Handbook of Spatial Statistics

Jesper Møller

May 7, 2008

Contents

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Parametric methods

4.1 Introduction

This chapter considers *inference* procedures for *parametric spatial point process models*. The widespread use of sensible but ad hoc methods based on summary statistics of the kind studied in Chapter 4.3 have through the last two decades been supplied by *likelihood based methods* for parametric spatial point process models. The increasing development of such likelihood based methods, whether frequentist or Bayesian, has lead to more objective and efficient statistical procedures. When checking a fitted parametric point process model, summary statistics and residual analysis (Chapter 4.5) play an important role in combination with simulation procedures.

Simulation free estimation methods based on *composite likelihoods* or *pseudo likelihoods* are discussed in Section 4.3. *Markov chain Monte Carlo (MCMC)* methods have had an increasing impact on the development of simulation-based likelihood inference, where *maximum likelihood inference* is studied in Section 4.4, and *Bayesian inference* in Section 4.5. On one hand, as the development in computer technology and computational statistics continues, computationally-intensive simulation-based methods for likelihood inference probably will play a increasing role for statistical analysis of spatial point patterns. On the other hand, since larger and larger point pattern dataset are expected to be collected in the future, and the simulation free methods are much faster, they may continue to be of importance, at least at a preliminary stage of a parametric spatial point process analysis, where many different parametric models may quickly be investigated.

Much of this review is inspired by the monograph Møller and Waagepetersen (2004) and the discussion paper Møller and Waagepetersen (2007). Other recent textbooks related to the topic of this chapter include Baddeley, Gregori, Mateu, Stoica and Stoyan (2006), Diggle (2003), Illian, Penttinen, Stoyan and Stoyan (2008), and Van Lieshout (2000). Readers interested in background material on MCMC algorithms for spatial point processes are referred to Geyer and Møller

(1994), Geyer (1999), Møller and Waagepetersen (2004), and the references therein. Notice the comments and corrections to Møller and Waagepetersen (2004) at `www.math.aau.dk/~jm`.

4.2 Setting and notation

The methods in this chapter will be applied to *parametric models of Poisson, Cox, Poisson cluster, and Gibbs (or Markov) point processes*. These models also play a major role in Chapter 4.2, but the reader will be reminded about the definitions and some of the basic concepts of these models. Chapter 5.3 studies spatio-temporal point process models specified in terms on a conditional intensity (of another kind than the Papangelou conditional density which is of fundamental importance in the present chapter), while other kinds of spatio-temporal point process models, which are closely related to the Cox point process models considered in this chapter, can be found in e.g. Brix and Diggle (2001) and Brix and Møller (2001).

We mostly confine attention to planar point processes, but many concepts, methods, and results easily extend to $\mathbb{R}^d$ or a more general metric space, including multivariate and marked point process models. Chapter 4.6 treats statistics for multivariate and marked point process models.

We illustrate the statistical methodology with various application examples, where most are examples of inhomogeneous point patterns. Often the `R` package `spatstat` has been used, see Baddeley and Turner (2005, 2006). Software in `R` and `C`, developed by Rasmus Waagepetersen in connection to our paper Møller and Waagepetersen (2007), is available at `www.math.aau.dk/~rw/sppcode`.

We consider a planar spatial point process X, excluding the case of multiple points, meaning that X can viewed as a random subset of $\mathbb{R}^2$. We assume also that X is locally finite, i.e. $X \cap B$ is finite whenever $B \subset \mathbb{R}^2$ is finite.

We let $W \subset \mathbb{R}^2$ denote a bounded observation window of area $|W| > 0$. In most application examples W is a rectangular region. Usually we assume that just a single realization $X \cap W = x$ is observed, i.e. the data

$$x = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$$

is a spatial point pattern. Here the number of points, denoted $n(x) = n$, is finite and considered to be a realization of a non-negative discrete random variable (if $n = 0$ then x is the empty point configuration). Sometimes, including two of our application examples, two or more spatial point patterns are observed, and sometimes a hierarchical point process model may then be appropriate as illustrated in Sections 4.4.2 and 4.5.1 (see also Chapter 4.3 where multivariate point patterns are discussed).

In order to account for edge effects, we may assume that $X \cap W = x \cup y$ is observed so that 'x conditional on y' is conditionally independent of X outside W. The details are given in Sections 4.3.4 and 4.4.1.

Finally, $\mathbb{I}[\cdot]$ is an indicator function, and $\|\cdot\|$ denotes the usual distance in $\mathbb{R}^2$.

4.3 Simulation free estimation methods

This section reviews simple and quick estimation procedures based on various *estimating equations* for parametric models of spatial point processes. The methods are *simulation free* and the estimating equations are derived from a composite likelihood (Sections 4.3.1-4.3.2), or by a minimum contrast estimation procedure (Section 4.3.2), or by considering a pseudo likelihood function (Section 4.3.4).

4.3.1 Methods based on first order moment properties

Consider a spatial point process X with a parametric *intensity function* $\rho_\beta(\mathbf{s})$, where $\mathbf{s} \in \mathbb{R}^2$ and β is an unknown real d-dimensional parameter which we want to estimate. We assume that $\rho_\beta(\mathbf{s})$ is expressible in closed form. This is the case for many Poisson, Cox and Poisson cluster point process models, while it is intractable for Gibbs (or Markov) point processes (Chapter 4.2). Below we consider a composite likelihood function (Lindsay, 1988) based on the intensity function.

Recall that we may interpret $\rho_\beta(\mathbf{s})\,\mathrm{d}\mathbf{s}$ as the probability that precisely one point falls in an infinitesimally small region containing the location $\mathbf{s}$ and of area $\mathrm{d}\mathbf{s}$. Let $C_i,\ i \in I$, be a finite partitioning of the observation window W into disjoint cells C_i of small areas $|C_i|$. Define $N_i = \mathbb{I}[X \cap C_i \neq \emptyset]$ and

$$p_i(\beta) = \mathrm{P}_\beta(N_i = 1) \approx \rho_\beta(\mathbf{u}_i)|C_i|$$

where $\mathbf{u}_i$ denotes a representative point in C_i. Consider the product of marginal likelihoods for the Bernoulli trials N_i,

$$\prod_{i\in I} p_i(\beta)^{N_i}(1-p_i(\beta))^{1-N_i} \approx \prod_{i\in I}(\rho_\beta(\mathbf{u}_i)|C_i|)^{N_i}(1-\rho_\beta(\mathbf{u}_i)|C_i|)^{1-N_i}. \qquad (4.1)$$

In the right hand side of (4.1) we may neglect the factors $|C_i|$ in the first part of the product, since they cancel when we form likelihood ratios. Then, as the cell sizes $|C_i|$ tend to zero, under suitable regularity conditions the limit of the product of marginal likelihoods becomes

$$L_c(\beta; x) = \exp\left(-\int_W \rho_\beta(\mathbf{s})\,\mathrm{d}\mathbf{s}\right)\prod_{i=1}^{n}\rho_\beta(\mathbf{s}_i). \qquad (4.2)$$

We call $L_c(\beta; x)$ the *composite likelihood* function based on the intensity function. If X is a Poisson point process with intensity function $\rho_\beta(\mathbf{s})$, then $L_c(\beta; x)$ coincides with the likelihood function.

If there is a unique β which maximizes $L_c(\beta; x)$, we call it the *maximum composite likelihood estimate* (based on the intensity function). The corresponding estimating function $s_c(\beta; x)$ is given by the derivative of $\log L_c(\beta; x)$ with respect to β,

$$s_c(\beta; x) = \sum_{i=1}^{n} \mathrm{d}\log\rho_\beta(\mathbf{s}_i)/\mathrm{d}\beta - \int_W (\mathrm{d}\log\rho_\beta(\mathbf{s})/\mathrm{d}\beta)\rho_\beta(\mathbf{s})\,\mathrm{d}\mathbf{s}. \qquad (4.3)$$

The estimating equation $s_c(\beta; x) = 0$ is unbiased (assuming in (4.3) that $(\mathrm{d}/\mathrm{d}\beta)$ $\int_W \cdots = \int_W (\mathrm{d}/\mathrm{d}\beta) \cdots$). Asymptotic properties of maximum composite likelihood estimators are investigated in Waagepetersen (2007) and Waagepetersen and Guan (2007). For a discussion of asymptotic results for maximum likelihood estimates of Poisson process models, see Rathbun and Cressie (1994) and Waagepetersen (2007).

The maximum composite likelihood estimate can easily be determined using `spatstat`, provided $\rho_\beta(\mathbf{s})$ is of the *log linear form*

$$\log \rho_\beta(\mathbf{s}) = \beta^T z(\mathbf{s}) \tag{4.4}$$

where $z(\mathbf{s})$ is a real function of the same dimension as β. In practice $z(\mathbf{s})$ is often a *covariate*. This covariate may only be partially observed on a grid of points, and hence some interpolation technique may be needed (Rathbun, 1996; Rathbun, Shiffman and Gwaltneyet, 2007; Waagepetersen, 2008). An example is considered in Section 4.3.3.

We refer to a *log linear Poisson process* when X is a Poisson process with intensity function of the form (4.4). For many *Cox process* models, the intensity function is also of the log linear form (4.4). Specifically, let $Y = \{Y(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$ be a spatial process where each $Y(\mathbf{s})$ is a real random variable with mean one, and let X conditional on $Y(\mathbf{s})$ be a Poisson process with intensity function

$$\Lambda(\mathbf{s}) = \exp(\beta^T z(\mathbf{s}))Y(\mathbf{s}). \tag{4.5}$$

Then (4.4) is satisfied. Usually Y is not observed, and the distribution of Y may depend on another parameter ψ, which may be estimated by another method as discussed in Section 4.3.2.

