Non-parametric Bayesian inference for inhomogeneous Markov point processes

Berthelsen, Kasper Klitgaard; Møller, Jesper; Johansen, Per Michael

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

AALBORG UNIVERSITY

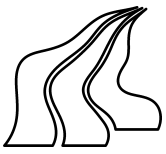
Non-parametric Bayesian inference for inhomogeneous Markov point processes

by

Kasper K. Berthelsen and Jesper Møller

R-2007-09 March 2007

Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7 G ▪ DK - 9220 Aalborg Øst ▪ Denmark
Phone: +45 96 35 80 80 ▪ Telefax: +45 98 15 81 29
URL: www.math.auc.dk/research/reports/reports.htm



ISSN 1399–2503 ▪ On-line version ISSN 1601–7811

Non-parametric Bayesian inference for inhomogeneous Markov point processes

Kasper K. Berthelsen Jesper Møller

March 2, 2007

Abstract

With reference to a specific data set, we consider how to perform a flexible non-parametric Bayesian analysis of an inhomogeneous point pattern modelled by a Markov point process, with a location dependent first order term and pairwise interaction only. A priori we assume that the first order term is a shot noise process, and the interaction function for a pair of points depends only on the distance between the two points and is a piecewise linear function modelled by a marked Poisson process. Simulation of the resulting posterior using a Metropolis-Hastings algorithm in the "conventional" way involves evaluating ratios of unknown normalising constants. We avoid this problem by applying a new auxiliary variable technique introduced by Møller, Pettitt, Reeves & Berthelsen (2006). In the present setting the auxiliary variable used is an example of a partially ordered Markov point process model.

Key words: Auxiliary variable method; Bayesian inference; hard core; Markov chain Monte Carlo; Markov point process; Non-parametric inference; Pairwise interaction point process; Partially ordered Markov point process; Perfect simulation; Shot noise process.

1 Introduction

Observed spatial point patterns often show signs of both inhomogeneity in the distribution of points and interaction between the points. Classical statistical models for inhomogeneous point patterns usually ignore or at least do not directly model the interaction. The focus has been on Poisson process models, where the intensity function may be random (i.e. a Cox process), or on Poisson cluster process models (Cox & Isham 1980, Ripley 1981, Cressie 1993, Stoyan, Kendall & Mecke 1995, Diggle 2003, Møller & Waagepetersen 2003). Usually the dimension of unknown parameters is low and moment or likelihood methods are used. A flexible non-parametric Bayesian model for the intensity function of a Poisson process is considered in Heikkinen & Arjas (1998).

In recent years, there has been an increasing interest in modelling both inhomogeneity and interaction, using likelihood methods for Markov point processes

or transformations of Markov point processes, still with a low-dimensional unknown parameter (Baddeley, Møller & Waagepetersen 2000, Jensen & Nielsen 2000, Nielsen 2000, Hahn, Jensen, van Lieshout & Nielsen 2003, Møller & Waagepetersen 2003). The objective in the present paper is to introduce and demonstrate the feasibility of a non-parametric Bayesian approach for inhomogeneous Markov point processes, using a new flexible model and a recent MCMC (Markov chain Monte Carlo) technique (Møller et al. 2006) for posterior simulation when the normalising constant of the observation model is not expressible on closed form and infeasible to calculate. Briefly, this MCMC technique is an auxiliary variable method based on perfect simulation in such a way that no calculation of normalising constants is required.

Specifically, our likelihood is the density of an inhomogeneous pairwise interaction point process Y with points in a bounded region $W \subset \mathbb{R}^2$. Thus a realisation of Y is a finite subset of W (a "point pattern"), and the density of Y with respect to the unit rate Poisson process on W is of the form

$$\pi(y|\beta,\varphi) = \frac{1}{Z_{(\beta,\varphi)}} \prod_i \beta(y_i) \prod_{i<j} \varphi(\|y_i - y_j\|). \tag{1}$$

Here $y = \{y_1, \ldots, y_n\} \subset W$ is an arbitrary point pattern, $n \in \{0, 1, \ldots\}$ is the number of points, $Z_{(\beta,\varphi)}$ is a normalising constant, $\beta : W \mapsto [0, \infty)$ is called the first order term, $\varphi : [0, \infty) \mapsto [0, \infty)$ the interaction function, and $\|\cdot\|$ denotes usual distance. Usually in applications, $\int_W \beta(\xi)\,\mathrm{d}\xi < \infty$ and φ is a repulsive interaction function, i.e. $\varphi \leq 1$ and φ is non-decreasing, in which case (1) is a well-defined density. In general, unless $\varphi = 1$, no closed form expression of $Z_{(\beta,\varphi)}$ is available which results in some difficulties considered in Section 4. Section 6 discusses the condition that $\varphi \leq 1$ and the possibility of analysing more general Markov point processes than (1).

Our non-parametric Bayesian approach specifies a prior for β by a shot noise process and a prior for φ by a piecewise linear function modelled by a marked Poisson process. Thus the model for Y could be called a shot noise Markov point process (in the special case $\varphi = 1$, Y is a shot noise Cox process (Møller 2003)). This is an extension of our previous approach in Berthelsen & Møller (2003), where a priori $\beta(\cdot)$ is assumed to be constant (the homogeneous case), but where we were not aware of the existence of the auxiliary variable method. This method was used in Berthelsen & Møller (2006), but a priori both β and φ were assumed to be of a simple parametric form. Though in the present paper we consider a much more complex Bayesian model, we manage to demonstrate the feasibility of our approach based on the auxiliary variable method. As in Berthelsen & Møller (2006), we let the auxiliary variable in the auxiliary variable method be a partially ordered Markov model point process approximating (1). It turns out that depending whether or not we include a hard core parameter h in the model (i.e. $\varphi(r) = 0$ for $r < h$), we need different kinds of partially ordered Markov model point processes.

The remainder of this paper is organised as follows. Section 2 presents a data set which is used throughout the paper for specificity and illustrative purposes.

Section 3.1 specifies the prior for β, and Section 3.2 the prior for φ. Section 4 considers the posterior and discusses the auxiliary variable method for posterior sampling in our first setting, where φ does not involve a hard core parameter. Results for the Bayesian analysis of the data are presented in Section 5, where mainly the case where φ involves a hard core parameter is considered. Section 6 contains concluding remarks.

2 Data example

The left panel of Figure 1 shows an inhomogeneous point pattern $y = \{y_1, \ldots, y_n\}$ consisting of $n = 617$ points observed within a rectangular window. The points specify the location of cells in a section of the mocous membrane of the stomach of a healthy rat, and the observation window is of size $W = [0, 1] \times [0, 0.893]$ (after some rescaling). The left hand side of the observation window corresponds to where the stomach cavity begins and the right hand side to where the muscle tissue begins. In the sequel we consider mainly the data y for illustrative purposes.

