Modern statistics for spatial point processes

Møller, Jesper; Waagepetersen, Rasmus

Publication date:
2006

Document Version
Accepted author manuscript, peer reviewed version

AALBORG UNIVERSITY

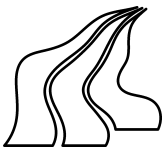
Modern statistics for spatial point processes

by

Jesper Møller and Rasmus P. Waagepetersen

R-2006-12 April 2006

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: http://www.math.aau.dk



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

Modern statistics for spatial point processes*

April 1, 2006

Jesper Møller and Rasmus P. Waagepetersen

Department of Mathematical Sciences, Aalborg University

Abstract: We summarize and discuss the current state of spatial point process theory and directions for future research, making an analogy with generalized linear models and random effect models, and illustrating the theory with various examples of applications. In particular, we consider Poisson, Gibbs, and Cox process models, diagnostic tools and model checking, Markov chain Monte Carlo algorithms, computational methods for likelihood-based inference, and quick non-likelihood approaches to inference.

Keywords: Bayesian inference, conditional intensity, Cox process, Gibbs point process, Markov chain Monte Carlo, maximum likelihood, perfect simulation, Poisson process, residuals, simulation free estimation, summary statistics.

1 Introduction

Spatial point pattern data occur frequently in a wide variety of scientific disciplines, including seismology, ecology, forestry, geography, spatial epidemiology, and material science, see e.g. Stoyan & Stoyan (1998), Kerscher (2000), Boots, Okabe & Thomas (2003), Diggle (2003), and Ballani (2006). The classical spatial point process textbooks like Ripley (1981, 1988), Diggle (1983), Stoyan & Stoyan (1995), and Stoyan, Kendall & Mecke (1995) usually deal with relative small point patterns, where the assumption of stationarity is central and

*Prepared for presentation as an special invited talk at the 21st Nordic Conference on Mathematical Statistics, June 11 - 15, 2006, and for submission to the Scandinavian Journal of Statistics.

non-parametric methods based on summary statistics play a major role. In recent years, fast computers and advances in computational statistics, particularly Markov chain Monte Carlo (MCMC) methods, have had a major impact on the development of statistics for spatial point processes. The focus has now changed to likelihood-based inference for flexible parametric models, often depending on covariates, and liberated from restrictive assumptions of stationarity. In short, 'Modern statistics for spatial point processes', where recent textbooks include Van Lieshout (2000), Diggle (2003), Møller & Waagepetersen (2003b), and Baddeley, Gregori, Mateu, Stoica & Stoyan (2006).

Much of the literature on spatial point processes is fairly technical with extensive use of measure theoretical terminology and statistical physics parlance. This has made the theory seem rather difficult. Moreover, in connection with likelihood-based inference, many statisticians may be unfamiliar with the concept of defining a density with respect to a Poisson process. It is our intention in Sections 3–9 to give a concise and non-technical introduction to the modern theory, making analogies with generalized linear models and random effect models, and illustrating the theory with various examples of applications introduced in Section 2. In particular, we discuss Poisson, Gibbs, and Cox process models, diagnostic tools and model checking, MCMC algorithms and computational methods for likelihood-based inference, and quick non-likelihood approaches to inference. Section 10 summarizes the current state of spatial point process theory and discusses directions for future research.

For definiteness, we mostly work with point processes defined in the plane $\mathbb{R}^2$, but most ideas easily extend to the general case of $\mathbb{R}^d$ or more abstract spaces. For ease of exposition, no measure theoretical details are given; see instead Møller & Waagepetersen (2003b) and the references therein. The computations for the data examples were done using the `R` package `spatstat` (Baddeley & Turner, 2005, 2006) or our own programmes in `C` and `R`, where the code is available at `www.math.aau.dk/~rw/sppcode`. Since we shall often refer to our own monograph, please notice the comments and corrections to Møller & Waagepetersen (2003b) at `www.math.aau.dk/~jm`.

2 Data examples

The following four examples of spatial point pattern data are from plant and animal ecology, and are considered for illustrative purposes in subsequent sec-

tions. In each example, the *observation window* refers to the area where points of the pattern can possibly be observed, i.e. when the point pattern is viewed as a realization of a spatial point process (Section 3.1). Absence of points in a region, where they could potentially occur, is a source of information complementary to the data on where points actually did occur. The specification of the observation window is therefore an integral part of a spatial point pattern data set.

Figure 1 shows positions of 55 minke whales (*balaneoptera acutorostrata*) observed in a part of the North Atlantic near Spitzbergen. The whales are observed visually from a ship sailing along predetermined so-called transect lines. The point pattern can be thought of as an incomplete observation of all the whale positions, since it is only possible to observe whales within the vicinity of the ship. Moreover, whales within sighting distance may fail to be observed due to bad weather conditions or if they are diving. The probability of observing a whale is a decreasing function of the distance from the whale to the ship and is effectively zero for distances larger than 2 km. The observation window is therefore a union of narrow strips of width 4 km around the transect lines. More details on the data set can be found in Skaug, Øien, Schweder & Bøthun (2004) and Waagepetersen & Schweder (2005). The objective is to estimate the abundance of the whales, or equivalently the whale intensity. The whales tend to cluster around locations of high prey intensity, and a point process model for all whale positions (including those not observed) should take this into account. The point process model used in Waagepetersen & Schweder (2005) is described in Example 4.2.

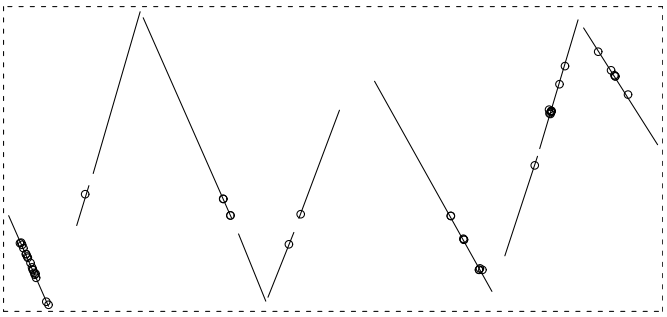


Figure 1: Observed whales along transect lines. The enclosing rectangle is of dimensions 263 km by 116 km.

In studies of biodiversity of tropical rain forests, it is of interest to study

whether the spatial patterns of the many different tree species can be related to spatial variations in environmental variables concerning topography and soil properties. Figure 2 shows positions of 3605 *Beilschmiedia pendula Lauraceae* trees in the tropical rain forest of Barro Colorado Island. This data set is a part of a much larger data set containing positions of hundreds of thousands of trees belonging to thousands of species, see Hubbell & Foster (1983), Condit, Hubbell & Foster (1996), and Condit (1998). In addition to the tree positions, covariate information on altitude and norm of altitude gradient is available, see Figure 3. Phrased in point process terminology, the question is whether the intensity of *Beilschmiedia* trees may be viewed as a spatially varying function of the covariates. In the study of this question, it is, as for the whales, important to take into account clustering, which in the present case may be due to tree reproduction by seed dispersal and possibly unobserved covariates. Different point process models for the tree positions are considered in Examples 4.1 and 4.3.

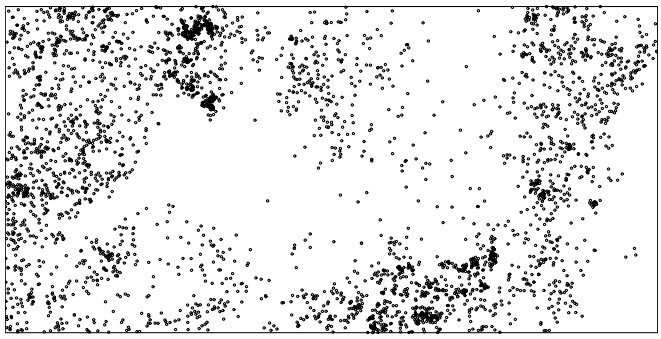


Figure 2: Locations of *Beilschmiedia pendula Lauraceae* trees observed in a 1000 m by 500 m rectangular window.

Another pertinent question in plant ecology is how trees interact due to competition. Figure 4 shows positions and stem diameters of 134 Norwegian spruces. This data set was first analyzed by Fiksel (1984), and it is an example of a marked point pattern, with points given by the tree locations and marks by the stem diameters. The discs in Figure 4 are of radii five times the stem diameters and may be thought of as 'influence zones' of the trees, see Penttinen, Stoyan & Henttonen (1992) and Goulard, Särkkä & Grabarnik (1996). From an ecological point of view it is of interest to study how neighbouring trees interact, i.e. when their influence zones overlap. It is then natural to model the conditional intensity, which roughly speaking determines the probability of

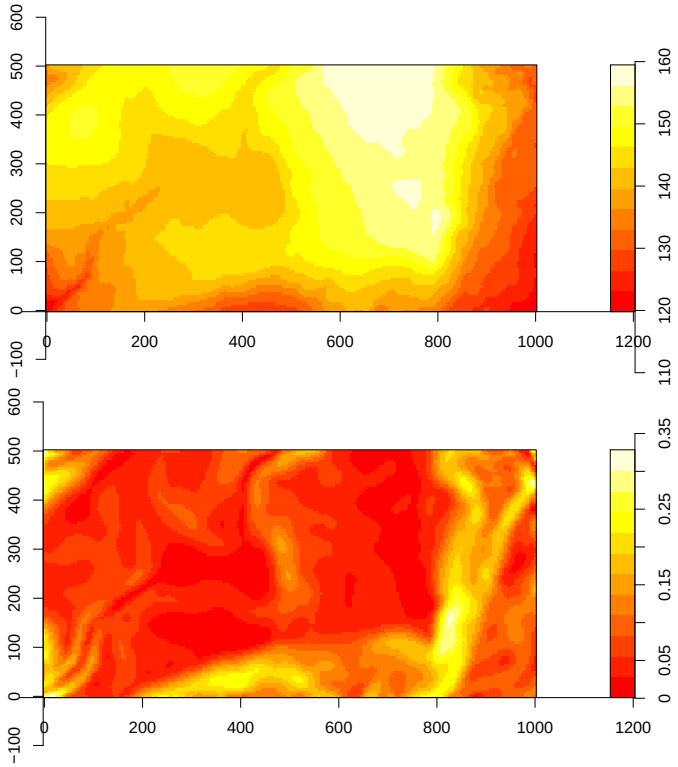


Figure 3: Altitude (upper plot) in meter and norm of altitude gradient (lower plot).

observing a tree at a given location and of given stem diameter conditional on the neighbouring trees. In Example 5.1, we consider a simple model where the conditional intensity depends on the amount of overlap between the influence zones of a tree and its neighbouring trees.

Our last data set is an example of a multitype point pattern, with two types of points specifying the positions of nests for two types of ants, *messor wasmanni* and *cataglyphis bicolor*, see Figure 5 and Harkness & Isham (1983). Note the rather atypical shape of the observation window. The interaction between the two types of ants is of main interest for this data set. Biological knowledge suggests that the *messor* ants are not influenced by presence or absence of *cataglyphis* ants when choosing sites for their nests. The *catagplyphis* ants, on the other hand, feed on dead *messors* and hence the positions of *messor* nest might affect the choice of sites for *cataglyphis* nests. Högmander and Särkkä (1999)

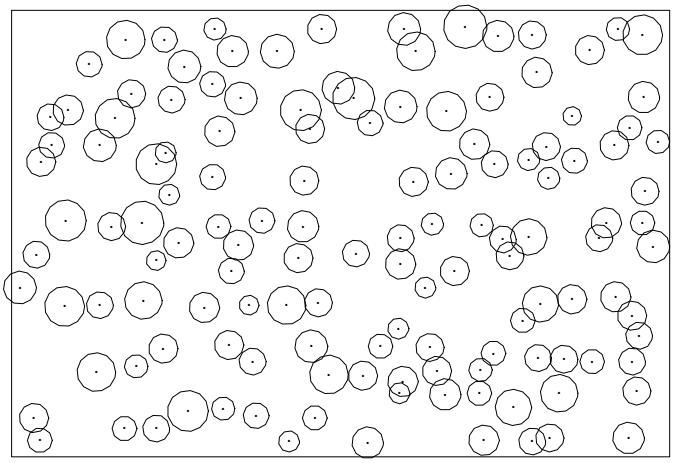


Figure 4: Norwegian spruces observed in a rectangular 56 m by 38 m window. The radii of the discs equal 5 times the stem diameters.

therefore specify a hierarchical model: first a model for the conditional intensity of a *messor* nest at a particular location given the neighbouring *messor nests*, and second a conditional intensity for a *cataglyphis* nest given the neighbouring *cataglyphis* nests *and* the neighbouring *messor* nests. Further details are given in Example 5.2.

These examples illustrate many important features of interest for spatial point process analysis: *clustering* due to e.g. seed dispersal or unobserved variation in prey intensity (as for the tropical rain forest trees and the whales), *inhomogeneity* e.g. caused by a thinning mechanism or covariates (as for the whales and tropical rain forest trees), and *interaction* between points, where the interaction possibly depends on marks associated with the points (as for the Norwegian spruces). The examples also illustrate different types of observation windows.

3 Preliminaries

3.1 Spatial point processes

In the simplest case, a *spatial point process* $\mathbf{X}$ is a *finite random subset* of a given bounded region $S \subset \mathbb{R}^2$, and a realization of such a process is a *spatial point pattern* $\mathbf{x} = \{x_1, \ldots, x_n\}$ of $n \geq 0$ points contained in S. We say that the *point*

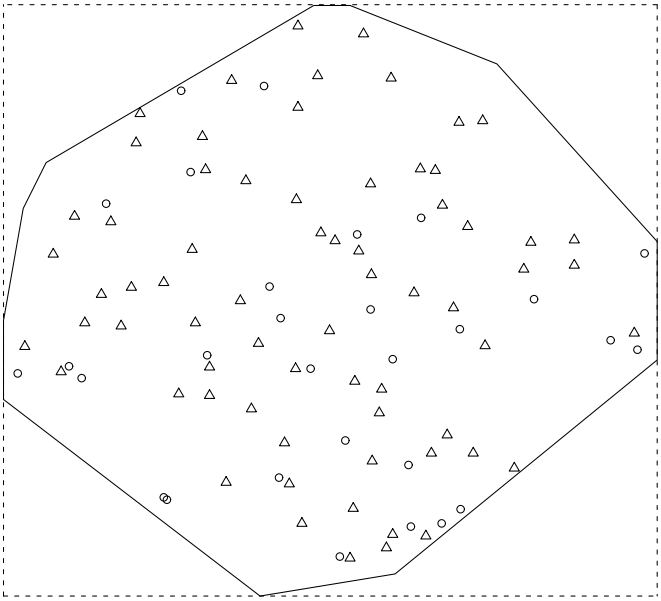


Figure 5: Locations of nests for *messor* (triangles) and *cataglyphis* (circles) ants. Enclosing rectangle for observation window is 829 ft by 766 ft.

process is defined on S, and we write $\mathbf{x} = \emptyset$ for the empty point pattern. The number of points, $n(\mathbf{X})$, is a random variable, and conditional on $n(\mathbf{X}) = n$, the joint distribution of the n points is exchangeable. An equivalent approach is to specify the distribution of the counts of points within subsets of S, i.e. the variables $N(B) = n(\mathbf{X}_B)$ for subsets $B \subseteq S$, where $\mathbf{X}_B = \mathbf{X} \cap B$.

If it is not known on which region the point process is defined, or if the process extends over a very large region, or if certain invariance assumptions such as stationarity are imposed, then it may be appropriate to consider an infinite point process on $\mathbb{R}^2$. We define a *spatial point process* $\mathbf{X}$ *on* $\mathbb{R}^2$ as a locally finite random subset of $\mathbb{R}^2$, i.e. $N(B)$ is a finite random variable whenever $B \subset \mathbb{R}^2$ is a bounded region. We say that $\mathbf{X}$ is *stationary* respective *isotropic* if its distribution is invariant under translations in $\mathbb{R}^2$ respective rotations about the origin in $\mathbb{R}^2$. Stationarity and isotropy may be reasonable assumptions for point processes observed within a homogeneous environment. These assumptions appeared commonly in the older point process literature, where typically rather small study regions were considered. We shall abandon these assumptions when spatial covariate information is available.

In most applications, the observation window W (see Section 2) is strictly contained in the region S, where the point process is defined. Since $\mathbf{X} \setminus W$ is unobserved, we face a missing data problem, which in the spatial point process literature is referred to as a problem of *edge effects*.

3.2 Moments

The mean structure of the count variables $N(B)$, $B \subseteq \mathbb{R}^2$, is summarized by the *moment measure*

$$\mu(B) = \mathrm{E}N(B), \quad B \subseteq \mathbb{R}^2. \tag{1}$$

In practice the mean structure is modelled in terms of a non-negative *intensity function* ρ, i.e.

$$\mu(B) = \int_B \rho(u)\,\mathrm{d}u$$

where we may interpret $\rho(u)\,\mathrm{d}u$ as the probability that precisely one point falls in an infinitesimally small region containing the location u and of area $\mathrm{d}u$.