4.3.2 Methods based on second order moment properties

Let the situation be as in Section 4.3.1. Consider a parametric model for the *pair correlation function* g_ψ or another second order characteristic such as the *(inhomogeneous) K-function* K_ψ (Baddeley, Møller and Waagepetersen, 2000; see also Chapter 4.3). We assume that β and ψ are variation independent, that is, $(\beta, \psi) \in B \times \Psi$, where $B \subseteq \mathbb{R}^p$ and $\Psi \subseteq \mathbb{R}^q$.

Recall that $\rho^{(2)}_{\beta,\psi}(\mathbf{s}, \mathbf{t}) = \rho_\beta(\mathbf{s})\rho_\beta(\mathbf{t})g_\psi(\mathbf{s}, \mathbf{t})$ is the *second order product density*, and we may interpret $\rho^{(2)}_{\beta,\psi}(\mathbf{s}, \mathbf{t})\,\mathrm{d}\mathbf{s}\,\mathrm{d}\mathbf{t}$ as the probability of observing a point in each of two infinitesimally small regions containing $\mathbf{s}$ and $\mathbf{t}$ and of areas $\mathrm{d}\mathbf{s}$ and $\mathrm{d}\mathbf{t}$, respectively. Using the same principle as in Section 4.3.1 but considering now pairs of cells C_i and C_j, $i \neq j$, we can derive a *composite likelihood* $L_c(\beta, \psi)$ based on the second order product density. Plugging in an estimate $\hat{\beta}$, e.g. the maximum composite likelihood estimate based on the intensity function, we obtain a function $L_c(\hat{\beta}, \psi)$ which may be maximized to obtain an estimate of ψ. See Møller and Waagepetersen (2007).

Minimum contrast estimation is a more common estimation procedure, where the idea is to minimize a 'contrast' (or 'distance') between e.g. K_ψ and its

non-parametric counterpart $\hat{K}(r)$ (Chapter 4.3), thereby obtaining a minimum contrast estimate. For instance, ψ may be estimated by minimizing the contrast

$$\int_a^b \left(\hat{K}(r)^\alpha - K_\psi(r)^\alpha\right)^2 \mathrm{d}r \tag{4.6}$$

where $0 \leq a < b < \infty$ and $\alpha > 0$ are chosen on an ad hoc basis, see e.g. Diggle (2003) and Møller and Waagepetersen (2004). Theoretical properties of minimum contrast estimators are studied in Heinrich (1992).

These 'simulation-free' estimation procedures are fast and computationally easy, but the disadvantage is that we have to specify tuning parameters such as a, b, α in (4.6).

4.3.3 Example: tropical rain forest trees

Figure 4.1 provides an example of an *inhomogeneous point pattern* where the methods described in Sections 4.3.1-4.3.2 apply. The figure shows the locations of rain forest trees in a rectangular observation window W of size 500×1000 m. This point pattern together with another point pattern of another kind of trees have previously been analyzed in Waagepetersen (2007) and Møller and Waagepetersen (2007). They are just a small part of a much larger data set comprising hundreds of thousands of trees belonging to hundreds of species (Hubbell and Foster, 1983; Condit, Hubbell and Foster, 1996; Condit, 1998). Figure 4.2 shows two kinds of covariates z_1 (altitude) and z_2 (norm of altitude gradient) which are measured on a 100×200 square grid, meaning that we approximate the altitude and the norm of altitude gradient to be constant on each of 100×200 squares of size 5×5 m.

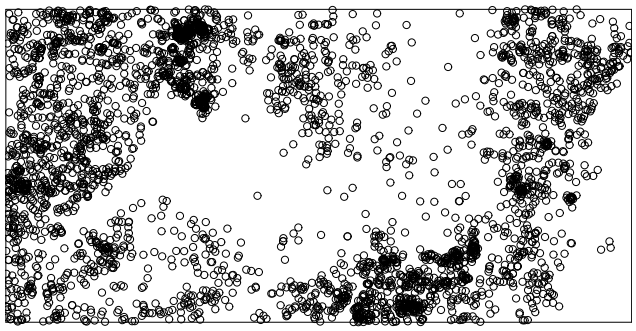

Figure 4.1: Locations of 3605 *Beilschmiedia pendula Lauraceae* trees observed within a 500×1000 m region at Barro Colorado Island.

A plot of a non-parametric estimate of the inhomogeneous K-function (omitted here) confirms that the point pattern in Figure 4.1 is clustered. This clustering may be explained by the covariates in Figure 4.2, by other unobserved covariates, and by tree reproduction by seed dispersal. We therefore assume an

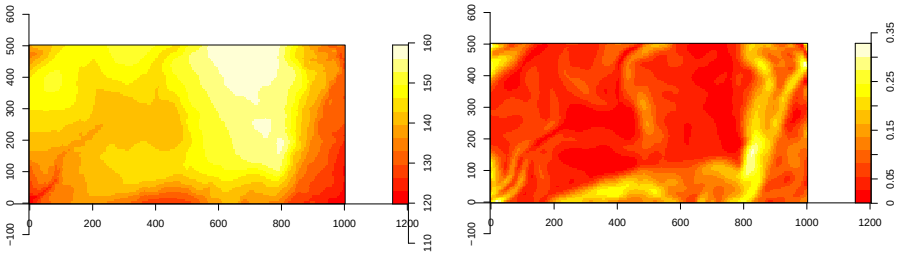

Figure 4.2: Rain forest trees: the covariates z_1 (altitude; left panel) and z_2 (norm of altitude gradient; right panel) are recorded on a 5 by 5 m grid (the units on the axes are in meters).

inhomogeneous Cox process model as specified by (4.5) with $\beta = (\beta_0, \beta_1, \beta_2)^T$ and $z = (z_0, z_1, z_2)^T$, where $z_0 \equiv 1$ so that β_0 is interpreted as an intercept. Moreover, Y in (4.5) is modelled by a stationary shot noise process with mean one, that is,

$$Y(\mathbf{s}) = \frac{1}{\omega\sigma^2} \sum_{\mathbf{t}\in\Phi} k((\mathbf{s}-\mathbf{t})/\sigma) \tag{4.7}$$

where Φ is a stationary Poisson process with intensity $\omega > 0$, $k(\cdot)$ is a density function with respect to Lebesgue measure, and $\sigma > 0$ is a scaling parameter. We call X an *inhomogeneous shot noise Cox process* (Møller, 2003; Waagepetersen, 2007; Møller and Waagepetersen, 2007). Finally, as in a modified Thomas process (Thomas, 1949), we assume that $k(x) = \exp(-\|x\|^2/2)/(2\pi)$ is a bivariate normal kernel. For short we then refer to X as an *inhomogeneous Thomas process*.

For β we obtain the maximum composite likelihood estimate $(\hat\beta_0, \hat\beta_1, \hat\beta_2) = (-4.989, 0.021, 5.842)$ (under the Poisson model this is the maximum likelihood estimate). Assuming asymptotic normality (Waagepetersen, 2007), 95% confidence intervals for β_1 and β_2 under the fitted inhomogeneous Thomas process are $[-0.018, 0.061]$ and $[0.885, 10.797]$, respectively, while much more narrow intervals are obtained under the fitted Poisson process ($[0.017, 0.026]$ and $[5.340, 6.342]$).

An unbiased estimate of the inhomogeneous K-function at distance $r > 0$ is given by

$$\sum_{i,j=1,\ldots,n:\, i\neq j} \frac{\mathbb{I}[\|\mathbf{s}_i - \mathbf{s}_j\| \le r]}{\rho(\mathbf{s}_i)\rho(\mathbf{s}_j)|W \cap (W + \mathbf{s}_i - \mathbf{s}_j)|}$$

where $W + \mathbf{s}$ denotes W translated by $\mathbf{s}$, and $|W \cap (W + \mathbf{s}_i - \mathbf{s}_j)|$ is an *edge correction factor*, which is needed since we sum over all pairs of points observed within W. In practice we need to plug in an estimate of $\rho(\mathbf{s}_i)\rho(\mathbf{s}_j)$, and we use the parametric estimate $\rho_{\hat\beta}(\mathbf{s}_i)\rho_{\hat\beta}(\mathbf{s}_j)$ with $\hat\beta$ the estimate obtained above. Let $\hat K(r)$ denote the resulting estimate of $K(r)$. Using the minimum contrast estimation procedure based on (4.6) with $a = 0$, $b = 100$, and $\alpha = 1/4$, we obtain $(\hat\omega, \hat\sigma) = (8 \times 10^{-5}, 20)$.

Estimation of this inhomogeneous Thomas process and an inhomogeneous *log Gaussian Cox process*, i.e. when $\log Y$ in (4.5) is a Gaussian process (see Møller, Syversveen and Waagepetersen, 1998, and Chapter 4.2), and their corresponding estimated K-functions are further considered in Møller and Waaagepetersen (2007).