To asses if this point pattern can be satisfactory modelled by an inhomogeneous Poisson process, we have simulated a number of realisations of an inhomogeneous Poisson processes, with the intensity estimated from the data by a non-parametric method (Diggle 1985). For these simulations and the cell data we compared estimated inhomogeneous K-functions and estimated pair correlation functions; these statistics summarise the second order properties of a spatial point process, see Baddeley et al. (2000) and the references therein. This comparison clearly showed that a Poisson process is an inadequate model for the data. For instance, the centre panel of Figure 1 shows a non-parametric estimate $\hat{g}(r),\ r > 0$, of the pair correlation function for the data and simulated 95%-envelopes (for details, see Sections 3.4.3-3.4.4 in Møller & Waagepetersen (2003)). Under the Poisson model the theoretical pair correlation function is constant 1. The low values of $\hat{g}(r)$ for distances $r < 0.01$ indicates repulsion between the points.

The cell data was originally analysed by Nielsen (2000) who modelled the data by transforming the first coordinates of the points of a Strauss point process (compared to Nielsen (2000) our data in Figure 1 are rotated 90 degrees, so Nielsen was actually transforming the second coordinates). Specifically, each data point $y_i = (y_{i1}, y_{i2})$ corresponds to a point $z_i = (z_{i1}, z_{i2})$ of the Strauss process, where $z_{i1} = (\mathrm{e}^{\theta y_{i1}} - 1)/(\mathrm{e}^{\theta} - 1)$ and $y_{i2} = z_{i2}$. Thus $z = \{z_1, \ldots, z_n\}$ is considered to be a realisation of a Strauss point process on W, with density of the form (1) (with y replaced by z) where $\beta(\xi) \equiv \beta$ is constant and $\varphi(r) = \gamma^{\ [r \le R]}$, with $0 < \gamma \le 1$, $R > 0$, and $[\cdot]$ being the indicator function. Nielsen (2000) found that the transformed Strauss point process with estimates $\theta = 1.3043$, $\beta = 760$, $\gamma = 0.09$, and $R = 0.007$ fit the data well. The right panel of Figure 1 shows a non-parametric estimate of the pair correlation function for the data, with simulated 95%-envelopes under the fitted transformed Strauss point process. The estimated pair correlation is almost within the 95% evelopes for

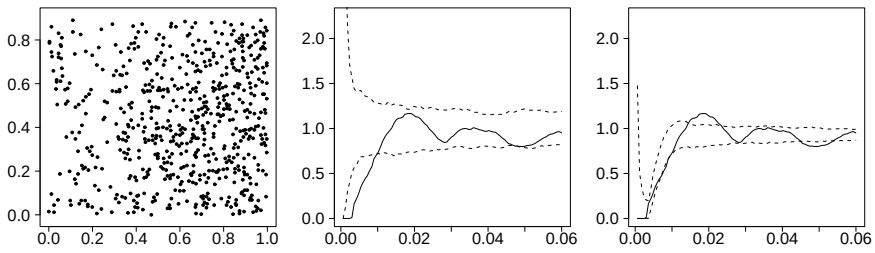


Figure 1: Left panel: Locations of 617 cells in a 2D section of the mocous membrane of the stomach of a healthy rat. Centre panel: Non-parametric estimate of the pair correlation function for the cell data (full line) and 95%-envelopes calculated from 200 simulations of a fitted inhomogeneous Poisson process. Right panel: Non-parametric estimate of the pair correlation function for the cell data (full line) and 95%-envelopes calculated from 200 simulations of the model fitted by Nielsen (2000).

small values of the distance r, suggesting that the transformed Strauss model captures the small scale inhibition in the data. Overall, the estimated pair correlation function follows the trend of the 95%-envelopes, but it falls outside the envelopes for some values. As the comparison with the envelopes can be considered as a multiple test problem, this is not necessarily reason to reject the transformed Strauss model.

3 Prior assumptions

This section specifies our prior model for β and φ when analysing the cell data. We assume a priori that all underlying random quantities in Section 3.1 specifying β are independent of all underlying random quantities in Section 3.2 specifying φ. We let $W = [0, a] \times [0, b]$ denote the observation window ($a = 1, b = 0.893$).

3.1 Prior for the first order term

A priori it is expected that the cell intensity only changes in the direction from the stomach cavity to the surrounding muscles tissue, so we assume that β only depends on its first coordinate (this assumption is also satisfied by the transformed Strauss process used in Nielsen (2000)). As a flexible prior for β we choose below a shot noise process, specified by a scale parameter $\gamma > 0$ and an infinite point process $\psi = \{c_1, c_2, \ldots\} \subset \mathbb{R}$ so that β becomes stationary (extending the definition of β to $\mathbb{R}^2$). For Bayesian inference we need a finite version $\{c_j \in [-\Delta, a + \Delta]\}$ obtained by restricting the process to the interval $[-\Delta, a+\Delta]$; we also denote this finite point process by ψ. Below we discuss how

to choose $\Delta > 0$ such that β on W is effectively the same for the two versions of ψ.

For either of the two versions of ψ, the shot noise process is assumed to be of the form

$$\beta(\xi) \equiv \beta(\xi; \psi, \gamma) = \gamma \sum_j \phi((\xi_1 - c_j)/\sigma_1)/\sigma_1, \qquad \xi = (\xi_1, \xi_2) \in \mathbb{R}^2, \quad (2)$$

where ϕ is a kernel, here chosen to be the density of the standard normal distribution, $\sigma_1 > 0$ is the kernel width, and ψ is independent of γ. The marginal distributions of ψ and γ are specified later, but these distributions play no important role in Section 3.1.1.

3.1.1 A useful result

The following proposition is useful when determining Δ in Section 3.1.2.

Recall that conditional on (β, φ), the Papangelou conditional intensity corresponding to the pairwise interaction point process density (1) is defined for points $\xi \in W \setminus y$ by

$$\lambda(y, \xi|\beta, \varphi) = \begin{cases} \beta(\xi) \prod_i \varphi(\|\xi - y_i\|) & \text{if } \pi(y|\beta, \varphi) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

and the intensity $\rho(\xi|\beta, \varphi)$ of the point process is given by the mean of the conditional intensity,

$$\rho(\xi|\beta, \varphi) = \int \lambda(y, \xi|\beta, \varphi)\pi(y|\beta, \varphi)\, \nu_W(\mathrm{d}y) \quad (4)$$

where ν_W denotes the unit rate Poisson process on W, see Møller & Waagepetersen (2003). Note that $\lambda(y, \xi|\beta, \varphi) \leq \beta(\xi)$ and $\int_W \rho(\xi|\beta, \varphi)\, \mathrm{d}\xi$ is the mean number of points in the pairwise interaction point process (when we condition on (β, φ)). Therefore, conditional on (ψ, γ),

$$B = \int_W \beta(\xi; \psi, \gamma)\, \mathrm{d}\xi$$

is an upper bound on the expected number of points.

Now, consider the infinite version of ψ and assume that it is a stationary point process on $\mathbb{R}$ (i.e. its distribution is invariant under arbitrary translations) with an intensity, κ_1 say, which is positive and finite. Then

$$M = \gamma \int_W \sum_{j:\, c_j \notin [-\Delta, a+\Delta]} \phi((\xi_1 - c_j)/\sigma_1)\, \mathrm{d}\xi$$

is the reduction of B when approximating the infinite version of ψ by its finite version. Finally, assume that the mean $\mathbb{E}\gamma$ exists, i.e. $0 < \mathbb{E}\gamma < \infty$, and let Φ denote the distribution function of the standard normal distribution.