The covariance structure of the count variables is most conveniently given in terms of the *second order factorial moment measure* $\mu^{(2)}$. This is defined by

$$\mu^{(2)}(A) = \mathrm{E} \sum_{u,v \in \mathbf{X}}^{\neq} \mathbf{1}[(u,v) \in A], \quad A \subseteq \mathbb{R}^2 \times \mathbb{R}^2, \tag{2}$$

where $\neq$ over the summation sign means that the sum runs over all pairwise different points u, v in $\mathbf{X}$, and $\mathbf{1}[\cdot]$ is the indicator function. For bounded regions $B \subseteq \mathbb{R}^2$ and $C \subseteq \mathbb{R}^2$,

$$\mathrm{Cov}[N(B), N(C)] = \mu^{(2)}(B \times C) + \mu(B \cap C) - \mu(B)\mu(C).$$

For many important model classes, $\mu^{(2)}$ is given in terms of an explicitly known *second order product density* $\rho^{(2)}$,

$$\mu^{(2)}(A) = \int \mathbf{1}[(u,v) \in A]\rho^{(2)}(u,v)\,\mathrm{d}u\mathrm{d}v$$

where $\rho^{(2)}(u,v)\mathrm{d}u\mathrm{d}v$ may be interpreted as the probability of observing a point in each of two regions of infinitesimally small areas $\mathrm{d}u$ and $\mathrm{d}v$ and containing u and v. More generally, for integers $n \geq 1$, the *nth order factorial measure* $\mu^{(n)}$

is defined by

$$\mu^{(n)}(A) = \mathrm{E} \sum_{u_1,\ldots,u_n \in \mathbf{X}}^{\neq} \mathbf{1}[(u_1,\ldots,u_n) \in A], \quad A \subseteq \mathbb{R}^{2n}, \tag{3}$$

with corresponding *nth order product density* $\rho^{(n)}$. From (3) we obtain *Campbell's theorem*

$$\mathrm{E} \sum_{u_1,\ldots,u_n \in \mathbf{X}}^{\neq} h(u_1,\ldots,u_n) = \int h(u_1,\ldots,u_n)\rho^{(n)}(u_1,\ldots,u_n)\,\mathrm{d}u_1,\cdots\mathrm{d}u_n \tag{4}$$

for non-negative functions h. The nth order moment measure is given by the right hand side of (3) without $\neq$. The reason for preferring the factorial moment measures are the nicer expressions for the product densities, cf. (6) and (16).

In order to characterize the tendency of points to attract or repel each other, while adjusting for the effect of a large or small intensity function, it is useful to consider the *pair correlation function*

$$g(u,v) = \rho^{(2)}(u,v)/(\rho(u)\rho(v)) \tag{5}$$

(provided $\rho(u) > 0$ and $\rho(v) > 0$). If points appear independently of each other, $\rho^{(2)}(u,v) = \rho(u)\rho(v)$ and $g(u,v) = 1$ (see also (6)). When $g(u,v) > 1$ we interpret this as *attraction* between points of the process at locations u and v, while if $g(u,v) < 1$ we have *repulsion* at the two locations. Translation invariance $g(u,v) = g(u-v)$ of g implies that $\mathbf{X}$ is *second order intensity reweighted stationary* (Baddeley, Møller & Waagepetersen, 2000 and Section 6.2.1), and in applications it is often assumed that $g(u,v) = g(\|u-v\|)$ depends only on the distance $\|u-v\|$. Notice that very different point process models can share the same g function (Baddeley & Silverman, 1984, Baddeley *et al.*, 2000, Møller & Waagepetersen, 2003b).

Suppose $\pi(u) \in [0,1]$, $u \in \mathbb{R}^2$, are given numbers. An *independent π-thinning* of $\mathbf{X}$ is obtained by independent retaining each point u in $\mathbf{X}$ with probability $\pi(u)$. It follows easily from (4) that $\pi(u_1)\cdots\pi(u_n)\rho^{(n)}(u_1,\ldots,u_n)$ is the nth order product density of the thinned process. In particular, $\pi(u)\rho(u)$ is the intensity function of the thinned process, while g is the same for the two processes.

3.3 Marked point processes

In addition to each point u in a spatial point process $\mathbf{X}$, we may have an associated random variable m_u called a mark. The mark often carries some information about the point, like for example the radius of a disc as in Figure 4, the type of ants as in Figure 5, or another point process (e.g. the clusters in a shot noise Cox process, see Section 4.2.2). The process $\mathbf{\Phi} = \{(u, m_u) : u \in \mathbf{X}\}$ is called a *marked point process*. This is 'just' a special type of point process, and many of the concepts for spatial point processes easily extend; see Stoyan & Stoyan (1995), Schlather (2001), and Møller & Waagepetersen (2003b). For the models presented later in this paper, the marked point process model of discs in Figure 4 will be viewed as a point process in $\mathbb{R}^2 \times (0, \infty)$, and the bivariate point process model of ants nests in Figure 5 will be specified by a hierarchical model so that no methodology specific to marked point processes is needed.

3.4 Generic notation

Unless otherwise stated,

$\mathbf{X}$ denotes a generic spatial point process defined on a region $S \subseteq \mathbb{R}^2$;

$W \subseteq S$ is a bounded observation window;

$\mathbf{x} = \{x_1, \ldots, x_n\}$ is either a generic finite point configuration or a realization of $\mathbf{X}_W$ (the meaning of $\mathbf{x}$ will always be clear from the context);

$z(u) = (z_1(u), \ldots, z_k(u))$ is a vector of covariates depending on locations $u \in S$ such as spatially varying environmental variables, known functions of the spatial coordinates themselves or distances to known environmental features, cf. Berman & Turner (1992) and Rathbun (1996);

$\beta = (\beta_1, \ldots, \beta_k)$ is a corresponding regression parameter;

θ is the vector of all parameters (including β) in a given parametric model.

4 Modelling the intensity function

This section discusses spatial point process models specified by a deterministic or random intensity function in analogy with generalized linear models and random effects models. Particularly, two important model classes, namely Poisson and

Cox/cluster point processes are introduced. Roughly speaking, the two classes provide models for no interaction and aggregated point patterns, respectively.

4.1 The Poisson process

A *Poisson process* $\mathbf{X}$ defined on S and with intensity measure μ and intensity function ρ satisfies for any bounded region $B \subseteq S$ with $\mu(B) > 0$,

(i) $N(B)$ is Poisson distributed with mean $\mu(B)$,

(ii) conditional on $N(B)$, the points in $\mathbf{X}_B$ are i.i.d. with density proportional to $\rho(u)$, $u \in B$.

Poisson processes are studied in detail in Kingman (1993). They play a fundamental role as a reference process for exploratory and diagnostic tools and when more advanced spatial point process models are constructed.

If $\rho(u)$ is constant for all $u \in S$, we say that the Poisson process is *homogeneous*. Realizations of the process may appear to be rather chaotic with large empty space and close pairs of points, even when the process is homogeneous. The Poisson process is a model for *'no interaction'* or *'complete spatial randomness'*, since $\mathbf{X}_A$ and $\mathbf{X}_B$ are independent whenever $A, B \subset S$ are disjoint. Moreover,

$$\rho^{(n)}(u_1, \ldots, u_n) = \rho(u_1) \cdots \rho(u_n), \quad g \equiv 1, \tag{6}$$

reflecting the lack of interaction. Stationarity means that $\rho(u)$ is constant, and it implies anisotropy of $\mathbf{X}$. Note that another Poisson process results if we make an independent thinning of a Poisson process.

Typically, a log linear model of the intensity function is considered,

$$\log \rho(u) = z(u)\beta^{\mathsf{T}}. \tag{7}$$

The independence properties of a Poisson process are usually not realistic for real data. Despite of this the Poisson process has enjoyed much popularity due to its mathematical tractability.

4.2 Cox processes

One natural extension of the Poisson process is a *Cox process* $\mathbf{X}$ driven by a non-negative process $\mathbf{\Lambda} = (\Lambda(u))_{u \in S}$, such that conditional on $\mathbf{\Lambda}$, $\mathbf{X}$ is a Poisson

process with intensity function $\mathbf{\Lambda}$ (Cox, 1955; Grandell, 1976; Daley & Vere-Jones, 2003).

Three points of statistical importance should be noticed. First, though $\mathbf{\Lambda}$ may be modelling a random environmental heterogeneity, $\mathbf{X}$ is stationary if $\mathbf{\Lambda}$ is stationary. Second, we cannot distinguish the Cox process $\mathbf{X}$ from its corresponding Poisson process $\mathbf{X}|\mathbf{\Lambda}$ when only one realization of $\mathbf{X}_W$ is available, cf. Bartlett (1964) and the discussion in Møller & Waagepetersen (2003b, Section 5.1). Third, the likelihood is in general unknown, while product densities may be tractable. The consequences of the latter point are discussed in Sections 7 and 8.

4.2.1 Log Gaussian Cox processes

In analogy with random effect models, as an extension of the log linear model (7), take

$$\log \Lambda(u) = z(u)\beta^{\mathsf{T}} + \Psi(u) \tag{8}$$

where $\mathbf{\Psi} = (\Psi(u))_{u \in S}$ is a zero-mean Gaussian process. Then we call $\mathbf{X}$ a *log Gaussian Cox process* (Møller, Syversveen & Waagepetersen, 1998). The covariance function $c(u, v) = \mathrm{Cov}[\Psi(u), \Psi(v)]$ typically depends on some lower-dimensional parameter, see e.g. Example 4.1 below. To ensure local integrability of $\Lambda(u)$, the covariance function has to satisfy certain mild conditions, which are satisfied for models used in practice.

The product densities are particularly tractable. The intensity function

$$\log \rho(u) = z(u)\beta^{\mathsf{T}} + c(u, u)/2 \tag{9}$$

is log linear, g and c are in a one-to-one correspondence as

$$g(u, v) = \exp(c(u, v))$$

and higher-order product densities are nicely expressed in terms of ρ and g (Møller *et al.*, 1998). Another advantageous property is that we have no problem with edge effects, since $\mathbf{X}_W$ is specified by the Gaussian process restricted to W.

Example 4.1. *(Log Gaussian Cox process model for tropical rain forest trees)* For the tropical rain forest trees in Figure 2, we consider in Example 7.5 inference for a log Gaussian Cox process with $z(u) = (1, z_2(u), z_3(u))$, where $z_2(u)$

and $z_3(u)$ denote the altitude and gradient covariates given in Figure 3. An exponential covariance function $c(u,v) = \sigma^2 \exp(-\|u-v\|/\alpha)$ is used for the Gaussian process, where σ and α are positive parameters.

4.2.2 Shot noise Cox processes

A *shot noise Cox process* $\mathbf{X}$ has

$$\Lambda(u) = \sum_{(c,\gamma)\in\mathbf{\Phi}} \gamma k(c,u) \tag{10}$$

where $c \in \mathbb{R}^2$, $\gamma > 0$, $\mathbf{\Phi}$ is a Poisson process on $\mathbb{R}^2 \times (0,\infty)$, and $k(c,\cdot)$ is a density for a two-dimensional continuous random variable (Møller, 2003). Note that $\mathbf{X}$ is distributed as the *superposition* (i.e. union) of independent Poisson processes $\mathbf{X}_{(c,\gamma)}$ with intensity functions $\gamma k(c,\cdot)$, $(c,\gamma) \in \mathbf{\Phi}$, where we interpret $\mathbf{X}_{(c,\gamma)}$ as a cluster with centre c and mean number of points γ. Thus $\mathbf{X}$ is an example of a *Poisson cluster process*, and provides a natural model for seed setting mechanisms causing clustering, see e.g. Brix & Chadoeuf (2002). Simple formulae for the intensity and pair correlation functions of a shot noise Cox process are provided in Møller (2003).

Example 4.2. *(Shot noise Cox process for minke whales)* In Waagepetersen & Schweder (2005), the positions of minke whales in Figure 1 are modelled as an independent thinning of a shot noise Cox process. Letting $p(u)$ denote the probability of observing a whale at location u, the process of observed whales is a Cox process driven by $\Lambda(u) = p(u)\sum_{(c,\gamma)\in\mathbf{\Phi}} \gamma k(c,u)$. The cluster centres are assumed to form a stationary Poisson process with intensity κ, the c's are independent of the γ's, and the γ's are i.i.d. gamma random variables with mean α and unit scale parameter. To handle edge effects, $k(c,\cdot)$ is the density of $N_2(c,\omega^2 I)$ restricted to $c + [-3\omega, 3\omega]^2$.

A particular simple case of a shot noise Cox process is a *Neyman-Scott process* $\mathbf{X}$, where the centre points form a stationary Poisson process with intensity κ and the γ's are all equal to a positive parameter α (Neyman & Scott, 1958). If furthermore $k(c,\cdot)$ is a bivariate normal density with mean c and covariance matrix $\omega^2 I$, then $\mathbf{X}$ is a *Thomas process* (Thomas, 1949). A Neyman-Scott process is stationary with intensity $\rho = \alpha\kappa$, and the Thomas process is also isotropic with

$$g(r) = 1 + \exp\left(-r^2/(4\omega^2)\right)/(4\pi\kappa\omega^2), \quad r > 0. \tag{11}$$

Shot noise Cox process can be extended in various interesting ways by allowing the kernel k to depend on a random band width b and replacing $\mathbf{\Phi}$ by a Poisson or non-Poisson process model for the points (c, γ, b) (Møller & Torrisi, 2005). In this paper, we consider instead an extension which incorporates covariate information in a multiplicative way, i.e. an inhomogeneous Cox process driven by

$$\Lambda(u) = \exp\left(z(u)\beta^{\mathsf{T}}\right) \sum_{(c,\gamma)\in\mathbf{\Phi}} \gamma k(c,u) \tag{12}$$

(Waagepetersen, 2005). A nice feature is that the pair correlation function of $\mathbf{X}$ is the same for (10) and (12), i.e. it does not depend on the parameter β.

Example 4.3. *(Inhomogeneous Thomas model for tropical rain forest trees)* In addition to the log Gaussian Cox process model for the tropical trees (Example 4.1), we consider an inhomogeneous Thomas process of the form (12),

$$\Lambda(u) = \frac{\alpha}{2\pi\omega^2} \exp\left(\beta_2 z_2(u) + \beta_3 z_3(u)\right) \sum_{c\in\mathbf{\Phi}} \exp\left(-\|u-c\|^2/(2\omega^2)\right)$$

where now $\mathbf{\Phi}$ denotes a stationary Poisson process with intensity κ. Then the intensity function is

$$\rho(u) = \kappa\alpha \exp\left(\beta_2 z_2(u) + \beta_3 z_3(u)\right) \tag{13}$$

while the pair correlation function is equal to (11).

5 Modelling the conditional intensity function

Gibbs point processes arose in statistical physics as models for interacting particle systems. The intensity function for a Gibbs process is unknown; instead, the *Papangelou conditional intensity* $\lambda(u, \mathbf{x})$ (Papangelou, 1974) becomes the appropriate starting point for modelling. The definition and interpretation of $\lambda(u, \mathbf{x})$ are given in Section 5.2 in terms of the density of a finite point process and in Section 5.4 using a more technical account for infinite point processes. The density of a Gibbs point process is specified in Section 5.3. Though the density has an unknown normalizing constant, likelihood inference based on MCMC methods is easier for parametric Gibbs point process models than for Cox processes; see Section 7. While Cox processes provide flexible models for aggregation or clustering in a point pattern, Gibbs point processes provide flexible models for

regularity or repulsion (Sections 5.3 and 10.3).

5.1 Finite point processes with a density

Throughout this section we assume that S is bounded and $\mathbf{X}$ is a finite point process defined on S. Moreover, $\mathbf{Y}_\rho$ denotes a Poisson process on S with intensity measure μ and intensity function ρ. In particular, $\mathbf{Y}_1$ is the *unit rate Poisson process* on S with intensity $\rho \equiv 1$. Before defining what is meant by the density of $\mathbf{X}$, we need the following useful Poisson expansion. If F denotes any event of spatial point patterns contained in S, by (i)-(ii) in Section 4.1,

$$\mathrm{P}(\mathbf{Y}_\rho \in F) = \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-\mu(S)}}{n!} \int_{S^n} \mathbf{1}[\mathbf{x} \in F]\, \rho(x_1) \cdots \rho(x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n \tag{14}$$

where $\mathbf{x} = \{x_1, \ldots, x_n\}$.

By (14), $\mathbf{X}$ has *density* f with respect to $\mathbf{Y}_1$ if

$$\begin{aligned} \mathrm{P}(\mathbf{X} \in F) &= \mathrm{E}\left[\mathbf{1}[\mathbf{Y}_1 \in F] f(\mathbf{Y}_1)\right] \\ &= \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-|S|}}{n!} \int_{S^n} \mathbf{1}[\mathbf{x} \in F]\, f(\mathbf{x})\, \mathrm{d}x_1 \cdots \mathrm{d}x_n \end{aligned} \tag{15}$$

where $|S|$ is the area of S. Combining (3) and (15) it follows that

$$\rho^{(n)}(u_1, \ldots, u_n) = \mathrm{E} f(\mathbf{Y}_1 \cup \{u_1, \ldots, u_n\}) \tag{16}$$

for any $n \in \mathbb{N}$ and pairwise different points $u_1, \ldots, u_n \in S$. Conversely, under mild conditions, f can be expressed in terms of the product densities $\rho^{(n)}$ (Macchi, 1975). Furthermore, conditional on $n(\mathbf{X}) = n$ with $n \geq 1$, the n points in $\mathbf{X}$ have a symmetric joint density

$$f_n(x_1, \ldots, x_n) \propto f(\{x_1, \ldots, x_n\}) \tag{17}$$

on S^n.

Apart from the Poisson process and a few other simple models such as the mixed Poisson process (Grandell, 1997), the density is not on a closed form. For the Poisson process $\mathbf{Y}_\rho$, (14) gives

$$f(\mathbf{x}) = \mathrm{e}^{|S| - \mu(S)} \prod_{i=1}^{n} \rho(x_i). \tag{18}$$

Thus, for a Cox process driven by $\mathbf{\Lambda} = (\Lambda(u))_{u \in S}$,

$$f(\mathbf{x}) = \mathrm{E}\left[\exp\left(|S| - \int_S \Lambda(u)\,\mathrm{d}u\right) \prod_{u \in \mathbf{X}} \Lambda(u)\right] \tag{19}$$

which in general is not on a closed form.

5.2 The conditional intensity

The usual conditional intensity of a one-dimensional point process does not extend to two-dimensional point processes because of the lack of a natural ordering in $\mathbb{R}^2$. Instead, the Papangelou conditional intensity becomes the appropriate counterpart (Papangelou, 1974); a formal definition is given below.

Let the situation be as in Section 5.1, and suppose that f is *hereditary*, i.e.

$$f(\mathbf{x}) > 0 \text{ and } \mathbf{y} \subset \mathbf{x} \quad \Rightarrow \quad f(\mathbf{y}) > 0 \tag{20}$$

for finite point configurations $\mathbf{x} \subset S$. This positivity condition is usually assumed in practice.