4.3.4 Pseudo likelihood

The *maximum pseudo likelihood estimate* is a simple and computationally fast but less efficient alternative to the maximum likelihood estimate. In the special case of a parametric Poisson point process model, the two kinds of estimates coincide. Since the pseudo likelihood function is expressed in terms of the Papangelou conditional intensity, pseudo likelihood estimation is particular useful for *Gibbs (or Markov) point processes*, while it is in general not useful for Cox and Poisson cluster processes.

We recall first the definition of the Papangelou conditional intensity in the case where X restricted to W has a *parametric density* $f_\theta(x)$ with respect to the Poisson process on W with unit intensity (Chapter 4.2). Let $x = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset W$ denote an arbitrary finite point configuration in W, and $\mathbf{s}$ an arbitrary location in $W \setminus x$. Assume that $f_\theta(x)$ is *hereditary*, meaning that $f_\theta(x \cup \{\mathbf{s}\}) > 0$ implies that $f_\theta(x) > 0$. For $f_\theta(x) > 0$, define the *Papangelou conditional intensity* by

$$\lambda_\theta(\mathbf{s}, x) = f_\theta(x \cup \{\mathbf{s}\})/f_\theta(x). \tag{4.8}$$

We may interpret $\lambda_\theta(\mathbf{s}, x)\,\mathrm{d}\mathbf{s}$ as the conditional probability that there is a point of the process in an infinitesimally small region containing $\mathbf{s}$ and of area $\mathrm{d}\mathbf{s}$ given that the rest of the point process coincides with x. How we define $\lambda_\theta(\mathbf{s}, x)$ if $f_\theta(x) = 0$ turns out not to be that important, but for completeness let us set $\lambda_\theta(\mathbf{s}, x) = 0$ if $f_\theta(x) = 0$. In the special case of a Poisson process with intensity function $\rho_\theta(\mathbf{s})$, we simply have $\lambda_\theta(\mathbf{s}, x) = \rho_\theta(\mathbf{s})$. In the case of a Gibbs (or Markov) point process, $\lambda_\theta(\mathbf{s}, x)$ depends only on x through its neighbours to $\mathbf{s}$ (see Chapter 4.2), and the intractable normalizing constant of the density cancel in (4.8).

The pseudo likelihood can then be derived by a limiting argument similar to that used for deriving the composite likelihood in (4.2), the only difference being that we replace $p_i(\beta)$ in (4.1) by the conditional probability

$$p_i(\theta) := \mathrm{P}_\theta(N_i = 1 | X \setminus C_i = x \setminus C_i) \approx \lambda_\theta(\mathbf{u}_i, x \setminus C_i)|C_i|.$$

Under mild conditions (Besag, Milne and Zachary, 1982; Jensen and Møller, 1991) the limit becomes the *pseudo likelihood* function

$$L_p(\theta; x) = \exp\left(-\int_W \lambda_\theta(\mathbf{s}, x)\,\mathrm{d}\mathbf{s}\right) \prod_{i=1}^{n} \lambda_\theta(\mathbf{s}_i, x) \tag{4.9}$$

which was first introduced in Besag (1977). Clearly, for a Poisson process with a parametric intensity function, the pseudo likelihood is the same as the likelihood.

The *pseudo score* is the derivative of $\log L_p(\theta; x)$ with respect to θ, that is,

$$s(\theta; x) = \sum_{i=1}^{n} \mathrm{d} \log \lambda_\theta(\mathbf{s}_i, x)/\mathrm{d}\theta - \int_W (\mathrm{d} \log \lambda_\theta(\mathbf{s}, x)/\mathrm{d}\theta)\lambda_\theta(\mathbf{s}, x)\, \mathrm{d}\mathbf{s}. \tag{4.10}$$

This provides an unbiased estimating equation $s(\theta; x) = 0$ (assuming in (4.10) that $(\mathrm{d}/\mathrm{d}\theta) \int_W \cdots = \int_W (\mathrm{d}/\mathrm{d}\theta) \cdots$). This can be solved using `spatstat` if λ_θ is of a log linear form similar to that in (4.4), that is,

$$\log \lambda_\theta(\mathbf{s}, x) = \beta^T t(\mathbf{s}, x) \tag{4.11}$$

(Baddeley and Turner, 2000).

Suppose that X may have points outside W, and we do not know its marginal density $f_\theta(x)$ on W. To account for *edge effects*, assume a *spatial Markov property* is satisfied. Specifically, suppose there is a region $W_{\ominus R} \subset W$ such that conditional on $X \cap (W \setminus W_{\ominus R}) = y$, we have that $X \cap W_{\ominus R}$ is independent of $X \setminus W$, and we know the conditional density $f_\theta(x|y)$ of $X \cap W_{\ominus R}$ given $X \cap (W \setminus W_{\ominus R}) = y$, where $f_\theta(\cdot|y)$ is hereditary. Here the notation $W_{\ominus R}$ refers to the common case where $\mathbf{X}$ is a Gibbs (or Markov) point process with a finite interaction radius R (see Chapter 4.2), in which case $W_{\ominus R}$ is naturally given by the W eroded by a disc of radius R, that is,

$$W_{\ominus R} = \{\mathbf{s} \in W : \|\mathbf{s} - \mathbf{t}\| \leq R \text{ for all } \mathbf{t} \in W\}. \tag{4.12}$$

For $\mathbf{s} \in W_{\ominus R}$, exploiting the spatial Markov property, the Papangelou conditional intensity is seen not to depend on points from $X \setminus W$, and it is given by replacing $f_\theta(x)$ by $f_\theta(x|y)$ in the definition (4.8). We denote this Papangelou conditional intensity by $\lambda_\theta(\mathbf{s}, x \cup y)$. Note that $\lambda_\theta(\mathbf{s}, x \cup y)$ depends only on $x \cup y$ through its neighbours to $\mathbf{s}$, and all normalizing constants cancel. Consequently, we need only to specify $f_\theta(\cdot|y)$ up to proportionality, and the pseudo likelihood $L_p(\theta; x \cup y)$ is given by (4.9) when $\lambda_\theta(\mathbf{s}, x)$ is replaced by $\lambda_\theta(\mathbf{s}, x \cup y)$. The pseudo score $s(\theta; x \cup y)$ is obtained as the derivative of $\log L_p(\theta; x \cup y)$ with respect to θ, and it provides an unbiased estimating equation $s(\theta; x \cup y) = 0$.

For an application example of maximum pseudo likelihood, see Section 4.4.2. Asymptotic results for maximum pseudo likelihood estimates are established in Jensen and Møller (1991), Jensen and Kunsch (1994), and Mase (1995, 1999). Alternatively a *parametric bootstrap* can be used, see e.g. Baddeley and Turner (2000).

4.4 Simulation-based maximum likelihood inference

For *Poisson process* models, computation of the likelihood function is usually easy, cf. Section 4.3.1. For Gibbs (or Markov) point process models, the likelihood contains an unknown normalizing constant, while for Cox process models,

the likelihood is given in terms of a complicated integral. Using MCMC methods, it is now becoming quite feasible to compute accurate approximations of the likelihood function for Gibbs and Cox process models as discussed in Sections 4.4.1 and 4.4.3. However, the computations may be time consuming and standard software is yet not available.

4.4.1 Gibbs point processes

Consider a parametric model for a spatial point process X, where X restricted to W has a *parametric density* $f_\theta(x)$ with respect to the Poisson process on W with unit intensity. For simplicity and specificity, assume that $f_\theta(x)$ is of *exponential family* form

$$f_\theta(x) = \exp(t(x)^T\theta)/c_\theta \tag{4.13}$$

where $t(x)$ is a real function of the same dimension as the real parameter θ, and c_θ is a normalizing constant. In general, apart from the special case of a Poisson process, c_θ is not 'known', i.e. c_θ has no closed form expression. Equation (4.13) holds if the Papangeleou conditional intensity $\lambda_\theta(\mathbf{s}, x)$ is of the log linear form (4.11). This is the case for many *Gibbs (or Markov) point processes* when the interaction radius $R < \infty$ is known. Examples include most *pairwise interaction point processes* such as the Strauss process, and more complicated interaction point processes such as the area-interaction point process, see Chapter 4.2.

From (4.13) we obtain the score function $u(\theta; x)$ and the observed information $j(\theta)$,

$$u(\theta; x) = t(x) - \mathrm{E}_\theta t(X), \quad j(\theta) = \mathrm{Var}_\theta t(X),$$

where E_θ and Var_θ denote expectation and variance with respect to $X \sim f_\theta$. Let θ_0 denote a fixed reference parameter value. The score function and observed information may be evaluated using the *importance sampling formula*

$$\mathrm{E}_\theta k(X) = \mathrm{E}_{\theta_0}\left[k(X)\exp\left(t(X)^T(\theta-\theta_0)\right)\right]/(c_\theta/c_{\theta_0}) \tag{4.14}$$

with $k(X)$ given by $t(X)$ or $t(X)t(X)^T$. For $k \equiv 1$, we obtain

$$c_\theta/c_{\theta_0} = \mathrm{E}_{\theta_0}\left[\exp\left(t(X)^T(\theta-\theta_0)\right)\right]. \tag{4.15}$$

Approximations of the likelihood ratio $f_\theta(x)/f_{\theta_0}(x)$, score, and observed information can be obtained by Monte Carlo approximation of the expectations $\mathrm{E}_{\theta_0}[\cdots]$ using MCMC samples from f_{θ_0}. Here, to obtain an approximate maximum likelihood estimate, Monte Carlo approximations may be combined with Newton-Raphson updates. Furthermore, if we want to test a submodel, approximate p-values based on the likelihood ratio statistic or the Wald statistic can be derived by MCMC methods. See Geyer and Møller (1994), Geyer (1999), and Møller and Waagepetersen (2004).