Proposition 1: We have

$$\mathbb{E}M/\mathbb{E}B = \frac{2}{a}\int_0^a \Phi((-s-\Delta)/\sigma_1)\,\mathrm{d}s. \tag{5}$$

Proof: Using first Fubini's theorem and the independence between γ and ψ, and second Campbell's theorem (see e.g. Møller & Waagepetersen (2003)), we obtain

$$\begin{aligned}\mathbb{E}B &= \mathbb{E}\gamma \int_W \mathbb{E}\sum_{j=1}^{\infty} \phi((\xi_1 - c_j)/\sigma_1)\sigma_1\,\mathrm{d}\xi \\ &= \kappa_1 \mathbb{E}\gamma \int_W \int_{-\infty}^{\infty} \phi((\xi_1 - c)/\sigma_1)/\sigma_1\,\mathrm{d}c\,\mathrm{d}\xi \\ &= ab\kappa_1\mathbb{E}\gamma.\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbb{E}M &= \kappa_1\mathbb{E}\gamma \int_W \int_{\mathbb{R}\setminus[-\Delta,a+\Delta]} \phi((\xi_1 - c)/\sigma_1)/\sigma_1\,\mathrm{d}c\,\mathrm{d}\xi \\ &= b\kappa_1\mathbb{E}\gamma \int_0^a [\Phi((-\Delta-\xi_1)/\sigma_1) + \Phi((-a-\Delta+\xi_1)/\sigma_1)]\,\mathrm{d}\xi_1.\end{aligned}$$

Combining the two results, we obtain (5).

3.1.2 Prior specification of β using the finite version of ψ

Henceforth we consider only the finite version of ψ, and assume that it is a homogeneous Poisson process on the interval $I = [-\Delta, a+\Delta]$ with intensity κ_1. Furthermore, γ is assumed to be gamma-distributed with shape parameter $\alpha_1 > 0$ and scale parameter $\alpha_2 > 0$. If ν_I denotes the unit rate Poisson process on I, then the joint distribution of ψ and γ has density

$$\pi(\psi,\gamma) \propto \kappa_1^{n(\psi)}\gamma^{\alpha_1-1}\mathrm{e}^{-\gamma/\alpha_2} \tag{6}$$

with respect to the product measure of ν_I and Lebesgue measure on $(0,\infty)$, where $n(\psi)$ denotes the number of points in ψ.

The remaining parameters $\sigma_1, \kappa_1, \alpha_1, \alpha_2, \Delta$ are specified by fixed values so that a satisfactory degree of flexibility for β is obtained, using practical considerations and simulation experiments as follows.

The kernel intensity κ_1 and the kernel width σ_1 determine how flexible the shape of β is. The higher κ_1 is, the more kernels, which in turn implies more flexibility. From a practical point of view a high value of κ_1 may lead to slow mixing in our MCMC algorithm for posterior simulations (Section 4). As a compromise, after some experimentation, we choose $\kappa_1 = 30$ and $\sigma_1 = 0.1$.

The parameter γ is scaling β and determined by the choice of α_1 and α_2. With the small scale inhibition in the data we expect the "true" value of $\mathbb{E}B$

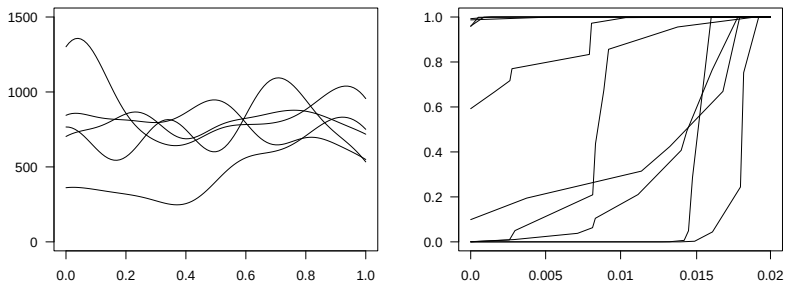


Figure 2: Left panel: Five independent realisations of β under its prior distribution. Right panel: Ten independent realisations of φ under its prior distribution.

to be slightly larger than 617 (the observed number of points). Hence, with the prior mean $\mathbb{E}B = ab\kappa_1\mathbb{E}\gamma$ in mind, we choose $\alpha_1 = 10$ and $\alpha_2 = 2.5$ so that $\mathbb{E}\gamma = 25$.

Finally, we choose Δ so that $\mathbb{E}M/\mathbb{E}B = 0.001$, where the integral in (5) is evaluated by numerical methods. Thereby we obtain $\Delta = 0.126$. The left panel in Figure 2 shows five independent realisations of β under its prior distribution, and ten independent realisations of φ under its prior distribution (specified in the next section). We find these realisations displaying a satisfactory degree of flexibility for both β and φ.

3.2 Prior for the interaction function

We consider first a similar approximation of φ as in Berthelsen & Møller (2003) but using an increasing, continuous, piecewise linear function:

$$\varphi(r) \equiv \varphi(r;\chi) = \ [r > r_p] + \sum_{i=1}^{p} \ [r_{i-1} < r \leq r_i] \left(\frac{r - r_{i-1}}{r_i - r_{i-1}} (\gamma_{i+1} - \gamma_i) + \gamma_i \right) \tag{7}$$

where $\chi = \{(r_1, \gamma_1), \ldots, (r_p, \gamma_p)\}$ is a finite point process, assuming $r_0 = 0 < r_1 < \ldots < r_p < \infty$ and $0 < \gamma_1 < \ldots < \gamma_p < \gamma_{p+1} = 1$. We refer to $r_1, \ldots, r_p$ as the change-points of φ and r_p as the range of interaction, since $r_p = \inf\{r > 0 : \varphi(s;\chi) = 1 \text{ for all } s > r\}$. Note that for $i = 1, \ldots, p$, $\{(r, \varphi(r;\chi)) : r_{i-1} < r \leq r_i\}$ is the line segment with endpoints (r_{i-1}, γ_i) and (r_i, γ_{i+1}), while in Berthelsen & Møller (2003), φ is the step function given by $\varphi(r;\chi) = \gamma_i$ if $r_{i-1} < r \leq r_i$. We find it a priori more natural to approximate φ by the continuous function (7) than by the step function in Berthelsen & Møller (2003), but expect that the difference in posterior results obtained using the two interaction functions will be minor. In Section 5, we consider a modification of (7) involving a hard core parameter, but for now we restrict attention to the interaction function in (7).