Now, for locations $u \in S$ and finite point configurations $\mathbf{x} \subset S$, the Papangelou conditional intensity is defined by

$$\lambda(u, \mathbf{x}) = f(\mathbf{x} \cup \{u\})/f(\mathbf{x} \setminus \{u\}) \tag{21}$$

if $f(\mathbf{x} \setminus \{u\}) > 0$, and $\lambda(u, \mathbf{x}) = 0$ otherwise. The precise definition of $\lambda(u, \mathbf{x})$ when $u \in \mathbf{x}$ is not that important, and (21) just covers this case for completeness. Note that $\lambda(u, \mathbf{x}) = \lambda(u, \mathbf{x} \setminus \{u\})$, and (20) implies that f and λ are in a one-to-one correspondence.

For a Poisson process, the Papangelou conditional intensity is simply the intensity: if $f(\mathbf{x}) > 0$ is given by (18), then $\lambda(u, \mathbf{x}) = \rho(u)$ does not depend on $\mathbf{x}$, again showing the absence of interaction in a Poisson process.

Combining (16) and (20)–(21),

$$\rho(u) = \mathrm{E}\lambda(u, \mathbf{X}). \tag{22}$$

Recall that the conditional probability $\mathrm{P}(A|\mathbf{x})$ of an event A given $\mathbf{X} = \mathbf{x}$ satisfies $\mathrm{P}(A) = \mathrm{E}[\mathrm{P}(A|\mathbf{X})]$. Thus due to the infinitesimal interpretation of $\rho(u)\,\mathrm{d}u$ (Section 3.2), it follows from (22) that $\lambda(u, \mathbf{x})\,\mathrm{d}u$ may be interpreted as the conditional probability that there is a point of the process in an infinitesimal

region containing u and of area du given that the rest of the point process coincides with $\mathbf{x}$.

Often a density f is specified by an *unnormalized density* h, i.e. $f \propto h$ where h is an hereditary function, for which the normalizing constant $\mathrm{E}h(\mathbf{Y}_1)$ is well defined but unknown. However,

$$\lambda(u, \mathbf{x}) = h(\mathbf{x} \cup \{u\})/h(\mathbf{x} \setminus \{u\})$$

does not depend on the normalizing constant. This is one reason why inference and simulation procedures are often based on the conditional intensity rather than the density of the point process.

In practically all cases of spatial point process models, an unnormalized density h is *locally stable*, that is, there is a constant K such that

$$h(\mathbf{x} \cup \{u\}) \leq K h(\mathbf{x}) \tag{23}$$

for all $u \in S$ and finite $\mathbf{x} \subset S$. Local stability implies both that h is hereditary and integrable with respect to unit rate Poisson process. Local stability also plays a fundamental role when studying stability properties of MCMC algorithms (Section 9.2).

5.3 Finite Gibbs point processes

Consider again a finite point process $\mathbf{X}$ defined on the bounded region S and with hereditary density f. This is a *Gibbs point process* (also called a canonical ensemble in statistical physics) if

$$\log \lambda(u, \mathbf{x}) = \sum_{\mathbf{y} \subseteq \mathbf{x}} U(\mathbf{y} \cup \{u\}) \quad \text{when } f(\mathbf{x}) > 0 \tag{24}$$

where the function $U(\mathbf{x}) \in [-\infty, \infty)$ is defined for all non-empty finite point configurations $\mathbf{x} \subset S$, and we set $\log 0 = -\infty$. In statistical mechanical terms, U is a *potential*.

A large selection of Gibbs point process models are given in Van Lieshout (2000) and Møller & Waagepetersen (2003b). Usually, a log linear model is considered, where the first order potential is either constant or depends on spatial covariates

$$U(u) \equiv U(\{u\}) = z(u)\beta^{\mathsf{T}}$$

and higher order potentials are of the form

$$U(\mathbf{x}) = V(\mathbf{x})\psi_{n(\mathbf{x})}^{\mathsf{T}}, \quad n(\mathbf{x}) \geq 2$$

where the ψ_n are so-called interaction parameters. Then λ is parameterized by $\theta = (\beta, \psi_2, \psi_3, \ldots)$ and is on log linear form

$$\log \lambda_\theta(u, \mathbf{x}) = (t(\mathbf{x} \cup \{u\}) - t(\mathbf{x}))\theta^{\mathsf{T}} \tag{25}$$

where

$$t(\mathbf{x}) = \left(\sum_{u \in \mathbf{x}} z(u), \sum_{\mathbf{y} \subseteq \mathbf{x}:\, n(\mathbf{y})=2} V(\mathbf{y}), \sum_{\mathbf{y} \subseteq \mathbf{x}:\, n(\mathbf{y})=3} V(\mathbf{y}), \ldots \right). \tag{26}$$

Combining (21) and (24), the Gibbs process has density

$$f(\mathbf{x}) \propto \exp\left(\sum_{\emptyset \neq \mathbf{y} \subseteq \mathbf{x}} U(\mathbf{y}) \right) \tag{27}$$

defining $\exp(-\infty) = 0$. Unless $\mathbf{X}$ is Poisson, i.e. when $U(\mathbf{y}) = 0$ whenever $n(\mathbf{y}) \geq 2$, the normalizing constant of the density is unknown. Usually for models used in practice, $U(\mathbf{y}) \leq 0$ if $n(\mathbf{y}) \geq 2$, which implies local stability (and hence integrability). This means that the points in the process repel one other, so that realizations of the process tend to be more regular than for a Poisson process. Most Gibbs models are *pairwise interaction processes*, i.e. $U(\mathbf{y}) = 0$ whenever $n(\mathbf{y}) \geq 3$, and typically the second order potential depends on distance only, $U(\{u, v\}) = U(\|u - v\|)$. A *hard core process* with hard core $r > 0$ has $U(\{u, v\}) = -\infty$ whenever $\|u - v\| < r$.

By the *Hammersley-Clifford theorem* (Ripley & Kelly, 1977), any hereditary density is of the form (27) and the following properties (I) and (II) are equivalent.

(I) $U(\mathbf{x}) = 0$ whenever there exist two points $\{u, v\} \subseteq \mathbf{x}$ such that $\|u - v\| > R$.

(II) If $f(\mathbf{x}) > 0$ and $u \in S \setminus \mathbf{x}$, then $\lambda(u, \mathbf{x}) = \lambda(u, \mathbf{x} \cap b(u, R))$.

Here $b(u, R)$ is the closed disc with centre u and radius R. When (I) or (II) is satisfied, $\mathbf{X}$ is said to be *Markov* with *interaction radius* R, or more precisely, Markov with respect to the R-close neighbourhood relation. This definition and the Hammersley-Clifford theorem can be extended to an arbitrary symmetric relation on S (Ripley & Kelly, 1977) or even a relation which depends on realizations of the point process (Baddeley & Møller, 1989). Markov point

processes constitute a particular important subclass of Gibbs point processes, since the *local Markov property* (II) very much simplifies the computation of the Papangelou conditional intensity in relation with parameter estimation and simulation.

The local Markov property (II) implies a *spatial Markov property*. If $B \subset S$ and $\partial B = \{u \notin B : b(u, R) \cap B \neq \emptyset\}$ is its R-close neighbourhood, then the process $\mathbf{X}_B$ conditional on $\mathbf{X}_{B^c}$ depends only on $\mathbf{X}_{B^c}$ through $\mathbf{X}_{\partial B}$. The conditional process $\mathbf{X}_B|\mathbf{X}_{\partial B} = \mathbf{x}_{\partial B}$ is also Gibbs, with density

$$f_B(\mathbf{x}_B|\mathbf{x}_{\partial B}) \propto \exp\left(\sum_{\emptyset \neq \mathbf{y} \subseteq \mathbf{x}_B} U(\mathbf{y} \cup \mathbf{x}_{\partial B})\right) \tag{28}$$

where the normalizing constant depends on $\mathbf{x}_{\partial B}$ (the conditional density may be arbitrarily defined if $U(\mathbf{y}) = -\infty$ for some non-empty point configuration $\mathbf{y} \subseteq \mathbf{x}_{\partial B}$). The corresponding Papangelou conditional intensity is

$$\lambda(u, \mathbf{x}_B|\mathbf{x}_{\partial B}) = \lambda(u, \mathbf{x}_B \cup \mathbf{x}_{\partial B}), \quad u \in B. \tag{29}$$

Example 5.1. *(Overlap interaction model for Norwegian spruces)* The conditional intensity for a Norwegian spruce with a certain influence zone should depend not only on the positions but also on the influence zones of the neighbouring trees, see Figure 4. A tree with influence zone given by the disc $b(u, m_u)$, where u is the spatial location of the tree and m_u is the influence zone radius, is treated as a point (u, m_u) in $\mathbb{R}^3$. Confining ourselves to a pairwise interaction process, we define the pairwise potential by

$$U(\{(u, m_u), (v, m_v)\}) = \psi|b(u, m_u) \cap b(v, m_v)|, \quad \psi \leq 0.$$

Hence, the strength of the repulsion between two trees (u, m_u) and (v, m_v) is given by ψ times the area of overlap between the influence zones of the two trees. We assume that the influence zone radii belong to a bounded interval $M = [a, b]$, where a and b are estimated by the minimal and maximal observed influence zone radii. We divide M into six disjoint subintervals of equal size, and define the first order potential by

$$U((u, m_u)) = \beta(m_u) = \beta_k \quad \text{if } m_u \text{ falls in the } k\text{th subinterval}$$

where β_k is a real parameter. This enables modelling the varying numbers

of trees in the six different size classes. However, the interpretation of the conditional intensity

$$\lambda_\theta((u, m_u), \mathbf{x}) = \exp\left(\beta(m_u) + \psi \sum_{(v,m_v)\in\mathbf{x}} |b(u, m_u) \cap b(v, m_v)|\right) \tag{30}$$

is not straightforward — it is for instance not in general a monotone function of m_u. On the other hand, for a fixed (u, m_u), the conditional intensity will always decrease if the neighbouring influence zones increase.

Example 5.2. *(Hierarchical model for ants nests)* The hierarchical model in Högmander & Särkkä (1999) for the positions of ants nests is based on so-called *Strauss processes with hard cores* and interaction range $R = 45$ (distances are measured in ft). Details follow below.

For distances $t > 0$, define

$$V(t; r) = \begin{cases} -\infty & \text{if } t \le r \\ 1 & \text{if } r < t \le 45 \\ 0 & \text{otherwise} \end{cases}$$

where $r \ge 0$ denotes a hard core distance (or no hard core if $r = 0$). For the *Messor* nests, the Strauss process process with hard core r_M is given by first and second order potentials

$$U_{M1}(u) = \beta_M, \quad U_{M2}(\{u, v\}) = \psi_M V(\|u - v\|; r_M),$$

and no higher order interactions. The conditional intensity for a putative nest at a location u is thus zero if an existing nest occur within distance r_M from u, and otherwise the log conditional density is given by the sum of β_M and ψ_M times the number of neighbouring nests within distance 45. Given the pattern $\mathbf{x}_M$ 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

$$U_{C1}(u) = \beta_C + \psi_{CM} \sum_{v\in\mathbf{x}_M} V(\|u - v\|; r_{CM}), \quad U_{C2}(\{u, v\}) = \psi_C V(\|u - v\|; r_C).$$

Finally, the hard cores are estimated by the observed minimum interpoint distances, i.e. $r_M = 9.35$, $r_C = 2.45$, and $r_{CM} = 6.1$.

The Strauss processes with hard cores $r_M > 0$ and $r_C > 0$ are well-defined for all real values of the parameters β_M, β_C, ψ_M, ψ_{CM}, and ψ_C. However, as noted by Møller (1994), 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. We later also use the estimates $r_M = 9.35$ and $r_C = 2.45$, but we find it more natural to consider a model with no hard core between the two types of ants nests, i.e. to let $r_{CM} = 0$. Figure 6 shows the support of the covariate function $z_2(u) = \sum_{v \in \mathbf{x}_M} V(\|u - v\|; 0)$ for the *Cataglyphis* model with $r_{CM} = 0$.

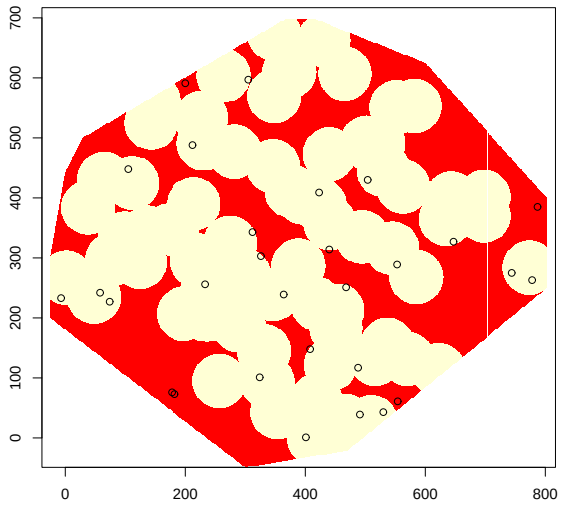


Figure 6: The white region is the set of $u \in W$ with distance less than 45 to a *Messor* nest. The dots show the locations of *Cataglyphis* nests.

5.4 Infinite Gibbs point processes

In general it is not possible to deal with densities of infinite point processes. For example, a stationary Poisson process has a density with respect to another stationary Poisson process if and only if their intensities are equal. However, the

Papangelou conditional intensity for a point process $\mathbf{X}$ on $\mathbb{R}^2$ can be indirectly defined as follows. If $\lambda(u, \mathbf{x})$ is a non-negative function defined for locations $u \in \mathbb{R}^2$ and locally finite point configurations $\mathbf{x} \subset \mathbb{R}^2$ such that

$$\mathrm{E} \sum_{u \in \mathbf{X}} h(u, \mathbf{X} \setminus \{u\}) = \int \mathrm{E}[\lambda(u, \mathbf{X}) h(u, \mathbf{X})] \, \mathrm{d}u \tag{31}$$

for all non-negative functions h, then λ is the Papangelou conditional intensity of $\mathbf{X}$. In fact two infinite point processes can share the same Papangelou conditional intensity; this phenomenon is known in statistical physics as *phase transition*.

The integral formula (31) is called the *Georgii-Nguyen-Zessin formula* (Georgii, 1976; Nguyen & Zessin, 1979), and this together with the Campbell theorem are basically the only known general formulae for spatial point processes. It is straightforward to verify (31) when $\mathbf{X}$ is defined on a bounded region, so that it is a finite point process with Papangelou conditional intensity (21). Using induction we obtain the *iterated GNZ-formula*

$$\begin{aligned} \mathrm{E} \sum_{x_1, \ldots, x_n \in \mathbf{X}}^{\neq} h(x_1, \ldots, x_n, \mathbf{X} \setminus \{x_1, \ldots, x_n\}) = \int \cdots \int \mathrm{E}[\lambda(x_1, \mathbf{X}) \lambda(x_2, \mathbf{X} \cup \{x_1\}) \\ \cdots \lambda(x_n, \mathbf{X} \cup \{x_1, \ldots, x_{n-1}\}) h(x_1, \ldots, x_n, \mathbf{X})] \, \mathrm{d}x_1 \cdots \mathrm{d}x_n \end{aligned} \tag{32}$$

for non-negative functions h. Combining (3) and (32), we see that

$$\rho^{(n)}(u_1, \ldots, u_n) = \mathrm{E}[\lambda(u_1, \mathbf{X}) \lambda(u_2, \mathbf{X} \cup \{u_1\}) \cdots \lambda(u_n, \mathbf{X} \cup \{u_1, \ldots, u_{n-1}\})]. \tag{33}$$

Notice that the iterated GNZ-formula (32) implies the Campbell theorem (4).

For instance, for a Cox process driven by Λ, the Papangelou conditional intensity is the Bayes estimator

$$\lambda(u, \mathbf{X}) = \mathrm{E}\left[\Lambda(u) \,|\, \mathbf{X}\right]. \tag{34}$$

However, this conditional expectation is usually unknown, and the GNZ-formula is more useful in connection with Gibbs point processes as described below.

The most common approach for defining a *Gibbs point process* $\mathbf{X}$ *on* $\mathbb{R}^2$ is to assume that $\mathbf{X}$ satisfies the spatial Markov property with respect to the R-close neighbourhood relation, and has conditional densities of a similar form as in the finite case. That is, for any bounded region $B \subset \mathbb{R}^2$, $\mathbf{X}_B | \mathbf{X}_{B^c}$ depends on $\mathbf{X}_{B^c}$

only through $\mathbf{X}_{\partial B}$, and (28) specifies the conditional density. An equivalent approach is to assume that $\mathbf{X}$ has a Papangelou conditional intensity, which in accordance with (28) and (29) satisfies $\lambda(u, \mathbf{X}) = \lambda(u, \mathbf{X} \cap b(u, R))$, where for finite point configurations $\mathbf{x} \subset \mathbb{R}^2$ and locations $u \in \mathbb{R}^2$,

$$\lambda(u, \mathbf{x}) = \exp\left(\sum_{\mathbf{y} \subseteq \mathbf{x}} U(\mathbf{y} \cup \{u\})\right) \quad \text{if } u \notin \mathbf{x}, \quad \lambda(u, \mathbf{x}) = \lambda(u, \mathbf{x} \setminus \{u\}) \quad \text{if } u \in \mathbf{x}.$$

Unfortunately, (33) is not of much use here, and in general a closed form expression for $\rho^{(n)}$ is unknown when $\mathbf{X}$ is Gibbs.

Questions of much interest in statistical physics are if a Gibbs process exists for λ specified by a given potential U as above, and if the process is unique (i.e. no phase transition) and stationary (even in that case it may not be unique); see Ruelle (1969), Preston (1976), Georgii (1976), Nguyen & Zessin (1979) or the review in Møller & Waagepetersen (2003b). These questions are of less importance in spatial statistics, where the process is observed within a bounded window W and, in order to deal with edge effects, we can use the so-called *border method*. That is, we base inference on $\mathbf{X}_{W_{\ominus R}} | \mathbf{X}_{\partial W_{\ominus R}}$, where $W_{\ominus R}$ is the clipped observation window

$$W_{\ominus R} = \{u \in W : b(u, R) \subset W\}$$

and the Papangelou conditional intensity is given by $\lambda(u, \mathbf{x}_{W_{\ominus R}} | \mathbf{x}_{\partial W_{\ominus R}}) = \lambda(u, \mathbf{x})$ when $\mathbf{X}_W = \mathbf{x}$ is observed. We return to this issue in Sections 6.1.3 and 7.2.