The *path sampling identity* (Gelman and Meng, 1998)

$$\log(c_\theta/c_{\theta_0}) = \int_0^1 \mathrm{E}_{\theta(s)} t(X)(\mathrm{d}\theta(s)/\mathrm{d}s)^T \mathrm{d}s \tag{4.16}$$

provides an alternative and often numerically more stable way of computing a ratio of normalizing constants. Here $\theta(s)$ is a differentiable curve, e.g. a straight line segment, connecting $\theta_0 = \theta(0)$ and $\theta = \theta(1)$. The log ratio of normalizing constants is approximated by evaluating the outer integral in (4.16) using e.g. the trapezoidal rule and the expectation using MCMC methods (Berthelsen and Møller, 2003; Møller and Waagepetersen, 2004).

For a Gibbs point process with unknown interaction radius R, the likelihood function is usually not differentiable as a function of R. Therefore maximum likelihood estimates of R are often found using a profile likelihood approach, where for each fixed value of R we maximize the likelihood as discussed above. Examples are given in Møller and Waagepetersen (2004).

If X may have points outside W, and we do not know its marginal density $f_\theta(x)$ on W, we may account for *edge effects* by exploiting the spatial Markov property (Section 4.3.4), using the smaller observation window $W_{\ominus R}$ given by (4.12). If $f_\theta(x|y)$ denotes the conditional density of $X \cap W_{\ominus R} = x$ given $X \cap (W \setminus W_{\ominus R}) = y$, the likelihood function

$$L(\theta; x) = \mathrm{E}_\theta f_\theta\left(x | X \cap (W \setminus W_{\ominus R})\right)$$

may be computed using a missing data approach, see Geyer (1999) and Møller and Waagepetersen (2004). A simpler but less efficient alternative is the *border method*, considering the conditional likelihood function

$$L(\theta; x|y) = f_\theta(x|y)$$

where the score, observed information, and likelihood ratios may be computed by analogy with the case above based on (4.14). See Møller and Waagepetersen (2004) for a discussion of these and other approaches for handling edge effects.

Asymptotic results for maximum likelihood estimates of Gibbs point process models are reviewed in Møller and Waagepetersen (2004) but these results are derived under restrictive assumptions of stationarity and weak interaction. According to standard asymptotic results, the inverse observed information provides an approximate covariance matrix of the maximum likelihood estimate, and log likelihood ratio and Wald statistics are asymptotically χ^2-distributed. If one is suspicious about the validity of the asymptotic approach, an alternative is to use a parametric bootstrap. See Møller and Waagepetersen (2004).

4.4.2 Example: ants nests

Figure 4.3 shows two point patterns of *ants nests* which are of two types, *Messor wasmanni* and *Cataglyphis bicolor*, see Harkness and Isham (1983). The interaction between the two types of ants nests is of main interest for this data set. Notice the rather atypical polygonal observation window W given in Figure 4.3.

The *Catagplyphis* ants feed on dead *Messors* and hence the positions of *Messor* nests might affect the choice of sites for *Cataglyphis* nests, while the *Messor* ants are believed not to be influenced by presence or absence of *Cataglyphis*

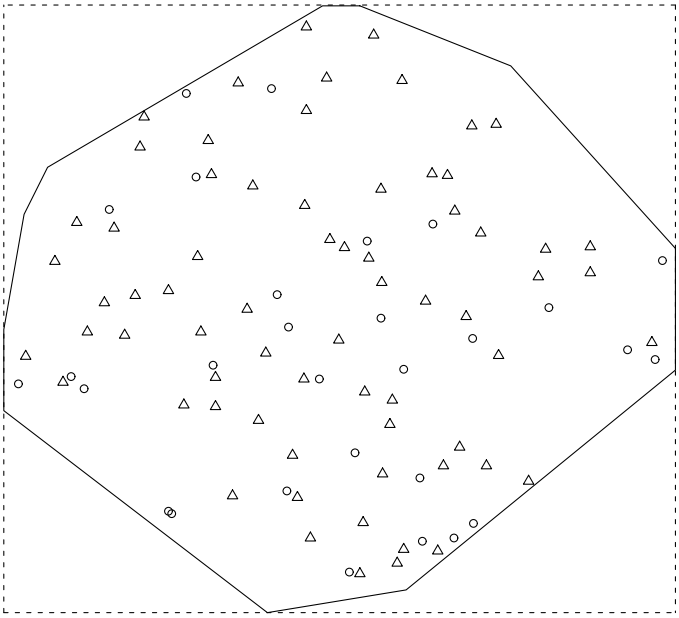

Figure 4.3: Locations of nests for *Messor* (triangles) and *Cataglyphis* (circles) ants. The observation window W is polygonal (solid line), and the enclosing rectangle for W (dashed line) is 414.5 ft by 383 ft.

ants when choosing sites for their nests. Högmander and Särkkä (1999) therefore specified a *hierarchical model* based on first a point process model for the *Messor* nests, and second a point process model for the *Cataglyphis* nests given the *Messor* nests. Both types of models are *pairwise interaction point process* models, with the log Papangelou conditional intensity of the form

$$\log \lambda(\mathbf{s}, x) = U(\mathbf{s}) + \sum_{i=1}^{n} V(\|\mathbf{s} - \mathbf{s}_i\|)$$

for $x = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset W$ and $\mathbf{s} \notin x$, where $U(\mathbf{s})$ and $V(\|\mathbf{s} - \mathbf{s}_i\|)$ are real functions called the first respective second order potential. In other words, if X is such a pairwise interaction point process, then X has density

$$f(x) \propto \exp\left(\sum_{i=1}^{n} U(\mathbf{s}_i) + \sum_{1 \le i < j \le n} V(\|\mathbf{s}_i - \mathbf{s}_j\|) \right)$$

with respect to the Poisson process on W with intensity one. Furthermore, the pairwise interaction process models are so-called *Strauss processes with hard*

cores specified as follows. For distances $t > 0$, define

$$V(t;r) = \begin{cases} -\infty & \text{if } t \leq r \\ 1 & \text{if } r < t \leq R \\ 0 & \text{otherwise} \end{cases}$$

where $R \geq 0$ is the interaction range, $r \in [0, R)$ denotes a hard core distance (or no hard core if $r = 0$), and $\exp(-\infty) = 0$. First, for the *Messor* nests, the Strauss process with hard core r_M is given by first and second order potentials

$$U_{M1}(\{\mathbf{s}\}) = \beta_M, \quad U_{M2}(\{\mathbf{s}_i, \mathbf{s}_j\}) = \psi_M V(\|\mathbf{s}_i - \mathbf{s}_j\|; r_M).$$

Thus the conditional intensity for a putative *Messor* nest at a location $\mathbf{s}$ is zero if an existing *Messor* nest occurs within distance r_M from $\mathbf{s}$, and otherwise the log conditional density is given by the sum of β_M and ψ_M times the number of neighbouring *Messor* nests within distance R. Second, conditional on the pattern x_M of *Messor* nests, the *Cataglyphis* nests are modelled as an inhomogeneous Strauss process with one hard core r_{CM} to the *Messor* nests and another hard core r_C between the *Cataglyphis* nests, i.e. using potentials

$$U_{C1}(\{\mathbf{s}\}) = \beta_C + \psi_{CM} \sum_{i=1}^{n} V(\|\mathbf{s} - \mathbf{s}_i\|; r_{CM}), \quad U_{C2}(\{\mathbf{s}_i, \mathbf{s}_j\}) = \psi_C V(\|\mathbf{s}_i - \mathbf{s}_j\|; r_C).$$

We use the maximum likelihood estimates $r_M = 9.35$ and $r_C = 2.45$ (distances are measured in ft), which are given by the observed minimum interpoint distances in the two types of point patterns. Using positive hard cores r_M and r_C may be viewed as an ad hoc approach to obtain a model which is well-defined for all real values of the parameters β_M, β_C, ψ_M, ψ_{CM}, and ψ_C, whereby both repulsive and attractive interaction within and between the two types of ants can be modelled. However, as noted by Møller (1994) and Geyer and Thompson (1995), the Strauss hard core process is a poor model for clustering due to the following 'phase transition property': for positive values of the interaction parameter, except for a narrow range of values, the distribution will either be concentrated on point patterns with one dense cluster of points or in 'Poisson-like' point patterns.

In contrast to Högmander and Särkkä (1999), we find it natural to let $r_{CM} = 0$, meaning there is no hard core between the two types of ants nests. Further, for comparison we fix R at the value 45 used in Högmander and Särkkä (1999), though pseudo likelihood computations indicate that a more appropriate interaction range would be 15. In fact, Högmander and Särkkä (1999) considered a subset of the data in Figure 4.3 within a rectangular region, and they conditioned on the observed number of points for the two species when computing maximum likelihood and maximum pseudo likelihood estimates, whereby the parameters β_M and β_C vanish. Instead we fit the hierarchical model to the full data set, and we do not condition on the observed number of points.