Along similar lines as in Berthelsen & Møller (2003), we make the following prior assumptions. Set $\gamma_0 = 0$, $\delta_i = \gamma_i - \gamma_{i-1}$, and $(\zeta_1, \ldots, \zeta_p) = (\ln(\delta_2/\delta_1), \ldots, \ln(\delta_{p+1}/\delta_p))$. Then $r_1 < \ldots < r_p$ are the events of a homogeneous Poisson process on $[0, r_{\max}]$ of rate κ_2; and conditionally on $(r_1, \ldots, r_p)$, we have that $\zeta_p, \ldots, \zeta_1$ is a Markov chain with $\zeta_p \sim \mathrm{N}(0, \sigma_2^2)$ and $\zeta_i | \zeta_{i+1} \sim \mathrm{N}(\zeta_{i+1}, \sigma_2^2)$, $i = p-1, \ldots, 1$. If μ denotes the unit rate Poisson process on $[0, r_{\max}] \times [0, 1]$, then $\chi = \{(r_1, \gamma_1), \ldots, (r_p, \gamma_p)\}$ with $r_1 < \ldots < r_p$ has density

$$\begin{aligned} \pi(\chi) &\propto \kappa_2^p \ [0 < \gamma_1 < \ldots < \gamma_p < 1]/(\delta_1 \times \cdots \times \delta_{p+1}) \\ &\times \left(2\pi\sigma_2^2\right)^{-p/2} \exp\left(-\sum_{i=1}^{p} \left(\zeta_i - \zeta_{i+1}\right)^2 / \left(2\sigma_2^2\right)\right) \end{aligned} \tag{8}$$

with respect to μ, where we set $\zeta_{p+1} = 0$.

As in the case of β, the remaining parameters are given fixed values chosen so that φ is satisfactory flexible. The parameter $r_{\max}$ specifies the largest interaction range allowed by the model. The estimated pair correlation function displayed in Figure 1 indicates that there is little interaction beyond an interpoint distance of 0.01. To be on the safe side, we let $r_{\max} = 0.02$. Further, $\mathbb{E}p = \kappa_2 r_{\max}$ specifies the "resolution". As in Berthelsen & Møller (2003), we let $\sigma_2 = 1$ and $r_{\max}\kappa_2 = 5$, whereby $\kappa_2 = 250$ is obtained.

4 Sampling from the posterior

Combining (1), (6), and (8), we obtain the posterior density for $\theta = (\psi, \gamma, \chi)$ with respect to the product measure of ν_I, Lebesgue measure on $(0, \infty)$, and μ:

$$\begin{aligned} \pi(\theta|y) \propto{}& \kappa_1^{n(\psi)} \gamma^{\alpha_1 - 1} \mathrm{e}^{-\gamma/\alpha_2} \kappa_2^p \ [0 < \gamma_1 < \ldots < \gamma_p < 1]/(\delta_1 \times \cdots \times \delta_{p+1}) \\ &\times \left(2\pi\sigma_2^2\right)^{-p/2} \exp\left(-\sum_{i=1}^{p} \left(\zeta_i - \zeta_{i+1}\right)^2 / \left(2\sigma_2^2\right)\right) \\ &\times \frac{1}{Z_\theta} \prod_i \beta(y_i; \psi, \gamma) \prod_{i<j} \varphi(\|y_i - y_j\|; \chi) \end{aligned} \tag{9}$$

where β and φ are given by (2) and (7). Clearly, this is intractable; in particular, the normalising constant Z_θ is not computable.

Inference will be based on Monte Carlo estimates obtained from simulations of the posterior distribution (9) using the Metropolis-Hastings algorithm. The main difficulty is that the Hastings ratio in a "conventional" Metropolis-Hastings algorithm involves a ratio $Z_\theta / Z_{\theta'}$ of normalising constants which we cannot compute. In this section we restrict attention to how to deal with this major problem. It is possible to approximate the ratio of unknown normalising constants in a number of ways, using e.g. path sampling (Gelman & Meng 1998), with recent examples of Green & Richardson (2002) and Berthelsen & Møller (2003). Instead we use the auxiliary variable method introduced in Møller et al.

(2006) which avoids such approximations. The idea is briefly explained below; the remaining details of the algorithm are given in Appendix A.

Let x be an auxiliary random variable which conditional on (θ, y) follows a density $f(\cdot|\theta, y)$ defined on the same space as the likelihood $\pi(\cdot|\theta) = \pi(\cdot|\beta, \varphi)$, i.e.

$$f(x|\theta, y) = 0 \quad \text{whenever } \pi(x|\theta) = 0. \tag{10}$$

This condition is automatically satisfied in the present setting where $\pi(\cdot|\theta) > 0$, but some care is needed in the case of the modified likelihood considered in Section 5. We sample (x, θ) from the joint density $\pi(x, \theta|y) \propto f(x|\theta, y)\pi(y|\theta)\pi(\theta)$, whereby the marginal distribution of θ is given by $\pi(\theta|y)$ in (9). This is done by the following Metropolis-Hastings algorithm. Suppose that (x, θ) is the current state. First, generate a proposal θ' from some proposal density $p(\theta'|\theta, y)$ from which we can easily make simulations (like in a "conventional" Metropolis-Hastings algorithm for sampling from (9)). Second, conditional on θ' generate a proposal x' from $\pi(x'|\theta')$ (the likelihood (1) evaluated for the proposed value θ'). The joint proposal (x', θ') is then accepted with probability $\min\{1, H(x', \theta'|x, \theta, y)\}$, and otherwise we retain (x, θ). Here

$$H(x', \theta'|x, \theta, y) = \frac{f(x'|\theta', y)\pi(y|\theta')\pi(\theta')\pi(x|\theta)p(\theta|\theta', y)}{f(x|\theta, y)\pi(y|\theta)\pi(\theta)\pi(x'|\theta')p(\theta'|\theta, y)} \tag{11}$$

is the Hastings ratio in which the unknown normalising constants cancel. Assuming that $p(\theta'|\theta, y)$ has been chosen well (if a "conventional" Metropolis-Hastings algorithm was used), the success of this method relies on how well the auxiliary density $f(\cdot|\theta, y)$ approximates $\pi(\cdot|\theta)$, cf. Møller et al. (2006).

In the present point process setting we use a partially ordered Markov model (POMM) point process x on W as the auxiliary point process. The POMM point process is best described by how to simulate x (see also Berthelsen & Møller (2006) and the references therein): Initially, W is partitioned into N disjoint subsets C_i, $i = 1, \ldots, N$. Sequentially, for $i = 1, \ldots, N$, conditional on (θ, y) and $x \cap C_1, \ldots, x \cap C_{i-1}$ (if $i > 1$), let $x \cap C_i$ be a realisation of a homogeneous Poisson process on C_i with intensity $\lambda_i(y, \theta, x \cap C_1, \ldots, x \cap C_{i-1})$ (which we read as $\lambda_1(y, \theta)$ if $i = 1$). This construction allows the POMM point process to incorporate some interaction while still having a normalised density as shown below.