6 Exploratory and diagnostic tools

It is often difficult to assess the properties of a spatial point pattern by eye. A realization of a homogeneous Poisson process may for example appear clustered due to points which happen to be close just by chance. This section explains how to explore the features of a spatial point pattern with the aim of suggesting an appropriate model, and how to check and critize a fitted model. The residuals described in Section 6.1 are useful to assess the adequacy of the specified (conditional) intensity function in relation to a given data set. The second order properties specified by the pair correlation function and the distribution of interpoint distances may be assessed using the more classical summary statistics

in Section 6.2.

In this section, $\hat{\rho}$ and $\hat{\lambda}$ denote estimates of the intensity function and the Papangelou conditional intensity, respectively. These estimates may be obtained by non-parametric or parametric methods. In the stationary case, or at least if ρ is constant on S, a natural unbiased estimate is $\hat{\rho} = n/|W|$. In the inhomogeneous case, a non-parametric kernel estimate is

$$\hat{\rho}(u) = \sum_{i=1}^{n} k(u - x_i) / \int_W k(v - u)\,\mathrm{d}v \tag{35}$$

where k is a kernel with finite band width, and where the denominator is an edge correction factor ensuring that $\int_W \hat{\rho}(u)\,\mathrm{d}u$ is an unbiased estimate of $\mu(W)$ (Diggle, 1985). If the intensity or conditional intensity is specified by a parametric model, $\rho = \rho_\theta$ or $\lambda = \lambda_\theta$, and θ is estimated by $\hat{\theta}(\mathbf{x})$ (Sections 7–8), we let $\hat{\rho} = \rho_{\hat{\theta}(\mathbf{x})}$ or $\hat{\lambda} = \lambda_{\hat{\theta}(\mathbf{x})}$.

6.1 Residuals

For a Gibbs point process with log Papangelou conditional intensity (24), the first order potential corresponds to the linear predictor of a generalised linear model (GLM), while the higher order potentials are roughly analogous to the distribution of the errors in a GLM. Recently, Baddeley, Turner, Møller & Hazelton (2005) developed a residual analysis for spatial point processes based on the GNZ-formula (31) and guided by the analogy with residual analysis for (non-spatial) GLM's. For a Cox process, the Papangelou conditional intensity (34) is usually not known on closed form, while the intensity function may be tractable. In such cases, Waagepetersen (2005) suggested residuals be defined using instead the intensity function. Whether we base residuals on the conditional intensity or the intensity, the two approaches are very similar.

6.1.1 Definition of innovations and residuals

For ease of exposition we assume first that the point process $\mathbf{X}$ is defined on the observation window W; the case where $\mathbf{X}$ extends outside W is considered in Section 6.1.3.

For non-negative functions $h(u, \mathbf{x})$, define the *h-weighted innovation* by

$$I_h(B) = \sum_{u \in \mathbf{X}_B} h(u, \mathbf{X} \setminus \{u\}) - \int_B \lambda(u, \mathbf{X}) h(u, \mathbf{X})\,\mathrm{d}u, \quad B \subseteq W. \tag{36}$$

We will allow infinite values of $h(u, \mathbf{x})$ if $u \in \mathbf{x}$, in which case we define $\lambda(u, \mathbf{x})h(u, \mathbf{x}) = 0$ if $\lambda(u, \mathbf{x}) = 0$. Baddeley *et al.* (2005) study in particular the *raw, Pearson, and inverse-λ* innovations given by $h(u, \mathbf{x}) = 1$, $1/\sqrt{\lambda(u, \mathbf{x})}$, $1/\lambda(u, \mathbf{x})$, respectively. Note that I_h is a signed measure, where we may interpret $\Delta I(u) = h(u, \mathbf{X} \setminus \{u\})$ as the innovation increment ('error') attached to a point u in $\mathbf{X}$, and $\mathrm{d}I(u) = -\lambda(u, \mathbf{X})h(u, \mathbf{X})\mathrm{d}u$ as the innovation increment attached to a background location $u \in W$. Assuming that the sum or equivalently the integral in (36) has finite mean, the GNZ-formula (31) gives

$$\mathrm{E}I_h(B) = 0. \tag{37}$$

The *h-weighted residual* is defined by

$$R_{\hat{h}}(B) = \sum_{u \in \mathbf{x}_B} \hat{h}(u, \mathbf{x} \setminus \{u\}) - \int_B \hat{\lambda}(u, \mathbf{x})\hat{h}(u, \mathbf{x})\,\mathrm{d}u, \quad B \subseteq W, \tag{38}$$

where, as the function h may depend on the model, $\hat{h}$ denotes an estimate. This also is a signed measure, and we hope that the mean of the residual measure is approximately zero. The raw, Pearson, and inverse-λ residuals are

$$R(B) = n(\mathbf{x}) - \int_B \hat{\lambda}(u, \mathbf{x})\,\mathrm{d}u,$$

$$R_{1/\sqrt{\hat{\lambda}}}(B) = \sum_{u \in \mathbf{x}_B} 1/\sqrt{\hat{\lambda}(u, \mathbf{x})} - \int_B \sqrt{\hat{\lambda}(u, \mathbf{x})}\,\mathrm{d}u,$$

$$R_{1/\hat{\lambda}}(B) = \sum_{u \in \mathbf{x}_B} 1/\hat{\lambda}(u, \mathbf{x}) - \int_B \mathbf{1}[\hat{\lambda}(u, \mathbf{x}) > 0]\,\mathrm{d}u.$$

In order that the Pearson and inverse-λ residuals be well defined, we require that $\hat{\lambda}(u, \mathbf{x}) > 0$ for all $u \in \mathbf{x}$. Properties of these innovations and residuals are analyzed in Baddeley, Møller and Pakes (2006).

Similarly, we define innovations and residuals based on ρ, where we in all expressions above replace λ and $\hat{\lambda}$ by ρ and $\hat{\rho}$, respectively, and $h(u, \mathbf{x})$ and $\hat{h}(u, \mathbf{x})$ by $h(u)$ and $\hat{h}(u)$, respectively. Here it is required that $\int_W h(u)\rho(u)\,\mathrm{d}u < \infty$, so that (37) also holds in this case.

6.1.2 Diagnostic plots

Baddeley *et al.* (2005) suggest various *diagnostic plots* for spatial trend, dependence of covariates, interaction between points, and other effects. In particular, the plots can check for the presence of such features when the fitted model does not include them. The plots are briefly described below in the case of residuals based on λ; if we instead consider residuals based on ρ, we use the same substitutions as in the preceding paragraph. Figures 7 and 8 show specific examples of the plots in the case of the *Cataglyphis* nests model (Examples 5.2) fitted in Example 8.2 and based on raw residuals ($h \equiv 1$). The plots are corrected for edge effects, cf. Section 6.1.3.

The *mark plot* is a pixel image with greyscale proportional to $\hat{\lambda}(u, \mathbf{x})\hat{h}(u, \mathbf{x})$ and a circle centred at each point $u \in \mathbf{x}$ with radius proportional to the residual mass $\hat{h}(u, \mathbf{x} \setminus \{u\})$. The plot may sometimes identify 'extreme points'. For example, for Pearson residuals and a fitted model of correct form, large/small circles and dark/light greyvalues should correspond to low/high values of the conditional intensity, and in regions of the same greylevel the circles should be uniformly distributed. The upper left plot in Figure 7 is a mark plot for the raw residuals obtained from the model fitted to the *Cataglyphis* nests in Example 8.2. In this case, the circles are by definition of the same radii and just show the locations of the nests. In the region of the large cluster of circles one could perhaps have expected larger values (more light grey scales) of the fitted conditional intensity.

The *smoothed residual field* at location $u \in W$ is

$$s(u, \mathbf{x}) = \frac{\sum_1^n k(u - x_i)\hat{h}(x_i, \mathbf{x} \setminus \{x_i\}) - \int_W k(u - v)\hat{\lambda}(u, \mathbf{x})\hat{h}(v, \mathbf{x})\,\mathrm{d}v}{\int_W k(u - v)\,\mathrm{d}v} \tag{39}$$

where k is a kernel and the denominator is an edge correction factor. For example, for raw residuals, the numerator of (39) has mean $\int_W k(u - v)\mathbb{E}[\lambda(v, \mathbf{X}) - \hat{\lambda}(v, \mathbf{X})]\,\mathrm{d}v$, so positive/negative values of s suggest that the fitted model under/overestimates the intensity function. The smoothed residual field may be presented as a greyscale image and a contour plot, see the lower right plot in Figure 7 which suggests some underestimation of the conditional intensity at the middle of the plot and overestimation in the top part of the plot.

For a given covariate $z : W \mapsto \mathbb{R}$ and numbers t, define $W(t) = \{u \in W : z(u) \leq t\}$. A plot of the 'cumulative residual function' $A(t) = R_{\hat{h}}(W(t))$ is called a *lurking variable plot*, since it may detect if z should be included in the

model. If the fitted model is correct, we expect $A(t) \approx 0$. The upper right and lower left plots in Figure 7 show lurking variable plots for the covariates given by the y and x spatial coordinates, respectively. The upper right plot indicates (in accordance with the lower right plot) a decreasing trend in the y direction, whereas there is no indication of trend in the x direction. The possible defects of the model indicated by the right plots in Figure 7 might to be related to inhomogeneity due to that the observation window consists of a 'field' and a 'scrub' part divided by a boundary which runs roughly along the diagonal from the lower left to the upper right corner (Harkness & Isham, 1983). Including covariates given by an indicator for the field and the spatial y-coordinate improved somewhat the appearance of the diagnostic plots.

Baddeley *et al.* (2005) also consider a *Q-Q plot* comparing empirical quantiles of $s(u, \mathbf{x})$ with corresponding expected empirical quantiles estimated from $s(u, \mathbf{x}^{(1)}), \ldots, s(u, \mathbf{x}^{(n)})$, where $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$ are simulations from the fitted model. This is done using a grid of fixed locations $u_j \in W$, $j = 1, \ldots, J$. For each $k = 0, \ldots, n$, where $\mathbf{x}^{(0)} = \mathbf{x}$ is the data, we sort $s_j^{(k)} = s(u_j, \mathbf{x}^{(k)})$, $j = 1, \ldots, J$ to obtain the order statistics $s_{[1]}^{(k)} \leq \ldots \leq s_{[J]}^{(k)}$. We then plot $s_{[j]}^{(0)}$ versus the estimated expected empirical quantile $\sum_{k=1}^{n} s_{[j]}^{(k)}/n$ for $j = 1, \ldots, J$. The Q-Q plot in Figure 8 shows some deviations between the observed and estimated quantiles but each observed order statistic fall within the 95% intervals obtained from corresponding simulated order statistics.

6.1.3 Edge effects

Substantial bias and other artifacts in the diagnostic plots for residuals based on λ may occur if edge effects are ignored. We therefore use the border method as follows (see also Baddeley, Møller and Pakes, 2006). Suppose the fitted model is Gibbs with interaction radius R (Sections 5.3-5.4). For locations u in $W \setminus W_{\ominus R} = \partial W_{\ominus R}$, $\lambda(u, \mathbf{x})$ may depend on points in $\mathbf{x}$ which are outside the observation window W. Since the Papangelou conditional intensity (29) with $B = W_{\ominus R}$ does not depend on points outside the observation window, we condition on $\mathbf{X}_{\partial W_{\ominus R}} = \mathbf{x}_{\partial W_{\ominus R}}$ and plot residuals only for $u \in W_{\ominus R}$. See e.g. the upper left plot in Figure 7.

For residuals based on ρ instead, we have no edge effects, so no adjustment of the diagnostic tools in Section 6.1.2 is needed.

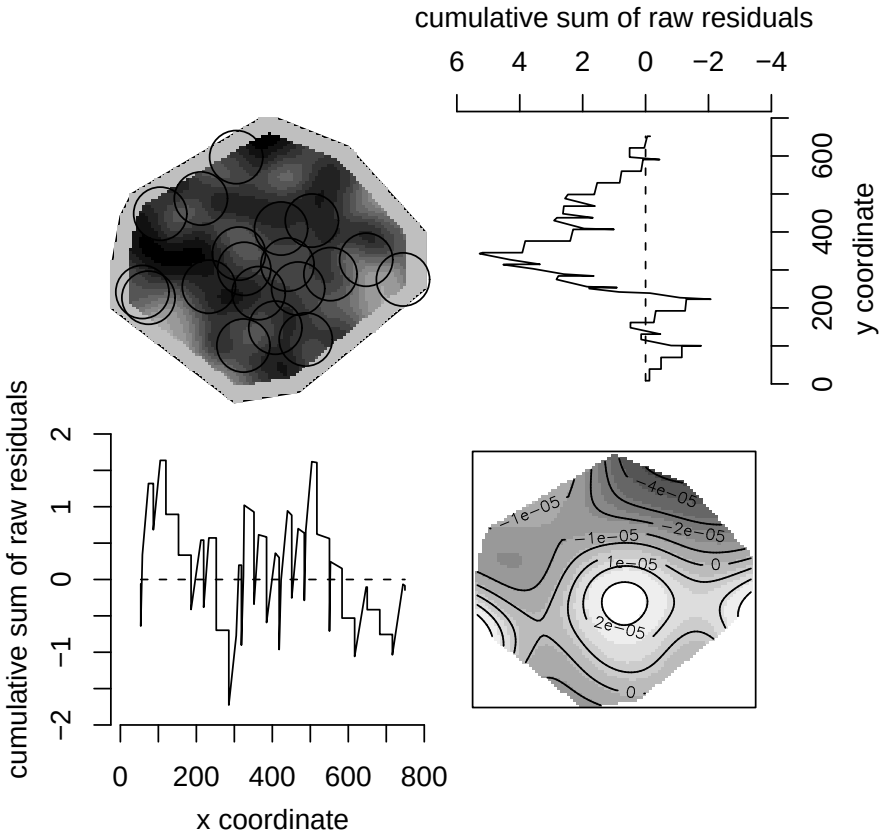


Figure 7: Plots for *Cataglyphis* nests based on raw residuals: mark plot (upper left), lurking variable plots for covariates given by y and x coordinates (upper right, lower left), and smoothed residual field (lower right). Dark grey scales correspond to small values.

6.2 Summary statistics

This section considers the more classical summary statistics such as Ripley's K-function and the nearest-neighbour function G. See also Baddeley, Møller and Waagepetersen (2006) who develop residual versions of such summary statistics.

6.2.1 Second order summary statistics

Second order properties are described by the pair correlation function g, where it is convenient if $g(u, v)$ only depends on the distance $\|u - v\|$ or at least the difference $u - v$ (note that $g(u, v)$ is symmetric). Kernel estimation of g

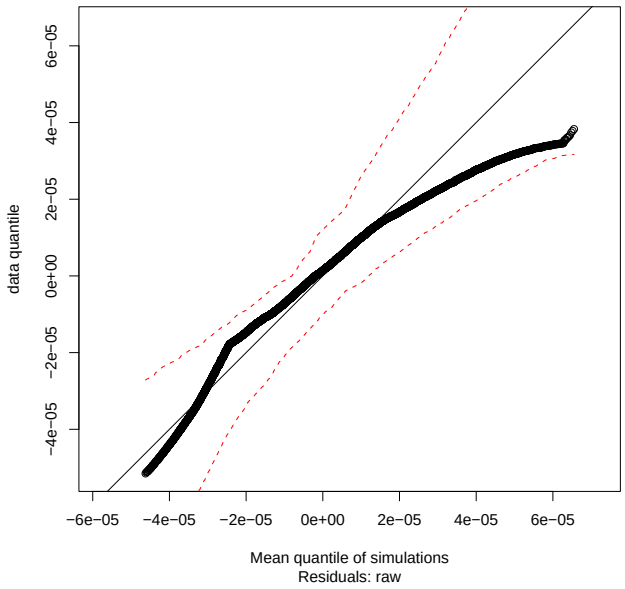


Figure 8: Q-Q plot for *Cataglyphis* nests based on smoothed raw residual field. The dotted lines show the 2.5% and 97.5% percentiles for the simulated order statistics

is discussed in Stoyan & Stoyan (2000). Alternatively, if $g(u,v) = g(u-v)$ is translation invariant, one may consider the *inhomogeneous reduced second moment measure* (Baddeley *et al.*, 2000)

$$\mathcal{K}(B) = \int_B g(u)\,\mathrm{d}u, \quad B \subseteq \mathbb{R}^2.$$

More generally, if g is not assumed to exist or to be translation invariant, we may define

$$\mathcal{K}(B) = \frac{1}{|A|}\mathrm{E}\sum_{u\in\mathbf{X}_A}\sum_{v\in\mathbf{X}\setminus\{u\}}\frac{\mathbf{1}[u-v\in B]}{\rho(u)\rho(v)} \tag{40}$$

provided that $\mathbf{X}$ is *second order reweighted stationary* which means that the right hand side of (40) does not depend on the choice of $A \subset \mathbb{R}^2$, where $0 < |A| < \infty$. Note that $\mathcal{K}$ is invariant under independent thinning.

The *(inhomogeneous) K-function* is defined by $K(r) = \mathcal{K}(b(0,r))$, $r > 0$. Clearly, if $g(u,v) = g(\|u-v\|)$, then $\mathcal{K}$ is determined by K, and $K(r) = 2\pi\int_0^r sg(s)\,\mathrm{d}s$, so that g and K are in a one-to-one correspondence. In the stationary case of $\mathbf{X}$, it follows from (40) that $\rho K(r)$ has the interpretation as

the expected number of further points within distance r from a typical point in $\mathbf{X}$, and $\rho^2 K(r)/2$ is the expected number of (unordered) pairs of distinct points not more than distance r apart and with at least one point in a set of unit area (Ripley, 1976). A formal definition of 'typical point' is given in terms of Palm measures, see e.g. Møller & Waagepetersen (2003b). For a Poisson process, $K(r) = \pi r^2$.