We first correct for edge effects by conditioning on the data in $W \setminus W_{\ominus 45}$, where $W_{\ominus 45}$ denotes the points within W with distance less than 45 to the

boundary of W. Using `spatstat`, the maximum pseudo likelihood estimate (MPLE) of (β_M, ψ_M) is $(-8.21, -0.09)$, indicating (weak) repulsion between the *Messor* ants nests. Without edge correction, we obtain a rather similar MPLE $(-8.22, -0.12)$. The edge corrected MPLE of $(\beta_C, \psi_{CM}, \psi_C)$ is $(-9.51, 0.13, -0.66)$, indicating a positive association between the two species and repulsion within the *Cataglyphis* nests. If no edge correction is used, the MPLE for $(\beta_C, \psi_{CM}, \psi_C)$ is $(-9.39, 0.04, -0.30)$. Högmander and Särkkä (1999) also found a repulsion within the *Cataglyphis* nests, but in contrast to our result a weak repulsive interaction between the two types of nests. This may be explained by the different modelling approach in Högmander and Särkkä (1999), where the smaller observation window excludes a pair of very close *Cataglyphis* nests, and where also the conditioning on the observed number of points in the two point patterns may make a difference.

No edge correction is used for our maximum likelihood estimates (MLE's). The MLE's $\hat{\beta}_M = -8.39$ and $\hat{\psi}_M = -0.06$ again indicate a weak repulsion within the *Messor* nests, and the MLE's $\hat{\beta}_C = -9.24$, $\hat{\psi}_{CM} = 0.04$, and $\hat{\psi}_C = -0.39$ also indicate positive association between *Messor* and *Cataglyphis* nests, and repulsion within the *Cataglyphis* nests. Confidence intervals for ψ_{CM}, when the asymptotic variance estimate is based on observed information or a parametric bootstrap, are $[-0.20, 0.28]$ (observed information) and $[-0.16, 0.30]$ (parametric bootstrap).

The differences between the MLE and the MPLE (without edge correction) seem rather minor. This is also the experience for MLE's and corresponding MPLE's in Møller and Waagepetersen (2004), though differences may appear in cases with a strong interaction.

4.4.3 Cluster and Cox processes

This section considers maximum likelihood inference for cluster and Cox process models. This is in general complicated and computionally more demanding than for Gibbs (or Markov) point processes.

For example, consider the case of an *inhomogeneous shot noise Cox process* X as defined by (4.5) and (4.7). We can interpret this as a *Poisson cluster process* as follows. The points in the stationary Poisson process Φ in (4.7) specify the centres of the clusters. Conditional on Φ, the clusters are independent Poisson processes, where the cluster associated to $\mathbf{t} \in \Phi$ has intensity function

$$\lambda_\theta(\mathbf{s}|\mathbf{t}) = \exp(\beta^T z(\mathbf{s})) \frac{1}{\omega\sigma^2} k((\mathbf{s}-\mathbf{t})/\sigma), \quad \mathbf{s} \in \mathbb{R}^2,$$

where $\theta = (\beta, \omega, \sigma)$. Finally, X consists of the union of all cluster points.

With probability one, X and Φ are disjoint. Moreover, in applications Φ is usually unobserved. In order to deal with *edge effects*, consider a bounded region $W_{\text{ext}} \supseteq W$ so that it is very unlikely that clusters associated to centres outside W_{ext} have points falling in W (see Brix and Kendall, 2002, and Møller, 2003). We approximate then $X \cap W$ by the union of clusters with centres in $\Psi := \Phi \cap W_{\text{ext}}$. Let $f(x|\psi)$ denote the conditional density of $X \cap W$ given $\Psi = \psi$,

where the density is with respect to the Poisson process on W with intensity one. For $x = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$,

$$f_\theta(x|\psi) = \exp\left(|W| - \int_W \sum_{\mathbf{t}\in\psi} \lambda_\theta(\mathbf{s}|\mathbf{t})\,\mathrm{d}\mathbf{s}\right) \prod_{i=1}^{n} \lambda_\theta(\mathbf{s}_i|\mathbf{t}) \tag{4.17}$$

and the likelihood based on observing $X \cap W = x$ is

$$L(\theta; x) = \mathrm{E}_\omega f_\theta(x|\Psi) \tag{4.18}$$

where the expectation is with respect to the Poisson process Ψ on W_{ext} with intensity ω. As this likelihood has no closed form expression, we may consider Ψ as *missing data* and use MCMC methods for finding an approximate maximum likelihood estimate, see Møller and Waagepetersen (2004). Here one important ingredient is an MCMC simulation algorithm for the conditional distribution of Ψ given $X \cap W = x$. This conditional distribution has density

$$f_\theta(\psi|x) \propto f_\theta(x|\psi) f_\omega(\psi) \tag{4.19}$$

where

$$f_\omega(\psi) = \exp\left(|W_{\mathrm{ext}}|(1-\omega)\right) \omega^{n(\psi)} \tag{4.20}$$

is the density of Ψ. For conditional simulation from (4.19), we use a *birth-death type Metropolis-Hastings algorithm* studied in Møller (2003).

For a *log Gaussian Cox process* model, the simulation-based maximum likelihood approach is as above except for the following. To specify the density of the Poisson process $X \cap W|Y$, since $\log Y$ in (4.5) is a Gaussian process, we need only to consider $Y(\mathbf{s})$ for $\mathbf{s} \in W$. Hence, in contrast to above, edge effects is not a problem, and the conditional density of $X \cap W$ given Y is

$$f(x|Y(\mathbf{s}),\, \mathbf{s} \in W) = \exp\left(|W| - \int_W \exp(Y(\mathbf{s}))\,\mathrm{d}\mathbf{s} + \sum_{i=1}^{n} Y(\mathbf{s}_i)\right). \tag{4.21}$$

However, when evaluating the integral in (4.21) and when simulating from the conditional distribution of Y on W given $X \cap W = x$, we need to approximate Y on W by a finite-dimensional log Gaussian random variable $Y_I = (Y(\mathbf{u}_i),\, i \in I)$ corresponding to a finite partition $\{C_i,\, i \in I\}$ of W, where $\mathbf{u}_i$ is a representative point of the cell C_i and we use the approximation $Y(\mathbf{s}) \approx Y(\mathbf{u}_i)$ if $\mathbf{s} \in C_i$. For simulation from the conditional distribution of Y_I given $X \cap W = x$, we use a *Langevin-Hastings algorithm* (also called a *Metropolis adjusted Langevin algorithm*), see Møller, Syversveen and Waagepetersen (1998) and Møller and Waagepetersen (2004).

For the shot noise Cox process model considered above, the likelihood (4.18) and its MCMC approximation are complicated functions of θ, possibly with many local modes. Similarly, in the case of a log Gaussian Cox process model. Careful maximization procedures are therefore needed when finding the (approximate) maximum likelihood estimate. Further details, including examples and specific algorithms of the *MCMC missing data approach* for shot noise and log Gaussian Cox processes, are given in Møller and Waagepetersen (2004, 2007).

4.5 Simulation-based Bayesian inference

A *Bayesian approach* often provides a flexible framework for incorporating prior information and analyzing spatial point process process models. Section 4.5.1 considers an application example of a Poisson process, where a Bayesian approach is obviously more suited than a maximum likelihood approach. Bayesian analysis for cluster and Cox processes is discussed in Section 4.5.2, while Section 4.5.3 considers Gibbs (or Markov) point processes. In the latter case a Bayesian analysis is more complicated because of the unknown normalizing constant appearing in the likelihood term of the posterior density.

4.5.1 Example: reseeding plants

Armstrong (1991) considered the locations of 6378 plants from 67 species on a 22 m by 22 m observation window W in the south western area of Western Australia. The plants have adapted to regular natural fires, where *resprouting species* survive the fire, while *seeding species* die in the fire but the fire triggers the shedding of seeds, which have been stored since the previous fire. See also Illian, Møller and Waagepetersen (2008), where further background material is provided and various examples of the point patterns of resprouting and reseeding plants are shown. Figure 4.4 shows the locations of one of the reseeding plants *Leucopogon conostephioides* (called seeder 4 in Illian, Møller and Waagepetersen, 2008). This and 5 other species of reseeding plants together with the 19 most dominant (influential) species of resprouters are analyzed in Illian, Møller and Waagepetersen (2008). Since it is natural to model the locations of the reseeding plants conditionally on the locations of the resprouting plants, we consider below a model for the point pattern x in Figure 4.4 conditional on the point patterns $y_1, \ldots, y_{19}$ corresponding to the 19 most dominant species of resprouters, as given in Figure 1 in Illian, Møller and Waagepetersen (2008). For a discussion of possible interaction with other seeder species, and the biological justification of the the covariates defined below, we refer again to Illian, Møller and Waagepetersen (2008).