Berthelsen & Møller (2006) conclude that the POMM point process is a good approximation of a pairwise interaction point process with finite interaction range, when the side lengths of rectangular subsets C_i are less than one 10th of the interaction range. In the present setting, where we expect the interaction range to be larger than 0.01, cf. Figure 1, we divide W into $N = 1000 \times 1000$ rectangular subsets of equal size and shape. For computational reasons we choose to index the subsets systematically row-wise. For $i = 1, \ldots, N$, let ξ_i be the centroid of C_i and let

$$\lambda_i(y, \theta, x \cap C_1, \ldots, x \cap C_{i-1}) = \omega\beta(\xi_i; \psi, \gamma)\Phi_i(x; \chi) \tag{12}$$

where

$$\Phi_i(x;\chi) = \prod_{j<i} \prod_{\eta \in x \cap C_j} \varphi(\|\eta - \xi_i\|;\chi) \tag{13}$$

(setting $\Phi_1(x;\chi) = 1$) and $\omega = \omega(\theta, y) > 0$. Then the POMM point process has density

$$f(x|\theta, y) = \exp\left[|W| - \sum_{i=1}^{N} |C_i| \lambda_i(y, \theta, x \cap C_1, \ldots, x \cap C_{i-1})\right] \times \prod_{i=1}^{N} \lambda_i(y, \theta, x \cap C_1, \ldots, x \cap C_{i-1})^{n_i(x)} \tag{14}$$

with respect to ν_W, where $|\cdot|$ denotes area and $n_i(x)$ denotes the cardinality of $x \cap C_i$. Note that compared to Berthelsen & Møller (2006) distances in $\Phi_i(x;\chi)$ are between the reference point ξ_i associated with C_i and points in $x \cap C_j$, $j < i$, and not between pairs of reference points. We expect this refinement to slightly improve how well the pairwise interaction point process can be approximated by the POMM point process.

The parameter ω in (12) is introduced because results in Berthelsen & Møller (2006) indicate that a POMM process equivalent to setting $\omega = 1$ is not the best choice. In Berthelsen & Møller (2006), instead of introducing the ω term, β in (12) and φ in equation (13) are essentially replaced by "optimised" choices that are functions of β and φ. These functions are estimated by MCMC methods prior to posterior simulations. This scheme is only feasible because Berthelsen & Møller (2006) consider the much simpler Strauss point process. In the present paper, we opt for the simpler solution of introducing the ω term, which we let be given by the maximum likelihood estimate when y given θ is assumed to follow (14), i.e.

$$\omega(\theta, y) = n(y) / \sum_{i=1}^{N} |C_i| \beta(\xi_i; \psi, \gamma) \Phi_i(y;\chi).$$

The auxiliary variable method requires exact samples from the point process specified by (1). Under the present assumption of repulsion this is possible using the dominated coupling from the past (dominated CFTP) algorithm in Kendall & Møller (2000).

5 Analysis

All posterior results in this section are Monte Carlo estimates based on the MCMC algorithm given in Section 4 and Appendix A.

5.1 Including a hard core parameter in the model

We first carried out a Bayesian analysis of the cell data in Figure 1, using the data model and prior specified in Sections 1 and 3. The results of this analysis

(not shown here) indicated the need for an explicit hard core assumption in the model. With the biological origin of the data in mind, this appears to be a reasonable assumption. Accordingly, we next modify the data distribution by replacing $\varphi(r;\chi)$ throughout by

$$\tilde{\varphi}(r;h,\chi) = \begin{cases} 0 & \text{if } r < h \\ \varphi\big(\frac{(r-h)r_{\max}}{r_{\max}-h};\chi\big) & \text{if } h \le r \le r_{\max} \\ 1 & \text{if } h > r_{\max} \end{cases}$$

where the hard core distance h is assumed to be smaller than $r_{\max}$ and a priori uniformly distributed on the interval $[0, r_{\max}]$. Notice that $\tilde{\varphi}(\cdot;h,\chi)$ has a discontinuity at $r = h$, but is otherwise continuous. Furthermore, we redefine $\theta = (\psi, \gamma, \chi, h)$.

The auxiliary variable method described in Section 4 remains unchanged apart from an modification of the POMM point process due to the hard core and the condition (10). Let $\bar{\varphi}(r;h,\chi) = \tilde{\varphi}(r;h+\epsilon,\chi)$, where ϵ is the largest possible distance from ξ_i to any point in C_i. Conditional on (θ, y) and $x\cap C_1, \ldots, x\cap C_{i-1}$, let $\lambda_i(\theta, y, x\cap C_1, \ldots, x\cap C_{i-1})$ be given by (12) when we replace $\Phi_i(x;\chi)$ by $\Phi_1(x;h,\chi) = \ [n_1(x) \le 1]$ if $i = 1$ and

$$\Phi_i(x;h,\chi) = \prod_{j<i} \prod_{\eta\in x\cap C_j} \bar{\varphi}(\|\eta - \xi_i\|;h,\chi)$$

if $i > 1$. Then, conditional on (θ, y) and $x\cap C_1, \ldots, x\cap C_{i-1}$ (if $i > 1$), we let $x \cap C_i = \emptyset$ with probability $(1 + |C_i|\lambda_i(y,\theta,x\cap C_1,\ldots,x\cap C_{i-1}))^{-1}$, and else we let $x \cap C_i$ consists of a single point which is uniformly distributed on C_i. Thus the density with respect to ν_W is given by

$$\tilde{f}(x|\theta,y) = \mathrm{e}^{|W|} \prod_{i=1}^{N} \ [n_i(x) \le 1] \frac{\lambda_i(y,\theta,x\cap C_1,\ldots,x\cap C_{i-1})^{n_i(x)}}{1 + |C_i|\lambda_i(y,\theta,x\cap C_1,\ldots,x\cap C_{i-1})} \qquad (15)$$

and it follows from the triangle inequality that (10) is satisfied. Finally, we let $\omega = \omega(\theta, y)$ be given by the maximum likelihood estimate (MLE) when y given θ is assumed to follow (15). The MLE can easily be determined, since it is computationally equivalent to finding the MLE of a logistic regression, where the $n_i = n_i(y)$ are independent Bernoulli variables and $n_i = 0$ with probability $(1 + \omega|C_i|\beta(\xi_i,\psi,\gamma)\Phi_i(y;h,\chi))^{-1}$, $i = 1, \ldots, N$.

Since the hard core of the POMM point process is slightly larger than that of the likelihood, it is expected that some proposals will be rejected simply because they do not fulfil the hard core condition of the POMM point process. On one hand we can reduce the differences in hard core conditions by increasing N, on the other hand this comes at a computational price as evaluating (15) has complexity linear in N. As a compromise we use $N = 2000 \times 2000$ rectangular subsets in which case $\epsilon = \sqrt{a^2 + b^2}/4000$.

Using the sampling scheme described in Appendix A we have generated a Markov chain of length 250,000 which has the posterior (9) as its equilibrium distribution. In the following we ignore the initial 5000 iterations of the Markov chain as burn-in.

5.2 Posterior results

The left panel of Figure 3 shows the posterior mean of β, $\mathbb{E}(\beta|y)$, together with pointwise 95% central posterior intervals. Also the smooth estimate of the first order term obtained by Nielsen (2000) is shown, where the main difference compared with $\mathbb{E}(\beta|y)$ is the abrupt change of $\mathbb{E}(\beta|y)$ in the interval $[0.2, 0.4]$. For locations near the edges of W, $\mathbb{E}(\beta|y)$ is "pulled" towards its prior mean as a consequence of the smoothing prior.