In our experience, non-parametric estimation of K is more reliable than that of g, since the latter involves kernel estimation, which is sensitive to the choice of the band width. Various edge corrections have been suggested, the simplest and most widely applicable being

$$\hat{K}(r) = \sum_{u,v \in \mathbf{x}}^{\neq} \frac{\mathbf{1}[\|u - v\| \leq r]}{\hat{\rho}(u)\hat{\rho}(v)|W \cap W_{u-v}|} \tag{41}$$

where W_u is W translated by u, and $\hat{\rho}$ is an estimate of the intensity function. One possibility is the non-parametric estimate of ρ given in (35) but the resulting estimate $\hat{K}(r)$ is very sensitive to the choice of kernel band width. In general we prefer to use a parametric estimate of the intensity function.

An estimate of the K-function for the tropical rain forest trees obtained with a parametric estimate of the intensity function (see Example 8.1) is shown in Figure 9. The plot also shows theoretical K-functions for fitted log Gaussian Cox, Thomas, and Poisson processes, where all three processes share the same intensity function (details are given later in Example 8.3). The trees seem to form a clustered point pattern since the estimated K-function is markedly larger than the theoretical K-function for a Poisson process.

One often considers the *L-function* $L(r) = \sqrt{K(r)/\pi}$, which is a variance stabilizing transformation when K is estimated by non-parametric methods (Besag, 1977). Moreover, for a Poisson process, $L(r) = r$. In general, at least for small distances, $L(r) > r$ indicates aggregation and $L(r) < r$ indicates regularity. Usually when a model is fitted, $\hat{L}(r) = \sqrt{\hat{K}(r)/\pi}$ or $\hat{L}(r) - r$ is plotted together with the average and 2.5% and 97.5% quantiles based on simulated $\hat{L}$-functions under the fitted model; we refer to these bounds as 95% *envelopes*. Examples are given in the right plots of Figures 11 and 12.

Estimation of third-order properties and of directional properties (so-called directional K-functions) is discussed in Stoyan & Stoyan (1995), Møller *et al.* (1998), and Schladitz & Baddeley (2000).

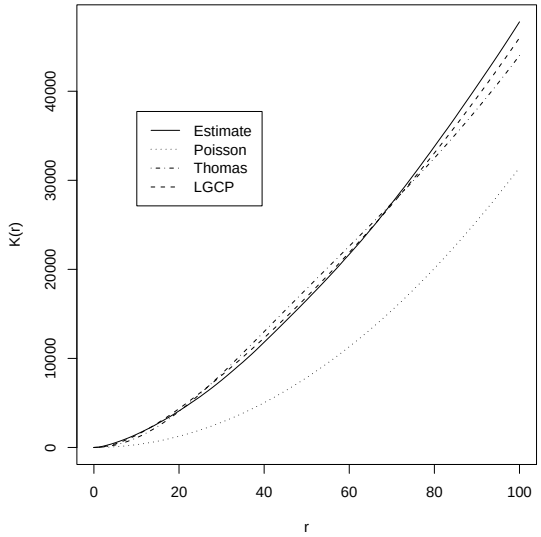


Figure 9: Estimated K-function for tropical rain forest trees and theoretical K-functions for fitted Thomas, log Gaussian Cox, and Poisson processes.

6.2.2 Interpoint distances

In order to interpret the following summary statistics based on interpoint distances, we assume stationarity of $\mathbf{X}$. The *empty space function* F is the distribution function of the distance from an arbitrary location to the nearest point in $\mathbf{X}$,

$$F(r) = \mathrm{P}(\mathbf{X} \cap b(0,r) \neq \emptyset), \quad r > 0.$$

The *nearest-neighbour function* is defined by

$$G(r) = \frac{1}{\rho|W|}\mathrm{E} \sum_{u \in \mathbf{X} \cap W} \mathbf{1}[(\mathbf{X} \setminus \{u\}) \cap b(u,r)], \quad r > 0,$$

which has the interpretation as the cumulative distribution function for the distance from a 'typical' point in $\mathbf{X}$ to its nearest-neighbour point in $\mathbf{X}$. Thus, for small distances, $G(r)$ and $\rho K(r)$ are closely related. For a stationary Poisson process, $F(r) = G(r) = 1 - \exp(-\pi r^2)$. In general, at least for small distances, $F(r) > G(r)$ indicates aggregation and $F(r) < G(r)$ indicates regularity. Van Lieshout & Baddeley (1996) study the nice properties of the *J-function* defined

by $J(r) = (1 - G(r))/(1 - F(r))$ for $F(r) < 1$.

Non-parametric estimation of F and G which account for edge effects is straightforward using border methods, see e.g. Møller & Waagepetersen (2003b). An estimate of J is obtained by plugging in the estimates of F and G in the expression for J. We combine the estimates to obtain an estimate of J. Estimates of F, G, and J for the positions of Norwegian spruces shown in Figure 10 provide evidence of repulsion.

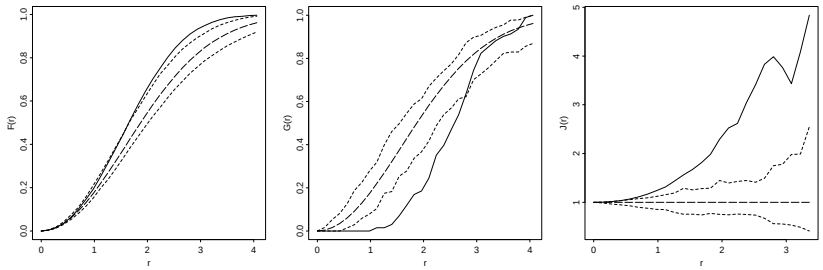


Figure 10: Left to right: estimated F, G, and J-functions for the Norwegian spruces (solid lines) and 95% envelopes calculated from simulations of a homogeneous Poisson process (dashed lines) with expected number of points equal to the observed number of points. The long-dashed curves show the theoretical values of F, G, and J for a Poisson process.

7 Likelihood-based inference and MCMC methods

Computation of the likelihood function is usually easy for Poisson process models (Section 7.1), while the likelihood contains an unknown normalizing constant for Gibbs point process models, and is given in terms of a complicated integral for Cox process models. Using MCMC methods, it is now becoming quite feasible to compute accurate approximations of the likelihood function for Gibbs and Cox process models (Sections 7.2 and 7.3). However, the computations may be time consuming and standard software is yet not available. Quick non-likelihood approaches to inference are reviewed in Section 8.

7.1 Poisson process models

For a Poisson process with a parameterized intensity function ρ_θ, the log likelihood function is

$$l(\theta) = \sum_{u \in \mathbf{x}} \log \rho_\theta(u) - \int_W \rho_\theta(u)\, \mathrm{d}u, \tag{42}$$

cf. (18), where in general numerical integration is needed to compute the integral. A clever implementation for finding the maximum likelihood estimate (MLE) numerically, based on software for generalized linear models (Berman and Turner, 1992), is available in `spatstat` when the intensity function is on the log linear form (7).

Rathbun & Cressie (1994) study increasing domain asymptotics for inhomogeneous Poisson point processes and provide fairly weak conditions for asymptotic normality of the MLE in the case of a log linear intensity function. Waagepetersen (2005) instead suggests asymptotics for a fixed observation window when the intercept in the log linear intensity function tends to infinity, and the only condition for asymptotic normality of the MLE of the remaining parameters is positive definiteness of the observed information matrix. Inference for a log linear Poisson process model is exemplified in Example 8.1.

7.2 Gibbs point process models

We restrict attention to parametric models for Gibbs point processes $\mathbf{X}$ as in Sections 5.3–5.4, assuming that the interaction radius R is finite and the conditional intensity is on the log linear form (25) (no matter whether $\mathbf{X}$ is finite or infinite). We assume to begin with that R is known.

First, suppose that the observation window W coincides with S. The density is then on exponential family form

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

where t is given by (26) and c_θ is the unknown normalizing constant. The score function and observed information are

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

where E_θ and Var_θ denote expectation and variance with respect to $\mathbf{X} \sim f_\theta$.

Consider a fixed reference parameter value θ_0. The score function and ob-

served information may then be evaluated using the *importance sampling formula*

$$\mathrm{E}_\theta k(\mathbf{X}) = \mathrm{E}_{\theta_0}\left[k(\mathbf{X})\exp\left(t(\mathbf{X})(\theta-\theta_0)^\mathsf{T}\right)\right]/(c_\theta/c_{\theta_0}) \tag{43}$$

with $k(\mathbf{X})$ given by $t(\mathbf{X})$ or $t(\mathbf{X})^\mathsf{T}t(\mathbf{X})$. The importance sampling formula also yields

$$c_\theta/c_{\theta_0} = \mathrm{E}_{\theta_0}\left[\exp\left(t(\mathbf{X})(\theta-\theta_0)^\mathsf{T}\right)\right]. \tag{44}$$

Approximations of the likelihood ratio $f_\theta(\mathbf{x})/f_{\theta_0}(\mathbf{x})$, score, and observed information are then obtained by Monte Carlo approximation of the expectations $\mathrm{E}_{\theta_0}[\cdots]$ using MCMC samples from f_{θ_0}, see Section 9.2.

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

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

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

Second, suppose that W is strictly contained in S and let $f_{W,\theta}(\mathbf{x}|\mathbf{x}_{\partial W})$ denote the conditional density of $\mathbf{X}_W$ given $\mathbf{X}_{\partial W} = \mathbf{x}_{\partial W}$. The likelihood function

$$L(\theta) = \mathrm{E}_\theta f_{W,\theta}\left(\mathbf{x}|\mathbf{X}_{\partial W}\right)$$

may be computed using a missing data approach, see Geyer (1999) and Møller & Waagepetersen (2003b). A simpler alternative is to consider the conditional likelihood function given by

$$f_{W_{\ominus R},\theta}(\mathbf{x}_{W_{\ominus R}}|\mathbf{x}_{\partial W_{\ominus R}})$$

where the score, observed information, and likelihood ratios may be computed in analogy with the $W = S$ case, cf. Sections 5.3–5.4.

For a fixed R, the approximated (conditional) likelihood function can be maximized with respect to θ using Newton-Raphson updates. In our experience the Newton-Raphson updates converge quickly, and in the examples below, the computing times for obtaining a MLE are modest — less than half a minute.

MLE's of R are often found using a profile likelihood approach, since the likelihood function is typically not differentiable and log concave as a function of R.

Asymptotic results for MLE's of Gibbs point process models are reviewed in Møller and Waagepetersen (2003b) 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 MLE, but if one is suspicious about the validity of this approach, an alternative is to use a parametric bootstrap.

Example 7.1. *(Maximum likelihood estimation for overlap interaction model)* For the overlap interaction model in Example 5.1, Møller & Waagepetersen (2003b) compute maximum likelihood estimates using both missing data and conditional likelihood approaches. Letting $W = [0, 56] \times [0, 38]$, the conditional likelihood approach is based on the trees with locations in $W_{\ominus 2b}$, since trees with locations outside W do not interact with trees located inside $W_{\ominus 2b}$. The conditional MLE is given by $(\hat{\beta}_1, \ldots, \hat{\beta}_6) = (-1.02, -0.41, 0.60, -0.67, -0.58, -0.22)$ and $\hat{\psi} = -1.13$. Confidence intervals for ψ obtained from the observed information and a parametric bootstrap are $[-1.61, -0.65]$ and $[-1.74, -0.79]$, respectively. As expected, due to the repulsive interaction term in the conditional intensity (30), the $\hat{\beta}_k$ tend to be larger than expected under the Poisson model with $\psi = 0$. This is illustrated in Figure 11 (left plot), where the $\exp(\hat{\beta}_k)$ are shown together with relative frequencies of trees within each of the six size classes (the frequencies are proportional to the MLE of the $\exp(\beta_k)$ under the Poisson model). The fitted overlap interaction process seems to capture well the second order characteristics for the point pattern of tree locations, see Figure 11 (right plot).

Example 7.2. *(Maximum likelihood estimation for ants nests)* Högmander & Särkkä (1999) consider a subset of the data in Figure 5 within a rectangular region, and they condition on the observed number of points for the two species when fitting the hierarchical model described in Example 5.2, whereby the parameters β_M and β_C vanish. Instead we fit the hierarchical model to the full data set, we do not condition on the observed number of points, and we set $r_{CM} = 0$. No edge correction is used for our MLE's, but in Example 8.2 we compare maximum pseudo likelihood estimates (Section 8.1) obtained both with and without edge correction. The MLE's $\hat{\beta}_M = -8.39$ and $\hat{\psi}_M = -0.41$ indicate a repulsion within the *Messor* nests, and the MLE's $\hat{\beta}_C = -10.3$, $\hat{\psi}_{CM} = 0.90$,

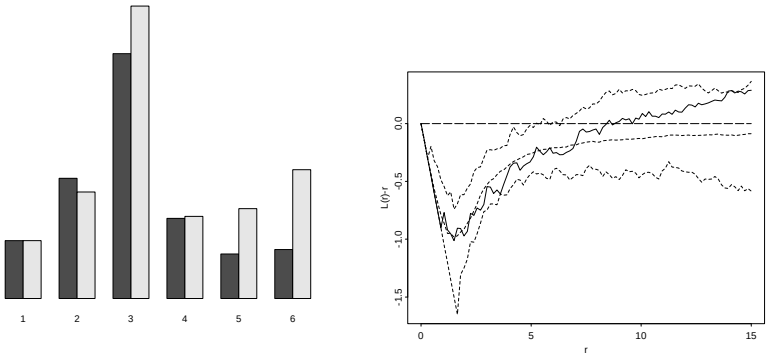


Figure 11: Dark grey bars: frequencies of trees for the six size classes (scaled so that light and dark bars are of the same height for the first class). Light gray bars: MLE of $\exp(\beta_k)$, $k = 1, \ldots, 6$. Right plot: estimated $L(r) - r$ function for spruces (solid line) and average and 95% envelopes computed from simulations of fitted overlap interaction model (dashed lines).

and $\hat{\psi}_C = -0.06$ indicates positive association between *Messor* and *Cataglyphis* nests, and a weak repulsion within the *Cataglyphis* nests. Confidence intervals for ψ_{CM} are $[-0.1, 1.9]$ (based on observed information) and $[0.3, 2.1]$ (parametric bootstrap). Due to the phase transition property of the Strauss hard core process (Example 5.2), we restrict $\psi_C \leq 0$ in the Newton-Raphson maximizations for the bootstrap simulated data sets. In this case, the two types of confidence intervals provide qualitatively different conclusions concerning the significance of the interspecies interaction. The results in Högmander & Särkkä (1999) differ from ours, since they estimate a strong repulsion within the *Cataglyphis* nests and a weak repulsion between the two species; see also Example 8.2.

7.3 Cox process models

We consider MLE for parametric models for first the case of a shot noise Cox process and second a log Gaussian Cox process.

In the case of a shot noise Cox process (Section 4.2.2), suppose that the parameter vector $\theta = (\alpha, \omega)$ consists of components α and ω parameterizing respectively the intensity function ζ_α of $\mathbf{\Phi}$ and the kernel $k(c, \cdot) = k(c, \cdot; \omega)$. Let $f(\mathbf{x}|\mathbf{\Lambda})$ denote the Poisson density of $\mathbf{X}_W$ given $\mathbf{\Lambda}(\cdot) = \mathbf{\Lambda}(\cdot; \mathbf{\Phi}, \omega)$. For simplicity assume that k is of bounded support, i.e. there exists a bounded region $\tilde{W} = \tilde{W}_\omega \supset W$ so that $k(c, u; \omega) = 0$ whenever $c \in \mathbb{R}^2 \setminus \tilde{W}$ and $u \in W$.

The likelihood

$$L(\theta) = \mathrm{E}_\alpha f\big(\mathbf{x}|\mathbf{\Lambda}(\cdot;\mathbf{\Phi},\omega)\big) = \mathrm{E}_\alpha f\big(\mathbf{x}|\mathbf{\Lambda}(\cdot;\mathbf{\Phi}_{\tilde{W}},\omega)\big)$$

is then given in terms of an expectation with respect to the Poisson process $\mathbf{\Phi}_{\tilde{W}} = \{(c,\gamma) \in \mathbf{\Phi} | c \in \tilde{W}\}$. We assume moreover that $\int_0^\infty \zeta_\alpha(c,\gamma)\mathrm{d}\gamma < \infty$ whenever $c \in \tilde{W}$. Thereby $\mathbf{\Phi}_{\tilde{W}}$ is finite and we let $f_{\tilde{W}}(\cdot;\alpha)$ denote the Poisson process density of $\mathbf{\Phi}_{\tilde{W}}$. Choose a reference parameter value $\theta_0 = (\alpha_0, \omega_0)$. Then $L(\theta)$ is the normalizing constant of $f(\mathbf{x}|\mathbf{\Lambda}(\cdot;\varphi,\omega))f_{\tilde{W}}(\varphi;\alpha)$ viewed as an unnormalized density for the conditional distribution of $\mathbf{\Phi}_{\tilde{W}}$ given $\mathbf{X}_W = \mathbf{x}$. Consequently, in analogy with (44),

$$L(\theta)/L(\theta_0) = \mathrm{E}_{\alpha_0}\left[\frac{f\big(\mathbf{x}|\Lambda(\cdot;\Phi_{\tilde{W}},\omega)\big)f_{\tilde{W}}(\Phi_{\tilde{W}};\alpha)}{f\big(\mathbf{x}|\Lambda(\cdot;\Phi_{\tilde{W}},\omega_0)\big)f_{\tilde{W}}(\Phi_{\tilde{W}};\alpha_0)}\middle|\mathbf{X}_W = \mathbf{x}\right] \tag{45}$$

which can be approximated using samples from the conditional distribution of $\mathbf{\Phi}_{\tilde{W}}$ given $\mathbf{X}_W = \mathbf{x}$ and $\theta = \theta_0$. Let

$$V_{\theta,\mathbf{x}}(\mathbf{\Phi}_{\tilde{W}}) = \mathrm{d}\log\big(f(\mathbf{x}|\mathbf{\Lambda}(\cdot;\mathbf{\Phi}_{\tilde{W}},\omega))f_{\tilde{W}}(\mathbf{\Phi}_{\tilde{W}};\alpha)\big)/\mathrm{d}\theta.$$

The score function and observed information are given by

$$u(\theta) = \mathrm{E}_\theta[V_{\theta,\mathbf{x}}(\mathbf{\Phi}_{\tilde{W}})|\mathbf{X}_W = \mathbf{x}]$$

and

$$j(\theta) = -\mathrm{E}_\theta\left[\mathrm{d}V_{\theta,\mathbf{x}}(\mathbf{\Phi}_{\tilde{W}})/\mathrm{d}\theta^{\mathsf{T}}|\mathbf{X}_W = \mathbf{x}\right] - \mathrm{Var}_\theta\left[V_{\theta,\mathbf{x}}(\mathbf{\Phi}_{\tilde{W}})|\mathbf{X}_W = \mathbf{x}\right]$$

where approximations of these conditional expectations can be obtained by applying importance sampling (Section 8.6.2 in Møller & Waagepetersen, 2003b). Samples from the conditional distribution of $\Phi_{\tilde{W}}$ can be generated using MCMC, see Section 9.2.