Let $\kappa_{\mathbf{t},i} \geq 0$ denote a parameter which specifies the radius of interaction of the ith resprouter at location $\mathbf{t} \in y_i$, and let κ denote the collection of all $\kappa_{\mathbf{t},i}$ for $\mathbf{t} \in y_i$ and $i = 1, \ldots, 19$. For $i = 1, \ldots, 19$, define covariates $z_i(\mathbf{s}) = z_i(\mathbf{s}; \kappa_{\mathbf{t},i}, \mathbf{t} \in y_i)$ by

$$z_i(\mathbf{s}; \kappa_{\mathbf{t},i}, \mathbf{t} \in y_i) = \sum_{\mathbf{t} \in y_i \colon \|\mathbf{s}-\mathbf{t}\| \leq \kappa_{\mathbf{t},i}} \left(1 - (\|\mathbf{s}-\mathbf{t}\|/\kappa_{\mathbf{t},i})^2\right)^2.$$

Conditional on $y_1, \ldots, y_{19}$, we assume that $x = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ is a realization of a *Poisson process* with log linear intensity function

$$\log \rho_{\theta, y_1, \ldots, y_n}(\mathbf{s}) = \beta_0 + \sum_{i=1}^{19} \beta_i z_i(\mathbf{s}; \kappa_{\mathbf{t},i}, \mathbf{t} \in y_i)$$

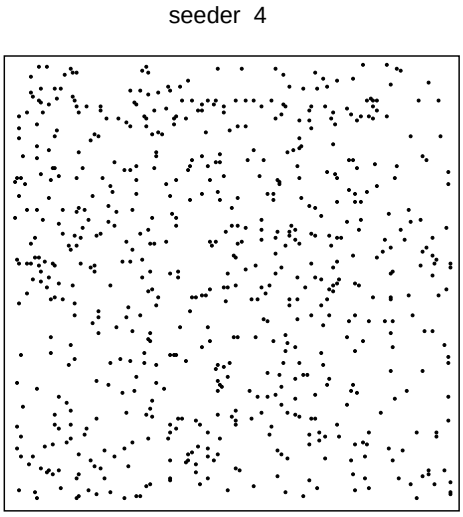

Figure 4.4: Locations of 657 *Leucopogon conostephioides* plants observed within a 22×22 m window.

where $\theta = (\beta, \kappa)$ and $\beta = (\beta_0, \ldots, \beta_{19})$ is a regression parameter, where β_0 is an intercept and β_i for $i > 0$ controls the influence of the ith resprouter. The likelihood depends on κ in a complicated way, and the dimension of κ is much larger than the size of the data x. This makes it meaningless to find maximum likelihood estimates.

Using a Bayesian setting we treat $\theta = (\beta, \kappa)$ as a random variable. Based on Table 1 in Illian, Møller and Waagepetersen (2008) and other considerations in that paper, we make the following prior assumptions. We let $\kappa_{\mathbf{t},i}$ follow the restriction of a normal distribution $\mathrm{N}(\mu_i, \sigma_i^2)$ to $[0, \infty)$, where (μ_i, σ_i^2) is chosen so that under the unrestricted normal distribution the range of the zone of influence is a central 95% interval. Furthermore, we let all the $\kappa_{\mathbf{t},i}$ and the β_i be independent, and each β_i is $\mathrm{N}(0, \sigma^2)$-distributed, where $\sigma = 8$. Combining these prior assumptions with the likelihood term, we obtain the posterior density

$$\pi(\beta, \kappa | x) \propto \exp\Big(- \beta_0/(2\sigma^2) - \sum_{i=1}^{19} \Big\{ \beta_i^2/(2\sigma^2) + \sum_{\mathbf{t} \in y_i} (\kappa_{\mathbf{t},i} - \mu_i)^2/(2\sigma_i^2) \Big\} \Big)$$
$$\times \exp\Big(- \int_W \rho_{\theta, y_1, \ldots, y_n}(\mathbf{s})\, \mathrm{d}\mathbf{s} \Big) \prod_{i=1}^{n} \rho_{\theta, y_1, \ldots, y_n}(\mathbf{s}_i), \quad \beta_i \in \mathbb{R}, \kappa_{\mathbf{t},i} \geq 0 \quad (4.22)$$

(suppressing in the notation $\pi(\beta, \kappa | x)$ that we have conditioned on $y_1, \ldots, y_{19}$ in the posterior distribution).

Simulations from (4.22) are obtained by a Metropolis-within-Gibbs algorithm (also called a hybrid MCMC algorithm, see e.g. Robert and Casella, 1999), where we alter between updating β and κ using random walk Metropolis up-

dates (for details, see Illian, Møller and Waagepetersen, 2008). Thereby various posterior probabilities of interest can be estimated. For example, a large (small) value of $\mathrm{P}(\beta_i > 0|\mathbf{x})$ indicates a positive/attractive (negative/repulsive) association to the ith resprouter, see Figure 2 in Illian, Møller and Waagepetersen (2008).

The model can be checked following the idea of *posterior predictive model assessment* (Gelman, Meng and Stern, 1996), comparing various summary statistics with their posterior predictive distributions. The posterior predictive distribution of statistics depending on X (and possibly also on (β, κ)) is obtained from simulations: we generate a posterior sample $(\beta^{(j)}, \kappa^{(j)})$, $j = 1, \ldots, m$, and for each j 'new data' $x^{(j)}$ from the conditional distribution of X given $(\beta^{(j)}, \kappa^{(j)})$. For instance, the grey scale plot in Figure 4.5 is a *residual plot* based on quadrant counts. We divide the observation window into 100 equally sized quadrants and count the number of plants within each quadrant. The grey scales reflect the probabilities that counts drawn from the posterior predictive distribution are less or equal to the observed quadrant counts where dark means small probability. The stars mark quadrants where the observed counts are 'extreme' in the sense of being either below the 2.5% quantile or above the 97.5% quantile of the posterior predictive distribution. Figure 4.5 does not provide evidence against our model. A plot based on the L-function (Chapter 4.3) and the posterior predictive distribution is also given in Illian, Møller and Waagepetersen (2008). Also this plot shows no evidence against our model.

4.5.2 Cluster and Cox processes

The simulation-based Bayesian approach exemplified above extends to cluster and Cox processes, where we include the 'missing data' η, say, in the posterior and use a Metropolis-within-Gibbs (or MCMC algorithm) algorithm, where we alter between updating θ and η. Examples are given below.

In case of the *Poisson cluster process* model for X considered in Section 4.4.3, $\eta = \Psi$ is the point process of centre points. Incorporating this into the posterior, we obtain the posterior density

$$\pi(\theta, \psi|x) \propto f_\theta(x|\psi) f_\omega(\psi) \pi(\theta)$$

where $f_\theta(x|\psi)$ and $f_\omega(\psi)$ are specified in (4.17) and (4.20), and $\pi(\theta)$ is the prior density. The Metropolis-within-Gibbs algorithm alters between updating from 'full conditionals' given by

$$\pi(\theta|\psi, x) \propto f_\theta(x|\psi) f_\omega(\psi) \pi(\theta) \tag{4.23}$$

and

$$\pi(\psi|\theta, x) \propto f_\theta(x|\psi) f_\omega(\psi). \tag{4.24}$$

Yet another Metropolis-within-Gibbs algorithm may be used when updating from (4.23), cf. Section 4.4.3. When updating from (4.24) we use the birth-death type Metropolis-Hastings algorithm mentioned in connection to (4.19).

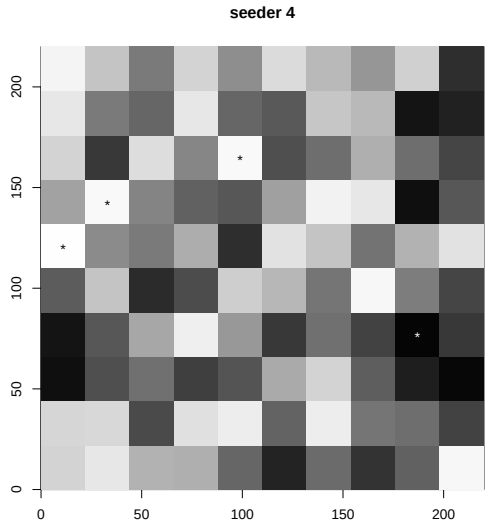

Figure 4.5: Residual plot based on quadrant counts. Quadrants with a '*' are where the observed counts fall below the 2.5% quantile (white '*') or above the 97.5% quantile (black '*') of the posterior predictive distribution. The grey scales reflect the probabilities that counts drawn from the posterior predictive distribution are less or equal to the observed quadrant counts (dark means small probability).

Similarly, for a *log Gaussian Cox process* model for X. Then we may approximate the log Gaussian process Y on W by the finite-dimensional log Gaussian random variable $\eta = Y_I$ specified in Section 4.4.3, and use a Langevin-Hastings algorithm for simulating from the conditional distribution of η given (θ, x). Rue, Martino and Chopin (2007) demonstrate that it may be possible to compute accurate Laplace approximations of marginal posterior distributions without MCMC simulations.

For instance, Møller and Waagepetersen (2007) considered a log Gaussian Cox process model for the rain forest trees considered in Section 4.3.3, and they used a 200×100 grid to index η, and imposed certain flat priors on the unknown parameters. Figure 4.6 shows the posterior means of the systematic part $\beta_0 + \beta_1 z_1(\mathbf{s}) + \beta_2 z_2(\mathbf{s})$ (left panel) and the random part $Y(\mathbf{s})$ (right panel) of the log random intensity function $\log \Lambda(\mathbf{s})$ given by (4.5). The systematic part seems to depend more on z_2 (norm of altitude gradient) than z_1 (altitude), cf. Figure 4.2. The fluctuations of the random part may be caused by small scale clustering due to seed dispersal and covariates concerning soil properties. The fluctuation may also be due to between-species competition.