Apart from boundary effects, since $\beta(\xi_1, \xi_2)$ only depends on ξ_1, we may expect that the intensity (4) only slightly depends on ξ_2, i.e. $\rho((\xi_1, \xi_2)|\beta, \varphi) \approx \rho(\xi_1|\beta, \varphi)$, where

$$\rho(\xi_1|\beta, \varphi) = \frac{1}{b} \int_0^b \rho((\xi_1, \xi_2)|\beta, \varphi) \, \mathrm{d}\xi_2.$$

We therefore refer to $\rho(\xi_1|\beta, \varphi)$ as the cell intensity, though it is more precisely the average cell intensity in W at $\xi_1 \in [0, a]$. A non-parametric estimate of $\rho(\xi_1|\beta, \varphi)$ is given by

$$\hat{\rho}(\xi_1) = \left[\sum_{i=1}^{n(y)} \phi((\xi_1 - y_{i1})/\sigma_k)/\sigma_k) \right] \Big/ \left[b \times (\Phi((1 - \xi_1)/\sigma_k) - \Phi(\xi_1/\sigma_k)) \right]$$

which is basically the 1-dimensional edge-corrected kernel estimator of Diggle (1985) with bandwidth $\sigma_k = 0.075$. The left panel of Figure 3 also shows this estimate. The posterior mean of $\beta(\xi_1)$ is not unlike $\hat{\rho}(\xi_1)$ except that $\mathbb{E}(\beta(\xi_1)|y)$ is higher as would be expected due to the repulsion in the likelihood, cf. Section 3.1.1.

The posterior mean of $\tilde{\varphi}$ is shown in the right panel of Figure 3 together with pointwise 95% central posterior intervals. The figure shows a distinct hard core on the interval from zero to the observed minimum inter-point distance $d = \min_{i \neq j} \|y_i - y_j\|$ (which is a little less than 0.006, see the left panel in Figure 5), and an effective interaction range which is no more than 0.015 (the posterior distribution of $\tilde{\varphi}(r)$ is concentrated close to one for $r > 0.015$). This further confirms that $r_{\max}$ was chosen sufficiently large. The corner at $r = d$ of the curve showing the posterior mean of $\tilde{\varphi}(r)$ is caused by that $\tilde{\varphi}(r)$ is often zero for $r < d$ (since the hard core is concentrated close to d), while $\tilde{\varphi}(r) > 0$ for $r > d$.

The prior and posterior distributions of the hard core distance h, the number of kernels $n(\psi)$, number of change points $n(\chi)$, and the parameter γ are shown in Figure 4. Moreover, the posterior distributions differ in various ways from the priors: The posterior distribution of the hard core distance is less than but concentrated close to the smallest inter-point distance in the data. The low number of change points suggests that a relatively simple interaction function suffices. Compared to the prior, the posterior distribution of $n(\psi)$ is shifted to the right, which could be interpreted as the data requires a more flexible first order term than a priori assumed. Furthermore, γ is a posteriori more

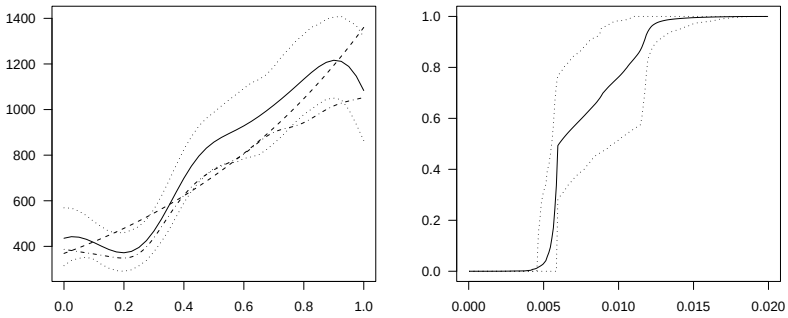


Figure 3: Posterior mean (solid line) and pointwise 95% central posterior intervals (dotted lines) for β (left panel) and $\tilde{\varphi}$ (right panel). The left panel also shows the first order term (dashed line) estimated by Nielsen (2000) and an estimate of the cell intensity (dot-dashed line).

concentrated around its mean than the prior, where the posterior and prior means are 25 and 20, respectively. This shift to the left for γ, as compared to the shift to the right for $n(\psi)$, is caused by a strong negative correlation between $n(\psi)$ and γ a posteriori (while they were assumed to be a priori independent). This is in accordance with the rather narrow 95% central posterior intervals in the left panel of Figure 3.

5.3 Posterior predictive results

Following Gelman, Meng & Stern (1996) we also performed a model check based on 246 samples from the posterior predictive distribution, where for $i = 5, \ldots, 250$, the ith sample $y^{(i)}$ was simulated from the pairwise interaction point process using the parameter values $(\beta^{(i)}, \varphi^{(i)})$ after update number $1000 \times i$ from the posterior distribution.

Figure 5 shows that both the observed minimum inter-point distance and observed number of points fall well within the support of the corresponding posterior predictive distributions.

The left and centre panels in Figure 6 compare the non-parametric summaries $\hat{\rho}(\xi_1)$ and $\hat{g}(r)$ based on the data y and the $y^{(i)}$, $i = 5, \ldots, 250$. The first order properties as represented by the intensity estimate $\hat{\rho}(\xi_1)$ from Figure 3 seem to be in good accordance with the posterior predictive distribution. For the second order properties as represented by the pair correlation estimate $\hat{g}(r)$, there seems to be some improvement as compared to the right panel in Figure 1. The observed pair correlation function is still for more than 5% of the considered r-values outside the simulated 95% envelopes, but less so than in Figure 1. This is partly due to wider 95% envelopes in Figure 6, which are a natural consequence of the fact that they are based on posterior predictive simulations (unlike Figure 1 which is based on simulations from a model with

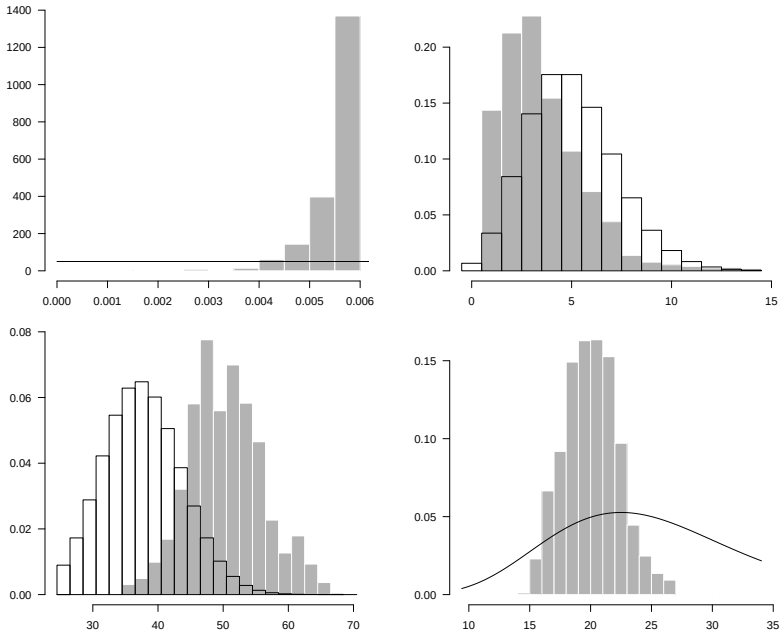


Figure 4: Row wise from top left panel, prior (solid line) and posterior (grey) distributions for h, $n(\chi)$, $n(\psi)$, and γ, respectively. For comparability only a selected range of each prior distribution is shown.