For a log Gaussian Cox process (Section 4.2.1), we consider a finite partition C_i, $i \in I$, of W and approximate the Gaussian process $(\Psi(u))_{u\in W}$ by a step function with value $\Psi(u_i)$ within C_i, where u_i is a representative point in C_i. We then proceed in a similar manner as for shot noise Cox processes, but now computing conditional expectations with respect to the finite Gaussian vector $(\Psi(u_i))_{i\in I}$ given $\mathbf{X}_W = \mathbf{x}$. Conditional samples of $(\Psi(u_i))_{i\in I}$ may be obtained

using Langevin-Hastings MCMC algorithms, see Section 10.2.3 in Møller & Waagepetersen, 2003b).

Asymptotic results for MLE's have been established for certain Cox process models defined on the real line, see Jensen (2005) and the references therein, but we are not aware of any such results for spatial Cox processes.

Example 7.3. *(Maximum likelihood estimation for North Atlantic whales)* For the shot noise Cox process model in Example 4.2, the unknown parameters are the intensity κ of the cluster centres, the mean number α of whales per cluster, and the standard deviation ω of the Gaussian density. Since it is difficult to evaluate the components of the score function and observed information corresponding to the parameter ω, Waagepetersen & Schweder (2005) compute the profile log likelihood function $l_p(\omega) = \max_{(\kappa,\alpha)} \log L(\theta)$ for a finite set of values ω_l. This is done using (45) repeatedly, i.e. by cumulating log likelihood ratios $\log L(\hat{\theta}_{l+1}) - \log L(\hat{\theta}_l)$, where $\hat{\theta}_l = (\hat{\kappa}_l, \hat{\alpha}_l, \omega_l)$ and $(\hat{\kappa}_l, \hat{\alpha}_l) = \arg\max_{(\kappa,\alpha)} \log L(\kappa, \alpha, \omega_l)$ is obtained using Newton-Raphson. The profile likelihood function is shown in Figure 12 (left plot) and gives $\hat{\omega} = 0.6$ with corresponding values $\hat{\kappa} = 0.025$ and $\hat{\alpha} = 2.4$. These estimates yield an estimated whale intensity of 0.06 whales per km^2 with a 95% parametric bootstrap confidence interval $[0.03, 0.08]$. Figure 12 (right plot) shows the fitted L-function; note the high variability of the non-parametric estimate of the L-function, cf. the envelopes computed from simulations of the fitted model. For this particular example, the computation of the profile likelihood function is very time consuming and Monte Carlo errors occasionally caused negative definite estimated observed information matrices. From a computational point of view, the Bayesian approach provides a more feasible alternative, see Example 7.4.

7.4 Bayesian inference

To compute posterior distributions for θ in a fully Bayesian approach to inference, we need to know the likelihood function for all values of θ. For a Gibbs point process, the computational problems due to the need for estimating the unknown normalizing constant are therefore even harder than for finding the MLE (Section 7.2) or the maximum a posteriori estimate (Heikkinen & Penttinen, 1999). Based on perfect simulation (Section 9.3) and auxiliary variable MCMC methods (Møller, Pettitt, Berthelsen & Reeves, 2006), progress on Bayesian inference for Markov point processes has been made in Berthelsen & Møller (2003, 2004, 2006a, 2006b).

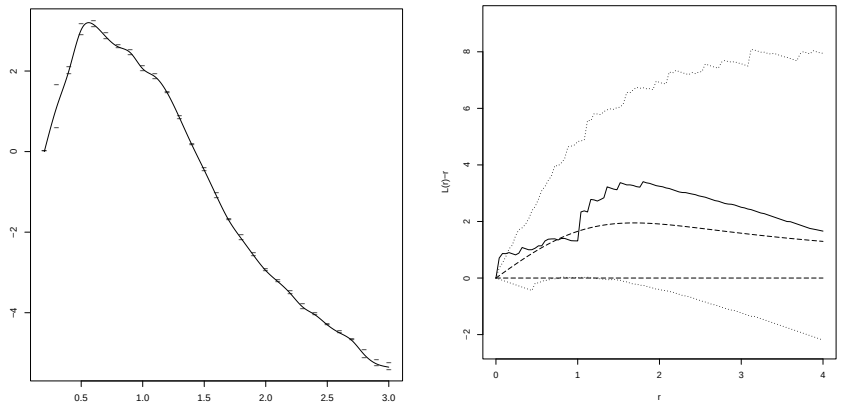


Figure 12: Fitting a shot noise Cox process model to the North Atlantic whales data set. Left: profile log likelihood function $l_p(\omega) = \max_{(\kappa,\alpha)} \log L(\theta)$ obtained by cumulating estimated log likelihood ratios, see text. The small horizontal bars indicate 95% Monte Carlo confidence intervals for the log likelihood ratios. Right: non-parametric estimate of $L(r)-r$ (solid line), 95% confidence envelopes based on simulations of fitted shot noise Cox process (dotted lines), $L(r)-r = 0$ for a Poisson process (lower dashed line), and $L(r) - r > 0$ for the fitted shot noise Cox process (upper dashed line).

In the examples below, we restrict attention to Cox processes for which the Bayesian approach is quite appealing from a computational point of view. The need for computing the likelihood function is eliminated by a demarginalization strategy, where the unknown random intensity function or cluster centre process is considered as an unknown parameter along with the original parameter θ. This simplifies computations, since the likelihood of the data given θ and the random intensity function is just a Poisson likelihood function. References on Bayesian inference for Cox processes include Wolpert & Ickstadt (1998), Best, Ickstadt & Wolpert (2000), Møller & Waagepetersen (2003a, 2003b), Beněs, Bodlák, Møller & Waagepetersen (2005), and Waagepetersen & Schweder (2005).

Example 7.4. *(Bayesian inference for North Atlantic whales)* In Waagepetersen & Schweder (2005), the unknown parameters κ, α, and ω (Examples 4.2 and 7.3) are assumed to be a priori independent with uniform priors on bounded intervals for κ and ω and an informative $N(2, 1)$ (truncated at zero) prior for α (the whales are a priori believed to appear in small groups of 1-3 animals). Posterior distributions are computed by extending an MCMC algorithm for

simulation of the cluster centres (see Section 9.2) with random walk MCMC updates for κ, α, and ω. The posterior means for κ, α, and ω are 0.027, 2.2, and 0.7, and the posterior mean of the whale intensity is identical to MLE. There is moreover close agreement between the 95% confidence interval (Example 7.3) and the 95% central posterior interval $[0.04, 0.08]$ for the whale intensity.

Example 7.5. *(Bayesian inference for tropical rain forest trees)* Considering the log Gaussian Cox process model for the tropical rain forest trees (Example 4.1), we assume that $\beta = (\beta_1, \beta_2, \beta_3)$, σ, and α are a priori independent, and use an improper uniform prior for β on $\mathbb{R}^3$, an improper uniform prior for σ on $[0.001, \infty)$, and a uniform prior for $\log \alpha$ with $1 \leq \alpha \leq 235$. For a discussion of posterior properity in similar models, see Christensen, Møller & Waagepetersen (2000). The Gaussian process is discretized to a 200×100 grid, and the posterior distribution of the discretized Gaussian process and the parameters is computed using MCMC with Langevin-Hastings updates for the Gaussian process (Section 7.3). The marginal posterior distributions of β, $\log \sigma$, and $\log \alpha$ are approximately normal. Posterior means and 95% central posterior intervals for the parameters of primary interest are 0.06 and $[0.02, 0.10]$ for β_2, 8.76 and $[6.03, 11.37]$ for β_3, 1.61 and $[1.44, 1.85]$ for σ, 42.5 and $[32.1, 56.45]$ for α. Figure 13 shows the posterior means of the systematic part $\beta_1 + \beta_2 z_2(u) + \beta_3 z_3(u)$ (left plot) and the random part $\Psi(u)$ (right plot) of the log random intensity function (8). The systematic part seems to depend more on z_3 (norm of altitude gradient) than z_2 (altitude), cf. Figure 3. The fluctuations of the random part may be caused by small scale clustering due to seed dispersal and covariates concerning soil properties.

Denote by $L(r; \mathbf{X}, \theta)$ the estimate of the L-function obtained from the point process $\mathbf{X}$ using (41) with $\hat{\rho}(u)$ replaced by the parametric intensity function $\rho_\theta(u) = \exp\left(z(u)\beta^{\mathsf{T}} + \sigma^2/2\right)$ for $\mathbf{X}$ given θ. Following the idea of posterior predictive model checking (Gelman *et al.*, 1996), we consider the posterior predictive distribution of the differences $\Delta(r) = L(r; \mathbf{x}, \theta) - L(r; \mathbf{X}, \theta)$, $r > 0$, i.e. the distribution obtained when $(\mathbf{X}, \theta)$ are generated under the posterior predictive distribution given the data $\mathbf{x}$. If zero is an extreme value in the posterior predictive distribution of $\Delta(r)$ for a range of distances r, we may question the fit of our model. Figure 14 shows 95% central envelopes obtained from posterior predictive simulations of $\Delta(r)$. The plot indicates that our model fails to accomodate clustering for small values of r less than 10 m.

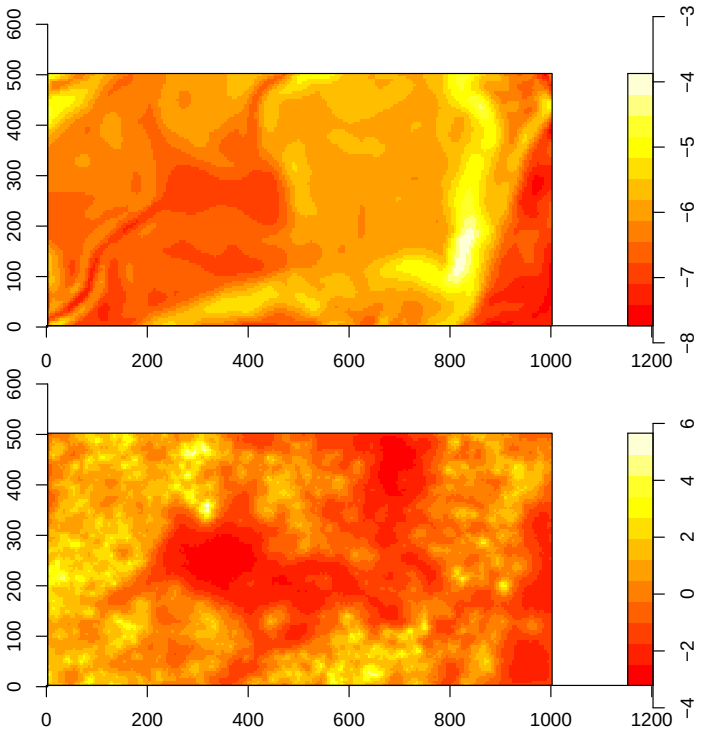


Figure 13: Posterior mean of $\beta_1 + \beta_2 z_2(u) + \beta_3 z_3(u)$ (left) and $\Psi(u)$ (right), $u \in W$, under the log Gaussian Cox process model for the tropical rain forest trees.

8 Simulation free estimation procedures

This section reviews quick non-likelihood approaches to inference using various estimating functions based on either first or second order properties of a spatial point process. In Section 8.1, estimating functions based on the (conditional) intensity function are motivated heuristically as limits of composite likelihood functions (Lindsay, 1988) for Bernouilli trials concerning absence or presence of points within infinitesimally small cells partitioning the observation window. Section 8.2 considers minimum contrast or composite log likelihood type estimating functions based on second order properties. In case of minimum contrast estimation, the parameter estimate minimizes the distance between a non-parametric estimate of a second order summary statistic and its theoretical expression. Yet another approach for obtaining estimating equations for spatial point process models is studied in Baddeley (2000).

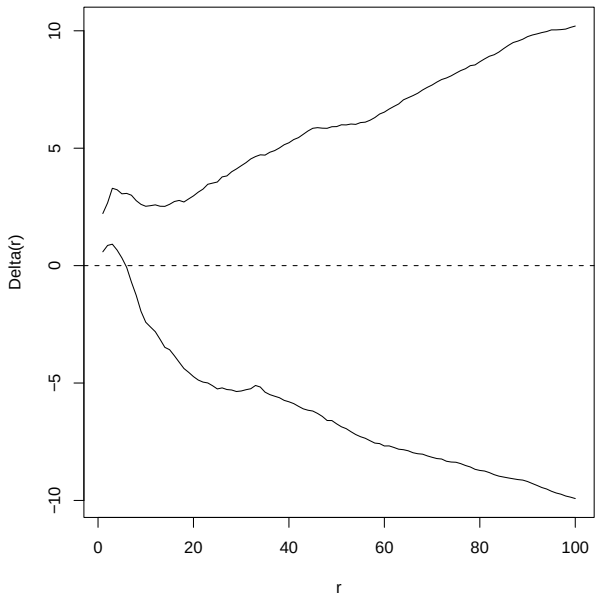


Figure 14: Tropical rain forest trees: 95% central envelopes obtained from posterior predictive simulations of $\Delta(r)$.

8.1 Estimating functions based on intensities

For a given parametric model with parameter θ, suppose that the intensity function ρ_θ is known on closed form. Consider a finite partitioning $C_i,\ i \in I$, of the observation window W into disjoint cells C_i of small areas $|C_i|$, and let u_i denote a representative point in C_i. Let $N_i = \mathbf{1}[N(C_i) > 0]$ and $p_i(\theta) = \mathrm{P}_\theta(N_i = 1)$. Then $p_i(\theta) \approx \rho_\theta(u_i)|C_i|$, and the composite likelihood based on the $N_i,\ i \in I$, is

$$\prod_{i \in I} p_i(\theta)^{N_i}(1 - p_i(\theta))^{(1-N_i)} \approx \prod_i (\rho_\theta(u_i)|C_i|)^{N_i}(1 - \rho_\theta(u_i)|C_i|)^{1-N_i}.$$

We neglect the factors $|C_i|$ in the first part of the product, since they cancel when we form likelihood ratios. In the limit, under suitable regularity conditions and when the cell sizes $|C_i|$ tend to zero, the log *composite likelihood* becomes

$$\sum_{u \in \mathbf{x}} \log \rho_\theta(u) - \int_W \rho_\theta(u)\,\mathrm{d}u$$

which coincides with the log likelihood function (42) in the case of a Poisson process. The corresponding estimating function is given by the derivative

$$\psi_1(\theta) = \sum_{u \in \mathbf{x}} \mathrm{d} \log \rho_\theta(u)/\mathrm{d}\theta - \int_W (\mathrm{d} \log \rho_\theta(u)/\mathrm{d}\theta) \rho_\theta(u)\, \mathrm{d}u. \tag{46}$$

By the Campbell theorem (4), $\psi_1(\theta) = 0$ is an unbiased estimating equation, and it can easily be solved using e.g. `spatstat`, provided ρ_θ is on a log linear form. For Cox processes, as exemplified in Example 8.1 below, the solution may only provide an estimate of one component of θ, while the other component may be estimated by another method.

For a Gibbs point process, it is more natural to consider the Papangelou conditional intensity λ_θ. Hence we redefine $p_i(\theta) = \mathrm{P}_\theta(N_i = 1 | \mathbf{X} \setminus C_i) \approx \lambda_\theta(u_i, \mathbf{X} \setminus C_i)|C_i|$. In this case the limit of $\log \prod_i (p_i(\theta)/|C_i|)^{N_i} (1 - p_i(\theta))^{(1-N_i)}$ becomes

$$\sum_{u \in \mathbf{x}} \lambda_\theta(u, \mathbf{x}) - \int_W \lambda_\theta(u, \mathbf{x})\, \mathrm{d}u$$

which is known as the log *pseudo likelihood function* (Besag, 1977; Jensen & Møller, 1991). By the GNZ formula (31), the *pseudo score*

$$s(\theta) = \sum_{u \in \mathbf{x}} \mathrm{d} \log \lambda_\theta(u, \mathbf{x})/\mathrm{d}\theta - \int_W (\mathrm{d} \log \lambda_\theta(u, \mathbf{x})/\mathrm{d}\theta) \lambda_\theta(u, \mathbf{x})\, \mathrm{d}u$$

provides an unbiased estimating equation $s(\theta) = 0$. This can be solved using `spatstat` if λ_θ is on a log linear form (Baddeley & Turner, 2000).