Møller and Waagepetersen (2004, 2007), Beněs, Bodlák, Møller and Waa-

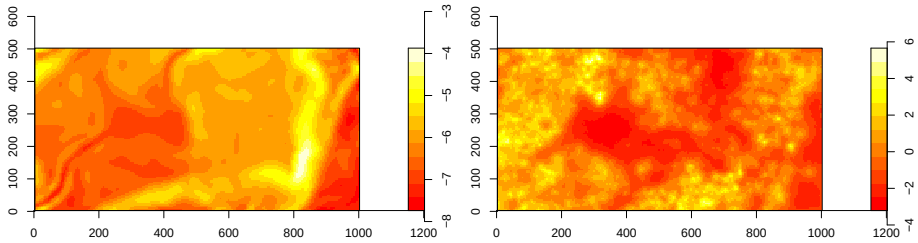

Figure 4.6: Posterior mean of $\beta_0 + \beta_1 z_1(\mathbf{s}) + \beta_2 z_2(\mathbf{s})$ (left panel) and $Y(\mathbf{s})$ (right panel), $\mathbf{s} \in W$, under the log Gaussian Cox process model for the tropical rain forest trees.

gepetersen (2005), and Waagepetersen and Schweder (2006) exemplified the simulation-based Bayesian approach for both Poisson cluster (or shot noise Cox) process and log Gaussian Cox process models. Other Cox models and application examples are considered in Heikkinen and Arjas (1998), Wolpert and Ickstadt (1998), Best, Ickstadt and Wolpert (2000), and Cressie and Lawson (2000).

4.5.3 Gibbs point processes

For a *Gibbs (or Markov) point process* the likelihood function depends on the unknown normalizing constant c_θ, cf. (4.13). Hence, in a Bayesian approach to inference, the posterior distribution for θ also depends on the unknown c_θ, and in an 'ordinary' Metropolis-Hastings algorithm, the Hastings ratio depends on a ratio of unknown normalizing constants. This ratio may be estimated using another method, see Section 4.4.1, but it is then unclear from which equilibrium distribution (if any) we are simulating and whether it is a good approximation of the posterior. Recently, the problem with unknown normalizing constants has been solved using an MCMC auxiliary variable method (Møller, Pettitt, Berthelsen and Reeves, 2006) which involves perfect simulations (Kendall, 1998; Kendall and Møller, 2000). The technique is applied for Bayesian inference of Markov point processes in Berthelsen and Møller (2004, 2006, 2008), where also the many technical details are discussed. Below we briefly demonstrate the potential of this technique when applied for *non/semi-parametric Bayesian inference* of a *pairwise interaction point process*.

4.5.4 Example: cell data

The left panel of Figure 4.7 shows the location of 617 cells in a section of the mocous membrane of the stomach of a healthy rat, where (after some rescaling) $W = [0,1] \times [0, 0.893]$ is the observation window. 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. The centre panel of Figure 4.7 shows a non-parametric estimate $\hat{g}(r),\ r > 0$, of the pair correlation function for

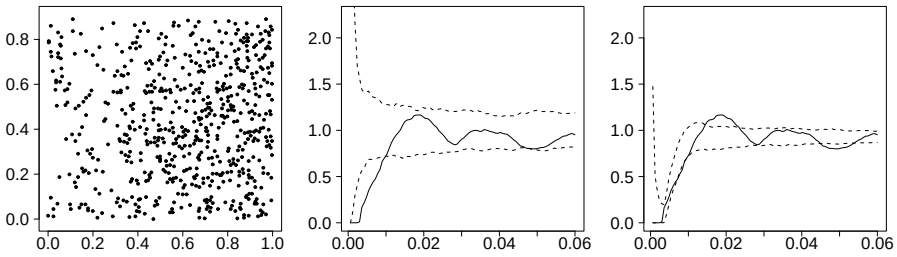

Figure 4.7: Left panel: locations of 617 cells in a 2D section of the mocous membrane of the stomach of a healthy rat. Centre panel: non-parametric estimate of the pair correlation function for the cell data (full line) and 95%-envelopes calculated from 200 simulations of a fitted inhomogeneous Poisson 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).

the data and simulated 95%-envelopes under an inhomogeneous Poisson process with a non-parametric estimate for its intensity function (Chapter 4.3). Under a Poisson process model the theoretical pair correlation function is constant one. The low values of $\hat{g}(r)$ for distances $r < 0.01$ indicates repulsion between the points. The point pattern looks inhomogeneous in the horizontal direction, and the data was originally analyzed by Nielsen (2000) using a Strauss point process model after transforming the first coordinates of the points. The right panel of Figure 4.7 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 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.

We consider an inhomogeneous pairwise interaction point process model for the point pattern $x = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ in Figure 4.7 (left panel). The density is

$$f_{\beta,\varphi}(x) = \frac{1}{c_{(\beta,\varphi)}} \prod_{i=1}^{n} \beta(\mathbf{s}_i) \prod_{1 \le i < j < \le n} \varphi(\|\mathbf{s}_i - \mathbf{s}_j\|) \tag{4.25}$$

with respect to the Poisson process on W with intensity one. Here the first order term β is a non-negative function which models the inhomogeneity, the second order term φ is a non-negative function which models the interaction, and $c_{(\beta,\varphi)}$ is a normalizing constant. A priori it is expected that the cell intensity only changes in the direction from the stomach cavity to the surrounding

muscles tissue. It is therefore assumed that $\beta(\mathbf{s})$ depends only on $\mathbf{s} = (t, u)$ through its first coordinate t. Further, partly in order to obtain a well-defined density and partly in order to model a *repulsive* interaction between the cells, we assume that $0 \leq \varphi(\|\mathbf{s}_i - \mathbf{s}_j\|) \leq 1$ is a non-decreasing function of the distance $r = \|\mathbf{s}_i - \mathbf{s}_j\|$. Furthermore, we specify a flexible prior for $\beta(\mathbf{s}) = \beta(t)$ by a shot noise process and a flexible prior for $\varphi(r)$ by a piecewise linear function modelled by a marked Poisson process. For details of these priors and how the auxiliary variable method from Møller, Pettitt, Berthelsen and Reeves (2006) is implemented to obtain simulations from the posterior distribution of (β, φ) given x, see Berthelsen and Møller (2008).

The left panel of Figure 4.8 shows the posterior mean of β, $\mathrm{E}(\beta|x)$, 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 $\mathrm{E}(\beta|x)$ is the abrupt change of $\mathrm{E}(\beta|x)$ in the interval $[0.2, 0.4]$. For locations near the edges of W, $\mathrm{E}(\beta|x)$ is 'pulled' towards its prior mean as a consequence of the smoothing prior.

The intensity $\rho_{\beta,\varphi}(\mathbf{s})$ of the point process is given by the mean of the Papangelou conditional intensity, that is,

$$\rho_{\beta,\varphi}(\mathbf{s}) = \mathrm{E}\left[\lambda_{\beta,\varphi}(\mathbf{s}, Y) f_{\beta,\varphi}(Y)\right] \tag{4.26}$$

where the expectation is with respect to the Poisson process Y on W with intensity one, see e.g. Møller and Waagepetersen (2004). Define

$$\rho_{\beta,\varphi}(t) = \frac{1}{b} \int_0^b \rho_{\beta,\varphi}(t, u)\, \mathrm{d}u$$

where $W = [0, a] \times [0, b] = [0, 1] \times [0, 0.893]$. Apart from boundary effects, since $\beta(\mathbf{s})$ only depends on the first coordinate of $\mathbf{s} = (t, u)$, we may expect that the intensity (4.26) only slightly depends on the second coordinate u, i.e. $\rho_{\beta,\varphi}(\mathbf{s}) \approx \rho_{\beta,\varphi}(t)$. We therefore refer to $\rho_{\beta,\varphi}(t)$ as the cell intensity, though it is more precisely the average cell intensity in W at $u \in [0, a]$. The left panel of Figure 4.8 also shows a non-parametric estimate $\hat{\rho}(t)$ of the cell intensity (the dot-dashed line). The posterior mean of $\beta(t)$ is not unlike $\hat{\rho}(t)$ except that $\mathrm{E}(\beta(t)|x)$ is higher as would be expected due to the repulsion in the pairwise interaction point process model.

The posterior mean of φ is shown in the right panel of Figure 4.8 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 $d = \min_{i \neq j} \|\mathbf{s}_i - \mathbf{s}_j\|$ which is a little less than 0.006, and an effective interaction range which is no more than 0.015 (the posterior distribution of $\varphi(r)$ is concentrated close to one for $r > 0.015$). The corner at $r = d$ of the curve showing the posterior mean of $\varphi(r)$ is caused by that $\varphi(r)$ is often zero for $r < d$ (since the hard core is concentrated close to d), while $\varphi(r) > 0$ for $r > d$.

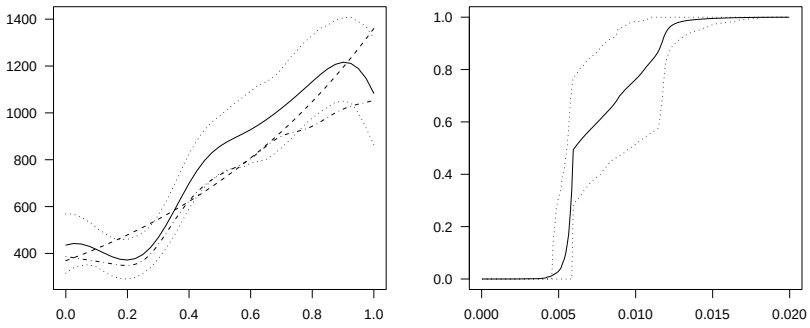

Figure 4.8: Posterior mean (solid line) and pointwise 95% central posterior intervals (dotted lines) for β (left panel) and φ (right panel). The left panel also shows the first order term (dashed line) estimated by Nielsen (2000) and an estimate of the cell intensity (dot-dashed line).