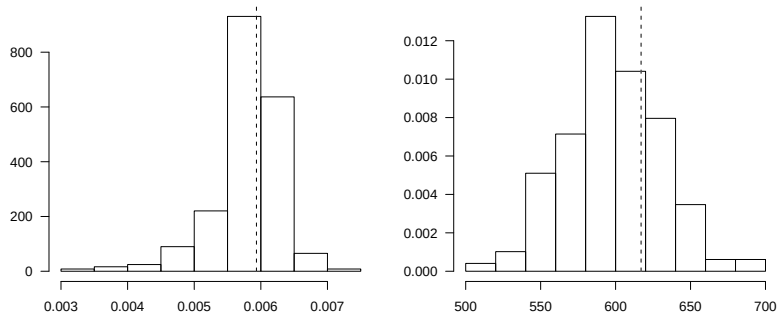


Figure 5: Observed value (dashed line) and posterior predictive distribution of minimum inter-point distance (left panel) and number of points (right panel).

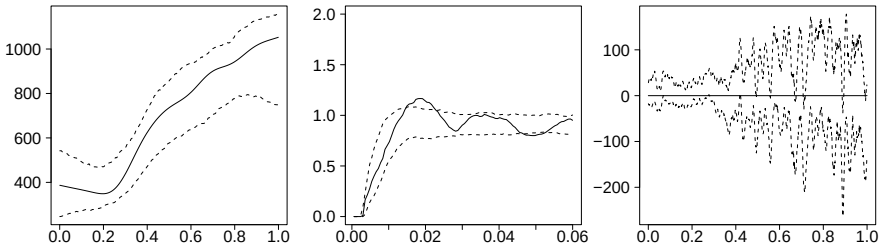


Figure 6: Left and centre panels: Observed (solid lines) $\hat{\rho}(\xi_1)$ (left panel) and $\hat{g}(r)$ (middle panel) together with pointwise 95% central posterior predictive intervals (dashed lines). Right panel: Pointwise 95% central posterior predictive intervals for the $s_i(\xi_1),\ i = 5, \ldots, 250$.

fixed parameters).

Apart from boundary effects, we may expect that the Papangelou conditional intensity (3) approximately only depends on $\xi = (\xi_1, \xi_2)$ through the first coordinate, $\lambda(y, (\xi_1, \xi_2)|\beta, \varphi) \approx \lambda(y, \xi_1|\beta, \varphi)$. Let $y^{(0)} = y$ be the data, and for $i = 0$ and $i = 5, \ldots, 250$, consider $q_i(\xi_1) = \sum_{\xi_2} \lambda(y^{(i)}, (\xi_1, \xi_2)|\beta^{(i)}, \varphi^{(i)})/1000$ as an estimate of $\lambda(y^{(i)}, \xi_1|\beta^{(i)}, \varphi^{(i)})$, using a 1000×1000 grid of $\xi = (\xi_1, \xi_2)$ values in W. The right panel in Figure 5 shows for each ξ_1-value the central 95% posterior predictive interval obtained from $s_i(\xi_1) = q_i(\xi_1) - q_0(\xi_1),\ i = 5, \ldots, 250$. There seems again to be a good accordance between the data and the posterior predictive distribution, since the horizontal line at zero is within the pointwise 95% envelopes for nearly all ξ_1-values.

6 Concluding remarks

In this paper we have considered how to perform Bayesian inference for pairwise interaction point processes using a flexible prior for both the first order term and the interaction function. The data has mainly been used for illustrative purposes, where the focus has been on demonstrating the feasibility and applicability of a non-parametric Bayesian MCMC approach based on the auxiliary variable method and a POMM point process.

Our approach can easily be extended to the case where the first order term $\beta(\xi)$ is believed to depend on both the first and second coordinates in $\xi = (\xi_1, \xi_2)$. This could be achieved by extending ψ to a point process on $\mathbb{R}^2$ and replacing the one dimensional normal kernels with bivariate normal kernels. Extending Proposition 1 to this case and estimating what corresponds to Δ are then straightforward. For some applications it may also be of interest to consider the case of a location dependent covariate $z(\xi)$ so that e.g. an exponential term $\exp(z(\xi) \cdot \alpha)$ is multiplied to the right hand expression in (2), where α is a parameter of the same dimension as the covariate, $\cdot$ is the usual inner product, and a prior is imposed on α. In such a setup it may be much more complicated to extend Proposition 1.

The non-parametric Bayesian approach for the estimation of the first order term may also be interesting as a tool for a Bayesian analysis of a location dependent interaction function. This would then be a Bayesian analysis of one of the models considered by Hahn et al. (2003) as approximations of a point process with local scaling.

Overall our model captures many of the main features of the data considered, albeit one may question whether a pairwise interaction point process of the kind considered is appropriate for describing the second order properties. Further, we have obtained a somewhat better fit than in Nielsen (2000). It could be of interest to see if other kinds of pairwise interaction point process models would provide an even better fit. For example, we imposed the condition that $\varphi \leq 1$, but would it help to relax this condition and thereby allow for attraction between points at certain distances apart? Indeed there exist a few parametric models for such pairwise interaction point processes, e.g. the Lennard-Jones process (see e.g. Ruelle (1969)), but we are not aware of any flexible non-parametric Bayesian model similar to that in the present paper unless $\varphi \leq 1$. Furthermore, one could easily imagine many kind of generalisations. For example, models where the interaction is not isotropic, i.e. where the interaction function φ in (1) not only depends on the distance between pairs of points, and perhaps also is very inhomogeneous. Markov point process models with third or higher order interaction terms (see e.g. Møller & Waagepetersen (2003)) could be yet another generalisation, but a flexible non-parametric Bayesian model might be very complicated and hard to analyse even by MCMC methods.

Acknowledgement

We thank Linda S. Nielsen for supplying the data. Part of this work was done while the first author was financed by an EPSRC grant at Lancaster University. We have also been supported by the Danish Natural Science Research Council, grant 272-06-0442, "Point process modelling and statistical inference".

Appendix A

This appendix provides the details of the Metropolis-Hastings algorithm used for the simulation of the posterior when φ in (9) is replaced by $\tilde{\varphi}$ in Section 5.

The algorithm is more precisely a hybrid Metropolis-Hastings algorithm (also known as Metropolis within Gibbs), where at each iteration we choose at random one of eight possible types of Metropolis-Hastings updates, where the ith type of update is applied with probability p_i, say, where $p_i > 0$ and $\sum_{i=1}^{8} p_i = 1$. Each update follows the procedure outlined in Section 4: Suppose that (x, θ) is the current state of the algorithm. First, generate a proposal θ' from a proposal distribution specific to the given type of update. Next, conditional on θ', generate x' from $\pi(x'|\theta')$ using dominated CFTP. Finally, accept the proposal (x', θ') with probability $\min\{1, H(x', \theta'|x, \theta, y)\}$, and otherwise retain the current state, where H is the Hastings ratio given by (11).