Example 8.1. *(Estimation of the intensity function for tropical rain forest trees)* For both the log Gaussian Cox process model in Example 4.1 and the inhomogeneous Thomas process model in Example 4.3, the intensity function is of the form $\exp(z(u)(\tilde{\beta}_1, \beta_2, \beta_3)^\mathsf{T})$, where $\tilde{\beta}_1 = \sigma^2/2 + \beta_1$ for the log Gaussian Cox process and $\tilde{\beta}_1 = \log(\kappa\alpha)$ for the inhomogeneous Thomas process. Using the estimating function (46) and `spatstat`, we obtain $(\hat{\tilde{\beta}}_1, \hat{\beta}_2, \hat{\beta}_3) = (-4.99, 0.02, 5.84)$, where $\hat{\beta}_2$ and $\hat{\beta}_3$ are smaller than the posterior means obtained with the Bayesian approach in Example 7.5. The estimate of course coincides with the MLE under the Poisson process with the same intensity function. Estimates of the clustering parameters, i.e. (σ^2, α) respectively (κ, ω), may be obtained using minimum contrast estimation, see Example 8.3.

Assuming $(\hat{\beta}_2, \hat{\beta}_3)$ is asymptotically normal (Waagepetersen, 2005), we obtain approximate 95% confidence intervals $[-0.02, 0.06]$ and $[0.89, 10.80]$ for β_2

and β_3, respectively. Under the Poisson process model much more narrow approximate 95% confidence intervals $[0.02, 0.03]$ and $[5.34, 6.34]$ are obtained.

Example 8.2. *(Maximum pseudo likelihood estimation for ants nests)* For the hierarchical model in Example 5.2, we first correct for edge effects by conditioning on the data in $W \setminus W_{\ominus 45}$. Using `spatstat`, the maximum pseudo likelihood estimate (MPLE) of (β_M, ψ_M) is $(-8.30, -0.44)$, indicating repulsion between the *Messor* ants nests. Without edge correction, a rather similar MPLE $(-8.48, -0.33)$ is obtained. The edge corrected MPLE of $(\beta_C, \psi_{CM}, \psi_C)$ is $(-26.19, 16.9, -0.43)$, indicating a positive association between the two species and repulsion within the *Cataglyphis* nests. As mentioned in Example 7.2, Högmander & Särkkä (1999) also found a repulsion within the *Cataglyphis* nests, but a weak repulsive interaction between the two types of nests. Baddeley & Turner (2006) modelled the *Messor* data conditional on the *Cataglyphis* data using an inhomogeneous Strauss hard core model and found that an apparent positive interspecies' interaction was not significant. Notice that this is a 'reverse' hierarchical model compared to our and Högmander & Särkkä's model.

The MPLE for *Cataglyphis* is very sensitive to whether edge correction is used or not (for our W, but not for the reduced observation window in Högmander & Särkkä, 1999). If no edge correction is used, the MPLE for $(\beta_C, \psi_{CM}, \psi_C)$ is $(-10.3, 0.89, 0.15)$. The large difference is due to that all *Cataglyphis* nests, which are not in the influence region of the *Messor* nests, are within the border region $W \setminus W_{\ominus 45}$, and two of these nest are moreover very close, cf. Figure 6. The differences between the MLE in Example 7.2 and the MPLE (without edge correction) seem rather minor. This is also the experience for MLE's and corresponding MPLE's in Møller & Waagepetersen (2003b), though differences may appear in cases with a very strong interaction.

8.2 Estimating functions based on the g or K-function

The pair correlation function g and the K-function in some sense describe the 'normalized' second order properties of a point process, cf. (5) and (40). For many Cox processes, g or K has a closed form expression depending on the 'clustering parameters' of the model. Examples include log Gaussian Cox processes (Section 4.2.1) and inhomogeneous Neyman-Scott processes with random intensity functions of the form (12) where k is a radially symmetric Gaussian density or a uniform density on a disc. Clustering parameter estimates may then be obtained using so-called *minimum contrast estimation*. That is, using

an estimating function given in terms of a distance between the theoretical expression for g or K and a non- or semi-parametric estimate $\hat{g}$ or $\hat{K}$, e.g. (41) where $\hat{\rho}$ could be a parametric estimate obtained from (46). This is illustrated in Example 8.3 for the K function. Minimum contrast estimation based on the g-function is considered in Møller *et al.* (1998). Asymptotic properties of minimum contrast estimates are derived in the case of stationary cluster processes in Heinrich (1992).

Alternatively, we may consider an estimating function based on the second order product density $\rho^{(2)}_\theta(u,v)$:

$$\psi_2(\theta) = \sum_{u,v\in\mathbf{x}}^{\neq} \mathrm{d}\log\rho^{(2)}_\theta(u,v)/\mathrm{d}\theta - \int_{W^2}\Big(\mathrm{d}\log\rho^{(2)}_\theta(u,v)/\mathrm{d}\theta\Big)\,\rho^{(2)}_\theta(u,v)\,\mathrm{d}u\mathrm{d}v. \tag{47}$$

This is the score of a limit of composite log likelihood functions based on Bernouilli observations $N_{ij} = \mathbf{1}[N(C_i) > 0, N(C_j) > 0]$, $i \neq j$. Unbiasedness of $\psi_2(\theta) = 0$ follows from Campbell's theorem (4). The integral in (47) typically must be evaluated using numerical integration. In the stationary case, Guan (2006) considers a related unbiased estimating function, where the integral in (47) is replaced by the number of pairs of distinct points times $\log\int_{W^2}\rho^{(2)}_\theta(u,v)\,\mathrm{d}u\mathrm{d}v$.

Example 8.3. *(Minimum contrast estimation of clustering parameters for tropical rain forest trees)* The solid curve in Figure 9 shows an estimate of the K-function for the tropical rain forest trees obtained using (41) with $\hat{\rho}$ given by the estimated parametric intensity function from Example 8.1. For the inhomogeneous Thomas process, a minimum contrast estimate $(\hat{\kappa},\hat{\omega}) = (8\times10^{-5}, 20)$ is obtained by minimizing

$$\int_0^{100}(\hat{K}(r)^{1/4} - K(r;\kappa,\omega)^{1/4})^2\mathrm{d}r \tag{48}$$

where

$$K(r;\kappa,\omega) = \pi r^2 + \big(1-\exp(-r^2/(4\omega)^2)\big)/\kappa$$

is the theoretical expression for the K-function. For the log Gaussian Cox process, we calculate instead the theoretical K-function

$$K(r;\sigma,\alpha) = 2\pi\int_0^r s\exp\big(\sigma^2\exp(-s/\alpha)\big)\,\mathrm{d}s$$

using numerical integration, and obtain the minimum contrast estimate $(\hat{\sigma}, \hat{\alpha}) = (1.33, 34.7)$. The estimated theoretical K-functions are shown in Figure 9.

Minimum contrast estimation is computationally very easy. A disadvantage is the need to choose certain tuning parameters like the upper limit 100 and the exponent 1/4 in the integral (48). Typically, these parameters are chosen on an ad hoc basis.

Example 8.4. *(Simultaneous estimation of parameters for tropical rain forest trees)* To estimate the parameters $(\tilde{\beta}_1, \beta_2, \beta_3)$ and (κ, ω) for the inhomogeneous Thomas process (see Example 8.1) simultaneously, we apply the estimating function ψ_2 (47). We solve $\psi_2(\theta) = 0$ using a grid search for ω combined with Newton-Raphson for the remaining parameters (Newton-Raphson for all the parameters jointly turns out to be numerically unstable). We then search for an approximate solution with respect to ω within a finite set of ω-values. The resulting estimates of $(\tilde{\beta}_1, \beta_2, \beta_3)$ and (κ, ω) are respectively $(-5.00, 0.02, 5.73)$ and $(7 \times 10^{-5}, 30)$. The estimate of ω differs considerably from the minimum contrast estimate in Example 8.3, while the remaining estimates are quite similar to those obtained previously for the inhomogeneous Thomas process in Examples 8.1 and 8.3. The numerical computation of ψ_2 and its derivatives is quite time consuming, and the whole process of solving $\psi_2(\theta) = 0$ takes about 75 minutes.

9 Simulation algorithms

As demonstrated several times, due to the complexity of spatial point process models, simulations are needed when fitting a model and studying the properties of various statistics such as parameter estimates and summary statistics. This section reviews the most applicable simulation algorithms.

9.1 Poisson and Cox processes

Even in the simple case of a Poisson point process, simulations are needed, see e.g. Figure 10. Simulation of a Poisson process within a bounded region is usually easy, using (i)–(ii) in Section 4.1 or other simple constructions (Section 3.2.3 in Møller & Waagepetersen, 2003b).

For simulation of a Cox process on a bounded region S, given a realization of the random intensity function $(\Lambda(u))_{u \in S}$, it is just a matter of simulating the Poisson process with intensity function $(\Lambda(u))_{u \in S}$. Details on how to simulate

$(\Lambda(u))_{u\in S}$ depend much on the particular type of Cox process model. For a log Gaussian Cox process, there are many ways of simulating the Gaussian process $(\log(\Lambda(u)))_{u\in S}$, see e.g. Schlather (1999) and Møller & Waagepetersen (2003b). For a shot noise Cox process, edge effects may occur since the Poisson process $\mathbf{\Phi}$ in (10) may be infinite, and so clusters associated to centre points outside S may generate points of the shot noise Cox process within S. Brix & Kendall (2002), Møller (2003) and Møller & Waagepetersen (2003b) discuss how to handle such edge effects.

9.2 Point processes specified by an unnormalized density

In this section, we consider simulation of a finite point process $\mathbf{X}$ with density $f \propto h$ with respect to the unit rate Poisson process defined on a bounded region S, where h is a 'known' unnormalized density. The normalizing constant of the density is not assumed to be known.

Simulation conditional on the number of points $n(\mathbf{X})$ can be done using a variety of Metropolis-Hastings algorithms, since the conditional process is just a vector of fixed dimension when we order the points as in the density (17). Most algorithms used in practice are a Gibbs sampler or a Metropolis-within-Gibbs algorithm, where at each iteration a single point given the remaining points is updated, see Møller & Waagepetersen (2003b, Section 7.1.1).

The standard algorithms (i.e. without conditioning on $n(\mathbf{X})$) are discrete or continuous time algorithms of the birth-death type, where each transition is either the addition of a new point (a birth) or the deletion of an existing point (a death). The algorithms can easily be extended to birth-death-move type algorithms, where e.g. in the discrete time case the number of points is retained in a move by using a Metropolis-Hastings update as discussed in the previous paragraph, see Møller & Waagepetersen (2003b, Section 7.1.2).

For instance, in the discrete time case, a simple *Metropolis-Hastings algorithm* updates a current state $\mathbf{X}_t = \mathbf{x}$ of the Markov chain as follows (Norman & Filinov, 1969; Geyer & Møller, 1994). Assume that h is hereditary, and define $r(u,\mathbf{x}) = \lambda(u,\mathbf{x})|S|/(n(\mathbf{x})+1)$ where, as usual, λ is the Papangelou conditional intensity. With probability 0.5 propose a birth, i.e. generate a uniform point u in S, and accept the proposal $\mathbf{X}_{t+1} = \mathbf{x}\cup\{u\}$ with probability $\min\{1, r(u,\mathbf{x})\}$. Otherwise propose a death, i.e. select a point $u \in \mathbf{x}$ uniformly at random, and accept the proposal $\mathbf{X}_{t+1} = \mathbf{x}\setminus\{u\}$ with probability $\min\{1, 1/r(u,\mathbf{x}\setminus\{u\})\}$. As usual in a Metropolis-Hastings algorithm, if the proposal is not accepted,

$\mathbf{X}_{t+1} = \mathbf{x}$.

This algorithm (like many other Metropolis-Hastings algorithms studied in Chapter 7 in Møller & Waagepetersen, 2003b) is irreducible and aperiodic with invariant distribution f; in fact it is time reversible with respect to f. In other words, the distribution of $\mathbf{X}_t$ converges towards f. Moreover, if h is locally stable, the rate of convergence is geometrically fast, and a central limit theorem holds for Monte Carlo errors (Geyer & Møller, 1994; Geyer, 1999).

An analogous continuous time algorithm is based on running a *spatial birth-death process* $\mathbf{X}_t$ with birth rate $\lambda(u, \mathbf{x})$ and death rate 1. This is also a reversible process with invariant density f (Preston, 1977; Ripley, 1977). Convergence of $\mathbf{X}_t$ towards f holds under weak conditions, and local stability of h implies geometrically fast convergence (Møller, 1989).

If h is highly multimodal, e.g. in the case of a strong interaction like in a hard core model with a high packing density, the birth-death (or birth-death-move) algorithms described above may be slowly mixing. The algorithms may then be incorporated into a simulated tempering scheme (Geyer & Thompson, 1995; Mase, Møller, Stoyan, Waagepetersen & Döge, 2001).

9.3 Perfect simulation

One of the most exciting recent developments in stochastic simulation is *perfect (or exact) simulation*, which turns out to be particular applicable for locally stable point processes (Kendall, 1998; Kendall & Møller, 2000). By this we mean an algorithm where the running time is a finite random variable and the output is a draw from a given target distribution (at least in theory — of course the use of pseudo random number generators and practical constraints of time imply that we cannot exactly return draws from the target distribution).

The most famous perfect simulation algorithm is due to Propp & Wilson (1996). It is based on a coupling construction called *coupling from the past (CFTP)*, which exploits the fact that any Markov chain algorithm, at least when it is implemented on a computer, can be viewed as a so-called stochastic recursive sequence $X_{t+1} = \phi(X_t, R_t)$, where ϕ is a deterministic function and the R_t are i.i.d. random variables. The updating function ϕ is supposed to be monotone with respect to some partial order $\prec$, that is, $x \prec y$ implies $\phi(x, r) \prec \phi(y, r)$. Further, it is assumed that there exist unique minimal and maximal states $\hat{0}$ and $\hat{1}$, so $\hat{0} \prec x \prec \hat{1}$ for any state x. The coupling construction is based on pairs of upper and lower dominating chains generated for $n = 1, 2, \ldots$

by $U_{t+1}^n = \phi(U_t^n, R_t)$ and $L_{t+1}^n = \phi(L_t^n, R_t)$, $t = T_n, T_n + 1, \ldots, -1$, where $U_{T_n}^n = \hat{1}$ and $L_{T_n}^n = \hat{0}$, and the starting times $T_n < 0$ decrease to $-\infty$ for $n = 1, 2, \ldots$. Note that the R_t are re-used for all $n = 1, 2, \ldots$. By monotonicity and the coupling construction, if $X_{T_n} = x$ for an arbitrary state x, we have the sandwiching property $L_t^n \prec X_t \prec U_t^n$ and the funneling property $L_t^n \prec L_t^{n+1} \prec U_t^{n+1} \prec U_t^n$, $t = T_n, \ldots, 0$. Moreover, if $L_s^n = U_s^n$ then $L_t^n = U_t^n$ for $s \leq t \leq 0$. Consequently, if the Markov chain is ergodic and with probability one, $L_0^n = U_0^n$ for some sufficiently large n, then we need only to generate the pairs of upper and lower chains (U^n, L^n) until we have coalescence at time 0, since $L_0^n = U_0^n$ will follow the equilibrium distribution of the chain.

The Propp-Wilson algorithm applies only for a few spatial point process models (Häggström, Van Lieshout & Møller, 1999; Møller & Waagepetersen, 2003, Chapter 11). For the natural partial ordering given by set inclusion, the empty point configuration is the unique minimal state, but there is no maximal element. This problem is solved by a modification of the Propp-Wilson algorithm, called *dominating CFTP* (Kendall & Møller, 2000), where the coupling construction is a dependent thinning from a dominating spatial birth-death process which is easy to simulate. The algorithm does not assume monotonicity, and it applies to perfect simulation for locally stable point processes. For instance, a spatial birth-death algorithm for a repulsive Gibbs point process is anti-monotone, but this problem can be fixed by a certain cross-over trick due to Kendall (1998). For an introduction to the dominated CFTP algorithm, including empirical findings and a discussion on how to choose the sequence of starting times T_n, see Berthelsen & Møller (2002) and Møller & Waagepetersen (2003, Chapter 11).

10 Directions for future research

10.1 Spatial point pattern data sets

In this paper, we have for illustrative purposes and to limit space considered relatively simple examples of data sets, where we could compare our results with results published elsewhere. We have also not discussed more complicated models involving, for example, both thinnings, movements, and superpositioning of points (Lund & Rudemo, 2000) or spatial point processes generating geometric structures such as Voronoi tessellations (Baddeley & Møller, 1989; Blackwell & Møller, 2003; Skare, Møller & Jensen, 2006).

Many scientific problems call for new spatial point process methodology for analyzing complex and large data sets (often with marks and possibly in time-space). For example, in the tropical rain forest example, the data for the *Beilschmiedia* trees is just a very small part of a very large data set containing positions and diameters at breast height for thousands of species recorded over several instances in time. Another example is the Sloan Digital Sky Survey with millions of galaxies, where it is of interest to model the clustering of galaxies.

10.2 Inhomogeneity

We have pointed out that stationarity is often not a reasonable assumption, and the focus on summary statistics based on the stationarity assumption (such as F, G, J) is therefore often out of place. In our experience, it is often sufficient to consider K or g, which also are well-defined in non-stationary situations. Residual analysis (Baddeley *et al.*, 2005) do not require stationarity, and using the GNZ-formula (31) or (32), it can be meaningful to use functionals related to F, G, J in inhomogeneous cases (Baddeley *et al.*, 2006).

Often we need to specify an inhomogeneous point process model. For Poisson and Cox processes, log linear modelling as in Examples 4.1 and 4.3 can be a useful approach. For Gibbs point processes, as exemplified by the model for the *Cataglyphis* ant's nests in Example 5.2, a simple way of modelling inhomogeneity is to introduce a non-constant first order potential (Ogata & Tanemura, 1986; Stoyan & Stoyan, 1998; Berthelsen & Møller, 2006b). Another possibility is to consider an independent thinning of a stationary Gibbs point process, in which case we have second order intensity reweighted stationarity (Baddeley *et al.*, 2000). Yet other constructions, using transformations of homogeneous Markov point processes and location dependent scaling, are studied in Jensen & Nielsen (2000), Hahn, Jensen, Van Lieshout & Nielsen (2003), and Nielsen & Jensen (2004). It is an open problem to extend asymptotic results for MLE (and to some extend also MPLE) to non-stationary situations.