Acknowledgments

Much of this contribution is based on previous work with my collaborators, particularly, Kasper K. Berthelsen, Janine Illian, Anne R. Syversveen, and last but not least, Rasmus P. Waagepetersen. Supported by the Danish Natural Science Research Council, grant no. 272-06-0442 ('Point process modelling and statistical inference').

Bibliography

[1] P. Armstrong. Species patterning in the heath vegetation of the northern sandplain. Honours thesis, University of Western Australia, 1991.

[2] A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, editors. *Case Studies in Spatial Point Process Modeling*. Springer Lecture Notes in Statistics 185, Springer-Verlag, New York, 2006.

[3] A. Baddeley, J. Møller, and R. Waagepetersen. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54:329–350, 2000.

[4] A. Baddeley and R. Turner. Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics*, 42:283–322, 2000.

[5] A. Baddeley and R. Turner. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12:1–42, 2005. URL: `www.jstatsoft.org`, ISSN: 1548-7660.

[6] A. Baddeley and R. Turner. Modelling spatial point patterns in R. In A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, editors, *Case Studies in Spatial Point Process Modeling*, pages 23–74. Springer Lecture Notes in Statistics 185, Springer-Verlag, New York, 2006.

[7] V. Benes, K. Bodlak, J. Møller, and R. P. Waagepetersen. A case study on point process modelling in disease mapping. *Image Analysis and Stereology*, 24:159–168, 2005.

[8] K. K. Berthelsen and J. Møller. Likelihood and non-parametric Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling. *Scandinavian Journal of Statistics*, 30:549–564, 2003.

[9] K. K. Berthelsen and J. Møller. An efficient MCMC method for Bayesian point process models with intractable normalising constants. In A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, editors, *Spatial Point Process Modelling and Its Applications*. Publicacions de la Universitat Jaume I, 2004.

[10] K. K. Berthelsen and J. Møller. Bayesian analysis of Markov point processes. In A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, editors, *Case Studies in Spatial Point Process Modeling*, pages 85–97. Springer Lecture Notes in Statistics 185, Springer-Verlag, New York, 2006.

[11] K. K. Berthelsen and J. Møller. Non-parametric Bayesian inference for inhomogeneous Markov point processes. *Australian and New Zealand Journal of Statistics*, 50, 2008. To appear.

[12] J. Besag. Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute*, 47:77–92, 1977.

[13] J. Besag, R. K. Milne, and S. Zachary. Point process limits of lattice processes. *Journal of Applied Probability*, 19:210–216, 1982.

[14] N. G. Best, K. Ickstadt, and R. L. Wolpert. Spatial Poisson regression for health and exposure data measured at disparate resolutions. *Journal of the American Statistical Association*, 95:1076–1088, 2000.

[15] A. Brix and P. J. Diggle. Spatio-temporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society Series B*, 63:823–841, 2001.

[16] A. Brix and W. S. Kendall. Simulation of cluster point processes without edge effects. *Advances in Applied Probability*, 34:267–280, 2002.

[17] A. Brix and J. Møller. Space-time multitype log Gaussian Cox processes with a view to modelling weed data. *Scandinavian Journal of Statistics*, 28:471–488, 2001.

[18] R. Condit. *Tropical Forest Census Plots*. Springer-Verlag and R. G. Landes Company, Berlin, Germany and Georgetown, Texas, 1998.

[19] R. Condit, S. P. Hubbell, and R. B. Foster. Changes in tree species abundance in a neotropical forest: impact of climate change. *Journal of Tropical Ecology*, 12:231–256, 1996.

[20] N. Cressie and A. Lawson. Hierarchical probability models and Bayesian analysis of minefield locations. *Advances in Applied Probability*, 32:315–330, 2000.

[21] P. J. Diggle. *Statistical Analysis of Spatial Point Patterns*. Arnold, London, second edition, 2003.

[22] A. Gelman and X.-L. Meng. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science*, 13:163–185, 1998.

[23] A. Gelman, X. L. Meng, and H. S. Stern. Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica*, 6:733–807, 1996.

[24] C. J. Geyer. Likelihood inference for spatial point processes. In O. E. Barndorff-Nielsen, W. S. Kendall, and M. N. M. van Lieshout, editors, *Stochastic Geometry: Likelihood and Computation*, pages 79–140, Boca Raton, Florida, 1999. Chapman & Hall/CRC.

[25] C. J. Geyer and J. Møller. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, 21:359–373, 1994.

[26] C. J. Geyer and E. A. Thompson. Annealing Markov chain Monte Carlo with applications to pedigree analysis. *Journal of the American Statistical Association*, 90:909–920, 1995.

[27] R. D. Harkness and V. Isham. A bivariate spatial point pattern of ants' nests. *Applied Statistics*, 32:293–303, 1983.

[28] J. Heikkinen and E. Arjas. Non-parametric Bayesian estimation of a spatial Poisson intensity. *Scandinavian Journal of Statistics*, 25:435–450, 1998.

[29] L. Heinrich. Minimum contrast estimates for parameters of spatial ergodic point processes. In *Transactions of the 11th Prague Conference on Random Processes, Information Theory and Statistical Decision Functions*, pages 479–492, Prague, 1992. Academic Publishing House.

[30] H. Högmander and A. Särkkä. Multitype spatial point patterns with hierarchical interactions. *Biometrics*, 55:1051–1058, 1999.

[31] S. P. Hubbell and R. B. Foster. Diversity of canopy trees in a neotropical forest and implications for conservation. In S. L. Sutton, T. C. Whitmore, and A. C. Chadwick, editors, *Tropical Rain Forest: Ecology and Management*, pages 25–41. Blackwell Scientific Publications, 1983.

[32] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley and Sons, Chichester, 2008.

[33] J. B. Illian, J. Møller, and R. P. Waagepetersen. Spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics*. (To appear), 2008.

[34] J. L. Jensen and H. R. Künsch. On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Annals of the Institute of Statistical Mathematics*, 46:475–486, 1994.

[35] J. L. Jensen and J. Møller. Pseudolikelihood for exponential family models of spatial point processes. *Annals of Applied Probability*, 3:445–461, 1991.

[36] W. S. Kendall. Perfect simulation for the area-interaction point process. In L. Accardi and C.C. Heyde, editors, *Probability Towards 2000*, pages 218–234. Springer Lecture Notes in Statistics 128, Springer Verlag, New York, 1998.

[37] W. S. Kendall and J. Møller. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*, 32:844–865, 2000.

[38] Lieshout, M. N. M. van. *Markov Point Processes and Their Applications*. Imperial College Press, London, 2000.

[39] B. G. Lindsay. Composite likelihood methods. *Contemporary Mathematics*, 80:221–239, 1988.

[40] S. Mase. Consistency of the maximum pseudo-likelihood estimator of continuous state space Gibbs processes. *Annals of Applied Probability*, 5:603–612, 1995.

[41] S. Mase. Marked Gibbs processes and asymptotic normality of maximum pseudo-likelihood estimators. *Mathematische Nachrichten*, 209:151–169, 1999.

[42] J. Møller. Contribution to the discussion of N.L. Hjort and H. Omre (1994): Topics in spatial statistics. *Scandinavian Journal of Statistics*, 21:346–349, 1994.

[43] J. Møller. Shot noise Cox processes. *Advances in Applied Probability*, 35:4–26, 2003.

[44] J. Møller, A. N. Pettitt, K. K. Berthelsen, and R. W. Reeves. An efficient MCMC method for distributions with intractable normalising constants. *Biometrika*, 93:451–458, 2006.

[45] J. Møller, A. R. Syversveen, and R. P. Waagepetersen. Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25:451–482, 1998.

[46] J. Møller and R. P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton, 2004.

[47] J. Møller and R. P. Waagepetersen. Modern spatial point process modelling and inference (with discussion). *Scandinavian Journal of Statistics*, 34:643–711, 2007.

[48] L. S. Nielsen. Modelling the position of cell profiles allowing for both inhomogeneity and interaction. *Image Analysis and Stereology*, 19:183–187, 2000.

[49] S. L. Rathbun. Estimation of Poisson intensity using partially observed concomitant variables. *Biometrics*, 52:226–242, 1996.

[50] S. L. Rathbun and N. Cressie. Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson processes. *Advances in Applied Probability*, 26:122–154, 1994.

[51] S. L. Rathbun, S. Shiffman, and C. J. Gwaltney. Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika*, 94:153–165, 2007.

[52] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, 1999.

[53] H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. Preprint Statistics No. 1/2007, Norwegian University of Science and Technology, 2007.

[54] M. Thomas. A generalization of Poisson's binomial limit for use in ecology. *Biometrika*, 36:18–25, 1949.

[55] R. Waagepetersen. An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, 63:252–258, 2007.

[56] R. Waagepetersen. Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika*, 95, 2008. To appear.

[57] R. Waagepetersen and Y. Guan. Two-step estimation for inhomogeneous spatial point processes. Submitted, 2007.

[58] R. Waagepetersen and T. Schweder. Likelihood-based inference for clustered line transect data. *Journal of Agricultural, Biological, and Environmental Statistics*, 11:264–279, 2006.

[59] R. L. Wolpert and K. Ickstadt. Poisson/gamma random field models for spatial statistics. *Biometrika*, 85:251–267, 1998.