparation.
- Ogata, Y. and Tanemura, M. (1986) Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In I.S. Francis, B.F.J. Manly and F.C. Lam (eds), *Proceedings of the Pacific Statistical Congress – 1985*, pp. 150–154. Amsterdam: Elsevier.
- Perrin, O. (1997) *Modèle de covariance d'un processus non-stationnaire par déformation de l'espace et statistique*. Doctoral thesis, Université de Paris I Panthéon-Sorbonne.
- Rathbun, S.L. (1996) Estimation of poisson intensity using partially observed concomitant variables. *Biometrics*, **52**, 226–242.
- Ripley, B.D. (1990) Gibbsian interaction models. In D.A. Griffith (ed.), *Spatial Statistics: Past, Present and Future*, pp. 3–28. New York: Institute of Mathematical Geography, Ann Arbor, MI.
- Ripley, B.D. and Kelly, F.P. (1977) Markov point processes. *J. London Math. Soc.*, **15**, 188–192.
- Sampson, P.D. and Guttorp, P. (1992) Nonparametric estimation of nonstationary spatial covariance structure. *J. Amer. Statist. Assoc.*, **87**, 108–119.
- Smith, R.L. (1997) Detecting signals in climatological data. In *Bulletin of the International Statistical Institute: Proceedings of the 51st Session (Istanbul)*, Book 1, pp. 211–214. Voorburg, Netherlands: ISI.
- Stoyan, D. and Stoyan, H. (1998) Non-homogeneous Gibbs process models for forestry – a case study. *Biometrical J.*, **40**, 521–531.

Stoyan, D., Kendall, W.S. and Mecke, J. (1995) *Stochastic Geometry and Its Applications*, 2nd edition. Chichester: Wiley.

Strauss, D.J. (1975) A model for clustering. *Biometrika*, **62**, 467–475.

Received July 1998 and revised August 1999