Often, when inhomogeneity is modelled in terms of covariates observed on a grid, we face a missing data problem, since the likelihood function depends on the covariate at any location in the observation window. In the tropical rain forest example, we assumed constant values of the covariates within grid cells, but this may not be appropriate when the covariates are observed on coarser grids.

10.3 Gibbs and other point processes

Markov or more generally Gibbs point processes originated naturally in statistical physics as models for the study of phase transition behaviour and other physical phenomena. We question their popularity in spatial statistics for the following reasons.

Markov point processes provide a flexible framework for modelling repulsive spatial interaction as exemplified by the overlap model for the Norwegian spruces. However, despite that the modelling using the influence zones is based on biological reasoning, one may object that the model fails to reflect that the observed spatial pattern is the result of an ongoing dynamic development of the forest. In fact, for this and many other application areas, we do not believe that spatial point patterns can be viewed as the equilibrium state of a reversible spatial birth-death process (Section 9.2).

The estimation of the interaction range R is a tricky issue which seems to require computation of a profile (pseudo) likelihood over a finite grid of R values. It is not clear how to obtain e.g. confidence intervals for this parameter. A parametric bootstrap is a computationally involved possibility although more research on the usefulness of this approach seems required. The same can be said about Bayesian inference with a prior imposed on R (Berthelsen & Møller, 2003, 2004, 2006a, 2006b).

There is a lack of satisfactory Gibbs point processes modelling attractive spatial interaction. As discussed in Example 5.2, Strauss point processes with a hard core are not so flexible due to a kind of phase transition behaviour. The Widom-Rowlinson model or area-interaction point process, which is another well known Gibbs point process with attractive spatial interaction (Widom & Rowlinson, 1970; Baddeley & Van Lieshout, 1995), may be inflexible as well because of a somewhat similar phase transition behaviour (Häggström *et al.*, 1999). The saturation and triplet point processes in Geyer (1999) are other examples of models for attraction between points, but in applications the interpretation of these models is not clear. In contrast, we find Cox process models more natural, and more generally we find hierarchical model constructions relevant, cf. Example 5.2 and Illian, Møller & Waagepetersen (2006).

As mentioned several times, closed form expressions for spatial point process densities are rare. Shirai & Takahashi (2003) and McCullagh & Møller (2005) study a large model class of non-Poisson point process models, called the permanent and determinant processes in McCullagh & Møller (2005), where both

the density of the process and the product densities are of an analytic form. The processes possess many appealing properties, where the permanent process models aggregation of points (in some cases it is a Cox process), while the determinant process models repulsion. It remains to investigate the processes thoroughly in connection to statistical inference.

10.4 Computational issues

Some of the earliest applications of computational methods and particularly MCMC methods in statistics are related to spatial point processes (Møller & Waagepetersen, 2003b, Section A.1). As discussed in Section 7, maximum likelihood inference is now quite feasible for Markov point processes. For Cox processes, likelihood-based inference is computationally more involved. The computing times can be discouraging and research to obtain more efficient Monte Carlo methods for Cox processes seems needed. For log Gaussian Cox processes, promising results in Rue & Martino (2005) suggest that it may be possible to compute accurate approximations of posterior distributions without MCMC.

In the future, we expect simulation-based methods to play an increasingly important role for spatial point process modelling and inference, though quick explorative tools and simulation free estimation procedures will still be useful. Due to the fast increase in computer power, we expect e.g. the development and use of perfect simulation techniques (Section 9.3) to become of much more practical relevance, especially in connection to Bayesian inference (Berthelsen Møller, 2003, 2004, 2006a, 2006b). Also the techniques in Møller & Mengersen (2006) for calculating ergodic averages, eliminating the problem of finding the burn-in by using upper and lower dominating processes but without the need of doing perfect simulation, should be studied in connection to spatial point process models. Simulation-based methods for the permanent and determinant processes mentioned in Section 10.3 have yet not been investigated. In this connection, an open problem is to develop and implement efficient algorithms for calculation of cyclic products and permanent polynomials.

In order to make statisticians familiar with spatial point process modelling and inference, there is an obvious need for userfriendly software. We therefore much appreciate the development of `spatstat` which offers a wide range of procedures for manipulation of point pattern data, residuals, summary statistics, and maximum pseudo likelihood estimation (Baddeley & Turner, 2005, 2006); see also the other references to software in Møller & Waagepetersen (2003b,

Appendix A.3).

10.5 Spatio-temporal point processes

Due to space constraints, apart from Sections 9 and 10.3, we have not considered time-space point processes. One challenge is to develop tractable and yet interesting continuous-time models, when spatial point process data is only available at discrete times, see e.g. the spatio-temporal extensions of log Gaussian Cox processes studied in Brix & Møller (2001) and Brix & Diggle (2001). For time-space point processes in general, we refer to Daley & Vere-Jones (2003), Møller & Waagepetersen (2003b, Section 2.4), Diggle (2005), and the references therein.

10.6 Conclusion

In conclusion, spatial point processes and their application areas have undergone major developments in recent years, and we expect they will continue to do so, as statisticians and scientists become aware of their importance and the tools for performing statistical analyses, and not at least, the challenges of developing new tools.

Acknowledgements We are grateful to Adrian Baddeley for assistance with `spatstat`.

References

Baddeley, A. & Møller, J. (1989). Nearest-neighbour Markov point processes and random sets. *International Statistical Review* **2**, 89–121.

Baddeley, A. & Silverman, B. W. (1984). A cautionary example for the use of second-order methods for analysing point patterns. *Biometrics* **40**, 1089–1094.

Baddeley, A. & Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics* **42**, 283–322.

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

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

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

Baddeley, A., Turner, R., Møller, J. & Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of Royal Statistical Society Series B* **67**, 617–666.

Baddeley, A., Gregori, P., Mateu, J., Stoica, R. & Stoyan, D., eds. (2006a). *Case Studies in Spatial Point Process Modeling*, Springer Lecture Notes in Statistics 185, Springer-Verlag, New York.

Baddeley, A., Møller, J. & Pakes, A. G. (2006b). Properties of residuals for spatial point processes. *Technical Report R-2006-03*, Department of Mathematical Sciences, Aalborg University.

Baddeley, A., Møller, J. & Waagepetersen, R. (2006c). In preparation.

Baddeley, A. J. (2000). Time-invariance estimating equations. *Bernoulli* **6**, 783–808.

Baddeley, A. J. & van Lieshout, M. N. M. (1995). Area-interaction point processes. *Annals of the Institute of Statistical Mathematics* **46**, 601–619.

Ballani, F. (2006). On modelling of refractory castables by marked Gibbs and Gibbsian-like processes. In: *Case Studies in Spatial Point Process Modeling* (eds. A. Baddeley, P. Gregori, J. Mateu, R. Stoica and D. Stoyan), Springer Lecture Notes in Statistics 185, Springer-Verlag, New York, 153–167.

Bartlett, M. S. (1964). The spectral analysis of two-dimensional point processes. *Biometrika* **51**, 299–311.

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

Berman, M. & Turner, R. (1992). Approximating point process likelihoods with GLIM. *Applied Statistics* **41**, 31–38.

Berthelsen, K. K. & Møller, J. (2002). A primer on perfect simulation for spatial point processes. *Bulletin of the Brazilian Mathematical Society* **33**, 351–367.

Berthelsen, K. K. & Møller, J. (2003). 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.

Berthelsen, K. K. & Møller, J. (2004). An efficient MCMC method for Bayesian point process models with intractable normalising constants. In: *Spatial point process modelling and its applications* (eds. A. Baddeley, P. Gregori, J. Mateu, R. Stoica and D. Stoyan), Publicacions de la Universitat Jaume I.

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

Berthelsen, K. K. & Møller, J. (2006b). Non-parametric Bayesian inference for inhomogeneous Markov point processes. In preparation.

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

Besag, J. E. (1977b). Discussion of the paper by Ripley (1977). *Journal of the Royal Statistical Society Series B* **39**, 193–195.

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

Blackwell, P. G. & Møller, J. (2003). Bayesian analysis of deformed tessellation models. *Advances in Applied Probability* **35**, 4–26.

Brix, A. & Chadoeuf, J. (2002). Spatio-temporal modeling of weeds and shot-noise G Cox processes. *Biometrical Journal* **44**, 83–99.

Brix, A. & Diggle, P. J. (2001). Spatio-temporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society Series B* **63**, 823–841.

Brix, A. & Kendall, W. S. (2002). Simulation of cluster point processes without edge effects. *Advances in Applied Probability* **34**, 267–280.

Brix, A. & Møller, J. (2001). Space-time multitype log Gaussian Cox processes with a view to modelling weed data. *Scandinavian Journal of Statistics* **28**, 471–488.

Christensen, O. F., Møller, J. & Waagepetersen, R. P. (2000). Analysis of spatial data using generalized linear mixed models and Langevin-type Markov chain Monte Carlo. *Technical Report R-00-2009*, Department of Mathematical Sciences, Aalborg University.

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

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

Cox, D. R. (1955). Some statistical models related with series of events. *Journal of the Royal Statistical Society Series B* **17**, 129–164.

Daley, D. J. & Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Springer-Verlag, New York, 2nd edition.

Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*. Academic Press, London.

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

Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*. Arnold, London, 2nd edition.

Diggle, P. J. (2005). Spatio-temporal point processes: Methods and applications. *Working Paper 78*, Department of Biostatistics, Johns Hopkins University.

Fiksel, T. (1984). Estimation of parameterized pair potentials of marked and nonmarked Gibbsian point processes. *Elektronische Informationsverarbeitung und Kypernetik* **20**, 270–278.

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

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

Georgii, H.-O. (1976). Canonical and grand canonical Gibbs states for continuum systems. *Communications of Mathematical Physics* **48**, 31–51.

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

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

Geyer, C. J. & Thompson, E. A. (1995). Annealing Markov chain Monte Carlo with applications to pedigree analysis. *Journal of the American Statistical Association* **90**, 909–920.

Goulard, M., Särkkä, A. & Grabarnik, P. (1996). Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method. *Scandinavian Journal of Statistics* **23**, 365–379.

Grandell, J. (1976). *Doubly Stochastic Poisson Processes*. Springer Lecture Notes in Mathematics 529. Springer-Verlag, Berlin.

Grandell, J. (1997). *Mixed Poisson Processes*. Chapman and Hall, London.

Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of American Statistical Association.* To appear.

Häggström, O., van Lieshout, M. N. M. & Møller, J. (1999). Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli* **5**, 641–659.

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

Harkness, R. D. & Isham, V. (1983). A bivariate spatial point pattern of ants' nests. *Applied Statistics* **32**, 293–303.

Heikkinen, J. & Penttinen, A. (1999). Bayesian smoothing in the estimation of the pair potential function of Gibbs point processes. *Bernoulli* **5**, 1119–1136.

Heinrich, L. (1992). 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*, Academic Publishing House, Prague, 479–492.

Högmander, H. & Särkkä, A. (1999). Multitype spatial point patterns with hierarchical interactions. *Biometrics* **55**, 1051–1058.

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

Illian, J. B., Møller, J. & Waagepetersen, R. P. (2006). Spatial point process analysis for a plant community with high biodiversity. *Technical Report R-2006-05*, Department of Mathematical Sciences, Aalborg University.

Jensen, A. T. (2005). *Statistical Inference for Doubly Stochastic Poisson Processes*. PhD thesis, Department of Applied Mathematics and Statistics, University of Copenhagen. Available at http://www.staff.kvl.dk/˜tolver/publications/Tiltryk.pdf.

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

Jensen, J. L. & Møller, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Annals of Applied Probability* **3**, 445–461.

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

Kendall, W. S. & Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability* **32**, 844–865.

Kerscher, M. (2000). Statistical analysis of large-scale structure in the Universe. In: *Statistical Physics and Spatial Statistics* (eds. K. R. Mecke and D. Stoyan), Lecture Notes in Physics, Springer, Berlin, 36–71.

Kingman, J. F. C. (1993). *Poisson Processes*. Clarendon Press, Oxford.

Lieshout, M. N. M. van (2000). *Markov Point Processes and Their Applications*. Imperial College Press, London.

Lieshout, M. N. M. van & Baddeley, A. J. (1996). A nonparametric measure of spatial interaction in point patterns. *Statistica Neerlandica* **50**, 344–361.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics* **80**, 221–239.

Lund, J. & Rudemo, M. (2000). Models for point processes observed with noise. *Biometrika* **87**, 235–249.

Macchi, O. (1975). The coincidence approach to stochastic point processes. *Advances in Applied Probability* **7**, 83–122.

Mase, S., Møller, J., Stoyan, D., Waagepetersen, R. P. & Döge, G. (2001). Packing densities and simulated tempering for hard core Gibbs point processes. *Annals of the Institute of Statistical Mathematics* **53**, 661–680.

McCullagh, P. & Møller, J. (2005). The permanent process. *Technical Report R-2005-29*, Department of Mathematical Sciences, Aalborg University.

Møller, J. (1989). On the rate of convergence of spatial birth-and-death processes. *Annals of the Institute of Statistical Mathematics* **3**, 565–581.

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

Møller, J. (2003). Shot noise Cox processes. *Advances in Applied Probability* **35**, 4–26.

Møller, J. & Mengersen, K. (2006). Ergodic averages for monotone functions using upper and lower dominating processes. *Technical Report R-2006-01*, Department of Mathematical Sciences, Aalborg University.

Møller, J. & Torrisi, G. L. (2005). Generalised shot noise Cox processes. *Advances in Applied Probability* **37**, 48–74.

Møller, J. & Waagepetersen, R. P. (2003a). An introduction to simulation-based inference for spatial point processes. In: *Spatial Statistics and Computational Methods* (ed. J. Møller), Springer Lecture Notes in Statistics 173, Springer-Verlag, New York, 143–198.

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

Møller, J., Syversveen, A. R. & Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics* **25**, 451–482.

Møller, J., Pettitt, A. N., Berthelsen, K. K. & Reeves, R. W. (2006). An efficient MCMC method for distributions with intractable normalising constants. *Biometrika* **93**, to appear.

Neyman, J. & Scott, E. L. (1958). Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society Series B* **20**, 1–43.

Nguyen, X. X. & Zessin, H. (1979). Integral and differential characterizations of Gibbs processes. *Mathematische Nachrichten* **88**, 105–115.

Nielsen, L. S. & Jensen, E. B. V. (2004). Statistical inference for transformation inhomogeneous Markov point processes. *Scandinavian Journal of Statistics* **31**, 131–142.

Norman, G. E. & Filinov, V. S. (1969). Investigations of phase transition by a Monte-Carlo method. *High Temperature* **7**, 216–222.

Ogata, Y. & Tanemura, M. (1986). Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In: *Proceedings of the Pacific Statistical Congress* (eds. I. S. Francis, B. F. J. Manly and F. C. Lam), Elsevier, Amsterdam, 150–154.

Papangelou, F. (1974). The conditional intensity of general point processes and an application to line processes. *Zeitschrift für Wahscheinlichkeitstheorie und werwandte Gebiete* **28**, 207–226.

Penttinen, A., Stoyan, D. & Henttonen, H. M. (1992). Marked point processes in forest statistics. *Forest Science* **38**, 806–824.

Preston, C. (1976). *Random Fields*. Lecture Notes in Mathematics 534. Springer-Verlag, Berlin.

Propp, J. G. & Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9**, 223–252.

Rathbun, S. L. (1996). Estimation of Poisson intensity using partially observed concomitant variables. *Biometrics* **52**, 226–242.

Rathbun, S. L. & Cressie, N. (1994). Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson processes. *Advances in Applied Probability* **26**, 122–154.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability* **13**, 255–266.

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

Ripley, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press, Cambridge.

Ripley, B. D. & Kelly, F. P. (1977). Markov point processes. *Journal of the London Mathematical Society* **15**, 188–192.

Rue, H. & Martino, S. (2005). Approximate inference for hierarchical Gaussian Markov random fields models. *Statistics Preprint 7/2005*, Norwegian University of Science and Technology.

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

Schladitz, K. & Baddeley, A. J. (2000). A third-order point process characteristic. *Scandinavian Journal of Statistics* **27**, 657–671.

Schlather, M. (1999). Introduction to positive definite functions and unconditional simulation of random fields. *Technical Report ST 99-10*, Lancaster University.

Schlather, M. (2001). On the second-order characteristics of marked point processes. *Bernoulli* **7**, 99–117.

Shirai, T. & Takahashi, Y. (2003). Random point fields associated with certain Fredholm determinants I: fermion, Poisson and boson point processes. *Journal of Functional Analysis* **205**, 414–463.

Skare, Ø., Møller, J. & Jensen, E. B. V. (2006). Bayesian analysis of spatial point processes in the neighbourhood of voronoi networks. *Technical Report R-2006-02*, Department of Mathematical Sciences, Aalborg University.

Skaug, H. J., Øien, N., Schweder, T. & Bøthun, G. (2004). Abundance of minke whales (*balaneoptera acutorostrata*) in the northeast Atlantic: variability in time and space. *Canadian Journal of Fisheries and Aquatic Sciences* **61**, 870–886.

Stoyan, D. & Stoyan, H. (1995). *Fractals, Random Shapes and Point Fields*. Wiley, Chichester.

Stoyan, D. & Stoyan, H. (1998). Non-homogeneous Gibbs process models for forestry — a case study. *Biometrical Journal* **40**, 521–531.

Stoyan, D. & Stoyan, H. (2000). Improving ratio estimators of second order point process characteristics. *Scandinavian Journal of Statistics* **27**, 641–656.

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

Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika* **36**, 18–25.

Waagepetersen, R. (2005a). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. Submitted for publication.

Waagepetersen, R. & Schweder, T. (2005). Likelihood-based inference for clustered line transect data. Submitted for publication.

Waagepetersen, R. P. (2005b). Discussion of the paper by Baddeley, Turner, Møller & Hazelton (2005). *Journal of the Royal Statistical Society Series B* **67**, 662.

Widom, B. & Rowlinson, J. S. (1970). A new model for the study of liquid-vapor phase transitions. *Journal of Chemical Physics* **52**, 1670–1684.

Wolpert, R. L. & Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251–267.