The eight types of updates can be divided into three groups: Kernel updates, change point updates, and parameter updates. Details are given as follows, when $\theta = (\psi, \gamma, \chi, h)$.

1. The proposal is to add a kernel c' which is uniformly distributed on $[-\Delta, a + \Delta]$. Then $\theta' = (\psi \cup \{c'\}, \gamma, \chi, h)$, $p(\theta'|\theta, y) = 1/(a + 2\Delta)$, and $p(\theta|\theta', y) = 1/(1 + n(\psi))$.

2. The proposal is to remove a kernel c' which is uniformly randomly selected from ψ (if ψ is empty we do nothing). Then $\theta' = (\psi \backslash \{c'\}, \gamma, \chi, h)$, $p(\theta'|\theta, y) = 1/n(\psi)$, and $p(\theta|\theta', y) = 1/(a + 2\Delta)$.

3. The proposal is to replace a uniformly randomly selected kernel $c \in \psi$ using a symmetric random walk Metropolis (RWM) update (if ψ is empty we do nothing). Specifically, the proposed replacement c' is distributed according to a normal distribution with mean c and variance 10^{-4}. Then $\theta' = ((\psi \backslash \{c\}) \cup \{c'\}, \gamma, \chi, h)$, noticing that $H = 0$ if $c' \notin [-\Delta, a + \Delta]$, and the proposal density terms cancel in (11).

4. The proposal is to add (r', γ') to χ, where r' is uniform on $[0, r_{\max}]$ and γ' given r' is uniform on $[\gamma_i, \gamma_{i+1}]$, where i is chosen so that $r_i \leq r' \leq r_{i+1}$. Then $\theta' = (\psi, \gamma, \chi \cup \{(r', \gamma')\}, h)$, $p(\theta'|\theta, y) = 1/((\gamma_{i+1} - \gamma_i) r_{\max})$, and $p(\theta|\theta', y) = 1/(n(\chi) + 1)$.

5. The proposal is to remove a uniformly randomly chosen point (r_i, γ_i) from χ (if χ is empty we do nothing). Then $\theta' = (\psi, \gamma, \chi \backslash \{(r_i, \gamma_i)\}, h)$, $p(\theta'|\theta, y) = 1/n(\chi)$, and $p(\theta|\theta', y) = 1/((\gamma_{i+1} - \gamma_{i-1}) r_{\max})$.

6. The proposal is to replace a uniformly randomly selected point (r_i, γ_i) from χ by a point (r', γ') uniformly distributed on $[r_{i-1}, r_{i+1}] \times [\gamma_{i-1}, \gamma_{i+1}]$ (if χ is empty we do nothing). Then $\theta' = (\psi, \gamma, \chi \backslash \{(r_i, \gamma_i)\} \cup \{(r', \gamma')\}, h)$, and the proposal density terms cancel in (11).

7. The proposal is to replace γ by γ' which is normal distributed with mean γ and standard deviation 1. Then $\theta' = (\psi, \gamma', \chi, h)$, noticing that $H = 0$ if $\gamma' \leq 0$, and the proposal density terms cancel in (11).

8. The proposal is to replace the hard core distance h by h' which is uniformly distributed on $[h - 0.0005, h + 0.0005]$. Then $\theta' = (\psi, \gamma, \chi, h')$, noticing that $H = 0$ if $h' \notin [0, r_{\max}]$, and the proposal density terms cancel in (11).

The updates 1., 2., 4., and 5., involving adding or removing either a kernel or a change point, are applications of the spatial birth-and-death algorithm given by Geyer & Møller (1994). To retain detail balance it is assumed that the probability of proposing either a birth or a death are equal, i.e. $p_1 = p_2$ and $p_4 = p_5$. The results in Section 5 are obtained using $p_1 = 0.15$, $p_2 = 0.15$, $p_3 = 0.10$, $p_4 = 0.15$, $p_5 = 0.15$, $p_6 = 0.10$, $p_7 = 0.10$, and $p_8 = 0.10$, and the mean acceptance probabilities for each of the eight updates are 0.076, 0.077, 0.098, 0.028, 0.028, 0.026, 0.066, and 0.076, respectively.

References

Baddeley, A., Møller, J. & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns, *Statist. Neerlandica* **54**: 329–350.

Berthelsen, K. K. & Møller, J. (2003). Likelihood and non-parametric Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling, *Scand. J. Statist.* **30**: 549–564.

Berthelsen, K. K. & Møller, J. (2006). Bayesian analysis of Markov point processes, *in* A. Baddeley, P. Gregori, J. Mateu, R. Stoica & D. Stoyan (eds), *Case Studies in Spatial Point Process Models*, Springer Lecture Notes in Statistics 185, Spinger-Verlag, New York, pp. 85–97.

Cox, D. R. & Isham, V. (1980). *Point Processes*, Chapman & Hall, London.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*, second edn, Wiley, New York.

Diggle, P. J. (1985). A kernel method for smoothing point process data, *Appl. Stat.* **34**: 138–147.

Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*, second edn, Arnold, London.

Gelman, A. & Meng, X.-L. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling, *Statist. Sci.* **13**: 163–185.

Gelman, A., Meng, X.-L. & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies, *Statist. Sinica* **6**: 733–807.

Geyer, C. J. & Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes, *Scand. J. Statist.* **21**: 359–373.

Green, P. J. & Richardson, S. (2002). Hidden Markov models and disease mapping, *J. Am. Statist. Assoc.* **97**: 1055–1070.

Hahn, U., Jensen, E. B. V., van Lieshout, M.-C. & Nielsen, L. S. (2003). Inhomogeneous spatial point processes by location dependent scaling, *Adv. Appl. Prob.* **35**: 319–336.

Heikkinen, J. & Arjas, E. (1998). Non-parametric Bayesian estimation of a spatial Poisson intensity, *Scand. J. Statist.* **25**: 435–450.

Jensen, E. B. V. & Nielsen, L. S. (2000). Inhomogeneous Markov point processes by transformation, *Bernoulli* **6**: 761–782.

Kendall, W. S. & Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes, *Adv. Appl. Prob.* **32**: 844–865.

Møller, J. (2003). Shot noise Cox processes, *Adv. Appl. Prob.* **35**: 614–640.

Møller, J., Pettitt, A. N., Reeves, R. & Berthelsen, K. K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants, *Biometrika* **93**: 451–458.

Møller, J. & Waagepetersen, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*, Chapman and Hall/CRC, Boca Raton.

Nielsen, L. S. (2000). Modelling the position of cell profiles allowing for both inhomogeneity and interaction, *Image Anal. Stereol.* **19**: 183–187.

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

Ruelle, D. (1969). *Statistical Mechanics: Rigorous Results*, W.A. Benjamin, Reading, Massachusetts.

Stoyan, D., Kendall, W. S. & Mecke, J. (1995). *Stochastic Geometry and Its Applications*, second edn, Wiley, Chichester.